Investigation of the protein and RNA interactions of a family of small CCCH zinc finger proteins involved in the life cycle development of *Trypanosoma brucei.*

Paul Craddy

Submitted for the degree of PhD
The University of Edinburgh
January 2007

Abstract

African trypanosomes (*Trypanosoma brucei spp.*) are the causative agents of Human African Trypanosomiasis and Nagana in cattle and livestock. The complex lifecycle of *T. brucei* involves differentiation from the mammalian host to the tsetse fly vector, a process that involves highly regulated changes in gene expression, morphology and surface antigen expression. Three members of the *T. brucei* Zinc Finger Protein (*Tb*ZFP1, 2 and 3) family have been characterised, which contain the unusual $C_{x8}C_{x5}C_{x3}H$ RNA binding zinc finger motif, as well as a PY or WW protein-protein interacting motif. The *Tb*ZFP family are important regulators of differentiation and all three molecules have been shown *in vivo* to be essential for effective differentiation, although no proteins or RNA transcripts involved in these mechanisms have as yet been identified.

Here, data obtained from screens for proteins and RNA transcripts that interact with the *Tb*ZFP family is presented. Using the yeast two-hybrid system, *Tb*ZFP1 was shown to interact directly with *Tb*ZFP2 and 3, this interaction being dependant on the WW domain. Furthermore, a yeast two-hybrid screen to identify novel protein interactions using *Tb*ZFP2 and 3 was undertaken, though no genuine interactions were identified. In addition, a systematic evolution of ligands by an exponential enrichment (SELEX) approach was used to identify the sequence UCAGU/C as a putative RNA binding motif for *Tb*ZFP3 and the ability of *Tb*ZFP3 to bind RNA was verified using an *in vitro* electrophoretic mobility shift assay. Finally, a screen was undertaken for RNA transcripts whose abundance was affected by the ablation of *Tb*ZFP2.

Declaration

The text of this thesis is my own work. The work presented is substantially my own, but would not have been possible without the help of others:

1. The creation of the *T. brucei* yeast two-hybrid Matchmaker library, a kind gift from Professor George Cross (Rockefeller University, USA).

2. The creation of the *T. brucei* yeast two-hybrid HybriZAP-2.1 GAL4 library, a kind gift from Professor Isobel Roditi, (University of Bern, Switzerland).

3. The manufacture of the *T. brucei* genomic macroarray, created by Professor Elisabetta Ullu (Yale University, USA).

4. The creation of the ZFP2.RNAi trypanosome cell line, a kind gift from Dr Ed Hendriks.

5. The creation of the recombinant *Tb*ZFP3 bacterial cell line, a kind gift from Dr Athina Paterou.

This work has not been submitted for any other degree or professional qualification.

Paul Craddy – January 2007

Acknowledgements

I am exceedingly grateful to my supervisor Keith Matthews for all his help and support, both financially and intellectually, throughout my time in the lab. I would also like to thank all members of the Matthews lab, both past and present. Specifically, I would like to thank Ricky and Ed for all the training and amusement they provided when I was new in the lab; Katelyn, for all her help with the EMSAs and computational work and Suzanne, for all the hard work and time spend working on the SELEX project. Debs and Pam deserve a special mention for all the sterling work and media preparation that they undertook and to ensure that every item was there when it was meant to be. I am also very grateful to both Matthew and Helen for all years we have spent together and all their support and help in getting through our respective PhDs. It was invaluable! I am grateful to the BBSRC for funding to allow me to undertake this project and to the University of Edinburgh for funding for the last two months.

I would like to thank everyone who has made my time in Edinburgh and Manchester so rewarding. Huge thanks goes to all the members of EUSAC, for all the great diving weekends away and providing such a great distraction to my PhD. I would like to single out Dave, Graham and Max for the numerous breaks, support, moans and discussions during our tea times. Thank you for keeping me sane when I needed you most! Many thanks must go to Paul B and Graeme, who were always up for a pint when I needed it most and to all the usual suspects for all the great times we have had. Additionally, thanks must go to the ex-MUSC members and all those

SWCC members for all those excellent caving weekends and copious amounts of ale. Also thanks must go to my close friends, Paul, Sara, Bernard, Jon, and Nathan for all their support over the last three years.

I have to thank my parents, Mark and Claire, for all their love and support throughout the last few years and believing in me and my sister, Angela. A special thanks goes to Peter for his guidance and wisdom in the early stages.

Finally, Anne-Marie, thank you for all your love, patience, companionship and for keeping a smile on my face though it all!

Abbreviations

aa	Amino acids
AD	Activation domain
ARE	AU rich element
ATP	Adenosine triposhphate
BSF, BS	Bloodstream from
DNA	Deoxyribonucleic acid
DNA BD	DNA binding domain
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electophoretic mobility shift assay
EP	Glutamic acid-proline procyclin
EtBr	Ethidium bromide
GPEET	Glycine-proline-glutamic acid-glutamic acidthreonine procyclin
HRP	Horseradish Peroxidase
kDNA	Kinetoplast DNA
LB	Luria broth
MOPS	3-(N-Morpholino)propanesulfonic acid
mRNA	Messenger RNA
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide phosphate
nt	nucleotide
ORF	Open reading frame
PBS	Phosphate buffered saline
PCF, PC	Procyclic form
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RNA	Ribonucleic acid
RNAi	RNA interference
rRNA	Ribosomal RNA
RSAT	Regulatory sequence analysis tools
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-Polyacrylmide gel
SELEX	Systematic evolution of ligands by expoential enrichment
SL	Splice leader
SMB	Single marker bloodstream
TAE	Tris-Acetate EDTA
TE	Tris-EDTA
UTR	Untranslated region
VSG	Variable surface glycoprotein coat
WT	Wild type
ZPFM	Zimmerman postfusion buffer
ZPFMG	Zimmerman postfusion buffer with glucose

TABLE OF CONTENTS

CHAPTER 2: MATERIALS AND METHODS.

Chapter 1:

Introduction.

1.1 Introduction.

The African trypanosome, *Trypanosoma brucei* spp., is a unicellular protozoan parasite that is the causative agent of African sleeping sickness (trypanosomiasis) in humans and Nagana in livestock. There are two subspecies of *T. brucei* which cause disease in humans: *T. brucei rhodesiense*, the cause of East African sleeping sickness and *T. brucei gambiense* which causes chronic West African trypanosomiasis (Figure 1.1). Both species are morphologically indistinguishable, but have different clinical presentations and epidemiology. Transmission of the *T. brucei* parasite is by the tsetse fly (*Glossina* spp.), either from an infected human or animal reservoir. Approximately 60 million people in 36 countries are at risk of developing trypanosomiasis, with 100,000–300,000 new cases being reported every year (Cattand *et al.*, 2001). It has been estimated that over half a million currently carry this disease and more than 50,000 deaths each year are attributed to *T. brucei spp.* (WHO, 2003).

After infection by tsetse fly bite, the disease passes through two clinically distinct early and late phases before progressing to death. West African sleeping sickness has a long, mostly asymptomatic, early phase that can last for several years or more. The lack of any obvious symptoms means that the disease is already very advanced by the time symptoms are observed, at which time the trypanosomes have crossed the blood-brain barrier and have entered the CNS. This denotes the start of the late phase. Finally, the characteristic somnolent state is reached which, if left unchecked, will lead to a coma and death (Stich *et al.*, 2002).

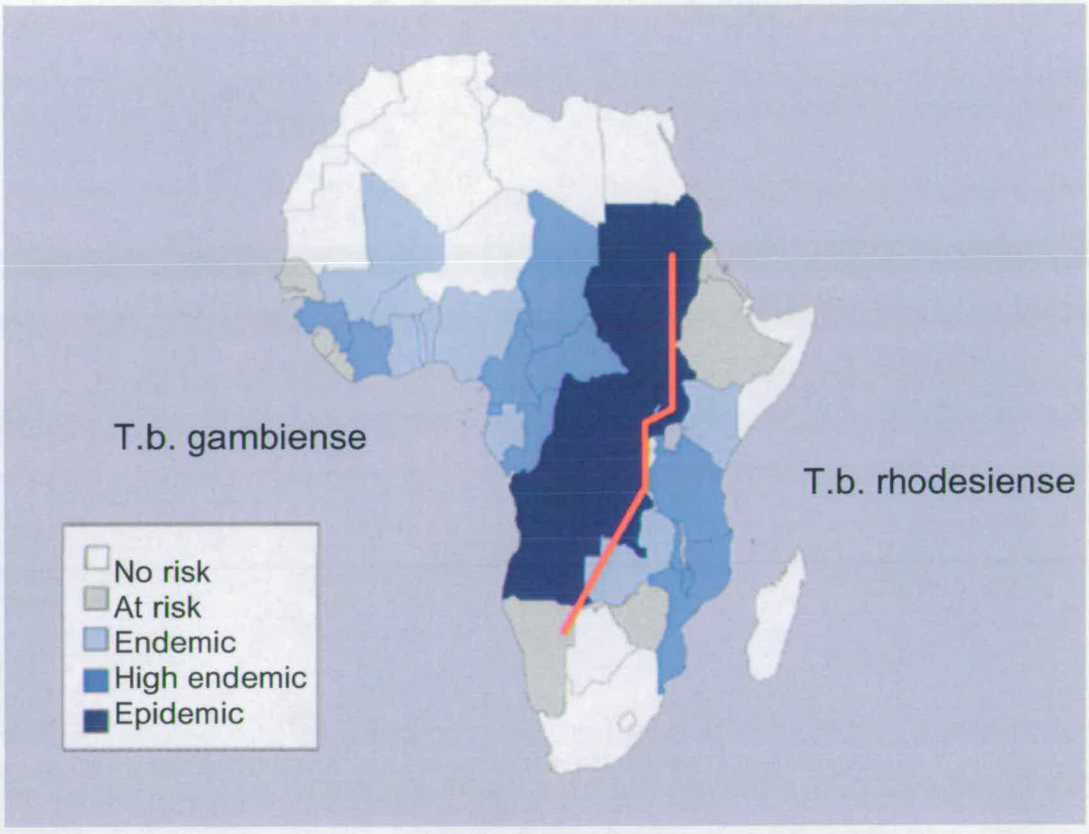

Figure 1.1 – Geographical distribution of Human African Trypanosomiasis (HAT)

T.b. gambiense is most prevalent in western and central Africa, whereas *T.b. rhodesiense* is mainly distributed in eastern African. The red line indicates the boundaries between the two species. The distribution of HAT closely mirrors that of the natural habitat of the tsetse fly vector. (Figure adapted from World Health Organisation, 2003)

Although East African trypanosomiasis follows a similar disease profile, onset of the late phase is much more rapid, with the result that 80% of people die within six months of contracting the disease (Welburn and Odiit, 2002).

1.2 Morphology of T. brucei.

The African trypanosome is an evolutionarily ancient organism and was one of the earliest organisms to diverge from the eukaryotic linage after the acquisition of a mitochondrion (Sogin *et al.*, 1986). The general ultrastructure of a trypanosome is shown in Figure 1.2. DNA is contained in a large central nucleus and within the mitochondrial genome, also known as the kinetoplast, which is located at the posterior end of the cell in the mammalian forms. The mitochondrion itself extends throughout the length of the trypanosome; in bloodstream forms it appears as a single tube lacking cristea, whereas in procyclic forms it is wider and more elaborate due to the presence of cristea and up regulation of mitochondrial enzymes required for energy production in that life cycle stage (Vickerman, 1985).

Intimately associated with the kinetoplast is the basal body, from where the single flagellum emerges that runs along the side of the cell. The flagellum consists of a `9+2' microtubule axoneme and a lattice-like structure that extends from the flagella pocket to the tip and is known as the paraflagellar rod (Bastin *et al.*, 1999). The flagellum is attached to the cell body by complex containing a single filament and 4 specialised microtubules situated in the subpellicular array and is known as the flagellum attachment zone (FAZ), (Kohl and Gull, 1998). Close to the base of the flagellum is the flagella pocket, where all vesicular transport into and out of the cell occurs. Similar to higher eukaryotes, trypanosomes possess the basic protein trafficking organelles, such as the Golgi body (situated close to the nucleus), endoplasmic reticulum (ER) and lysosomes. They also possess specialised

compartments within the cytoplasm called glycosomes, where glycolysis occurs (Clayton *et al.*, 1995; Parsons, 2004).

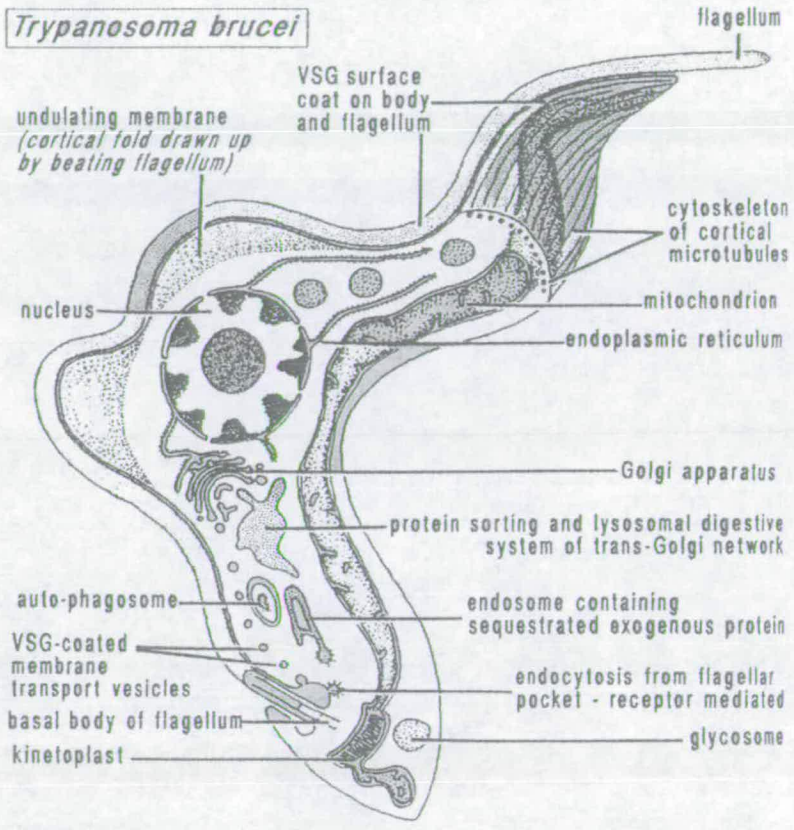

Figure 1.2 - Diagrammatic representation of the basic morphology of an intermediate bloodstream form of *T. brucei*.

The basic features of a trypanosome are highlighted here and are discussed further in the text. Adapted from Vickerman *et al.*, (1992).

The surface of the bloodstream trypanosome is covered with a stage-specific molecule, called Variable Surface Glycoprotein (VSG), which undergoes antigenic variation to protect the organism from antibody lysis (refer to Section 1.3). The cell maintains its structure by a subpellicular corset of microtubules, which are cross-linked to each other and to the plasma membrane and are situated just beneath the

surface of the membrane. The microtubules of the subpellicular corset all have the same polarity and are assembled at the posterior end of the cell, this in contrast to the microtubules of the flagella axoneme, which are assembled at the distal end of the flagellum (Robinson *et al.*, 1995). The major differences between the bloodstream and procyclic forms will be discussed below in further detail.

1.3 Life cycle of T. brucei.

T. brucei has a complicated life cycle and undergoes a number of developmental stages in the mammalian host and tsetse fly (Figure 1.3). Infection in mammals starts when the biting tsetse fly injects mature metacyclic trypanosomes beneath the skin. This causes the formation of a trypanosomal chancre, a local inflammatory response (Barry and Emergy, 1984). From here, the trypanosomes migrate via the lymphatics to the bloodstream, where rapid replication occurs enabling the proliferative slender form to predominate in the population (Vickerman, 1985). Slender forms are long (approximately 29μm) and thin with a free flagellum and express a specific VSG coat. The fluctuating waves of parasitaemia that characterise the bloodstream disease are due to the host raising an antibody to the prevalent VSG protein in the population, which causes a rapid decrease in cell numbers as the trypanosomes are lysed. However, as the population of trypanosomes are heterogeneous for the VSG, trypanosomes will always be present which have no antibodies raised to their VSG and then proliferate to form the next wave of parasitaemia (Gray, 1965).

At the height of each wave of parasitaemia, the proliferative slender population is replaced by non-dividing stumpy trypomastigotes (Vickerman, 1985). Stumpy cells

are usually shorter (approximately 18μl in length), thicker than slender forms and lack a free flagellum. However, these trypanosomes tend to be more resistant to antibody lysis (McLintock *et al.*, 1993; Nolan *et al.*, 2000). Stumpy cells are arrested in the G0 stage of the cell cycle and up-regulate components of the mitochondrial machinery and other pathways to ensure differentiation can occur rapidly and effectively in the tsetse fly (Shapiro *et al.*, 1984; Tyler *et al.*, 1997). The differentiation stimulus for induction of the stumpy form is known as Stumpy Induction Factor (SIF). It is believed to be cell density dependant, though the SIF itself has yet to be characterised (Vassella *et al.*, 1997). Stumpy trypanosomes tend to predominate at each peak of parasitaemia in order to prolong the length of infection; otherwise the slender cells would rapidly colonise and kill the host, interrupting the lifecycle (Matthews, 1999). Interestingly, once a slender cell has differentiated into the stumpy form, the stumpy cell cannot revert back to a slender cell (Hamm *et al.*, 1990). The stumpy cell is committed to continue differentiating to a procyclic cell, but if this does not occur, it will deteriorate or be killed by the antibody response (Turner *et al.*, 1995).

When the Tsetse fly takes an infected blood meal, it contains a heterogeneous population of slender and stumpy trypanosomes. Once in the fly midgut, the slender cells rapidly die off, but the pre-adapted stumpy cells emerge from cell cycle arrest and differentiate into procyclic trypanosomes (Matthews and Gull, 1994; Vickerman, 1985). During the differentiation process, the VSG coat is lost within 6 hours and replaced by a tsetse fly specific procyclin coat, expression of which is detectable within 2 hours (Matthews and Gull, 1994; Ziegelbauer and Overath, 1990).

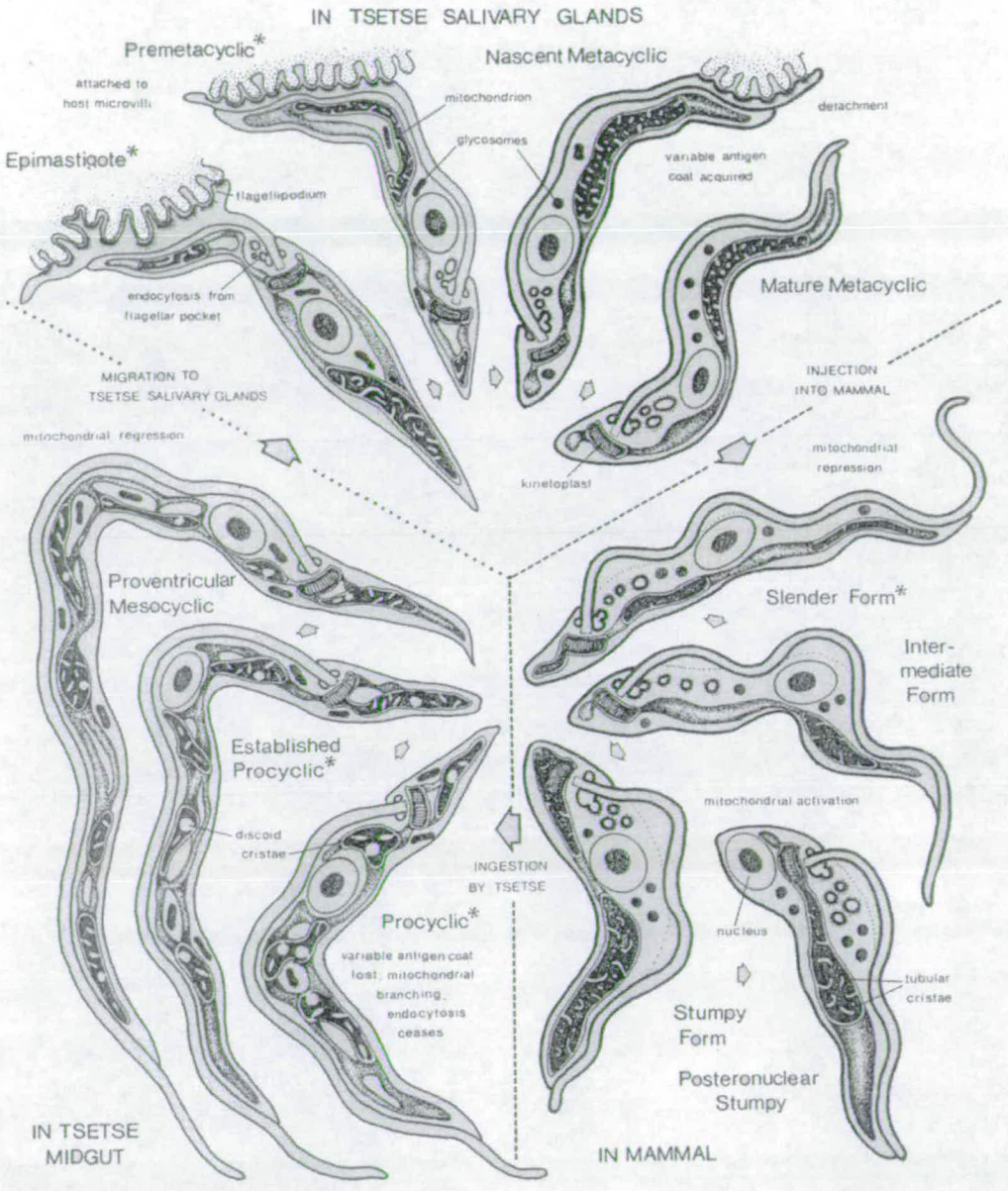

Figure 1.3 – Diagrammatic representation showing the lifecycle of the African trypanosome.

T. brucei undergoes a complex biphasic lifecycle, with developmental stages occurring within the bloodstream and tsetse fly midgut and salivary glands. Mammalian forms possess a VSG coat, whereas the tsetse fly forms contain a procyclin coat. Figure adapted from Vickerman (1985).

The organism also undergoes remodelling of its cytoskeleton, which causes an increase in the dimensions of the posterior end of the cell that leads to repositioning of its kinetoplast (Matthews *et al.*, 1995). Finally, enzymatic components of the Krebs cycle and oxidative phosphorylation are up-regulated, although a fully functional Krebs cycle is not essential to the procyclics' survival (Durieux *et al.*, 1991; van Weelden *et al.*, 2003). These activities are absent from bloodstream forms as all energy requirements can be met by glycolysis, due to the presence of abundant glucose in the bloodstream (Van Hellemond *et al.*, 1998). After a week, the established procyclic cells increase in size to become proventricular mesocyclic cells, which leave the midgut and migrate via the oesophagus and salivary ducts to the salivary glands where the next stage in the life cycle occurs (Vickerman, 1985).

Once in the salivary gland, the trypanosomes undergo division to become an epimastigote - another proliferative stage. The epimastigotes attach themselves to the epithelial cells via their flagellum, which undergoes dendritic outgrowth to ensure effective attachment (Tetley and Vickerman, 1985). Next, the epimastigotes divide into premetacyclics, which are still proliferative and are attached to the epithelial cells. Once cell division is complete and a surface coat has developed, the cells are classed as nascent metacyclic. Finally, the trypanosomes are released from the epithelium and the mature metacyclic trypanosomes are ready for introduction into the mammalian host (Vickerman, 1985).

1.3.1 Cultivation of trypanosomes.

Some life cycle stages of *T. brucei* can be cultivated *in vitro* in the laboratory, mainly the bloodstream and procyclic forms. Bloodstream cell lines, which have been serially passaged through animals until no stumpy forms develop and then culture adapted, are called monomorphic lines (Vickerman, 1965). Monomorphic lines show reduced VSG switching and proliferate so rapidly that an uncontrolled parasitaemia develops if these lines are reintroduced into an animal host (Barry and McCulloch, 2001; Turner *et al.*, 1995). Although wild type (also known as pleomorphic) cell lines rich in stumpy forms undergo synchronous differentiation to procyclic forms with high efficiency (roughly 95%), monomorphic trypanosomes differentiate asynchronously and with a much lower efficiency (Matthews *et al.*, 2004; Matthews and Gull, 1994). It is possible to induce differentiation *in vitro* from bloodstream trypanosomes to procyclic cells by a reduction in temperature from 37^0C to 27^0C and the addition of *cis*-aconitate and/or citrate (Czichos *et al.*, 1986). *Cis*-aconitate alone will induce differentiation, but a reduction in temperature is necessary for cell cycle re-entry and DNA synthesis (Matthews and Gull, 1997).

1.4 Differentiation events from Bloodstream to Procyclic trypanosomes.

The process of differentiation from bloodstream to procyclic trypanosomes follows a defined temporal pattern. Stumpy cell populations differentiate to procyclic cells efficiently and synchronously, which has enabled mapping of the major events. The first noticeable effect of differentiation from a stumpy cell is the gain of the procyclin surface coat within two hours of the onset of differentiation (Figure 1.4) (Matthews and Gull, 1994; Roditi *et al.*, 1989). Two major forms of procyclin are expressed, EP and GPEET. EP contains three different molecular isoforms that vary in glycosylation status and Glu-Pro peptide repeats (named EP1, 2 and 3), whereas GPEET is classified by the Gly-Pro-Glu-Glu-Thr repeats (Roditi and Clayton, 1999; Roditi *et al.*, 1998). GPEET is initially highly expressed, along with EP, but is subsequently down regulated, with EP becoming the prevalent surface protein (Vassella *et al.*, 2001). However, the Variant Surface Glycoprotein (VSG) coat is not lost until 4-6 hours have elapsed, which means that both stage-specific antigens are expressed for a short time (Matthews and Gull, 1994; Roditi *et al.*, 1989). Loss of the VSG coat is very rapid and occurs primarily due to zinc metalloprotease activity ((MSP B - major surface protease B), though proteolytic cleavage of the VSG molecule by a glycosylphosphatidylinositol-specific phospholipase C (GPI-PLC) is also required (Gruszynski *et al.*, 2006; Webb *et al.*, 1997; Ziegelbauer *et al.*, 1993). This is then followed by a pulse of adenylate cyclase activity between 7-9 hours (Rolin *et al.*, 1993) and the remodelling of the cytoskeleton, as shown by the expression of the microtubule-binding protein Cap5.5 (Hertz-Fowler *et al.*, 2001).

During remodelling of the cytoskeleton, there is outgrowth at the cell posterior end of the cell and slight movement of the kinetoplast towards the nucleus, in order to establish the correct organelle positioning that is so vital to the cell (also known as kinetoplast repositioning) (Matthews *et al.*, 1995; Vickerman, 1965). At a similar time of 6 hours, a burst of RNA synthesis occurs, enzymatic components of the Krebs cycle are expressed and detectable events in cell cycle re-entry occur (Matthews and Gull, 1994). Between 6-12 hours, DNA replication of the kinetoplast and then nucleus occurs, followed closely by the separation of the basal bodies and the growth of a new daughter flagellum. Finally, from 10-20 hours, the kinetoplasts segregate and mitosis and then cytokinesis occurs. These events provide useful markers to assess the progress of differentiation.

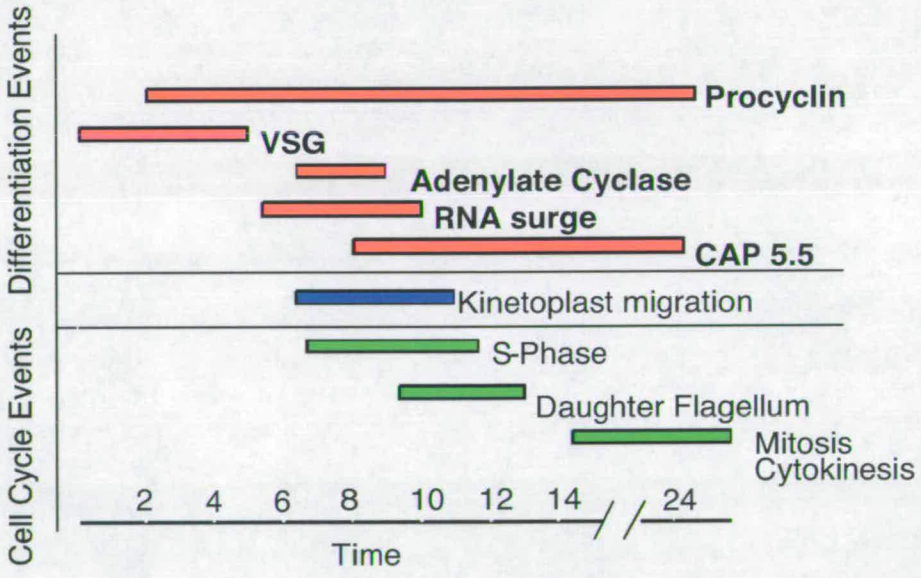

Figure 1.4 - A diagram showing the temporal events that occur during synchronous differentiation from bloodstream to procyclic trypanosomes.

The major events of differentiation from stumpy forms are shown in relation to the time when they occur (Time – X axis) and events in the cell cycle. Figure adapted from Matthews and Gull (1994).

1.5 Surface Proteins.

The two major classes of surface proteins expressed by trypanosomes are the Variant Surface Glycoprotein (VSG), which is expressed in bloodstream stages and the procyclin coat expressed by the insect midgut forms.

1.5.1 VSG.

The VSG is the major surface protein of the bloodstream and metacyclic forms of the parasite. It forms a tightly packed immunogenic coat containing approximately 10^7 molecules per cell and the constant antigenic variation of this coat enables the trypanosome population to survive the host immune response. The molecular weight of a VSG protein is around 55kDa and the proteins form homodimers that are attached to the plasma membrane via a glycosylphophatidylinositol (GPI) anchor (Ferguson *et al.*, 1999). Although only one VSG gene is expressed at any one time, the whole genome is thought to contain in excess of 1000 VSG genes, of which 806 have currently been characterised (Berriman *et al.*, 2005; Barry, 1997; Vanhamme and Pays, 1995). The switching of the VSG coat occurs rapidly and spontaneously at a rate of 10^{-2} per cell and per generation (Turner and Barry, 1989). In order for a VSG gene to be expressed, it must be located in an active Expression Site (ES). There are approximately 20-40 expression sites in the trypanosome genome and these are located at the telomeres and are bloodstream or metacyclic stage-specific (Figure 1.5) (Vanhamme *et al.*, 2001). However, like the VSG gene, only one ES is active at any one time. Bloodstream expression sites are long (40 - 60 kb)

polycistronic units containing up to 12 Expression Site Associated Genes (ESAGs) and an RNA Polymerase I promoter, and are flanked by telomeric repeats (TTAGGG) at one end and 50 bp repeats at the 5' end. The VSG gene is the last gene in the cluster before the telomere and is separated from the rest of the ES by a group of 70 bp repeats.

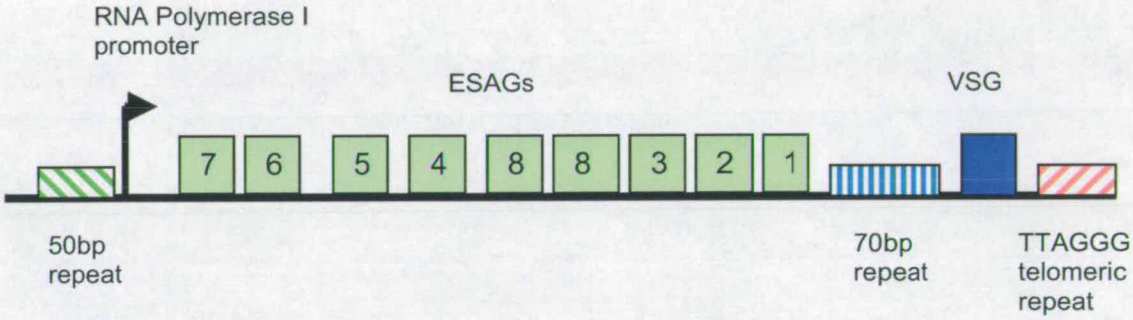

Figure 1.5 – Schematic representation of a bloodstream form VSG expression site.

Bloodstream VSG expression sites are situated at the ends of chromosomes. Each expression site contains a distant promoter that transcribes the ESAGs as well as the VSG itself. The VSG gene is flanked upstream by a series of 70 nucleotide repeats that enable silent genes to be switched into the ES. Downstream, the VSG gene is flanked by the telomeric repeats.

In order to express a different VSG gene, either gene rearrangement has to occur whereby an internal chromosomal or ectopic VSG gene is switched into the active ES or transcriptional switching occurs that results in another ES being activated and the present one silenced (Vanhamme *et al.*, 2001). The replication of the VSG gene and its insertion into the ES, replacing the old VSG gene, is called duplicative transposition. Recent studies tend to suggest that duplicative transposition is favoured over ES switching (Robinson *et al.*, 1999). Normally, the whole VSG gene is replaced; however, on occasions mosaic genes are formed, which range from just a

few kilobases to large segments (Pays, 1985). The formation of these mosaic genes tends to occur late on in a trypanosome infection. However, the recent data from the trypanosome genome sequencing project has shown that the majority of silent chromosomal VSG genes are non-functional, with only 7% of VSG genes analysed being fully functional (Berriman *et al.*, 2005). This potentially implies a more important role for mosaic formation than previously thought.

It has been proposed that in order for a VSG ES to be expressed, it has to be associated with the Expression Site Body (ESB) (Navarro and Gull, 2001). The ESB is an extranucleolar RNA polymerase I transcriptional factory that is located within the nucleus and is only present in bloodstream forms, despite the fact that the procyclin genes are also transcribed by RNA Polymerase I. As only the active VSG ES is associated with the ESB, this would explain the inability to activate other ESs simultaneously, although the mechanism for how the ESB achieves this has yet to be established (Navarro and Gull, 2001).

1.5.2 Procyclin.

There are two types of procyclin proteins expressed on the surface of a procyclic cell, EP and GPEET, which are named after the internal peptide repeats. EP contains three different molecular isoforms that vary in glycosylation status and Glu-Pro peptide repeats (named EP1, 2 and 3), whereas GPEET is classified by the Gly-Pro-Glu-Glu-Thr repeats (Roditi and Clayton, 1999; Roditi *et al.*, 1998). The procyclin glycoproteins are densely packed (approximately $3x10^6$ proteins/cell) and cover the entire surface of the procyclic cell in the fly midgut. Like the VSG proteins, all forms

of procyclin are anchored to the cell membrane via a GPI anchor (Field *et al.*, 1991). During differentiation, both EP and GPEET are up regulated within 2 hours of the start of differentiation, but after a few days GPEET expression is down regulated, with EP1 and EP3 becoming the prevalent procyclin isoform (Vassella *et al.*, 2001). Expression of GPEET procyclin is regulated by a 25 nucleotide glycerol-response element located within the 3' UTR, with the result that when glycerol is added to the media, GPEET expression is not down-regulated (Vassella *et al.*, 2004). Additionally, cold shock (a drop of temperature to below 15°C) is required to induce EP expression and hyposensitizes stumpy trypanosomes to *cis*-aconitate (Section 1.3.1) (Engstler and Boshart, 2004). Both EP and GPEET have a highly repetitive structure, which cannot be cleaved by proteases present in the midgut (Acosta-Serrano *et al.*, 2001). It has been suggested that this may explain the procyclin structure as it allows protection against proteolytic bombardment.

Expression of procyclin is tightly regulated as inappropriate expression of procyclin within the bloodstream would lead to rapid recognition and lysis by the immune system. In bloodstream forms, transcription of procyclin is down regulated 10 fold, while a 26mer sequence within the 3' UTR causes rapid degradation of the transcript if transgenically expressed within the bloodstream stages (Hotz *et al.*, 1998; Hotz *et al.*, 1997). Also, deletion of a 16 nucleotide stem-loop, leads to a 3 fold decrease in translation in procyclic as well as bloodstream forms, indicating that this is a positive regulatory sequence controlling procyclin expression (Hehl *et al.*, 1994; Hotz *et al.*, 1997).

1.6 Genome Organisation.

1.6.1 The nuclear genome.

The *T. brucei* nuclear genome is approximately 26Mb in size and divided into three chromosome classes: megabase, intermediate and minichromosomes (Berriman *et al.*, 2005). There are 11 megabase chromosomes, ranging in size from 5Mb to 1Mb and which contain approximately 9086 predicted genes (Berriman *et al.*, 2005). Intermediate chromosomes are between 200 - 900kb in length, while the size of minichromosomes ranges from 50kb – 150kb (El-Sayed *et al.*, 2000; Melville *et al.*, 1998). The megabase chromosomes contain mainly housekeeping and trypanosome-specific genes, which are ordered into polycistronic gene clusters near the centre of the chromosome (Berriman *et al.*, 2005; Hall *et al.*, 2003). On chromosome 1, these directional clusters contain roughly 400 genes, some of which are tandemly arrayed (Hall *et al.*, 2003). Trypanosome genes do not contain introns, with the only exceptions characterised to date being the poly-A polymerase (Mair *et al.*, 2000) and the tRNA (Tyr) genes (Schneider *et al.*, 1994). Interestingly, there is great variation in size (~ 1MB) between homologous chromosome pairs. The cause of this variation is unknown, but possible factors include rearrangement of VSG and ES sites, telomeric repeats, tandemly arrayed housekeeping genes and other unknown repetitive sequences (Melville, 1997; Melville *et al.*, 1999).

Intermediate chromosomes vary in size and number between different *T. brucei* stocks. They contain few housekeeping genes, although several VSG genes have

been identified. Minichromosomes are 30-150kb linear DNA molecules that contain a 177 bp tandem repeat and silent VSG genes (Sloof *et al.*, 1983; Wickstead *et al.*, 2003). As no ES are associated with the minichromosomes, the DNA must either be duplicated or undergo telomeric exchange in order for these genes to be expressed (Wickstead *et al.*, 2003). It appears that the function of the minichromosomes is to act as a repository for the VSG genes, although why trypanosomes devote such a large part of their genome (~20%) to this function remains unclear (Wickstead *et al.*, 2003).

1.6.2 The kinetoplast.

The kinetoplast is a mass of DNA composed of approximately 50 maxicircles and ~10 000 interlinked minicircles. As there is only one copy of the kinetoplast in each cell, careful replication and segregation of this organelle is vital to the trypanosome's survival (Klingbeil *et al.*, 2001). The maxicircles are roughly 22kb in size and are homologous to other eukaryotic mitochondrial DNA in that they encode mitochondrial proteins. However, the majority of transcripts need the addition or deletion of nucleotides to encode a functional mRNA, a process known as RNA Editing (See Section 1.6.3). RNA editing requires the use of guide RNAs (gRNAs), which are encoded on the 1kb minicircles. Each minicircle can encode three to four guide RNAs, which can give a complexity of over 1200 gRNAs (Stuart *et al.*, 2005). Not all genes required in the mitochondrion are encoded within the kinetoplast, genes for tRNAs and cytochrome oxidase subunits IV – X are encoded within the nucleus and must be imported into the mitochondrion (Tan *et al.*, 2002; Timms *et al.*, 2002).

1.6.3 RNA Editing.

In order for the trypanosome to produce functional mitochondrial proteins from its mitochondrial genome, some transcripts require the posttranslational insertion or deletion of uridine bases at specific sites within their sequence (Figure 1.6). The process requires guide RNAs (gRNAs) that are complementary to the potential mRNA sequence (Blum and Simpson, 1990). Each gRNA is between 55-70 nucleotides in length, contains an "anchor" sequence that allows complementary interaction between the gRNA and the pre-mRNA, an editing sequence that contains the U insertions and deletions and a final 3' oligo (U) tail (Blum and Simpson, 1990). Initially, each pre-mRNA forms an anchor duplex with the 5' region of the gRNA. Next, the pre-mRNA is cleaved by an endonuclease 8-10 bases upstream of the anchor complex, with U residues being added or deleted as specified by the gRNA using either the TUTase (terminal uridylyl transferase) or 3'-ExoUase respectively. The resultant mRNA is then ligated together by an RNA ligase to complete the step. Editing then proceeds in a 3' to 5' direction, relative to the mRNA, which results in the anchor sequence of the gRNA binding to an edited sequence that did not previously exist (Maslov *et al.*, 1992). Multiple gRNAs need to be coordinated along the whole pre-mRNA to create a functional mature mRNA. A multiprotein complex called the editosome catalyses the whole process, which while the subunits have yet to be fully characterised, sediments at approximately 20S as analysed by glycerol gradients (Stuart *et al.*, 2005). Components of the editosome identified to date and their putative functions are reviewed by Stuart *et al.* (2005).

It has recently been proposed that the RNA editing complex has an alternative function – that of generating diversity in open reading frames and the creation of novel proteins (Ochsenreiter and Hajduk, 2006). In support of this, an alternatively edited copy of the cytochrome oxidase III was identified and the protein localised to the mitochondrial membrane fractions (Ochsenreiter and Hajduk, 2006). The extent of this mechanism in generating novel proteins remains to be established, though it does raise interesting possibilities as a mechanism of generating protein diversity.

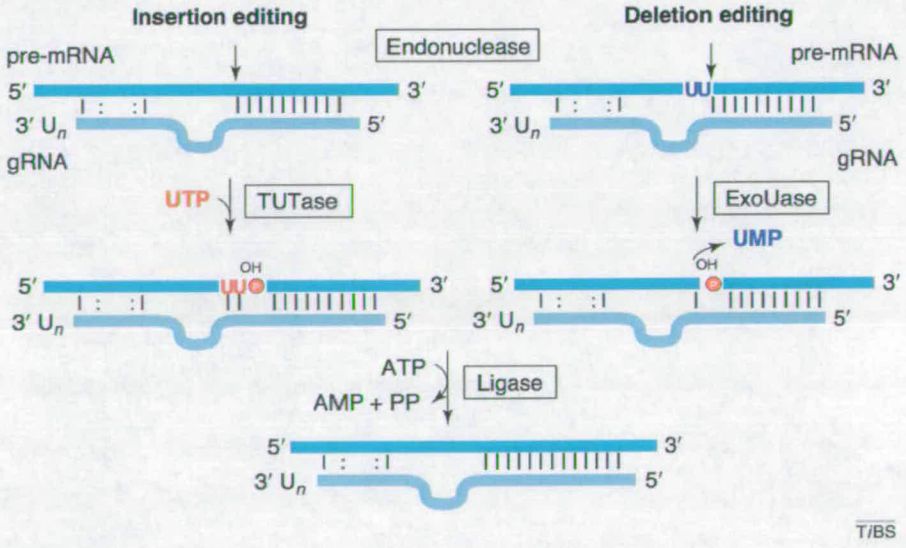

Figure 1.6 – Schematic representation of the mechanism of RNA Editing in trypanosomes.

The gRNA binds via the anchor region to the pre-mRNA. Next the pre-mRNA is cleaved by an RNA endonuclease. Next the U residues are either added by a TUTase or removed using an ExoUase, according to the template provided by the gRNA. Finally, the 5' and 3' RNA fragments are ligated using an RNA ligase. Figure adapted from Stuart *et al.* (2005).

1.7 Gene Expression.

T. brucei has novel and unusual ways of regulating gene expression, quite different from other eukaryotes. While most eukaryotic organisms adapt to changes in their environments by altering gene expression, usually by regulating transcription initiation, trypanosomes appear to regulate gene expression at the posttranscriptional level (Vanhamme and Pays, 1995). Although three types of RNA polymerases have been identified, which share significant homology to those in higher eukaryotes, some differences remain. RNA Polymerase II transcribes mainly protein coding genes and splice leader (SL) genes (Clayton, 2002). However, no RNA Polymerase II promoters have been identified for protein coding genes, the only exception being the promoters identified for the spliced leader genes (See Section 1.7.2)(Gilinger and Bellofatto, 2001). Also, RNA Polymerase I transcribes not only the ribosomal RNAs, but also the protein coding VSG and procyclin genes, an unusual feature of the kinetoplastida (Hotz *et al.*, 1998). A putative explanation for the use of RNA Polymerase I for transcription of the VSG and procyclin genes is that it may allow strong initiation of transcription, which results in large quantities of these transcripts and hence proteins, being produced (Hotz *et al.*, 1998; Lee and Van der Ploeg, 1997). RNA polymerase III transcribes the tRNAs and U small nuclear RNAs (Clayton, 2002). In addition, substantially fewer transcriptional regulators have been identified in trypanosomes, lending more substance to posttranscriptional control being the major gene control mechanism in these organisms (Ivens *et al.*, 2005).

1.7.1 Polycistronic Transcription.

The mechanisms of RNA transcription and maturation in trypanosomes are very interesting and although they share some homology with higher eukaryotes, many differences exist. Trypanosomes produce long pre-mRNAs, derived from polycistronic transcription units that contain multiple unrelated gene coding regions, which are then spliced into functional mRNAs. Polycistronic transcription units can span up to 1.2Mb and are separated by strand switch regions that contain an unusual composition of bases (Ivens *et al.*, 2005; Tosato *et al.*, 2001). The result of this is that many genes are likely to be under the control of a single upstream promoter (Clayton, 2002). It has been shown experimentally that RNA Polymerase II can initiate transcription within these strand-switch regions, though no specific sequence motif has been identified (Martinez-Calvillo *et al.*, 2004). Some polycistronic transcription units contain related genes, such as the genes encoding phosphoglycerate kinase pathway, while other polycistronic mRNA contain completely unrelated genes (Gibson *et al.*, 1988).

The pre-mRNA derived from a polycistronic transcription unit must be processed into individual mature mRNAs. This is achieved by the addition of a capped 39 nucleotide RNA sequence, called the splice leader (SL) sequence, which is added upstream of each coding region (Murphy *et al.*, 1986). The acceptor site for the *trans*-splicing of the SL is an AG dinucleotide, which is preceded by a polypyrimidine tract (Liang *et al.*, 2003). Next, polyadenylation at the 3' end allows the production of a translatable mRNA (Ullu *et al.*, 1993). The process of splice leader addition and polyadenylation are physically coupled as the site of

polyadenylation is determined by trans-splicing of the downstream mRNA (LeBowitz *et al.*, 1993; Matthews *et al.*, 1994). Unlike the majority of higher eukaryotes where the polyadenylation site is denoted by the AAUAAA sequence, no consensus polyadenylation signal has been characterised as yet (Hendriks *et al.*, 2003). However, recent data that tried to map the mRNA processing sites identified the trans-splicing site as generally occurring at the first AG following a polypyrimidine tract of 8-25 nt (Benz *et al.*, 2005).

1.7.2 RNA degradation.

The machinery for degradation of mRNA in trypanosomes shares considerable similarity with other eukaryotes, although some factors involved in the pathway remain to be identified. The first stage of RNA degradation involves the Poly (A) tail being trimmed into short oligo (A) fragments (deadenylation) before decapping at the 5' end of the mRNA occurs (Figure 1.7). The remaining mRNA is then degraded by the exosome in the 3'-5' direction and by cytoplasmic exonucleases in the 5'-3' direction (including the XRNA, XRNB, XRNC and XRND proteins)(Wilusz *et al.*, 2001). The exosome is a multimeric protein complex with ribonuclease activity and is similar to that of higher eukaryotes, being composed of at least 11 subunits in trypanosomes (CSL4, RRP4, RRP6, RRP40, RRP41A, RRP41B, RRP43, RRP45) (Estevez *et al.*, 2001; Estevez *et al.*, 2003).

As most genes are regulated at the posttranscriptional level, mRNA stability is of great importance. As yet, only a handful of short linear sequences have been identified which affect the half life of the mRNA and it appears that this important

role may lie within the secondary structure of the 3' untranslated region (UTR). For example, the 3'-UTRs of the VSG genes do possess an 8 mer and 14 mer linear sequence that regulates mRNA levels and translation, suppressing expression in the insect stages and stimulating expression in the bloodstream form (Berberof *et al.*, 1995).

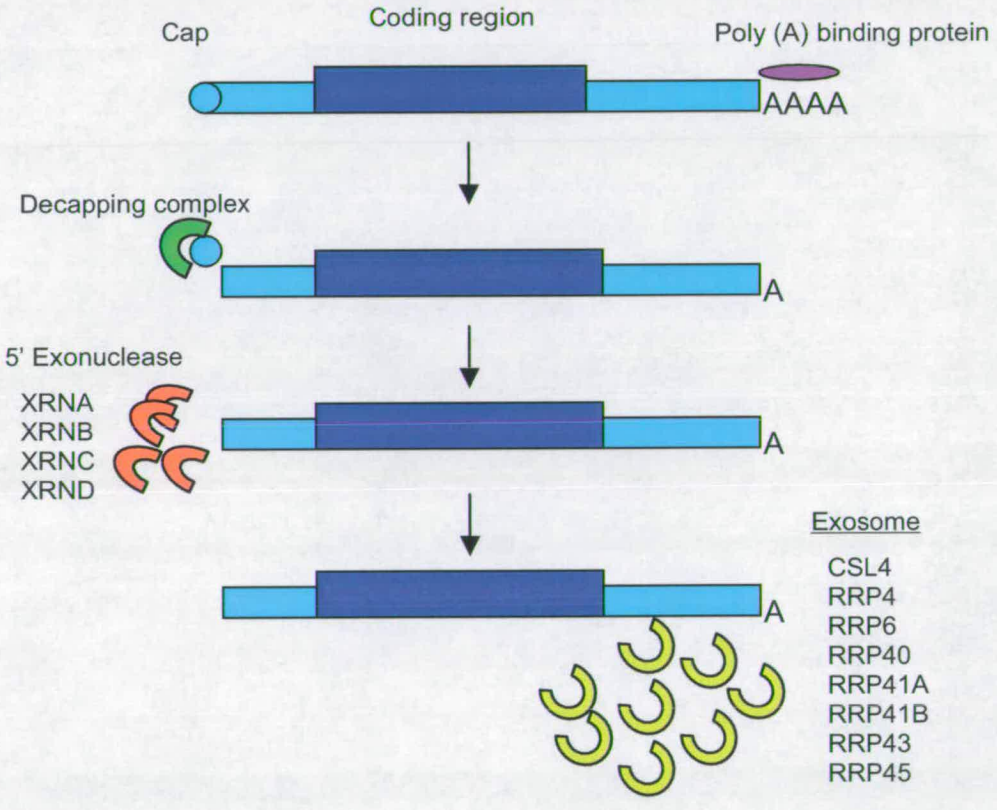

Figure 1.7 - Schematic diagram demonstrating the suggested process of mRNA degradation in trypanosomes.

The transcript is stabilized by Poly (A) binding proteins and regulatory proteins in the 3' UTR, before deadenylation by the Poly (A) nucleases occurs, causing decapping to occur at the 5' end. The remaining mRNA is degraded by the exosome and 5'-3' exonucleases. Adapted from Clayton (2002)

Regulatory motifs in the secondary structure of the 3' UTR are best understood in the EP and GPEET procyclin molecules. These 3' UTRs share little obvious homology except for a U-rich 26 mer and a 16 mer stem loop structure described earlier. The 16 mer structure causes enhanced translation to occur (Furger *et al.*, 1997), while deletion or alteration of the 26 mer, leads to eliminated or reduced developmental regulation (Hotz *et al.*, 1997; Schurch *et al.*, 1997). Structural mapping of the U-rich 26 mer has shown that it is normally present as a single stranded structure as no stable base pairing interactions have been identified (Drozdz and Clayton, 1999). Similar U-rich sequences have been found in other 3' UTRs and abolition of these single stranded U-rich domains caused the cessation, at both the RNA and protein level, of stage specific regulation. These sequences share a strong homology with the AU-rich elements (AREs) found in mammalian cells, which destabilize the mRNAs involved in differentiation and cell growth (Mitchell and Tollervey, 2000). Recently, an analysis of sequences within the 3' UTR of the cytochrome oxidase complex identified sequences important within many procyclin-specific transcripts as well as amongst a subset of regulated mitochondrial proteins (Mayho *et al.*, 2006). A U-rich sequence was identified that shared significant homology with the 26 mer identified by Furger *et al.*, (1997) and was present in over 30% of procyclic-enriched transcripts (Furger *et al.*, 1997) (Mayho *et al.*, 2006).

1.8 Regulation of mRNA stability in higher eukaryotes

The mechanisms controlling mRNA stability in higher eukaryotes are many and varied. While trypanosomes possess the basic core machinery for mRNA regulation found throughout higher eukaryotes, many elements and processes still remains to be fully characterised.

In higher eukaryotes such as *S. cerevisae*, the majority of mRNA follows a similar decay pathway to that found in trypanosomes (Section 1.7.2). That is to say the initial stage of mRNA degradation is the shortening of the 3' polyA tail, followed by the removal of the 5' 7-methyguanosine cap and then degradation of the transcript in a 5' to 3' direction by the 5' XRN1 ribonuclease or by the exosome in a 3' to 5' direction (Amrani *et al.*, 2006). However, the exact trigger that causes the initial deadenylation has yet to be identified (Amrani *et al.*, 2006).

Some higher eukaryotes contain specialist cytoplasmic site called P bodies where mRNA degradation can occur (Anderson and Kerdersha, 2006). However, not all of the mRNA decay machinery is localised to P bodies and P bodies are not always found with the cytoplasm, unless the cells are exposed to stress (Kedersha *et al.*, 2005). P bodies are most prevalent when there is an excess of mRNA within the cell. However, if the levels of RNA decrease (or transcription is inhibited), the levels of P bodies within the cytoplasm decreases. Recent data from *S. cerevisae* has proposed that mRNA can be released from P bodies if the cells are deprived of glucose, implying a role in mRNA regulation, not just mRNA degradation (Teixeira *et al.*,

2006). Thus, it appears that P bodies are dynamic structures, which potentially regulate the interaction between mRNA decay and translation.

While the majority of the mRNA undergoes degradation via the deadenylation pathway described above, other pathways do exist to process specific mRNA subsets. The RPS28B protein in *S. cerevisae* is able to autoregulate its own decay, by binding to a stem-loop structure in its 3' UTR and recruiting decapping factors. This process is initially independent of polyA degradation (Badis *et al.*, 2004). Similarly, EDC1 mRNA (which encodes a decapping regulatory protein) contains a polyU stretch within its 3' mRNA which prevents deadenylation, so the initial stage in its degradation is removal of the cap structure (Muhlrad and Parker, 2005). Additionally, some mRNAs are cleaved directly into two fragments by cellular endonuclease, such as PMR1 and IRE1 (Yang *et al.*, 2006). Other mRNAs are regulated by the Argonaute protein 2 (AGO2), which requires short interfering RNAs (siRNAs) as templates. The AGO2 protein is an important constituent of the RNA interference pathway (Liu *et al.*, 2004).

Several pathways exist in all eukaryotes to prevent mRNA transcripts that contain mistakes or premature stop codon from being translated. One of these pathways is the nonsense-mediated decay pathway and contains 3 highly conserved core proteins – UPF1, UPF2 and UPF3, though the pathway itself is quite divergent across eukaryotes (Conti and Izaurrable, 2005). The identification of abhorrent mRNAs in mammalian cells is usually identified by the presence of the exon junction complex (EJC). In mRNA transcripts that have been correctly processed, all EJCs will have

been removed. However, in transcripts that contain abhorrent stop codons, these EJC will remain and are then detected by UPF1, which targets the mRNA for degradation (Buhler *et al.*, 2006). The nonsense-mediated pathway does not only recognise aberrant transcripts, it is also believed to play a role in the regulation of mRNA. For example, mammalian cells depleted in UPF-1 show up-regulation of "normal" transcripts lacking any mutations (Mendell *et al.*, 2004).

Non-stop decay is used to target mRNAs that lack a stop codon, which are usually formed from frame shift mutations or through breakage of the mRNA. The pathway can operate in two mechanisms; either the translating ribosome dislodges the PolyA binding protein (PABP), which exposes the RNA to exosome degradation, or by the recruitment of the SKI complex (Frischmeyer *et al.*, 2002). The SKI complex (consisting of Ski2, Ski3, and Ski8) is recruited to the stalled ribosome, causing it to dislodge before the exosome and the SKI complex initiate deadenylation (Frischmeyer *et al.*, 2002). Whilst some of the mechanisms described here have been partially characterised in trypanosomes, many more remain to be investigated and the physiological role in mRNA stability remains to be identified.

1.9 Regulation of Development.

The most likely level of control of gene expression and developmental processes in trypanosomes is at the posttranscriptional level. Indeed, analysis of the genome has identified almost 6 times the number of RNA binding proteins containing the CCCH motif in trypanosomes than in *S. cerevisiae* (Ivens *et al.*, 2005). Several molecules have been identified within trypanosomes that can regulate mRNA stability and modify abundance by targeting the AU rich elements (AREs), located within the 3' UTRs. In *Trypanosoma cruzi*, over expression of a U-rich RNA-binding protein (*Tc*UBP-1) has been shown to destabilise the small mucin gene mRNA by interacting with the AREs (D'Orso and Frasch, 2002). Other examples include a pumilio family member (*Tb*PUF1), which belongs to the large PUF family found throughout higher eukaryotes and which was first identified in *Drosophila melanogaster* and *C. elegans* (Hoek *et al.*, 2002). *Tb*PUF1 has been shown to affect the stability of Expression Site Associated Gene (ESAG) 8, presumably in a similar way to other members of the PUF family that bind to the 3' UTR and regulate mRNA translation and stability (Hoek *et al.*, 2002). Ectopic expression of the human RNA binding protein (HuR) in trypanosomes has shown that it can bind to the AREs in the mRNA of phosphoglycerate B kinase(PGKB), leading to altered abundance and stabilisation of the mRNA (Quijada *et al.*, 2002).

1.10 The small CCCH zinc finger protein family in T. brucei.

A differential display procedure undertaken to identify transcripts involved in differentiation from bloodstream to procyclic cells identified a novel small RNA binding protein, named *Tb*ZFP1 (Hendriks *et al.*, 2001). Two further members of this family (*Tb*ZFP2 and 3) were later identified by genome analysis and all contain the characteristic $C_{x8}C_{x5}C_{x3}H$ (CCCH) zinc finger motif, hence the name *Trypanosoma brucei* Zinc Finger Protein (*Tb*ZFP)(Figure 1.8) (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). The $C_{x8}C_{x5}C_{x3}H$ zinc motif is a known RNA binding element and has been found to regulate mRNA stability and localisation in a wide variety of eukaryotes (Blackshear, 2002).

*Tb*ZFP1 has a small molecular weight (being only 101 amino acids long) and is encoded by an mRNA approximately 3.5 kb in length. In addition to the CCCH RNA binding motif, it possesses the Proline-Proline- $X_{(\text{any amino acid})}$ -Tyrosine amino acid sequence (PY) protein interaction domain (Hendriks *et al.*, 2001; Pirozzi *et al.*, 1997). *Tb*ZFP1 is transiently expressed during differentiation, appearing 2-8 hours after the initiation of differentiation from bloodstream to procyclic forms, and is also expressed at low levels in procyclic forms. This is in contrast to *Tb*ZFP2 and 3 which are constitutively expressed in both bloodstream and procyclic forms (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). While *Tb*ZFP1 has been shown to be non-essential in bloodstream cells, it is apparently essential for survival of procyclic cells since ablation of the gene was not tolerated (Hendriks and Matthews, 2005). Similarly, *Tb*ZFP3 is apparently essential for survival in both bloodstream and procyclic

trypanosomes, unlike *Tb*ZFP2 which can be ablated in both life cycle stages without obvious phenotype (Hendriks *et al.*, 2001; Paterou *et al.*, 2006).

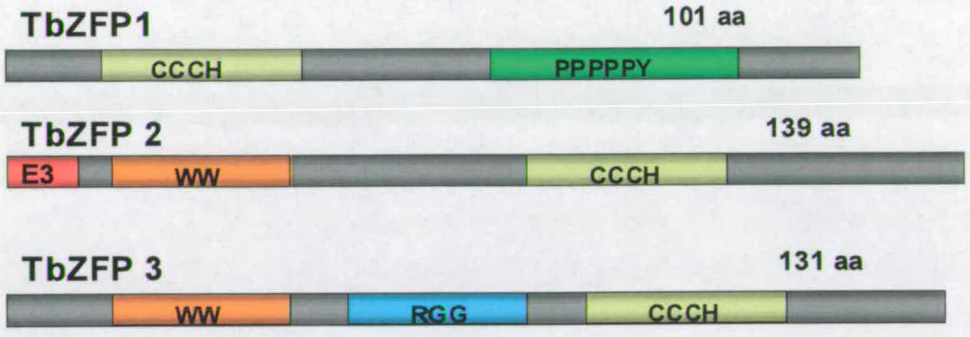

Figure 1.8 - Schematic diagram of TbZFP1, 2 and 3 showing the approximate positions of the motifs.

All three proteins contain the CCCH RNA binding motif, with *Tb*ZFP2 and 3 containing the WW protein interaction motif, while TbZFP1 contains the complementary PY domain. Additionally TbZFP3 possesses an RGG RNA binding domain, while TbZFP2 contains a motif associated with E3 ubiquitin ligases.

In addition to the CCCH RNA binding domain, the 139 amino acid molecule *Tb*ZFP2 contains a highly conserved WW protein interaction domain (characterized by the presence of two tryptophan residues) and a motif previously only associated with E3 Ubiquitin ligases (Hendriks *et al.*, 2001). Similarly, *Tb*ZFP3 also possesses the WW domain and CCCH RNA binding motif, but additionally also possesses an RGG motif (characterised by Arginine-Glycine-Glycine repeats), which has been implicated in RNA binding (Kiledjian and Dreyfuss, 1992; Paterou *et al.*, 2006). Intriguingly, the WW protein-protein interaction motif is complementary to, and interacts with, the PY motif, raising the possibility of *Tb*ZFP1 being able to interact with *Tb*ZFP2 and 3 (Sudol *et al.*, 1995).

Ectopic overexpression of both *Tb*ZFP2 and 3 in procyclic cells leads to an extension of the posterior dimensions of the cell, which has been termed the "nozzle" phenotype (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). Mutation of the CCCH or WW motifs abolished the formation of the nozzle phenotype in *Tb*ZFP3, whereas elimination of these domains in *Tb*ZFP2 resulted in reduced nozzle size (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). Additionally, the deletion of the first 10 amino acids in *Tb*ZFP2, which contains an E3 ubiquitin ligase motif, caused total ablation of the nozzle phenotype (Hendriks *et al.*, 2001). In contrast, over-expression of *Tb*ZFP1 in procyclic forms showed no such phenotype, whereas ectopic expression in bloodstream trypanosomes could not be achieved (Hendriks and Matthews, 2005). This implies that *Tb*ZFP1 is posttranscriptionally regulated in such a way as to prevent protein expression.

When trypanosomes expressing *Tb*ZFP3 ectopically were induced to differentiate from bloodstream to procyclic forms, differentiation was enhanced, with twice as many cells expressing procyclin as the uninduced control samples (Paterou *et al.*, 2006). Conversely, when *Tb*ZFP2 was ablated by RNA interference, differentiation from bloodstream to procyclic cells was significantly compromised; with cells unable to undergo kinetoplast repositioning, morphological restructuring or other differentiation associated events (Hendriks *et al.*, 2001). Similarly, a bloodstream form null mutant of *Tb*ZFP1 displayed a compromised differentiation phenotype, whereby the kinetoplast did not undergo migration towards the nucleus and there was no outgrowth of the posterior end; instead the kinetoplast remained terminal at the posterior end (Hendriks and Matthews, 2005). As differentiation follows a specific

temporal pattern and the trypanosomes had exchanged their surface antigens and entered the cell cycle, it suggests that cell cycle progression is possible in the absence of kinetoplast repositioning (Hendriks and Matthews, 2005). This provides further evidence that these novel, kinetoplastid-specific molecules are potential regulators of trypanosome differentiation.

Recently, *Tb*ZFP2 and 3 have been shown to be associated with polyribosomes in procyclic forms, but not bloodstream or stumpy forms. The association with polysomes by *Tb*ZFP2 and 3 is lost if either of the CCCH RNA binding or WW protein interaction domains are mutated or deleted (Paterou *et al.*, 2006). Additionally, data recently published investigating the interactions of the homologous *Trypanosoma cruzi* Zinc Finger Protein (*Tc*ZFP) family suggests that *Tc*ZFP1 A and B interact with *Tc*ZFP2A (orthologue of *Tb*ZFP2) and *Tc*ZFP2B (orthologue of *Tb*ZFP3) (Caro *et al.*, 2005). It is also proposed that *Tc*ZFP2A and *Tc*ZFP2B can homodimerise, this being revealed using the yeast two-hybrid system (Caro *et al.*, 2005).

Recent RNA:protein binding studies on the *Trypanosoma cruzi* homologue of *Tb*ZFP1 (*Tc*ZFP1) have also shown a preference for poly-C oligonucleotide sequences, though binding was also observed with the U-rich 26-mer identified in the 3' UTR of EP/GPEET (Hotz *et al.*, 1997; Morking *et al.*, 2004). However, no direct association was observed with ARE element. Clearly, identification of the RNA targets for these proteins is a vital part of understanding the functions and role of these molecules in differentiation in the trypanosome.

1.11 The role of CCCH RNA binding zinc finger proteins in other eukaryotes.

The CCCH motif has been best characterised in the tristetraprolin (TTP) protein family, which is found in mammals. The cytosolic TTP protein has been shown to destabilise the mRNA of tumour necrosis factor α (TNF-α) and granulocyte-macrophages colony stimulating factor (GM-CSF) by binding to class II AU-rich elements within the mRNA (Blackshear, 2002; Carballo *et al.*, 1998; Carballo *et al.*, 2000; Lai *et al.*, 1999). Specifically, TTP binds to the UUAUUUAUU RNA sequence and it has been proposed that each individual CCCH zinc finger in the protein (at least two are always found in tandem) binds to a UAUU motif (Blackshear *et al.*, 2003; Brown, 2005; Worthington *et al.*, 2002). This binding then stimulates mRNA degradation by promoting deadenylation of the TNF-α mRNA which results in decreased levels of the encoded protein (Carballo *et al.*, 2000; Lai *et al.*, 1999). Importantly, two copies of the CCCH motif are located within the TTP protein and the presence of two or more CCCH motifs is the precedent for higher eukaryotes (Blackshear, 2002). It is highly unusual for only one copy of the CCCH motif to be present in a protein, as occurs in the *Tb*ZFP family (Brown, 2005).

As well as affecting mRNA stability, proteins containing the CCCH motif are involved in a wide range of RNA processing events. The polyadenylation and cleavage of RNA transcripts is undertaken by a protein complex called the cleavage and polyadenylation specificity factor (CPSF) (Keller *et al.*, 1991). The 30kDa subunit, known as CPSF30, contains five CCCH zinc finger motifs and is considered

essential for the effective cleavage and polyadenylation of the RNA target (Barabino *et al.*, 1997). CPSF30 shows significant homology to a protein named YHT1, with mutants lacking the YHT1 in *S. cerevisiae* being unable to polyadenylate cleavage products (Barabino *et al.*, 1997). Further dissection of this complex has dissected the domains and functions of the CPSF30 homologue, highlighting the interactions between the RNA and cleavage and polyadenylation complex (Tacahashi *et al.*, 2003). Recently, the *T. brucei* orthologue of CPSF30 was identified and was characterised as performing a homologous function, although the trypanosome protein could not complement a YHT1 yeast mutant (Hendriks *et al.*, 2003).

Other examples of CCCH motif-containing proteins include the components of the *Caenorhabditis elegans* 'germ plasm', PIE-1, POS-1, MEX-1, MEX-5 and MEX-6 (Reese *et al.*, 2000; Schubert *et al.*, 2000; Seydoux *et al.*, 1996; Tabara *et al.*, 1999). These proteins are involved in the asymmetric segregation of specialized cytoplasm during early development (Houston and King, 2000). Again, the role of these proteins is implied in RNA regulation and all of these molecules contain two zinc fingers. Another molecule, Zona binding inhibitory factor-1 (ZIF-1), has been shown to target the components of the 'germ plasm', such as PIE-1, for degradation. ZIF-1 also contains a CCCH motif and it is believed that the mechanism it uses for degradation relies on it recognising the specific CCCH motif present in these molecules (DeRenzo *et al.*, 2003). However ZIF-1 requires other factors to form a complex which is able to degrade CCCH fingers *in vivo*, one of which is the E3 ubiquitin ligase complex (DeRenzo *et al.*, 2003). Whether this process has any implications in trypanosomes remain to be seen.

1.12 The WW and PY protein interaction domains.

Members of the *Tb*ZFP family also contain WW (*Tb*ZFP2, 3) or PY domains (*Tb*ZFP1). The WW domain, one of the smallest known protein interaction motifs, is approximately 40 amino acids long and is characterized by the presence of two tryptophan residues which are spaced between 20 to 22 amino acids apart (Bork and Sudol, 1994; Macias *et al.*, 1996). This forms a stable triple beta sheet structure containing a hydrophobic pocket, which is able to recognise proline rich sequences (Macias *et al.*, 1996). The WW domains have been classified into four groups based on their preferences for ligand binding. The largest group, Group I, recognises the Proline-Proline- $X_{(\text{any amino acid})}$ -Tyrosine amino acid sequence and is known as the PY motif (Hu *et al.*, 2004; Sudol *et al.*, 1995). Group II bind ligands with Proline-Proline-Leucine-Proline (PPLP) motifs, while Group III have a preference for proline motifs flanked by serine or Arginine (PP-R)(Hu *et al.*, 2004; Sudol *et al.*, 1995). Group IV WW domains bind ligands containing phospho-Serine-proline or phospho-Threonine-Proline (p-S)P or (p-T)P (Hu *et al.*, 2004; Sudol *et al.*, 1995).

The WW domain has been identified in a wide range of proteins, especially those involved in signalling, regulatory and structural pathways (Chen and Sudol, 1995). It is found in most eukaryotic species and has been directly implicated in disease ranging from Alzheimer's to cancer and even in latent Epstein-Barr virus infections (Sudol, 1996). WW domains have been identified as being an important factor in Liddle's syndrome, where mutation of one of the critical PPxY residues prevents binding to neural precursor cell expressed developmentally down-regulated 4 (Nedd-

4). Interestingly, Nedd-4 also contains an ubiquitin ligase domain, which bears resemblance to the motif identified in *Tb*ZFP2 (Hendriks *et al.*, 2001; Schild *et al.*, 1996). PPxY domains have also been observed in a number of transcription factors, such as Nuclear factor, erythroid-derived 2 (NF-E2), where it may have a role in transcriptional activation as deletion of one of the CCCH motifs inhibits the molecule's ability to promote gene expression (Mosser *et al.*, 1998). As well as activating transcription, the interaction of the WW and PPxY domains may also have a role in inhibiting transcriptional activation, as is postulated for PQBP-1 (Waragai *et al.*, 1999).

1.13 The RGG box.

In addition to the CCCH motif, *Tb*ZFP3 (and in degenerate form, *Tb*ZFP2) contains another RNA binding motif, the RGG box. The RGG box was initially characterised in hnRNP and consisted of a series of Arginine-Glycine-Glycine (RGG) amino acid repeats (Kiledjian and Dreyfuss, 1992). The number of repeats varies from six in heterogeneous nuclear ribonucleoprotein (hnRNP) to 18 in the yeast GAR1 protein, though interestingly, *Tb*ZFP3 only contains 3 repeats (Burd and Dreyfuss, 1994) (Paterou *et al.*, 2006). The RGG domain usually occurs in combination with other RNA binding motifs and it has been postulated that the function of the RGG box was to bind non-specific sequences and hence increase the overall RNA binding affinity of the protein (Ghisolfi *et al.*, 1992). However, a recent RNA selection procedure has shown that the RGG box recognises G-quartet structures specifically within mRNAs bound by the fragile X mental retardation protein (FMRP) (Darnell *et al.*, 2001; Ramos *et al.*, 2003; Schaeffer *et al.*, 2001). This suggests that there is a specific target recognised by the RGG motif. In the case of *Tb*ZFP3, however, the presence of an RGG motif in the absence of the CCCH domain was not able to sustain the association with the polysomes (Paterou *et al.*, 2006).

As well as RNA, the RGG motif has also been implicated in protein binding. A ribonuclearprotein, RAP55, which contains two RGG motifs, has shown potential binding to the polyadenylation binding protein (PABP). As no evidence for RNA binding was found, the implication was that RAP55 was interacting directly with the polyadenylation binding protein (Lieb *et al.*, 1998). The RGG domain has also been

shown to be the substrate for acetyltransferase activity in the hnRNP A2 protein. Normally, the hnRNP A2 protein is localised in the nucleus, though if acetyltransferase activity is suppressed, the protein is localised to the cytoplasm (Nichols *et al.*, 2000). In addition, it has been speculated that the RGG domain may be involved in the homodimerisation of *Tc*ZFP3, though the significance of this *in vivo* remains to be confirmed (Caro *et al.*, 2005).

1.14 Aims and objectives.

Due to its complex lifecycle, the African trypanosome requires exquisite control of its gene expression to ensure that each stage specific molecule is expressed within the correct lifecycle stage. However, unlike the majority of higher eukaryotes, the trypanosome has evolved to regulate gene expression almost exclusively at the posttranscriptional level. Hence, RNA stability mechanisms and the proteins associated with these pathways are likely to be of considerable importance to the trypanosome. Thus, the identification of a novel family of small RNA binding proteins (*Tb*ZFP1, 2 and 3) implicated in the differentiation process from bloodstream to procyclic forms is of considerable importance. Subsequent work has identified the *in vivo* phenotypes associated with each molecule, though the pathways through which these molecules act has yet to be established. The aims of this project were to identify which RNA transcripts and protein molecules interact with the *Tb*ZFP family of molecules in *T. brucei*.

Specific aims of the work were as follows:

- To investigate if *Tb*ZFP2 affects the abundance of RNA transcripts by undertaking a genome wide screen for regulated transcripts.
- Identification of the RNA binding motif bound by *Tb*ZFP3 using a systematic enrichment of ligands by exponential enrichment (SELEX) approach.
- To investigate potential protein interactions between *Tb*ZFP1, 2 and 3 using the Yeast two-hybrid system.

- To identify novel proteins which interact directly with members of the *Tb*ZFP family by undertaking a Yeast two-hybrid screen of a *T. brucei* cDNA library.

Chapter 2:

Materials and Methods.

2.1 Trypanosome strains.

The bloodstream and procyclic form trypanosomes that were used in this thesis are presented below.

BSF SMB (S16)

Trypanosoma brucei brucei Lister s427 cell line, expressing the tetracycline repressor and T7 RNA polymerase, a kind gift of Professor George Cross, Rockefeller University, USA (Wirtz *et al.*, 1999).

PCF

Trypanosoma brucei brucei s427 cell line, transfected with pHD449 expressing the tetracycline repressor (Biebinger *et al.*, 1997).

2.2 Bacterial strains.

The bacterial strains used in this thesis are presented below.

XL1 Blue (Stratagene, Cambridge, UK)

E. coli: *rec*A1 *end*A1 *gyf*A96 *thi*-1 *hsd*A17 *sup*E44 *rel*A1 *lac*[F' *pro*AB *lac* $1^{9}\Delta$M15Tn(Tet^{r})]

JM109 (DE3) (Sigma-Aldrich, UK)

E. coli: e14 (McrA) *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17(r_k-m_k+) *sup*E44 *rel*A1 Δ(*lac-pro*AB) [F' *tra*D36 *pro*AB *lac* 1^{9}ZΔM15]

ElectroMAX DH10B cells (Invitrogen, UK)

E. coli: F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL nupG

2.3 Yeast strains.

The yeast strains used in this thesis are listed below.

YRG-2 (Stratagene, Cambridge, UK)

MATa *ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3 112 gal4-542 gal80-538* LYS2::UAS_{GAL1}-$TATA_{GAL1}$-HIS3 URA3::$_{UASGAL4\ 17mers(x3)}$-$TATA_{CYC1}$-lacZ3,, GAL1::LacZ, $HIS3_{UAS\ GAL1}$::HIS3@LYS2, $can1^R$, $cyh2^R$

AH109 (Clontech, France)

MATa, *trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ,* LYS::$GAL1_{UAS}$-$GAL1_{TATA}$-HIS3, MEL1 $GAL2_{UAS}$-$GAL2_{TATA}$-ADE2, URA3:: $MEL1_{UAS}$-$MEL1_{TATA}$-lacZ

L40 (A kind gift of Professor Rolf Sternglanz, Stony Brook University, USA)

MATa, *trp1-901, leu2-3, 112, his3-200 ade2* LYS:(*4LexAop* - HIS3), URA3:: *(8LexAop – lacZ)* GAL4

2.4 Chemicals and reagents.

All reagents used within this project were supplied by Sigma-Aldrich (Poole, UK), ROCHE Applied Science (Falcon way, Hertfordshire, UK), VWR International Ltd (Magna Park, Leicester, UK), GE Healthcare (Amersham Place, Buckinghamshire, UK), Fisher Scientific UK (Loughborough, UK) and Greiner Bio-One Ltd (Brunel Way, Gloucestershire, UK).

Unless otherwise stated, all chemicals are analytical grade and supplied by Sigma-Aldrich (Poole, UK) and all restriction enzymes and DNA Taq Polymerase were from Promega (Chilworth Science Park, UK). All disposable plastic ware was obtained from Greiner Bio-One Ltd (Brunel Way Gloucestershire, UK). A detailed description of specific materials is included in each section.

2.5 Antibodies.

Antibody	Source	Animal source	Application	Dilution
Anti-*Tb*ZFP2	(Hendriks *et al.*, 2000)	Rabbit, polyclonal	Western blotting	1:100
Anti-*Tb*ZFP3	(Paterou, 2006)	Rabbit, polyclonal	Western blotting	1:500
Anti-His tag	GE Healthcare	Mouse	Western blotting	1:3000
Anti-c-Myc tag	Clontech	Mouse	Western blotting	1:5000
Anti-mouse IgG (whole molecule) Peroxidase Conjugate	Sigma	Goat	Western blotting	1:5000
Anti-rabbit IgG (whole molecule) Peroxidase Conjugate	Sigma	Goat	Western blotting	1:5000

Table 2.1 - List of antibodies used within this thesis.

2.6 Oligonucleotides.

Name	Sequence	Application
5' EcoR 1 ZFP1	5' AAA **GAA TTC** ATG CAC ATT TCA ACA CCC 3'	Anneals at the 5' end of the *Tb*ZFP1 ORF and inserts the *Eco*R1 site upstream of START codon.
5' EcoR 1 ZFP2	5' AAA **GAA TTC** ATG GCC TTC AAC CAA 3'	Anneals at the 5' end of the *Tb*ZFP2 ORF and inserts the *Eco*R1 site upstream of START codon.
5' EcoR 1 ZFP3	5' AAA **GAA TTC** ATG CAG GGC TAT TTT 3'	Anneals at the 5' end of the *Tb*ZFP3 ORF and inserts the *Eco*R1 site upstream of START codon.
3' Sal 1 ZFP1	5' GTA G**GT CGA C**GT CAA TCC TCC AAC CTT AC 3'	Anneals at the 3' end of the *Tb*ZFP1 ORF and inserts the *Sal*1 site downstream of STOP codon.
3' Sal 1 ZFP2	5' GTA G**GT CGA C**GC TAC TGC TGC AGA TGG TT 3'	Anneals at the 3' end of the *Tb*ZFP2 ORF and inserts the *Sal*1 site downstream of STOP codon.
3' Sal 1 ZFP3	5' AAA G**TC GAC G**TT ATG CCA TGG GCG G 3'	Anneals at the 3' end of the *Tb*ZFP3 ORF and inserts the *Sal*1 site downstream of STOP codon.
3' Xho 1 ZFP1	5' AAA **CTC GAG** TTC AAT GGT GGA ACC TTA C 3'	Anneals at the 3' end of the *Tb*ZFP1 ORF and inserts the *Xho*1 site downstream of STOP codon.
3' Xho 1 ZFP2	5' AAA **CTC GAG** TCT ACT GCT GCA GAT GGT 3'	Anneals at the 3' end of the *Tb*ZFP2 ORF and inserts the *Xho*1 site downstream of STOP codon.
3' Xho 1 ZFP3	5' AAA **CTC GAG** TTT ATG CCA TGG GCG 3'	Anneals at the 3' end of the *Tb*ZFP3 ORF and inserts the *Xho*1 site downstream of STOP codon.
5' Nco 1 ZFP1	5' CAT G**CC ATG G**AG ATC GAC ATT TCA ACA CCC 3'	Anneals at the 5' end of the *Tb*ZFP1 ORF and inserts the *Nco*1 site upstream of START codon.
5' Nco 1 ZFP2	5' CAT A**CC ATG G**AG ATG	Anneals at the 5' end of the

	GCC TTC AAC CAA CGC TTT 3'	*Tb*ZFP2 ORF and inserts the *Nco* 1 site upstream of START codon.
Splice Leader	5' CTA TTA TTA GAA CAG TTT CTG TAC TAT ATT G 3'	Anneals to the 5' Splice leader sequence.
Oligo dT	5' CCC TTT TTT TTT TTT 3'	Anneals to the 3' polyA tail.
pGAD10 Sequencing primer	5' TAC CAC TAC AAT GGA TG 3'	Anneals at the 3' end of the GAL4 AD ORF, for sequencing DNA inserts.
5' pGAD10 Insert F	5' CCA CTG TCA CCT GGT TGG ACG G 3'	Anneals at the 3' end of the GAL4 AD ORF.
3' pGAD10 Insert R	5' TGC ACA GTT GAA GTG AAC TTG CGG GG 3'	Anneals at the 3' end of the MCS.
SELEX core	5' GGG AAG ATC TCG ACC AGA AG N50 TAT GTG CGT CTA CAT GGA TCC TCA 3'	The random 50-nucleotide core flanked by a 5' and 3' constant sequence.
5' SELEX	5' CGG *ATG TAA TAC GAC TCA CTA TAG* GGA AGA TCT CGA CCA GAA G 3'	Anneals at the 5' end to the constant region and addition of a T7 RNA Polymerase initiation site (shown in italics)
3' SELEX	5' TGA GGA TCC ATG TAG ACG CAC ATA 3'	Anneals at the 3' end to the constant region.
5' CLONE 6 MACRO	5' CCC **AAG CTT** ATG CAG CGG GAA GTG ATA CG 3'	5' Insertion of a *Hind* III site upstream of the START codon in *Tb*.10.61.1200.
5' CLONE 8 MACRO	5' CCC **AAG CTT** ATG TAC TGT CAC GCT GGC G 3'	5' Insertion of a *Hind* III site upstream of the START codon in *Tb*.10.389.0680.
5' CLONE 9 MURF	5' CCC **AAG CTT** TAT ATA AAT TGA TAA TAA CTG T 3'	5' Insertion of a *Hind* III site upstream of the START codon in MURF.
5' CLONE 9 CO1	5' CCC **AAG CTT** ATG TCT TGT GTG CTT AAG 3'	5' Insertion of a *Hind* III site upstream of the START codon in Cox I.
5' CLONE 10 N19B2.075	5' CCC **AAG CTT** ATG AAT AAG AAC GGC GGT TGT GTG G 3'	5' Insertion of a *Hind* III site upstream of the START codon in N19B2.075.
5' CLONE 10 N19B2.080	5' CCC **AAG CTT** ATG AAA ACA AAA CGT TAC AAT G 3'	5' Insertion of a *Hind* III site upstream of the START codon in N19B2.080.
5' CLONE 10 N19B2.085	5' CCC **AAG CTT** TTG CTT CTT TGC CAT CCG TGG 3'	5' Insertion of a *Hind* III site upstream of the START codon in N19B2.085.
3' CLONE 6 MACRO	5' CCC **GGA TCC** TTA CAT TGC ACT GAG CCA CCC 3'	3' Insertion of a *Bam*H 1 site downstream of STOP

		codon in *Tb*.10.61.1200.
3' CLONE 8 MACRO	5' CCC **GGA TCC** TCA CTC GTT ATT CTC CAT ATT GG 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in *Tb*.10.389.0680.
3' CLONE 9 MURF	5' CCC **GGA TCC** GCT TGG CTT TGA TTG AGT CGT G 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in MURF.
3' CLONE 9 CO1	5' CCC **GGA TCC** TTG ATA TAT TCG GAT CAT TA 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in Cox I.
3' CLONE 10 N19B2.075	5' CCC **GGA TCC** TCA TCT CCT TCC ACG TCG 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in N19B2.075.
5' CLONE 10 N19B2.080	5' CCC **GGA TCC** TTA AAG AGG AGG GAA CCA ACC 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in N19B2.080.
5' CLONE 10 N19B2.085	5' CCC **GGA TCC** TTA CCC ACC GAC AAA TAA TTG 3'	3' Insertion of a *Bam*H 1 site downstream of STOP codon in N19B2.085.

Table 2.2 - List of oligonucleotides used within this thesis and their function.

All oligonucleotides were synthesised by Sigma-genosys (Poole, UK). Restriction enzyme sites present within the oligonucleotides are highlighted in bold.

2.7 Plasmids.

The plasmids used within this thesis are listed below.

pGEM-T Easy (Promega, UK)

Used for cloning of PCR products, generation of riboprobes and DNA sequencing.

pBluescript SK+/- (Stratagene, UK)

Used for cloning PCR products and riboprobe manufacture.

p2T7i

Trypanosome RNAi vector containing two opposing T7 promoters modified by inclusion of the tetracycline operator sequence. A kind gift from Professor J. Donelson (LaCount *et al.*, 2002).

pGAD10 (Clontech, France)

Yeast two-hybrid library vector that creates a fusion protein containing the GAL4 activation domain fused to the 5' end of the MCS. The plasmid is selected in yeast using the LEU2 nutritional marker and in *E. coli* using the *bla* gene, which confers ampicillin resistance.

pGADT7 (Clontech, France)

Yeast two-hybrid vector that creates a fusion protein containing the GAL4 activation domain fused to the 5' end of the MCS. Included in the MCS is the HA epitope Tag. The plasmid is selected in yeast using the LEU2 nutritional marker and in *E. coli* using the *bla* gene that confers ampicillin resistance.

pGBKT7 (Clontech, France)

Yeast two-hybrid vector that creates a fusion protein containing amino acids 1-147 of the GAL4 DNA binding domain fused to the 5' end of the MCS. The MCS also contains the c-Myc epitope tag. Selection in yeast is using the TRP1 nutritional marker and kanamycin resistance in *E. coli*.

pSTT91 (based on pBTM116 (Sutton *et al.*, 2001))

Yeast two-hybrid vector that forms a fusion protein with the LexA DNA repeats fused to the 5' end of the MCS. The plasmid contains the TRP1 and ADE2 yeast nutritional markers and is selected for in *E. coli* using ampicillin resistance. A kind gift of Professor Rolf Sternglanz, Stony Brook University, USA.

pBD-GAL4 Cam (Stratagene, UK)

Yeast two-hybrid vector containing the GAL4 DNA binding domain fused to the 5' end of the MCS. Selection in yeast is using the TRP1 nutritional marker and chloroamphenicol resistance in *E. coli*.

pAD-GAL4-2.1 (Stratagene, UK)

Yeast two-hybrid vector that creates a fusion protein containing the GAL4 activation domains fused to the 5' end of the MCS. Selection in yeast is using the LEU2 nutritional marker and in *E. coli* using the *bla* gene, which confers ampicillin resistance.

2.8 Buffers and Media

Name	Constituents
10% RNA DIG Blocking solution	10% RNA DIG Block dissolved in Maleic acid buffer.
6 x DNA loading buffer	0.25% Bromophenol blue, 0.25% Xylene cyanol FF, 15% Ficoll.
Blocking Buffer	A 1:10 dilution of liquid block with H_2O
Buffer A (SELEX)	25mM Hepes, 150mM NaCl.
Buffer A	100m Tris, 300mM NaCl
Buffer B	25mM Hepes (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 150mM NaCl, 1M Imidazole.
Complete Cracking Buffer	1ml Cracking Buffer stock solution, 10µl β-mercaptoethanol, 50µl PMSF.
Coomassie Blue Stain	50% Methanol, 10% Acetic acid, 40% H_20, 2% Coomassie brilliant blue R250.
Cracking Buffer stock solution	8M Urea, 5% w/v SDS, 40mM Tris-HCl (pH 6.8), 0.1mM EDTA, 0.4mg/ml Bromophenol blue.
Destain solution	40% Methanol, 10% Acetic acid, 50% H_20.
Detection Buffer	100mM Tris-HCl (pH 9.5), 100mM

	NaCl.
Hybridisation Buffer (RNA)	5 x SSC (0.75M NaCl, 0.075M trisodium citrate), 50% Formamide, 0.02% SDS, 2% Blocking solution.
Hybridisation buffer	5% dextran sulphate, 5 x SSC, 10 fold dilution of liquid block, 0.1% SDS.
Hydration Buffer	500mM Ammonium acetate (pH 6), 0.1% SDS, 10mM Magnesium acetate, 0.1mM EDTA (pH 8).
LB (Luria-Bertani) media	10 g Tryptone, 5 g Yeast Extract, 5 g NaCl was dissolved in 1L H_2O and autoclaved.
Laemmli sample buffer	62.5mM Tris-HCl (pH 6.8), 2% SDS, 10% Glycerol, 2mg Bromophenol blue, 5% β-Mercaptoethanol.
Lysis buffer	25mM Hepes, 150mM NaCl, 0.1% Triton X 100, 5mM Imidazole.
Maleic acid buffer	100mM Maleic acid, 150mM NaCl, adjust to pH 7.5.
PBS (Phosphate buffer solution)	137mM NaCl, 3mM KCl, 7mM Na_2HP0_4, 1mM KH_2PO_4.
PEG/yeast transformation solution (Polyethylene glycol)	44% PEG, 0.1M LiAc, 0.1M TE

Ponceau stain	0.4% Ponceau S, 3% Trichloroacetic acid.
RF1 solution	10mM $RbCl_2$, 50mM $MnCl_24H_2O$, 30mM KOAc, 10mM $CaCl_2$, 10% w/v Glycerol, pH to 0.5M with 0.2M acetic acid.
RF2 solution	10mM $RbCl_2$, 10mM MOPS, 75mM $CaCl_2$, 10% w/v Glycerol, adjust to pH 6.8 with NaOH.
RNA gel loading buffer	30% Formamide, 16% Formaldehyde, 10% Glycerol, 0.01% Bromophenol blue, 1 x MOPS (40mM 3-(N-Morpholino) propanesulfonic acid), 33.4% H_20.
SDS-PAGE Lower gel buffer	1.5M Tris-HCl (pH 8.8), 0.4% SDS.
SDS-PAGE Running buffer [10x]	0.25M Tris-HCl (pH 8.3), 1.92M glycine, 1% SDS.
SDS-PAGE Upper gel buffer	0.5M Tris-HCl (pH 6.8), 0.4% SDS.
SELEX Binding Buffer.	0.5 M LiCl, 20 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$.
Solution I	50mM Glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0).
Solution II	1% SDS, 0.2M NaOH.
Solution III	5M KOAc, 11% (v/v) Glacial Acetic

	acid.
SSC [20x]	3M NaCl, 0.3M trisodium citrate.
TAE (Tris acetate EDTA)	40mM Tris-Acetate, 1mM EDTA.
TBE (Tris borate EDTA)	45mM Tris-borate, 1mM EDTA.
TBS-TNT (Tris-buffered saline)	20mM Tris-HCl (pH 7.4), 0.9% NaCl (w/w), 0.3% Tween 20, 0.2% NP40, 0.05% Triton X 100.
TE (Tris EDTA)	10mM Tris-HCl (pH 8.0), 1mM EDTA.
TELT Lysis buffer	50mM Tris-HCl (pH 8.0), 62.5mM EDTA, pH 9.0, 2.5M LiCl, 4% (v/v) Trition X 100.
Transfer buffer	25mM Tris (pH 8.3), 0.15M glycine, 20% (w/v) methanol 0.02% SDS.
Wash Buffer	0.3% Tween 20 dissolved in Maleic acid buffer.
Yeast transformation solution	0.1M TE, 0.1M LiAc
Z Buffer	60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$, 50mM β-mercaptoethanol.
Zimmerman post fusion medium plus glucose (ZPFMG)	132mM NaCl, 8mM KCl, 8mM Na_2HP0_4, 1.5mM KH_2PO_4, 0.5mM $MgOAc.4H_2O$, 90μM $CaOAc/Cl_2$, Adjust pH to 7.0 with acetic acid.

Table 2.3 - List of solutions and media used within this thesis

2.9 Cell manipulation.

2.9.1 Bloodstream form trypanosomes.

The SMB bloodstream form trypanosomes were cultured in HMI-9 (See Appendix 1) that was supplemented with 2.5µg/ml G418 and incubated at 37°C in an incubator supplemented with 5% CO_2 (Hirumi and Hirumi, 1989). The cultures were maintained at mid-log growth phase (approximately 1-2 x 10^6 cells/ml).

2.9.2 Procyclic form trypanosomes.

The 427 procyclic trypanosomes transfected with the pHD449 plasmid were maintained in SDM-79 (See Appendix 1) (Brun and Schonenberger, 1979) at 28°C and at a cell density of between 2-4 x 10^6 cells/ml.

2.10 DNA manipulations.

2.10.1 DNA extraction from trypanosomes.

Cells in mid-log growth phase (1-2 x 10^6 cells/ml) were centrifuged at 700g at 4°C for 10 mins and the pellet obtained was then resuspended in 300µl of TELT lysis buffer and incubated at room temperature for 5 mins. Next, 300µl of water-equilibrated phenol/chloroform was added and the tube gently mixed. The tube was then centrifuged at 13000g for 10 mins and the upper aqueous phase was removed to a clean tube. To this, 600µl of 100% ethanol was added and left at room temperature for 5 mins. The solution was then centrifuged at 13000g for 10 mins and the pellet washed with 70% ethanol. The pellet was left to air-dry for 10 mins before being resuspended in sterile water containing DNase-free RNase A (20µg/ml) and stored at 4°C.

2.10.2 Polymerase Chain Reaction amplification.

A 50µl PCR reaction consisted of 10% PCR buffer (supplied by Promega), 2.5mM dNTPs, 100ng DNA (plasmid or genomic), 100pM of 5' and 3' oligonucleotide (see Table 2.2 for primers used), 10U Taq DNA Polymerase (Promega), with the remaining volume made up with sterile water. PCR conditions used were: 95°C for 5 min, followed by 30 cycles of 95°C for 15s, 55°C for 15 sec and 72°C for 1 min, with a final cycle of 5 min at 72°C. PCR products were then purified using a Nucleospin column (AB gene) according to the manufacturers instructions.

2.10.3 DNA agarose gel electrophoresis.

DNA was visualised on a 100ml 1% TAE agarose gel to which 50µg of ethidium bromide was added. Samples to be loaded onto the gel were mixed with 6 x DNA loading buffer and the gel was run in 1 x TAE solution at 115 volts for approximately 30 mins to ensure good fragment separation.

2.10.4 Purification of DNA from PCR products and agarose gels.

To purify DNA amplified from PCR reactions and agarose gels, a Nucleospin column (AB Gene) was used according to the manufacturers instructions and the product eluted in distilled water (typically 30-50µl).

2.10.5 Digestion of DNA using restriction enzymes.

Each reaction contained 0.2 – 1µg DNA, 1 x restriction enzyme-specific reaction buffer and 5% (v/v) restriction enzyme (typically 10 units). The remaining volume was made up using nuclease free water. The reaction was then incubated at 37°C for a minimum of 1 hour.

2.10.6 Ligation of DNA inserts into plasmids.

The purified DNA was ligated into the appropriate vector using a 3:1 end molar ratio of insert: vector. The reaction volume contained 1 x ligase buffer (Roche), 1U T4 DNA ligase, insert and 100ng vector to give a total volume of 10µl, and was incubated at room temperature for 3 hours or 16°C overnight.

2.10.7 Competent bacterial cell preparation.

E. coli XL-1 Blue bacterial cells were inoculated into a 1ml culture and incubated at 37°C overnight. The following day, the cultures were inoculated into 100ml of LB broth and grown until OD_{600nm}=0.4 was obtained. The cells were harvested at 1000g for 15min at 15°C. The pellet was resuspended in 50ml cold RF1 and incubated on ice for one hour. The mixture was then centrifuged at 1000g for 5mins at 4°C and the pellet was resuspended in 18ml of RF2. The cells were then aliquoted, snap frozen and stored at -70°C.

2.10.8 Transformation of DNA into competent *E. coli* bacteria.

Approximately 2µl of the ligation reaction was incubated with 100µl of competent *E. coli* cells on ice for 25 mins. The mixture was then heat shocked at 42°C for 2 mins before being plated onto the appropriate selective LB agar media and incubated overnight at 37°C to allow colonies to develop.

2.10.9 Transformation of DNA by electroporation into competent *E. coli* bacteria.

To a microcentrifuge tube, 2µl DNA and 20µl DH10B *E. coli* cells (Invitrogen) were added and the mixture transferred into a chilled cuvette. The cells were then transformed using a Bio-Rad GenePulser and the following conditions: 2.0kV, 200Ω 25µF. 1ml of LB Broth was then added and the cells incubated at 37°C for one hour before being plated onto the appropriate selective LB media.

2.10.10 Small-scale preparation of DNA from bacterial cells.

A bacterial colony was inoculated into 2mls of LB broth containing the appropriate selective antibiotic and incubated at 37°C for 16 hours. The following day, 1ml of the culture was centrifuged at 10000g for 5 mins, and the pellet resuspended in 100µl of Solution I. This was incubated on ice for 5 mins, before the addition of 200µl of Solution II and another 5 mins incubation period on ice. Finally 150µl of solution III was added and the lysate left for 5 mins on ice. The solution was then centrifuged at 10000g for 10 mins and the supernatant promptly removed into a clean tube containing 900µl of 100% ethanol. This was incubated on ice for 15 mins, centrifuged for 10 mins and the pellet washed with 70% ethanol, before being air-dried for 10 mins. The pellet

was resuspended in 40µl sterile water containing DNase-free pancreatic RNase A (20µg/ml) and stored at -20°C.

2.10.11 Large-scale preparation of DNA from bacterial cells.

A 250ml LB broth culture containing the appropriate selective antibiotic was incubated with a 100µl aliquot from a small-scale bacterial culture and then incubated with agitation overnight. The following day, the culture was centrifuged at 3000g and the supernatant removed. The pellet was then processed according to the manufacturer's instructions for large-scale preparation of DNA using a Plasmid Maxi kit (Qiagen).

2.11 Trypanosome manipulations.

2.11.1 Stable transfection of Trypanosomes.

Approximately 1-2 x 10^7 trypanosomes, grown to mid-log growth phase, were centrifuged at 700g for 10 min and the pellet washed in half a volume of chilled ZPFMG. This was then centrifuged at 700g for 10 mins before being resuspended in 500µl ZPFMG. The cells were transferred into a 0.4cm electroporation cuvette along with 15µg of the linearised plasmid DNA and transfected with a BTX 830 Electroporator at 1700v, 100µs, 200ms for 3 pulses and the cells were then transferred into flasks containing HMI – 9 and incubated at 37°C with 5% CO_2.

After 24 hours growth, the cells were diluted to give 1 x 10^5 cells/ml in HMI-9 containing the appropriate antibiotic. The transfections were then analysed after 7-10 days for transfectants that were resistant to the selective antibiotic and transformants were diluted using a 1:2 dilution with fresh HMI-9. Once a sufficient concentration of cells was obtained (1 x 10^6 cells/ml), the cells were transferred to 25ml flasks and cultured as described previously (Section 2.8.1).

2.11.2 Protein extraction from Trypanosomes.

Trypanosomes were grown to the mid-log growth phase (1-2 x 10^6 cells/ml) and then $2x10^7$ cells were centrifuged at 700g for 10 mins and washed in 1 x PBS. The pellet was then resuspended in 100µl of boiling Laemmli buffer and the sample boiled for 10 mins. The sample was then stored at -20°C.

2.11.3 RNA extraction from Trypanosomes.

Small-scale preparations of RNA from trypanosomes were achieved from $1-2x10^7$ cells using an RNeasy mini kit (Qiagen) according to the manufacturer's instructions for 'animal cells in suspension'. Larger volumes of RNA were obtained from $1-4x10^8$ trypanosomes using an RNeasy midi kit (Qiagen) according to the manufacturer's instructions for 'animal cells in suspension'.

2.12 Protein protocols.

2.12.1 SDS-PAGE electrophoresis.

The protein samples were analysed on a 15% denaturing polyacrylamide gel run in 1 x running buffer at 150 volts for 70 mins (BioRad mini protean III electrophoresis apparatus). The gels were then used for Western blotting or stained with 50ml Coomassie brilliant blue and incubated with agitation for 30 mins. The gel was then destained in 50ml of destain solution and dried.

2.12.2 Western blotting.

The gel was incubated, along with nitrocellulose membrane and two blotting pads (Bio-Rad) in 1 x transfer buffer for 30 mins. Transfer was achieved using a Trans-Blot SD semi-dry Electrophoretic transfer cell (Bio-Rad) at 15 volts for 35 mins. Once completed, the membrane was incubated in Ponceau S red stain for 5 mins to ensure efficient transfer before being destained with distilled water.

The membrane was then blocked for a minimum of one hour in 5mls of 5% non-fat milk–PBS solution with agitation, before the membrane was hybridised for one hour with 5mls of 5% milk–PBS with the appropriate dilution of primary antibody. The membrane was then washed for 15 mins with 1 x PBS and twice more with TBS-TNT to remove any unbound antibody. Next, the secondary antibody conjugated to

horseradish peroxidase (HRP) was diluted in 5mls of 5% milk–PBS and left to incubate for one hour, before being washed once with 1 x PBS for 15 mins and then a further two times with of TBS-TNT for 15 mins. Detection of the proteins was performed using the ECL (Enhanced Chemiluminescence) system (GE Healthcare), according to the manufacturer's instructions, before exposing the membrane to blue sensitive X-Ray film (KODAK).

2.12.3 Expression of recombinant His-tag *Tb*ZFP3 in JM109 (DE3) *E. coli* cells.

A starter culture consisting of 10ml LB broth with 30µg/ml kanamycin was inoculated with a colony containing the pET30a-*Tb*ZFP3-N-His tag plasmid and incubated overnight at 37°C with agitation. The overnight culture was then inoculated into 500mls of LB containing 30µg/ml kanamycin and grown until an OD_{600}= 0.5-0.6 was obtained. At this cell density, the cells were induced to express the recombinant protein by the addition of IPTG to a final concentration of 1mM and incubated at 37°C for 3 hours. After this time, the cells were centrifuged at 1000g for 25 mins at 4°C, the supernatant removed and the pellet stored at -20°C.

2.12.4 Purifying recombinant His-tagged *Tb*ZFP3.

For FPLC purification, cells were resuspended in 1ml lysis buffer with the addition of 1 x protease inhibitor tablet (Roche). Cells were sonicated (6 x 30s pulses) and the insoluble cell debris pelleted by centrifugation at 12500g. The supernatant was applied to a 1 ml Ni HiTrap chelating HP column (GE Healthcare) which had first been charged with Ni^{2+} ions and equilibrated with Buffer A. The column was washed with Buffer A until the absorbance $O.D._{280nm} = 0$. The column was then washed with 15 column volumes (CVs) Buffer A. The bound protein was eluted using a linear 10mM to 500mM imidazole gradient and 1 ml fractions were collected. Protein concentration was determined using Bradford reagent (Bio-Rad) and the purity of the recombinant protein was checked on an SDS-PAGE gel.

2.13 RNA manipulations.

2.13.1 Preparation of Riboprobes.

A PCR reaction using the M13 Forward and Reverse primers was undertaken on either pGEM-T Easy or pBluescript that contained the DNA insert. This added either a T7, T3 or SP6 RNA promoter onto the PCR product. Once the resulting PCR had been purified using a Nucleospin column, an *in vitro* transcription reaction was performed using the RNA labelling kit (ROCHE)(the correct RNA promoter being selected to produce an

anti-sense probe). This incorporated digoxigenin ribonucleotides into the newly synthesised RNA transcript The reaction mix contained 1µg purified DNA, 2µl dNTP mix, 1µl RNase inhibitor, 1 x reaction buffer, 2µl of the appropriate RNA Polymerase (T7, T3 or SP6) and distilled water to a total reaction volume of 20µl.

The reaction was incubated at 37°C for 2 hours, before 2µl of DNase was added to digest the DNA template. After 15 minutes, 2µl of 0.2M EDTA was added to stop the reaction. The RNA was then ethanol precipitated using 2.5 volumes of 100% ethanol and 2.5µl of a 4M LiCl solution and this placed at -70°C for 30 mins. Next, the reaction was centrifuged at 12000g for 15 mins and the pellet washed with 70% cold ethanol. The pellet was air-dried for 10 mins, before being resuspended in 50µl RNase free water containing 20U RNase inhibitor and stored at -70°C

2.13.2 RNA agarose gel electrophoresis.

A 1.2% agarose gel was prepared containing 10% (v/v) MOPS and 3% (v/v) formaldehyde and placed in a tank containing 1 x MOPS. Each RNA sample was prepared by the addition of 2µl RNA loading dye, 2µl 10 x MOPS, 9µl formamide, 3µl formaldehyde and then heated at 65°C for 5 min prior to loading. The gel was then electrophoresied at 150V for 90 min on a 12cm flat bed agarose gel apparatus (BioRad) before being stained with 20µl ethidium bromide (10 mg/ml) in 200 ml of 1 x MOPS for 15 min. The gel was destained twice with water for 30 min before visualising on an U.V. illuminator.

2.13.3 Northern Blotting.

For Northern blotting, the gel, positively charged nylon membrane (ROCHE) and 2 extra thick blotting pads (Bio-Rad) were soaked for 30 min in 0.5 x TBE for blotting on a semi-dry blotter (Bio-Rad Trans-Blot SD). A pad was placed onto the blotter, on top of which the nylon membrane and then the gel were placed before finally the top pad. Any air bubbles were removed and the blotter set to a constant 200mA for 30 min. After UV cross-linking, the blot was ready to be hybridised. To hybridise, the blot was prehybridised at 68°C for one hour in 7mls hybridisation buffer. After this time, the hybridisation solution was exchanged and 2µl of riboprobe was added to the solution. The blot was then hybridised overnight at 68°C.

The following day, the blot was washed twice at 68°C for 30 mins with 2 x SSC/0.1% SDS, then once at 68°C with 0.5 x SSC/0.1% SDS. The blot was blocked in 50mls Maleic Acid buffer containing 1% RNA blocking solution (Roche) for one hour, then incubated with 50mls Maleic Acid buffer containing 1% RNA blocking solution and a 1:25000 dilution of Anti-digoxigenin antibody (Roche) for 30 mins. The blot was then washed 6 times with wash buffer for 5 mins before being soaked for 2 mins in detection buffer. Next, a 1:100 dilution of CDP-Star detection reagent (Roche) was applied to the blot and left for 2 mins, before being exposed to blue sensitive autoradiography X-Ray film (Kodak).

2.14 Yeast Manipulations.

2.14.1 Transformation of DNA into Yeast.

A 50ml culture of YPD or appropriate SD dropout media was inoculated with a medium sized yeast colony and incubated at 30°C overnight. The following day the culture was diluted to OD_{600}=0.1 in 300mls YPD or SD media (or 600mls SD media for library screens) and incubated for 3 hours until OD_{600}=0.6 was reached, when the cells were centrifuged at 1000g for 5 mins. The cell pellet was washed with half a volume of yeast transformation solution and the pellet resuspended in 1ml yeast transformation solution (or 2mls for library screens). In a round-bottomed tube, 100µl of the yeast cell suspension, 55µg salmon sperm carrier DNA, 100ng of each plasmid and 1.2 ml PEG/yeast transformation solution were mixed. For library screens, 2ml of the yeast cell suspension, 110µl salmon sperm carrier DNA, 20µl of each plasmid and 24ml 44% PEG/ 0.1M LiAc/ 0.1M TE were mixed. The solution was then incubated at 30°C for 1 hour. After this, 100µl (or 1ml) DMSO was added and the tube heat shocked at 42°C for 15 mins. The solution was then centrifuged, the supernatant removed and the pellet resuspended in 1ml YPD for 1 hour at 30°C, before being plated on the appropriate SD media.

2.14.2 Mini-scale preparation of DNA from Yeast.

From a 5ml overnight culture grown in the appropriate SD media, 1.5ml of yeast was centrifuged at 12000g for 5 mins and the supernatant removed. The pellet was resuspended in 0.2 ml lysis buffer, 0.3mls glass beads and 0.2mls of phenol chloroform. This was vortexed for 2 mins, before the addition of 0.2ml of 1 x TE and then centrifuged at 12000g for 5 mins. Next, the aqueous phase was transferred to a clean tube and 2 volumes of 100% ethanol were added. The mixture was then centrifuged again and the pellet washed with 70% ethanol, before being air-dried for 5 mins. The pellet was then resuspended in 10μl water.

2.14.3 Protein extraction from Yeast.

A 5ml culture containing the yeast in the appropriate SD media was incubated at 30°C overnight, before being centrifuged at 1000g for 5 mins. The pellet was then washed in ice-cold ethanol and resuspended in complete cracking buffer, which had been pre-warmed to 60°C. 100μl of complete cracking buffer was used per 7.5 OD_{600} units of yeast. The solution was then transferred to a centrifuge tube containing glass beads (80μl of beads per 7.5 OD_{600} units) and heated to 70°C for 10 mins. The tube was then vortexed for 1 min before centrifugation at 12000g for 5 mins. The supernatant was removed, briefly boiled and then loaded onto a SDS-PAGE gel (see Section 2.12.1) for analysis.

2.14.4 X-Gal assay for the production of β-Galactosidase.

The colonies to be assayed for β-galactosidase production were replica plated onto nitrocellulose membrane and placed on the surface of a YPD plate. The plate was then incubated at 30°C overnight. The next day, 10mg X-gal was dissolved in 100μl DMF and added to 10mls Z buffer and 60μl β-mercaptoethanol. This solution was used to soak two Whatman No.1 filter papers in a petri dish. Next, the nitrocellulose membrane was removed and immersed in liquid nitrogen for 20s, and then placed onto the pre-soaked filter paper, colony side up and incubated at 37°C for 7 hours. A blue colour change indicated a positive result.

2.15 Selection of Ligands by exponential enrichment (SELEX).

2.15.1 Selection of Ligands using His-tagged *Tb*ZFP3.

RNA selection was performed according to the method of Hori *et* al., (2005). A 94 base oligonucleotide, consisting of a central core of 50nt random sequence flanked by primer binding sites was synthesized (Sigma-genosys). 1nmol of this oligonucleotide was PCR amplified using a forward primer also containing a T7 promoter sequence (5' T7 SELEX), and a reverse primer (3' SELEX). Five reactions each with 6μg DNA were carried out. Each PCR reaction contained 2.5mM $MgCl_2$, 1mM each 5' T7 and 3' SELEX primers, 0.2mM dNTPs and 50 Units Taq polymerase (Promega) and consisted

of 10 cycles of 30s 94°C, 30s 59°C and 5s 72°C. 30µg library DNAs were transcribed *in vitro* with T7 RNA polymerase, the transcription reaction stopped by addition of DNase I, and the transcribed RNA ethanol precipitated. The purified RNA was then heat-denatured at 95°C for 1 min and pre-absorbed with Ni-NTA agarose resin (Qiagen) to remove nonspecifically binding RNAs from the pool. 2nmol RNA was then heat-denatured once again at 95°C for 1 min and mixed with 20µl Ni-NTA agarose beads prefilled with 200µg His-tagged *Tb*ZFP3 in 1 ml SELEX binding buffer. The binding reaction was incubated with mixing at 4°C for 60 min and washed five times in binding buffer. Agarose resin-protein-RNA complexes were extracted with phenol/chloroform and the RNA in the aqueous phase precipitated with ethanol. The selected RNA was reverse transcribed in a reaction containing 3' SELEX primer, 1mM dNTPs, 40U RNase inhibitor and 30U AMV Reverse Transcriptase (Promega, UK). The reaction mixture was incubated for 60 min at 42°C. PCR amplification of the DNA was repeated and the RT-PCR product used for the next round of the selection procedure.

2.15.2 Transcription of radiolabelled RNA probes.

Transcription of a probe was undertaken using DNA containing the T7 RNA polymerase binding site. The transcription reaction contained 2µl each of 10mM rATP, rGTP, rCTP, rUTP , 1 x transcription buffer, 1µl RNase inhibitor, 4µl DNA, 2µl RNA Polymerase and 5µl water. For radiolabelled RNA, 1µl 1mM rUTP and 5µl α ^{32}P-rUTP (GE Healthcare) was substituted for the 10mM rUTP and water. The reaction was then

incubated at 37°C for 2 hours before the addition of 2µl DNase (RNase free) and a further incubation for 15 minutes.

2.15.3 Purification of a radioactive RNA probe.

The RNA probe was then purified using a Urea based gel. This contained 1 x TBE 23.7g Urea, 12.5ml Acrylamide 40 % (29:1), 300µl APS, 30µl TEMED and 11.25ml water. An equal volume of RNA loading dye was added to the probe and the gel run at 1800v, 80W for 90 minutes in 0.5 x TBE.

The probe was located and excised using X-ray film, then the gel fragment incubated overnight in 600µl Hydration buffer. To this, two volumes of 100% ethanol was added, and incubated on ice for 30 minutes. The sample was then centrifuged at 12000g for 10 minutes before the pellet was washed with 70% ethanol and air dried for 10 minutes. The probes were then stored at -20°C or 10µl water added and used directly in an electrophoretic mobility shift assay.

2.15.4 Electrophoretic mobility shift assay (EMSA).

A native acrylamide gel was made containing 1 x TBE, 7.5ml 30% Acrylamide (37.5:1), 2.5ml, 20% Glycerol, 32.5ml water, 375µl APS and 37.5µl TEMED. The gel was pre-run for 30 minutes in 0.5 x TBE before loading and was then run at 45mA for 90 minutes.

Next, the gel was removed from the glass plates and placed on a piece of filter paper, before being covered with Saran wrap. This was then placed on a gel drier and dried for 3 hours. The dried gel was then exposed to blue sensitive autoradiography X-Ray film (KODAK).

2.16 Macroarray Methods.

2.16.1 PolyA selection of RNA and generation of cDNA probe.

RNA was isolated from trypanosomes according to Section 2.11.3. Next, the isolation of mRNA was undertaken using the PolyATtract mRNA isolation system IV (Promega), according to the manufacturer's protocol.

The isolated mRNA was then added to 100pmole Oligo dT, 1x Reverse transcription buffer, 5μl 0.1M DTT, 20U RNase inhibitor, 25U AMV Reverse Transcriptase (Promega), with the total reaction volume being made up with distilled water. The reaction was incubated at 42°C for 2 hours before the cDNA was cleaned up using a G50 Sephadex column (GE Healthcare) and ethanol precipitated.

The pellet was then resuspended in 30μl distilled water and labelled using the Gene Images Kit (GE Healthcare), which involved the addition of 5μl primer mix, 10μl of green labelling nucleotide mix and 1μl of Klenow Polymerase and incubation at 37°C for 1 hour. The reaction was stopped by the addition of 2μl 0.5M EDTA.

2.16.2 Hybridisation of cDNA to the *T. brucei* macroarray.

The blot was prehybridised in 15 ml of hybridisation buffer at 65°C for 2 hours. Next, 25 µl of probe was added to 75µl of hybridisation buffer and boiled for 10 minutes, while the hybridisation buffer was replaced with 15 ml of fresh pre-warmed hybridisation buffer and the boiled probe. The tube was then left to hybridize overnight at 65°C. The blots were then washed twice at 65°C in 2 x SSC / 0.1% SDS for 30 minutes before being washed in 0.1 x SSC / 0.1% SDS for another 30 minutes. The blot was gently agitated with 165 ml Blocking Buffer for 1 hour. Next, the anti-fluorescein-AP conjugate was diluted 1:25000 with Blocking Buffer and incubated for 1 hour. The unbound conjugate was removed by washing 5 times with 0.3% (w/v) Tween 20 in Buffer A for 10 minutes. Finally, 5 ml of CDP Star detection reagent was spread across the blot. The membrane was then placed against Blue Sensitive Autoradiography film (KODAK) and developed after the required exposure.

Chapter 3:

Does the ablation of *Tb*ZFP2 by RNAi affect the abundance of mRNA transcripts within *Trypanosoma brucei*?

3.1 Introduction.

All organisms adapt to changes in their environmental surroundings, usually by the regulation of gene expression. The most important method for eukaryotic organisms to regulate differential gene expression is via the initiation and control of transcription. However, trypanosomes are unusual, as they appear to rarely (if ever) regulate the initiation of transcription and almost exclusively rely on regulation of gene expression at the post-transcriptional level (Clayton, 2002). Indeed, the organisation of the *T. brucei* genome into polycistronic transcription units, where many functionally unrelated genes are transcribed together from a distant promoter, makes regulation of gene expression by transcription initiation highly unlikely (Berriman *et al.*, 2005). Also, despite extensive investigation, no defined promoters for RNA Polymerase II have been characterised for protein coding genes within the trypanosome genome (Clayton, 2002).

Therefore, it is likely that regulation of RNA stability at the post-transcriptional level is key to the control of gene expression within *T. brucei*. The result is that molecules that contain RNA binding motifs have the potential to act as trans-acting factors that can regulate gene expression. Several trans-acting proteins have been identified that can regulate mRNA stability and modify RNA abundance by targeting the AU rich elements (AREs) found in the 3' UTRs of mRNAs. In *T. cruzi*, over expression of *Tc*UBP1 (U-rich binding protein) has been shown to destabilise the small mucin mRNA (*Tc*SMUG) by an interaction with an ARE present in the 3' UTR (D'Orso and Frasch, 2001). Also,

ectopic expression of the human RNA binding protein (HuR) in trypanosomes has shown that it can bind to the ARE in the mRNA of phosphoglycerate kinase B (PGKB), leading to altered abundance and stabilisation of the mRNA transcript (Quijada *et al.*, 2002). Other examples of RNA binding proteins include *Tb*PUF1 (Pumilio/fem-3 binding protein), a member of the PUF family found throughout higher eukaryotes that cause mRNA instability and translational repression (Caro *et al.*, 2006). In trypanosomes, *Tb*PUF1 has been shown to affect the stability of Expression Site Associated Gene (ESAG) 8 mRNA, which resulted in retarded growth and differentiation of the parasite (Caro *et al.*, 2006; Hoek *et al.*, 2002).

Analysis of the *T. brucei* genome has also shown an unusually large number of predicted proteins containing the $C_{X8}C_{X5}C_{X3}H$ RNA binding motif. So far 65 putative proteins that contain the $C_{X8}C_{X5}C_{X3}H$ motif have been identified in *T. brucei*, in comparison to 12 in *S. pombe* (Ivens *et al.*, 2005). Proteins that contain the $C_{X8}C_{X5}C_{X3}H$ motif have been implicated in a variety of biological systems in higher eukaryotes, including RNA stability, mRNA localisation, translation control and differentiation (Houston and King, 2000; Lai *et al.*, 2000; Reese *et al.*, 2000). The prototype protein for the $C_{X8}C_{X5}C_{X3}H$ motif family is tristetraproline (TTP). TTP has been shown to destabilize mRNA transcripts such as tumour necrosis factor-α, by recognition of an ARE motif within the 3' UTR (Blackshear, 2002). Therefore, these putative $C_{X8}C_{X5}C_{X3}H$ containing proteins encoded within the trypanosome genome are likely to play an important role in the post-translational regulation of gene expression within the kinetoplastid family.

Although 65 putative $C_{X8}C_{X5}C_{X3}H$ proteins have been identified in trypanosomes, to date only six proteins containing the $C_{X8}C_{X5}C_{X3}H$ RNA binding motif have been characterised, these being *Tb*CPSF30, *Tb*CSBP A and B and *Tb*ZFP1, 2 and 3. *Tb*CPSF30 is a homologue of the cleavage and polyadenylation specificity factor found in higher eukaryotes and is involved in polycistronic RNA processing (Hendriks *et al.*, 2003). *Tb*CSBP A and B are cycling sequence binding proteins that bind to sequence elements in mRNAs that accumulate periodically during the cell cycle and were initially identified in *Crithidia fasciculata* (Mahmood *et al.*, 2001). The *Tb*ZFP family consists of three proteins, *Tb*ZFP1, 2 and 3, which possess the $C_{X8}C_{X5}C_{X3}H$ RNA binding motif and either the WW or complementary PY protein interaction motifs (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou, 2004). *Tb*ZFP1, 2 and 3 have all been implicated in differentiation from bloodstream to procyclic forms, although the mode of action of the *Tb*ZFP family has yet to be elucidated (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou, 2004).

The *Tb*ZFP2 RNA transcript has been effectively ablated in bloodstream forms using RNA interference (RNAi), with no obvious phenotype (Hendriks *et al.*, 2001). However, once trypanosomes lacking *Tb*ZFP2 are induced to differentiate to procyclic forms, the trypanosomes show impaired differentiation (Hendriks *et al.*, 2001). Hence, the presence of the $C_{X8}C_{X5}C_{X3}H$ motif and its role in RNA stability in other proteins and organisms, combined with the genomic organisation of the kinetoplastida and the large numbers of

RNA binding proteins, led to the hypothesis that *Tb*ZFP2 may regulate the abundance of RNA transcripts involved in differentiation to procyclic forms.

Therefore, it was decided to undertake a screen for mRNA transcripts whose abundances were regulated by *Tb*ZFP2. If *Tb*ZFP2 regulated mRNA transcripts, either through ARE motifs or within other regions of the 3' UTRs, then the mRNA expression profile should differ between trypanosomes with the *Tb*ZFP2 transcript depleted by RNAi and wild type cells. By comparing these cell lines using a *T. brucei* genome macroarray, RNA transcripts that are regulated by *Tb*ZFP2 should be identified.

In this chapter, work is presented that investigated if *Tb*ZFP2 was able to regulate the abundance of mRNA transcripts. This was achieved by the regeneration and verification of cell lines in which *Tb*ZFP2 was effectively ablated using RNAi. Two trypanosome cell lines (the SMB cell line containing wild type levels of the *Tb*ZFP2 protein and the induced 2T7i.*Tb*ZFP2 2.3 cell line lacking the *Tb*ZFP2 protein) were hybridised on two separate *T. brucei* genomic macroarrays and then compared to identify differences in mRNA abundance. Candidate genes that were isolated from the macroarray through showing a differential expression profile were then verified by northern blot to confirm that *Tb*ZFP2 regulated the abundance of the identified transcript. The results showed that *Tb*ZFP2 regulated none of the mRNA transcripts identified from the macroarray.

3.1 Ablation of the TbZFP2 RNA transcript using RNAi.

Ablation of the *Tb*ZFP2 RNA transcript using RNAi had previously been achieved, where a >90% decrease in the level of the protein was observed within bloodstream form trypanosomes (Hendriks *et al.*, 2001). The ZFP2.RNAi cell line was created by transfecting the SMB trypanosome line, which had been engineered to express the T7 RNA Polymerase and the tetracycline repressor, with the p2T7i.*Tb*ZFP2 RNAi plasmid. To verify the phenotype obtained and ensure reproducible and effective ablation of the *Tb*ZFP2 transcript, the ZFP2.RNAi cell line (a kind gift of Dr Ed Hendricks), was assayed for its ability to induce ablation of the *Tb*ZFP2 transcript. Induction of RNAi was achieved by the addition of 2.5μg tetracycline to the culture media of the ZFP2.RNAi trypanosome cell line for 72 hours. The parental SMB line was cultured in parallel as a control sample. RNA was then harvested using the RNeasy kit (Qiagen) and analysed on a northern blot using a riboprobe specific to *Tb*ZFP2 (Figure 3.1). SMB and ZFP2.RNAi trypanosomes lacking tetracycline were used as controls. The induced ZFP2.RNAi cell line showed effective ablation of the *Tb*ZFP2 transcript, in comparison to the uninduced ZFP2.RNAi, where the transcript was still present in abundance. No ablation was observed in the SMB parental strain, in the presence or absence of tetracycline. The level of *Tb*ZFP2 ablation achieved (>90%) was similar to that published by Hendriks *et al* (2001). The presence of a small molecular size band (Figure 3.1, indicated by the asterisk) is likely to be the double stranded RNA produced by the p2T7i.ZFP2 plasmid. Its presence within the induced and uninduced samples of the

ZFP2.RNAi cell line indicated that repression of the p2T7i.ZFP2 plasmid was incomplete in the uninduced samples and that even in the absence of tetracycline, a small amount of the endogenous transcript was probably ablated.

Ablation of *Tb*ZFP2 using the same ZFP2.RNAi cell line was repeated under identical conditions to ensure a reproducible result. However, the effectiveness of RNAi to ablate the *Tb*ZFP2 transcript appeared to have decreased (Figure 3.2, A). As before, the SMB line showed no ablation of the *Tb*ZFP2 transcript, but the induced ZFP2.RNAi cell line was only able to achieve a 50% decrease in the levels of the transcript when compared to the uninduced sample. Further repetition of this experiment again showed a similar partial ablation of the *Tb*ZFP2 transcript (data not shown). To confirm this was not a cell line specific phenomenon, an independent ZFP2.RNAi cell line created by Dr Ed Hendriks (a kind gift of Dr Ed Hendriks) was assayed for its ability to ablate the *Tb*ZFP2 transcript. Again, only a 50% ablation of the transcript was observed, when assayed by northern blot, in the induced ZFP2.RNAi cell line when compared to the uninduced ZFP2.RNAi cell line (Figure 3.2, B). Hence, it was decided that the ZFP2.RNAi cell lines generated by Hendriks *et al* were not suitable for use in analysing differences in the expression profile of mRNA transcripts affected by *Tb*ZFP2.

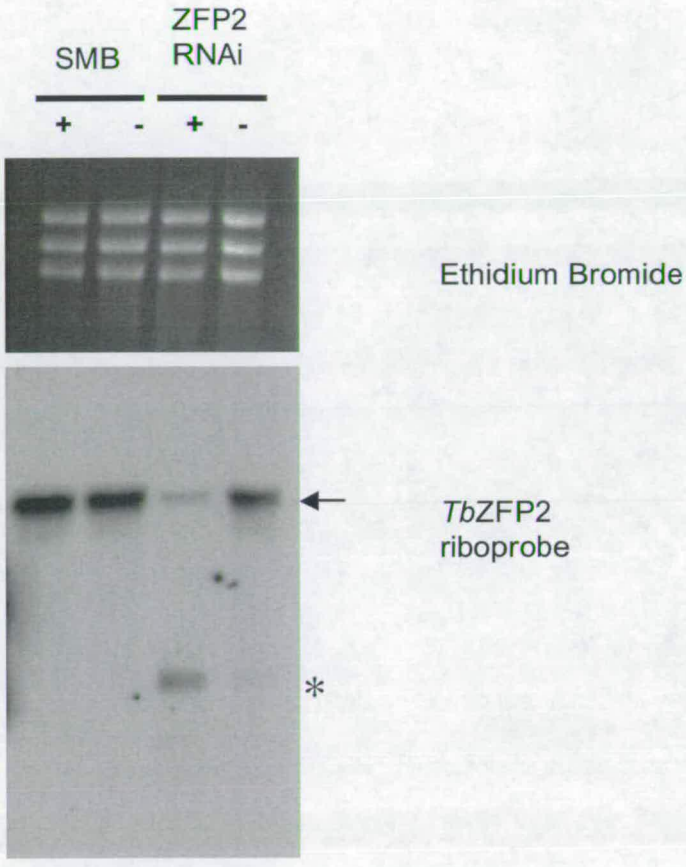

Figure 3.1 – RNAi ablation of the endogenous *Tb*ZFP2 transcript within bloodstream form *T. brucei*.

Northern analysis of the parental trypanosome line (SMB) transfected with the p2T7i.ZFP2 RNAi plasmid. The trypanosomes were then induced for 3 days with (+) or without (-) tetracycline to induce RNAi, before RNA samples were prepared and the northern blot analysed with a riboprobe specific to *Tb*ZFP2. The endogenous *Tb*ZFP2 is indicated by the arrow, with a significant decrease (~90%) in RNA levels observed in the induced p2T7i.ZFP2 lane. The asterisk indicates the double stranded *Tb*ZFP2 derived from the transfected construct. The ribosomal RNA is included as a loading control (Ethidium Bromide).

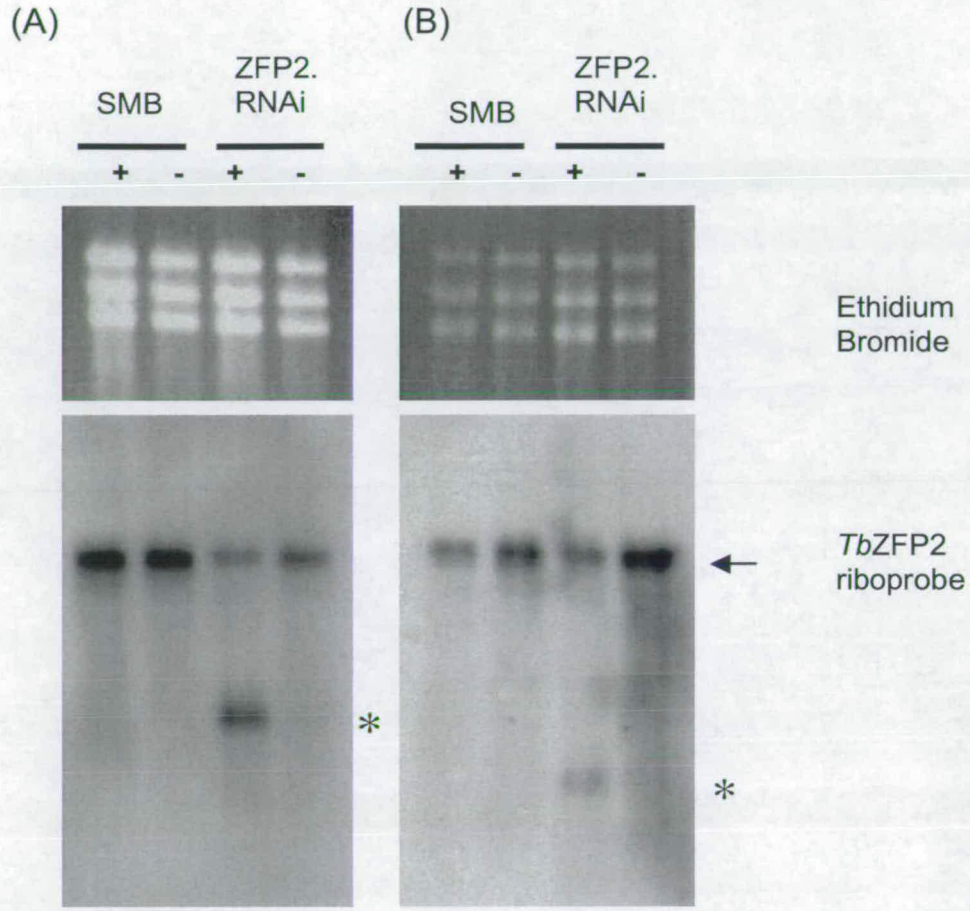

Figure 3.2 – Ineffective ablation of the endogenous *Tb*ZFP2 transcript by RNAi.

(A) The ZFP2.RNAi cell line was subjected to a repeated induction under identical conditions to Figure 3.1. However, northern analysis undertaken with a riboprobe specific to *Tb*ZFP2 after induction for 3 days with (+) or without (-) tetracycline, indicated that ablation of *Tb*ZFP2 was less effective than previously observed (Figure 3.1). (B) An independent clone of the ZFP2.RNAi cell line was cultured and induced as previously described. Again, a northern blot probed with a riboprobe to *Tb*ZFP2 indicated that ineffective ablation of the *Tb*ZFP2 transcript had occurred when the induced ZFP2.RNAi (+) was compared to the uninduced lanes (-). The ribosomal RNA is included as a loading control. Asterisk denotes the dsRNA derived from the construct.

3.2 Regeneration of p2T7i.TbZFP2 bloodstream form trypanosome cell lines.

In order to investigate the role of *Tb*ZFP2 in the regulation of mRNA transcripts, effective ablation of the *Tb*ZFP2 transcript was essential. Therefore, it was decided to repeat the initial transfection to recreate the ZFP2.RNAi cell line, as previously undertaken by Hendriks *et al* (2001). Hence, the SMB trypanosome cell line was transfected with the linearised p2T7i.*Tb*ZFP2 plasmid (a kind gift of Dr Ed Hendriks). Resistance to the drug phleomycin identified successful transformants that contained the p2T7i.*Tb*ZFP2 plasmid stably integrated into the genome.

Four cell lines were obtained and named 2T7i.*Tb*ZFP2 2.1-4 (Figure 3.3A). Each cell line was induced with tetracycline for 3 days and assayed for the ablation of the *Tb*ZFP2 transcript. Protein samples were also taken to analyse the effect of ablation of the *Tb*ZFP2 RNA transcript at the protein level. Figure 3.3A showed that ablation of the *Tb*ZFP2 transcript was highly effective in all cell lines, with similar levels observed to those published by Hendriks *et al,* (2001). A slight decrease in RNA levels was observed when the uninduced *Tb*ZFP2 cell lines were compared to the uninduced SMB, indicating incomplete repression of the p2T7i.*Tb*ZFP2 plasmid. Furthermore, a decrease of >90% of the endogenous *Tb*ZFP2 protein was observed, as shown by the western blot probed with the anti-*Tb*ZFP2 antibody in Figure 3.3B. These experiments were repeated and the phenotype was found to be stable and reproducible (data not shown). Hence,

reproducible and effective ablation of the *Tb*ZFP2 RNA transcript was achieved using RNAi.

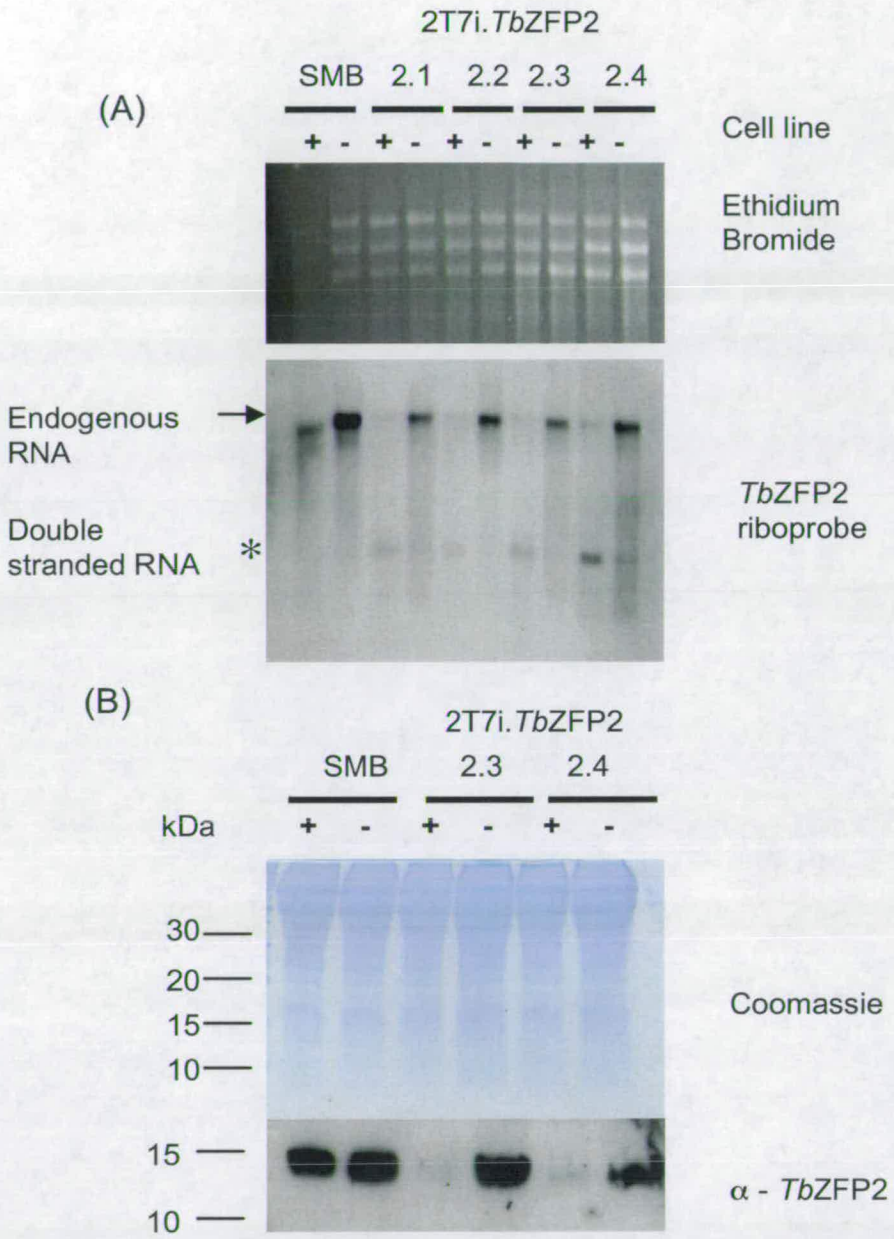

Figure 3.3 – Verification of RNAi ablation of the *Tb*ZFP2 transcript and protein depletion in transfected trypanosomes.

The SMB cell line was transfected with the p2T7i.ZFP2 plasmid and the four cell lines obtained (2T7i.*Tb*ZFP2 2.1-2.4) were induced with tetracycline for three days, after which RNA and protein samples were prepared to test the ability of RNAi to ablate *Tb*ZFP2. (A) Northern blotting analysis of the SMB and 2T7i.*Tb*ZFP2 2.1-2.14 transformants hybridised with a *Tb*ZFP2 specific riboprobe. (B) Coomassie and western blot of the SMB and 2T7i.*Tb*ZFP2 2.3 and 2.4 hybridised with anti-*Tb*ZFP2 antibody.

3.3 Isolation of total trypanosome mRNA by PolyA selection.

The next stage was to isolate mRNA from the selected cell lines in order to make labelled cDNA. It was decided to use the uninduced SMB cell line and compare this to the induced 2T7i.*Tb*ZFP2 2.3 cell line as this presented the greatest change in the expression profile of *Tb*ZFP2. A comparison between the induced and uninduced 2T7i.*Tb*ZFP2 2.3 cell line would have been preferential. However, due to the incomplete repression of the p2T7i.*Tb*ZFP2 plasmid in the uninduced 2T7i.*Tb*ZFP2 2.3, the level of *Tb*ZFP2 RNA appeared decreased when compared to the SMB cell line, though this observation was not tested at the protein level (Figure 3.3 and data not shown). Thus, it was hypothesised that this effect may have lead to a reduction in the levels of the *Tb*ZFP2 transcript and potentially a reduced effect on RNA transcript abundance. Therefore, it was decided to use the uninduced SMB cell line as these contained endogenous levels of *Tb*ZFP2 transcript and compare this to the induced 2T7i.*Tb*ZFP2 2.3 cell line.

To isolate the RNA, $2x10^9$ uninduced SMB and induced 2T7i.*Tb*ZFP2 2.3 trypanosomes were cultured for 3 days and the RNA extracted using an RNeasy midi kit (Qiagen). As enrichment of the mRNA population was required, the rRNA was removed by using PolyA selection of mRNA via the PolyATtract mRNA isolation system kit (Promega). Briefly, this involved incubation of the total RNA sample with a biotinylated oligo dT primer, which annealed to the 3' polyA tail of the mature mRNA transcript. The oligo

dT primer/mRNA complex was then captured to magnetic beads that contained antibodies to biotin and stringently washed to remove the rRNA, before the mRNA was finally eluted from the magnetic beads. Removal of rRNA was important for accurate analysis of the mRNA expression profile.

To assay the effectiveness of the polyA selection procedure, a northern blot was undertaken using an aliquot of the mRNA samples eluted from the PolyATtract kit. Total SMB trypanosome RNA was used as a control. The three trypanosome rRNA bands were absent in the lanes that contained the PolyA selected SMB and induced 2T7i.*Tb*ZFP2 2.3 RNA on the agarose gel stained with ethidium bromide, when compared to the SMB total RNA control (Figure 3.4). The northern blot was then probed with a riboprobe specific for the *Tb*ZFP3 transcript, which showed that *Tb*ZFP3 mRNA was present within all three lanes and indicated that the PolyATtract system had removed the majority of the rRNA. *Tb*ZFP3 is a molecule related to *Tb*ZFP2 that is constitutively expressed. This acted as a control for RNA integrity in the selected mRNA populations.

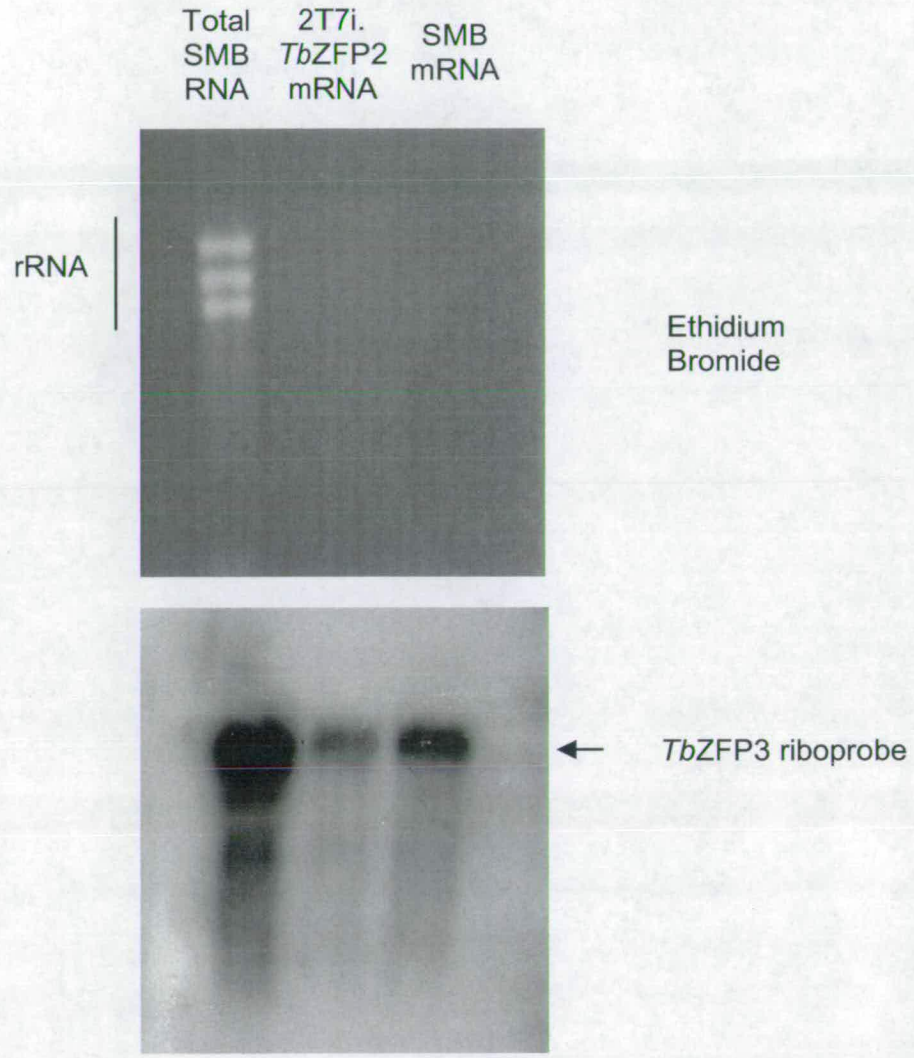

Figure 3.4 – Isolation of mRNA from total trypanosome RNA.

SMB and induced 2T7i.*Tb*ZFP2.RNAi RNA were subjected to PolyA selection to remove the rRNA and the subsequent samples analysed by northern blotting. Total SMB RNA was used as a control. The three rRNA bands are visible in the ethidium stained RNA gel of the total SMB RNA lane, but absent from the PolyA selected SMB and induced ZFP2.RNAi RNA lane (top panel). Hybridisation of the northern blot with a riboprobe specific to *Tb*ZFP3 (bottom panel) demonstrated that intact mRNA is present across all three lanes. TbZFP3 is a molecule related to TbZFP2 and acts as a control for mRNA integrity.

3.4 Comparison of mRNA abundance using a genomic *T. brucei* macroarray.

The next stage was the creation of labelled probes with which to probe the macroarray. Hence, the PolyA selected SMB and 2T7i.*Tb*ZFP2 2.3 mRNA was transcribed into cDNA using oligo dT as the primer and the reaction purified though a G50 Sephadex column to remove the unincorporated nucleotides. The cDNA was then labelled using the Gene images detection kit (GE Healthcare), which involved the incorporation of a dUTP nucleotide base conjugated to fluorescein into the DNA. The reaction products were tested to ensure that the probe had been successfully labelled (data not shown).

The genomic macroarray membranes used were created from the *T. brucei* reference strain 927 and contained genomic DNA sheared into 1- 2 Kb sized fragments and ligated into the pUC18 plasmid. The plasmids were then transformed into *E. coli* and spotted onto the membrane. Each macroarray contain 15 000 bacterial colonies spotted in duplicate onto the membrane, which represents a coverage of 80% of the trypanosome genome. The membranes were constructed by Professor Elisabetta Ullu, Yale University, USA.

For the hybridisation procedure, two macroarray membranes were used. One membrane was hybridised with the labelled SMB probe while the other was hybridised with the induced 2T7i.*Tb*ZFP2 2.3 probe (*Tb*ZFP2 RNAi cDNA). Both membranes were

hybridised overnight and then washed to remove the unbound probe and the signal was detected using the CDP-Star chemiluminescent system. The next stage was the analysis of the membranes for differences in RNA expression. Both membranes were systematically analysed by eye for the presence or absence of hybridisation spots, which indicated a difference in transcript expression. In the manufacture of the membranes, each colony spot was duplicated in a specific pattern. So for a spot difference to be considered genuine, both colony spots had to show a hybridisation signal on one membrane but show no signal on the other membrane. As no data is present to show if *Tb*ZFP2 stabilised or destabilised mRNA transcripts, all hybridisation differences between the two membranes were identified and analysed. For a pair of spots to be considered for further analysis, the putative hybridisation spots also had to be close to another pair of clearly hybridised spots (Figure 3.5C). Initially, 15 spots were identified as showing the strongest hybridisation difference between the macroarrays hybridised with SMB and TbZFP2 RNAi cDNA and named clones 1-15 (Figure 3.5C).

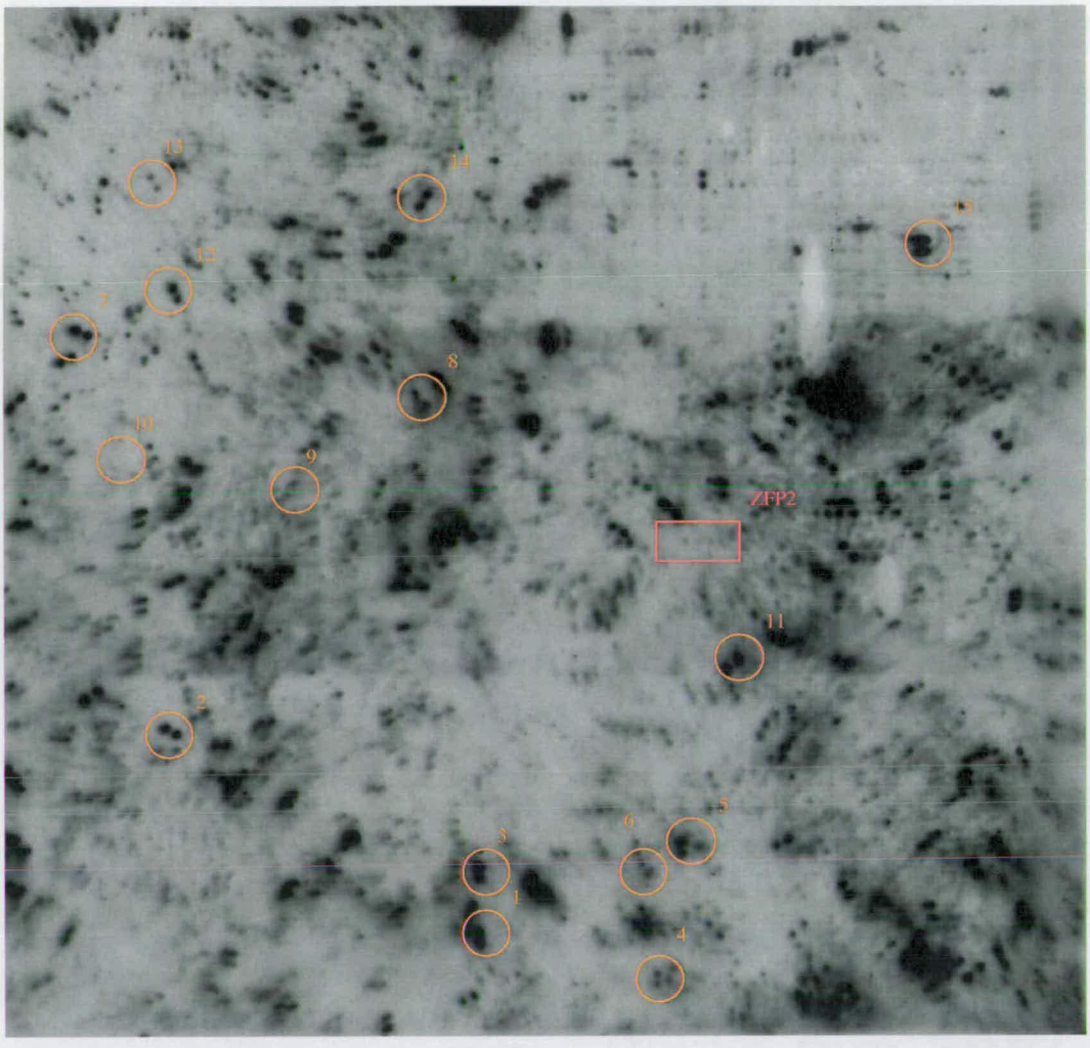

Figure 3.5A – A *T. brucei* genomic macroarray hybridised with SMB cDNA.

Labelled SMB cDNA was hybridised with the genomic macroarray and the spots analysed against the ZFP2 macroarray for differences in expression profile. The 15 colonies with the largest change in hybridisation profile as assayed by eye between the two membranes are highlighted in orange on both membranes. The clone containing the *Tb*ZFP2 gene is highlighted with a red square.

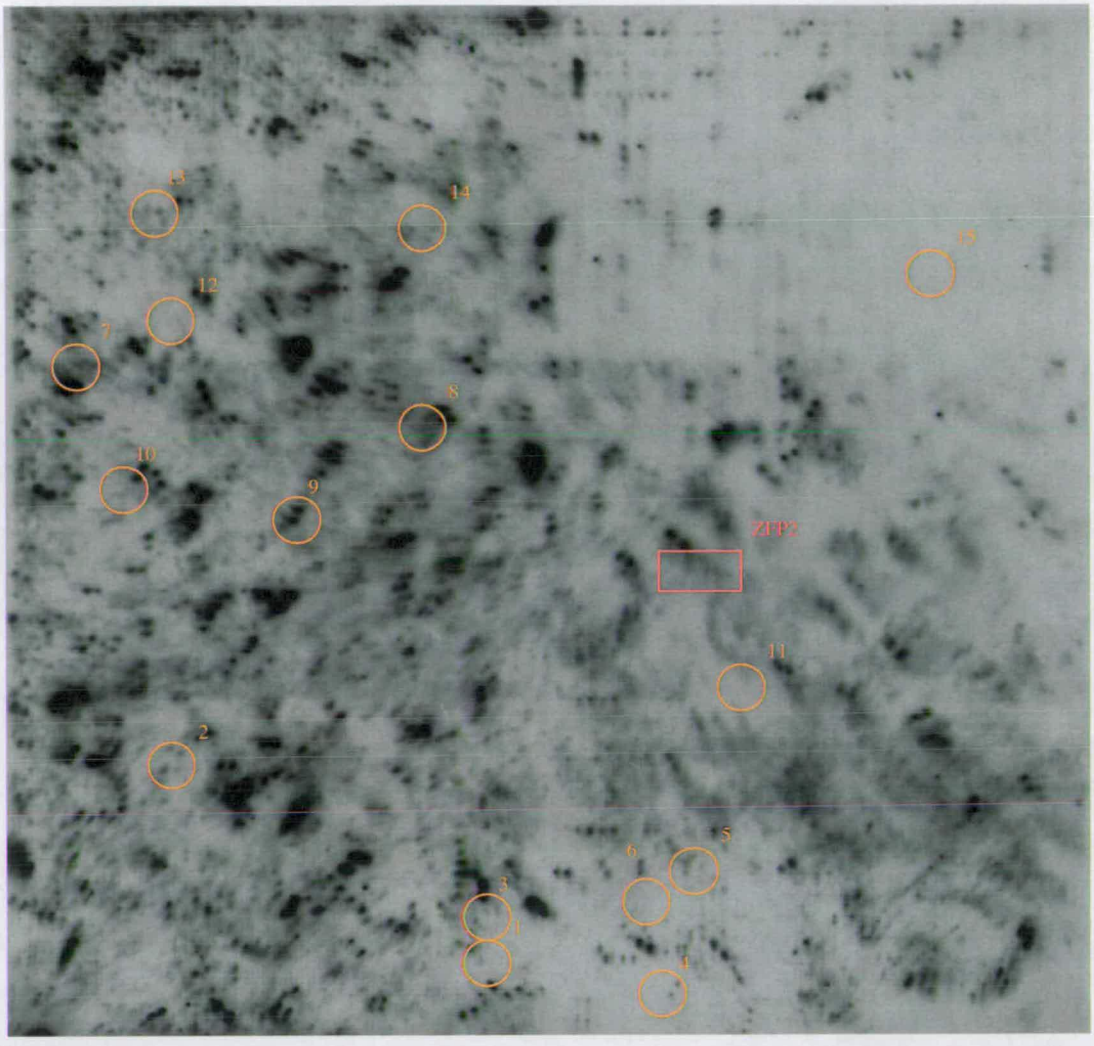

Figure 3.5B – A *T. brucei* genomic macroarray hybridised with induced 2T7i.*Tb*ZFP2 2.3 cDNA.

Labelled 2T7i.*Tb*ZFP2 2.3 cDNA was hybridised with the genomic macroarray and the spots analysed against the SMB macroarray for differences in expression profile. The 15 colonies with the largest change in hybridisation profile as assayed by eye between the two membranes are highlighted in orange on both membranes. The clone containing the *Tb*ZFP2 gene is highlighted with a red square.

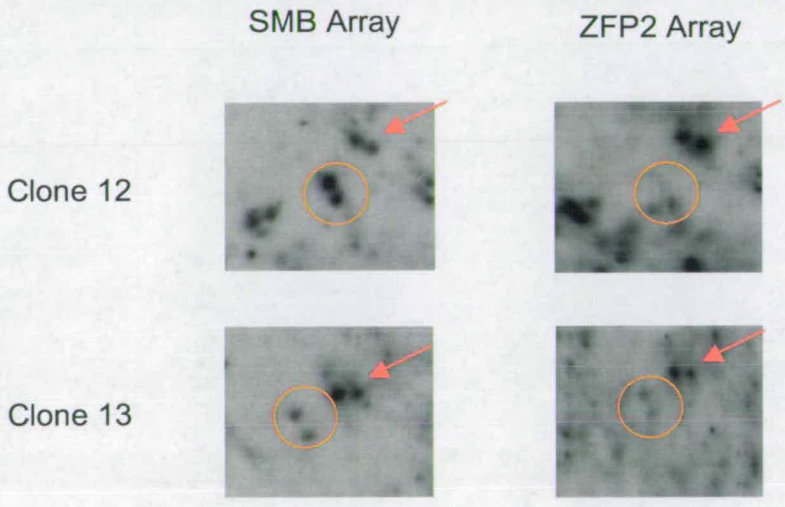

Figure 3.5C – Highlighted regions of the SMB and ZFP2 macroarray indicating hybridisation differences between two different clones across the membranes.

In order to highlight the hybridisation differences, two representative clones (Clones 12 and 13) were chosen which differ between the SMB and ZFP2 macroarray. The hybridisation differences are highlighted in orange, while an adjacent pair of clones which shown no hybridisation differences are indicated with the arrow.

3.5 Identification of mRNA transcripts that differ in abundance between SMB bloodstream form cell lines and p2T7i.TbZFP2 bloodstream form cell lines.

Each of the 15 hybridisation differences identified as having a differential expression pattern between macroarrays are highlighted in orange on both the SMB and ZFP2 macroarray in Figure 3.5 A-C. Each spot represented a bacterial colony transformed with a plasmid containing a section of *T. brucei* genomic DNA inserted. For all 15 hybridisation differences, each plasmid was sequenced using M13 forward and reverse primers (a kind gift of Professor Elisabetta Ullu, Yale University, USA) and the DNA sequences were then analysed against the *T. brucei* genome database to identify any genes within that clone (Table 3.1). Any clones containing ribosomal sequence were analysed futher, as it was probable that these represented false positive results. Supporting this, no consistency was observed in the rRNA genes identified. Clones that coded for a transposon or other mobile genetic element were also discarded, as preliminary data had indicated that ablation of *Tb*ZFP2 did not affect the levels of these transcripts (P Craddy, data not shown). In total, out of the 15 clones initially identified, 10 contained DNA sequences whose primary function was to encode rRNA, while another clone coded for a transposon. Only four clones (6, 8, 9, and 10) contained a DNA insert that coded for at least one protein coding gene fragment.

Clone	GeneDB identification	Size (nt)	Spot presence on		Function
			SMB	*Tb*ZFP2	
1			Y	N	rRNA
2			Y	L	VSG pseudogene/ Ingi hypothetical protein
3			Y	L	rRNA/transposon
4			Y	L	rRNA
5			Y	L	rRNA
6	*Tb*10.61.1200	3594	Y	N	Conserved hypothetical protein
7			Y	N	rRNA
8	*Tb*10.389.0680	1149	Y	L	Hypothetical protein
9		979	L	Y	Maxicircle unidentified reading frame 2 (Murf2)
9		1620	L	Y	Cytochrome oxidase 1 (Co1)
10	N19B2.075	2131	N	Y	RHS4 pseudogene
10	N19B2.080	349	N	Y	Hypothetical protein
10	N19B2.085	327	N	Y	Hypothetical transmembrane protein
11			Y	N	rRNA
12			Y	N	rRNA
13			Y	L	rRNA
14			Y	L	rRNA
15			Y	N	rRNA

Table 3.1 – Analysis of the 15 clones isolated from the comparison of the macroarrays hybridised with SMB and *Tb*ZFP2 RNAi cDNA.

Each of the 15 spot differences identified on the macroarray represented a clone containing a *T. brucei* genomic DNA insert. These were sequenced and the data analysed against GeneDB to identify the gene fragment and to assign function, where possible. The table shows the GeneDB identification tag for the gene fragments isolated, the size of the predicted gene in nucleotides as indicated by GeneDB and the predicted function of the gene. The expression profile of each clone on both the SMB and TbZFP2 macroarray is indicated; Y – spot strongly present; N – no spot present; L – low intensity of spot present.

As the array was made from sheared genomic DNA, the possibility existed that more than one gene was present within the same clone. This phenomenon occurred for clones 9 and 10, where two and three gene fragments were contained, respectively. The

expression profile of clones 6 and 8 was 'present' as spots on the macroarray hybridised with SMB cDNA, but 'absent' (or low) on the macroarray hybridised with *Tb*ZFP2 RNAi cDNA. This is in contrast to clones 9 and 10, where there was a strong hybridisation signal on the macroarray hybridised with *Tb*ZFP2 RNAi cDNA, but a low or absent hybridisation signal on the macroarray hybridised with SMB cDNA. Interestingly, all of the 10 rRNA clones identified were strongly present on the macroarray hybridised with SMB cDNA but were low or absent on the macroarray hybridised with *Tb*ZFP2 RNAi cDNA, suggesting that the mRNA purification in the SMB sample may have been less efficient. Thus, 7 DNA sequences (*Tb*10.61.1200, *Tb*10.389.0680, Co1, MURF2, N19B2.075, N19B2.080 and N19B2.085) were further investigated as potential candidates whose mRNA transcripts may have been regulated (positively or negatively) by *Tb*ZFP2.

3.6 Verification of mRNA abundance levels.

The expression of each gene identified from the macroarray screen was verified to ensure that *Tb*ZFP2 ablation regulated its abundance. Therefore, two independently generated cell lines were induced for three days to undergo RNAi to ablate *Tb*ZFP2 (2T7i.*Tb*ZFP2 2.1 and 2T7i.*Tb*ZFP2 2.2), after which RNA was extracted. The RNA samples were then analysed by Northern blotting to confirm that *Tb*ZFP2 had been successfully ablated by RNAi (data not shown). Confirmation of *Tb*ZFP2 ablation was undertaken on all RNA samples used in this section. Uninduced cell lines and the SMB cell line were also included as controls. With the ablation of *Tb*ZFP2 confirmed, the RNA was then analysed by Northern blot and the abundance of the candidate RNA transcripts investigated by hybridisation of a riboprobe specific to each.

Figure 3.6 shows that ablation of *Tb*ZFP2 does not appear to affect the abundance of either the N19B2.075 or N19B2.080 transcripts found on clone 10, as no difference was observed between cells in which *Tb*ZFP2 ablation was induced or not. Although there appeared to be a difference in the level of the N19B2.080 between the SMB and ZFP2 RNAi lines, subsequent Northern blots did not show the same phenomenon (data not shown). Thus, it is probable that these differences were due to loading differences or variation in the RNA levels of this transcript unrelated to the presence or absence of *Tb*ZFP2. The third gene fragment identified on clone 10, N19B2.085, was also not regulated by *Tb*ZFP2, as an invariant hybridisation signal was observed across all three

cells lines (Figure 3.7, A, indicated by the arrow). The probe to N19B2.080 appeared to cross react with several RNA transcripts present within the sample, but no increased or decrease in abundance was observed upon *Tb*ZFP2 ablation. The RNA transcript of gene *Tb*10.61.1200, which was isolated on clone 6, was not detected in any of the three cell lines tested (Figure 3.7, B).

Of the two gene fragments found on clone 9, no transcript was detected for maxicircle unidentified reading frame 2 (MURF2), despite repeated attempts (Figure 3.8A, data not shown). MURF2 is encoded from maxicircle DNA within the kinetoplast and undergoes minimal RNA editing (26 U residues are inserted and 4 deleted within the first 60 bases, the remaining transcript of 979nt being unchanged). As the riboprobe was capable of hybridising to the entire ORF transcript, the minimal editing of the 5' end was not likely to limit hybridisation efficiency. No change in RNA abundance was observed for the cytochrome oxidase 1 (CO1) gene when *Tb*ZFP2 was ablated (Figure 3.8 B). Again, while RNA variability was observed between cell lines, no specific effect due to ablation of the *Tb*ZFP2 transcript was observed. Unlike MURF2, CO1 does not undergo any RNA editing to create the mature mRNA. Figure 3.8C shows that the abundance *of* *Tb*10.389.0680 was also not affected by *Tb*ZFP2 ablation. Again, as there was no difference in transcript abundance, it implies that *Tb*ZFP2 is not affecting the abundance of the RNA transcript. Hence, of these 7 transcripts studied, no difference in abundance between cell lines in which *Tb*ZFP2 RNA was ablated, or not, could be confirmed.

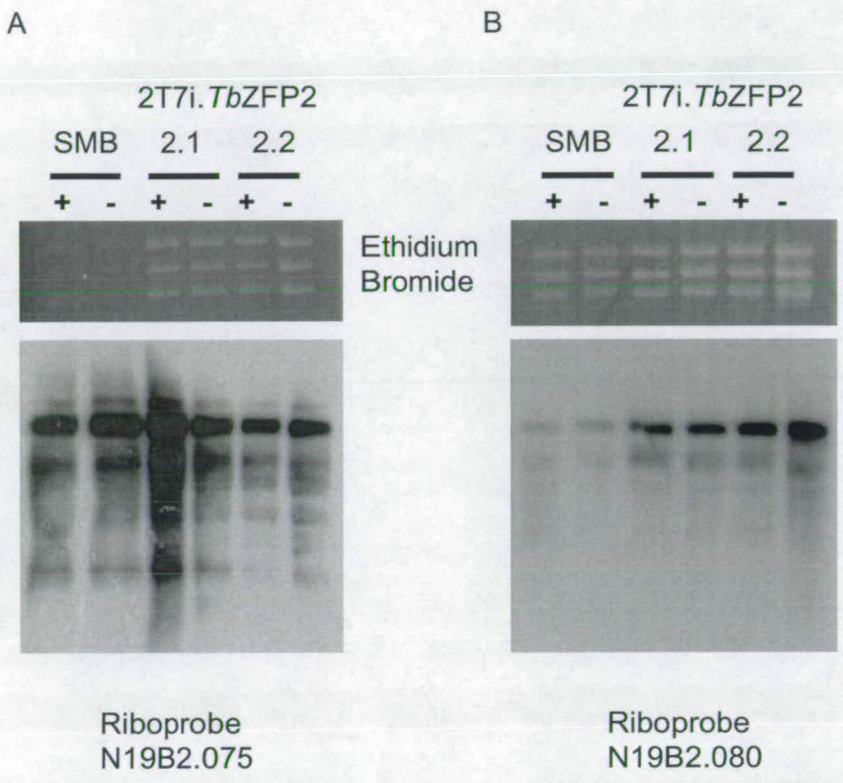

Figure 3.6 –RNA transcript abundance in the presence and absence of *Tb*ZFP2 ablation using RNAi.

Northern analysis showing the effect of the presence or absence of *Tb*ZFP2 on two RNA transcripts, N19B2.075 and N19B2.080, in comparison to the parental cell line (SMB). 2T7i.*Tb*ZFP2 2.1 and 2.2 indicates two independent *Tb*ZFP2 RNAi cell lines, in the presence (+) or absence (-) of tetracycline to induce the RNAi phenotype. The RNA was then hybridised using the N19B2.075 riboprobe (A) or N19B2.080 riboprobe (B) to determine if *Tb*ZFP2 ablation had an effect on transcript abundance. The ethidium bromide stained rRNA (Ethidium Bromide) is included as a loading control.

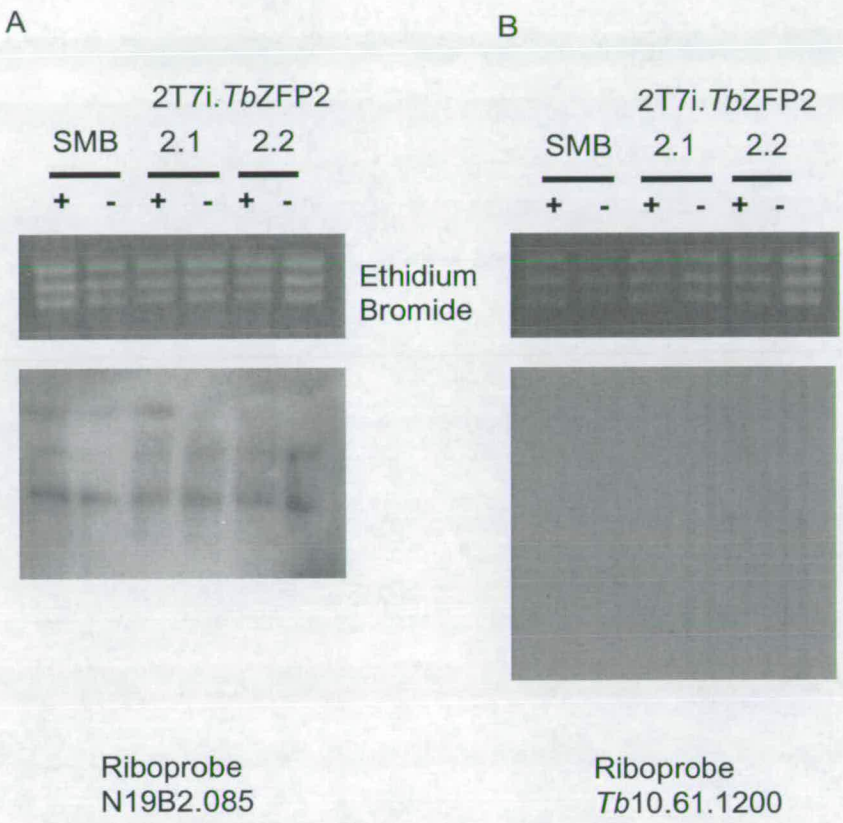

Figure 3.7 – RNA transcript abundance in the presence and absence of *Tb*ZFP2 ablation using RNAi.

Northern analysis showing the effect of the presence or absence of *Tb*ZFP2 on two RNA transcripts, N19B2.085 and *Tb*10.61.1200, in comparison to the parental cell line (SMB). 2T7i.*Tb*ZFP2 2.1 and 2.2 indicates two independent *Tb*ZFP2 RNAi cell lines, in the presence (+) or absence (-) of tetracycline to induce the RNAi phenotype. The RNA was then hybridised using the N19B2.085 riboprobe (A) or *Tb*10.61.1200 riboprobe (B) to determine if *Tb*ZFP2 ablation had an effect on transcript abundance. The ethidium bromide stained rRNA (Ethidium Bromide) is included as a loading control.

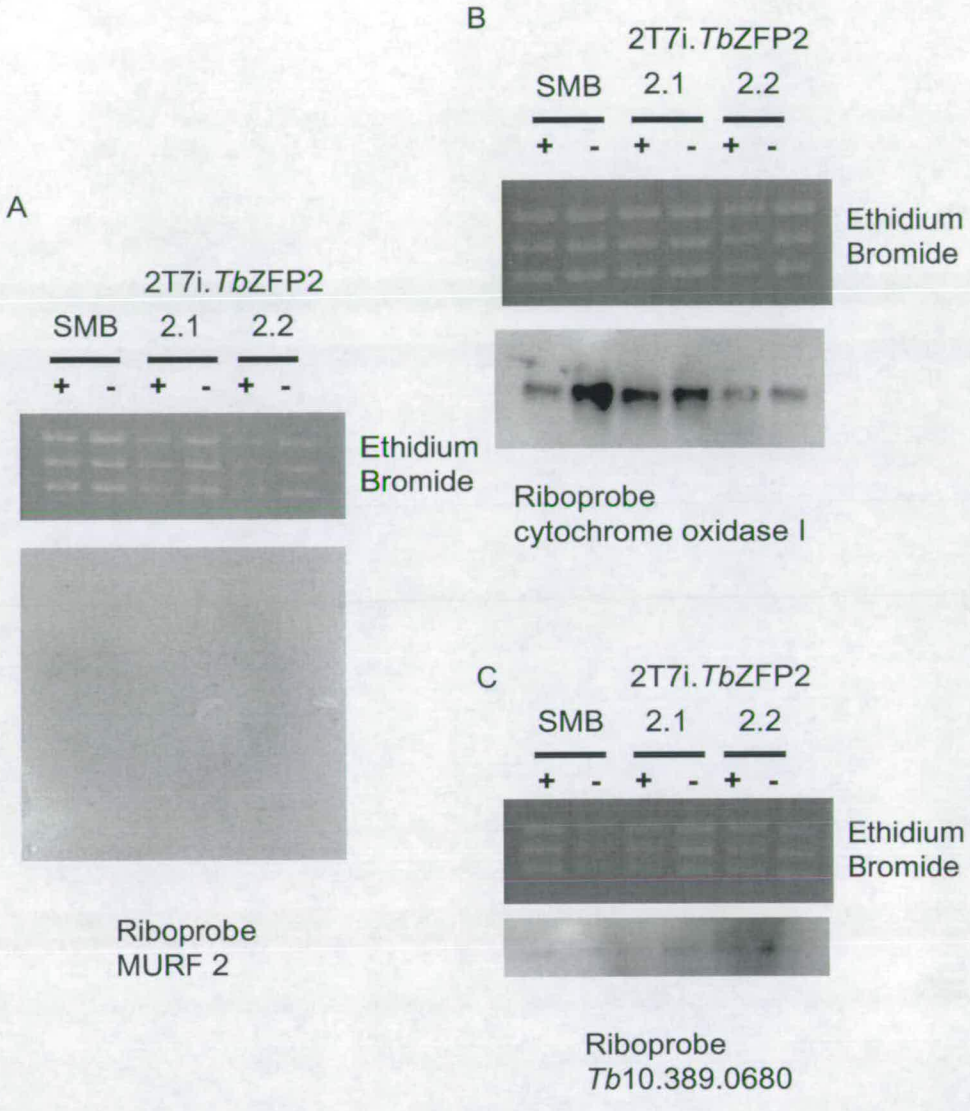

Figure 3.8 – RNA transcript abundance in the presence and absence of *Tb*ZFP2 ablation using RNAi.

Northern analysis showing the effect of the presence or absence of *Tb*ZFP2 on three RNA transcripts, MURF2, cytochrome oxidase I and *Tb*10.389.0680, in comparison to the parental cell line (SMB). 2T7i.*Tb*ZFP2 2.1 and 2.2 indicates two independent *Tb*ZFP2 RNAi cell lines, in the presence (+) of tetracycline to induce the ablation of the *Tb*ZFP2 transcript. (-) indicates that the *Tb*ZFP2 transcript is present. The RNA was then hybridised using the MURF2 riboprobe (A), cytochrome oxidase I riboprobe (B) or *Tb*10.389.0680 riboprobe (C) to determine if *Tb*ZFP2 ablation had an effect on transcript abundance. The ethidium bromide stained rRNA (Ethidium Bromide) is included as a loading control.

3.7 Identification of TbZFP2 DNA on the T. brucei macroarray.

As the *Tb*ZFP2 transcript was targeted by the RNAi pathway, the transcript levels of *Tb*ZFP2 were expected to be differentially expressed between the two hybridised macroarrays, This would be manifested as the presence of hybridisation spots on the macroarray hybrisied with SMB cDNA and the absence of a hybridisation signal on the membrane hybridised with *Tb*ZFP2 RNAi cDNA. Therefore, this provided a useful control for assaying hybridisation effectiveness.

To identify which clone contained the *Tb*ZFP2 gene fragment, the macroarray hybridised with *Tb*ZFP2 RNAi cDNA was washed in 0.5% SDS for 1 hour at 68°C, before being washed for 1 hour in TE to strip the membrane. Thereafter, the membrane was hybridised with a *Tb*ZFP2 specific probe overnight and the membrane processed according to the hybridisation protocol (Section 2.15.2). *Tb*ZFP2 was found to be located within one of the macroarray clones (Figure 3.5, indicated by the red square). However, when the intensities of this spot were compared between the macroarray hybridised with *Tb*ZFP2 RNAi and SBM cDNA, no difference was observed.

3.8 Discussion.

Since the majority of gene expression in trypanosomes is regulated at the post-transcriptional level, novel kinetoplastid specific RNA binding molecules are likely to play an important role in the control of gene expression. Hence, RNA binding proteins implicated in the process of differentiation, such as the *Tb*ZFP family, that contain no homologues in higher eukaryotes and which contain RNA binding motifs proven to regulate mRNA stability in other systems, are good candidates for regulators of differentiation.

Many examples exist of RNA binding proteins regulating mRNA though ARE regions within the 3' UTR, with the TTP protein being the prototype for this in higher eukaryotes. TTP binds to the ARE in the 3' UTR of cytokine transcripts and causes destabilisation of the transcript. As the *Tb*ZFP family are probable RNA binding proteins due to the presence of the $C_{X8}C_{X5}C_{X3}H$ motif and are involved in differentiation, it was hypothesised that *Tb*ZFP2 would affect the abundance of RNA transcripts, perhaps in a similar manner to TTP. Hence, a comparative macroarray screen was undertaken to identify RNA transcripts whose abundance was regulated by *Tb*ZFP2. This was achieved by comparing two trypanosome cell lines on a macroarray, one with *Tb*ZFP2 mRNA present, the other with *Tb*ZFP2 mRNA ablated by RNAi. From the macroarray analysis, 15 clones were identified as being potentially regulated by *Tb*ZFP2, which corresponded to 6 putative RNA transcripts, once false positives had

been removed. Subsequent analysis of the abundance of these transcripts by northern hybridisation showed no corresponding validation in response to *Tb*ZFP2 ablation.

The failure to detect transcripts regulated by *Tb*ZFP2 may be due to a variety of reasons. While northern hybridisation has shown that >90% *Tb*ZFP2 transcript was ablated and the protein level was significantly reduced (also ~ 90%), the small quantities of remaining *Tb*ZFP2 may have been sufficient for physiological function. There is currently no data on the level of *Tb*ZFP2 protein required for correct functioning in trypanosomes. The low level of *Tb*ZFP2 transcript present might also have been reason that the *Tb*ZFP2 gene itself showed no significant difference between the macroarrays hybridised with SMB and *Tb*ZFP2 RNAi cDNA. Hence, if it were not possible to detect this known difference in *Tb*ZFP2 abundance between the macroarrays, it would render the identification of RNA targets regulated by *Tb*ZFP2 a challenging prospect. However, the clone containing the *Tb*ZFP2 gene was in a region that contained poor cDNA hybridised and the lack of a detectable difference in this region may represent a consequence of this. Also, the possibility existed that there were multiple genes contained within the genomic clone containing the *Tb*ZFP2 gene. The nearest gene to *Tb*ZFP2 is ~1100 nucleotides away, so it is quite possible that the clone contained parts of this gene. Therefore, if this gene were constitutively expressed at high levels, it would mask the decreased *Tb*ZFP2 expression abundance in the induced RNAi line. Only by sequencing the *Tb*ZFP2 clone, which was not undertaken, would it be possible to identify if this was correct.

The SMB cell line was used in preference to the uninduced ZFP2.RNAi for the macroarray comparison due to the incomplete repression of the p2T7i.ZFP2 RNAi plasmid. As regulation of the RNAi construct was not complete and a slight drop in transcript level of the endogenous *Tb*ZFP2 in the uninduced ZFP2.RNAi was observed when compared to the SMB line, it was decided that this might affect the RNA abundance of transcripts regulated by *Tb*ZFP2. Hence, the use of the SMB and 2T7i.*Tb*ZFP2 2.3 cell line should have emphasised any differences in the RNA transcript levels due to the large expression profile change in *Tb*ZFP2.

The presence of 10 rRNA clones identified from the macroarrays indicated that, while the PolyA selection was effective at removing the majority of rRNA, there was still significant amount of rRNA present, particularly in the SMB sample. Improved mRNA selection and more stringent washes of the selected material would decrease these false positives. In addition, the data appears to suggest that PolyA selection was less efficient in the SMB sample than in the *Tb*ZFP2 sample. This would lead to the differences observed between the rRNA transcripts on each macroarray. Also, if the effect of *Tb*ZFP2 ablation were a subtle up-regulation or down-regulation of transcripts required for differentiation, these differences might have been obscured against the background of rRNA. Coupled to the fact that some areas on the macroarrays were poorly hybridised, it would make it hard to identify genuine changes in transcript abundance between the samples. Also, as the macroarrays only contained 80% of the trypanosome

genome, the possibility remains that transcripts regulated by *Tb*ZFP2 were not present on the arrays. Had a full analysis of all the clones been undertaken and even small differences analysed, it may have isolated clones that were regulated by *Tb*ZFP2. However, the large number of false positives observed suggested that identification of any genuine transcripts would have been very challenging.

The variability of the effectiveness of the RNAi mediated *Tb*ZFP2 ablation was intriguing. Storage of the ZFP2.RNAi trypanosomes within liquid nitrogen appeared to decrease the effectiveness of the RNAi effect. This has been observed in other laboratories (Keith Matthews, personal communication), Hence to ensure optimal efficiency of the RNAi phenotype and maximal depletion of the *Tb*ZFP2 transcript, all ZFP2.RNAi lines were freshly transfected and tested to ensure that *Tb*ZFP2 knockdown was obtained.

Beyond technical differences, an alternative explanation for why no changes in RNA abundance were observed is that *Tb*ZFP2 may have no function in the bloodstream form trypanosome, as no obvious phenotype was observed when the *Tb*ZFP2 transcript was ablated using RNAi (Hendriks *et al.*, 2001). The *Tb*ZFP2 phenotype was originally observed during the process of differentiation and in procyclic cells (Hendriks *et al.*, 2001). Hence, using the two cell lines 4 hours into differentiation may have made a better comparison. However, since the RNAi lines are monomorphic, asynchronous differentiation would occur. In this context, a potential problem would be that it could be

hard to identify which changes in mRNA expression profile were due to the ablation of *Tb*ZFP2, when set against the significant changes in RNA expression profile that occur after the onset of differentiation. Alternatively, it may have been better to undertake the screen in procyclic trypanosomes where *Tb*ZFP2 has been shown to be functionally active, as assayed by the "Nozzle" phenotype (Hendriks *et al.*, 2001).

A more relevant experiment to identify *Tb*ZFP2-regulated transcripts would be to create a cell line where both copies of *Tb*ZFP2 gene have been genetically disrupted (i.e. a *Tb*ZFP2 KO line). The *Tb*ZFP2 KO line should be viable in bloodstream forms, as RNAi has achieved effective ablation of the *Tb*ZFP2 transcript with no obvious phenotype. A comparison of the parental and bloodstream KO line on a trypanosome oligonucleotide microarray containing each trypanosome gene as an individual oligo could then be undertaken and might give a much cleaner mRNA expression profile comparison. This would allow a clearer analysis of which transcripts, if any, are regulated by *Tb*ZFP2. It would also overcome the problems of having multiple gene fragments present within the same clone. If it were possible to create the *Tb*ZFP2 KO trypanosome line using pleomorphic trypanosomes, it should also be possible to look at the effect of *Tb*ZFP2 on the mRNA profile during synchronous differentiation using the microarray. This should highlight the important transcripts associated with *Tb*ZFP2 during this process.

Work on the *T. cruzi* homologue of *Tb*ZFP1 (*Tc*ZFP1) has shown that *Tc*ZFP1 is unable to bind ARE sequences *in vitro* (Morking *et al.*, 2004). Indeed, data indicates that *Tc*ZFP1 has a strong affinity for poly-C oligoribonucleotides, although a weak affinity was also seen with the U-rich 26 identified in EP procyclin (Hotz *et al.*, 1997; Morking *et al.*, 2004). Similarly, *in vitro* RNA binding assays undertaken for *Tb*ZFP3 (Chapter 5) also show no affinity for ARE sequences, with the preference being for sequences containing the TCAGT pentamer. Therefore, the *in vitro* RNA binding studies suggests that the *Tb*ZFP family is not acting through AREs within the 3' UTR of the mRNA.

The major limitation of the approach used in this chapter is that the control of differentiation by *Tb*ZFP2 might not be at the level of RNA stability, but instead operate either through translational repression/activation or mRNA localisation. Indeed, recent data has demonstrated that *Tb*ZFP2 and 3 associates with the polysomal apparatus in procyclic trypanosomes (Paterou *et al.*, 2006), which suggests that this family of molecules operate together to translationally activate stage specific transcripts. Certainly, the results from this macroarray do not support a role for *Tb*ZFP2 in wide scale mRNA regulation.

Chapter 4:

A Systematic Evolution of Ligands by Exponential Enrichment (SELEX) approach to identify the RNA binding motif and RNA targets for *Tb*ZFP3.

4.1 Introduction

The process of differentiation from bloodstream form trypanosomes to procyclic forms is of great interest from a cell biological perspective. Hence, the *Tb*ZFP family containing the $C_{X8}C_{X5}C_{X3}H$ RNA binding motif is of particular interest as these molecules are important to the process of differentiation (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou, 2004). One member of this family, *Tb*ZFP3, enhances the process of differentiation when ectopically over-expressed, though the mechanisms by which this occurs have yet to be identified. Although extensive work has been carried out on the phenotypes generated by perturbed expression of *Tb*ZPF3 *in vivo* (Paterou, 2004; Paterou, 2006), it has not yet been shown that *Tb*ZFP3 binds RNA nor have the RNA targets been identified.

Eukaryotic proteins that contain the $C_{X8}C_{X5}C_{X3}H$ zinc finger RNA binding motif generally bind the 3' UTRs of mRNA through AU rich elements (AREs), leading to mRNA instability (Blackshear, 2002). The prototype protein of the mammalian TIS family, where this motif was originally identified, is Tristetraprolin (TTP), and it has been demonstrated by SELEX that UUAUUUAUU is the optimal RNA binding sequence for this molecule (Blackshear *et al.*, 2003; Worthington *et al.*, 2002). Similar $C_{X8}C_{X5}C_{X3}H$ motifs are involved in differentiation, control of translation and post-transcriptional gene regulation throughout the eukaryotic world (Lai *et al.*, 2000; Reese *et al.*, 2000; Tabara *et al.*, 1999). Therefore, one may postulate that the mode of action for *Tb*ZFP3 could be through binding AREs within the 3' UTRs of its mRNA targets. However, all other RNA binding proteins containing the

$C_{X8}C_{X5}C_{X3}H$ motif identified to date contain at least two copies of the $C_{X8}C_{X5}C_{X3}H$ motif, unlike *Tb*ZFP3 that only possesses one.

*Tb*ZFP3 contains another RNA binding motif, the RGG box (Kiledjian and Dreyfuss, 1992), a domain that usually occurs in combination with other RNA binding motifs. An RNA selection procedure has shown that the RGG box recognises G-quartet structures specifically within mRNAs bound by the fragile X mental retardation protein (FMRP) (Darnell *et al.*, 2001; Schaeffer *et al.*, 2001). Its presence has thus been postulated to bind RNA, with other RNA motifs present providing the RNA sequence specificity (Burd and Dreyfuss, 1994).

Recent RNA:protein binding studies on the *Trypanosoma cruzi* homologue of *Tb*ZFP1 (TcZFP1) have shown a preference for poly-C oligonucleotide sequences, though weak binding was also observed with the U-rich 26-mer identified in the 3' UTR of EP/GPEET procyclins (Hotz *et al.*, 1997; Morking *et al.*, 2004). However, no direct association was observed with ARE elements (Morking *et al.*, 2004).

Identification of the RNA targets for these proteins is a vital part of understanding the functions and role of these molecules in differentiation in the trypanosome. Approaches are currently underway to identify mRNA sequences that interact with *Tb*ZFP3 *in vivo* by co-immunoprecipitation. However, while this may identify the full-length mRNA sequences, it is less likely to identify the precise RNA motif that is bound by *Tb*ZFP3 and any identified RNA interactions may be the result of indirect binding. Therefore, a complementary approach of Systematic Evolution of

Ligands by EXponential enrichment (SELEX) was used as a method to identify putative motifs that may interact with *Tb*ZFP3 directly. This *in vitro* method is predicted to allow the identification of the optimal binding motif that interacts with *Tb*ZFP3. The approach has been used successfully for identifying RNA sequence motifs in a wide range of eukaryotic systems including identification of mRNA targets of the proline-rich RNA binding protein in mice (Prrp) (Hori *et al.*, 2005), the optimal RNA sequence bound by Tristetraproline (Worthington *et al.*, 2002) and the *in vitro* selection of ligands to cell adhesion receptors in *T. cruzi* (Ulrich *et al.*, 2002).

In this chapter, the results from using a SELEX methodology to identify RNA binding motifs for *Tb*ZFP3 will be presented. The ability of *Tb*ZFP3 to bind RNA will be shown, and a putative RNA binding motif identified. Identification of the RNA sequence motif that interacts with *Tb*ZFP3 has the potential to provide an insight into the function of this molecule by identification of the RNAs that are regulated by it. This is crucial to understanding the role that *Tb*ZFP3 plays during differentiation and in the trypanosome's lifecycle as a whole.

4.2 Systematic Evolution of Ligands by Exponential Enrichment (SELEX).

The approach used here is based on that published by Hori *et al* (Hori *et al.*, 2005). The method involved taking a large DNA pool containing a random central core of 50 nucleotides flanked by common sequences comprising an RNA polymerase binding site, and transcribing them *in vitro* to generate RNA. The random pool of RNA was then incubated with recombinant His-tagged *Tb*ZFP3 immobilized onto Ni NTA-agarose beads. Any unbound RNA was removed by washing and any RNA, which was bound to the *Tb*ZFP3 protein, was then eluted, reverse transcribed and amplified by PCR to complete one round of selection. These aptamers were then transcribed and formed the input RNA for the next round of selection. This procedure is repeated between 3-10 times using the amplified RNA as the new input RNA each time. The final population was then sequenced and putative motifs analysed by oligonucleotide frequency scoring (van Helden *et al.*, 1998), these being validated by electrophoretic mobility shift assay (EMSA). A schematic representation of the SELEX procedure is shown in Figure 4.1.

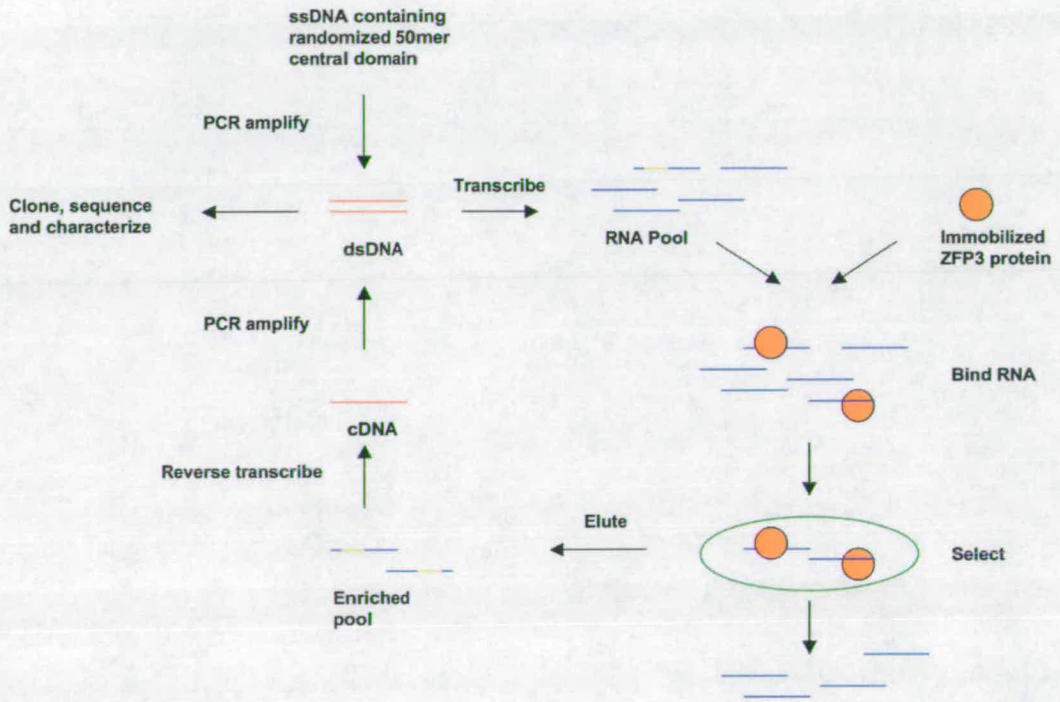

Figure 4.1 – Schematic representation of the SELEX procedure.

A single stranded DNA aptamer containing a random 50mer core surrounded by conserved flanking regions was used as the template to transcribe a pool of single stranded RNA. This pool was then incubated with recombinant His-tagged TbZFP3 anchored to Ni NTA agarose beads to allow the RNA to bind and any unbound RNA was removed by washing. The RNA was extracted from the beads by phenol:chloroform extraction and reverse transcribed into cDNA, using the specific 3' primer This product underwent PCR amplification to form the pool for the next round of selection. After between 3 -10 rounds of selection, the DNA is cloned and analysed for any motifs present.

4.3 Induction and purification of recombinant His-TbZFP3 from JM109 *E. coli* bacterial cells.

The expression and purification of a recombinant His-tagged *Tb*ZFP3 has been previously achieved (Paterou, 2004). Thus, the *Tb*ZFP3 open reading frame was successfully ligated into the pET30a expression vector and expressed in JM109 (DE3) *E. coli* cells. The protein was then successfully purified by FPLC to give a fraction containing recombinant *Tb*ZFP3 possessing an N terminal 6-Histidine tag. As large quantities of His-tagged *Tb*ZFP3 were required for the SELEX protocol, an aliquot containing pET30a-His-*Tb*ZFP3 in JM109 (DE3) *E. coli* cells and a 1L culture of bacterial cells was induced at OD_{600}= 0.6 with 1mM IPTG at 37°C, as previously described (Paterou, 2004). Figure 4.2 shows the recombinant protein was expressed at high levels after induction, matching the expression levels seen previously (Paterou, 2004). As expected, the recombinant protein migrates at approximately 20kDa as opposed to the native protein, which has a molecular weight of 14.1kDa. This size difference is due to the addition of the 6 - Histidine tag.

Purification of the recombinant *Tb*ZFP3 from the JM109 (DE3) cells was undertaken by FPLC according to Paterou (2004). However, when using this procedure, many contaminating bands were present within the elutant, which could adversely affect the SELEX procedure (data not shown). Thus, the FPLC purification conditions were optimised to increase the purity of the recombinant *Tb*ZFP3. In particular, the constituents of the binding buffers (Buffer A and B) were altered from 20mM Na_2PO4 to 25mM HEPES and the concentration of NaCl decreased from 0.5M to

150mM. In addition, 0.1% Trition X and 5mM Imidazole was added to the lysis buffer to aid removal of non-specific contaminants during selection of the recombinant protein. Purification was then carried out on an AKTA FPLC system with a 1ml Hitrap column, using an increasing imidazole gradient to elute the recombinant protein.

In Figure 4.3, purification of recombinant *Tb*ZFP3 using Ni^{2+} chelating chromatography is shown, with the blue line indicating the protein absorbance at 280nm. The far left blue peak corresponds to the unbound lysed cell extract, while the second peak at fraction 12 corresponds to the elution of bacterial histidine rich proteins that bind the column. The peak encompassing fractions 14 -18 is that of the purified recombinant *Tb*ZFP3, as demonstrated by Figure 4.3B which shows the Coomassie stained gel fractions and the equivalent western blots, probed with an anti-His tag and anti-*Tb*ZFP3 antibody. *Tb*ZFP3 was consistently eluted by an imidazole concentration of 400 - 500mM. Interestingly, it appears that there was rapid degradation of the recombinant *Tb*ZFP3, evident by the 15kDa band observed on the western blot probed with the anti-His tag antibody (indicated by the asterisk). This was surprising as a protease inhibitor cocktail was included throughout the purification process and in the final elutant, which should have minimised the level of protein degradation that occurred. However, a similar sized band was observed within the purifications undertaken by Dr Paterou, implying that a low level of degradation is usual when purifying recombinant *Tb*ZFP3 (Paterou, 2004).

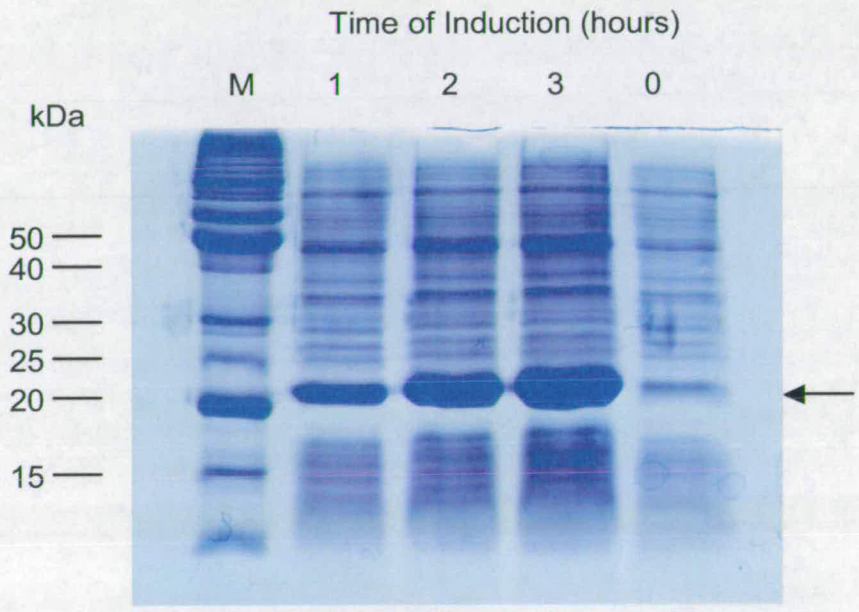

Figure 4.2 – Induction and expression of recombinant His-tagged TbZFP3 in JM109 (DE3) *E. coli* cells

A small scale overnight culture of the JM109 (DE3) *E. coli* cells containing the pET30a-His-TbZFP3 plasmid was diluted into 250mls LB and incubated at 37^0C until an OD_{600nm} = 0.6. A sample was taken (Time 0) and the culture was then induced with 1mM IPTG, with further samples taken each hour for 3 hours after induction. The samples were then analysed on a 15 % SDS page gel and stained with Coomassie. The expressed His-tagged TbZFP3 is evident as the large band at 20 kDa, indicated by the arrow.

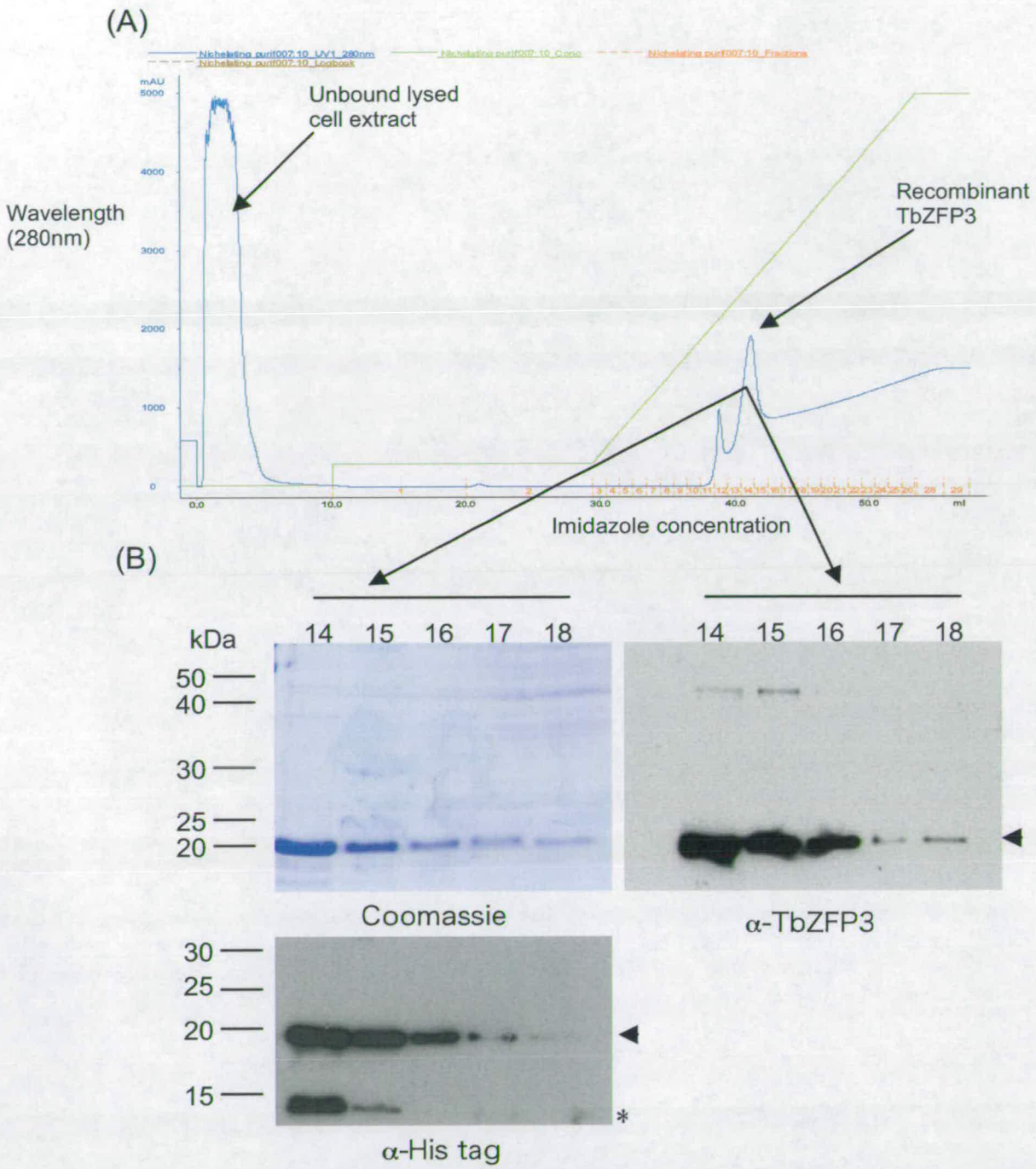

Figure 4.3 – Purification of recombinant His-tagged TbZFP3 from JM109 (DE3) *E. coli* cells.

Recombinant His-tagged TbZFP3 was purified as described in Chapter 2. The absorbance at 280nm of the protein purification is shown by the blue line in (A). (B) Fractions 14 – 18 from the purification in (A) were analysed on a 15% SDS PAGE gel and stained with Coomassie or transferred onto nitrocellulose membrane and probed with anti-His tag or anti-TbZFP3 antibody. The arrowhead indicates the recombinant TbZFP3 while the asterisk identifies the degraded TbZFP3.

4.4 Binding of recombinant His-TbZFP3 to Ni NTA agarose beads.

An essential part of the SELEX protocol is the immobilisation of any protein:RNA complex formed and the ability for the unbound RNA to be removed. Therefore, in the protocol used here, Ni-NTA Agarose beads (Qiagen) were used to bind and immobilise the recombinant His-tagged *Tb*ZFP3. However, because the purification of *Tb*ZFP3 resulted in a high concentration (~ 400mM) of imidazole remaining in the elution buffer, it was necessary to decrease the concentration of imidazole prior to progression to the next stage. Therefore, several methods were attempted in order to exchange buffers, including the use of vivaspin columns to concentrate the protein. However, even using a HEPES based buffer, minimal amounts of concentrated protein were eluted from the vivaspin columns. Moreover, buffer exchange methods were tried using capillary tubing, which resulted in the rapid precipitation of *Tb*ZFP3 (data not shown).

Nonetheless, since the initial purification of *Tb*ZFP3 had resulted in the isolation of highly concentrated fractions of *Tb*ZFP3 (1μg/μl), it was possible to use a small amount of this material and to dilute it 5 fold with binding buffer. This reduced the final concentration of imidazole to lower levels (~ 80mM), thereby enabling the His-tagged *Tb*ZFP3 to bind directly to the Ni NTA agarose beads. Since the protein binding capacity of the beads is 5-10mg/ml, 200μg of *Tb*ZFP3 protein was added to 20μl of packed beads, thereby ensuring that the beads would be completely saturated

with *Tb*ZFP3. Figure 4.4 shows successful binding of *Tb*ZFP3 to the Ni NTA beads, with a major band at 20kDa. In addition a band at 50kDa is also detected, this being potentially due to the dimerisation of *Tb*ZFP3 protein, though this is likely to be due to an interaction via the His tag, rather than a direct interaction between the *Tb*ZFP3 proteins themselves. This is because no homodimerisation of TbZFP3 was observed in the yeast two-hybrid analysis (see Chapter 5).

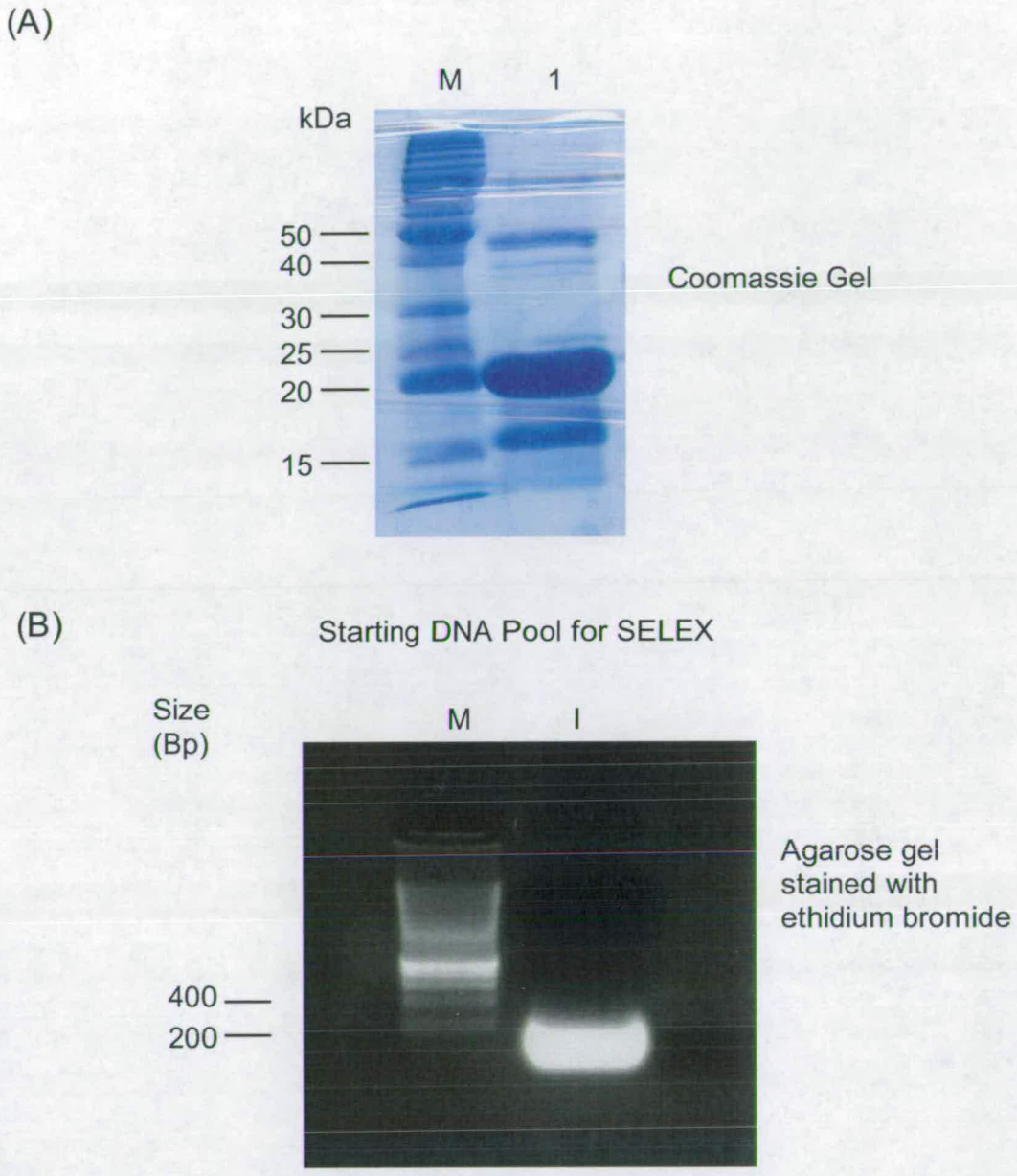

Figure 4.4 – Binding of His-*Tb*ZFP3 to Ni-NTA Agarose beads and the initial DNA pool for the SELEX procedure.

(A) His-tagged *Tb*ZFP3 (1) was incubated for an hour with Ni-NTA agarose beads and the supernatant removed before the beads were boiled in Laemmli buffer and analysed on a 15% SDS PAGE gel. The gel was then stained with Coomassie Blue. (M) = Protein marker. (B) The initial PCR amplification to produce the DNA pool for transcription for the first round of selection (I). The DNA was run on a 1% agarose gel stained with ethidium bromide and a DNA marker (M)

4.5 Creation of the initial population of RNA for the SELEX procedure.

The population of RNA aptamers is of paramount importance in SELEX, as ideally it should contain every possible sequence combination to ensure a successful outcome. The DNA oligonucleotide pool used to generate this contained a random 50-nucleotide core sequence flanked by two specific amplification sequences and will be referred to as an aptamer. The corresponding 5' primer contained a T7 RNA polymerase promoter sequence, while the 3' primer was used to amplify the aptamer and for reverse transcription (Figure 4.5).

The DNA oligonucleotide population was synthesized by Sigma-Genosys, who used an equimolar quantity of all 4 bases when inserting each base to ensure no possible bias when creating the random 50mer core. The initial DNA pool was created by amplifying the synthesised DNA oligos with the 5' and 3' SELEX primers to create a 30 μg DNA pool (Figure 4.4B). This library was then transcribed into RNA using T7 RNA polymerase to produce the initial RNA population. From this population, 1nmol of RNA was used as the input RNA, which was expected to be composed of 10^{14} molecules.

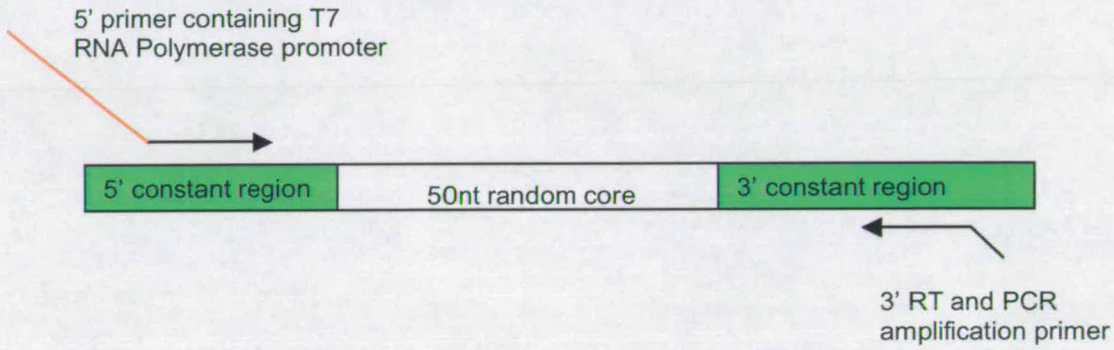

Figure 4.5 – Schematic representation of the SELEX aptamer.

Each random 50 nucleotide core was flanked by a 5' and 3' region of constant sequence. A 5' primer was designed to bind to the 5' constant region and contained a T7 RNA Polymerase site integrated into the primer. The 3' primer bound to the 3' end and was used for PCR and RT reactions.

4.6 SELEX enrichment of the RNA population.

The next stage of the protocol entailed enrichment of the RNA population. For this, 20µl of Ni-NTA agarose beads charged with His-tagged *Tb*ZFP3 were incubated with 1nmol of the initial input RNA for one hour. The beads were then washed 5 times with binding buffer before the RNA was extracted by phenol chloroform extraction and ethanol precipitation. The resulting resuspended RNA was then reverse transcribed and amplified with the 3' and 5' SELEX specific primers and an aliquot analysed on a 1% agarose gel. This pool was then transcribed *in vitro* into RNA using T7 RNA polymerase and this pool of RNA was to be used as the input RNA for the next round of selection.

To identify any non-specific binding of RNA to either the Ni NTA agarose beads or to the recombinant *Tb*ZFP3, two controls were included. As one control, empty Ni NTA agarose beads were used and, as the other, Ni NTA agarose beads loaded with His-tagged recombinant Protein Tyrosine Phosphatase 1 (*Tb*PTP1, a kind gift of Dr Balazs Szoor). The inclusion of *Tb*PTP1, a *T. brucei* protein tyrosine phosphatase that lacks any known RNA binding motifs, was to recognise the potential for nonspecific interactions of RNA with the His-tag or any other non-specific protein interactions. The ability of the empty Ni NTA agarose beads to bind RNA nonspecifically was also a potential concern; hence the inclusion of empty Ni NTA agarose beads as a control. This population could be used to remove any bias from within the selected *Tb*ZFP3 population caused by the presence of the Ni NTA agarose beads alone.

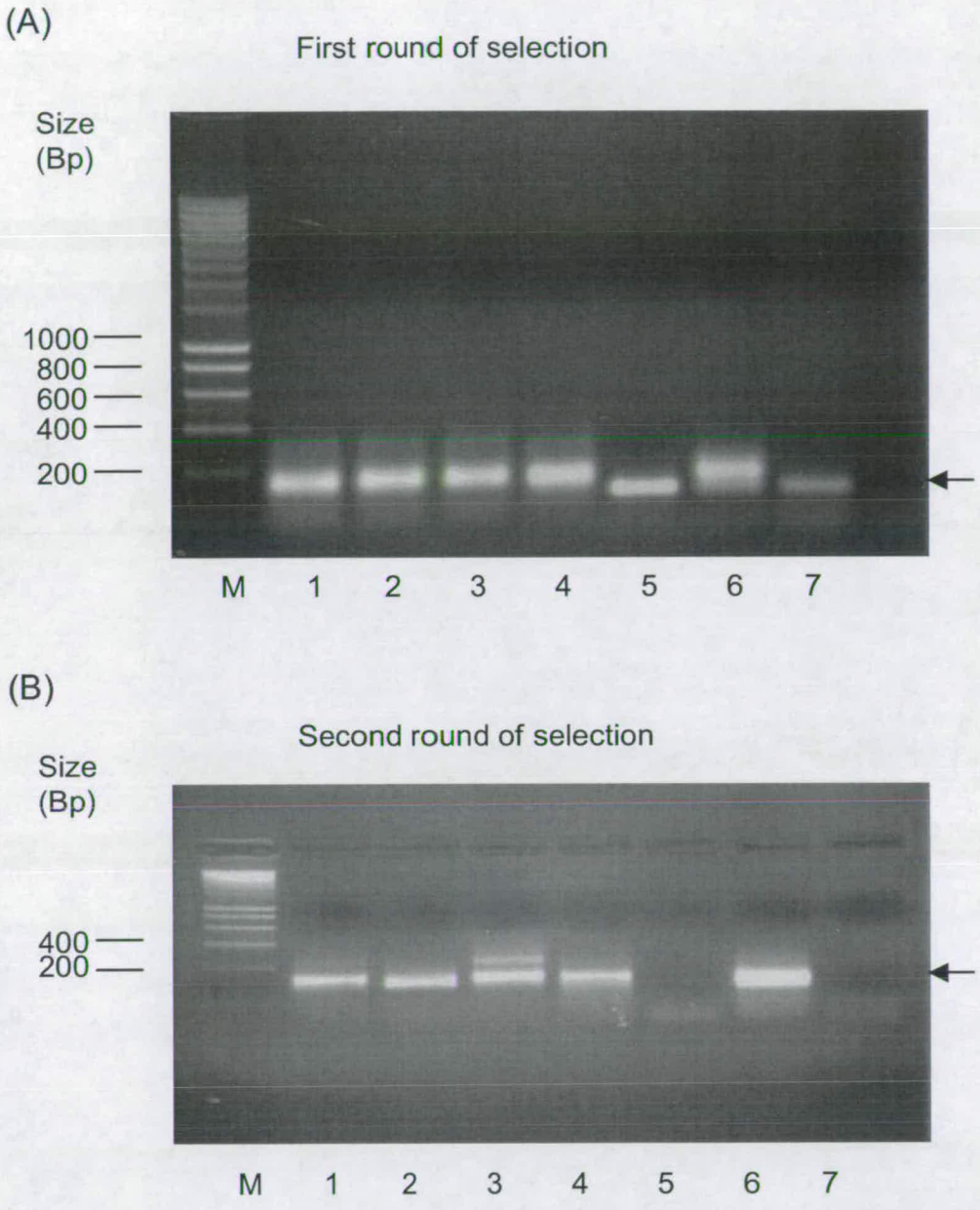

Figure 4.6 – Agarose gel electrophoresis of rounds one and two of the SELEX procedure after reverse transcription and PCR amplification.

The results of RT and PCR amplifications as analyzed on a 1% agarose gel stained with ethidium bromide after one round (A) and two rounds (B) of selection. M: Marker; 1: *Tb*ZFP3 beads; 2: *Tb*PTP1 beads; 3: Empty beads; 4: Round 0 RNA (positive control); 5: No RNA (Negative control); 6: Wash 5 of *Tb*ZFP3 beads; 7: PCR water control. PCR water control.

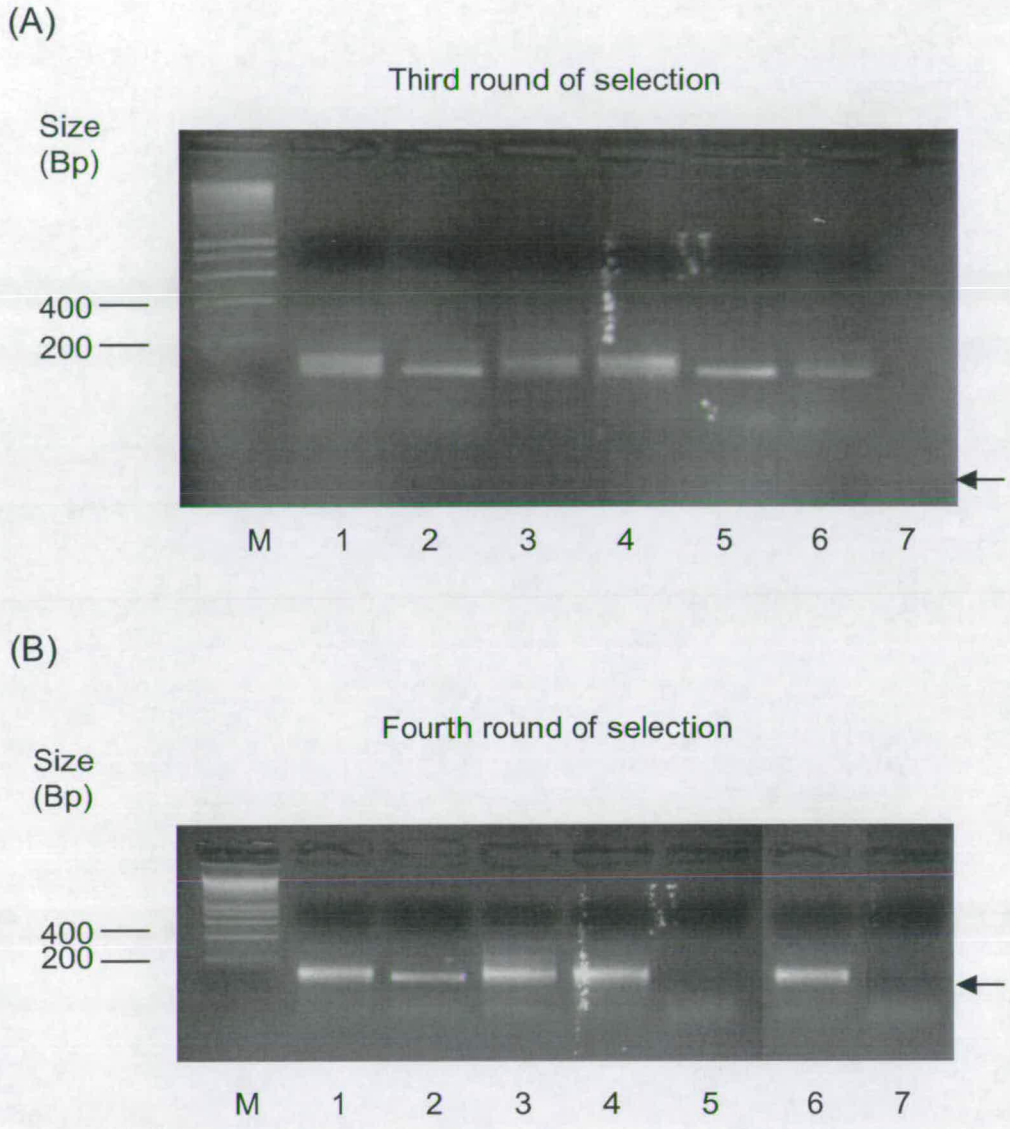

Figure 4.7 – Agarose gel electrophoresis of rounds three and four of the SELEX procedure after reverse transcription and PCR amplification.

The results of RT and PCR amplifications as analyzed on a 1% agarose gel stained with ethidium bromide after three rounds (A) and four rounds (B) of selections. M: Marker; 1: *Tb*ZFP3 beads; 2: *Tb*PTP1 beads; 3: Empty beads; 4: Round 0 RNA (positive control); 5: No RNA (Negative control); 6: Wash 5 of *Tb*ZFP3 beads; 7: PCR water control.

The results of the reverse transcription and amplification from the first round of SELEX are shown in Figure 4.6A. It is clear that the Ni NTA agarose/*Tb*ZFP3 complex was binding RNA, as indicated by the band present at ~ 100bp (Figure 4.6A, lane 1). However, as RNA was also eluted from the Ni NTA agarose/*Tb*PTP1 complex (lane 2) and the Ni NTA agarose beads alone (lane 3), the implication is that these complexes were binding RNA nonspecifically. The occurrence of RNA within the wash stage is to be expected (lane 6), as it would be almost impossible to remove all of the unbound RNA. The RT control (lane 5) contained RNA transcribed from the initial pool, while the negative control contained water only. The occurrence of a contaminating band within the RT control (lane 5) and to a lesser degree, the PCR control (lane 6) was concerning, but these bands are a slightly smaller size than those obtained by the selected RNA, and bands of this size are not seen within the selected RNA.

The DNA reverse transcribed from the RNA selected by *Tb*ZFP3 was transcribed using T7 polymerase and used as the input for the second round of selection. Again the His-tagged recombinant *Tb*ZFP3 Ni NTA agarose beads were used for selection as were Ni-agarose beads containing *Tb*PTP1 and empty Ni-agarose beads, as controls. As in round 1, *Tb*ZFP3 appeared to bind RNA, as shown by the ~100bp band amplified in Figure 4.6B, lane 1. However, as before, the empty beads and the *Tb*PTP1/beads complex also resulted in amplified bands (lanes 2 and 3). In this case no contamination was observed in either the RT or PCR controls (lanes 5 and 7).

The amplified DNA was next used for the third round of selection, the results of which are shown in Figure 4.7A. This produced similar results to the earlier rounds of selection. Thus, charged beads containing either *Tb*ZFP3 (lane 1), *Tb*PTP1 (lane 2) or empty beads (lane 3) all appeared to bind RNA, while RNA was still present in the wash stages (lane 6), albeit in very low levels. However, a low level of contamination was observed within the reverse transcription control (lane 5), although not in the PCR control (lane 7).

Finally, a fourth round of selection was carried out, with similar results to those previously described (Figure 4.7B). Thus, the RNA aptamers bound not only the *Tb*ZFP3 beads (lane 1) but also the *Tb*PTP1 beads (lane 2) and the empty beads (lane 3). Although amplification was consistently observed in the control lanes containing Ni NTA agarose beads, it remains unclear at this stage whether the RNA pool selected by *Tb*ZFP3 represented enrichment for a particular motif, or non-specific RNA binding.

One potential consideration was contamination of DNA within the RNA pools. It has been shown that $C_{X8}C_{X5}C_{X3}H$ zinc finger proteins are able to bind DNA, such that any DNA contamination would have the potential to significantly bias the outcomes of selection (Brown, 2005). Therefore, to evaluate this possibility, a small aliquot of each RNA pool from the first round of selection was treated with RNase H for an hour and the products analysed on an agarose gel. Figure 4.8 clearly shows the RNA before and after treatment with RNase H and confirms there was no obvious

contamination with DNA. However, amplification by PCR was not undertaken to confirm this.

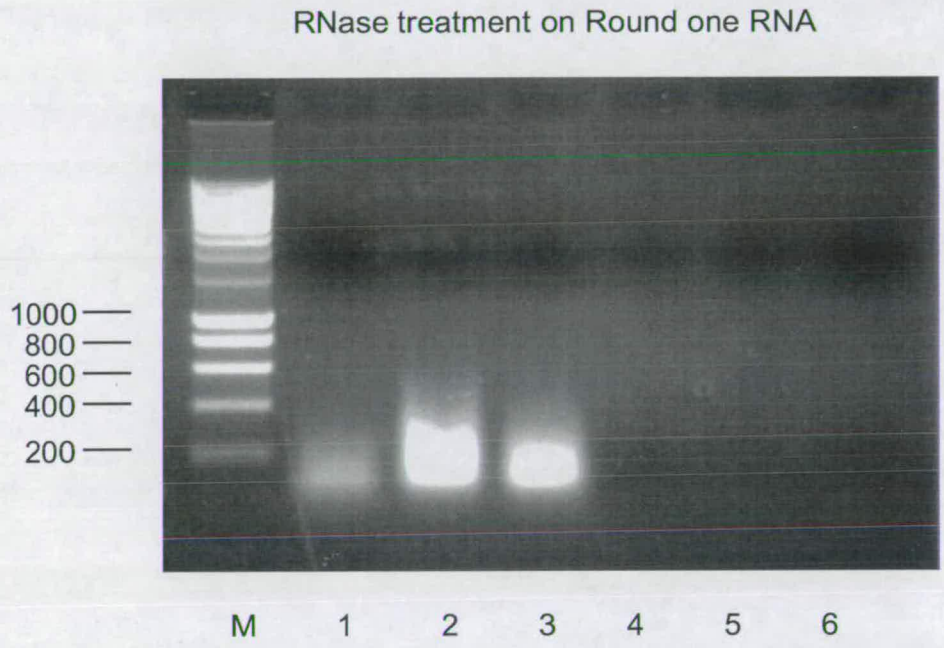

Figure 4.8 – RNase A treatment of RNA transcribed after one round of selection.

DNA that had been subjected to one round of selection was transcribed into using T7 RNA Polymerase. The resulting product was subject to treatment with RNase A and run on a 1% agarose gel. (M) Marker; (1) RNA selected by *Tb*ZFP3, and after treatment with RNase A (4); (2) RNA selected by *Tb*PTP1, and after treatment with RNase A (5); (3) RNA selected by Ni-NTA beads, and after treatment with RNase A (6). No evidence for contaminating DNA was observed in lanes 4-6.

4.7 Electrophoretic Mobility Shift Assay (EMSA) analysis of protein: RNA interactions between His-TbZFP3 and RNA after four rounds of selection.

To determine whether the RNA obtained after 4 rounds of SELEX enrichment would bind to recombinant *Tb*ZFP3 with high affinity, an electrophoretic mobility shift assay (EMSA) was undertaken. The basis of this assay is that the migration of a radiolabelled RNA:protein complex will be retarded, when run on a native acrylamide gel, in comparison to that of the free radiolabelled RNA. A schematic representation of the EMSA is shown in Figure 4.9.

To carry out EMSA analysis, cDNA from the initial (unselected) pool (round 0) and RNA subjected to 4 rounds of SELEX enrichment were radiolabelled by transcription in the presence of ^{32}P UTP and the labelled transcripts incubated with or without His-tagged *Tb*ZFP3 for 15 minutes in binding buffer. Also, the specificity of any complex formed was assayed by the addition of increasing amounts of tRNA, a non-specific competitor. The samples were run onto an 8% native gel, the gel dried and exposed to X-ray film.

Figure 4.10 shows that the incubation of *Tb*ZFP3 with initial round 0 RNA (Lane 2) caused a decreased mobility in the labelled probe (as indicated by the arrow), when compared to the free RNA in Lane 1 (arrowhead). Similarly, free RNA that has undergone 4 rounds of selection with charged *Tb*ZFP3 Ni-NTA agarose beads, also generated a mobility shift when incubated with recombinant *Tb*ZFP3 (compare lanes

4 and 5) when *Tb*ZFP3 is incubated with the RNA. This complex was similar in size and appearance to the complex formed by the round 0 RNA/*Tb*ZFP3 in Lane 2. This demonstrates that *Tb*ZFP3 is able to bind RNA. However, both the round 0 and selected RNA:protein complexes can be competed by the addition of 100μg tRNA, suggesting the binding affinity of the protein to the RNA is low.

It was to be hoped that having undergone 4 rounds of selection, the RNA in lanes 4-6 would bind *Tb*ZFP3 with a much higher affinity than round 0, as the SELEX procedure should have selected for sequences that bind *Tb*ZFP3 with greater specificity. This may have occurred, but with the high affinity RNA ligands representing a minor subset of the round 4 RNA pool. Alternatively, it was possible that the binding conditions of the EMSA were too stringent, with the result that the interactions generated by the SELEX procedure were not stable. Interestingly, both lanes 2 and 5 also contained another weak bandshift, as indicated by the asterisk, which ran significantly higher up the gel. The presence of this bandshift may have resulted from dimerisation of the His-tagged *Tb*ZFP3 protein or the formation of large aggregated RNA complexes. Regardless of its composure, this high molecular weight complex was consistently observed between experiments.

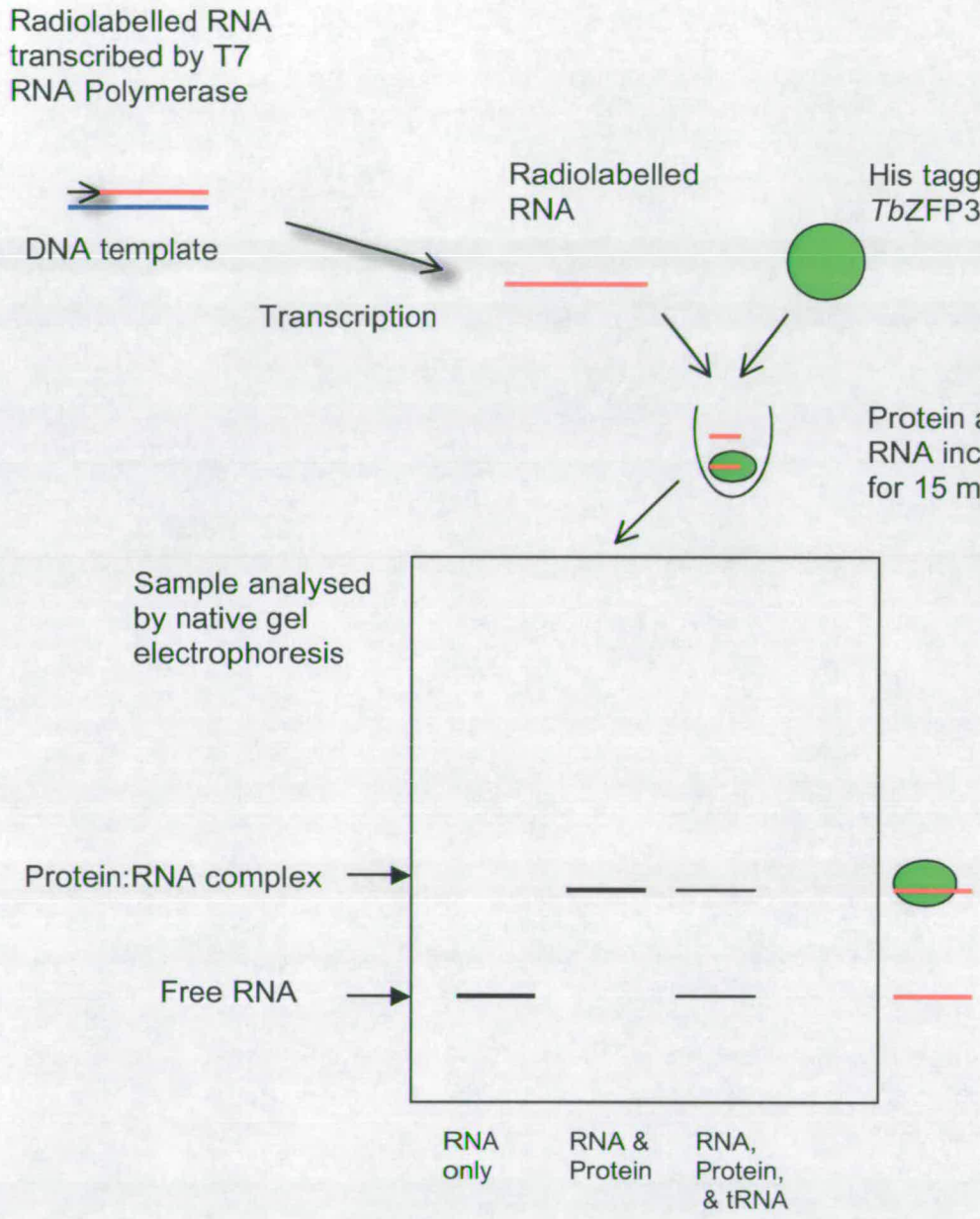

Figure 4.9 – Schematic representation of the Electrophoretic Mobility Shift Assay (EMSA).

Radiolabelled RNA is transcribed from the products of selection using ^{32}P UTP and T7 RNA Polymerase. The labelled RNA and His-tagged *Tb*ZFP3 were then incubated for 15 minutes with the addition of competitors, before being analysed on an 8% native gel. Once run, the gel was dried and exposed to autoradiographic film. Any RNA:Protein complexes are retarded on the acrylamide gel and hence are observed as a larger molecular weight band, when compared to the free RNA.

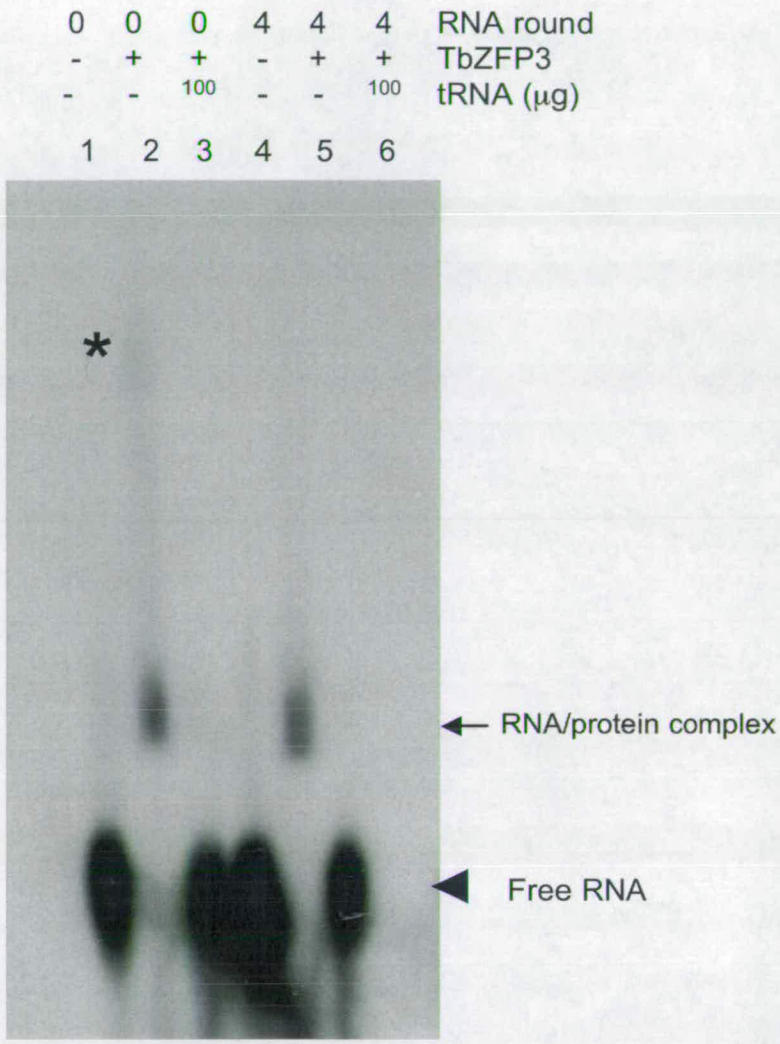

Figure 4.10 – EMSA showing the His-tagged TbZFP3:RNA complex formed with RNA from the initial SELEX pool and RNA that has been subjected to 4 rounds of selection with TbZFP3 Ni NTA agarose beads.

Radiolabelled ^{32}P RNA from SELEX round 0 and round 4 was incubated with His-tagged TbZFP3 and the complex analysed by native gel electrophoresis. A schematic representation of the contents of each lane is shown above. Lane 1: Round 0 RNA only; Lane 2: Round 0 RNA and TbZFP3; Lane 3: Round 0 RNA, TbZFP3 and tRNA; Lane 4: Round 4 RNA only; Lane 5: Round 4 RNA and TbZFP3; Lane 6: Round 4 RNA, TbZFP3 and tRNA. The arrowhead indicates the free RNA probes; the arrow indicates the TbZFP3 protein/RNA complex and the asterisk points to a large unknown molecular weight protein/RNA complex.

4.8 An Oligonucleotide counting method to identify potential RNA binding motifs for TbZFP3.

To further investigate any RNA sequence enrichment that had occurred during the 4 rounds of SELEX, the sequence of the aptamers that had been selected using *Tb*ZFP3 were compared with those bound by the empty beads. Hence, 80 aptamers were sequenced, 40 from the 4th round of *Tb*ZFP3 selection and 40 from the 4th round of selection with the empty Ni-NTA agarose beads. Since all of the sequences were transcribed in a given orientation using T7 RNA polymerase, it was possible to orientate each sequence to ensure that they would be presented in the same strand and orientation as they would have been during the SELEX procedure. These sequences are to be referred to as the '*Tb*ZFP3 dataset' and the 'beads dataset' and are shown in Figures 4.11 and 4.12.

Initially, 40 sequences from each pool were aligned and scanned manually for any enriched motifs, though no obviously selected sequences were identified. Therefore, to identify any motifs present within the raw sequence data, an oligonucleotide counting approach was used. This approach searches each dataset for oligonucleotides that are statistically overrepresented when compared to a control population or to an equiprobable frequency.

>S6 AUUUUGUUUAAUGUGUUUAUUGUGCUUGUAACCGUUUCGAUACUGAGGUU
>S4 UUAAAGCACUGCGGUCCCGCUACGCAGCUGAGGUCAAGUCGCGCCGUGA
>S3 UUUUCCUUUCCGCGCGGCUCGUUUCCUAAGGUCUGGUCUAUUUCUUUUGU
>S1 UUAAUGUGUAGAAACGGUUUUUAUUUGUGUUGAUUGGAUUUUGUGGGUGU
>S16 GUGACUUCACAGUAGUUGUUGCGGGUUUUCACUUGCCCCAUAUCCAGGU
>S14 AUCCUCUACUCGAAUUCCUCAGCCAAAAUUGAAGCGUUAUAUAUGUCGUG
>S13 UUUAGCAUGUUGGGUUAGGUCGUAGUCUCCGCGUUGUUGACGUUGCUGUU
>S11 CUUUAUUUUUUCGAGAUUGCGUGCUUAGACCUGUAUGGACUCCGGUGUU
>S10 GUUGCCUUUUUUGCGGUUUUAUGAGGUCAGCCUUGGAUGCACGGUUGGA
>S24 GUUAACUAGUCACGCUGCAUCAGTCCUCUCCUUGACAUGGUAAUGGUGAU
>S23 CUAUUUUCACGCUUCUGAUUUCUUCUACUUACCGAGGAGUUUGCUGU
>S22 CUGGGCUUUUUGUUUACGUGUGUUGGGUGUUUGGCUGUUAAUGGGUCAGT
>S20 AUCGUGAUUCUAAUGGUCAGTCAAGGGAAGUGACGUCCGUUGUUGAUUGC
>S19 GUAUAUUUUGACAUUUUUCUGUUUCAAAACUCAUCGCGCUUUCAGGCUUU
>S18 GUUUUCAUUUCCUAUCGCUACGGUUUUGAUAUGCCUUUUGAUUGCCUGU
>S17 AUUUUUCCGCAUACAUAUUAUCAGTUGUCAGAGAAUGUUCAGTCGUUGGU
>S33 ACCAACGACUGAACAUUCUCUGACAACUGAUAAUAUGUAUGCGGAAAAAU
>S32 CUAAUAUAAAUUGGGGCUAUCGUGAUACCUUAACUUCAUGUGAUUUCGAG
>S30 AGUAAAAACCUGCCCCUUACUGGUCUCUAGGUUAUACCUGUGUUUGUCAA
>S29 GUUCAGAUUAUCAGCUUUGCGUUUACAUUUCGGGCCGGUCCUGUUGCGGU
>S37 CUAUUAGUUCCCCUCCAUUUGGUGCACUGAACGUGGUGACUGUUGCGUAU
>S28 UACAAAUUCCACCUAUCUAAGGUAAUUUUGUGGCGCACGGCAUUUGGGUU
>S27 UAGCCAUGUCGGUAAAGAUUGGUCGGUUUCACCGGUGAUAUAGUGAUAU
>S26 CUUUAUCAGCUGGUGUGUUUUGUUAUACAGUCAUAUGUAUUUCGGUUGU
>S34 GCUAAAUUAUAAGUUGUGACUCAGTGUUCCAGUUAACCCAAGGCUCGGG
>S45 UCAAGUUUUUUCGCUGCACUGAAAGUGGCGAUAAUGGCCAGUGUCAGGAU
>S53 GUUAAUAGCCACGUUGUACUGGCCCCCUCAAGGUCGUGUCAAUAUGUGUC
>S44 UUCAAUAUUCUCACGCAUGUUGAUUUGUCGUUAGGGGUUGUUGUUUCAUG
>S41 AUACGGUGGGUGCUCUCGGUACCCGUUGAUAAACUAGUGUCGUGGUGGUUG
>S40 AAUUACCGCUAAAAGGGACCUUAAGAUCGGUAGAAUUCGAUGUGUAUGCU
>S60 UUCUUCUUGUUGUUAGCAUUUUCCGUUGUAGUAGUGUCUGAUUGCGUGG
>S59 UAUAGUAUCAGAAUGGGCUGCUUAUGGGGCGUCAGCGUGGGGUUUCGUG
>S58 UUUCUUGUUUUUCCGUGGUGGUCAACUUCAUGUUGUCACAAGAGACGUUU
>S57 CUUUCUUUCUAGCACUAAAAAAUUGAUUAUGGCUUGGGGUUUUUUGUGGG
>S56 CUCCUCUAAUCCCGGUUAGGCGCUACCUACUGAUGUUCUUACUUUCCGGGG
>S55 CUCCUCUAAUCCCGGUUAGGCGCUACCUACUGAUGUUCUUACUUUCCGGGG
>S54 ACUUUUCCCUACAUCGGGCGUAACCUCUUUUCGACGAAACAGUGCUGU
>S53 GCACUGUUGGCUCGGGUGCAUGUAGUUUCGCCAUCGAUAUAAGUCGGGGA
>SV CAGUCCUUUCGUGGCGUGUUUGUAAGGAUCAGCCUAGGGUGAUAGGAUUU
>SW AGCUUCUACUUAUGAGGUACUCCAGUGGUCGAUUGAUGUUUUGAGUAUGG

Figure 4.11 – Sequence data obtained after four rounds of SELEX selection using prefilled Ni-NTA agarose beads containing His-tagged *Tb*ZFP3.

After four rounds of selection using His-tagged *Tb*ZFP3, the selected PCR products were cloned into pGEM T-easy and 40 clones sequenced. Each sequence was aligned to show the correct orientation of the random 50mer as it would be represented within the RNA. The UCAG motif is shown in red, with the final U/C shown in green if present. UUUUCC is highlighted in blue.

```
>B23 GUUUUGAUCUAGACUAUAGUAAGUCGGUAAUUUUUAACGUUUUCGCGUGU
>B1  CUUUCACUACGCGUGUCGACCCGUUUUAACCAAGCUUAUUGCCGCGUGUU
>B13 AUUUCGGGAUUAAGGUUGCGGGUGAUUUCUUUUUUCAUGGGGUGUUGUUU
>B21 GAAGCAUGUGCCGGUGGUGGGUUUUGGGGAAAUUUUUGUGUACAGCCUCG
>B12 AUUUACUUCUCACAGUGGUCCGGUCUACCUCCGUAGGAUGUGUGGAUGGU
>B20 ACCACUCGGAACCUCUUCUUUUCGGUCGGUCGUGUGGUAUGGCGUGGUA
>B19 UUACUAGUAUAGCAUUGACGUUCAUCGGCUAUAUGCGUCGGUACCGGUGG
>B18 GUAUUUUUCUUCCCACCGGUGUCACAUCGGAUUGGGUUGUUGUAUUAGUUA
>B9  UUAAGUACUCUCCGCAAUAGCGUUAUUGUCUCAUUGUUAAUAUCUGGUAG
>B17 UUUCGCUAUAUGGAAGUUCCGCCAGCUACUAUAGUUUCGUCGUCUCUGAG
>B8  AAGAAAAUUCUCAUUCGCUUGCGAAUGCGUUUGAUACUUAUGUGGAUAG
>B16 UUCCGUGUUUAGGUUUGUCGGCGGAUUUUGGACCGUCUGCUGUCAUUUUU
>B25 AUAAUAUAUUGAACGUUCUAUGGGCCUGACUAAUGCGUUGACGCACUUAU
>B24 AUUGUUCUUAACGUUUCUGCUGUUGGGCGGUAUGUAGUUGUGUGCUUUG
>B22 UUAUCAUUCUACGGGUAUUCUAAUUGGUUGUGAUAUUUGAUUAUUUACCU
>B21 GUUAAGUAUAGGCCAUGAUAGUUGAACAAGGUUGAGGCCGAUCGUGUUA
>B28 UUUAGCGUUUGUUAAUACCGUAAAUUCUAGCGUGUCUGGCGCGUUGUUU
>B19 UUCCGAAUAUUUCAAUCUGCUUUUUUAUUUCUUCGGCACAGACGCGGGU
>B18 UCAAUAUCACCCAGUUGGUUUCCUGACACUUCUAUNGCUGGUCGGUGGAU
>B17 UAGCCGAAUAACCCUCCAUGUUUAAUCGGUGUUACUGUUGAAGAUGUUG
>B24 CUACUUUCUUACGAUCUGGAUGGUGAUCGUGGCGUUCGCCUGGUGUAGGGU
>B23 CGCUGCACUAUGCCAGCGGGUAAUGCGUGCCCAUUCAGACAGACCAUUC
>B38 UUUGGCAGUUUUGAGUCCUGGGGACCGUCUGCCGUGGUGAUCGUUUGUGU
>B37 CUUUAGCUGCGGUUUUCGGUGGUCUUGUGGUUUCCAGCUUGUACGUGUGU
>B37 UUGAACUUUACCAAGGGGACAGCUCUUCUAUGUCUGUUUUUAUGUAACCG
>B36 UCACGUUUCAUUACUUAUUCGUGGUUCUGUUUAUACAGGUACGAUUUUUUU
>B35 UAACCUCACAGCGUUUUGUUGAACUCAGAUGUUGANCUGUAGGCUGUUUUU
>B33 GUCAUAAUAUAAAGGCGUUUCUCUCGAAGUUCGACCCAUAGUUUGCAUA
>B32 UUAAAUUCCUCUAUUUUUACGUCAUCGUUGGAAUUAUCCCAGCCUGUGUU
>B31 UUCGUAUUGCGCUACUCUCUUACGUCCUGUCUCCGCAACAUGGUCUUUUU
>BH  AAACUUUAGCUUUGUGCGUUUACCUGUAACUUAAUGGUAAUGCGGUUGA
>BG  UUAGAUGGCUCGGGGGCAGGCCGUGAUCCCAGUUUGCCCGCAUGUAACGA
>BF  CUGAUAAUUUUCUUUUUACAGCGGACAAUGGGCCACGUUAGGUGUGCGAU
>BE  UUGAAUAUGCUUUUAGUCUGUGGCUCUGGGUAACAGGUGUGUCAUGUCUAU
>BD  CAGGAUUCCAGAGUAAUGCUUGAGUCUUGGGUUCUCUGCCCAGGUAUGU
>BL  CUUUCCGAACUGGGGUACAUUAUCCUAGGAUUCAUUCGUAGGCUGGAGUU
>BK  UUAUUUUUGACCAGUUAUACGCUUAAGUGUGGCGCCGCGCUUCUGCCGAU
>BJ  CAUUCUAUAUCAGAUAUUAAGUUUUGCCAAGAACCGUACGCGUUGAAGU
>BA  UUCACUUUUAGCGCGACACGAACCCGUUCUUAGUUUCGUCUAUCUUUGU
>BI  UUGUCUUUUUGUGAGCGGCUGUCUAUUUUCUGCGAUUCGCCGGGUUGUUU
```

Figure 4.12 – Sequence data obtained after four rounds of SELEX selection using empty Ni-NTA agarose beads.

After four rounds of selection using empty Ni-NTA agarose beads, the selected PCR products were cloned into pGEM T-easy and 40 clones sequenced. Each sequence was aligned to show the correct orientation of the random 50mer as it would be represented within the RNA. The UCAG motif is shown in red. No UUUUCC motifs were present in this dataset.

The Regulatory Sequence Analysis Tools (RSAT) programme (van Helden *et al.*, 1998) has successfully identified mitochondrial genes that were regulated in response to catabolite repression in *S. cerevisiae* (Jacobs Anderson and Parker, 2000) and putative motifs involved in the stability of the cytochrome oxidase mRNAs in *T. brucei* (Mayho, 2006).

Motif	Motif Length (nt)	Frequency	Predicted frequency	P Value	E Value	Significance index
U	1	712	472	$2.3e^{-34}$	$9.1e^{-34}$	33.04
GU	2	197	115	$5.5e^{-13}$	$8.7e^{-12}$	11.06
UU	2	191	115	$1.7e^{-11}$	$2.7e^{-10}$	9.57
UG	2	179	115	$8.9e^{-09}$	$1.4e^{-07}$	6.58
GUU	3	74	28	$3.8e^{-13}$	$2.4e^{-11}$	10.61
UUU	3	69	28	$4.8e^{-11}$	$3.1e^{-9}$	8.52
UUG	3	67	28	$3e^{-10}$	$1.9e^{-8}$	7.72
UGT	3	64	28	$4.1e^{-9}$	$2.7e^{-7}$	6.58
UUUC	4	32	7	$3.2e^{-12}$	$7.9e^{-10}$	9.10
TGTT	4	30	7	$6.8e^{-11}$	$1.7e^{-8}$	7.77
GUUU	4	26	7	$2.1e^{-8}$	$5.3e^{-6}$	5.27
UUUU	4	26	7	$2.1e^{-8}$	$5.3e^{-6}$	5.37

Table 4.1 – RSAT Oligocounting analysis comparing the *Tb*ZFP3 dataset against an equiprobable nucleotide model.

The dataset containing 40 clones from the 4 rounds of selection using His-tagged *Tb*ZFP3 (Figure 4.12) was analyzed using the oligo-analysis programme available from Regulatory Sequence Analysis Tools (RSAT). This counted the frequency of oligo over-representation within the *Tb*ZFP3 dataset when compared to an equal probability nucleotide model, the results of which are shown in the above figure. Only results that gave a significance index of over 5 (i.e. occurred over 1 in 100 000 times) are included.

Initial RSAT analysis of the *Tb*ZFP3 dataset was performed against a theoretical 'control' population based upon an equiprobable nucleotide frequency, i.e. each nucleotide has an equal probability of being in each position within the oligonucleotide. According to the manufacturers, Sigma-Genosys, the original input

DNA oligonucleotides were synthesised with an equiprobable nucleotide frequency. Hence, any motifs present within the 50mers should have been present in equal proportions. The analysis in Table 4.1 demonstrated that the round 4 selected sequences were highly U rich, these occurring 1.5 times more within the round 4 selected sequences than within the control population. The probability of this result occurring by chance is $2.3e^{-34}$. Thus, this implies that *Tb*ZFP3 may be binding U rich sequences. Other sequences identified with high probabilities include UUUC ($3.2e^{-12}$), GUU ($2.4e^{-11}$), UU ($2.7e^{-10}$) and GU ($8.7e^{-12}$). All the motifs identified were very U rich and short. While these sequences identified are not the classical ARE's, the overrepresentation of U within this population suggested that *Tb*ZFP3 may be binding a similar sequence.

While the equiprobable frequency method is a valid analysis method, it fails to account for any non-specific experimental selection that may have occurred during the SELEX procedure. Hence, a better analysis was to use the empty 'beads dataset' that had undergone 4 rounds of selection as a background control and compare this against the '*Tb*ZFP3 dataset'. This approach should eliminate any selected motifs that are caused by the nonspecific binding of the RNA pool, e.g. to the Ni-NTA agarose beads themselves or any of the other reagents used in the SELEX procedure. Therefore the 'beads dataset' was used to create a frequency table of oligomers, which could then be used as the background control. The *Tb*ZFP3 dataset was then analysed against this background set to identify oligomers that were overrepresented. The results of the analysis are shown in Table 4.2.

Motif	Motif Length (nt)	Frequency	Predicted frequency	P Value	E Value	Significance index
UCAG	4	17	3	$2.7e^{-8}$	$6.9e^{-6}$	5.16
UCAGU	5	6	0.08	$4.8e^{-10}$	$4.8e^{-7}$	6.32
UCAGC	5	6	0.08	$4.8e^{-10}$	$4.8e^{-7}$	6.32
UUUUCC	6	5	0.02	$3.1e^{-11}$	$8.7e^{-8}$	7.06

Table 4.2 – RSAT Oligocounting analysis comparing the *Tb*ZFP3 dataset against the empty beads dataset

The '*Tb*ZFP3 dataset' was analysed against the 'beads dataset' as a background model to remove any motifs that were over represented. This analysis counted the frequency of oligonucleotide sequences over-represented within the *Tb*ZFP3 dataset when compared to the empty beads dataset. Only results that gave a significance index of over 5 (i.e. occurred over 1 in 100 000 times) are included.

Strikingly, the most prevalent oligonucleotide was UCAG, which occurred 17 times in the selected population in comparison to its predicted representation of 3, a 5-fold enrichment. An extension of this UCAG sequence, in the form of UCAGU and UCAGC were also highly over represented in comparison to their predicted frequency in the background dataset, with 6 occurrences compared to the predicted frequency of 0.08. The only other sequence identified was UUUUCC, which was again highly overrepresented in comparison to the control population. However due to the high frequency of U's present in the background dataset, this may be a carry over from the control population selected with Ni NTA agarose beads alone. Unlike the equiprobable model, no overrepresentation of U's was observed within this dataset, implying that both the '*Tb*ZFP3 dataset' and the 'beads dataset' must be very U rich. Indeed, RSAT analysis of the beads dataset against an equiprobable model also showed a significant overrepresentation of U residues (data not shown). This

implies that the high numbers of U nucleotides within the selected material are not due to the presence of *Tb*ZFP3, but to some other experimental factor.

Within the *Tb*ZFP3 dataset, there were 17 occurrences of the UCAG motif within 13 separate aptamers, such that 1/3 of all aptamers contained this motif, with 4 aptamers containing multiple copies. In contrast, UCAG only occurred 3 times within the beads dataset. Therefore, there is a 4.4 fold enrichment for this motif between these datasets, which given that only 4 rounds of selection have been undertaken, was very interesting. The distribution of these motifs within the datasets is highlighted within Figures 4.11 and 4.12. Although the UCAG motif lacks any similarity with other ARE elements or $C_{X8}C_{X5}C_{X3}H$ RNA binding motifs previously published, the putative RNA binding motif for *Tb*ZFP3 might be a core containing the UCAG with either a U or C as the final base pair.

4.9 Investigation of the putative UCAG motif within the RNA: TbZFP3 complex as assayed by an Electrophoretic Mobility Shift Assay (EMSA).

In order to identify if the UCAG C/U sequence was a putative RNA binding motif for *Tb*ZFP3, an EMSA was undertaken to investigate RNA binding *in vitro*. Two clones were identified from the *Tb*ZFP3 dataset, one containing both the UCAGU and the UUUUCC motifs (S17) and one lacking the UCAG motif (S13). These would be used to test the binding efficiency of *Tb*ZFP3 to each RNA in an EMSA. Each clone was amplified and then transcribed using radiolabelled ^{32}P UTP. The probes were then incubated in a binding reaction with 5μg of His-tagged *Tb*ZFP3 with specific and non-specific competitors for 15 minutes before being analysed on a native acrylamide gel. Cold RNA from each clone was also added in a competition assay, while tRNA was added to ensure specificity of any protein:RNA complexes formed.

As S17 possessed both putative RNA binding domains, it was the first to be investigated. As indicated by the arrow in Figure 4.13, there was a clear shift to a larger molecular weight complex, implying that when incubated with *Tb*ZFP3, *Tb*ZFP3 was binding the labelled RNA (lane 2) when compared with the free RNA (Lane 1). However, the presence of increasing amounts of tRNA abolished this complex (lanes 3-5), suggesting the interaction may be non-specific. Similarly, the addition of unlabelled S17 RNA successfully competed away the binding (lanes 6-10). Interestingly, only unlabelled RNA could compete away the binding, since the

addition of 10μg of unlabelled DNA was unable to disrupt the complex (lane 11). Consistent with analysis of the RNA pool (Figure 4.9), an additional high molecular weight bandshift was observed in lane 2 (indicated by an asterisk). However, the addition of tRNA (lanes 3-5) or cold RNA (lanes 8-10) caused a downward shift or ablation of this high molecular weight complex.

The EMSA for S13, which lacks the UCAGU motif, showed a similar binding profile to that of S17 (Figure 4.14). Thus, there was a clear shift to a defined higher molecular weight complex (Lane 2, indicated by the arrow), which was abolished by the addition of increasing amounts of tRNA (Lanes 1-5). In this case, however the presence of a very pronounced high molecular weight shift was also seen (asterisk), which was competed away by the addition of unlabelled RNA (lanes 6-10). However, there appeared to be little difference between the EMSAs undertaken with S17 and S13, suggesting that binding of *Tb*ZFP3 to these sequences is non-specific.

Thus, the data indicated that recombinant His-tagged *Tb*ZFP3 will bind RNA, but in a non-specific manner. It appeared that the UCAGU motif identified from the SELEX procedure is unable to be verified by EMSA. S17 also contained the UUUUCC sequence, which was significantly overrepresented within the SELEX dataset (Table 4.2). However, as no increased binding affinity was observed with S17 when compared to S13, it could not be confirmed that these are the putative RNA binding motif for TbZFP3, at least when analysed under the current *in vitro* conditions.

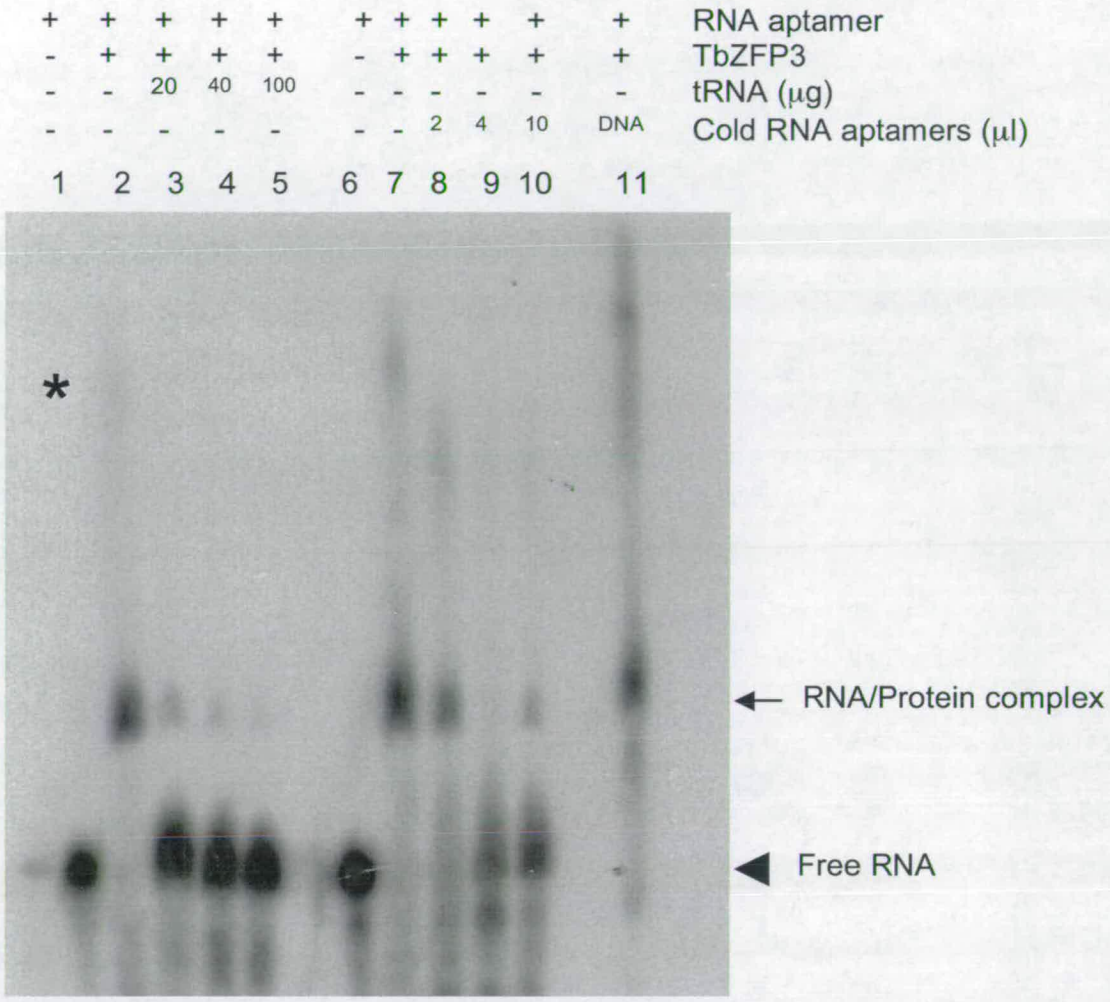

Figure 4.13 – EMSA Competition binding experiments containing the putative TbZFP3 motif UCAG, which show the strength of the RNA/protein complex in the presence of cold RNA and tRNA.

EMSAs were run as described in 2.14.4. Competition experiments using specific (unlabelled RNA) and non-specific (tRNA) competitors of S17. A schematic representation of the contents of each lane is shown above. Before the addition of the labelled RNA to the His-tagged TbZFP3, the reactions were incubated with increasing amounts of tRNA (0, 20, 40 and 100μg, lanes 2-5) or unlabelled RNA (0, 2, 4 and 10μl, lanes 7-10). Free RNA probe was included as a control (lanes 1 and 6). Lane 11 contains 10μl DNA that was used as a specific competitor. The arrowhead indicates the free RNA probes; the black arrow indicates the RNA/protein complex while the asterisk points to a large unknown molecular weight protein/RNA complex.

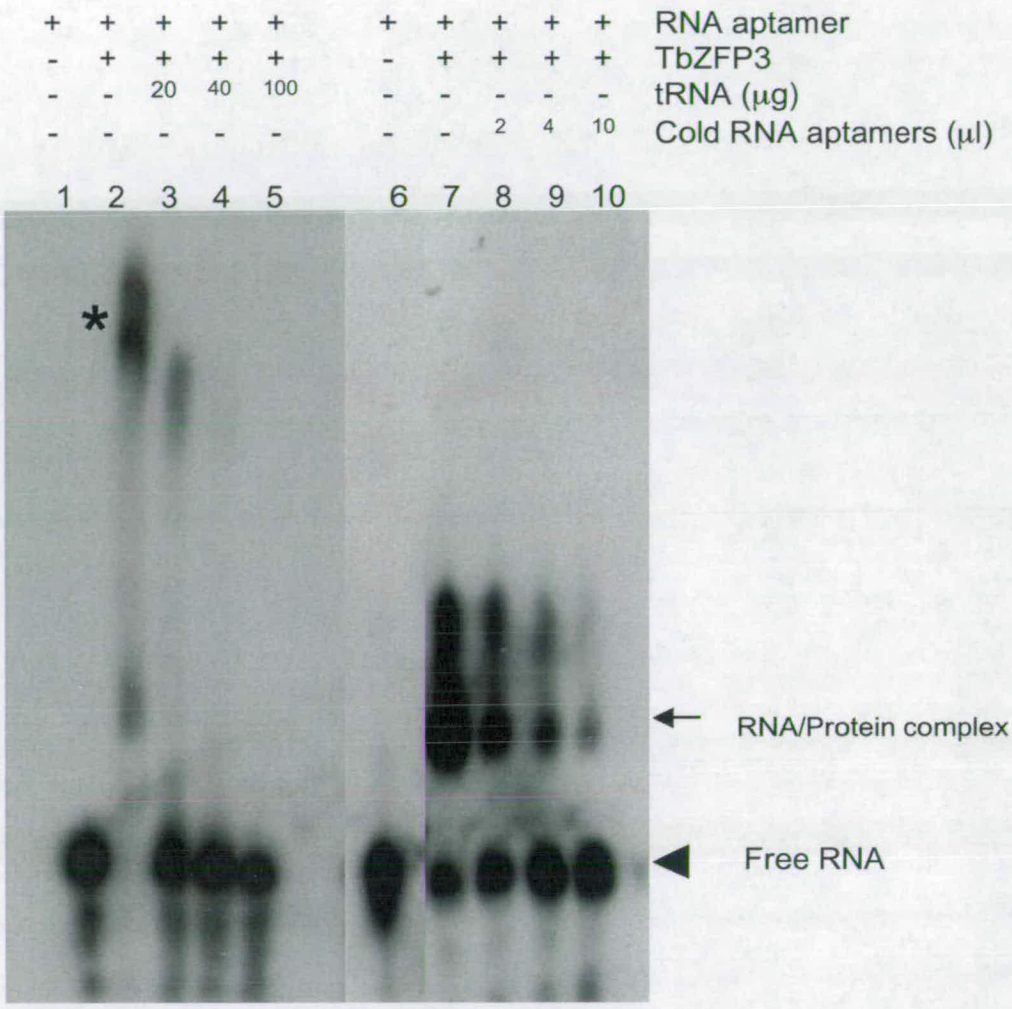

Figure 4.14 – EMSA Competition binding experiments lacking the putative TbZFP3 motif, which shows the strength of the RNA/protein complex in the presence of cold RNA and tRNA.

Competition experiments using specific (unlabelled RNA) and non-specific (tRNA) competitors of clone 13. A schematic representation of the contents of each lane is shown above. Before the addition of the labelled RNA to the His-tagged TbZFP3, the reactions were incubated with increasing amounts of tRNA (0, 20, 40 and 100μg, lanes 2-5) or unlabelled RNA (0, 2, 4 and 10μl, lanes 7-10). Free RNA probe was included as a control. The arrowhead indicates the free RNA probes; the arrow indicates the RNA/protein complex, while the asterisk points to a large unknown molecular weight protein/RNA complex.

4.10 Genomic context of the UCAG and UUUUCC motifs identified via SELEX.

Although the EMSA could not confirm the interaction of *Tb*ZFP3 with UCAG U/C or UUUUCC, it was decided to analyse the location of these motifs within the trypanosome genome in order to investigate if there was any overrepresentation in either bloodstream or procyclic enriched transcripts. Microarray data (kindly provided by Stefanie Brems, Joerg Hoheisel and Christine Clayton, University of Heildelberg and available at http://www.zmbh.uni-heidelberg.de/Clayton/default.shtml) was used to identify transcripts that were up regulated within bloodstream and procyclic stages. Once identified, the first 300nt after the stop codon (approximately the median 3' UTR length in *T. brucei* (Benz *et al.*, 2005)) were screened using PERL scripts (kindly aided by Dr Katelyn Fenn) for presence of UCAG U/C and UUUUCC. This was then extended to analyse all the 3' UTR's of genes predicted within the *T. brucei* genome. Table 4.3 shows the frequency with which these motifs occur within the bloodstream specific, procyclic specific and total predicted 3' UTRs present in the genome.

Table 4.3 shows that there is no over-enrichment of UCAG U/C within either the bloodstream or procyclic UTRs. However, there is a slight enrichment of the UUUUCC motif (a 15% enrichment) within the procyclic UTRs in comparison to the bloodstream and total 3' UTRs.

	Bloodstream UTRs		Procyclic UTRs		ALL UTRs	
Motif	Number	%	Number	%	Number	%
UCAG	168	58.9	115	63.9	7422	68.9
UCAGU	56	19.5	48	26.8	2633	24.5
UCAGC	30	10.5	27	15	1743	16.2
UCAGU/C	86	30	75	41	4372	40.6
UUUUCC	99	34	89	49.7	3729	34.6
Total	287		179		10765	

Figure 4.3 – Screening of the UTRs of the trypanosome genome for the presence and frequency of motifs identified from the SELEX screen.

The motifs identified from the SELEX procedure were used to screen a trypanosome 3' UTR database containing bloodstream enriched, procyclic enriched and total UTR's, as identified by microarray data. Bloodstream UTRs – transcripts up regulated in bloodstream forms. Procyclic UTRs – transcripts up regulated in procyclic forms. All UTRs – All UTRs from predicted genes within the *T. brucei* genome.

4.11 Discussion.

*Tb*ZFP3 is a member of the small $C_{X8}C_{X5}C_{X3}H$ RNA binding zinc finger protein family and has been reported to enhance differentiation, though its function and mode of action remain largely unknown (Paterou, 2006). The $C_{X8}C_{X5}C_{X3}H$ motif has previously been shown to bind RNA within other organisms (Lai *et al.*, 2002). Hence, a RNA selection method (SELEX) was undertaken to identify any putative RNA sequence motifs that may bind *Tb*ZFP3 and thereby to gain an insight into its function. After 4 rounds of selection, a UGAC U/C motif was identified as a putative RNA binding motif and its ability to form an RNA:protein complex tested using an EMSA. Although a band shift was observed, this was rapidly abolished by the addition of competing tRNA, suggesting that *Tb*ZFP3 was binding the RNA nonspecifically. While it has been speculated that *Tb*ZFP3 is an RNA binding protein, this is the first evidence to show that it binds RNA directly, albeit nonspecifically.

The SELEX procedure has been widely used to identify protein:RNA interactions, with the number of rounds of selection ranging from 3 to 10(Hori *et al.*, 2005; Worthington *et al.*, 2002) (Ulrich *et al.*, 2002). For this study, 4 rounds of selection were undertaking (due to external factors), after which the RNA pool was assayed for an enrichment. The EMSA seen in Figure 4.10 imply that there is no clear enrichment in the population for *Tb*ZFP3 binding after 4 rounds, since no difference in RNA binding between round 0 and round 4 was observed. Thus, more rounds of selection may needed to be undertaken for further enrichment of the pool enriched

for any *Tb*ZFP3 motif. However, although the analysis does reveal enrichment for particular motifs, like UGACU, the nonspecific binding of the input RNA by *Tb*ZFP3 suggest that a specific enrichment for a certain motif maybe unlikely, even with more rounds of selection.

Another consideration with this method is the presence of RNA selected within the control reactions of recombinant *Tb*PTP1 and the empty beads, throughout the rounds of selection. This suggests that RNA was binding the beads in a non-specific manner, be it to the Sepedex beads themselves or as an interaction between the negatively charged RNA and the positively charged nickel ions. Such nonspecific binding to the beads would also have biased the *Tb*ZFP3 selection procedure and may have contributed to the enrichment of U's within the selected population. The presence of the N terminal Histidine tag may also have affected the folding and binding ability of *Tb*ZFP3, though being situated at the distal end from the RGG and $C_{X8}C_{X5}C_{X3}H$ motifs, this should have been minimised. While the His-tag itself possesses no charge, there is the potential that its presence caused some RNA binding.

The binding buffers and conditions used are of paramount importance in these types of *in vitro* selection procedures, as ionic conditions can affect the RNA binding affinity as well as protein folding. The buffers used were based on those used by (Hori *et al.*, 2005) for the SELEX and those of (D'Orso and Frasch, 2001) though how similar these buffers are to the *in vivo* conditions is unknown. It is possible that these binding conditions were dissimilar to those *in vivo*, explaining the observed

nonspecific binding. The variation in buffers between the SELEX and the EMSAs is also a putative reason why the UCAG C/U motif was clearly enriched after SELEX but not specifically bound by the EMSA. Hence, optimisation of the binding conditions within the EMSA may be required before the correct protein:RNA complex is observed.

The appearance of the large macromolecular complex in all the EMSAs is potentially due to the aggregation of the RNA aptamers, especially if bound by *Tb*ZFP3, which would then migrate at the much higher molecular weight observed. The addition of tRNA to these reactions caused a slight decrease in the size of these complexes, reinforcing the suggestion that it is a conglomeration of RNA aptamers. A range of different monovalent ionic concentrations was tested to observe any effect on this complex, however none was seen, the complex always being observed (data not shown). Another contributing factor might have been the dimerisation or coagulation of *Tb*ZFP3, either by the His-tag, the WW protein interaction motifs or via nonspecific intramolecular forces. Evidence has suggested that *Tb*ZFP3 may interact via the WW motif (Caro *et al.*, 2005), though data from Chapter 5 and other colleagues suggest this is unlikely (Paterou, 2006)(Athina Paterou, personal comminication).

The lack of specificity for *Tb*ZFP3 is intriguing, as although there are many factors that affect protein:RNA complexes, the bandshift observed by binding *Tb*ZFP3 to the random initial pool of RNA suggests that it is binding non-specifically to the RNA under these conditions. Other eukaryotic proteins containing the $C_{X8}C_{X5}C_{X3}H$ motif

tend to have a defined motif with which they bind such as TTP that binds the UUAUUUAUU nonamer (Blackshear *et al.*, 2003). However all these molecules contain at least one tandem copy of the $C_{X8}C_{X5}C_{X3}H$ motif, which may impart a higher degree of specificity. For example, each of the TTP $C_{X8}C_{X5}C_{X3}H$ zinc fingers interacts with one UAUU motif, from the full UUAUUUAUU sequence (Hudson *et al.*, 2004). It is possible that the UCAG C/U motif identified is only part of the RNA binding sequence and that an interacting partner protein will complete the RNA binding motif, such as *Tb*ZFP1 (Chapter 5).

The presence of the RGG box within *Tb*ZFP3 also complicates matters, as this is another putative RNA binding domain. While it has been shown that the RGG box recognises G-quartet structures (Darnell *et al.*, 2001; Schaeffer *et al.*, 2001), it is also suggested to be a non-specific motif present to enhance the RNA binding affinities of other RNA motifs (Burd and Dreyfuss, 1994). Thus, the possibility exists that the RGG motif was the important motif in binding RNA during the SELEX and EMSA procedures, which would explain the nonspecific binding. Future work could involve repeating the SELEX procedure with a truncated protein lacking the RGG or $C_{X8}C_{X5}C_{X3}H$ motif, which may shed further light on which motif was specific for the UCAG U/C sequence. However, data from polysomes analysis using an ectopic overexpression *Tb*ZFP3 mutant, lacking the $C_{X8}C_{X5}C_{X3}H$ motif, which is unable to associate with polysomes, unlike the wild type *Tb*ZFP3. Therefore the implication is that the $C_{X8}C_{X5}C_{X3}H$ motif is integral for polysomes association.

The presence of the WW motif within *Tb*ZFP3 allows interactions with other proteins, such as *Tb*ZFP1, which has been shown to interact with *Tb*ZFP3 *in vitro* and *in vivo* by yeast two-hybrid and co-immunoprecipitation respectively (Paterou, 2006; Caro *et al.*, 2005). Hence, an interacting partner protein, such as *Tb*ZFP1, may provide the specificity for RNA targets with *Tb*ZFP3 simply tethering the molecule, or it maybe that the combination of both $C_{X8}C_{X5}C_{X3}H$ motifs that provides the specificity. This presents the possibility that the *Tb*ZFP1:*Tb*ZFP2 complex may interact with RNA via some or none of the following: the 26mer, poly C rich motifs (Morking *et al.*, 2004), a G quartet (Darnell *et al.*, 2001) or the UCAG U/C motif isolated from this study. Finally, recent findings suggest that *Tb*ZFP3 is associated with the translational apparatus (Paterou, 2006), whether directly or indirectly, and it is this machinery that could provide the specific signals. The lack of enrichment of either UCAG U/C or UUUUCC from either the bloodstream or procyclic UTRs suggests that these motifs are unlikely to be biologically relevant, although the context or presence within 2^{o} structure may be important.

The data presented in this chapter shows that RNA binds *Tb*ZFP3 in a non-specific manner *in vitro*, although whether this is the case *in vivo* remains to be investigated. The SELEX procedure produced two motifs which were enriched within the *Tb*ZFP3 population, UCAG U/C and UUUUCC, though under the conditions used within this study it was not possible to confirm by EMSA that these were specific RNA binding targets for *Tb*ZFP3. Identification of target RNAs may shed light on the specificity of *Tb*ZFP3's $C_{X8}C_{X5}C_{X3}H$ RNA binding motif, while identification of further interacting partner proteins would provide clues as to whether *Tb*ZFP3 was acting as

part of a complex or in isolation. A deeper understanding of these interactions would enable a more thorough dissection of the RNA binding mechanism of *Tb*ZFP3 and its role in the life cycle regulation of the trypanosome.

Chapter 5:

Using a Yeast Two-Hybrid approach to screen for novel proteins that interact with the *Tb*ZFP family in *Trypanosoma brucei*.

5.1 Introduction.

Direct protein-protein interactions are vital in biological systems to create large molecular weight complexes from smaller subunits and to allow for a more flexible modular approach. For example, it enables proteins to undertake multiple roles within different developmental stages within the organism, or in different tissues or under different physiological conditions. The *Tb*ZFP family is an example of this, as all 3 proteins have been implicated in the process of differentiation from bloodstream to procyclic forms (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou *et al.*, 2006), yet only one, *Tb*ZFP1, is differentially expressed during this process. As both *Tb*ZFP2 and 3 are constitutively expressed throughout the lifecycle, there is a strong probability that *Tb*ZFP2 and 3 would need to interact with other proteins in order for effective differentiation to occur. This is due to *Tb*ZFP2 and 3 being able to interact with the polysome apparatus in procyclics, but not in the bloodstream trypanosome forms.

Members of the TbZFP family were initially characterised on the basis of their small molecular weight (less than 16kDa) and the presence of a $C_{x8}C_{x5}C_{x3}H$ zinc finger RNA binding motif. However, all other eukaryotic $C_{x8}C_{x5}C_{x3}H$ proteins identified to date contain at least two copies of the $C_{x8}C_{x5}C_{x3}H$ motif, with the majority of proteins containing over 3 copies of the motif (Brown, 2005; Lai *et al.*, 2000). As well as the possession of the $C_{x8}C_{x5}C_{x3}H$ RNA binding domain, all 3 members of the *Tb*ZFP family also contain a protein-protein interaction motif (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). *Tb*ZFP2 and *Tb*ZFP3 contain a WW protein interaction

domain, while *Tb*ZFP1 contains the complementary proline rich PY protein interaction domain.

The WW domain family is characterised by the presence of two tryptophan residues that are spaced 20 to 22 amino acids apart in order to form a stable triple stranded beta sheet with a hydrophobic pocket, able to bind proline rich ligands (Bork and Sudol, 1994; Macias *et al.*, 1996). The WW domains are classified into four groups (Groups I to IV) on the basis of the ligand that is recognised. The largest group, Group I, recognises the Proline-Proline- X $_{(\text{any amino acid})}$ -Tyrosine amino acid sequence and is known as the PY motif (Hu *et al.*, 2004; Sudol *et al.*, 1995). Therefore, it is possible to postulate that *Tb*ZFP1 may interact with a protein containing a group I WW domain, due to the presence of the PY motif. Proteins containing the WW domain are involved in a wide variety of cellular processes, including cell cycle control (Kay *et al.*, 2000), transcriptional control (Mosser *et al.*, 1998) and ubiquitination (Sudol and Hunter, 2000).

It is known that ablation of the *Tb*ZFP2 RNA transcript by RNAi inhibits the differentiation process and that ectopic overexpression of *Tb*ZFP2 or *Tb*ZFP3 causes an increase in the kinetoplast-posterior dimension (the Nozzle phenotype) (Hendriks *et al.*, 2001; Paterou, 2004). Moreover, this kinetoplast-posterior extension is dependant on the presence of the WW and $C_{x8}C_{x5}C_{x3}H$ motifs, as no nozzle phenotype is observed in trypanosomes expressing a mutated *Tb*ZFP2 or 3 protein that lack the WW and $C_{x8}C_{x5}C_{x3}H$ domains (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). Additionally, polyribosome analysis has shown that *Tb*ZFP3 exhibits an

association with the polysomes in procyclic cells, which is lost if either the $C_{x8}C_{x5}C_{x3}H$ or WW motifs are disrupted or deleted (Paterou *et al.*, 2006). Although *Tb*ZFP2 has also been shown to associate with polysomes, no mutational analysis has yet been carried out (Paterou *et al.*, 2006). Nevertheless, since the integrity of the WW and $C_{x8}C_{x5}C_{x3}H$ domains are vital for the formation of Nozzle cells and the interaction of *Tb*ZFP3 with polysomes, it can be hypothesised that the interaction with polysomes by *Tb*ZFP2 is also WW and $C_{x8}C_{x5}C_{x3}H$ dependant. Hence, the presence of the WW and $C_{x8}C_{x5}C_{x3}H$ RNA binding domains appear to be vital for the correct function of the endogenous *Tb*ZFP2 and 3 proteins.

Further evidence for association of *Tb*ZFP1, 2 and 3 was provided by co-immunoprecipitation. Thus, using an anti-*Tb*ZFP3 antibody, it was possible to co-precipitate *Tb*ZFP1, 2 and 3 from a procyclic trypanosome cell extract (Paterou, 2004; Paterou *et al.*, 2006). No association was observed using bloodstream trypanosome extract (Paterou, 2004). This implies that *Tb*ZFP1, 2 and 3 are associated within the same complex only in procyclic trypanosomes. However, whether *Tb*ZFP1 interacts directly with *Tb*ZFP2 or 3 remains to be established. Data recently published investigating the interactions of the homologous *Trypanosoma cruzi* Zinc Finger Protein (TcZFP) family suggests that TcZFP1 A and B interact with TcZFP2A (*Tb*ZFP2) and TcZFP2B (*Tb*ZFP3) (Caro *et al.*, 2005). It is also proposed that TcZFP2A and TcZFP2B can homodimerise, as analysed using the yeast two-hybrid system.

In this chapter, a yeast two-hybrid approach was used to investigate the protein-protein interactions that may occur between the *Tb*ZFP molecules. The yeast two-hybrid system was also used to screen a *T. brucei* cDNA library for novel potential interacting partner proteins for the *Tb*ZFP proteins. The aim was to identify interacting proteins and to analyse their protein-protein interactions, in order to further investigate the role and function of the *Tb*ZFP family within the trypanosome.

5.2 The Yeast two-hybrid principle.

The yeast two-hybrid system was developed to identify and verify direct protein-protein interactions *in vitro* and has been used successfully to identify protein interactions in many organisms. The principle behind the system is that the GAL4 yeast transcription factor can be separated into two discrete parts – the DNA binding domain (DNA BD) (also known as the 'bait') and the transcriptional activating domain (or 'prey') (AD) (Chien *et al.*, 1991; Fields and Song, 1989). Individually, the AD and DNA BD are unable to drive transcription of the reporter genes. However, once brought into close proximity by a protein interaction, a functional transcription factor is generated which can initiate transcription.

A fusion protein is created containing the test protein fused to the DNA binding domain, while a second test gene is expressed as a fusion protein to the activation domain. Both fusion proteins are then expressed in *S. cerevisiae*, where the HIS3 and LacZ reporter genes are under the control of an upstream activating sequence, which is bound by the GAL4 DNA BD. If there is a direct interaction between the two proteins, the DNA BD and AD are able to reconstitute the transcription factor and transcription of the reporter genes occurs. This is manifested as growth on media lacking histidine and adenine (SD – His – Ade) and the production of β-galactosidase as assayed by the X-Gal assay. Both selectable markers need to be positive for it to indicate a putative interaction. However, if there is no interaction, the transcription factor is not formed and no transcription of the reporter genes occurs. Both the HybriZAP-2.1 yeast Two-hybrid system (Stratagene) and the

Matchmaker GAL4 two-hybrid System 3 (Clontech) are based on the GAL4 transcription factor principle. A schematic representation of this is shown in Figure 5.1.

An alternate yeast two-hybrid system exists, which is based on the prokaryotic LexA protein (Fields and Song, 1989; Hollenberg *et al.*, 1995). In this system, the entire LexA protein replaces the GAL4 DNA BD and a series of LexA DNA repeats are inserted into the genome in place of the GAL4 upstream activating element. The GAL4 AD is still required to activate transcription. The advantage of this system is that it does not rely on an endogenous GAL4 yeast protein for control of the reporter genes, which should result in a reduced number of false positive results. The L40 yeast strain and pSTT91 plasmid constitute a LexA Yeast two-hybrid system.

The Yeast two-hybrid system has been used successfully in many organisms to identify a range of direct protein-protein interactions, including *T. brucei.* Protein interactions identified include the interaction of CYC2 with CRK3 (Van Hellemond *et al.*, 2000), the interaction of exosome components (Estevez *et al.*, 2003) and the interaction of Expression-site-associated-gene (ESAG) 8 with *Tb*PUF1, a member of the pumilio family (Hoek *et al.*, 2002).

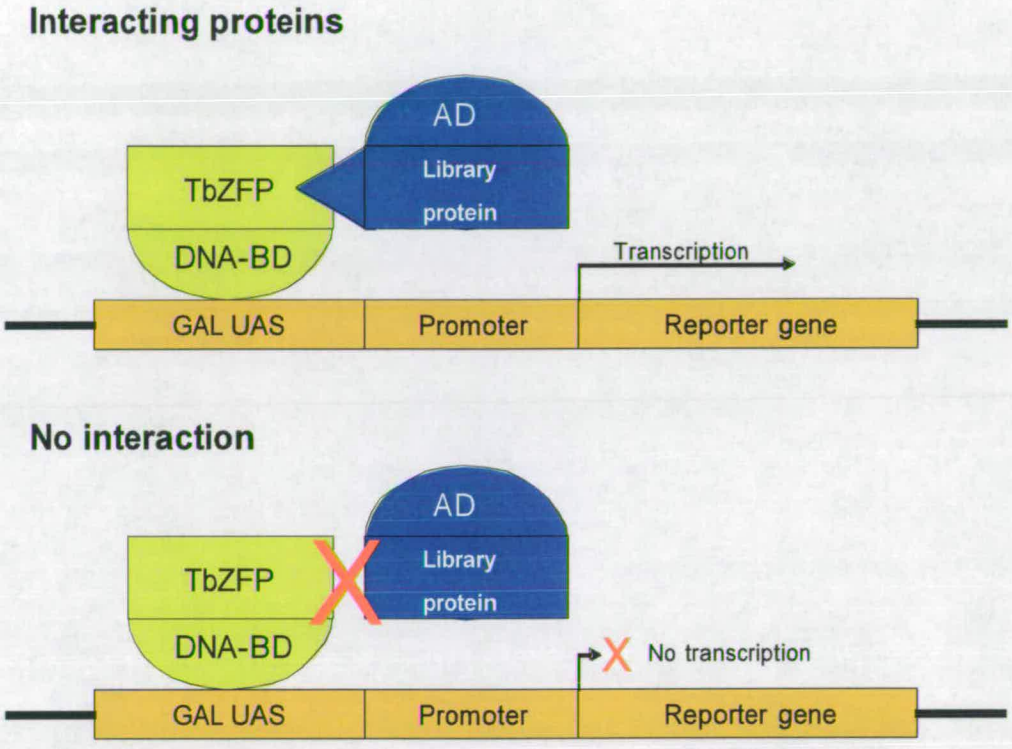

Figure 5.1 – Schematic representation of the GAL4 yeast two-hybrid system for identifying direct protein interactions.

Two fusion protein are created, one containing a protein of interest (*Tb*ZFP) fused to the GAL4 DNA BD, the other containing a gene of interest fused to the GAL4 AD. If a protein interaction occurs between the two proteins of interest, the DNA BD and AD are brought into close proximity and will allow transcription of the reporter genes. If no protein interaction occurs, then no transcription of the reporter genes occurs.

5.3 Creation of the TbZFP fusion proteins for use in the Yeast Two-hybrid system.

With a range of yeast two-hybrid systems available, it was initially decided to use the Matchmaker GAL4 two-hybrid system 3 (Clontech), to identify protein-protein interactions that occur within the *Tb*ZFP family. This was to ensure compatibility with a *Trypanosoma brucei* cDNA library fused to the pGAD10 Matchmaker AD plasmid, which was to be used for the library screen (a kind gift from Professor George Cross, Rockefeller University, USA). This library had been previously used to identify a protein interaction between ESAG8 and *Tb*PUF1, a member of the pumilio family (Hoek *et al.*, 2002). Hence, *Tb*ZFP 1 and 2 were amplified in a PCR reaction with primers that resulted in the integration of a 5' *Nco1* and 3' *Sal1* restriction site at each end of the insert. The PCR product was then ligated directionally in frame into the pGBKT7 DNA BD using the corresponding restriction sites. *Tb*ZFP3 was ligated into pGBKT7 DNA BD using an identical methodology, however the 5' *Nco1* was substituted for a 5' *EcoR1* restriction site for PCR amplification and insertion into the vector. The use of these restriction sites created a fusion protein containing an N-terminal GAL4 DNA BD fused to the C terminal *Tb*ZFP. All constructs were sequenced to ensure the correct insertion of the gene and that the correct reading frame was maintained to produce a functional fusion protein.

To assay whether the fusion protein was correctly translated and expressed within the AH109 yeast strain, the pGBKT7.*Tb*ZFP2 plasmid was transformed into the AH109 yeast strain using a high efficiency polyethylene glycerol/lithium acetate

transformation procedure (Gietz *et al.*, 1992). The yeast colonies were selected on media lacking tryptophan (SD – Trp) and incubated at 30°C for 3 days. Several colonies were then used to inoculate 5ml overnight cultures. The yeast cells were harvested and protein samples prepared, which were analysed by western blot (Figure 5.2). As the pGBKT7 plasmid contains an integral c-Myc epitope tag situated at the 3' end of the GAL4 DNA BD, an anti c-Myc antibody can be used to detect the fusion protein, in addition to the anti-*Tb*ZFP2 antibody. Figure 5.2 clearly shows the GAL4 DNA BD *Tb*ZFP2 fusion protein (indicated by the arrow) as a 32kDa band in lanes 1-6, as detected by both the anti-c-Myc (Panel A) and anti-*Tb*ZFP2 antibody (Panel B). Total trypanosome protein extract was used as a positive control (lane 7), with the endogenous *Tb*ZFP2 clearly visible at 16kDa (indicated by the arrowhead), this not being detected by the anti-c-Myc antibody, as expected.

To create the Activation Domain (AD) (prey) plasmids, the *Tb*ZFP1, 2 and 3 genes were amplified in a PCR reaction with primers that resulted in the integration of a 5' *EcoR1* and 3' *Xho1* restriction site at each end of each gene. The PCR products were then ligated into the pGADT7 plasmid using the 5' *EcoR1* and 3' *Xho1* restriction sites. Again, use of these sites created a protein with the N-terminal GAL4 AD fused to the *Tb*ZFP gene. All constructs were sequenced to verify that the gene had been inserted into the correct reading frame.

Test transformations were then undertaken by pair wise transformation of the DNA BD and AD using the standard polyethylene glycerol/lithium acetate method. Briefly, this involved incubating a 5ml culture at 30°C overnight before diluting the

culture into 20ml YPDa media and growing until an OD_{600}= 0.6 was obtained. The cells were then washed with 1 x TE before the addition of 5µg DNA BD, 5µg AD plasmids and 5µg carrier DNA, along with polyethylene glycerol/lithium acetate. The mixture was incubated at 30°C for 1 hour and then 15 minutes at 42°C. The mixture was then plated onto minimal media lacking tryptophan and leucine (SD – Trp – Leu) to select for yeast transformants. Co-transformation of pGBKT7.ZFP1 and pGADT7.ZFP2 into the yeast strain AH109 gave a transformation efficiency of ~300 cfu/µg DNA. Sequential transformation of the same plasmids resulted in a slightly improved efficiency (3000 cfu/µg DNA), though not the $1x10^{4-5}$ cfu/µg DNA required to screen a library. Despite use of a variety of approaches to improve transformation efficiency, no significant increase was achieved.

Hence, another yeast strain was identified which had excellent transformation efficiency. This strain, L40 (a kind gift of Professor Rolf Sternglanz, Stony Brook University, USA), was based on the LexA yeast two hybrid principle and as a result *Tb*ZFP1, 2 and 3 were cloned into pSTT91 (a derivative of the pBTM116 vector) using the 5' *EcoR1* and 3' *Sal1* restriction site. Mutated copies of *Tb*ZFP2 and 3 in which the WW domain had been deleted were also cloned into pSTT91 using the 5' *EcoR1* and 3' *Sal1*. A mutated copy of *Tb*ZFP1 containing the second proline substituted for an arginine amino acid to give an amino acid sequence of PRPPY (a kind gift of Dr Athina Paterou) was also ligated into the pSTT91 vector. The transformation efficiency was in the order of 4000 cfu/µg DNA for co-transformation and 40 000 cfu/µg for sequential transformations.

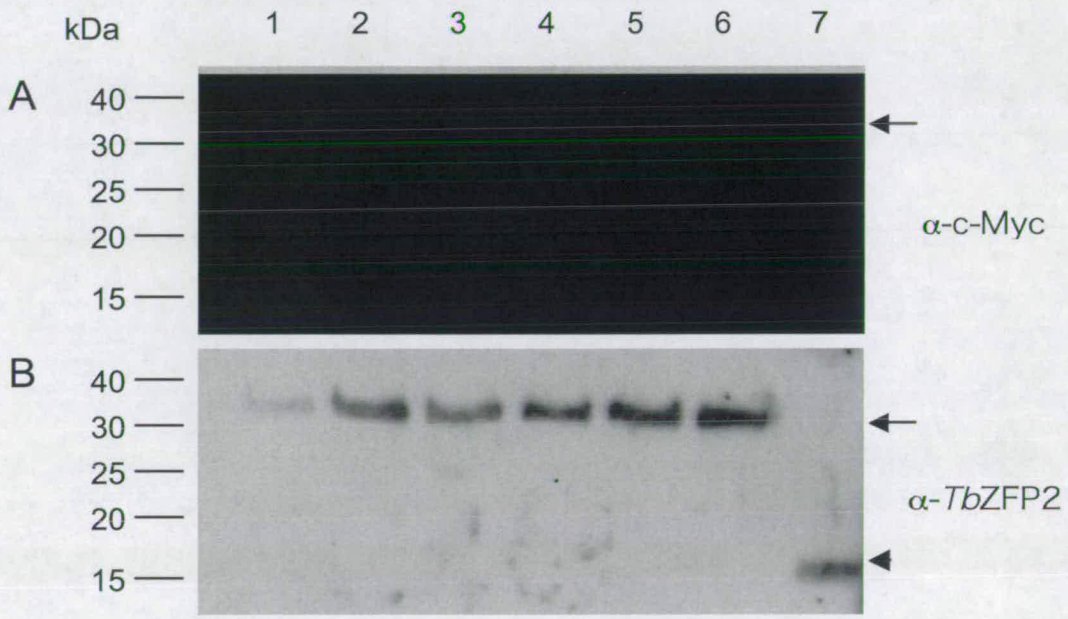

Figure 5.2 – Expression of the pGBKT7.ZFP2 DNA BD fusion protein in AH109 *S. cerevisiae* cells as detected by western blotting.

A 5ml overnight culture of AH109 *S. cerevisiae* cells transformed with the pGBKT7.ZFP2 DNA BD plasmid was grown in SD - Trp and incubated at 30°C until an OD_{600mn}=0.6 was achieved. The cells were then harvested, protein samples were prepared and an equal volume of sample loaded onto a 15% SDS PAGE gel (lanes 1-6). A total trypanosome protein extract sample was loaded in lane 7 as a positive control. The gels were then transferred to nitrocellulose membrane and blotted with anti-c-Myc (A) and anti-*Tb*ZFP2 (B) antibody. The arrows point to the *Tb*ZFP2 DNA BD recombinant protein at ~ 32kDa, while the arrowhead indicates the endogenous *Tb*ZFP2 protein within the trypanosome extract.

5.4 Protein interactions between TbZFP1, TbZFP2 and TbZFP3 as assayed by the Yeast Two-Hybrid system.

The presence of the compatible WW and PY protein interaction motifs within the *Tb*ZFP family implies that these molecules may interact with each other. To test this hypothesis, *Tb*ZFP1, 2 and 3 were expressed as fusion proteins combined with either the LexA DNA BD (pSTT91) or the GAL 4 AD (pGADT7). Mutated copies of *Tb*ZFP2 and 3 lacking the WW domain and *Tb*ZFP1 mutated within the PY motif were also used to investigate complementary interactions of these proteins. These constructs were then transformed pair-wise using a standard polyethylene glycerol/lithium acetate transformation (see Section 2.13.1) into the L40 strain and analysed for the activation of reporter genes.

Figure 5.3 suggests that there was a direct protein interaction between *Tb*ZFP2 DNA BD and *Tb*ZFP1 AD, as well as *Tb*ZFP3 DNA BD and *Tb*ZFP1 AD, as both colonies grew on selective media lacking histidine and adenine (SD – His – Ade, Panel D) and turn blue rapidly during the Lac Z assay (Panel C). This interaction appears to be mediated through the WW motif, as none of the selective markers were activated when the mutated *Tb*ZFP2 and 3 that lacked the WW domain were used in the assay. No other combination of constructs activated both selectable markers, though weak activation of the LacZ gene was observed between *Tb*ZFP2 DNA BD and *Tb*ZFP3 AD. No self-activation of the selectable markers was observed when each DNA BD was transformed with an empty AD plasmid as a control.

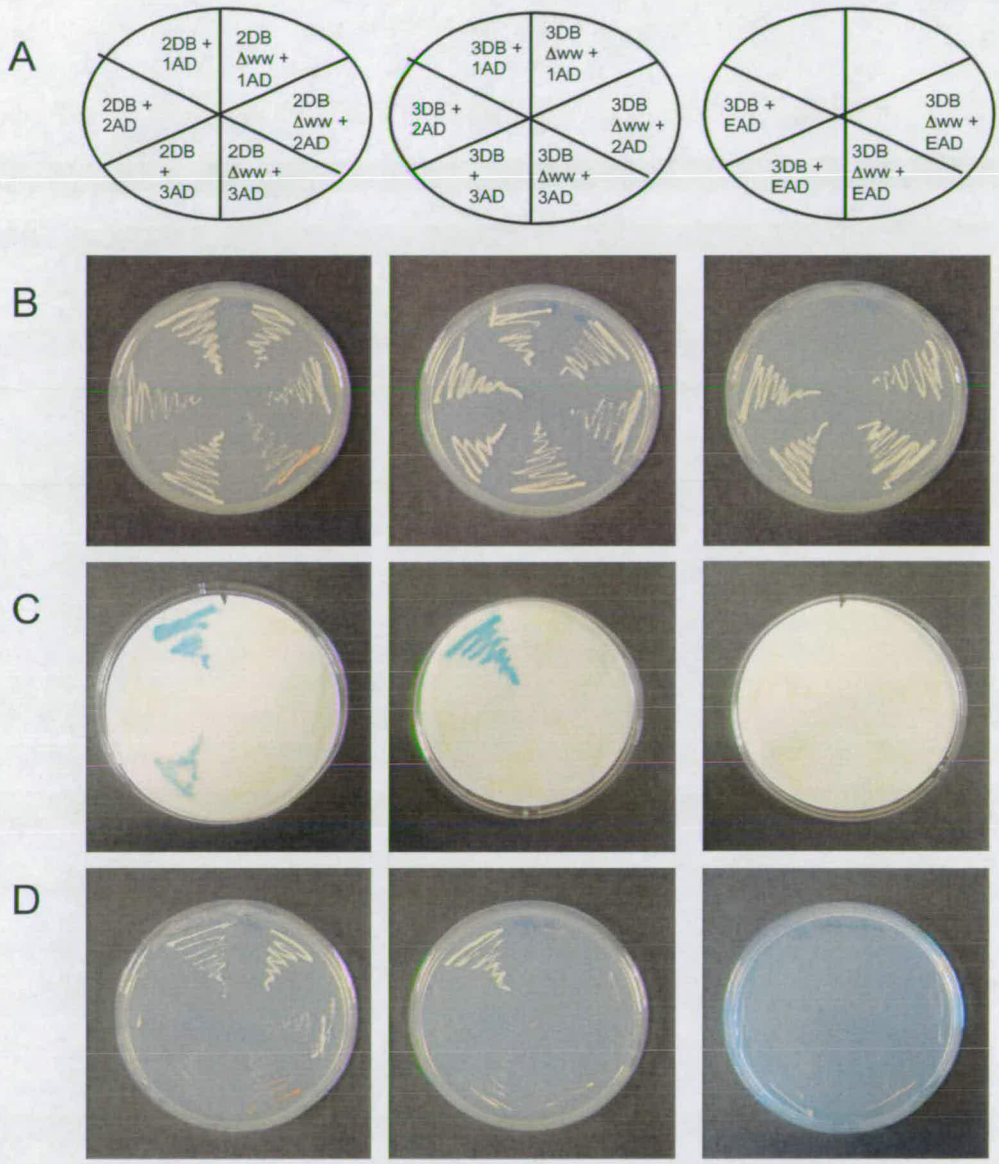

Figure 5.3 – Interactions of the *Tb*ZFP family as assayed by the Yeast two-hybrid system.

Direct protein interaction were assayed between *Tb*ZFP2 or 3 (2DB, 3DB) or mutants lacking the WW motif (2DB ΔWW; 3DB ΔWW) and *Tb*ZFP1, 2 or 3 fused to the AD (1AD, 2AD, 3AD) or empty AD (EAD). A: Schematic representation of colonies; B: Growth on SD – Trp – Leu media; C: X-Gal assay testing expression of the LacZ gene. D: Growth on SD – His – Ade media. A blue colouration in the X–Gal assay and growth on SD – His – Ade is indicative of a protein interaction.

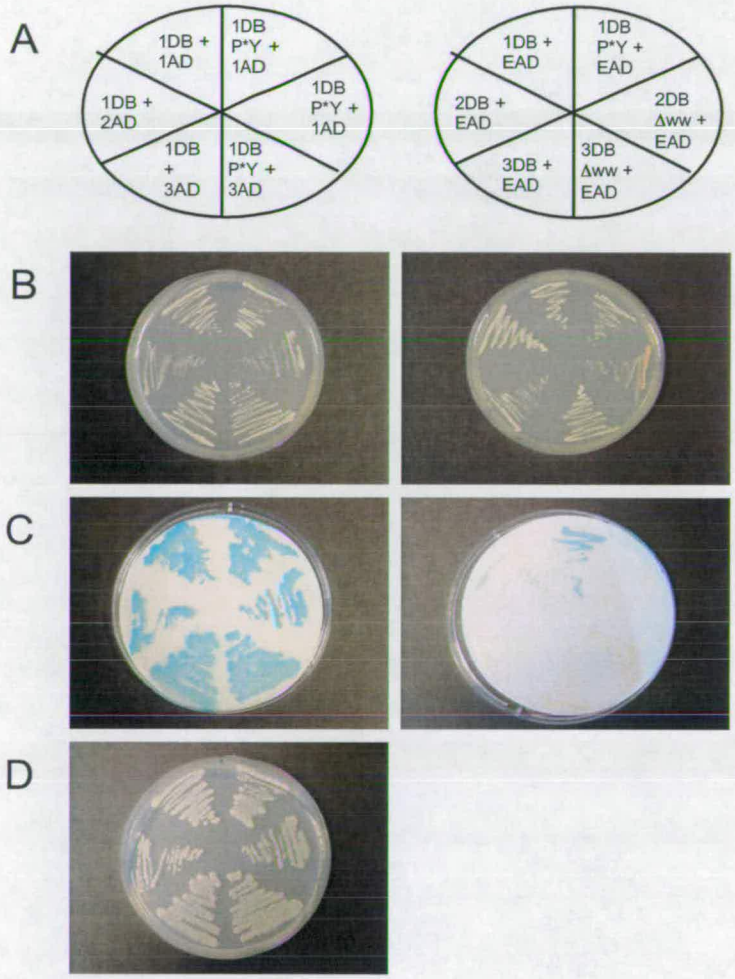

Figure 5.4 – Interaction of *Tb*ZFP1 with other members of the *Tb*ZFP family as assayed by yeast 2-hybrid analysis.

*Tb*ZFP1 and a P*Y mutant version of *Tb*ZFP1 were fused to the DNA BD (1DB, 1DB P*Y) and assayed for protein interactions with *Tb*ZFP1, 2 or 3 fused to the AD (1AD, 2AD, 3AD) or empty AD (EAD). *Tb*ZFP1 and *Tb*ZFP1 P*Y activated the reporter genes when transformed with the empty AD. Hence, pSTT91.*Tb*ZFP1 causes auto-activation of the reporter genes. A: Schematic representation of colonies; B: growth on SD – Trp – Leu media; C: X-gal assay testing expression of the LacZ gene. D: Growth on SD – His – Ade media.

Identical transformations were attempted with *Tb*ZFP1 DNA BD in combination with the other *Tb*ZFPs in the pGADT7 AD plasmid. However, when *Tb*ZFP1 DNA BD was transformed with an empty pGADT7 AD, it was found to activate the LacZ gene (Figure 5.4, right column). Similarly, when transformed into L40 yeast cells in combination with *Tb*ZFP1, 2 and 3 in the pGADT7 AD, all the reporter genes were activated to similar levels as the controls (Figure 5.4 and data not shown). Hence, *Tb*ZFP1 self activates the reporter genes, and was not used further for this analysis.

Thus, the data indicated that *Tb*ZFP2 and 3 interact directly with *Tb*ZFP1 through the WW domain in this *in vitro* yeast two-hybrid screen. The results shown in Figure 5.3 concur with the findings published in *T. cruzi* (Caro *et al.*, 2005), although no homodimerisation of *Tb*ZFP3 (TcZFP2B) was detected. No other interactions were detected, although the presence of a weak blue colouration between pSTT91.*Tb*ZFP2 DNA BD and pGBKT7.*Tb*ZFP3 AD may suggest a weak protein interaction. Alternatively, it may have been a false positive result, due to leaky transcription of the LacZ gene.

5.5 Screening of a Yeast Two-Hybrid Matchmaker library containing T. brucei fusion proteins with TbZFP2.

While attempts have been made to dissect the morphological effects of the *Tb*ZFP family within *T. brucei*, little is known about the molecular interaction of each of these molecules. The ability of *Tb*ZFP1 to interact with *Tb*ZFP2 and *Tb*ZFP3 directly implies an interaction *in vivo* between these molecules during differentiation and within procyclic cells. This is based on the expression profile of *Tb*ZFP1. Yet, *Tb*ZFP2 and 3 are constitutively expressed throughout the lifecycle stages and are known not to associate with polysomes within the bloodstream forms (Hendriks *et al.*, 2000; Paterou, 2004). Therefore, are there other unknown proteins that interact with *Tb*ZFP1, 2 and 3 and what can these molecules tell us about the function of the *Tb*ZFP family?

In order to investigate the molecular interactions of *Tb*ZFP2, a yeast two-hybrid library screen was undertaken to identify proteins that interact directly with *Tb*ZFP2. A *T. brucei* cDNA library containing inserts fused to the pGAD10 AD was used for the screen. The library was created by ligating adapter primers to the ends of the trypanosome cDNA inserts and then ligating the inserts non-directionally into the pGAD10 vector (Hoek *et al.*, 2002). The compatibility of the library with the L40 yeast strain and pSTT91.ZFP2 was verified by test transformations and selection using the nutritional markers (data not shown). Also, the transformation efficiency of this strain was much improved in comparison to the AH109 yeast strain, giving on average 30 000/cfu per µg DNA. This was sufficient for a successful screen of the

library. The controls also indicated that there was no autoactivation of the reporter genes.

The ability of *Tb*ZFP2 to interact directly with *Tb*ZFP1 in the yeast two-hybrid system provides a positive control for identification of unknown interactions from within the library. It would be anticipated that *Tb*ZFP1 should be isolated from the library, as the library was created using a mixture of bloodstream and procyclic RNA, where *Tb*ZFP1 is expressed.

5.5.1 Screening the library for potential interacting proteins.

Test transformations had shown that the transformation efficiency was significantly increased (up to 10 fold increase) when each plasmid was consecutively transformed into the yeast strain, in comparison to co-transformation. Hence, all the transformations for the library screens were done sequentially. Initially, pSTT91.ZFP2 was transformed into the L40 yeast strain using a small-scale polyethylene glycerol/lithium acetate transformation protocol and transformants were selected on SD - Trp media. To confirm the presence of the pSTT91.ZFP2 plasmid, eight yeast colonies were boiled for 10 minutes and the supernatant was used in a PCR reaction using *Tb*ZFP2-specific primers. The amplification of a 400 nucleotide band indicated that the L40 yeast strain had been successfully with the pSTT91.ZFP2 plasmid. This yeast strain was named L40.pSTT91.ZFP2 (Figure 5.5).

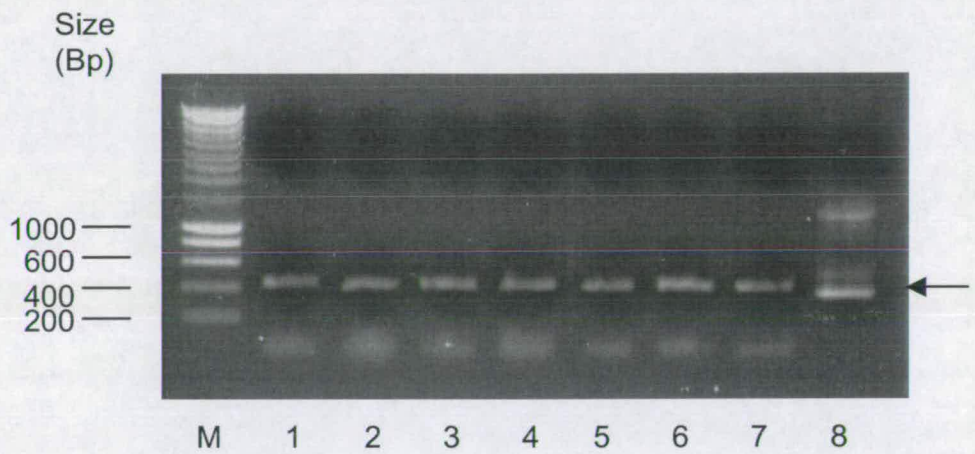

Figure 5.5 – PCR amplification of *Tb*ZFP2 from L40 yeast transfected with the pSTT91.ZFP2 plasmid.

To confirm the presence of *Tb*ZFP2 within the L40 yeast strain, a colony boil PCR analysis was undertaken on 7 colonies using *Tb*ZFP2 specific primers. Lanes 1 – 7 contain yeast colonies transformed with the pSTT91.ZFP2 plasmid; the 400 nucleotide *Tb*ZFP2 insert is clearly shown (arrow). Lane 8 contains pGBKT7.ZFP2 as a positive control for the *Tb*ZFP2 gene.

As L40.pSTT91.ZFP2 had previously been shown not to constitutively activate the reporter genes within the L40 strain and that *Tb*ZFP2 was expressed as a fusion protein (Figure 5.3), a small-scale test library screen was undertaken to optimise the transformation efficiency. Thus, a 50ml overnight culture containing SD - Trp media was inoculated with L40.pSTT91.ZFP2. The following day, the culture was diluted to an OD_{600}= 0.2 and incubated until an OD_{600}= 0.6 was obtained. The cells were then harvested and transformed with the library using a polyethylene glycerol/lithium acetate transformation procedure. Three different concentrations of library vector were used, 0.1μg, 0.5μg and 1μg, in three different experiments.

In order to test the transformation efficiency, 1μl of each experiment was plated onto SD –Trp - Leu media, while the remaining sample was plated out onto 25cm SD – His - Ade plates to screen for any positive interactions. From analysis of the colonies growing on the SD – Trp – Leu media, it was established that the optimum transformation efficiency was obtained at a concentration of 0.1μg of library DNA, which produced 52 000 cfu/μg DNA. The total number of colonies screened was 32000, although no colonies grew on the SD – His - Ade plates, thus indicating that there were no putative interacting proteins within this small test transformation.

The next stage was to undertake a full library screen. Ideally, a minimum of $1x10^6$ colonies had to be screened to generate a one-fold coverage of the library. Hence, based on the previous transformation efficiencies obtained, a 20μg aliquot of the library was anticipated to generate the equivalent of $1x10^6$ transformants, assuming the transformation efficiency was 52 000 cfu/μg DNA. The L40.pSTT91.ZFP2 yeast

strain was incubated overnight in a 50ml culture before dilution to an OD_{600}= 0.2. This 600ml culture was incubated for 3 hours until an OD_{600}= 0.6 was obtained, when the cells were harvested and used in a polyethylene glycerol/lithium acetate library transformation with 20μg of the trypanosome cDNA library. The resulting yeast was then plated over 50 25cm SD –His - Ade plates to select for colonies containing activated reporter genes and incubated for 4 days at 30^0C.

After 4 days, 71 colonies were identified as growing on SD – His - Ade plates. These were then streaked onto SD – Trp - Leu media and incubated for 3 days, before being replica plated onto nylon membrane to test for the production of β-galactosidase and again onto SD -His-Ade plates. Only colonies that grew on SD – His – Ade plates and turned blue in the X-Gal assay would be investigated further as these contained active reporter genes (Figure 5.6). Out of the original 71 colonies identified after 4 days, only 13 tested positive for both growth on SD -His-Ade plates and the production of β-galactosidase. These were colonies 1, 2, 3, 5, 9, 11, 18, 19, 31, 35, 37, 42 and 43. In total the transformation efficiency was calculated to be 42 000 cfu/μg DNA. Therefore $8.4x10^5$ colonies were screened using 20μg, which represents a 0.85 x coverage of the library.

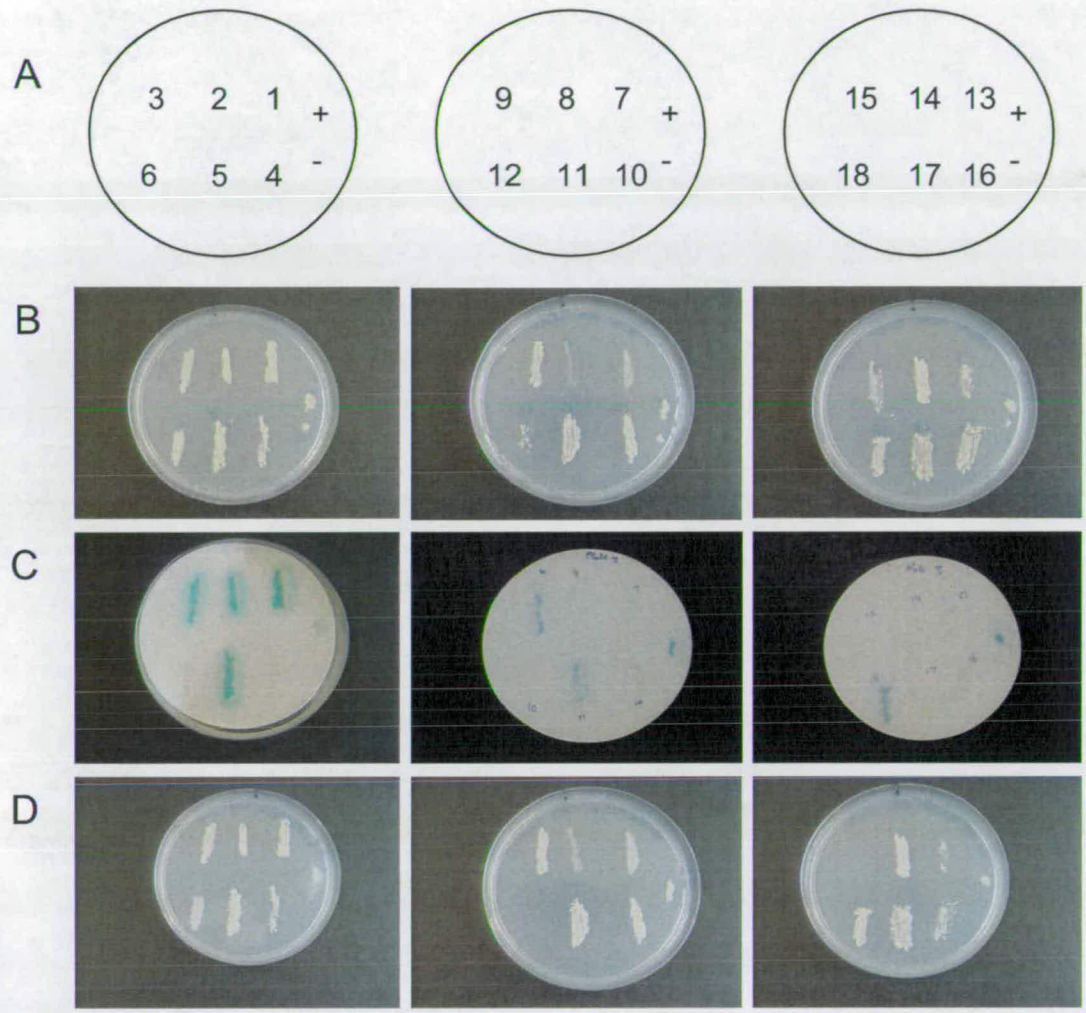

Figure 5.6A – Analysis of the colonies identified from the *T. brucei* cDNA Matchmaker library screen using pSTT91.*Tb*ZFP2 as the bait plasmid.

All colonies that were identified from the library screen as possessing the ability to grow on SD – His – Ade media were tested for expression of the Lac Z and HIS3 genes. Only those that grew on SD – His – Ade media and displayed β-galactosidase expression (colonies 1, 2, 3, 5, 9, 11 and 18) represented colonies containing potential protein interactions. A: Schematic representation of colonies, Controls + = pSTT91.*Tb*ZFP1 and pGADT7.*Tb*ZFP2, - = pGADT7.Lamin and pSTT91.*Tb*ZFP2; B: Growth on SD – Trp – Leu media; C: X-Gal assay testing expression of the LacZ gene; D: Growth on SD – His – Ade media.

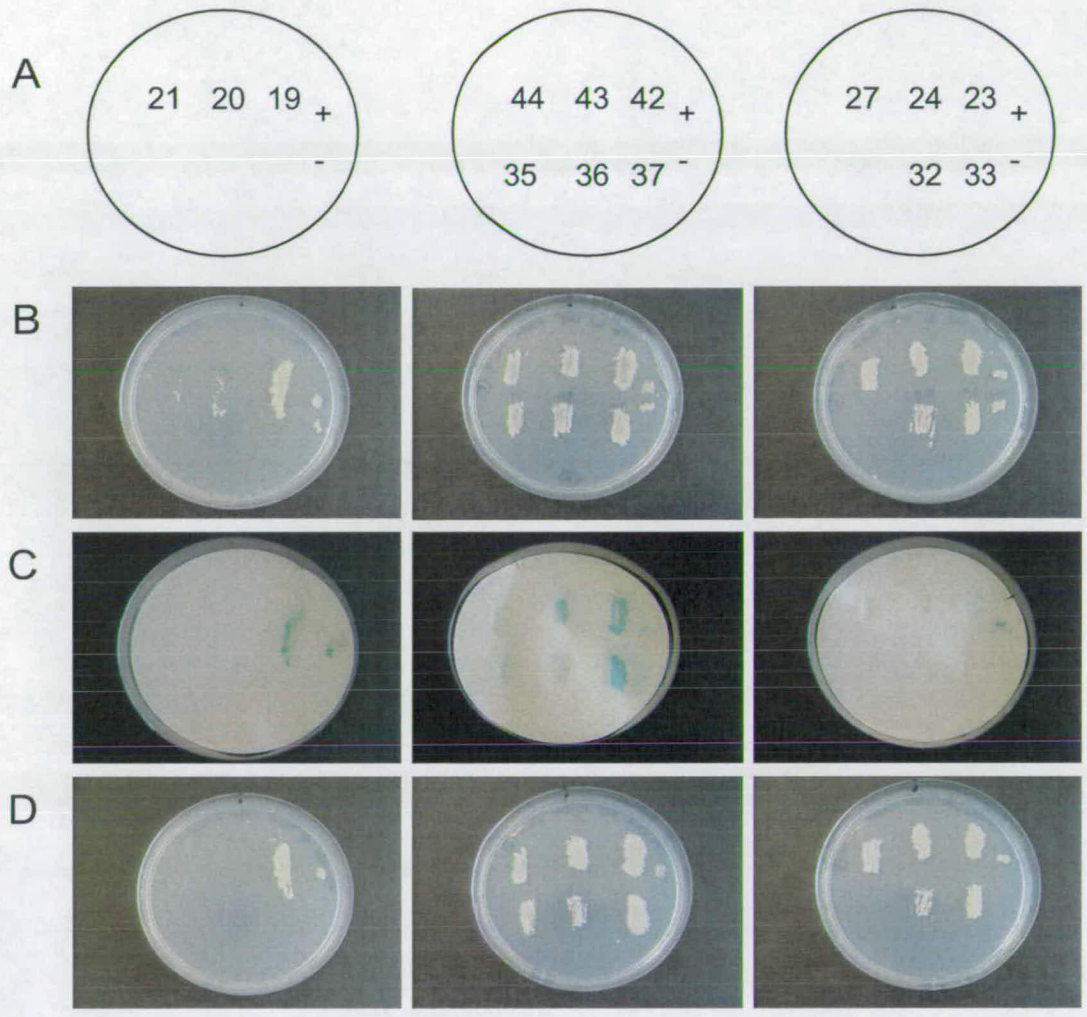

Figure 5.6B – Analysis of the colonies identified from the *T. brucei* cDNA Matchmaker library screen using pSTT91.*Tb*ZFP2 as the bait plasmid (continued).

All colonies that were identified from the library screen as possessing the ability to grow on SD – His – Ade media were tested for expression of the Lac Z and HIS3 genes. Only those that grew on SD – His – Ade media and displayed β-galactosidase expression (colonies 35, 37, 42 and 43) represented colonies containing potential protein interactions. A: Schematic representation of colonies, Controls + = pSTT91.*Tb*ZFP1 and pGADT7.*Tb*ZFP2, - = pGADT7.Lamin and pSTT91.*Tb*ZFP2; B: Growth on SD – Trp – Leu media; C: X-Gal assay testing expression of the LacZ gene; D: Growth on SD – His – Ade media.

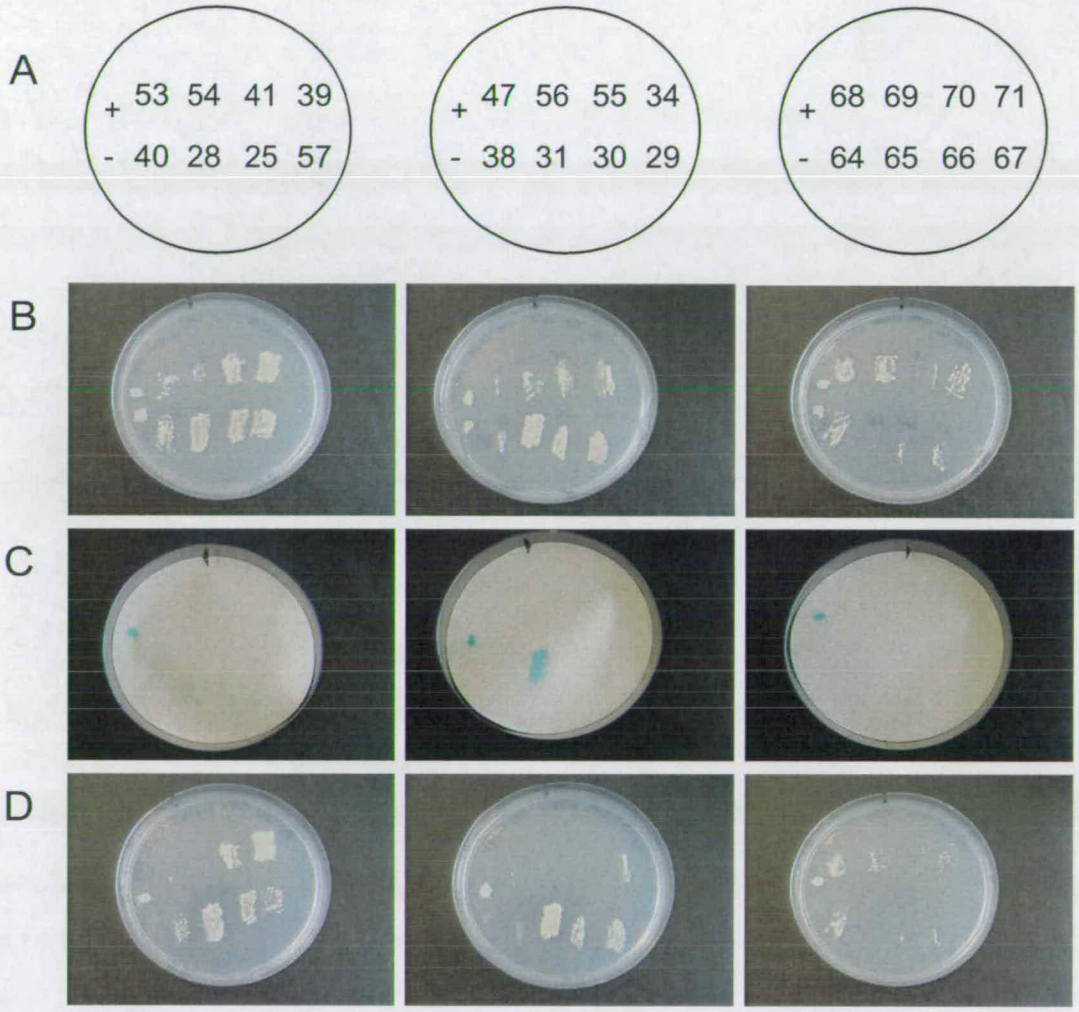

Figure 5.6C – Analysis of the colonies identified from the *T. brucei* cDNA Matchmaker library screen using pSTT91.*Tb*ZFP2 as the bait plasmid (continued).

All colonies that were identified from the library screen as possessing the ability to grow on SD – His – Ade media were tested for expression of the Lac Z and HIS3 genes. Only those that grew on SD – His – Ade media and displayed β-galactosidase expression (colony 31) represented colonies containing potential protein interactions. A: Schematic representation of colonies, Controls + = pSTT91.*Tb*ZFP1 and pGADT7.*Tb*ZFP2, - = pGADT7.Lamin and pSTT91.*Tb*ZFP2; B: Growth on SD – Trp – Leu media; C: X-Gal assay testing expression of the LacZ gene; D: Growth on SD – His – Ade media.

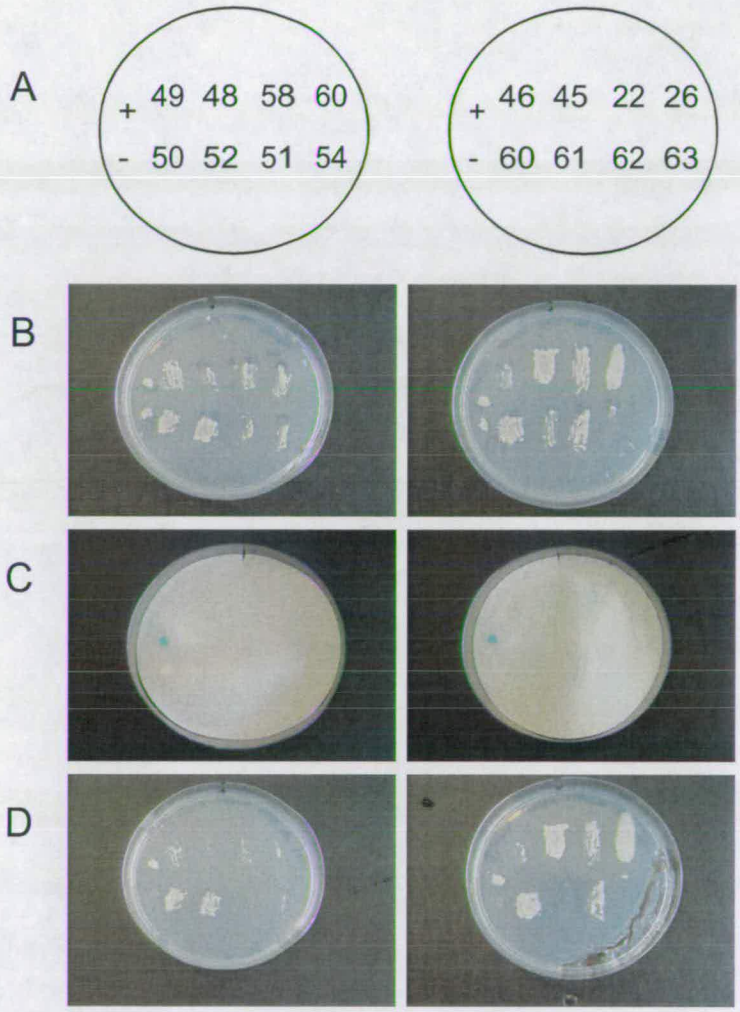

Figure 5.6D – Analysis of the colonies identified from the *T. brucei* cDNA Matchmaker library screen using pSTT91.*Tb*ZFP2 as the bait plasmid (continued).

All colonies that were identified from the library screen as possessing the ability to grow on SD – His – Ade media were tested for expression of the Lac Z and HIS3 genes. Only those that grew on SD – His – Ade media and displayed β-galactosidase expression (no colonies) represented colonies containing potential protein interactions. A: Schematic representation of colonies, Controls + = pSTT91.*Tb*ZFP1 and pGADT7.*Tb*ZFP2, - = pGADT7.Lamin and pSTT91.*Tb*ZFP2; B: Growth on SD – Trp – Leu media; C: X-Gal assay testing expression of the LacZ gene; D: Growth on SD – His – Ade media.

5.5.2 Analysis of colonies that putatively interact with *Tb*ZFP2.

Each of the 13 candidate yeast colonies, which had both reporter genes active, were lysed and the plasmid DNA extracted. The plasmid DNA obtained from the yeast colonies was then transformed using electroporation into *E. coli* and selected for on ampicillin plates. As both the pSTT91 and pGAD10 plasmids contain the ampicillin resistance gene, it was impossible to distinguish between transformants produced by each plasmid based on their antibiotic growth. Hence, PCR analysis was undertaken on 8 individual bacterial colonies of each of the 13 colonies to identify which bacterial colony possessed the pGAD10 plasmid. This was repeated for each of the yeast colonies that demonstrated a positive interaction. The primers used flanked the MCS of the AD plasmid and were specific only to the pGAD10 plasmid. Also, pGAD10 plasmid lacking an insert gave a PCR product size of approximately 200 nucleotides. Thus, any PCR products with a molecular weight larger than 200 nucleotides indicated that an insert was present within the pGAD10 vector. Any lanes that contained no specific PCR product indicated an *E. coli* colony containing the pSTT91.ZFP2 plasmid, as the primers were unable to bind to pSTT91. The results are seen in Figure 5.7. Strikingly, it appears that the majority of plasmids contained no inserts as colonies 2, 5, 11, 19, 31, 42 and 43 all produced inserts of approximately 200 nucleotides in size. However, the possibility exists that these plasmids may contain short sequences of DNA only a few nucleotides long which were able to interact with the pSTT91.ZFP2 fusion protein. A number of colonies did contain inserts; colony 1 had an insert size of 2kb, colony 18 had an insert size of 1.5kb while colonies 3, 9 and 35 all possessed an insert of 0.8kb.

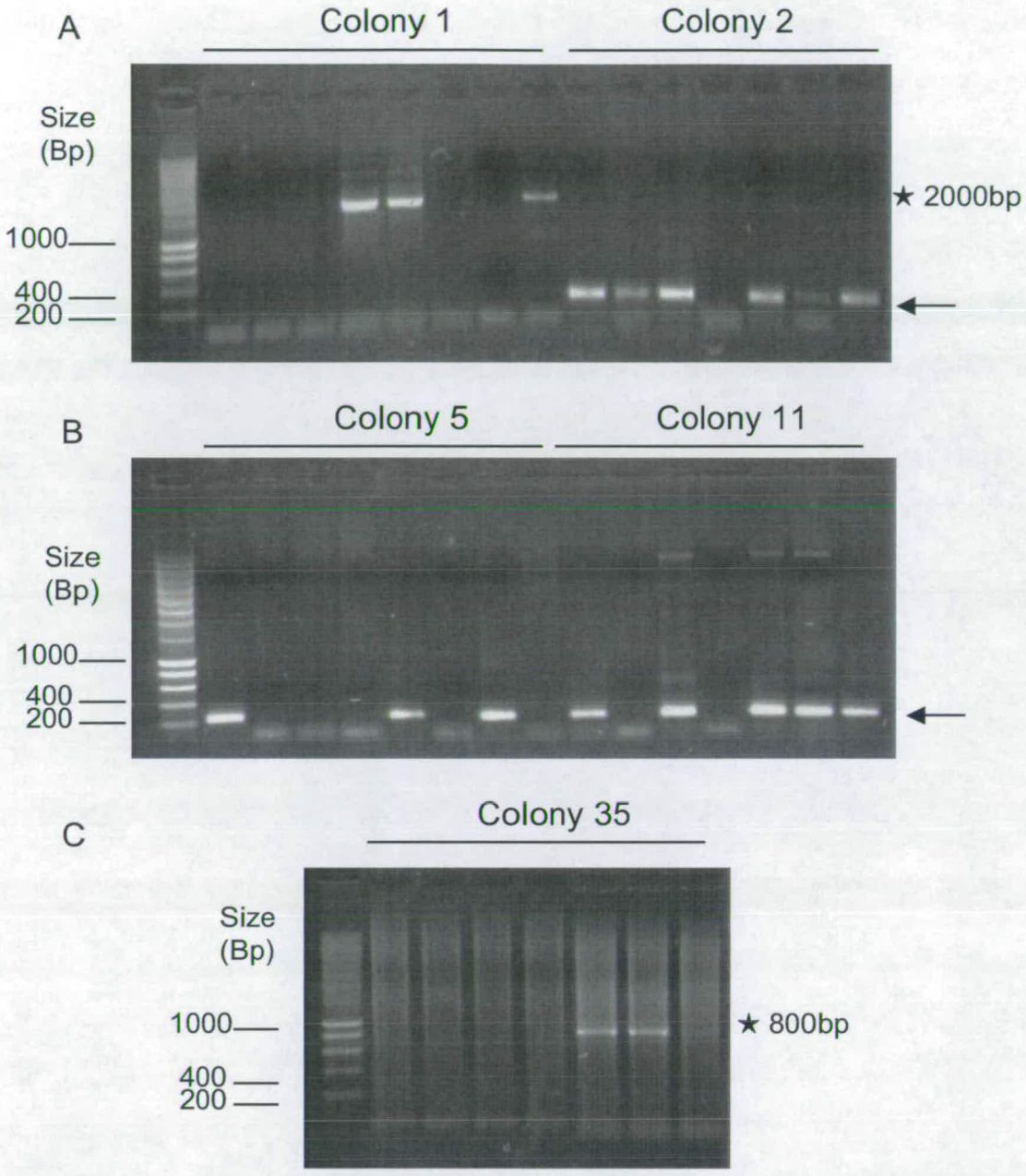

Figure 5.7A – PCR amplification of DNA inserts from the pGAD10 activation domain from *E. coli* colonies.

Initially, *E. coli* was transformed with DNA from yeast colonies positive for reporter gene activation. Eight *E. coli* transformations were analysed by PCR, using pGAD10 specific primers, for pGAD10 plasmids containing trypanosome cDNA. An insert size of ~200bp indicates a pGAD10 plasmid containing no significant insert (arrow). No PCR product implies that the colony is the DNA BD vector, as the primers used were unable to bind this plasmid. The asterisk indicates pGAD10 plasmids containing inserts. The relative size of these products is listed in nucleotides.

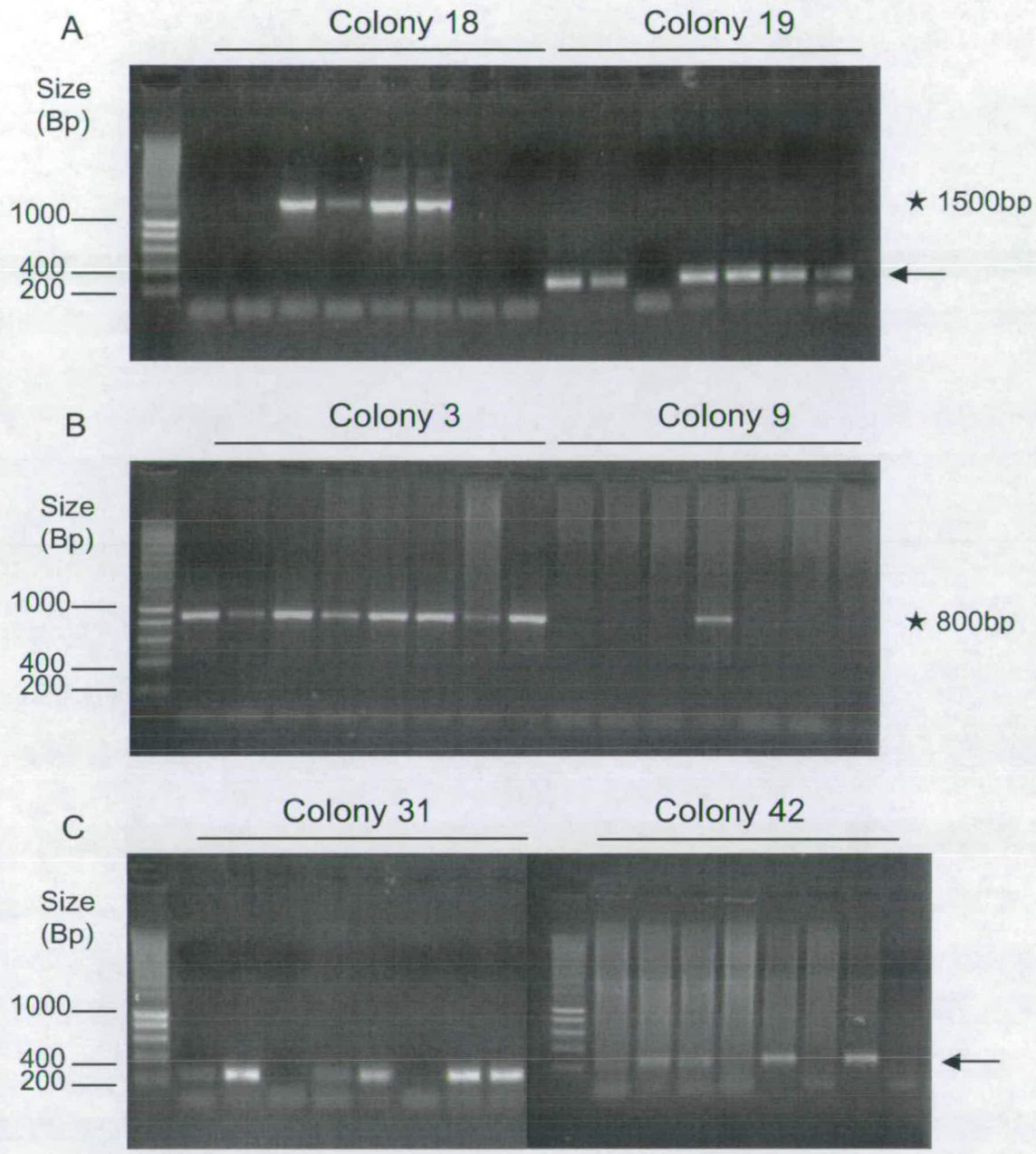

Figure 5.7B – PCR amplification of DNA inserts from the pGAD10 activation domain from *E. coli* colonies (continued).

Initially, *E. coli* was transformed with DNA from yeast colonies positive for reporter gene activation. Eight *E. coli* transformations were analysed by PCR, using pGAD10 specific primers, for pGAD10 plasmids containing trypanosome cDNA. An insert size of ~200bp indicates a pGAD10 plasmid containing no significant insert (arrow). No PCR product implies that the colony is the DNA BD vector, as the primers used were unable to bind this plasmid. The asterisk indicates pGAD10 plasmids containing inserts. The relative size of these products is listed in nucleotides.

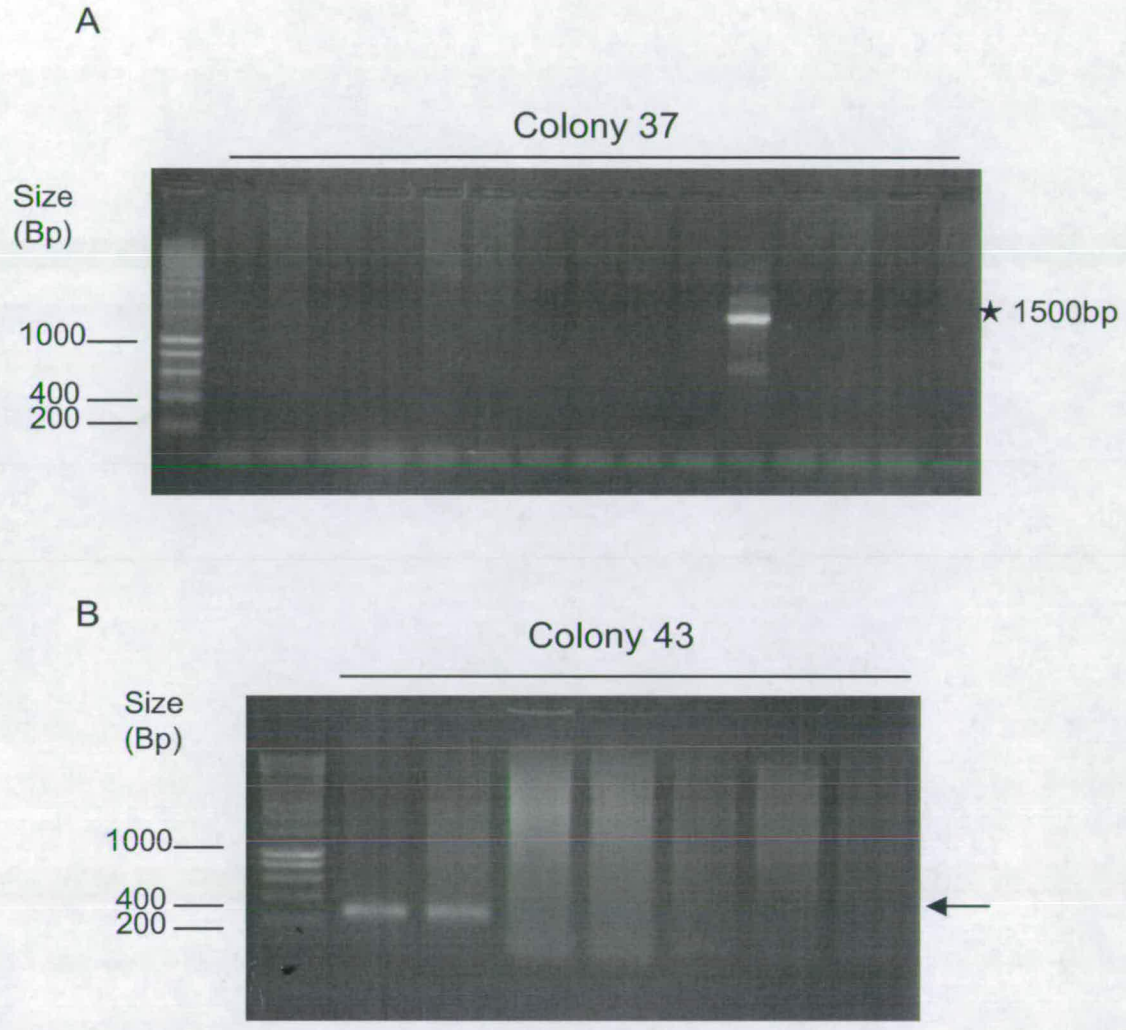

Figure 5.7C – PCR amplification of DNA inserts from the pGAD10 activation domain from *E. coli* colonies (continued).

Initially, *E. coli* was transformed with DNA from yeast colonies positive for reporter gene activation. Eight *E. coli* transformations were analysed by PCR, using pGAD10 specific primers, for pGAD10 plasmids containing trypanosome cDNA. An insert size of ~200bp indicates a pGAD10 plasmid containing no significant insert (arrow). No PCR product implies that the colony is the DNA BD vector, as the primers used were unable to bind this plasmid. The asterisk indicates pGAD10 plasmids containing inserts. The relative size of these products is listed in nucleotides.

Figure 5.8 – Verification of the number of different pGAD10 plasmids present within yeast colony 31 by PCR amplification.

To verify if multiple copies of the pGAD10 plasmids exists within yeast colony 31, 30 bacterial colonies were analysed for the presence of an insert within the pGAD10 plasmid by PCR amplification. The pGAD10 specific primers used gave a fragment size of ~200bp from an empty pGAD10 vector, as indicated by the arrow. No PCR product implies that it is the DNA BD vector, as the primers are unable to function. The data indicates that only one species of pGAD10 AD plasmid was present.

As *S. cerevisiae* possesses the ability to maintain several episomal plasmids, the possibility existed that multiple copies of the pGAD10 plasmid, that contain different inserts (or no insert) could be maintained within one yeast colony. Hence, a PCR screen of 30 bacteria colonies from a single yeast (colony 31) was undertaken to identify if other pGAD10 plasmids species were present beyond the 'no insert' construct identified in the primary screen. The data in Figure 5.8 shows that there was only one pGAD10 plasmid present within the cell.

Colony	PCR insert size (bp)	Insert
1	2000	Genomic DNA
2	200	None
3	800	*Tb*10.70.1640
5	200	None
9	800	*Tb*10.70.1640
11	200	None
18	1500	MITat1.2
19	200	None
31	200	None
35	800	*Tb*10.70.1640
37	1500	*Tb*07.2813.190
42	200	None
43	200	None

Table 5.1 – Analysis and identification of the inserts within colonies indicating a putative interaction with *Tb*ZFP2.

All colonies identified as activating the LacZ and HIS3 genes were PCR amplified, sequenced and the inserts aligned to the *T. brucei* genome using GeneDB. The table shows the colony number, insert size and GeneDB ID number if an insert was present. Plasmids that contained no insert are indicated by 'none'.

Any bacterial colony that gave a PCR product of 200 nucleotides or more was inoculated overnight in a 3ml LB culture and the DNA extracted by mini-preparation extraction. The DNA was then sent for sequencing using a pGAD10 sequencing primer, which bound specifically to the 3' end of the GAL4 AD of the pGAD10

plasmid. The sequences were then analysed against the trypanosome genome database (GeneDB) and the inserted genes or gene fragments identified, as shown in Table 5.1. As predicted by the PCR screen, 7 out of the 13 colonies sequenced contained pGAD10 plasmids that had no trypanosome DNA insert present. These were colonies 2, 5, 11, 19, 31, 42 and 43 and are the same colonies as identified by the insert size analysis.

Colonies 3, 9 and 35 all contained a DNA fragment that coded for an open reading frame (ORF) located on the complementary DNA strand to that of the *Tb*10.70.1640 gene. The short ORF contained no homology to any known protein, no obvious protein interaction motifs and was lacking a start (ATG) codon. Hence, it would appear that this interaction with *Tb*ZFP2 is unlikely to be biologically relevant.

MITat1.2 represents a VSG gene expressed in the trypanosome strain used to make the yeast two-hybrid library and is thus likely to be expressed in very high levels. Indeed, the levels of VSG within the library had previously been analysed and 5-10% of the library was composed of VSG cDNA (Hoek *et al.*, 2002). Hence, it is likely that Colony 18 was a false positive result, since many other clones should have been isolated, were this interaction real. Colony 1 contained a length of intergenic genomic DNA situated on chromosome X between genes *Tb*10.70.5970 and *Tb*10.70.5980. Its size and location made any functional fusion protein unviable.

The most promising candidate was Colony 37, which contained *Tb*07.2813.190, a 33.7 kDa hypothetical protein situated on chromosome 7. It shares 26% homology

with the mammalian DEK protein, which contains two DNA binding motifs and is believed to function as an architectural chromatin protein by binding to the DNA. The AD fusion protein formed was on the same DNA strand and within the correct reading frame to form a functional fusion protein. Therefore, this interaction had to be verified to ensure that that it was a genuine protein-protein interaction.

5.5.3 Verification of the putative candidates identified from the Matchmaker library screen using *Tb*ZFP2.

In order to confirm a genuine protein interaction, the AD plasmid containing the putative insert had to be co-transformed into the L40 yeast strain along with the pSTT91.ZFP2 DNA BD and the same phenotype as the original screen reproducibly observed (growth on SD – His – Ade plates and activation of the LacZ gene). Hence 5μg of each candidate plasmid was co-transformed using the small-scale polyethylene glycerol/lithium acetate protocol along with 5μg of pSTT91.ZFP2 DNA BD into the L40 yeast strain. The colonies were then selected on SD –Trp – Leu plates and SD – His – Ade plates and incubated for 3 days at 30°C. Each colony was then replica plated onto nylon membrane and assayed for the production of β-galactosidase, the results of which are shown in Figure 5.9.

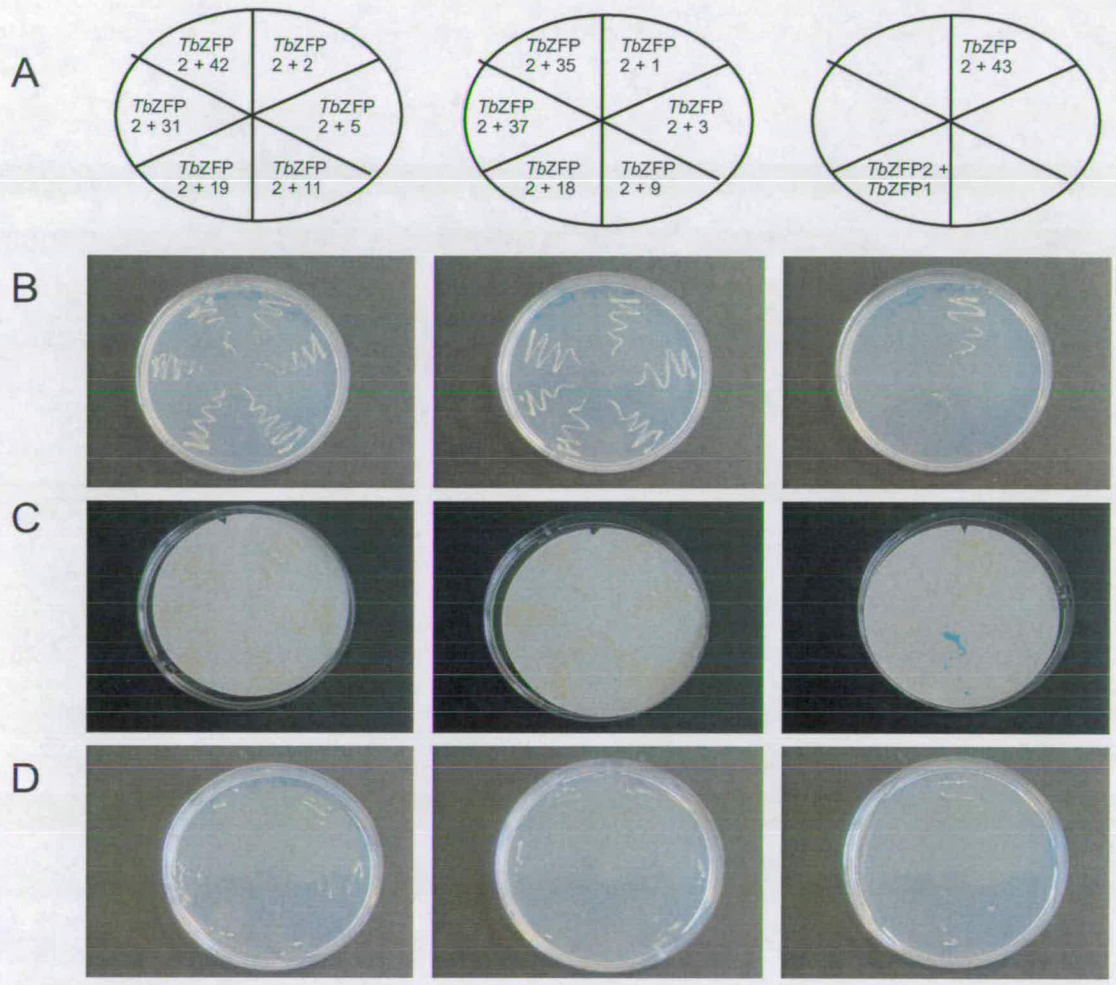

Figure 5.9 – Verification of phenotypes observed from the Matchmaker library screen using pSTT91.*Tb*ZFP2.

To confirm a genuine protein interaction, each pGAD10 AD colony obtained from the Matchmaker library screen was co-transformed along with pSTT91.*Tb*ZFP2 into L40. The expression of the LacZ and HIS3 reporter genes were then assayed after 4 days incubation. The activation of these genes suggests a genuine protein interaction. A: Schematic representation of the colony streaks; B: Growth on plates lacking SD – Trp – Leu; C: X-gal analysis – a blue colouration indicates an active LacZ gene; D: Growth on plates lacking SD - His –Trp - Leu.

As is obvious from the figure, none of the protein interactions identified from the screen were reproducible. No colonies were able to grow on SD – His – Ade, nor was any β-galactosidase produced. The implication therefore, was that all of these candidates were false positives and that it was mutations within the yeast genome that permitted growth of the colonies on SD – His – Ade. This was supported by the number of colonies that were originally isolated, of which only 6 contained inserts.

5.5.4 Testing the Matchmaker yeast two-hybrid library for the frequency of plasmids that contained inserts.

The presence of the large number of empty plasmids identified from the screen was intriguing. It should not be possible for the empty activation domain to activate the reporter genes in such a way. More important was the fact that 61% (8/13) of the plasmids analysed from the library screen did not contain any cDNA inserts. This suggested that the library might have a very low level of inserts, limiting its usefulness. Thus, if 60% of the plasmids from the library contained no inserts, it would mean that at least $2x10^6$ colonies would have to be screened to get a minimum of one fold coverage.

To investigate the frequency of inserts in the library, 0.1μg of the library was transformed into *E. coli* and selected on LB ampicillin media. Next, 96 of the resulting colonies were analysed by PCR and this used to evaluate the number of plasmid-containing inserts. In addition, the plasmid-containing inserts were sequenced to determine how many genes were in the correct frame. Figure 5.10

shows that 64 out of 96 colonies investigated contained PCR products of 200bp, which is the expected size of the empty plasmid (as indicated by the arrow). This indicated that 67% of the library contained pGAD10 plasmids with no insert.

The cDNA library had been manufactured by using non-directional adapters. Therefore, had the library contained plasmids with 80% inserts (as indicated by Clontech, the original manufacturers of the library) only 13% of the AD protein produced would have been fused to the gene in the correct reading frame. However, in reality as only 33% of pGAD10 plasmids contain inserts, this meant only 5% of plasmids were likely to contain inserts in the correct frame to produce a functional fusion protein. Thus, screening $1x10^6$ colonies would result in screening of only 50 000 plasmids containing a functional fusion protein with the correct reading frame. In light of this, it was decided that a continuation of the screening procedure using the matchmaker library did not represent a sensible use of resources and that we did not possess the quantities of library required to obtain a good coverage of the genome. Therefore, it was decided to use another trypanosome library from another source.

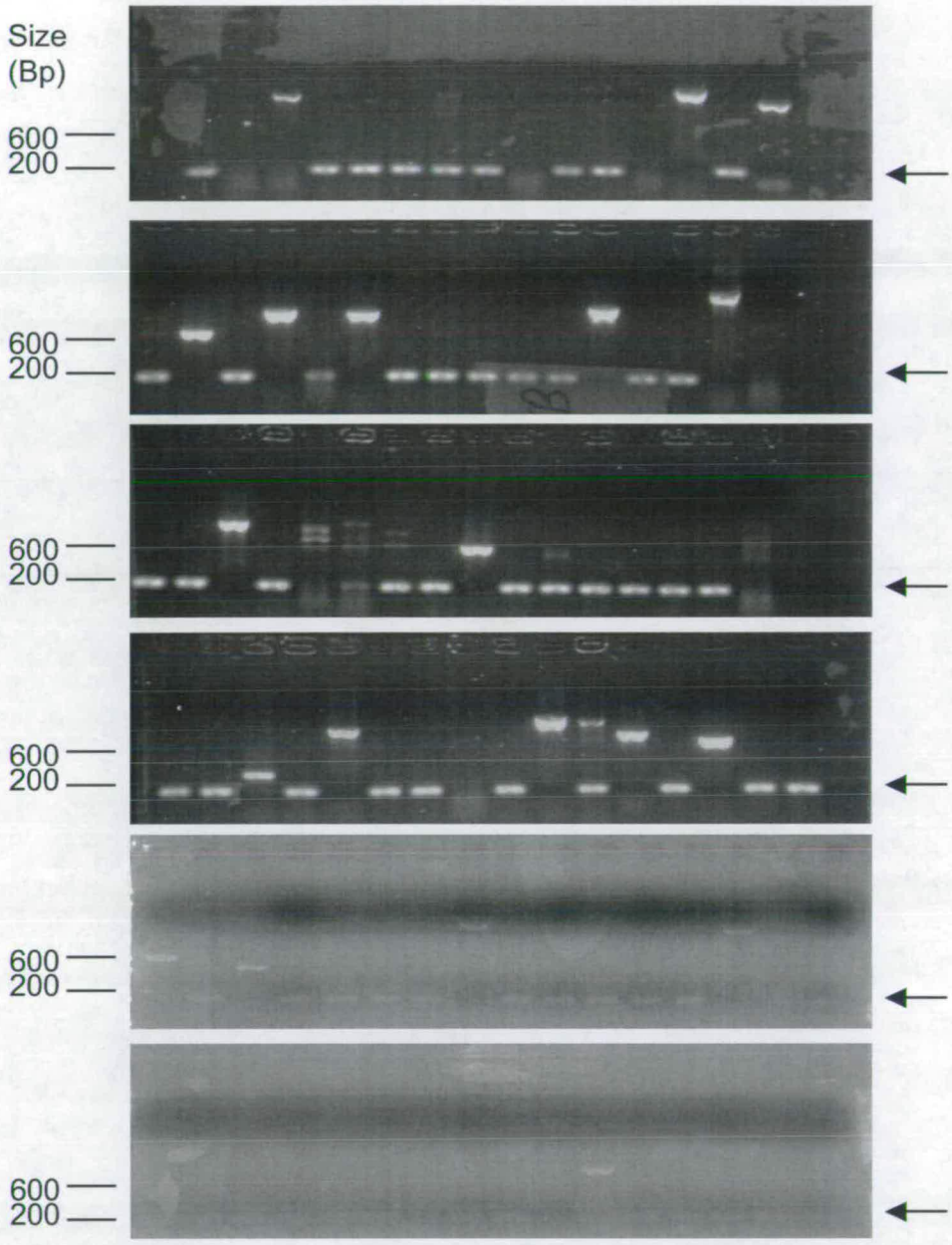

Figure 5.10 – Analysis of 96 *E. coli* colonies transformed with the Matchmaker library plasmid for the presence of a trypanosome cDNA insert.

To control for the quality of the Matchmaker library, 0.1μg library DNA was transformed into *E. coli*. 96 colonies were then PCR amplified using primers specific to the pGAD10 plasmid to analyse their insert size. An insert size of 200 nucleotides indicated a plasmid containing no insert (as indicated by the arrow). Any plasmid over 200 nucleotides was deemed to contain an insert.

5.6 Screening of a T. brucei Yeast Two-Hybrid HybriZAP library with TbZFP3.

5.6.1 Verification of the HybriZAP library using *Tb*ZFP3.

As the probability of isolating a genuine interaction from within the Matchmaker library was low, other options were investigated. Hence, another trypanosome cDNA yeast library was obtained (a kind gift from Professor Isobel Roditi, University of Bern, Switzerland). This library was created using AnTat1.1 procyclic cDNA in the HybriZAP-2.1 GAL4 yeast two-hybrid system (Stratagene). Whilst the HybriZAP-2.1 was based on the GAL4 system and should be compatible with the Matchmaker system, there was no guarantee of compatibility. As the major focus of the work within our group was currently on *Tb*ZFP3, it was decided to screen this library using *Tb*ZFP3 as the 'bait' plasmid. This would allow any potential interactions to be further investigated *in vivo* using the anti-*Tb*ZFP3 antibody, which worked very effectively for the co-immunoprecipitation. Hence, *Tb*ZFP3 was cloned into the pBD-GAL4 Cam plasmid to ensure consistency with the HybriZAP system. The *Tb*ZFP3 gene was amplified using primers that added a 5' *EcoR1* and 3' *Xho1* restriction site onto the end of the insert. The insert was then ligated into the pBD-GAL4 Cam plasmid using the *EcoR1* and *Xho1* sites within the MCS.

The pBD-GAL4 Cam.ZFP3 was transformed into the YGR-2 yeast strain using a small-scale polyethylene glycerol/lithium acetate transformation, and the resulting

colonies selected on SD – Trp media. In order to confirm the correct integration of *Tb*ZFP3 into the pBD-GAL4 Cam plasmid, the YGR-2.pBD-GAL4 Cam.ZFP3 colonies were amplified by PCR using specific *Tb*ZFP3 primers. The results are shown in Figure 5.11A, where a 400 nucleotide band is present (lanes 1-7), indicating that *Tb*ZFP3 had successfully integrated into the plasmid. To confirm the production of the GAL4 *Tb*ZFP3 fusion protein, a 5ml overnight culture inoculated with YGR-2.pBD-GAL4 Cam.ZFP3 was harvested and protein samples made. The samples were then analysed on a 15% SDS-PAGE gel, blotted onto nylon membrane and then probed with anti-*Tb*ZFP3 antibody. Figure 5.11B clearly shows the recombinant pBD-GAL4 Cam.ZFP3 protein at ~ 30kDa (as indicated by the arrow) when compared to the endogenous *Tb*ZFP3 (Lane S16, asterisk). Hence, a functional recombinant protein had been successfully created.

5.6.2 Screening the HybriZAP2.1 library for potential interacting partners for *Tb*ZFP3.

In order to undertake another library screen, a test library transformation was carried out to optimise the transformation efficiency. The YGR-2.pBD-GAL4 Cam.ZFP3 yeast strain was inoculated in a 50ml overnight culture, then diluted to an OD_{600}= 0.2 with 600mls to SD – Trp media. The cells were harvested once an OD_{600}= 0.6 was obtained and used within a polyethylene glycerol/lithium acetate transformation procedure. Again, 0.1 µg DNA was found to be the optimal concentration, which gave a transformation efficiency of 65 000 cfu/µg DNA.

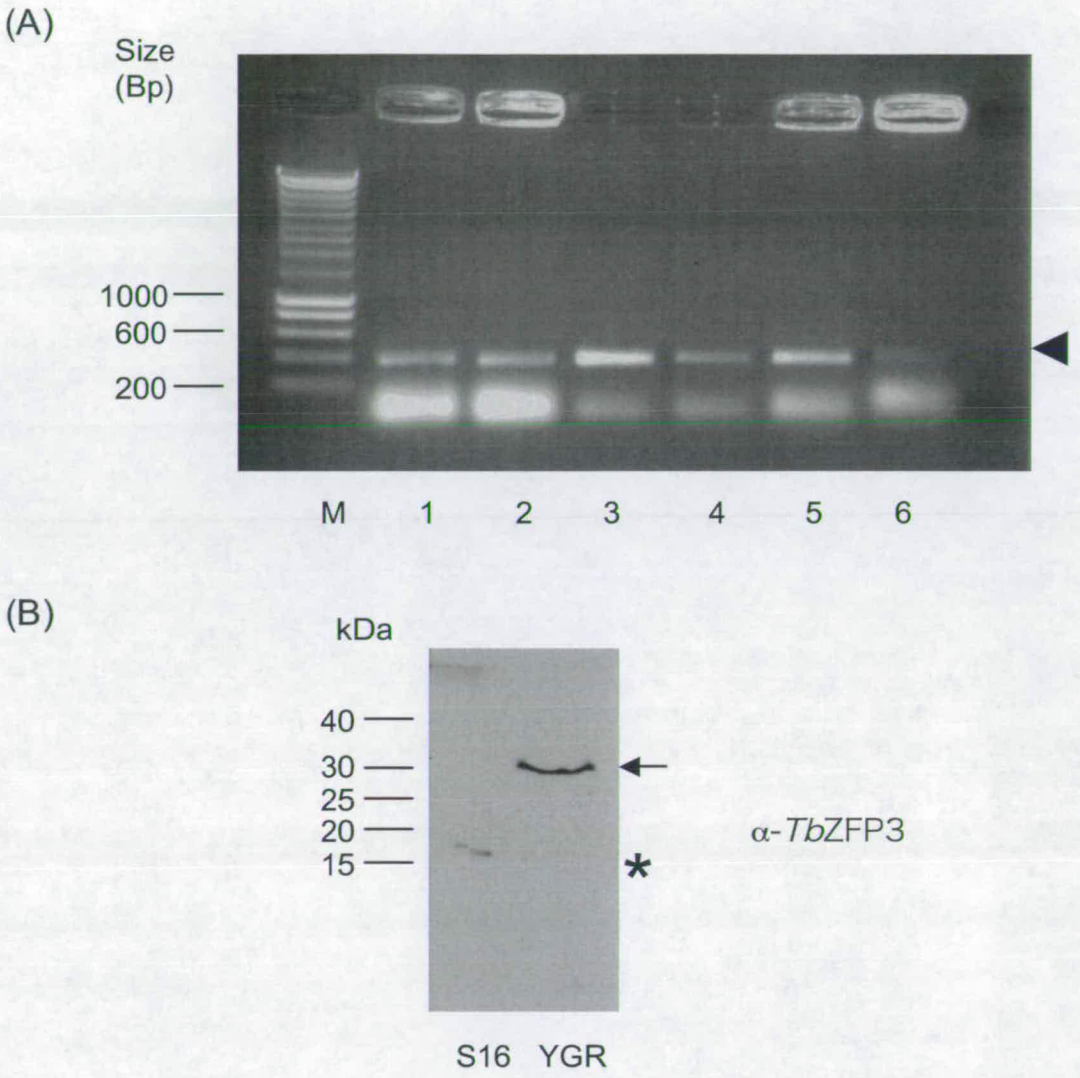

Figure 5.11 – Expression of *Tb*ZFP3 fused to pAD-GAL4-2.1 Gal4 DNA BD within the YRG-2 yeast strain.

(A). Confirmation of the correct insertion of *Tb*ZFP3 into pAD-GAL4-2.1 was obtained by PCR analysis of 6 yeast colonies using primers specific to *Tb*ZFP3. The 400 nucleotide insert of *Tb*ZFP3 is shown by the arrowhead. (B) Total YGR-2.pAD-GAL4-2.1.*Tb*ZFP3 protein extract (YGR) and total trypanosome extract (S16) were analysed on a 15% SDS – PAGE gel, which was then blotted onto nitrocellulose membrane and probed with anti–*Tb*ZFP3 antibody. The asterisk indicates the 15kDa *Tb*ZFP3 protein in trypanosome extract (S16 lane), while the arrow indicates the recombinant Gal4 *Tb*ZFP3 protein at 30kDa in the yeast extract (YGR Lane).

The next stage was to undertake a library screen of the HybriZAP-2.1 library to identify novel interacting partners for *Tb*ZFP3. It was decided that 30μg of library DNA would be used for the screen as this would give at least $1x10^6$ coverage if the transformation efficiency was lower (i.e. 35 000 cfu/μg DNA). Hence, an overnight culture inoculated with YGR-2.pBD-GAL4 Cam.ZFP3 was diluted into 600ml SD - Trp media and incubated until an OD_{600}= 0.6 was achieved. The yeast cells were then harvested and added to 30μg HybriZAP library DNA, which was transformed into the YGR-2.pBD-GAL4 Cam.ZFP3 cells using the standard polyethylene glycerol/lithium acetate protocol. The resulting solution was then plated over 50 25cm SD - His plates (the YRG-2 yeast strain used only the HIS3 gene for nutritional selection) to select for colonies containing activated reporter genes and incubated for 4 days at 30°C.

After 4 days, only 2 colonies had grown on the SD –His media. However the transformation efficiency was low, being only 17 500 cfu/μg DNA, implying that only $0.5x10^6$ colonies were screened. These colonies, named Y and Z, were assayed for the production of β-galactosidase and to retest the ability to grow on SD – His media. Neither colony produced β-galactosidase as assayed by the X-Gal assay (Figure 5.12) though both retained the ability to grow on SD – His media. Thus, it is likely that these colonies had undergone a spontaneous mutation, allowing them to grow on SD – His media, as only the HIS3 reporter gene was activated.

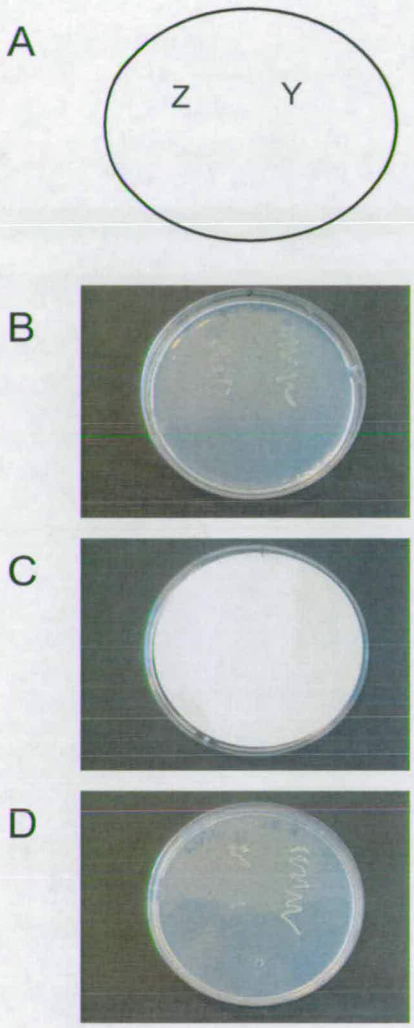

Figure 5.12 – Analysis of colonies identified from the second yeast two-hybrid HybriZAP library screen with pBD-GAL4 Cam.ZFP3 in YGR-2.

Colonies identified from the library screen, after 4 days incubation, as being putative interacting proteins were analysed for the presence of active β-galactosidase and histidine reporter genes. The activation of these genes suggests a genuine protein interaction. A: Schematic representation of the colony streaks; B: Growth on SD plates – Trp – Leu; C: X-gal analysis – a blue colouration indicates an active LacZ gene; D: Growth on SD plates - His –Trp, - Leu.

5.6.3 Second screen of the HybriZAP library using *Tb*ZFP3.

Due to the low transformation efficiency in the first HybriZAP library screen, we screened more colonies, to enhance the chances of identifying other candidates. Hence, another library screen was undertaken, however this time only 20µg DNA was used. The YGR-2.pBD-GAL4 Cam.ZFP3 yeast strain was again inoculated overnight and grown until the OD_{600}= 0.6 was reached, whereupon it was harvested and transformed with 20µg DNA according to the standard polyethylene glycerol/lithium acetate protocol. The solution was then plated over 50 25cm SD - His plates and incubated for 4 days at 30^0C.

Although the transformation efficiency was again lower than expected, 13 200 cfu/µg DNA, 22 colonies were obtained after 4 days growth. The activity of the HIS3 and LacZ genes were assayed using the X-Gal assay and are shown in Figure 5.13. While the majority of the 22 colonies obtained showed growth on SD - His media, no colonies produced β-galactosidase, the only exception being *Tb*ZFP1, which was included as a positive control. Hence, it appears that no colonies contained a genuine protein-protein interaction and all the colonies obtained were false positives.

Thus, in summary, three yeast two- hybrid library screens were undertaken, using two different *T. brucei* cDNA libraries. The matchmaker library was screened once with *Tb*ZFP2, while the HybriZAP library was screened twice using *Tb*ZFP3. In all

three cases, no genuine protein interactions were observed between either *Tb*ZFP2 or 3, despite the appropriate controls working.

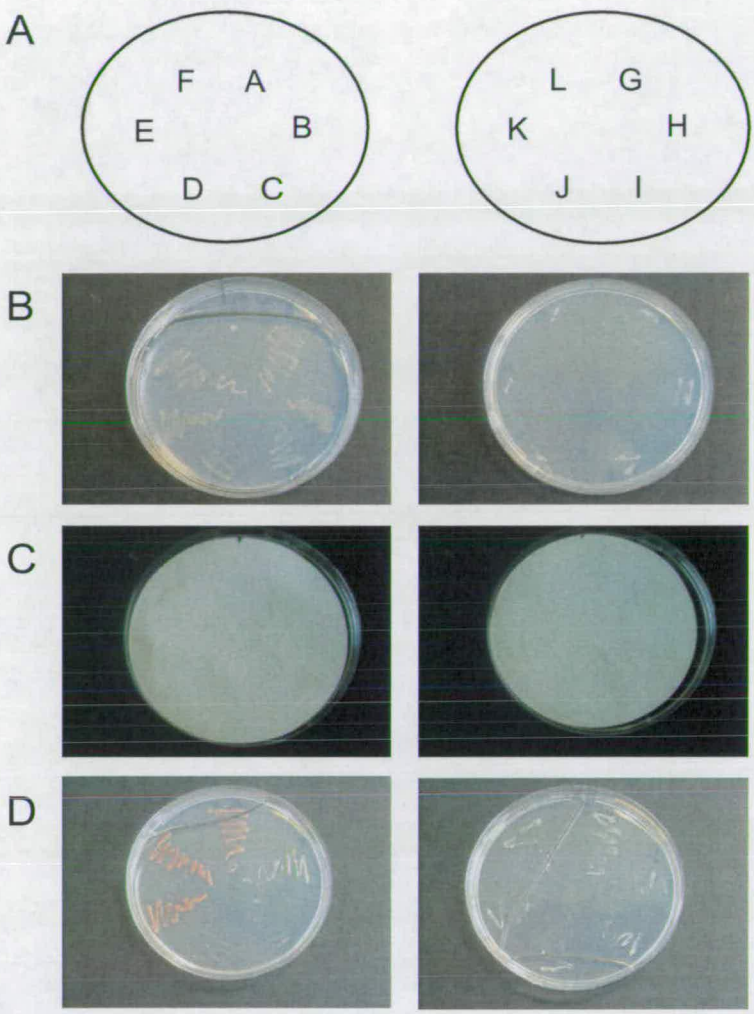

Figure 5.13A – Analysis of colonies identified from the yeast two-hybrid HybriZAP library screen with pBD-GAL4 Cam.ZFP3 in YGR-2.

Colonies identified from the library screen, after 4 days incubation, as being putative interacting proteins were analysed for the presence of active β-galactosidase and histidine reporter genes. The activation of these genes suggests a genuine protein interaction. A: Schematic representation of the colony streaks; B: Growth on SD plates – Trp – Leu; C: X-Gal analysis – a blue colouration indicates an active LacZ gene; D: Growth on SD plates - His –Trp, - Leu.

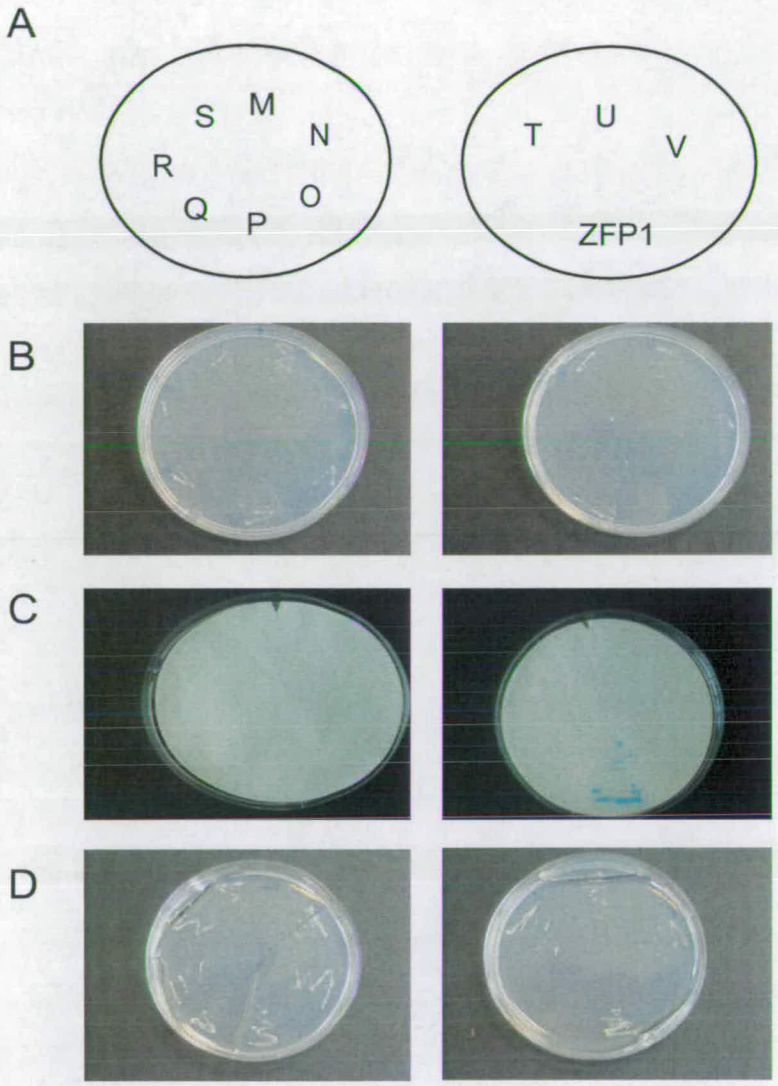

Figure 5.13B – Analysis of colonies identified from the yeast two-hybrid HybriZAP library screen with pBD-GAL4 Cam.ZFP3 in YGR-2 (continued).

Colonies identified from the library screen, after 4 days incubation, as being putative interacting proteins were analysed for the presence of active β-galactosidase and histidine reporter genes. The activation of these genes suggests a genuine protein interaction. A: Schematic representation of the colony streaks; B: Growth on SD plates – Trp – Leu; C: X-Gal analysis – a blue colouration indicates an active LacZ gene; D: Growth on SD plates - His –Trp, - Leu. pGBKT7.ZFP1 was transformed together with pBD-GAL4 Cam.ZFP3 as a positive control.

5.7 Discussion.

Direct protein-proteins interactions provide organisms with a mechanism for creating a wide range of protein complexes that allow increasingly elaborate systems from a relatively small group of proteins. Thus, the identification of proteins which directly interact with other molecules can provide information on the complexes which they act within and, by association, may suggest a functional role for the identified proteins. While the morphological effects of the *Tb*ZFP family have previously been published (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou *et al.*, 2006), the protein interactions of these molecules have yet to be studied. Hence, a yeast two-hybrid screen was undertaken to identify any protein interaction between members of the *Tb*ZFP3 family and to identify their interactions with novel partner proteins. The data within this chapter demonstrates that *Tb*ZFP1 can interact with *Tb*ZFP2 and 3 directly within the yeast two-hybrid system and that this interaction is dependant on the WW domain contained within *Tb*ZFP2 and 3. Based on this data, *Tb*ZFP2 and 3 can be provisionally classed as Class I WW domains due to their ability to interact with the PY motif of *Tb*ZFP1 (Sudol *et al.*, 1995). However, no interactions between the *Tb*ZFP family and other trypanosome proteins were identified.

The results obtained here are broadly supported by the work of Caro *et al.* (2005), in so far as *Tc*ZFP1 (*Tb*ZFP1) was shown to interact with *Tc*ZFP2A (*Tb*ZFP2) and *Tc*ZFP2B (*Tb*ZFP3) and that this interaction was mediated through the WW domain (Caro *et al.*, 2005). However, no data was obtained to support the hypothesis that

*Tb*ZFP3 can homodimerise, be it through the RGG or WW motif, as proposed by Caro *et al.* (2005). Little evidence has been found so far *in vitro* or *in vivo* to support the hypothesis that *Tb*ZFP3 is able to homodimerise (P. Walrad and A. Paterou, personal communication).

The transient expression profile of *Tb*ZFP1 during differentiation suggests that its interaction with *Tb*ZFP2 and 3 is likely to be important for effective differentiation. *Tb*ZFP2 and 3 are only associated with the polyribosomes in procyclic cells and as *Tb*ZFP1, 2 and 3 are only able to co-precipitate within procyclic trypanosomes, it is possible that *Tb*ZFP1 is essential for the association of *Tb*ZFP2 and 3 with the polyribosomes. This is supported by mutation analysis of *Tb*ZFP2 and 3 whereby both the $C_{x8}C_{x5}C_{x3}H$ and WW domains are vital for nozzle formation and polysome association. An interesting experiment would be to observe whether *Tb*ZFP2 and 3 were able to associate with the polyribosomes within the *Tb*ZFP1 null mutant created (Hendriks and Matthews, 2005). If this association were lost, it would strongly support the idea that *Tb*ZFP1 interacts directly with *Tb*ZFP2 and *Tb*ZFP3 *in vivo* and is essential for stage specific polysome association. However, this experiement would have to be done early in differentiation as null mutant procyclic forms are not viable (Hendriks and Matthews, 2005).

Sucrose gradient analysis has shown that *Tb*ZFP2 and 3, although expressed, are not present within the polysomes of slender and stumpy bloodstream forms, implying that these proteins may associate with different complexes during different life stages. As RNAi can effectively ablate *Tb*ZFP2 during the bloodstream and procyclic

life stages without any obvious phenotype, it implies that *Tb*ZFP2 is not essential to survival. However, no effective *Tb*ZFP3 RNAi lines have been generated to date, suggesting that *Tb*ZFP3, in contrast to *Tb*ZFP2, may be essential in bloodstream stages. Thus, the implication is that *Tb*ZFP3 has a vital role in trypanosome survival within the bloodstream stages, whereas *Tb*ZFP2 is not critical for survival. Although no new protein interactions have been identified with *Tb*ZFP3 as a result of the yeast two-hybrid library screens, it is likely that *Tb*ZFP3 has a major role within the bloodstream forms of the trypanosome that is still to be elucidated.

As to the question of why no other protein interactions were identified from either library screen, one possibility is in the strategy used for the manufacture of the libraries. Both libraries were created using trypanosome RNA which had undergone polyA selection and, thereafter, the cDNA was primed using oligo dT, which binds to the 3' poly A tail. So, in consequence, it is likely that these libraries are biased towards the 3' ends of trypanosome transcripts, which may have resulted in under representation of long mRNA transcripts. Along a similar line, it was expected that *Tb*ZFP1 should have been isolated from the library screen. However, *Tb*ZFP1 has an mRNA of at least 6kb in length and within this is a long 3' UTR (>5kb), likely to contain complex stage regulatory signals. As a result, it is unlikely that *Tb*ZFP1 would be present in either library. Therefore the possibility exists that if *Tb*ZFP2 and 3 do interact with stage specific proteins, such as *Tb*ZFP1, these are likely to be under strict regulation and therefore may contain long 3' UTRs (like *Tb*ZFP1) and would be less likely to be represented within the library.

While the product information obtained with the Matchmaker yeast library stated that the library had not undergone amplification (information obtained from Professor George Cross, Rockefeller University, USA), this seems surprising given the presence of the large number of plasmids containing no inserts. If the library had been amplified, the empty AD plasmids may have had a selection advantage compared to the plasmids containing inserts, thereby decreasing the representation of inserted clones. This may also explain the presence of three identical copies of the *Tb*10.70.1640 containing clone identified from the screen. Another possibility is a failure in the correct manufacture of the library. Immediately before the library was made, Clontech had halted commercial production of yeast two-hybrid libraries due to unspecified problems. Hence, it may have been that insert ligation into the pGAD10 vector was inefficient, although this contradicts the manufacturers claims. Why several empty plasmid constructs activated the reporter genes is unknown, especially considering the large number of pGAD10 plasmids without any inserts in the library as a whole. Possibly this was due to spontaneous compensatory mutations in the host yeast cell, in the similar way to all the other false positive colonies. Although the library has been previously used to identify an interaction between ESAG8 and *Tb*PUF1, it appears that either *Tb*PUF1 was present with high copy numbers or the Cross group were extremely fortunate in the interaction that they obtained.

The difference between transformation efficiencies obtained from the small and library scale transformation was likely to be due to the volumes used. While the small-scale transformation protocol was scaled up to larger volumes for the library

screen, it is possible that during the library screens the larger volumes were unable to mix effectively or that incubating the cells at 42^0C had less effect. All these factors may have resulted in the lower transformation efficiencies observed when the screen was performed.

While the yeast two-hybrid approach has been successfully used to identify many hundreds of direct protein interactions, there are many limitations with the system. The system is only able to identify direct interactions between two proteins. If a protein needs accessory factors, or must interact with a partner protein before it can interact with another protein, it is unlikely to be identified. This possibility may have occurred with the *Tb*ZFP molecules. If *Tb*ZFP1 must interact with *Tb*ZFP2 or 3 before it can interact with a third protein, this would not be detected. Also, proteins that contain an RNA or DNA binding domain may act as transcriptional activators and unintentionally activate the reporter genes when transformed into yeast. The pSTT91.*Tb*ZFP1 plasmid appeared to show this phenotype as the LacZ gene was activated in the presence of an empty AD, though it is surprising this occurred in the DNA BD rather than the AD. Other limitations may include the plasmid in which the library was created, for example it has been shown that some interactions only occur with the bait and prey protein in the AD and DNA BD respectively. Thus, the interaction between *Tc*ZFP1 and 2 was shown only to function if *Tc*ZFP1 was present within the DNA BD (Caro *et al.*, 2005). Also, the yeast two-hybrid system relies on proteins interacting within the nucleus. Some types of proteins, such as those containing hydrophobic domains or localisation signals, may be excluded from the nucleus and therefore unable to interact. While the *Tb*ZFP family have been

shown to be cytosolic, if an interaction occurred with proteins containing localisation signals or hydrophobic domains, it would be unlikely to be detected within this system.

Data presented within this chapter suggests two putative interactions between *Tb*ZFP1 and *Tb*ZFP2 and 3, and the interactions imply association and function within the polyribosome complex. However, only further work will conclusively demonstrate these interactions *in vivo* and the precise role of these molecules within the trypanosome.

Chapter 6:

Overview.

6.1 Summary

The currently characterised *Tb*ZFP family consists of three members, *Tb*ZFP1, 2 and 3. These proteins are defined by the presence of the $C_{x8}C_{x5}C_{x3}H$ RNA binding motif and their small size, all having a molecular weight of less than 17kDa (Hendriks and Matthews, 2005; Hendriks *et al.*, 2001; Paterou *et al.*, 2006). In addition to the $C_{x8}C_{x5}C_{x3}H$ RNA binding motif, all three molecules possess a protein interaction domain. *Tb*ZFP1 contains the PY domain, while *Tb*ZFP2 and 3 possess the WW domain, a potential ligand for the PY domain (Hendriks *et al.*, 2001; Paterou *et al.*, 2006; Sudol *et al.*, 1995). This suggests that *Tb*ZFP1 may be able to directly interact with *Tb*ZFP2 and 3. Finally, *Tb*ZFP3 contains another RNA binding domain, the RGG motif (Kiledjian and Dreyfuss, 1992; Paterou *et al.*, 2006).

*Tb*ZFP1 was initially identified during a screen for molecules up-regulated during differentiation and *Tb*ZFP2 and 3 have subsequently been shown to play a vital role in differentiation. *Tb*ZFP1 is transiently up-regulated during differentiation, being up-regulated between 2 - 8 hours after the initiation of this process, before being expressed in mature procyclic trypanosomes (Hendriks *et al.*, 2001). In contrast, *Tb*ZFP2 and 3 are constitutively expressed throughout the life cycle (Hendriks *et al.*, 2001; Paterou *et al.*, 2006).

Extensive work has been undertaken to identify the function of the *Tb*ZFP family *in vivo*. Differentiation of a null mutant of *Tb*ZFP1, where both *Tb*ZFP1 genes were disrupted, displayed incomplete kinetoplast repositioning (Hendriks and Matthews,

2005). The kinetoplast remained at the posterior tip of the cell (Hendriks and Matthews, 2005). Ablation of the *Tb*ZFP2 mRNA by RNAi compromised differentiation, with the majority of cells being unable to effectively complete differentiation (Hendriks *et al.*, 2001). Ablation of *Tb*ZFP3 using RNAi has been attempted, though no cell lines have been successfully created to date. However, ectopic expression of *Tb*ZFP3 in bloodstream forms caused enhanced differentiation to procyclic forms (Paterou *et al.*, 2006). This implies that *Tb*ZFP2 and 3 may be undertaking a similar function, although ectopic expression of *Tb*ZFP2 in bloodstream forms did not show the same enhanced differentiation phenotype as seen with TbZFP3. Additionally, when *Tb*ZFP2 and *Tb*ZFP3 were ectopically over-expressed in procyclic forms, a significant increase in the kinetoplast-posterior dimension was observed (called the "nozzle" phenotype) (Hendriks *et al.*, 2001; Paterou *et al.*, 2006). The presence of the WW and $C_{x8}C_{x5}C_{x3}H$ motifs were vital for the formation of "nozzle" phenotype, as mutation of these domains disrupted its formation (Hendriks *et al.*, 2001; Paterou *et al.*, 2006).

Recently, *Tb*ZFP2 and 3 have been shown to associate with polyribosomes in procyclic forms, though no interaction was observed with the polysomes in either bloodstream slender or stumpy trypanosomes (Paterou *et al.*, 2006). Association of *Tb*ZFP2 and 3 with the polysomes requires functional copies of the $C_{x8}C_{x5}C_{x3}H$ and WW motifs, as mutation studies have shown that the interaction with the polysomes is lost if these domains are mutated (Paterou *et al.*, 2006).

6.2 Protein interactions of the TbZFP family

As the *Tb*ZFP proteins have a small molecular weight and contain only one RNA binding motif, the potential exists that these proteins may act in a modular function. This argument is further strengthened by the presence of the PY protein interaction domain in *Tb*ZFP1 and the complementary interacting WW domain within *Tb*ZFP2 and 3. Hence, a yeast two-hybrid screen was undertaken to analyse if any interactions occurred *in vitro*. Data presented in this thesis (Chapter 4) showed a direct protein interaction between *Tb*ZFP1 and either *Tb*ZFP2 or *Tb*ZFP3. The interaction was dependant on the presence of the WW domain, as the interaction was disrupted in molecules lacking the WW domain. This data was supported by the work of Caro *et al*, (2005) who also demonstrated an interaction between *Tc*ZFP1 and *Tc*ZFP2A or *Tc*ZFP2B, the homologous molecules to the *Tb*ZFP proteins in *T. cruzi* (Caro *et al.*, 2005). However, Caro *et al* made an additional observation that *Tc*ZFP2B was able to homodimerise, a finding that was not supported by our investigations (Caro *et al.*, 2005).

Further analysis *in vivo* has supported an interaction between *Tb*ZFP1 and *Tb*ZFP2 and 3. Immunoprecipitation of *Tb*ZFP3 using an anti-*Tb*ZFP3 antibody in procyclic cells managed to co-select *Tb*ZFP1 and 2, suggesting that *Tb*ZFP1, 2 and 3 are associated within the same complex in the trypanosome (Paterou *et al.*, 2006). Additionally, *Tb*ZFP2 and 3 have both been shown to associate with the polyribosome apparatus in procyclic trypanosomes. However, as no association with the polysomes apparatus is observed in slender or stumpy trypanosomes, the

presence of the differentially expressed *Tb*ZFP1 may regulate the constitutively expressed *Tb*ZFP2 and 3 proteins and cause them to associate with the polysomes complex, once differentiation is initiated. Whether any direct protein interactions occur *in vivo,* either amongst members of the *Tb*ZFP family or directly with the ribosomal components remains to be confirmed.

While recent work has suggested a putative role for *Tb*ZFP1, 2 and 3 as stage specific regulators of translational activation (or repression) within procyclic trypanosomes, the exact function of *Tb*ZFP2 and 3 within bloodstream form trypanosomes remains to be identified. Therefore, several yeast two-hybrid trypanosome cDNA library screens were undertaken using *Tb*ZFP2 and 3 to identify any potential unknown protein interactions (Chapter 5). Despite several successful library screens, no genuine interactions were identified. This implies that either the yeast two-hybrid system was unable to detect the interactions of these proteins or that *Tb*ZFP 2 or 3 only interact with *Tb*ZFP1. Perhaps, the role of *Tb*ZFP2 and 3 within the bloodstream forms is to tether specific developmentally regulated mRNA subsets until differentiation is induced, upon which these proteins associate with the polysomes. Only further work will establish the validity of this hypothesis.

6.3 RNA interactions of the TbZFP family

The signature motif of the *Tb*ZFP family is the presence of the $C_{x8}C_{x5}C_{x3}H$ RNA binding motif. Initial identification of *Tb*ZFP2 and 3 was by a screen of the genome database for molecules that contained the $C_{x8}C_{x5}C_{x3}H$ motif. In higher eukaryotes, the $C_{x8}C_{x5}C_{x3}H$ motif has been implicated in a wide variety of biological processes, including mRNA stability, control of translation and mRNA localisation. All proteins containing the $C_{x8}C_{x5}C_{x3}H$ motif that have been identified to date, possess a minimum of two copies of the $C_{x8}C_{x5}C_{x3}H$ motif within the protein. However, *Tb*ZFP1, 2 and 3 only possess one copy of the motif. Therefore, the presence of an interaction *in vivo* between *Tb*ZFP1 and *Tb*ZFP2 or 3 would create a complex that contained the two $C_{x8}C_{x5}C_{x3}H$ fingers.

Work on the *T. cruzi* homologue of *Tb*ZFP1 (*Tc*ZFP1) has shown that *Tc*ZFP1 is unable to bind ARE sequences *in vitro* (Morking *et al.*, 2004). However, further experiments have indicated that *Tc*ZFP1 has a strong affinity for poly-C oligoribonucleotides, although a weak affinity was also observed with the U-rich 26mer sequence located within the EP procyclin 3' UTR of *T. brucei* (Hotz *et al.*, 1997; Morking *et al.*, 2004). Interestingly, a sequence similar to the U-rich 26mer was also identified in the cytochrome oxidase V mRNA of *T. brucei* and has been implicated as a potential general procyclic stage-specific signal (Mayho *et al.*, 2006).

While recent data has indicated that *Tb*ZFP2 and 3 are associated with the polyribosome apparatus and with EP procyclin, whether *Tb*ZFP2 or 3 interact direct

with the EP procyclin mRNA remains to be established. Indeed, only limited information is available on whether the *Tb*ZFP family interact directly with RNA and what motifs may be involved with this interaction. In order to identify if *Tb*ZFP3 was able to bind RNA and identify any RNA motifs involved, a selection of ligands by exponential enrichment (SELEX) approach was used (Chapter 5). This identified the sequence UCAGU as a putative motif bound by *Tb*ZFP3. RNA binding assays undertaken using recombinant *Tb*ZFP3 demonstrated the ability of *Tb*ZFP3 to bind RNA, though it was not dependant on the UCAGU motif. The UCAGU motif identified bears no resemblance to the classical ARE bound by other CCCH zinc finger proteins, nor does it corroborate the work of Morking *et al.*, although this was undertaken on the *T. cruzi* homologue of *Tb*ZFP1. This may be because *Tb*ZFP3 requires a partner protein that provides the specificity to bind the RNA sequence. In this instance, it could be *Tb*ZFP1 and the combination of the two molecules that would increase *Tb*ZFP1's affinity for the U rich 26mer sequence, or another as yet unidentified sequence. Identification of the RNA targets of the *Tb*ZFP family should provide more insight into the function and targets of this novel complex during differentiation.

6.4 Implications for future research

The interactions of the *Tb*ZFP family are vital to the dissection of the function of these molecules. The identification of EP procyclin RNA being associated with the *Tb*ZFP complex implies a function for the complex, though whether *Tb*ZFP1, 2 or 3 interact directly or indirectly with this mRNA remains to be established. The identification of RNA targets bound by the *Tb*ZFP molecules may provide further insight into the functions of these molecules. If it was established that *Tb*ZFP1 and 2 or 3 did interact directly *in vivo*, then a SELEX approach could be used, using *Tb*ZFP1 and 2 or 3 in combination to select for the RNA motif bound by this complex. An alternate method would be the use of a yeast three-hybrid system to screen an oligonucleotide library for RNA sequences that interact with the *Tb*ZFP complex. These methods should identify the consensus sequence that the *Tb*ZFP complex would bind to, which could then be used to screen the genome database and identify putative RNA substrates. The protein:RNA interaction would be confirmed using the *in vitro* EMSA.

EP procyclin has been identified as a putative target for further work (Paterou *et al.*, 2006). By undertaking RNA:protein binding studies using *Tb*ZFP1 and 2 or 3, it should be possible to identify if the *Tb*ZFP proteins interact directly with the EP procyclin mRNA. If a direct interaction is found, mutational analysis of the 3' UTR of EP procyclin could be undertaken to identify the consensus sequence that binds

the *Tb*ZFP proteins. It could be speculated that this would be the U-rich 26mer which has been shown to weakly bind *Tb*ZFP1 (Hotz *et al.*, 1997; Morking *et al.*, 2004).

Analysis of the polysomes within the *Tb*ZFP1 null mutant using glycerol gradients may provide further information if *Tb*ZFP1 can interact with the polysome machinery. It would be of interest to see if *Tb*ZFP2 and 3 still associated with the polysomes in a *Tb*ZFP1 null mutant that was differentiated to procyclic cells, though this could only be done transiently, as the *Tb*ZFP1 null mutant is not viable in procyclic trypanosomes. If the interaction were lost, it would suggest that *Tb*ZFP1 was required for the association of *Tb*ZFP2 and 3 with polysomes and may provide further evidence of a direct interaction between *Tb*ZFP1 and 2 or 3. Additionally, it would be interesting to undertake coimmunoprecipitation using the anti-*Tb*ZFP3 antibody to test the integrity of the complex in the absence of *Tb*ZFP1. Again, disruption of the complex would support the theory that *Tb*ZFP1 is interacting directly with *Tb*ZFP2 and 3.

To ascertain if the *Tb*ZFP family are directly or indirectly interacting with the polyribosomes, a Tandem Affinity Purification (TAP) method could be used to isolate the complexes. Purification and analysis by mass spectrometry of any complex identified using this method may provide additional data on the composition of the complex. This data would also support recent coimmunoprecipitation data obtained from analysis of complexes associated with *Tb*ZFP3 (P. Walrad, personal communication). Information gained from these two

methods could then be used to dissect the molecular interactions of the *Tb*ZFP family *in vivo*.

While significant inroads have been made into the function of the *Tb*ZFP family, many questions remain. Specifically, what are the RNA targets and how does the *Tb*ZFP family regulate these targets? Answers to these and many other questions would provide a detailed insight into the function of the *Tb*ZFP family and its role during differentiation to procyclic forms in the trypanosome.

Bibliography.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I. and Englund, P.T. (2001). The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proc Natl Acad Sci U S A*, **98**, 1513-1518.

Anderson, P. & Kedersha, N. (2006).RNA granules. *J. Cell Biol.* **172**, 803–808

Badis, G., Saveanu, C., Fromont-Racine, M. & Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. *Mol. Cell* **15**, 5–15

Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L. and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev*, **11**, 1703-1716.

Barry, J.D. (1997). The relative significance of mechanisms of antigenic variation in African trypanosomes. *Parasitol Today*, **13**, 212-218.

Barry, J.D. and Emergy, D.L. (1984). Parasite development and host responses during the establishment of *Trypanosoma brucei* infection transmitted by tsetse fly. *Parasitology*, **88 (Pt 1)**, 67-84.

Barry, J.D. and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol*, **49**, 1-70.

Bastin, P., Pullen, T.J., Sherwin, T. and Gull, K. (1999). Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant snl-1. *J Cell Sci*, **112 (Pt 21)**, 3769-3777.

Benz, C., Nilsson, D., Andersson, B., Clayton, C. and Guilbride, D.L. (2005). Messenger RNA processing sites in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **143**, 125-134.

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

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., Bohme, U., Hannick, L., Aslett, M.A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U.C., Arrowsmith, C., Atkin, R.J., Barron, A.J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T.J., Churcher, C., Clark, L.N., Corton, C.H., Cronin, A., Davies, R.M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M.C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B.R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A.X., Koo, H., Larke, N., Landfear, S., Larkin, C.,

Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P.J., Moule, S., Martin, D.M., Morgan, G.W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C.S., Peterson, J., Quail, M.A., Rabbinowitsch, E., Rajandream, M.A., Reitter, C., Salzberg, S.L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A.J., Tallon, L., Turner, C.M., Tait, A., Tivey, A.R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M.D., Embley, T.M., Gull, K., Ullu, E., Barry, J.D., Fairlamb, A.H., Opperdoes, F., Barrell, B.G., Donelson, J.E., Hall, N., Fraser, C.M., Melville, S.E. and El-Sayed, N.M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416-422.

Biebinger, S., Wirtz, L.E., Lorenz, P. and Clayton, C. (1997). Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol Biochem Parasitol*, **85**, 99-112.

Blackshear, P.J. (2002). Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem Soc Trans*, **30**, 945-952.

Blackshear, P.J., Lai, W.S., Kennington, E.A., Brewer, G., Wilson, G.M., Guan, X. and Zhou, P. (2003). Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates. *J Biol Chem*, **278**, 19947-19955.

Blum, B. and Simpson, L. (1990). Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the pre-edited region. *Cell*, **62**, 391-397.

Bork, P. and Sudol, M. (1994). The WW domain: a signalling site in dystrophin? *Trends Biochem Sci*, **19**, 531-533.

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol*, **15**, 94-98.

Brun, R. and Schonenberger. (1979). Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication. *Acta Trop*, **36**, 289-292.

Buhler, M., Steiner, S., Mohn, F., Paillusson, A. & Muhlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin- mRNA depends on 3' UTR length. Nature Struct. Mol. Biol. **13**, 462–464

Burd, C.G. and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615-621.

Carballo, E., Lai, W.S. and Blackshear, P.J. (1998). Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science*, **281**, 1001-1005.

Carballo, E., Lai, W.S. and Blackshear, P.J. (2000). Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability. *Blood*, **95**, 1891-1899.

Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M.J. and Vazquez, M.P. (2005). Protein interactions within the *Tc*ZFP zinc finger family members of *Trypanosoma cruzi*: implications for their functions. *Biochem Biophys Res Commun*, **333**, 1017-1025.

Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M.J. and Vazquez, M.P. (2006). *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. *Exp Parasitol*, **113**, 112-124.

Cattand, P., Jannin, J. and Lucas, P. (2001). Sleeping sickness surveillance: an essential step towards elimination. *Trop Med Int Health*, **6**, 348-361.

Chen, H.I. and Sudol, M. (1995). The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci U S A*, **92**, 7819-7823.

Chien, C.T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A*, **88**, 9578-9582.

Clayton, C., Hausler, T. and Blattner, J. (1995). Protein trafficking in kinetoplastid protozoa. *Microbiol Rev*, **59**, 325-344.

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

Conti, E. & Izaurralde, E (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17**, 316–325

Czichos, J., Nonnengaesser, C. and Overath, P. (1986). *Trypanosoma brucei*: *cis*-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes *in vitro*. *Exp Parasitol*, **62**, 283-291.

D'Orso, I. and Frasch, A.C. (2001). *Tc*UBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes. *J Biol Chem*, **276**, 34801-34809.

D'Orso, I. and Frasch, A.C. (2002). *Tc*UBP-1, an mRNA destabilizing factor from trypanosomes, homodimerizes and interacts with novel AU-rich element and Poly(A)-binding proteins forming a ribonucleoprotein complex. *J Biol Chem*, **277**, 50520-50528.

Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489-499.

DeRenzo, C., Reese, K.J. and Seydoux, G. (2003). Exclusion of germ plasm proteins from somatic lineages by cullin-dependent degradation. *Nature*, **424**, 685-689.

Drozdz, M. and Clayton, C. (1999). Structure of a regulatory 3' untranslated region from *Trypanosoma brucei*. *RNA*, **5**, 1632-1644.

Durieux, P.O., Schutz, P., Brun, R. and Kohler, P. (1991). Alterations in Krebs cycle enzyme activities and carbohydrate catabolism in two strains of *Trypanosoma brucei* during *in vitro* differentiation of their bloodstream to procyclic stages. *Mol Biochem Parasitol*, **45**, 19-27.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E. and Donelson, J.E. (2000). The African trypanosome genome. *Int J Parasitol*, **30**, 329-345.

Engstler, M. and Boshart, M. (2004). Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. *Genes Dev*, **18**, 2798-2811.

Estevez, A.M., Kempf, T. and Clayton, C. (2001). The exosome of *Trypanosoma brucei*. *Embo J*, **20**, 3831-3839.

Estevez, A.M., Lehner, B., Sanderson, C.M., Ruppert, T. and Clayton, C. (2003). The roles of intersubunit interactions in exosome stability. *J Biol Chem*, **278**, 34943-34951.

Ferguson, M.A., Brimacombe, J.S., Brown, J.R., Crossman, A., Dix, A., Field, R.A., Guther, M.L., Milne, K.G., Sharma, D.K. and Smith, T.K. (1999). The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness. *Biochim Biophys Acta*, **1455**, 327-340.

Field, M.C., Menon, A.K. and Cross, G.A. (1991). A glycosylphosphatidylinositol protein anchor from procyclic stage *Trypanosoma brucei*: lipid structure and biosynthesis. *Embo J*, **10**, 2731-2739.

Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, **340**, 245-246.

Frischmeyer, P. A. *et al.* (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).

Furger, A., Schurch, N., Kurath, U. and Roditi, I. (1997). Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of

Trypanosoma brucei by modulating RNA stability and translation. *Mol Cell Biol*, **17**, 4372-4380.

Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F. and Erard, M. (1992). Concerted activities of the RNA recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with pre-ribosomal RNA. *Eur J Biochem*, **209**, 541-548.

Gibson, W.C., Swinkels, B.W. and Borst, P. (1988). Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. *J Mol Biol*, **201**, 315-325.

Gietz, D., St Jean, A., Woods, R.A. and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res*, **20**, 1425.

Gilinger, G. and Bellofatto, V. (2001). Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res*, **29**, 1556-1564.

Gray, A.R. (1965). Antigenic variation in a strain of *Trypanosoma brucei* transmitted by *Glossina morsitans* and *G. palpalis*. *J Gen Microbiol*, **41**, 195-214.

Gruszynski, A.E., van Deursen, F.J., Albareda, M.C., Best, A., Chaudhary, K., Cliffe, L.J., del Rio, L., Dunn, J.D., Ellis, L., Evans, K.J., Figueiredo, J.M., Malmquist, N.A., Omosun, Y., Palenchar, J.B., Prickett, S., Punkosdy, G.A., van Dooren, G., Wang, Q., Menon, A.K., Matthews, K.R. and Bangs, J.D. (2006). Regulation of surface coat exchange by differentiating African trypanosomes. *Mol Biochem Parasitol*, **147**, 211-223.

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., Tait, A., Gull, K., Barrell, B. and Melville, S.E. (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Res*, **31**, 4864-4873.

Hamm, B., Schindler, A., Mecke, D. and Duszenko, M. (1990). Differentiation of *Trypanosoma brucei* bloodstream trypomastigotes from long slender to short stumpy-like forms in axenic culture. *Mol Biochem Parasitol*, **40**, 13-22.

Hehl, A., Vassella, E., Braun, R. and Roditi, I. (1994). A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*, **91**, 370-374.

Hendriks, E., van Deursen, F.J., Wilson, J., Sarkar, M., Timms, M. and Matthews, K.R. (2000). Life-cycle differentiation in *Trypanosoma brucei*: molecules and mutants. *Biochem Soc Trans*, **28**, 531-536.

Hendriks, E.F., Abdul-Razak, A. and Matthews, K.R. (2003). *Tb*CPSF30 depletion by RNA interference disrupts polycistronic RNA processing in *Trypanosoma brucei*. *J Biol Chem*, **278**, 26870-26878.

Hendriks, E.F. and Matthews, K.R. (2005). Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of *Tb*ZFP1, a differentiation-enriched CCCH protein. *Mol Microbiol*, **57**, 706-716.

Hendriks, E.F., Robinson, D.R., Hinkins, M. and Matthews, K.R. (2001). A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *Embo J*, **20**, 6700-6711.

Hertz-Fowler, C., Ersfeld, K. and Gull, K. (2001). CAP5.5, a life-cycle-regulated, cytoskeleton-associated protein is a member of a novel family of calpain-related proteins in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **116**, 25-34.

Hirumi, H. and Hirumi, K. (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J Parasitol*, **75**, 985-989.

Hoek, M., Zanders, T. and Cross, G.A. (2002). *Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a Pumilio family protein. *Mol Biochem Parasitol*, **120**, 269-283.

Hollenberg, S.M., Sternglanz, R., Cheng, P.F. and Weintraub, H. (1995). Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol Cell Biol*, **15**, 3813-3822.

Hori, T., Taguchi, Y., Uesugi, S. and Kurihara, Y. (2005). The RNA ligands for mouse proline-rich RNA-binding protein (mouse Prrp) contain two consensus sequences in separate loop structure. *Nucleic Acids Res*, **33**, 190-200.

Hotz, H.R., Biebinger, S., Flaspohler, J. and Clayton, C. (1998). PARP gene expression: control at many levels. *Mol Biochem Parasitol*, **91**, 131-143.

Hotz, H.R., Hartmann, C., Huober, K., Hug, M. and Clayton, C. (1997). Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res*, **25**, 3017-3026.

Houston, D.W. and King, M.L. (2000). Germ plasm and molecular determinants of germ cell fate. *Curr Top Dev Biol*, **50**, 155-181.

Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., James, M., Davis, R., Sudol, M., Rodwell, J. and Herrero, J.J. (2004). A map of WW domain family interactions. *Proteomics*, **4**, 643-655.

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004). Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol*, **11**, 257-264.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S.M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C.V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R.M., Cronin, A., Cruz, A.K., Davies, R.M., De Gaudenzi, J., Dobson, D.E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A.C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J.C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T.M., Price, C., Purnelle, B., Quail, M.A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J.C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D.C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D.F., Blackwell, J.M., Stuart, K.D., Barrell, B. and Myler, P.J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436-442.

Jacobs Anderson, J.S. and Parker, R. (2000). Computational identification of *cis*-acting elements affecting post-transcriptional control of gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, **28**, 1604-1617.

Kay, B.K., Williamson, M.P. and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *Faseb J*, **14**, 231-241.

Kedersha, N. *et al.* (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* **169**, 871–884

Keller, W., Bienroth, S., Lang, K.M. and Christofori, G. (1991). Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *Embo J*, **10**, 4241-4249.

Kiledjian, M. and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *Embo J*, **11**, 2655-2664.

Klingbeil, M.M., Drew, M.E., Liu, Y., Morris, J.C., Motyka, S.A., Saxowsky, T.T., Wang, Z. and Englund, P.T. (2001). Unlocking the secrets of trypanosome kinetoplast DNA network replication. *Protist*, **152**, 255-262.

Kohl, L. and Gull, K. (1998). Molecular architecture of the trypanosome cytoskeleton. *Mol Biochem Parasitol*, **93**, 1-9.

LaCount, D.J., Barrett, B. and Donelson, J.E. (2002). *Trypanosoma brucei* FLA1 is required for flagellum attachment and cytokinesis. *J Biol Chem*, **277**, 17580-17588.

Lai, W.S., Carballo, E., Strum, J.R., Kennington, E.A., Phillips, R.S. and Blackshear, P.J. (1999). Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA. *Mol Cell Biol*, **19**, 4311-4323.

Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A. and Blackshear, P.J. (2000). Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J Biol Chem*, **275**, 17827-17837.

Lai, W.S., Kennington, E.A. and Blackshear, P.J. (2002). Interactions of CCCH zinc finger proteins with mRNA: non-binding tristetraprolin mutants exert an inhibitory effect on degradation of AU-rich element-containing mRNAs. *J Biol Chem*, **277**, 9606-9613.

LeBowitz, J.H., Smith, H.Q., Rusche, L. and Beverley, S.M. (1993). Coupling of poly(A) site selection and trans-splicing in Leishmania. *Genes Dev*, **7**, 996-1007.

Lee, M.G. and Van der Ploeg, L.H. (1997). Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol*, **51**, 463-489.

Liang, X.H., Haritan, A., Uliel, S. and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell*, **2**, 830-840.

Lieb, B., Carl, M., Hock, R., Gebauer, D. and Scheer, U. (1998). Identification of a novel mRNA-associated protein in oocytes of *Pleurodeles waltl* and *Xenopus laevis*. *Exp Cell Res*, **245**, 272-281.

Liu, J. *et al.* (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature*, **382**, 646-649.

Mahmood, R., Mittra, B., Hines, J.C. and Ray, D.S. (2001). Characterization of the *Crithidia fasciculata* mRNA cycling sequence binding proteins. *Mol Cell Biol*, **21**, 4453-4459.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I., Stern, L.S., Wang, Z., Ullu, E. and Tschudi, C. (2000). A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA. *Rna*, **6**, 163-169.

Martinez-Calvillo, S., Nguyen, D., Stuart, K. and Myler, P.J. (2004). Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell*, **3**, 506-517.

Maslov, D.A., Sturm, N.R., Niner, B.M., Gruszynski, E.S., Peris, M. and Simpson, L. (1992). An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. *Mol Cell Biol*, **12**, 56-67.

Matthews, K.R. (1999). Developments in the differentiation of *Trypanosoma brucei*. *Parasitol Today*, **15**, 76-80.

Matthews, K.R., Ellis, J.R. and Paterou, A. (2004). Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol*, **20**, 40-47.

Matthews, K.R. and Gull, K. (1994). Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J Cell Biol*, **125**, 1147-1156.

Matthews, K.R. and Gull, K. (1997). Commitment to differentiation and cell cycle re-entry are coincident but separable events in the transformation of African trypanosomes from their bloodstream to their insect form. *J Cell Sci*, **110 (Pt 20)**, 2609-2618.

Matthews, K.R., Sherwin, T. and Gull, K. (1995). Mitochondrial genome repositioning during the differentiation of the African trypanosome between life cycle forms is microtubule mediated. *J Cell Sci*, **108 (Pt 6)**, 2231-2239.

Matthews, K.R., Tschudi, C. and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev*, **8**, 491-501.

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S. and Matthews, K. (2006). Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res*, **34**, 5312-5324.

McLintock, L.M., Turner, C.M. and Vickerman, K. (1993). Comparison of the effects of immune killing mechanisms on *Trypanosoma brucei* parasites of slender and stumpy morphology. *Parasite Immunol*, **15**, 475-480.

Melville, S.E. (1997). Parasite genome analysis. Genome research in *Trypanosoma brucei*: chromosome size polymorphism and its relevance to genome mapping and analysis. *Trans R Soc Trop Med Hyg*, **91**, 116-120.

Melville, S.E., Gerrard, C.S. and Blackwell, J.M. (1999). Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. *Chromosome Res*, **7**, 191-203.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A. and Blackwell, J.M. (1998). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol Biochem Parasitol*, **94**, 155-173.

Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genet.* **36**, 1073–1078

Mitchell, P. and Tollervey, D. (2000). mRNA stability in eukaryotes. *Curr Opin Genet Dev*, **10**, 193-198.

Morking, P.A., Dallagiovanna, B.M., Foti, L., Garat, B., Picchi, G.F., Umaki, A.C., Probst, C.M., Krieger, M.A., Goldenberg, S. and Fragoso, S.P. (2004). *Tc*ZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides *in vitro*. *Biochem Biophys Res Commun*, **319**, 169-177.

Mosser, E.A., Kasanov, J.D., Forsberg, E.C., Kay, B.K., Ney, P.A. and Bresnick, E.H. (1998). Physical and functional interactions between the transactivation domain of the hematopoietic transcription factor NF-E2 and WW domains. *Biochemistry*, **37**, 13686-13695.

Murphy, W.J., Watkins, K.P. and Agabian, N. (1986). Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell*, **47**, 517-525.

Navarro, M. and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*, **414**, 759-763.

Nichols, R.C., Wang, X.W., Tang, J., Hamilton, B.J., High, F.A., Herschman, H.R. and Rigby, W.F. (2000). The RGG domain in hnRNP A2 affects subcellular localization. *Exp Cell Res*, **256**, 522-532.

Nolan, D.P., Rolin, S., Rodriguez, J.R., Van Den Abbeele, J. and Pays, E. (2000). Slender and stumpy bloodstream forms of *Trypanosoma brucei* display a

differential response to extracellular acidic and proteolytic stress. *Eur J Biochem*, **267**, 18-27.

Ochsenreiter, T. and Hajduk, S.L. (2006). Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep*, **7**, 1128-1133.

Parsons, M. (2004). Glycosomes: parasites and the divergence of peroxisomal purpose. *Mol Microbiol*, **53**, 717-724.

Paterou, A. (2004). Morphological and developmental modulation by a novel CCCH zinc finger protein in *Trypanosoma brucei*. PhD thesis. University of Manchester.

Paterou, A., Walrad, P., Craddy, P., Fenn, K. and Matthews, K. (2006). Identification and stage-specific association with the translational apparatus of *Tb*ZFP3, a CCCH Protein that promotes Trypanosome Life-cycle development. *J Biol Chem*, **281**, 39002-39013.

Pays, E. (1985). Gene conversion in trypanosome antigenic variation. *Prog Nucleic Acid Res Mol Biol*, **32**, 1-26.

Pirozzi, G., McConnell, S.J., Uveges, A.J., Carter, J.M., Sparks, A.B., Kay, B.K. and Fowlkes, D.M. (1997). Identification of novel human WW domain-containing proteins by cloning of ligand targets. *J Biol Chem*, **272**, 14611-14616.

Quijada, L., Guerra-Giraldez, C., Drozdz, M., Hartmann, C., Irmer, H., Ben-Dov, C., Cristodero, M., Ding, M. and Clayton, C. (2002). Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic Acids Res*, **30**, 4414-4424.

Ramos, A., Hollingworth, D. and Pastore, A. (2003). G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*, **9**, 1198-1207.

Reese, K.J., Dunn, M.A., Waddle, J.A. and Seydoux, G. (2000). Asymmetric segregation of PIE-1 in *C. elegans* is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. *Mol Cell*, **6**, 445-455.

Robinson, D.R., Sherwin, T., Ploubidou, A., Byard, E.H. and Gull, K. (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *J Cell Biol*, **128**, 1163-1172.

Robinson, N.P., Burman, N., Melville, S.E. and Barry, J.D. (1999). Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Mol Cell Biol*, **19**, 5839-5846.

Roditi, I. and Clayton, C. (1999). An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. *Mol Biochem Parasitol*, **103**, 99-100.

Roditi, I., Furger, A., Ruepp, S., Schurch, N. and Butikofer, P. (1998). Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol Biochem Parasitol*, **91**, 117-130.

Roditi, I., Schwarz, H., Pearson, T.W., Beecroft, R.P., Liu, M.K., Richardson, J.P., Buhring, H.J., Pleiss, J., Bulow, R., Williams, R.O. and Overath, P. (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *J Cell Biol*, **108**, 737-746.

Rolin, S., Paindavoine, P., Hanocq-Quertier, J., Hanocq, F., Claes, Y., Le Ray, D., Overath, P. and Pays, E. (1993). Transient adenylate cyclase activation accompanies differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Mol Biochem Parasitol*, **61**, 115-125.

Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C. and Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *Embo J*, **20**, 4803-4813.

Schild, L., Lu, Y., Gautschi, I., Schneeberger, E., Lifton, R.P. and Rossier, B.C. (1996). Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. *Embo J*, **15**, 2381-2387.

Schneider, A., McNally, K.P. and Agabian, N. (1994). Nuclear-encoded mitochondrial tRNAs of *Trypanosoma brucei* have a modified cytidine in the anticodon loop. *Nucleic Acids Res*, **22**, 3699-3705.

Schubert, C.M., Lin, R., de Vries, C.J., Plasterk, R.H. and Priess, J.R. (2000). MEX-5 and MEX-6 function to establish soma/germline asymmetry in early *C. elegans* embryos. *Mol Cell*, **5**, 671-682.

Schurch, N., Furger, A., Kurath, U. and Roditi, I. (1997). Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma brucei*. *Mol Biochem Parasitol*, **89**, 109-121.

Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Priess, J.R. and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature*, **382**, 713-716.

Shapiro, S.Z., Naessens, J., Liesegang, B., Moloo, S.K. and Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Trop*, **41**, 313-323.

Sloof, P., Menke, H.H., Caspers, M.P. and Borst, P. (1983). Size fractionation of *Trypanosoma brucei* DNA: localization of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. *Nucleic Acids Res*, **11**, 3889-3901.

Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986). Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc Natl Acad Sci U S A*, **83**, 1383-1387.

Stich, A., Abel, P.M. and Krishna, S. (2002). Human African trypanosomiasis. *Bmj*, **325**, 203-206.

Stuart, K.D., Schnaufer, A., Ernst, N.L. and Panigrahi, A.K. (2005). Complex management: RNA editing in trypanosomes. *Trends Biochem Sci*, **30**, 97-105.

Sudol, M. (1996). Structure and function of the WW domain. *Prog Biophys Mol Biol*, **65**, 113-132.

Sudol, M., Chen, H.I., Bougeret, C., Einbond, A. and Bork, P. (1995). Characterization of a novel protein-binding module - the WW domain. *FEBS Lett*, **369**, 67-71.

Sudol, M. and Hunter, T. (2000). NeW Wrinkles for an old domain. *Cell*, **103**, 1001-1004.

Sutton, A., Heller, R.C., Landry, J., Choy, J.S., Sirko, A. and Sternglanz, R. (2001). A novel form of transcriptional silencing by Sum1-1 requires Hst1 and the origin recognition complex. *Mol Cell Biol*, **21**, 3514-3522.

Tabara, H., Hill, R.J., Mello, C.C., Priess, J.R. and Kohara, Y. (1999). pos-1 encodes a cytoplasmic zinc-finger protein essential for germline specification in *C. elegans*. *Development*, **126**, 1-11.

Tacahashi, Y., Helmling, S. and Moore, C.L. (2003). Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. *Nucleic Acids Res*, **31**, 1744-1752.

Tan, T.H., Bochud-Allemann, N., Horn, E.K. and Schneider, A. (2002). Eukaryotic-type elongator tRNAMet of *Trypanosoma brucei* becomes formylated after import into mitochondria. *Proc Natl Acad Sci U S A*, **99**, 1152-1157.

Tetley, L. and Vickerman, K. (1985). Differentiation in *Trypanosoma brucei*: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. *J Cell Sci*, **74**, 1-19.

Timms, M.W., van Deursen, F.J., Hendriks, E.F. and Matthews, K.R. (2002). Mitochondrial development during life cycle differentiation of African trypanosomes: evidence for a kinetoplast-dependent differentiation control point. *Mol Biol Cell*, **13**, 3747-3759.

Tosato, V., Ciarloni, L., Ivens, A.C., Rajandream, M.A., Barrell, B.G. and Bruschi, C.V. (2001). Secondary DNA structure analysis of the coding strand switch regions of five *Leishmania major* Friedlin chromosomes. *Curr Genet*, **40**, 186-194.

Turner, C.M., Aslam, N. and Dye, C. (1995). Replication, differentiation, growth and the virulence of *Trypanosoma brucei* infections. *Parasitology*, **111 (Pt 3)**, 289-300.

Turner, C.M. and Barry, J.D. (1989). High frequency of antigenic variation in *Trypanosoma brucei rhodesiense* infections. *Parasitology*, **99 Pt 1**, 67-75.

Tyler, K.M., Matthews, K.R. and Gull, K. (1997). The bloodstream differentiation-division of *Trypanosoma brucei* studied using mitochondrial markers. *Proc R Soc Lond B Biol Sci*, **264**, 1481-1490.

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

Ulrich, H., Magdesian, M.H., Alves, M.J. and Colli, W. (2002). *In vitro* selection of RNA aptamers that bind to cell adhesion receptors of *Trypanosoma cruzi* and inhibit cell invasion. *J Biol Chem*, **277**, 20756-20762.

van Helden, J., Andre, B. and Collado-Vides, J. (1998). Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol*, **281**, 827-842.

Van Hellemond, J.J., Neuville, P., Schwarz, R.T., Matthews, K.R. and Mottram, J.C. (2000). Isolation of *Trypanosoma brucei* CYC2 and CYC3 cyclin genes by rescue of a yeast G(1) cyclin mutant. Functional characterization of CYC2. *J Biol Chem*, **275**, 8315-8323.

Van Hellemond, J.J., Opperdoes, F.R. and Tielens, A.G. (1998). Trypanosomatidae produce acetate via a mitochondrial acetate:succinate CoA transferase. *Proc Natl Acad Sci U S A*, **95**, 3036-3041.

van Weelden, S.W., Fast, B., Vogt, A., van der Meer, P., Saas, J., van Hellemond, J.J., Tielens, A.G. and Boshart, M. (2003). Procyclic *Trypanosoma brucei* do not use Krebs cycle activity for energy generation. *J Biol Chem*, **278**, 12854-12863.

Vanhamme, L. and Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiol Rev*, **59**, 223-240.

Vanhamme, L., Pays, E., McCulloch, R. and Barry, J.D. (2001). An update on antigenic variation in African trypanosomes. *Trends Parasitol*, **17**, 338-343.

Vassella, E., Acosta-Serrano, A., Studer, E., Lee, S.H., Englund, P.T. and Roditi, I. (2001). Multiple procyclin isoforms are expressed differentially during the development of insect forms of *Trypanosoma brucei*. *J Mol Biol*, **312**, 597-607.

Vassella, E., Probst, M., Schneider, A., Studer, E., Renggli, C.K. and Roditi, I. (2004). Expression of a major surface protein of *Trypanosoma brucei* insect forms is controlled by the activity of mitochondrial enzymes. *Mol Biol Cell*, **15**, 3986-3993.

Vassella, E., Reuner, B., Yutzy, B. and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci*, **110 (Pt 21)**, 2661-2671.

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

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

Vickerman, K., Myler, P. and Stuart, K. (1992). African Trypanosomiasis. In : *Immunology and Biology of Parasitic infections*, 3rd Edition, pp197-212 Blackwell Scientific Publishers, Oxford.

Waragai, M., Lammers, C.H., Takeuchi, S., Imafuku, I., Udagawa, Y., Kanazawa, I., Kawabata, M., Mouradian, M.M. and Okazawa, H. (1999). PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum Mol Genet*, **8**, 977-987.

Webb, H., Carnall, N., Vanhamme, L., Rolin, S., Van Den Abbeele, J., Welburn, S., Pays, E. and Carrington, M. (1997). The GPI-phospholipase C of *Trypanosoma brucei* is nonessential but influences parasitemia in mice. *J Cell Biol*, **139**, 103-114.

Welburn, S.C. and Odiit, M. (2002). Recent developments in human African trypanosomiasis. *Curr Opin Infect Dis*, **15**, 477-484.

Wickstead, B., Ersfeld, K. and Gull, K. (2003). The mitotic stability of the minichromosomes of *Trypanosoma brucei*. *Mol Biochem Parasitol*, **132**, 97-100.

Wilusz, C.J., Wormington, M. and Peltz, S.W. (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol*, **2**, 237-246.

Wirtz, E., Leal, S., Ochatt, C. and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol*, **99**, 89-101.

Worthington, M.T., Pelo, J.W., Sachedina, M.A., Applegate, J.L., Arseneau, K.O. and Pizarro, T.T. (2002). RNA binding properties of the AU-rich element-binding recombinant Nup475/TIS11/tristetraprolin protein. *J Biol Chem*, **277**, 48558-48564.

Yang, F. *et al.* (2006). Polysome-bound endonuclease PMR1 is targeted to stress granules via stress-specific binding to TIA-1. *Mol. Cell Biol.* **26**, 8803–8813

Ziegelbauer, K. and Overath, P. (1990). Surface antigen change during differentiation of *Trypanosoma brucei*. *Biochem Soc Trans*, **18**, 731-733.

Ziegelbauer, K., Stahl, B., Karas, M., Stierhof, Y.D. and Overath, P. (1993). Proteolytic release of cell surface proteins during differentiation of *Trypanosoma brucei*. *Biochemistry*, **32**, 3737-3742.

Appendices

Appendix I

HMI – 9
Preparation for 500ml:

IMDM	365ml
Heat inactivated FBS	10%
Serum plus	10%
Hypoxanthine	68mg
Bathocuproine sulfonate	14.1mg
Cysteine	91mg
β -Mercaptoethanol	7µl
Thymidine	19.5mg
Penicillin/Streptomycin (10mg/ml)	5ml

The media was sterilised by filtration though a 0.22 micron filter.

SDM – 79
Preparation for 1 litre:

MEM amino acid solution	8ml
MEM non-essential amino acids	6ml
S-MEM (Eagle)	7g
HEPES	8g
MOPS	5g
Medium 199	2g
Glucose	1
Sodium pyruvate	100mg
L-proline	600mg
L-alanine	200mg
L-arginine	100mg
L-glutamine	300mg
L-threonine	350mg
L-tyrosine	100mg
L-taurine	160mg
L-methionine	70mg
L-phenylalanine	80mg
L-serine	60mg
Adenosine	10mg
Guanosine	10mg
Glucosamine HCL	50mg
Folic acid	4mg
p-aminobenzoic acid	2mg
Biotin	0.2mg
$NaHCO_3$	2g

The media was adjusted to pH 7.3 with NaOH and sterilised with a 0.22 filter. Before use, the media was then supplemented with 10% heat inactivated foetal bovine serum and bovine haemin was added to a final concentration of 7.5mg/l.

Appendix II

Work previously published by Paul Craddy.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 51, pp. 39002–39013, December 22, 2006

Identification and Stage-specific Association with the Translational Apparatus of *Tb*ZFP3, a CCCH Protein That Promotes Trypanosome Life-cycle Development*[S]

Received for publication, May 4, 2006, and in revised form, September 26, 2006 Published, JBC Papers in Press, October 16, 2006, DOI 10.1074/jbc.M604280200

Athina Paterou[1], **Pegine Walrad**[1], **Paul Craddy, Katelyn Fenn, and Keith Matthews**[2]

From the Institute of Immunology and Infection Research, School of Biological Sciences, Ashworth Laboratories, King's Buildings, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland, United Kingdom

The post-transcriptional control of gene expression is becoming increasingly important in the understanding of regulated events in eukaryotic cells. The parasitic kinetoplastids have a unique reliance on such processes, because their genome is organized into polycistronic transcription units in which adjacent genes are not coordinately regulated. Indeed, the number of RNA-binding proteins predicted to be encoded in the genome of kinetoplastids is unusually large, invoking the presence of unique RNA regulators dedicated to gene expression in these evolutionarily ancient organisms. Here, we report that a small CCCH zinc finger protein, *Tb*ZFP3, enhances development between life-cycle stages in *Trypanosoma brucei*. Moreover, we demonstrate that this protein interacts both with the translational machinery and with other small CCCH proteins previously implicated in trypanosome developmental control. Antibodies to this protein also co-immunoprecipitate EP procyclin mRNA and encode the major surface antigen of insect forms of *T. brucei*. Strikingly, although *Tb*ZFP3 is constitutively expressed, it exhibits developmentally regulated association with polyribosomes, and mutational analysis demonstrates that this association is essential for the expression of phenotype. *Tb*ZFP3 is therefore a novel regulator of developmental events in kinetoplastids that acts at the level of the post-transcriptional control of gene expression.

Regulatory events in prokaryotic and eukaryotic organisms are most commonly governed by transcriptional control. In this way transcription factors and their co-regulators can target the promoters for individual genes or coordinately regulated gene networks enabling either a specific response to an environmental change or a programmed and more widespread response through cell differentiation.

Gene regulation can also operate post-transcriptionally, either at the level of RNA processing, mRNA stability, translational efficiency, protein localization, or turnover (1). The extent of post-transcriptional control of gene expression differs between different genes, networks, cells, and organisms. One group of organisms that has an unusual emphasis on post-transcriptional control is the kinetoplastid protozoa, comprising the important human pathogens, *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania* sp. (2). These organisms, among the most evolutionarily ancient eukaryotes, are unusual in that their genome is organized into long polycistronic transcription units with tens of genes being coordinately transcribed from distant, as yet unidentified, promoters (3). Such co-transcribed gene clusters do not seem, in general, to be organized as co-regulated operons; instead adjacent genes can show differential gene expression, for example in different life-cycle stages. This places the emphasis of gene regulation in these organisms almost entirely at the post-transcriptional level. However, although the kinetoplastids conserve the conventional eukaryotic machinery for mRNA turnover (4), little is known about the specific trans-acting factors that govern developmental events at the level of gene expression. Moreover, transcriptional control has only been observed for two protein-coding transcription units in *T. brucei*: those encoding the variant surface glycoprotein gene (in bloodstream forms of the parasite) and procyclin genes, which encode a protein coat on the forms of the parasite found in its tsetse fly vector (reviewed in Ref. 2). In contrast to the remainder of trypanosome transcription units these are unique in being transcribed by RNA polymerase I (5, 6).

Differentiation between bloodstream and tsetse-midgut (procyclic) forms of *T. brucei* entails a large number of fundamental changes in these unicellular protozoa (7). In addition to the aforementioned surface antigen exchange, these include changes in cell morphology (8), organelle development and activity (9), metabolism (10, 11), and cell-cycle control (12–14). Usefully, if differentiation is initiated with a subtype of the bloodstream parasite population that accumulates to near homogeneity at the peak of each parasitemia (*i.e.* stumpy forms), then cell differentiation is almost completely synchronous in the population (15). This has allowed a mapping of the events in this developmental pathway, thus revealing a high order of temporal and spatial regulation. It has also allowed the identification of molecules transiently enriched during this process. One such protein, *Tb*ZFP1, has been found to be important for repositioning of the kinetoplast (an elaborated mitochondrial genome) from the posterior end of the cell to a

* This work was supported by the Greek State scholarship foundation and the Wellcome Trust (to A. P.) and by a program grant from the Wellcome Trust (to K. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental text and Fig. S1.

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 44-131-651-3639; Fax: 44-131-651-3670; E-mail: keith.matthews@ed.ac.uk.

237

position midway between the nucleus and cell posterior (16, 17). This is one of the major morphological events that characterize differentiation between bloodstream and procyclic forms and is mediated by posterior outgrowth of the trypanosome microtubule corset (8). A related protein, *Tb*ZFP2, also generates a procyclic stage-specific morphological phenotype when overexpressed and is required in bloodstream forms for efficient differentiation, as revealed by transcript-specific RNAi[3] (16).

*Tb*ZFP1 and *Tb*ZFP2 are members of a family of proteins characterized by their possession of the zinc finger motif $C_{7-8}C_5C_3H$ (CCCH) and small size (<150 amino acids). Although *Tb*ZFP1 and -2 are unique to kinetoplastids, the CCCH motif is an RNA binding zinc finger found in a range of eukaryotic proteins involved at all levels of post-transcriptional gene expression control (18–21). In particular, the mammalian tristetraprolin (TTP) family of proteins recognize AU-rich elements in the 3′-untranslated region of mRNAs causing regulated instability (18), whereas in *Caenorhabditis elegans* regulated control of translation and mRNA localization during embryonic development is conferred by POS1 (22), MEX3 (23), and PIE-1 (19). Consistent with this, the homologue of *Tb*ZFP1 in *T. cruzi* has been shown to bind RNA and interact with identified regulatory RNA elements that control developmental gene expression (24). Notably, *Tb*ZFP1 and -2 also contain protein-interaction domains with *Tb*ZFP2 possessing a WW motif closely related to one of those found in E3 ubiquitin ligases.

The recent completion of the genome sequence for three kinetoplastids reveals an unusual number of CCCH proteins encoded in each organism, with 65 predicted CCCH proteins encoded in *T. brucei*, compared with 12 in *Schizosaccharomyces pombe* (4). Among this set, a third small CCCH protein has been identified, which enhances trypanosome life-cycle differentiation. Moreover, analysis of the interactions of *Tb*ZFP3 *in vivo* demonstrates its capacity to complex with both *Tb*ZFP1 and *Tb*ZFP2. This supports a modular model for functional interactions among this unusual protein family, each of which has been implicated in the regulation of differentiation events. Finally, we demonstrate that *Tb*ZFP3 exhibits a procyclic stage-specific association with polyribosomes despite its constitutive expression through the life cycle and that this association is necessary for the expression of phenotype. This supports a role for this unique protein family in the post-transcriptional control of developmental events in the trypanosome life cycle.

EXPERIMENTAL PROCEDURES

Trypanosomes—For ectopic expression, *T. brucei* 427 bloodstream and procyclic forms were used, each being previously engineered to express the tetracycline repressor protein, enabling regulated gene expression. Slender and stumpy form RNA, protein extracts, and polysomal extracts were prepared from *T. brucei rhodesiense* EATRO 2340. Stumpy forms were isolated 5–6 days after inoculation of mice with 1×10^5 *T. brucei rhodesiense* EATRO 2340 GUP2962 when the population was >80% stumpy by morphology. Parasite transfection was carried out as previously described (16), with procyclic cells being cultured in SDM-79 and selected with 20 μg/ml hygromycin, whereas bloodstream forms were cultured in HMI-9 (at 37 °C, 5% CO_2) and selected with 2 μg/ml hygromycin.

Plasmid constructs were each generated in the trypanosome expression vector pHD451 (30). Deletion and mutation of *Tb*ZFP3 were carried out either via the Stratagene mutate-a-base mutagenesis kit or by a PCR mutagenesis strategy. Complete details of plasmid constructions in this report, and associated primers, are available from the authors upon request.

Northern and Western Blotting, Immunoprecipitation, and Polysome Fractionation—RNA was prepared from bloodstream or procyclic forms using an RNeasy RNA isolation kit (Qiagen), and RNAs were resolved on 1% formaldehyde gels. Hybridization was carried out using digoxigenin-labeled Riboprobes with post hybridization washes being carried out at 68 °C, 0.1× SSC. Signal detection was via chemiluminescence using CDP-Star (Roche Applied Science) as a reaction substrate.

Protein extracts were prepared from 1×10^6 bloodstream or procyclic forms in Laemmli sample buffer. Samples were transferred to nitrocellulose membranes via electroblotting, and effective transfer was verified by Ponceau staining. Blots were blocked with 5% powdered milk in phosphate-buffered saline overnight and then incubated with 1/500 dilution (in 5% milk in phosphate-buffered saline) of anti-*Tb*ZFP3 antibody, this being raised against the peptide antigen N-DSSQMQQVGHDVP-PMA-C in rabbits (by Eurogentec, Belgium). After incubation with a 1/5000 dilution of anti-rabbit IgG-horseradish peroxidase conjugate (Dako), signals were detected via chemiluminescence.

For immunoprecipitation, $1–5 \times 10^8$ cells were pelleted and then snap frozen before resuspension in 0.5 ml of IP-150 lysis buffer (10 mM Tris-Cl, pH 8, 150 mM NaCl, 1% Nonidet P-40, 0.5 mM EDTA). After 25–30 strokes in a Dounce homogenizer, the cell extract was cleared by microcentrifugation (10,000 × *g* for 10 min at 4 °C) and used directly for immunoprecipitation, or centrifuged at 100,000 × *g* in a Beckman TLA100.3 rotor to yield a S100 supernatant or pellet. For immunoprecipitation cell extracts were incubated at 4 °C 1 h overnight with antibody (1:500), or with antibody preincubated with the peptide antigen (40 μg) to which it was raised (as a negative control). Washed protein G beads were then incubated with the extract for a further 1 h and pelleted by centrifugation at 10,000 × *g* for 10 min, yielding the flow-through. The beads were washed 6–9 times with IP-150 lysis buffer, and then bound proteins were extracted into boiling Laemmli sample buffer. For RNA immunoprecipitation, the same procedure was used with RNA being isolated in IP-150 lysis buffer plus 1% SDS at 65 °C, and then purified by phenol/chloroform extraction and ethanol precipitation. RNA retrieved from each sample was divided into two reverse transcription reactions using dT_{18}Anchor (GCGCCG-GCGCCTCAGCG) primer at 42 °C for 1 h. Equivalent quantities of cDNA obtained from all samples was amplified using the primers 5′-EP1 GGTGCTGCAACGCTGAAATCTGTTGC, or Actin 5′-GTATAGCGTGTGGATTGGCGGTTCC in combination with dT_{18}Anchor at 65 °C for up to 35 cycles.

[3] The abbreviations used are: RNAi, RNA interference; TTP, tristetraprolin; E3, ubiquitin-protein isopeptide ligase; BD, binding domain; AD, activation domain.

For polysome fractionation $0.5–1 \times 10^9$ cells were incubated for 10 min with 100 μg/ml cycloheximide and then centrifuged and resuspended in 25 ml of phosphate-buffered saline containing cycloheximide and washed once. The cells were then resuspended in 1 ml of polysome buffer (20 mM Tris-Cl, pH 7.5, 120 mM KCl, 2 mM $MgCl{\cdot}6H_2O$, 1 mM dithiothreitol, 22 μM leupeptin, with 2 μl of RNasin per milliliter) containing cycloheximide, centrifuged at $1,000 \times g$, resuspended in 0.5 ml of polysome buffer, and subjected to 20 strokes in a Dounce homogenizer with Nonidet P-40 being added to 1.2%. The lysate was cleared by centrifugation at $1,000 \times g$. The lysate was then fractionated on a 15–50% sucrose gradient at $240,000 \times g$ in a Beckman SW40Ti rotor, and fractions were collected using a peristaltic pump connected to an AKTA Basic high-performance liquid chromatograph and fraction collector, with RNA concentration being monitored at 254 nm.

Cell Image Acquisition and Morphometric Analysis—Cells were processed for immunofluorescence as previously described. Kinetoplast-posterior measurements were taken using Scion image 1.62. Cell images were captured using a Zeiss Axioscop 2 and processed using Adobe Photoshop CS.

Yeast Two-hybrid Analysis—The Matchmaker system 3 (Clontech) was used throughout, and protocols were followed according to the manufacturer's instructions with the following exceptions: the LexA DNA binding domain plasmid pSTT91 (25) was digested with EcoR1/BamH1to allow for in-frame ligation of the relevant *Tb*ZFP coding region. The activation domain plasmid pGADT7 was created using an EcoR1/XhoI digestion allowing the insertion of the appropriate *Tb*ZFP open reading frame. Co-transformation of 1 μg of both these fusion plasmids into the L40 yeast strain (26) was achieved using the high efficiency polyethylene glycol/lithium acetate method (27), and double transformants were selected on plates lacking Leu and Trp. Activation of the His3 reporter gene by reconstitution of the GAL4 transcription factor encoded in the fusion proteins was characterized by growth on plates lacking Leu/Trp/His/Ade. Activation of the β-galactosidase reporter gene was observed by replica-plating the colonies onto nylon membrane (Amersham Biosciences) on a YPDa plate, incubating for 16 h, and then performing a colony lift assay (28).

RESULTS

We searched the *T. brucei* genome sequence data base (www.geneDB.org) using the *Tb*ZFP1 and *Tb*ZFP2 protein sequences to identify further members of this protein family. This identified an open reading frame (Tb927.3.720) encoding a 130-amino acid predicted protein comprising a CCCH domain toward its C terminus and a WW protein-interaction domain toward its N terminus (Fig. 1*A*). This matched the overall motif organization of *Tb*ZFP2 (Tb11.01.6590), which has been previously characterized to have a role in trypanosome differentiation and posterior end morphogenesis (16). Similar to *Tb*ZFP2, the WW domain in the newly identified protein was most similar to a particular WW domain in E3 ubiquitin ligases. Unlike *Tb*ZFP2, however, *Tb*ZFP3 also possessed three intact copies of the RGG-predicted RNA-binding motif positioned centrally between the WW and CCCH domain. The newly identified gene has been recently described in the context of a bioinformatic analysis of homologous genes in *T. cruzi* and has been named *Tc*ZFP2b (29). However, although they share several motifs associated with molecular interactions, the primary amino acid sequences of *Tb*ZFP2a and -b are dissimilar (Fig. 1*A*), and their genes are unlinked. Therefore, we prefer to denote this by renaming of the new molecule *Tb*ZFP3.

The previously characterized differentiation regulator *Tb*ZFP1 is silenced in bloodstream forms at the protein level but is transiently elevated during differentiation to procyclic forms where expression is then retained (16, 17). In contrast, *Tb*ZFP2 is expressed equally in bloodstream and procyclic forms and exhibits uniform expression throughout synchronous differentiation between stumpy and procyclic forms (16). To address whether *Tb*ZFP3 is regulated during the parasite life cycle, we examined the mRNA and protein levels for this gene in bloodstream and procyclic forms. Fig. 1*B* demonstrates that this 390-bp-coding region is contained within an mRNA of ~1.9 kb that is expressed equally in slender and stumpy bloodstream forms and in procyclic forms. To verify that the gene was translated into protein and to examine its expression profile, an anti-peptide antibody was raised to the sequence N-DSSQMQQVGHDVPPMA-C, this being located at the extreme C terminus of *Tb*ZFP3 protein. When reacted with protein derived from bloodstream monomorphic forms and procyclic forms, a band was detected at the predicted molecular mass for *Tb*ZFP3 (14 kDa) in each life-cycle stage, and at approximately equal levels, consistent with the RNA profile (Fig. 1*B*). The profile of *Tb*ZFP3 protein expression was also assayed during synchronous differentiation between bloodstream stumpy forms and procyclic forms, after culture in SDM-79 medium at 27 °C containing 6 mM *cis*-aconitate (Fig. 1*C*). This demonstrated approximately equivalent expression throughout differentiation with no evidence for a transient elevation in expression during this synchronous transition. This matched the expression profile of *Tb*ZFP2 but differs from *Tb*ZFP1, which is elevated in expression level around 4–8 h through differentiation (16). Therefore, *Tb*ZFP3 is consistently expressed in both bloodstream and procyclic forms and during synchronous differentiation between these forms.

Ectopic Expression of TbZFP3 Induces Enhanced Differentiation—Prior knockdown of *Tb*ZFP2 in bloodstream forms by RNAi had resulted in decreased efficiency of differentiation as assessed by gain of the insect-stage-specific antigen, EP procyclin (16). This implicated *Tb*ZFP2 in the control of early differentiation events, contrasting with *Tb*ZFP1, which appears to act later in the differentiation program (17). To determine whether *Tb*ZFP3 also had a role in developmental control, we initially attempted RNAi for this molecule in bloodstream forms. Although viable cell lines were generated using several available RNAi vectors, effective reduction of *Tb*ZFP3 protein was not achieved (a maximum reduction of 60% was observed, data not shown). However, cell lines were generated in which tetracycline-regulated ectopic overexpression of *Tb*ZFP3 was possible. Specifically, a copy of *Tb*ZFP3 encoding a C-terminal Ty1 epitope tag (denoted *Tb*ZFP3-Ty) was inserted into the trypanosome expression vector pHD451 (30) and transfected into *T. brucei* 427 monomorphic bloodstream forms, which express the tetracycline repressor protein. The resulting cell

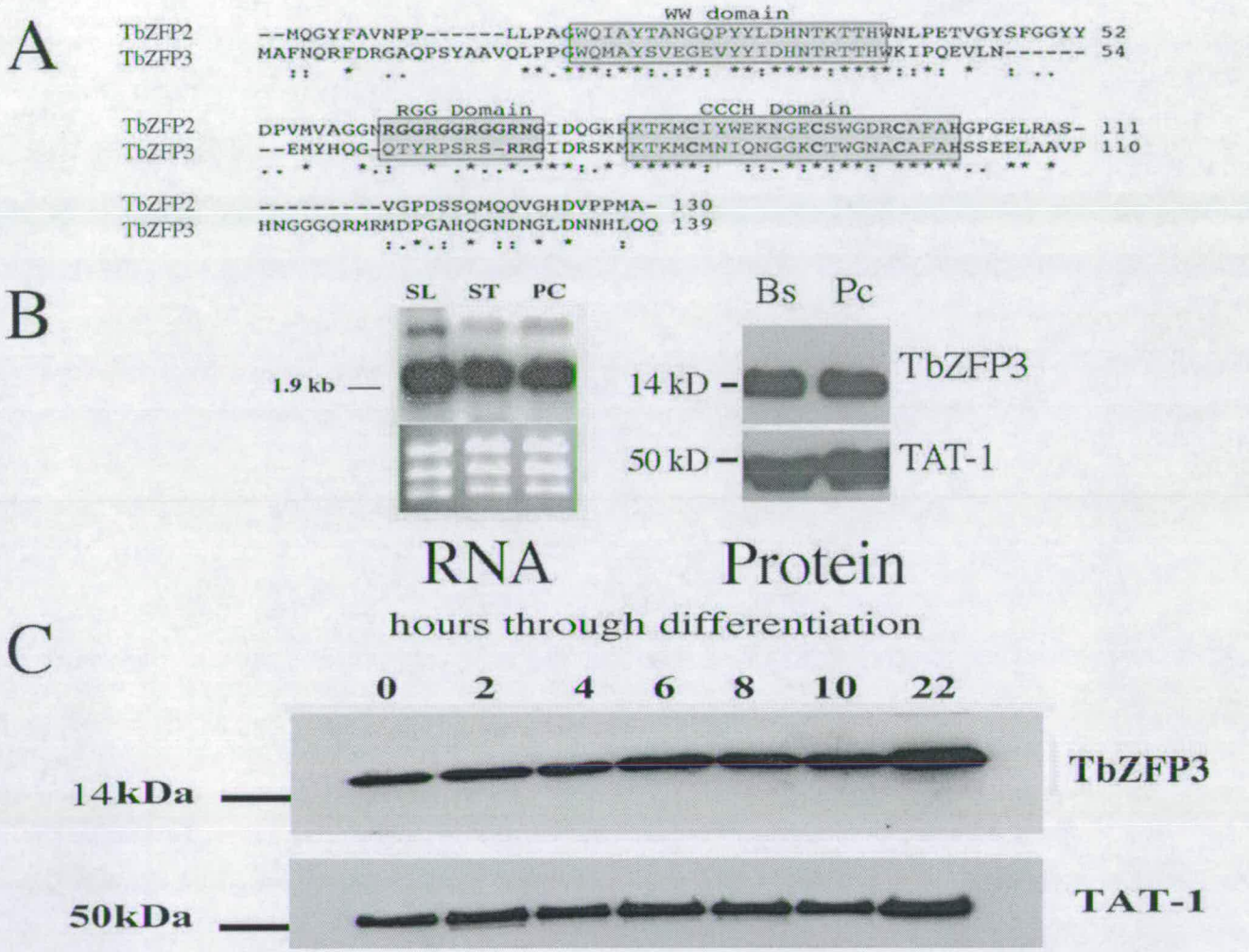

FIGURE 1. **Identification and expression of a novel CCCH protein.** *A*, sequence alignment of *Tb*ZFP3 with *Tb*ZFP2, a previously identified small CCCH protein in trypanosomes. The WW, RGG, and CCCH domains are *boxed* in each case, highlighting the similarity in structural organization of the two sequences, despite their lacking obvious primary sequence identity. *B*, expression profile of *Tb*ZFP3 during the trypanosome life cycle. The *left-hand panel* shows a Northern blot of RNA derived from bloodstream slender (*SL*), bloodstream stumpy (*ST*), and procyclic (*PC*) forms hybridized with a *TbZFP3* mRNA-specific Riboprobe. The transcript migrates at 1.9 kb, considerably larger than the coding region of *TbZFP3*, suggesting the presence of untranslated regions extending for ~1.5 kb. The *right-hand panel* shows protein expression of *Tb*ZFP3 in monomorphic bloodstream forms (*Bs*) or procyclic (*Pc*) forms. The protein migrates at the expected molecular mass (14 kDa). Staining with an antibody to trypanosome α-tubulin (*TAT1*) provides a control for loading. *C*, *Tb*ZFP3 protein expression during synchronous differentiation between a uniform population of bloodstream stumpy forms and procyclic forms. Protein samples were prepared at 2-h intervals during differentiation to procyclic forms. *Tb*ZFP3 expression was relatively uniform during the differentiation time course, matching the expression of α-tubulin, which is detected with the antibody TAT1.

lines were then induced to express *Tb*ZFP3-Ty by the addition of tetracycline, this being monitored by Western blotting using the *Tb*ZFP3-specific anti-peptide antibody allowing simultaneous detection of the endogenous and transgenic protein. Fig. 2*B* (*Day 0* samples) shows inducible ectopic expression of *Tb*ZFP3-Ty in bloodstream forms was achieved, generating a combined 1- to 2-fold overexpression of *Tb*ZFP3 compared with wild-type cells.

Having successfully generated a cell line capable of perturbed *Tb*ZFP3 expression we examined the consequences of this for the efficiency of differentiation to procyclic forms. Specifically, parental or *Tb*ZFP3-Ty-transfected bloodstream forms were induced, or not, with tetracycline for 3 days in culture, and then differentiation was initiated under standard conditions. Thereafter, EP procyclin expression was followed daily via immunofluorescence, with protein samples also being prepared to monitor *Tb*ZFP3-Ty expression. Fig. 2*A* (*days 1–3* samples) shows the result of four such experiments, which revealed that ectopic expression of *Tb*ZFP3-Ty resulted in enhanced EP procyclin expression in the population. Thus, at each time point after the induction of differentiation the proportion of EP procyclin expressers was consistently up to 2-fold higher than in uninduced or parental controls (either in the presence or absence of tetracycline). This correlated with the detection of the ectopically expressed protein by Western blotting (Fig. 2*B*). To complement this IFA-based counting assay (which qualitatively scores cells as being "+" or "−" on an individual cell basis), differentiation was also scored in an independent experiment by flow cytometry for EP expression at the population level (Fig. 2*C*). This confirmed enhanced differentiation upon

240

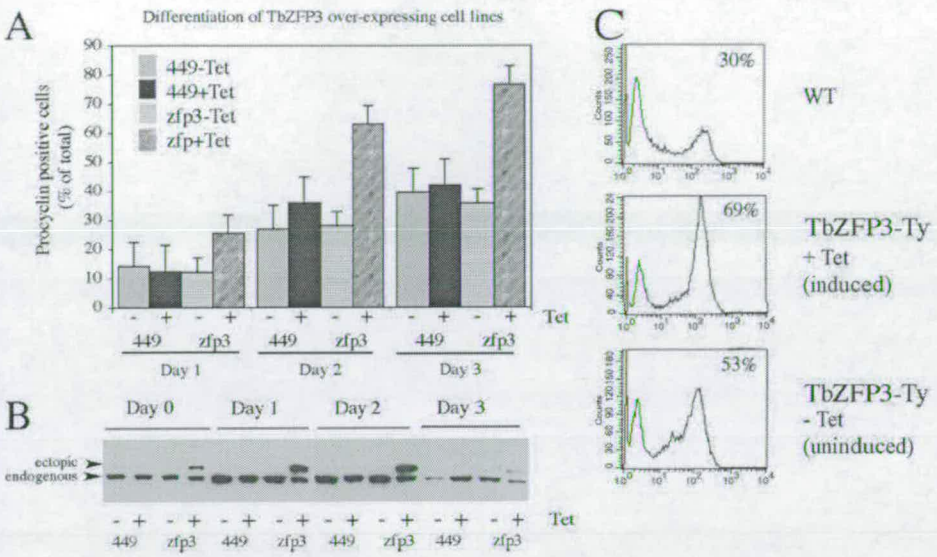

FIGURE 2. **TbZFP3 ectopic expression generates an enhanced differentiation phenotype.** *A*, a graph shows the percent EP procyclin expressers in either parental cells in the presence or absence of tetracycline, or in a cell line in which *Tb*ZFP3-Ty expression is inducible. Procyclin expression was monitored for 3 days after exposure of the cells to 6 mM *cis*-aconitate by immunofluorescence. The data represent means of four independent experiments, with the standard error of the mean being presented for each *bar*. *B*, during the differentiation experiments shown in *A*, the expression of endogenous and ectopic *Tb*ZFP3 was monitored using the antibody to *Tb*ZFP3. In each case *Tb*ZFP-Ty expression was detected only in the induced population. *C*, an analysis of EP procyclin expression by flow cytometry. Cells were processed for staining with antibody to EP procyclin at 48 h after the initiation of differentiation, and then the proportion of cells expressing this surface marker was assayed by fluorescence-activated cell sorting. Expression of *Tb*ZFP3-Ty enhances the proportion of cells expressing EP procyclin.

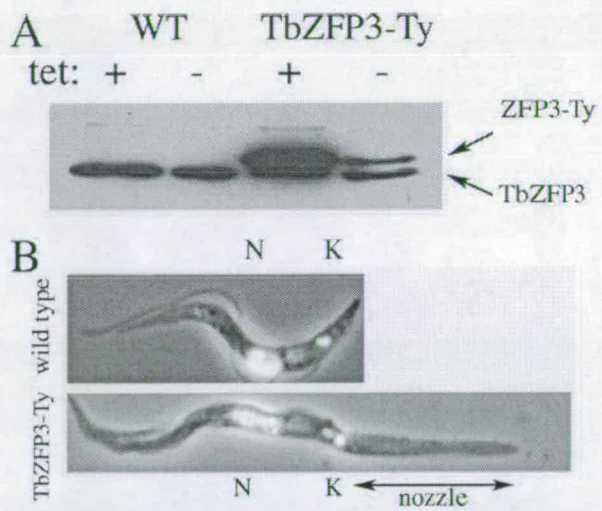

FIGURE 3. **Ectopic expression of *Tb*ZFP3-Ty in procyclic forms.** *A*, shows cells in which *Tb*ZFP3-Ty expression is induced (or not) by the presence of tetracycline, together with wild-type controls. In each case *Tb*ZFP3 is detected with the *Tb*ZFP3-specific antibody; ectopic expression of *Tb*ZFP3-Ty is highly expressed in the induced cell population, and also present more weakly in the uninduced cell line (due to incomplete repression). *B*, shows either a wild-type cell, or a cell that has undergone morphological extension at its posterior end following induction of the expression of *Tb*ZFP3-Ty. Note that cell dimensions are unchanged, excepting in the distance between the kinetoplast and cell posterior. *N* = nucleus, *K* = kinetoplast.

*Tb*ZFP3-Ty expression with the induced cell line demonstrating 69% EP-procyclin expression compared with 53% in the uninduced control at 72 h after the initiation of differentiation. We verified that this effect was not due to the increased expression of stumpy characteristics via a number of assays for that life-cycle stage and found no consistent difference from uninduced cells (refer to the supplemental data). Therefore, ectopic expression of *Tb*ZFP3 results in enhanced differentiation to procyclic forms as assessed by EP procyclin expression, consistent with the reduced efficiency of this process upon knockdown of *Tb*ZFP2 by RNAi (16). Combined with the specific inhibition of kinetoplast repositioning induced by genetic ablation of *Tb*ZFP1 (17), this supports the role for this family of proteins in differentiation processes.

Ectopic Expression in Procyclic Forms Causes a Characteristic Morphological Phenotype—In addition to the bloodstream form cell lines, ectopic expression of *Tb*ZFP3 was also investigated in procyclic forms (Fig. 3*A*). In this case, upon induction of *Tb*ZFP3-Ty expression, the cell population was found to produce elongated cells reminiscent of the nozzle phenotype previously observed upon ectopic expression of *Tb*ZFP2 (16) (Fig. 3*B*). As with *Tb*ZFP2 expression, cell elongation upon *Tb*ZFP3 expression involved posterior microtubule extension as revealed by YL1/2 staining of tyrosinated α-tubulin in the trypanosome cytoskeleton (data not shown). However, unlike with ectopic *Tb*ZFP2 expression, where nozzle formation was entirely restricted to cells with 1 kinetoplast and 1 nucleus (*i.e.* G_1 and S phase cells), in cells expressing ectopic *Tb*ZFP3-Ty ~15% of cells with an extended posterior had segregated their kinetoplast. This is indicative of cell cycle progression into G_2, consistent with a similar phenotype observed upon RNAi ablation of the cyclin, CYC2 (31).

The manifestation of the nozzle morphology provided a convenient and quantifiable phenotypic assay with which to dissect important domains in *Tb*ZFP3. Specifically we wanted to evaluate whether the WW domain and CCCH RNA binding domain in *Tb*ZFP3 were essential for morphological extension. Therefore, we generated a series of cell lines in which either the WW or CCCH domain in *Tb*ZFP3 were deleted from the ectopically expressed protein, or in which a single amino acid change was introduced to mutate the CCCH motif to the sequence CCAH. Finally, we generated a cell line in which the Ty tag was absent from the transgenic protein thereby ensuring that phenotypic effects were a result of the native protein and not the introduction of the tag. Fig. 4*A* shows that in each case the ectopic proteins were effectively expressed in the transgenic cell lines. The relative expression between induced and uninduced samples varied somewhat, but in each the level of the induced protein expression was similar. Having verified effective and comparable expression levels for each protein, we then

241

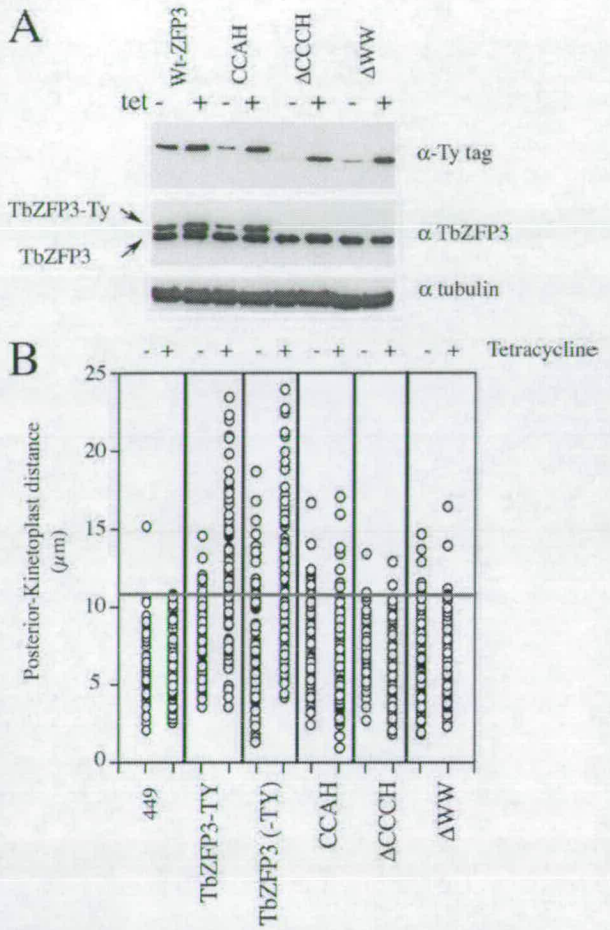

FIGURE 4. **Nozzle formation as a phenotypic assay for important domains in TbZFP3.** *A*, ectopic expression of either *Tb*ZFP3-Ty, or mutant derivatives of this in which the WW domain is deleted, the CCCH domain is deleted, or in which the third cysteine in the CCCH domain is mutated to alanine. In each case, expression of the ectopic proteins is detected with the anti-Ty tag antibody, BB2 (*α Ty tag*). Also shown are the same samples hybridized with the *Tb*ZFP3-specific anti-peptide antibody, which detects both the transgenic protein and endogenous *Tb*ZFP3. Note that deletion of the WW or CCCH domains causes the transgenic and endogenous *Tb*ZFP3 to co-migrate. In each case equivalent protein loading is indicated by staining for trypanosome α-tubulin. *B*, the kinetoplast-posterior dimension for each cell line either when induced, or not, by the presence of tetracycline. In each case 100 cells were scored, with a *horizontal line* being placed at the position of the longest kinetoplast-posterior dimension in the parental population (*449*) to assist visualization of the appearance of the nozzle phenotype. One cell in the parental population also showed an unusually long kinetoplast-posterior dimension: this is seen very occasionally in such populations but is at a frequency far lower than manifestation of the nozzle phenotype.

assayed the degree of morphological extension by measuring the kinetoplast to posterior dimension in 100 cells in each population. This analysis, shown in Fig. 4*B*, demonstrated that, whereas the presence or absence of the Ty tag was not important in nozzle formation, deletion of the WW or CCCH domain, or introduction of a mutation to disrupt the CCCH domain, each significantly abrogated nozzle formation. Thus ectopic expression of either native or Ty-tagged *Tb*ZFP3 generated a mean kinetoplast-posterior dimension of 10.31 μm (±0.35 μm (S.E.)) and 9.52 μm (±0.24 μm) respectively, whereas the values generated for each mutant were: ΔWW, 5.24 μm (±0.21 μm), ΔCCCH 5.40 μm (±0.21 μm), and CCAH, 5.88 μm (±0.19 μm), with these approximating to the kinetoplast-posterior dimension in wild-type procyclic forms of 5.41 μm (± 0.18 μm). We conclude that the CCCH and WW domains in *Tb*ZFP3 are both individually important for generation of the nozzle phenotype.

Molecular Interactions of the TbZFP Protein Family—We have proposed that members of the *Tb*ZFP family may interact on the basis of complementary protein interaction domains whereby *Tb*ZFP2 and -3 each contain a WW domain, and *Tb*ZFP1 contains the complementary target PPPPPY motif (16). To investigate this potential, the ability of *Tb*ZFP1, *Tb*ZFP2, and *Tb*ZFP3 to interact in a pairwise yeast two-hybrid interaction screen was assayed. Thus fusion proteins of *Tb*ZFP1, *Tb*ZFP2, and *Tb*ZFP3 were generated in bait and prey plasmids, which allowed expression of each *Tb*ZFP fused to either the LEXA DNA binding domain (BD) or the GAL4 activation domain (AD). In each case mutants, in which the WW domain in both *Tb*ZFP2 and -3 were deleted, were also generated to establish their function in potential interactions with *Tb*ZFP1. These molecules were then co-transformed pairwise into yeast and colonies streaked onto nonselective (−*TRP* and −*LEU*) or selective (−*ADE* and −*HIS*) plates, as well as being assayed for LacZ production.

The data in Fig. 5 indicate that *Tb*ZFP1 is able to directly interact with *Tb*ZFP2 and -3, this evidenced by the production of LacZ and activation of nutritional markers on −*TRP*, −*LEU*, −*HIS*, and −*ADE* plates (*2DB+1AD* and *3DB+1AD*, Fig. 5). Moreover, this interaction is mediated by the WW motif alone, because deletion of this motif in *Tb*ZFP2 (*2DBΔww+1AD*, Fig. 5) or *Tb*ZFP3 (*3DBΔww+1AD*, Fig. 5) abolishes the interaction of each with *Tb*ZFP1. Although a weak potential interaction was observed between the *Tb*ZFP2-LexA DNA BD fusion and the *Tb*ZFP3-Gal4AD fusion (*2DB+3AD*, Fig. 5), this was not reciprocated when the bait and prey were reversed and is likely to be artifactual. Thus, there appears to be no direct interaction between *Tb*ZFP2 and *Tb*ZFP3 and we found no evidence that these proteins can homodimerize. However, *Tb*ZFP1 can interact with *Tb*ZFP2 or *Tb*ZFP3 with the WW protein binding motifs being essential in this.

Although two-hybrid analysis can prove informative for direct protein interactions *in vitro*, it was also important to validate potential interactions between these proteins *in vivo*. Therefore, the ability of the antibody against *Tb*ZFP3 to co-immunoprecipitate *Tb*ZFP2 and *Tb*ZFP1 was investigated in procyclic cell extracts, with specificity being verified by blocking with the peptide antigen used to raise the antibody. Fig. 6*A* demonstrates that *Tb*ZFP1, -2, and -3 were each co-immunoprecipitated with anti-*Tb*ZFP3 antibody and that this was efficiently blocked by the peptide immunogen. These experiments confirm that *Tb*ZFP1 can interact directly with *Tb*ZFP2 and *Tb*ZFP3 *in vitro* and that all three proteins can be co-immunoprecipitated *in vivo* suggesting their involvement in overlapping or the same protein complex(es).

Cell Fractionation of TbZFP3 Reveals Stage-specific Association with Polyribosomes—To analyze further the cellular location and interactions of *Tb*ZFP3, the distribution of this protein among different cellular compartments was investigated by cell lysis and differential detergent extraction. This resolved

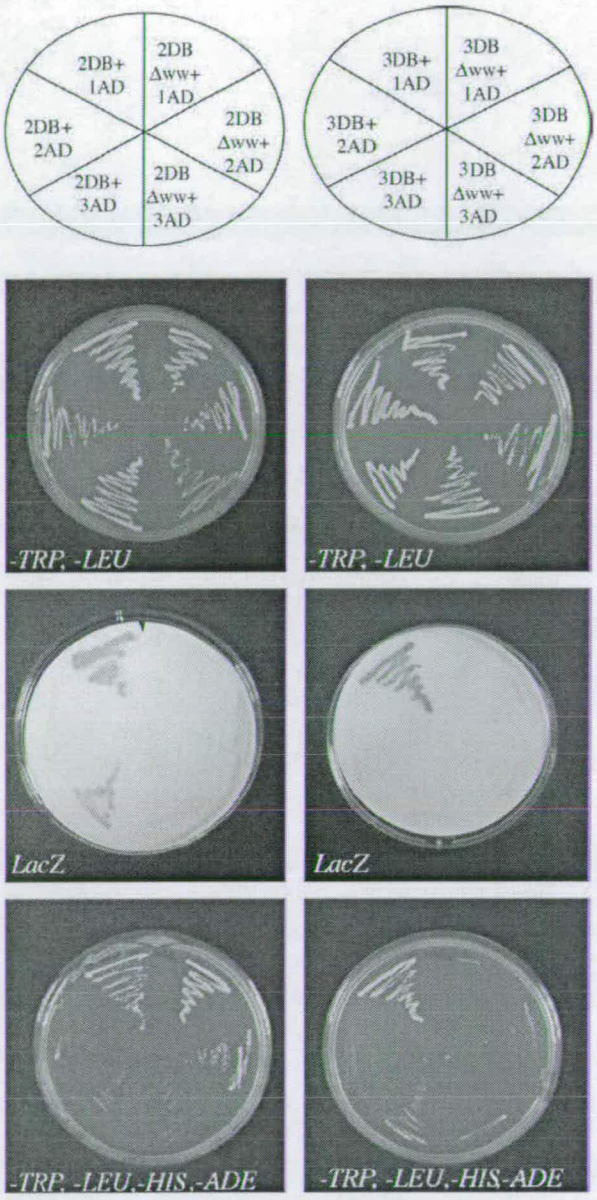

FIGURE 5. **Interactions among the TbZFP family assessed by yeast two-hybrid analysis.** In each case interactions between intact *Tb*ZFP2 or -3 (*2DB* and *3DB*) or mutant derivatives of these lacking the WW domain (*2DB ΔWW* and *3DB ΔWW*) fused to the LexA DNA binding domain are shown in combination with *Tb*ZFP1, -2, or -3 fused to the Gal4 activation domain (*1AD*, *2AD*, and *3AD*). *Tb*ZFP1-LexA DNA binding domain constructs are not included in this analysis, because they autoactivate. *Tb*ZFP2 and *Tb*ZFP3 LexA DNA binding domain constructs show no autoactivation (not shown).

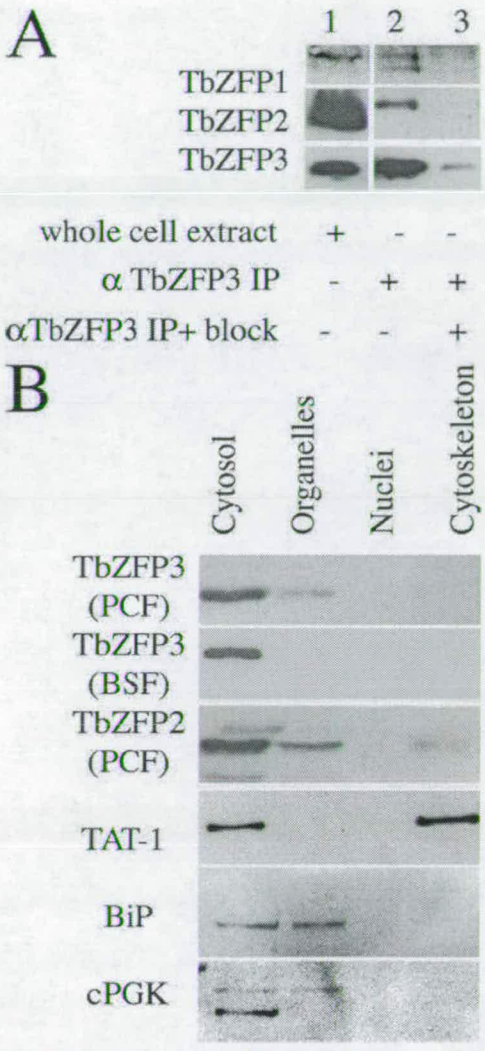

FIGURE 6. **Protein interactions and cellular distribution of TbZFP3.** *A*, co-immunoprecipitation of *Tb*ZFP1, -2, and -3 using anti-*Tb*ZFP3 antibody. In each case procyclic cell extracts were subject to immunoprecipitation using anti-ZFP3 antibody and protein G beads. *Tb*ZFP1, -2, and -3 were each immunoprecipitated by *Tb*ZFP3-antibody, the specificity of this interaction being confirmed by inclusion, as a block, of the peptide immunogen against which the antibody was raised. Consistent with earlier results (16) *Tb*ZFP1 signal is detected as a doublet, the relative intensity of which differs between experiments for unknown reasons. *B*, distribution of *Tb*ZFP3 in *T. brucei* bloodstream (*BSF*) or procyclic form (*PCF*) extracts. *Tb*ZFP3 associates with the cytosolic fraction, as does *Tb*ZFP2. Analysis of the same extracts with antibodies for alpha tubulin (cytoskeletal and cytosolic), BiP (endoplasmic reticulum-associated, but with leakage into the cytosol upon cell fractionation), cytosolic phosphoglycerate kinase (cytosolic) confirmed the identity of each fraction. No protein associated with the nuclear fraction.

trypanosome proteins into cytosolic, nuclear, organellar vesicles/plasma membrane, or cytoskeletal fractions. Fig. 6*B* demonstrates that the majority of *Tb*ZFP3 co-fractionated with cytosolic proteins in both bloodstream and procyclic forms, with a small proportion fractionating with cell organelles/plasma membrane, this most probably representing some cytosolic contamination of this fraction. This distribution was matched by *Tb*ZFP2. Consistent with a cytosolic location, immunofluorescence of *Tb*ZFP3 in procyclic forms (fixed by paraformaldehyde in the presence of 0.1% Triton X-100) revealed a uniform staining pattern through the cell, without concentration in any recognizable organelle (data not shown).

Cell extracts were also analyzed by differential centrifugation, revealing a distribution of *Tb*ZFP3 between S100 supernatant and pellet (data not shown). This suggested the possibility of association of *Tb*ZFP3 with polysomes in procyclic forms. To investigate this further, procyclic cell extracts prepared in polysome buffer in the presence of cycloheximide were separated on 15–50% sucrose gradients, and fractions were analyzed

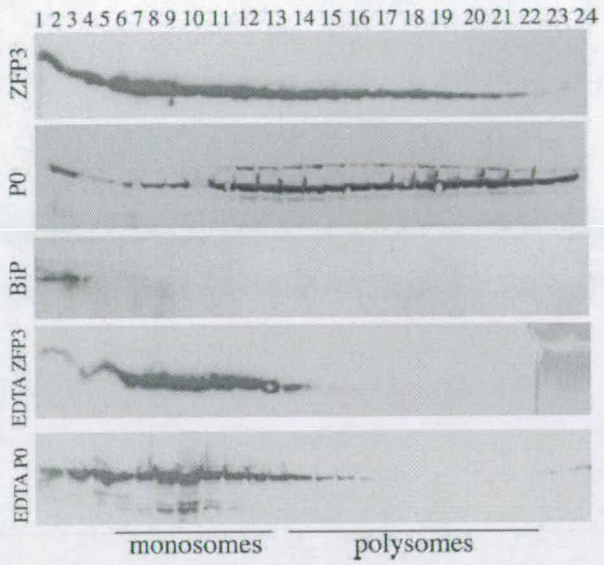

FIGURE 7. **TbZFP3 is polysome-associated in procyclic forms.** Procyclic form cell extracts were separated on 15–50% sucrose gradients, and the presence of *Tb*ZFP3, the ribosomal protein P0, and the non-ribosomal endoplasmic reticulum protein BiP were detected via specific antibodies against each protein. Lower molecular weight fractions are to the *left* of each profile, high molecular weight polysomal fractions are to the *right*. Extracts were prepared in the presence of cycloheximide or in the presence of EDTA, which disassociates polysomes causing redistribution of *Tb*ZFP3 and P0 into lower molecular weight fractions.

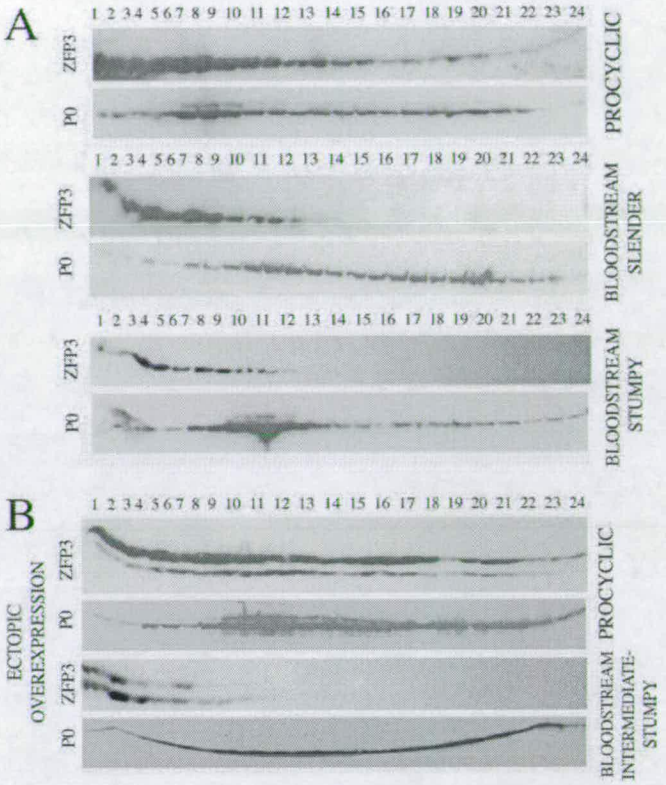

FIGURE 8. **Stage-specific polysome association of TbZFP3.** *A*, polysome association of TbZFP3 is procyclic-enriched. The distribution of TbZFP3 or the ribosomal protein P0 is shown in either procyclic forms, bloodstream slender forms, or bloodstream stumpy forms. *B*, ectopic expression of TbZFP3 in procyclic and bloodstream forms. In each case the ectopic protein (which runs slightly larger than the endogenous protein due to the incorporation of a C-terminal Ty epitope tag) co-sediments with the endogenous protein, this being polysomal in procyclic forms, and non-polysomal in bloodstream forms.

for the presence of *Tb*ZFP3, BiP (a non-polysomal endoplasmic reticulum protein) and ribosomal protein P0 (a ribosomal stalk component that fractionates with monosomes and polysomes). Fig. 7 shows that *Tb*ZFP3 distributed into monosomal (fractions 6–12) and polysomal fractions (fractions 13–22) matching the distribution of P0, although *Tb*ZFP3 was also abundant in the non-polysomal fractions, consistent with the bimodal distribution of *Tb*ZFP3 between the S100 pellet and supernatant. Confirming the association of *Tb*ZFP3 with polysomes in procyclic forms, *Tb*ZFP3 and P0 redistributed to lower molecular weight fractions in EDTA-treated cell extracts (Fig. 7).

The stage-specific phenotypes resulting from *Tb*ZFP3 ectopic expression prompted us to investigate the polysome association of this protein in bloodstream forms, as well as procyclic forms. Interestingly, although *Tb*ZFP3 is expressed at similar levels in bloodstream and procyclic forms, clear evidence for polysome association was only observed in the latter (Fig. 8*A*). Thus, in bloodstream slender forms, *Tb*ZFP3 sedimented in the low molecular weight fractions with a distinct distribution from the polysomal marker P0. Similarly, in homogenous populations of bloodstream stumpy forms *Tb*ZFP3 was also disassociated from polyribosomes, although the abundance of polysomes was considerably reduced as expected in this quiescent life-cycle stage (32). This demonstrated that the association of *Tb*ZFP3 with polysomal fractions was developmentally regulated, being enriched in procyclic forms and absent in bloodstream slender and stumpy forms.

The stage-specific association of *Tb*ZFP3 with polyribosomes prompted us to further determine whether ectopic expression of this protein perturbed its normal associations with the translational apparatus in either bloodstream or procyclic forms. Thus, we examined the distribution of endogenous and ectopic *Tb*ZFP3 in either transgenic procyclic forms (where the nozzle phenotype is generated) or in bloodstream forms (which exhibit enhanced differentiation). More specifically, *Tb*ZFP3 was expressed in either procyclic forms for 48 h (coinciding with the appearance of nozzle cells in the population) or in a bloodstream line that can generate division-arrested intermediate/stumpy forms when grown in mice (33). These cells were grown for 8 days in mice that had been supplied with doxycycline in their drinking water (to induce *Tb*ZFP3-Ty expression), and then parasites were harvested for analysis of the distribution of *Tb*ZFP3 on sucrose gradients. Fig. 8*B* demonstrates that in both bloodstream and procyclic forms the ectopically expressed and endogenous *Tb*ZFP3 exactly co-sedimented, this being polysomal for procyclic forms and non-polysomal for the bloodstream forms. We conclude that ectopically expressed *Tb*ZFP3 does not alter the normal polysomal distribution of this protein, association of the translational apparatus being enriched in procyclic forms regardless of the expression of ectopic *Tb*ZFP3.

Polysome Association Precisely Correlates with Both Nozzle Formation and Enhanced Differentiation—In procyclic forms the morphological phenotype generated by *Tb*ZFP3 ectopic expression was dependent upon the integrity of the WW and

244

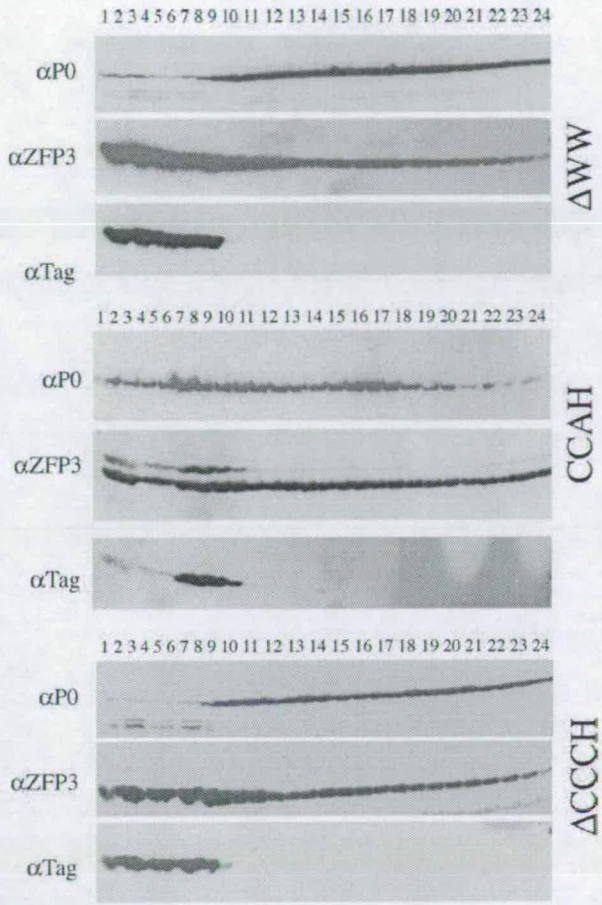

FIGURE 9. **Polysomal association of TbZFP3 mutants.** Ectopic expression of either *Tb*ZFP-ΔWW, *Tb*ZFP3-CCAH, or *Tb*ZFP3-ΔCCCH was induced in procyclic forms, and the distribution of endogenous and ectopic *Tb*ZFP3 was detected using the *Tb*ZFP3-specific antibody (α-*Tb*ZFP3). Specific detection of the ectopic protein was via antibody to the Ty1-eptiope tag incorporated into each mutant protein (α-Ty1). The polysomal distribution was determined by antibody to the P0 protein (α-P0). Note that the ΔCCCH and ΔWW mutants co-migrate with endogenous TbZFP3, whereas the CCAH mutant runs slightly larger than endogenous TbZFP2 (due to incorporation of the Ty1 epitope tag).

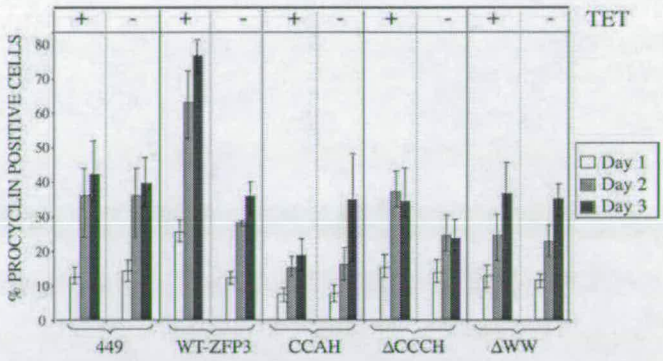

FIGURE 10. **Differentiation capacity of bloodstream forms expressing mutant TbZFP3.** Differentiation efficiency of transgenic bloodstream lines expressing under tetracycline regulation either intact *Tb*ZFP3 or mutants disrupted for the CCCH domain (CCAH or ΔCCCH) or WW domain. The expression of EP procyclin was scored by immunofluorescence after 24, 48, or 72 h. Data represent the mean of four independent experiments, with *error bars* demonstrating the standard error.

CCCH domain. To investigate whether disruption of these domains resulted in disassociation of the ectopic protein from polysomes, the distribution of endogenous and ectopically expressed *Tb*ZFP3 was analyzed on sucrose gradients of procyclic cell extracts. In each case the ectopic and endogenous proteins were detected using either the anti-Ty1 epitope tag antibody BB2 (to detect the ectopic protein only) or the *Tb*ZFP3 antibody (detecting both endogenous and ectopic *Tb*ZFP3). Fig. 9 demonstrates that, when the ΔWW, ΔCCCH, or CCAH mutants were expressed in procyclic forms, the ectopic protein fractionated into the non-polysomal or monosomal region of the gradient, whereas endogenous *Tb*ZFP3 and the ribosomal marker P0 each remained polysome-associated. This demonstrates that *Tb*ZFP3 associates with polysomes only if its WW or CCCH domains are intact. This exactly correlates with the capacity of each mutant to induce morphological extension as determined in Fig. 4.

Having determined that mutations that prevented polysome association also prevented nozzle formation in procyclic forms, we investigated whether the same mutants also disrupted enhanced differentiation of bloodstream forms. Therefore, stable bloodstream trypanosome cell lines were generated expressing each of the *Tb*ZFP3 mutants (ΔWW, ΔCCCH, and CCAH), and the expression of these was induced for 3 days by growth in tetracycline. Thereafter, each was induced to initiate differentiation with *cis*-aconitate, and the percentage of procyclin expressers quantified over the following 72 h by immunofluorescence. Fig. 10 shows that, whereas ectopic expression of wild-type *Tb*ZFP3 enhanced differentiation as expected, the mutants with a disrupted WW or CCCH domain each differentiated with an equivalent efficiency to wild-type parental cells. This demonstrated that the phenotypes generated by ectopic expression of *Tb*ZFP3 in both procyclic forms (nozzle formation) and bloodstream forms (enhanced differentiation) were each dependent upon the integrity of the WW and CCCH domain in the protein. Moreover, manifestation of these phenotypes exactly correlated with the ability of each protein to associate with polysomes. This provides evidence for a cause-and-effect relationship between the function of *Tb*ZFP3 and its association with the translational apparatus.

The enhanced differentiation phenotype induced by TbZFP3 ectopic expression was detected by assay for the expression of EP procyclin protein. Therefore, we assessed whether antibodies to TbZFP3 could co-immunoprecipitate EP procyclin mRNA. Fig. 11*A* demonstrates that this was the case: EP procyclin mRNA was selected by antibody to TbZFP3 (Fig. 11*A*, *lane 1*), and this was prevented in the presence of blocking peptide (Fig. 11*A*, *lane 2*). The specificity of this was confirmed by the absence of actin mRNA, a constitutively expressed transcript, in the immunoprecipitated material (Fig. 11*B*). We conclude that TbZFP3 protein and EP procyclin mRNA co-precipitate in procyclic cells, either directly or indirectly.

DISCUSSION

The control of cell differentiation in *T. brucei* is highly regulated, involving a series of developmental steps that are temporally coordinated. This ensures the generation of a viable cell upon differentiation from the bloodstream to the very different environment of the tsetse midgut. The control of these cellular

245

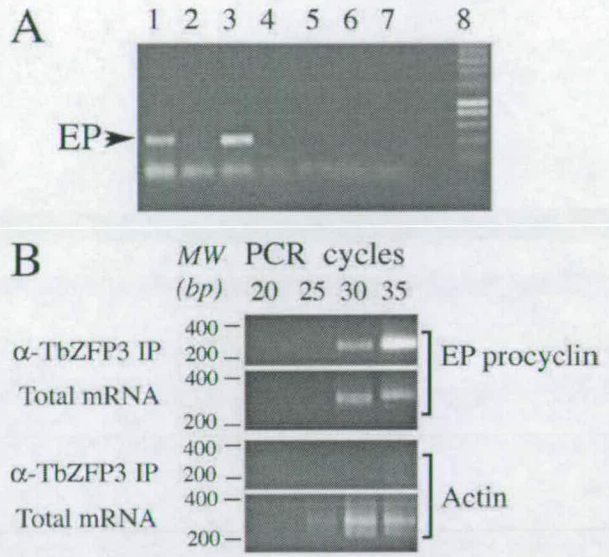

FIGURE 11. **Immunoprecipitation of EP procyclin transcripts with TbZFP3.** *A*, EP procyclin mRNA is selected by immunoprecipitation of TbZFP3. *Lane 1*, EP procyclin cDNA amplification after immunoprecipitation with TbZFP3-specific antibody; *lane 2*, as *lane 1*, but with immunoprecipitation being blocked by the *Tb*ZFP3 peptide immunogen; *lane 3*, EP procyclin cDNA amplified from reverse transcribed procyclic form total RNA; *lane 4*, as *lane 3* but without reverse transcriptase; *lane 5*, as *lane 3* but without inclusion of procyclic form mRNA; *lane 6*, *T. brucei* genomic DNA in place of procyclic form total RNA; *lane 7*, water in place of procyclic form mRNA; *lane 8*, marker. *B*, immunoprecipitation of TbZFP3 differentially selects EP procyclin mRNA with respect to a constitutively expressed transcript. cDNA was derived after RNA immunoprecipitation with anti-TbZFP3 antibody, or derived from unselected total RNA. In each case, an equivalent amount of cDNA was amplified with primers specific for EP procyclin or the constitutively expressed transcript actin in combination with an oligo(dT) anchor, to amplify the 3′-end of each transcript. All amplification reactions were carried out simultaneously with the same source material being used for each. Reaction products were withdrawn every 5 cycles for 35 cycles. EP procyclin mRNA was selected in the immunoprecipitation, whereas actin mRNA was not. In total RNA both transcripts were amplified, with actin mRNA being detected before EP procyclin.

events requires changes in gene expression, this effected by unknown regulatory factors that must operate at the post-transcriptional level. In this report we demonstrate that a new member of a previously characterized family of unusual small RNA-binding proteins is able to enhance differentiation when overexpressed. Moreover, we demonstrate that *Tb*ZFP3 exhibits developmentally regulated co-fractionation with polyribosomes. To our knowledge this is the first regulatory protein identified in these organisms that controls differentiation capacity and shows life-cycle stage-specific phenotypes via involvement at a specific step in the gene expression pathway, the translational apparatus.

*Tb*ZFP3 was identified on the basis of its possession of a CCCH zinc finger domain. This motif is characterized by its interaction with RNA in a number of regulatory proteins in eukaryotic cells, including trypanosomes (24). The best characterized of these is the TTP family of mRNA regulators, which destabilize specific cytokine transcripts important in regulating immune responses. In TTP knock-out mice, dysregulation of tumor necrosis factor-α and granulocyte macrophage-colony stimulating factor in macrophages is observed as are changes in the abundance of at least two interleukin transcripts (reviewed in Ref. 34). These changes are caused by binding of TTP to AU-rich elements in the 3′-untranslated region of target mRNAs, resulting in their destabilization. In addition to TTP family proteins, a number of different classes of CCCH-containing molecule with a diversity of functions have subsequently been characterized in eukaryotic cells. These include the regulation of flower development by HUA1 (35) and FES1 (36), proteins that protect against viral infection by destabilization of retroposon transcripts (37), components of the mRNA processing machinery (U2AF, CPSF30, suppressor of sable), and a number of developmental regulators in *C. elegans*, which function via temporal and spatial regulation of maternal mRNAs at several levels (transcription, mRNA processing, RNA localization, and translation) (19, 38, 39).

The binding of CCCH-containing proteins to RNA is normally dependent upon the presence of two tandemly arranged CCCH zinc finger motifs (40, 41), these being at precisely conserved spacing. In contrast to this organization, *Tb*ZFP3, like the previously identified *Tb*ZFP1 and *Tb*ZFP2, possesses only a single CCCH domain. Although one CCCH domain is sufficient for RNA binding *in vitro* (27),[4] we investigated whether *Tb*ZFP1, -2, or -3 could interact to generate a multicomponent complex (possibly an mRNP complex) possessing more than one CCCH finger. Initially, yeast two-hybrid studies confirmed the ability of *Tb*ZFP2 and -3 to interact with *Tb*ZFP1, this being mediated via the WW domain in each protein. This matches the findings of a related study with an analogous protein identified in *T. cruzi* (29), although in contrast to that study we did not find evidence that *Tb*ZFP2 or *Tb*ZFP3 could homodimerize. More importantly, when we investigated the association of *Tb*ZFP3 with *Tb*ZFP1 and *Tb*ZFP2 *in vivo* via co-immunoprecipitation we found evidence for direct or indirect interaction between each protein, providing evidence that these proteins can co-associate in the cell.

Analysis of the polysomal distribution of *Tb*ZFP3 revealed that this protein (and *Tb*ZFP2)[4] showed stage-enriched association with the translational apparatus. Based on this surprising observation, we propose that *Tb*ZFP3 operates as a trans-acting regulator to translationally activate (or repress, if the protein acts on a negative regulator) specific mRNA subsets in a developmentally regulated manner. This scenario is compatible with the phenotypes observed in both procyclic and bloodstream forms when *Tb*ZFP3 is ectopically overexpressed. Thus, in procyclic forms, posterior extension to generate nozzle forms represents an exaggerated manifestation of the normal morphological changes that occur during differentiation. In bloodstream forms, *Tb*ZFP3 ectopic expression would have no consequence until differentiation was initiated, and the protein associates with the translational apparatus. As in established procyclic forms, this would then result in the robust expression of procyclic characteristics (*e.g.* procyclin expression and kinetoplast positioning), this being manifested as more efficient differentiation. Therefore, direct or indirect regulation of differentiation markers such as EP procyclin is possible, and indeed we found that TbZFP3 is able to co-immunoprecipitate EP procyclin transcripts. This does not indicate direct binding, although *Tc*ZFP1 from *T. cruzi* (related to *Tb*ZFP1 in *T. brucei*)

[4] P. Walrad and K. Matthews, unpublished observations.

246

can directly interact with a procyclin regulatory element *in vitro* (24). Consistent with the association of *Tb*ZFP3 with polysomes only after the initiation of differentiation to procyclic forms, we observed no pre-expression of mRNA or protein for EP procyclin or the procyclic-specific cytoskeletal protein CAP 5.5 upon *Tb*ZFP3 ectopic expression in bloodstream forms.[5]

Several RNA-binding proteins regulate developmental pathways via translational control. Examples include the *Drosophila* Bruno protein that interacts with the Oskar embryonic mRNA and silences its translation via sequestration into silencing particles (42). Similarly, in *C. elegans* the CCCH protein POS-1 controls embryonic development via translational regulation of the maternal mRNA glp-1 (39). In another case, PIE-1 regulates the association of mRNAs with stress granules, such that they are repressed (38). The paucity of transcriptional control, and the polycistronic organization of the genome, make post-transcriptional mechanisms the most important level for gene regulation in trypanosomes. To date, most attention has focused on the role of differential mRNA stability for stage-regulated transcripts. However, it is increasingly clear that translational control is also a major component of regulated gene expression in trypanosomatids. This matches the emerging picture for genes in other organisms in which developmental processes are controlled via pre-existing mRNAs being held in a translationally inactive state (34, 43) often by sequestration into ribonuclear granules (P bodies).

*Tb*ZFP3, like *Tb*ZFP2, is constitutively expressed. However, its association with other RNA-binding proteins may modify its activity or substrate specificity. One of these modulators may be *Tb*ZFP1. *Tb*ZFP1 is synthesized only after the initiation of differentiation between bloodstream and procyclic forms such that association with *Tb*ZFP2, or *Tb*ZFP3 (or a complex containing both proteins), may alter their target specificity. This proposed modularization of function may allow these proteins to control the expression of distinct cohorts of genes in different life-cycle stages, or at stages through development, in a form of post-transcriptional operonic regulation (44). A general perturbation of protein translation induced by *Tb*ZFP3 expression is ruled out by the observation that P0 does not redistribute into monosomes upon ectopic *Tb*ZFP3 expression. Furthermore, general translational inhibition in procyclic forms by hygromycin does not induce comparable nozzle formation.[5]

The *Tb*ZFP family of RNA-binding proteins are present in kinetoplastid parasites but not in other eukaryotes for which there is complete genome sequence information. Indeed, in kinetoplastids there is over five times the number of CCCH proteins as in yeast. It is likely that the unique reliance upon post-transcriptional control by the trypanosome genome necessitates this plethora of novel RNA-binding proteins, which permit stringent and rapid control of gene expression during the parasite life cycle. Although many of the core components of the gene expression machinery in these parasites are conserved with their hosts and vectors, it is those components that are not present in other organisms that provide the key to understanding the unique complexity of gene regulation in these organisms. The *Tb*ZFP family of CCCH proteins provides an excellent route into dissecting the molecular interactions required for coordinating a developmental pathway almost completely dependent on the post-transcriptional control of gene expression.

[5] A. Paterou, P. Walrad, P. Craddy, K. Fenn, and K. Matthews, unpublished observations.

Acknowledgments—We thank Jay Bangs and Steven Reed for the gifts of antibodies to BiP and T. cruzi P0, respectively. The L40 and AMR70 yeast strains and pSTT91 yeast vector were provided courtesy of R. Sternglanz, Center for Yeast Genetics, Stony Brook University, New York, NY.

REFERENCES

1. Moore, M. J. (2005) *Science* **309,** 1514–1518
2. Clayton, C. E. (2002) *EMBO J.* **21,** 1881–1888
3. Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville S. E., and El-Sayed, N. M. (2005) *Science* **309,** 416–422
4. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley S. M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J. (2005) *Science* **309,** 436–442
5. Rudenko, G., Bishop, D., Gottesdiener, K., and Van der Ploeg, L. H. T. (1989) *EMBO J.* **8,** 4259–4263
6. Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Lee, P. I., and Lee, M. G. (2003) *Eukaryot. Cell* **2,** 542–551
7. Matthews, K. R. (2005) *J. Cell Sci.* **118,** 283–290
8. Matthews, K. R., Sherwin, T., and Gull, K. (1995) *J. Cell Sci.* **108,** 2231–2239
9. Priest, J. W., and Hajduk, S. L. (1994) *J. Bioenerg. Biomembr.* **26,** 179–191
10. Tielens, A. G. M., and VanHellemond, J. J. (1998) *Parasitol. Today* **14,** 265–271
11. van Weelden, S. W., Fast, B., Vogt, A., van der Meer, P., Saas, J., van Hellemond, J. J., and Tielens, A. G. (2003) *J. Biol. Chem.* **278,** 12854–12863

247

12. Hammarton, T. C., Clark, J., Douglas, F., Boshart, M., and Mottram, J. C. (2003) *J. Biol. Chem.* **278,** 22877–22886
13. Tu, X., Mancuso, J., Cande, W. Z., and Wang, C. C. (2005) *J. Cell Sci.* **118,** 4353–4364
14. Tu, X., and Wang, C. C. (2005) *Mol. Biol. Cell* **16,** 97–105
15. Ziegelbauer, K., Quinten, M., Schwarz, H., Pearson, T. W., and Overath, P. (1990) *Eur. J. Biochem.* **192,** 373–378
16. Hendriks, E. F., Robinson, D. R., Hinkins, M., and Matthews, K. R. (2001) *EMBO J.* **20,** 6700–6711
17. Hendriks, E. F., and Matthews, K. R. (2005) *Mol. Microbiol.* **57,** 706–716
18. Blackshear, P. J. (2002) *Biochem. Soc. Trans.* **30,** 945–952
19. Reese, K. J., Dunn, M. A., Waddle, J. A., and Seydoux, G. (2000) *Mol. Cell* **6,** 445–455
20. Tacahashi, Y., Helmling, S., and Moore, C. L. (2003) *Nucleic Acids Res.* **31,** 1744–1752
21. Wentz-Hunter, K., and Potashkin, J. (1995) *Nucleic Acids Symp. Ser.* **33,** 226–228
22. Tabara, H., Hill, R. J., Mello, C. C., Priess, J. R., and Kohara, Y. (1999) *Development* **126,** 1–11
23. Guedes, S., and Priess, J. (1997) *Development* **124,** 731–739
24. Morking, P. A., Dallagiovanna, B. M., Foti, L., Garat, B., Picchi, G. F., Umaki, A. C., Probst, C. M., Krieger, M. A., Goldenberg, S., and Fragoso, S. P. (2004) *Biochem. Biophys. Res. Commun.* **319,** 169–177
25. Sutton, A., Heller, R. C., Landry, J., Choy, J. S., Sirko, A., and Sternglanz, R. (2001) *Mol. Cell Biol.* **21,** 3514–3522
26. Hollenberg, S. M., Sternglanz, R., Cheng, P. F., and Weintraub, H. (1995) *Mol. Cell Biol.* **15,** 3813–3822
27. Gietz, D., St Jean, A., Woods, R. A., and Schiestl, R. H. (1992) *Nucleic Acids Res.* **20,** 1425
28. Breeden, L., and Nasmyth, K. (1985) *Cold Spring Harbor Symp. Quant. Biol.* **50,** 643–650
29. Caro, F., Bercovich, N., Atorrasagasti, C., Levin, M. J., and Vazquez, M. P. (2005) *Biochem. Biophys. Res. Commun.* **333,** 1017–1025
30. Biebinger, S., Wirtz, L. E., Lorenz, P., and Clayton, C. (1997) *Mol. Biochem. Parasitol.* **85,** 99–112
31. Hammarton, T. C., Engstler, M., and Mottram, J. C. (2004) *J. Biol. Chem.* **279,** 24757–24764
32. Brecht, M., and Parsons, M. (1998) *Mol. Biochem. Parasitol.* **97,** 189–198
33. van Deursen, F. J., Shahi, S. K., Turner, C. M., Hartmann, C., Guerra-Giraldez, C., Matthews, K. R., and Clayton, C. E. (2001) *Mol. Biochem. Parasitol* **112,** 163–171
34. Barreau, C., Paillard, L., and Osborne, H. B. (2005) *Nucleic Acids Res.* **33,** 7138–7150
35. Chen, X., and Meyerowitz, E. M. (1999) *Mol. Cell* **3,** 349–360
36. Schmitz, R. J., Hong, L., Michaels, S., and Amasino, R. M. (2005) *Development* **132,** 5471–5478
37. Gao, G., Guo, X., and Goff, S. P. (2002) *Science* **297,** 1703–1706
38. Tenenhaus, C., Subramaniam, K., Dunn, M. A., and Seydoux, G. (2001) *Genes Dev.* **15,** 1031–1040
39. Ogura, K., Kishimoto, N., Mitani, S., Gengyo-Ando, K., and Kohara, Y. (2003) *Development* **130,** 2495–2503
40. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A., and Blackshear, P. J. (2000) *J. Biol. Chem.* **275,** 17827–17837
41. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2004) *Nat. Struct. Mol. Biol.* **11,** 257–264
42. Chekulaeva, M., Hentze, M. W., and Ephrussi, A. (2006) *Cell* **124,** 521–533
43. Wickens, M., and Goldstrohm, A. (2003) *Science* **300,** 753–755
44. Keene, J. D., and Tenenbaum, S. A. (2002) *Mol. Cell* **9,** 1161–1167

247

5312–5324 Nucleic Acids Research, 2006, Vol. 34, No. 18 Published online 29 September 2006
doi:10.1093/nar/gkl598

Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements

Matthew Mayho[1,2], Katelyn Fenn[1], Paul Craddy[1], Susan Crosthwaite[2] and Keith Matthews[1,*]

[1]Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK and [2]Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK

Received June 14, 2006; Revised July 24, 2006; Accepted August 2, 2006

ABSTRACT

Trypanosomes represent an excellent model for the post-transcriptional regulation of gene expression because their genome is organized into polycistronic transcription units. However, few signals governing developmental stage-specific expression have been identified, with there being no compelling evidence for widespread conservation of regulatory motifs. As a tool to search for common regulatory sequences we have used the nuclear-encoded components of the cytochrome oxidase (COX) complex of the trypanosome respiratory chain. Components of this complex represent a form of post-transcriptional operon because trypanosome mitochondrial activity is unusual in being developmentally programmed. By genome analysis we identified the genes for seven components of the COX complex. Each mRNA exhibits bloodstream stage-specific instability, which is not mediated by the RNA silencing pathway but which is alleviated by cycloheximide. Reporter assays have identified regulatory regions within the 3′-untranslated regions of three *COX* mRNAs operating principally at the translational level, but also via mRNA stability. Interrogation of the mapped regions via oligonucleotide frequency scoring provides evidence for genome-wide conservation of regulatory sequences among a large cohort of procyclic-enriched transcripts. Analysis of the co-regulated subunits of a stage-specific enzyme is therefore a novel approach to uncover cryptic regulatory sequences controlling gene expression at the post-transcriptional level.

INTRODUCTION

The analysis of cohorts of mRNAs involved in functionally co-ordinated processes is providing a useful tool for the identification of regulatory signals controlling gene expression in eukaryotes (1). Of particular value has been the characterization of the regulatory signals and RNA binding proteins that control nuclear-encoded genes responsible for mitochondrial function. For example, in *Saccharomyces cerevisiae*, mitochondrial function is regulated dynamically in response to a shift between a fermentable and non-fermentable carbon source, oxygen and haem. This has enabled the experimental analysis (2) and computational prediction (3,4) of cryptic regulatory motifs governing expression.

The requirement to identify control elements in mRNAs which are subject to regulated expression is particularly acute in the parasite *Trypanosoma brucei*. These organisms, the causative agents of human and animal African trypanosomiasis, are evolutionarily divergent protozoa that are transmitted between mammalian hosts by the tsetse fly (*Glossina* spp) (5). The different environments that the parasite encounters during its complex life-cycle represent extreme challenges that the cell must overcome, this necessitating regulation of such fundamental processes as metabolism, cell morphology and cell cycle control (6). Unusually, however, these processes are not controlled by regulated transcription. This is because the genome is organized into polycistronic transcription units in which co-transcribed genes are not necessarily co-regulated (7). This genome organization requires resolution of polycistronic pre-mRNAs via the RNA processing reactions of *trans*-splicing and polyadenylation, generating mRNAs with a 39 nt 5′ spliced leader sequence and 3′ poly(A) tail. Interestingly, *trans*-splicing and polyadenylation appear to be mechanistically coupled, with the polyadenylation of the upstream gene in an array being co-ordinated with *trans*-splicing for the downstream gene (8,9). This linkage of RNA processing

*To whom correspondence should be addressed. Tel: +44 131 651 3639; Fax: +44 131 651 3670; Email: keith.matthews@ed.ac.uk

events for adjacent genes which can display differential mRNA abundance make it unlikely that RNA processing efficiency is a major contributor to the control of gene expression. Instead, the emphasis of trypanosome gene regulation is at the level of mRNA stability, although the importance of translational control is becoming increasingly clear.

In several cases the regulation of mRNA stability has been examined for genes that show differential expression during the life-cycle. The best-characterized genes are those of the major surface antigens of the bloodstream and procyclic form. In bloodstream forms, the variant surface glycoprotein (VSG) mRNA is specifically stabilized through signals in its 3′-untranslated region (3′-UTR), this also mediating destabilization upon transformation to procyclic forms *in vitro* (10,11). More detailed characterization has been carried out for the insect stage mRNAs encoding EP and GPEET procyclin. In both cases regulatory motifs can be identified in their 3′-UTR which regulate both the stability of each transcript and also their translational competence (12). For example, EP procyclin contains both 16mer and 26mer elements that contribute to mRNA stability and translation efficiency (13–16). For GPEET procyclin an element has also been identified in the 3′-UTR that confers regulation in response to glycerol and hypoxia (17). Although well characterized, VSG, EP and GPEET procyclin genes are all unusual in being transcribed by RNA polymerase I (pol I) (18) and their expression is very stringently regulated in response to life-cycle differentiation signals (19). This contrasts with the majority of regulated trypanosome genes, which are transcribed by RNA polymerase II (pol II). Although evidence for mRNA regulation of several stage-regulated pol II transcripts has been obtained, regulatory signals remain cryptic (7). This contrasts with at least some stage-regulated genes in the related kinetoplastids, *Trypanosoma cruzi* and *Leishmania donovani* where common regulatory elements can be identified (20,21).

In an attempt to decipher the complex regulatory information among stage-regulated genes in *T.brucei*, we have focussed on the cytochrome oxidase complex (COX; complex IV of the electron transport chain). Unusually, in trypanosomes the COX complex is developmentally regulated, being absent in bloodstream forms and induced upon transformation to the insect form (22). This reflects the differential biochemistry of bloodstream forms (where glycolysis driven by blood glucose provides sufficient ATP) and procyclic forms (where mitochondrial elaboration is required in the glucose-sparse tsetse midgut) (23). The COX complex is also a good model for the analysis of regulated gene expression because at least 10 components of the complex are nuclear-encoded (24), with several identified *COX* genes being dispersed throughout the *T.brucei* genome (25). In consequence each component must be co-incidentally up-regulated during differentiation in order to construct a functional multi-component enzyme complex. Therefore, analysis and comparison of the regulatory mechanisms governing each component has the potential to uncover hitherto cryptic signals not recognizable among larger groups of stage-regulated genes, or in individual genes.

Here we report the identification and verification of seven nuclear-encoded subunits of the *T.brucei* cytochrome oxidase complex. We demonstrate for each component that the regulation involves differential mRNA stability and, for at least three subunits, translational control mechanisms. Moreover, we show that differential mRNA abundance results from specific destabilization of each transcript in bloodstream forms. By reporter assays in stably transformed parasites we have mapped the signals governing differential gene expression in the 3′-UTR of three of the *COX* genes. This has uncovered the presence in the *COX V* mRNA of a regulatory element related to that found in procyclin transcripts, the so-called 26mer sequence. Moreover, genome-wide computational analysis of predicted 3′-UTRs demonstrates that the core of this motif is significantly overrepresented in procyclic-enriched transcripts. An element common in developmentally regulated transcripts encoding mitochondrial proteins is also identified. This reveals hitherto unexpected conservation of common regulatory signals between pol I and pol II transcribed protein coding genes during the life-cycle of *T.brucei* which may assist bioinformatic prediction of expression profile.

MATERIALS AND METHODS

Trypanosomes

Bloodstream and procyclic form trypanosomes were *T.brucei brucei* s427. For the analysis of Argonaute null mutants, bloodstream forms of *T.brucei* STIB 247 were used. Parasites were grown routinely in HMI-9 at 37°C in 5% CO_2 (bloodstream forms) or at 27°C in SDM-79 (procyclic forms). For parasite transfection, 10–15 μg of linearized DNA was electroporated using a BTX ECM830 electroporator. Parasites were selected using 0.5 μg/ml phleomycin (CAT reporter constructs; into bloodstream forms).

Northern blotting and signal quantification; transcriptional and translational inhibition

RNA was extracted from ~5×10^7 to 1×10^8 trypanosomes using an RNAeasy RNA purification kit (Qiagen). Approximately 3 μg of purified RNA was resolved on formaldehyde-agarose gels and transferred to nylon membrane by capillary blotting. For transcript detection, digoxygenin-labelled riboprobes were used, signal detection being achieved via a Bio-Rad Fluor-S imager. For each probe used for quantitation, the linearity of signal was verified over the detection range, generating R^2 values between 0.94 and 0.99. To accurately quantify the respective levels of *COX* transcripts in bloodstream and procyclic forms, serial dilutions of RNA from each life-cycle stage were hybridized with each *COX*-specific riboprobe, thereby determining the relative expression level, and confirming linearity of detection at least over a log range of signal. Loading was normalized to rRNA levels in each lane, this being determined also via a Bio-Rad Fluor-S imager. Decay measurements for mRNAs were carried out after treatment of either bloodstream or procyclic form cultures with 5 μg/ml actinomycin D, samples being taken at 0, 30, 60 and 90 min (for bloodstream forms) or 0, 120, 240 and 480 min (for procyclic forms), these time points being established after preliminary analysis of the decay kinetics in each life-cycle stage. For cycloheximide treatment, samples were incubated with 50 μg/ml

cycloheximide for 0, 60, 120 and 240 min. In both cases RNA from parallel untreated samples was also prepared and analysed.

Analysis of mRNA processing sites

To map sites of polyadenylation RT–PCR was used, with 3′ ends being amplified for each transcript via a gene-specific primer hybridizing to the 5′ end of the coding region and a generic oligo-dT ADAPT 3′ primer recognizing the poly(A) tail (5′ GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TT 3′). Amplified products were then subjected to a second round of amplification using an oligonucleotide recognizing the 3′ end of each *COX* coding region and the primer AUAP (5′ GGC CAC GCG TCG ACT AGT AC 3′), which binds to the specific oligonucleotide sequence incorporated into the 5′ end of the ADAPT primer. The resulting products were gel purified and then sequenced to determine the site of polyadenylation in each life-cycle stage for each gene. For 5′ end mapping, a spliced leader specific primer was used in combination with a primer binding in the 5′ end of each *COX* gene.

Construct generation and CAT assays

Reporter constructs were based on the *T.brucei* expression vector pHD449, with the tetracycline resistance cassette being replaced by the coding region for choramphenicol acetyl transferase (CAT). This was achieved by digesting pHD449 with HindIII and BamHI to excise the TETR gene, this being replaced by a CAT gene PCR amplicon provided with a 5′HindIII site and 3′ BamHI site. To insert each *COX* gene 3′-UTR, each intact *COX* intergenic region was amplified using primers binding immediately after the *COX* gene stop coding and immediately upstream of the ATG codon of the downstream gene. An exception to this was *COX V* where the downstream ('hypothetical unlikely') gene was 1.4 kb downstream and the next open reading frame was over 2 kb from the *COX V* stop codon. In this case, a 3′ primer-binding site was chosen ~30 bp downstream of the last of three prominent polypyrimidine tracts in the intergenic region. Each 5′ and 3′ primer had, respectively, BamHI and Bbs1 restriction sites incorporated to allow cloning into CAT449. The resulting PCR products were digested with BamHI and Bbs1 and these inserted into BamHI/Bbs1 digested CAT 449 to allow replacement of the endogenous truncated aldolase 3′-UTR in the vector. To insert regulatory regions in to intact CAT449 for analysis of function out of context, CAT 449 was digested with BamHI and each *COX* regulatory region, amplified with primers containing BamHI at each end, inserted. The resulting constructs were sequenced to ensure fidelity of insertion and to verify insertion copy number and orientation.

Levels of CAT protein derived from each reporter construct were determined by CAT ELISA assay (Roche) according to the manufacturer's instructions, values being determined on a Dynex technologies MRX II ELISA plate reader. In each case a CAT standard curve was constructed using known concentrations of CAT protein and these used to determine CAT levels in each lysate, this being verified to be in the linear range by lysate dilution over a 1000-fold range. CAT standard curves had a linear regression value typically of 0.998, providing accurate determination of CAT to as low as 0.001 ng/nl.

Oligonucleotide frequency analysis

Oligomer counting was performed using the oligo-analysis tool, a web-based tool forming part of the Regulatory Sequence Analysis Tools (RSAT) suite (http://rsat.scmbb.ulb.ac.be/rsat/). The tool identifies oligomers ranging in length from 3 to 8 nt that are more frequent within the UTR sequence of a group of co-regulated genes compared to a set of non-regulated genes. To identify possible regulatory elements within the 3′-UTR regions of the *COX* components from *T.brucei* and *T.congolense*, counts were performed on either the 3′-UTR sequence up to the experimentally mapped polyadenylation site or, if this site had not been mapped, the first 300 nt from the *COX* gene stop codon. Oligomer frequency counts were performed at various oligomer lengths from 3 to 8 and compared to 'background counts' performed on a database of 3′-UTR sequences from a mixed population of genes considered to be non-directionally regulated. For the background dataset, 300 nt of 3′-UTR sequence were retrieved from every annotated gene on *T.brucei* chromosomes 1 and 2, totalling 883 sequences. The same method was used for analysis of the microarray data with the genes shown to be upregulated in either the procyclic stage or bloodstream stage considered to be co-regulated. Again 300 nt of the 3′-UTR were retrieved for each gene and counted in the analysis. All statistics and significance values were determined according to (3).

RESULTS

Identification of the nuclear-encoded *T.brucei* COX genes

Components of the cytochrome oxidase complex of the kinetoplastid *Crithidia fasciculata*, have been biochemically purified and N-terminal sequences derived for six subunits (26). We used these peptide sequences to search the *T.brucei* genome database (http://www.genedb.org/genedb/tryp/) in order to identify genes encoding putative orthologues in that organism [(27) and this study]. The largest nuclear-encoded subunit (COX IV) has been characterized subsequently from *Leishmania tarentolae* and its homologue in *T.brucei* identified previously (28).

Table 1 shows each of the predicted *T.brucei* proteins aligned with the N-terminal sequence derived from *C.fasciculata* or *L.tarentolae*. Notably alignment of the derived open reading frame for subunits V–X with the determined N-terminal sequence of each from *C.fasciculata* predicts a cleaved leader sequence of between 1 [COX VI, VII; (27)] and 27 amino acids (COX VIII) whereas comparison of the *T.brucei COX IV* gene with the *L.tarentolae* COX IV protein predicts a 44 amino acid leader. The genes for these seven predicted nuclear-encoded cytochrome oxidase components are unlinked in *T.brucei*, being distributed over six distinct chromosomes and each has been annotated onto the *T.brucei* genome database (GeneDB co-ordinates for each gene are given in Table 1).

Table 1. Identification of COX subunits in the genome of *T.brucei*

Predicted leader sequence	Mature N-terminus (actual or predicted)	Organism	Protein	GeneDB identifier	Reference
MFTRRAVSSVVGVTGSAAVVT-SSPLSVQRRY	DHDRWYGHALELDTHNYKFNGEP	*L.tarentolae*	COX IV		(28)
MFARRSLIATVAAATATKPTSSA-AQSNANGTAATQSTLLQQRRY	DHDRWYGHALELDSHNYKFTGEP	*T.brucei*	COX IV	Tb927.1.4100	
	FFGKGWDNASLDTIFSSML	*C.fasciculata*	COX V		(26)
MKRFVTPFLATVPLSRN	FFGKGWDNAALDTIFSCML	*T.brucei*	COX V	Tb09.160.1820	
	PHADHRKYKIQREEMP--PHFSDFNDP RF	*C.fasciculata*	COX VI		(26)
M	PFVDHNKYKIQREDLPALPHFTDFNDPRF	*T.brucei*	COX VI	Tb10.100.0160	
	PFVDHNKYKIQREDLPALPHFTDFNDPRF	*T.brucei*	COX VI-TY		(27)
	PRPFGVWAPATTLAEYRARIPNPFAYSFK-WVYSMKKEIFY	*C.fasciculata*	COX VII		(26)
M	PRPFGVWAPATTLAEYRARIPGMSNFKLR-WVFGARREVYY	*T.brucei*	COX VII	Tb927.3.1410	
	GGDMHSSDRFKAAWDEIPLHM	*C.fasciculata*	COX VIII		(26)
MIRRTAPAVSFTTSHRALMLR TNRPLL	SADMHSLERFKVAWDEMPVH	*T.brucei*	COX VIII	Tb927.4.4620	
	YMLAFNSKAKARPNFGLRGVGYWH-EVY nKPGQsY	*C.fasciculata*	COX IX		(26)
MFSCALRTSRRT	YINAFNAKAKARPNFGLRGVGYWTSEVYH-KPGQNY	*T.brucei*	COX IX	Tb10.6k15.2180	
	LHFPISAPPIEIDYLDNDPLEFAVRTEArKwGF	*C.fasciculata*	COX X		(26)
MLRRAGSRVACACSVPQARS	LHFPITPPPIEIEYLDNDPLEFAVRTEARKWRF	*T.brucei*	COX X	Tb11.01.4702	

In each case the biochemically derived protein sequence of corresponding subunits from either *C.fasciculata* (26) or *Leishmania tarantole* (28) was used as a search tool to identify the *T.brucei* gene. Analysis of the determined N-terminal protein sequence and genomic analysis of the translated gene sequence allowed prediction of the N-terminal leader sequence on each subunit. For *T.brucei COX VI* this has been verified experimentally (27).

To verify the identity of each subunit we used an epitope-tagging approach to visualize their predicted mitochondrial location. Thus, each of the identified genes was amplified from *T.brucei* genomic DNA and inserted into the *T.brucei* expression vector pHD451 (29) modified to incorporate a Ty1 epitope tag (30) at the C-terminus of the encoded transgenic proteins (Figure 1A, schematic representation). Each construct was transiently transfected into procyclic form cells, which were subsequently analysed for expression and localization of the ectopically expressed protein after paraformaldehyde fixation and detergent permeabilization (27). Figure 1A shows cells transfected with tagged *COX IV-COX X* (panels C–H) and reveals, in each case, lattice-like staining characteristic of the procyclic form mitochondrion and matching the distribution of the mitochondrial vital dye, Mitotracker Red (Figure 1A, panel B). This analysis confirmed the predicted mitochondrial localization of each identified subunit, supporting their identification as components of the biochemically characterized kinetoplastid cytochrome oxidase complex.

The COX subunit mRNAs are regulated at the level of stage-specific differential stability

Activity of the cytochrome oxidase complex is developmentally regulated in *T.brucei*, and the nuclear-encoded subunit COX VI has been shown previously to be stage-regulated, its mRNA being enriched in procyclic forms (31). In order to determine if regulation at the RNA level was a general feature of the nuclear-encoded *COX* genes, northern blots of bloodstream and procyclic RNA were hybridized with riboprobes specific for each. Figure 1B demonstrates that although low levels of mRNA are detectable for each subunit in bloodstream forms, there is increased abundance in procyclic forms. Quantitative analysis of the level of several of the *COX* transcripts in bloodstream and procyclic forms revealed a mean differential expression of ~3-fold between bloodstream and procyclic forms (*COX V*, 4-fold; *COX VI*, 2.5-fold; *COX IX*, 2.6-fold).

Because of the importance of post-transcriptional regulation in trypanosome gene expression, the most likely mechanism for regulated mRNA abundance is via developmental changes in mRNA stability. In order to investigate the stability of each *COX* transcript in bloodstream and procyclic forms we used actinomycin D to inhibit transcription and, thereafter, followed the decay of each mRNA in each life-cycle stage by northern blotting. In each case transcript levels were quantitated with respect to the amount of rRNA loaded and after verifying the linearity of detection with each probe. Figure 2 (upper panel) shows a representative northern blot of *COX V* mRNA decay in bloodstream (upper left hand panel) and procyclic forms (upper right hand panel) at time points after transcriptional inhibition with actinomycin D. Analysis of the decay of transcripts in procyclic cells revealed that even after 4 h *COX V* transcripts were still abundant (50% of *COX V* mRNA remained with respect to its abundance before the addition of actinomycin D; Figure 2 lower right panel). This contrasts with bloodstream forms where after only 30 min, *COX V* transcripts were barely detectable (Figure 2, lower left panel; *COX V* mRNA is <10% of the initial abundance after 30 min). The same analysis for each of the other identified *COX* transcripts is shown in Figure 2 (lower panels) with the abundance of each being quantified at time points after actinomycin D addition to either bloodstream or procyclic forms. In each case the *COX* transcripts were relatively unstable in bloodstream forms with respect to procyclic forms, with a mean half-life of ~4 h in procyclic forms whereas in bloodstream cells the half-life was <15 min.

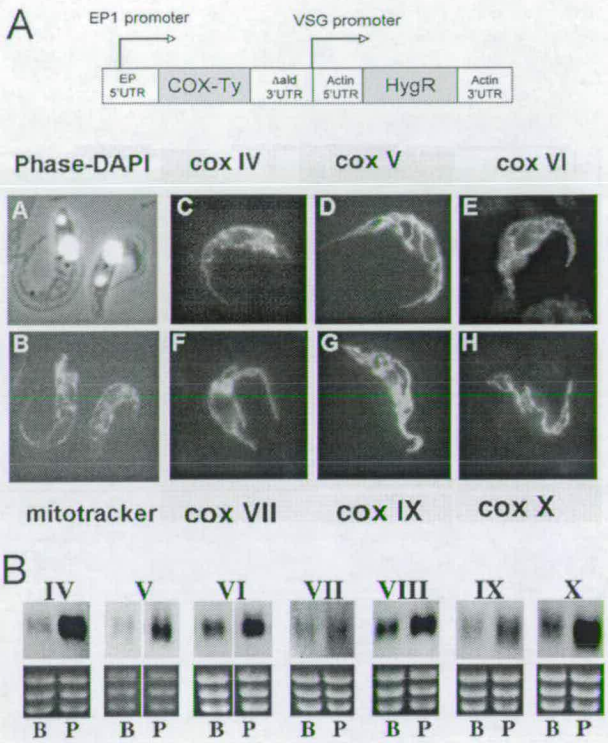

Figure 1. (**A**) Mitochondrial localization of identified COX subunits. For each subunit, the coding region was inserted into a trypanosome expression vector such that the expressed protein included a C-terminal Ty1 epitope tag. Transient transfection of each construct into procyclic forms resulted in expression of an epitope-tagged protein which, in each case, localized to the mitochondrion of the procyclic form (panels C–H). This localization also matched the staining observed with the mitochondrial vital dye, Mitotracker red (panel B). A phase contrast image of the cells in panel B stained with DAPI is shown in panel A, this visualizing the position of the cell nucleus and the mitochondrial genome (kinetoplast). (**B**) Northern blots demonstrating the relative expression of the identified *COX* subunit transcripts in bloodstream (B) or procyclic (P) form trypanosomes. In each case, a section of the ethidium stained gel is shown (containing the rRNAs) to demonstrate relative loading.

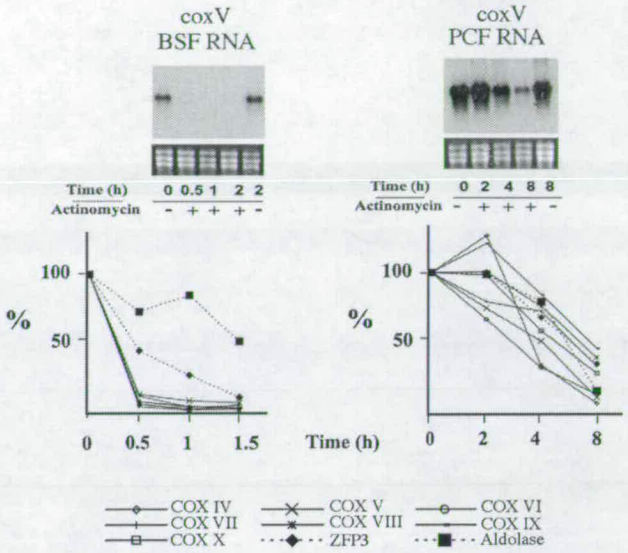

Figure 2. Stability of *COX* subunit mRNAs in bloodstream and procyclic form trypanosomes. The upper panels show representative northern blots of *COX V* transcripts from cells untreated (−) or treated (+) with 5 μg/ml actinomycin D. RNA was isolated at the times shown. The lower panels show a quantitation of all *COX* subunit mRNAs in each life-cycle stage in the presence of actinomycin D, with each data point representing the level of mRNA with respect to its abundance before the addition of actinomycin D. Control transcripts of *TbZFP3* (constitutively expressed) and aldolase (bloodstream enriched) are also shown.

Contrasting with this, analysis of a constitutively expressed control transcript, *TbZFP3* (A. Paterou, P. Walrad, P. Craddy, K. Fenn and K. R. Matthews, manuscript submitted), or bloodstream-specific transcripts such as aldolase, revealed far greater stability in bloodstream forms ($t_{1/2}$ = 40 min or 1.5 h, respectively). Although these transcripts were less stable in bloodstream forms than in procyclic forms, consistent with previous observations (10,16), this instability was far less than the >16-fold difference in half-life observed with each *COX* transcript. Together this analysis demonstrated that each of the *COX* mRNAs is differentially regulated between bloodstream and procyclic forms, with differences in mRNA stability being a significant contributor to this.

To investigate the basis of the differential mRNA stability of *COX* transcripts between bloodstream and procyclic forms we assayed three potential regulatory mechanisms: (i) the use of distinct mRNA processing sites in each life-cycle stage (32,33); (ii) the destabilization of regulated transcripts via the RNA silencing pathway (34); and (iii) control via labile protein regulators. With respect to the differential usage of RNA processing sites in distinct life-cycle stages, RT–PCR was performed on bloodstream and procyclic RNA using primers specific for the 3′ end or 5′ end of the *COX V* and *IX* mRNA. This revealed that although several sites of polyadenylation are used for each gene, these did not differ between different life-cycle stages (Figure 3A). Similarly spliced leader addition sites were identical in bloodstream and procyclic forms (data not shown). To investigate potential involvement of the RNA silencing machinery in regulated control of the *COX* transcripts, the relative abundance of *COX V* and *COX IX* mRNA was assayed in an Argonaute null mutant bloodstream form line and the corresponding wild-type parent (35). Argonaute is a functional homologue of slicer in the trypanosome RNA silencing machinery (36), with null mutants being incapable of RNAi. Figure 3B reveals that no difference in the abundance of either transcript was detected in the Argonaute null mutant when compared to wild-type parents, eliminating a role for the RNAi machinery in stage-regulation of these transcripts.

To determine if *COX* transcripts are actively destabilized in bloodstream cells or actively stabilized in procyclic cells via labile protein factors, bloodstream and procyclic cells were incubated in 50 μg/ml cycloheximide to inhibit *de novo* translation. Under these conditions any labile regulatory proteins controlling *COX* transcripts would be rapidly depleted leading either to a decrease in the abundance of *COX* transcripts in procyclic cells or an increase in bloodstream cells (37–40). The upper panels in Figure 4 show a northern blot of RNA extracted from bloodstream and procyclic cells at various time points after translational inhibition and hybridized to detect *COX V* mRNA. Figure 4

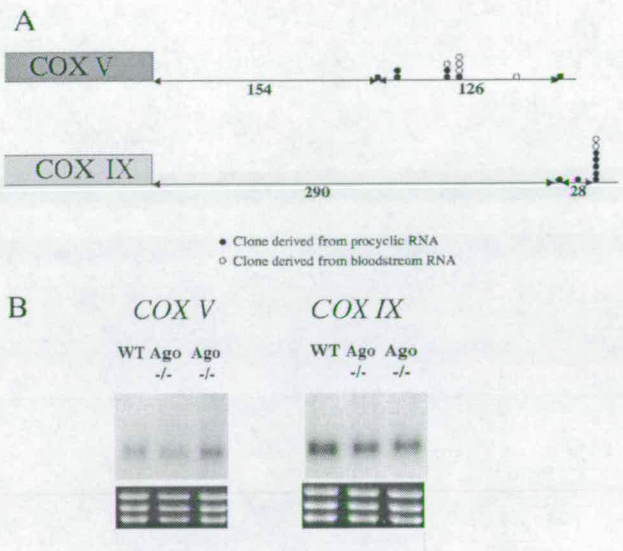

Figure 3. (**A**) Mapping of the site of polyadenylation for two *COX* subunit mRNAs (*COX V* and *IX*) in bloodstream (open circles) or procyclic forms (closed circles). The sites of polyadenylation are distributed, but there is no consistent difference between the sites used in each life-cycle stage. (**B**) mRNA abundance of *COX V* and *COX IX* in wild-type bloodstream forms of *T.brucei brucei* STIB 247 and in two independently derived Ago null mutants of the same strain (35).

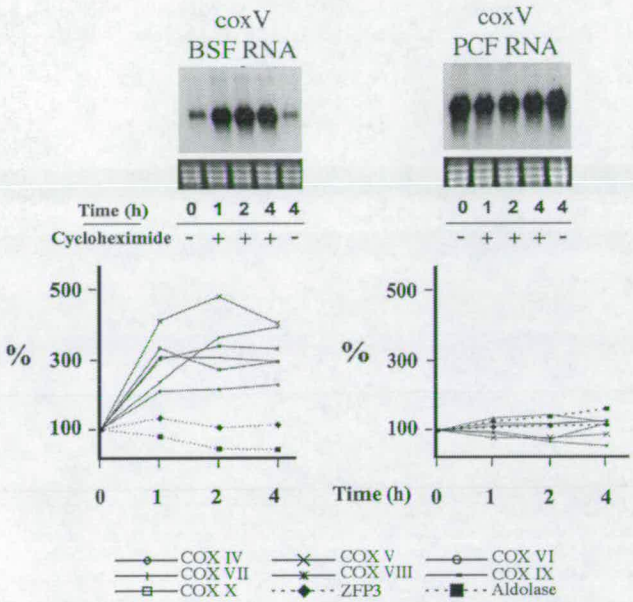

Figure 4. mRNA abundance of *COX* subunit mRNAs after treatment of bloodstream or procyclic form trypanosomes with cycloheximide. The upper panels show representative northern blots of *COX V* transcripts from bloodstream (BSF) or procyclic forms (PCF) untreated (−) or treated (+) with 50 μg/ml cycloheximide. RNA was isolated at the times shown. The lower panels show a quantification of all *COX* subunit mRNAs in each life-cycle stage in the presence of cycloheximide, with each data point representing the level of mRNA with respect to its abundance before the addition of cycloheximide. Control transcripts of *TbZFP3* (constitutively expressed) and aldolase (bloodstream-enriched) are also shown.

(lower panels) show quantification of corresponding northern blots probed for each of the remaining nuclear-encoded *COX* transcripts. This analysis revealed a marked increase in the abundance of all *COX* transcripts tested in bloodstream cells when treated with cycloheximide (3.7-fold, ±0.4 SE), but not in procyclic cells (1.2-fold, ±0.1 SE). In contrast to this, the constitutively expressed transcript TbZFP3 and bloodstream-specific aldolase transcript showed no significantly enhanced abundance in the presence of translational inhibition either in bloodstream or procyclic forms (Figure 4). These experiments establish that *COX* mRNAs are specifically destabilized in bloodstream forms, this being alleviated by translational inhibition.

CAT reporter assays identify elements that repress COX gene expression

In order to dissect the sequences contributing to the bloodstream stage-specific repression of *COX* gene expression, three distinct and arbitrarily chosen *COX* genes were analysed for their regulatory signals. Thus, we created a series of CAT reporter constructs each bearing sequential deletions of the *COX V*, *VI* and *IX* 3′-UTRs (Figure 5A). These constructs were integrated into the *T.brucei* tubulin gene cluster, enabling read-through transcription of the reporter gene via pol II and accessory factors. This ensured that each operated in a transcription unit matching that of the endogenous gene. In each case the CAT reporter gene was provided with either a deleted 3′-UTR of the aldolase gene [which results in constitutive expression; (29)] or with ∼50 nt sequential deletions of the *COX V*, *VI* or *IX* downstream intergenic regions (Supplementary Figure 1 shows the sequence and deletion limits for each 3′-UTR). The deletions were progressively made from the 5′ end of each 3′-UTR toward the mapped region of polyadenylation for each gene, such that all constructs retained the endogenous 3′ end processing site and downstream intergenic sequences extending either to the next gene in the polycistron, or in the case of *COX V* (where the next gene is over 1.5 kb downstream) 164 nt downstream of the major site of polyadenylation (this region contains at least three prominent polypyrimidine tracts). Our strategy was to transfect these constructs into bloodstream cells and assay for increases in CAT production as the 3′-UTRs were progressively deleted, thereby mapping regions responsible for mRNA instability or translational repression in bloodstream forms. Figure 5B shows CAT activity assays from a minimum of two assays performed on protein lysates extracted from at least two independently derived stably transformed clonal cell lines for each construct. These assays revealed specific regions that when deleted cause a marked increase in CAT expression. For *COX V*, an overall 92-fold increase in CAT expression was observed during progressive truncation of the intergenic region, this comprising a 13-fold increase in CAT protein when nt 33–78 from the stop codon (i.e. the *COX V*Δ2 region) were deleted (compare Δ1–33 with Δ1–78) and a further 2.9-fold increase with deletion of nt 119–160 (*COX V*Δ4 region). For *COX IX*, a single deletion positioned between nt 181 and 257 from the stop codon (*COX IX* Δ4) caused a 5-fold increase in CAT activity with respect to the intact

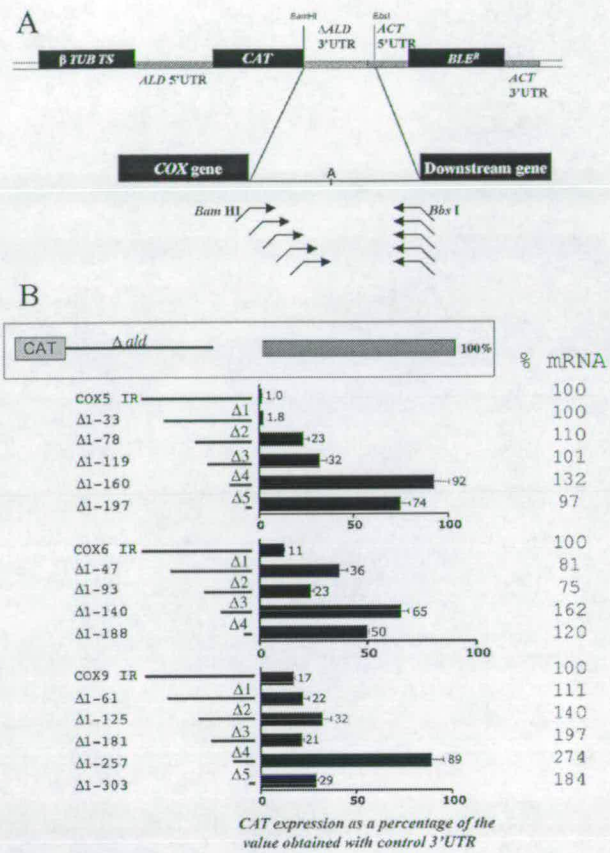

Figure 5. (**A**) Investigation of regulatory signals within *COX V*, *VI* and *IX* 3′-UTRs. The intergenic region for each *COX* subunit was positioned downstream of the CAT gene, this being encompassed within a construct comprising a phleomycin resistance gene and targeted to integrate into the tubulin gene array. The site of polyadenylation is indicated within the intergenic region (denoted 'A'). At least two stable cell lines were generated for each construct and two independent CAT assays performed on cell extracts from each cell line. (**B**) Resulting CAT expression for each *COX* intergenic region deletion are shown as a percentage of the CAT expression from an identical construct comprising a deleted aldolase gene 3′-UTR, which provides constitutive expression in bloodstream and procyclic forms. The CAT mRNA abundance for each construct is shown at the extreme right, as a percentage of the CAT mRNA abundance derived from constructs comprising each intact *COX* subunit intergenic region (IR). These values were derived from northern blots, representatives of which are shown in Figure 6.

intergenic region, although this increase was abolished by further deletion close to the major site of polyadenylation (*COX IX*Δ5). In contrast to *COX V* and *COX IX* 3′-UTRs, deletions of the *COX VI* 3′-UTR sequence did not clearly define particular regions having a major impact on CAT gene expression, although truncation of nt 93–140 increased CAT expression 2.8-fold with respect to the intact intergenic region.

mRNA abundance is not a major regulator of CAT production for COX V and VI

In order to distinguish the relative contribution of mRNA abundance to the observed CAT activity, we used northern blotting to determine the level of CAT mRNA derived from each truncation of the *COX V*, *VI* and *IX* intergenic regions (Figures 5B and 6A). This revealed that the CAT mRNA abundance derived from those constructs furnished with the *COX V* intergenic region differed little, exhibiting only a maximal 32% increase over the intact intergenic region despite a nearly 100-fold increase in CAT protein upon truncation of the 3′-UTR. This indicates that this 3′-UTR conferred regulation almost exclusively at the translational/post-translational level. Similarly, for *COX VI*, the CAT mRNA abundance exhibited a maximal 62% increase over the intact intergenic region (contrasting with a 6-fold protein difference), whereas for *COX IX* CAT mRNA abundance increased by 274% in the *COX IX*Δ4 truncation. This approximates to the differential mRNA abundance for the endogenous *COX IX* transcript between bloodstream and procyclic forms (2.6-fold; Figure 1B). Analysis of the *COX IX* Δ4 deletion further by actinomycin D treatment confirmed that the mRNA stability of the CAT transcript was increased over that of the intact intergenic region, although the rapid turnover of the latter prevented accurate quantitation of the relative increase in stability (Figure 6B). Importantly, RT–PCR analysis (for *COX V*) or CAT transcript sizing (for *COX V, COX VI and COX IX*) for mRNAs derived from all truncations revealed that no alteration of the polyadenylation site occurred with respect to the intact intergenic region for each gene (Figure 6A and data not shown). This eliminated perturbance in the site of 3′ end formation being responsible for the observed changes in CAT expression.

Since little or no increase in the level of CAT mRNA was observed upon deletion of the *COX V* 3′-UTR, we examined whether this intergenic region exhibited elevation of CAT mRNA after exposure to cycloheximide, thereby matching the response of the endogenous *COX V* transcript. Figure 6C demonstrates that this was the case: the intact intergenic region for *COX V* showed a 300% elevation in CAT mRNA after cycloheximide treatment, contrasting with the response of the same construct bearing the aldolase intergenic region (which showed a 0–30% increase after cycloheximide treatment; data not shown). Moreover the response of the *COX V* 3′-UTR to cycloheximide was retained even in the *COX V*Δ4 truncation which lacks all but 50 nt of the *COX V* 3′-UTR, this construct exhibiting a 230% elevation (Figure 6C). These observations implicate sequences in the intergenic region downstream of the polyadenylation site as contributing to the response to cycloheximide. This matches previous observations with constructs containing the procyclin intergenic region, where deletion of the complete 3′-UTR also did not ablate the super-induction of mRNA in response to cycloheximide (41).

To conclude this section, we showed enhanced expression of CAT upon deletion of each *COX* 3′-UTR implicating the presence of negative control elements. For *COX IX* mRNA stability is an important component, whereas for *COX V* and *COX VI* control operates almost exclusively at the translational/post-translational level. This is consistent with analysis of endogenous *COX VI* expression where protein levels are stringently stage-regulated in a mechanism dominant to differential mRNA regulation (27). Combined, this highlights the complexity of regulation among components of even the same enzyme complex.

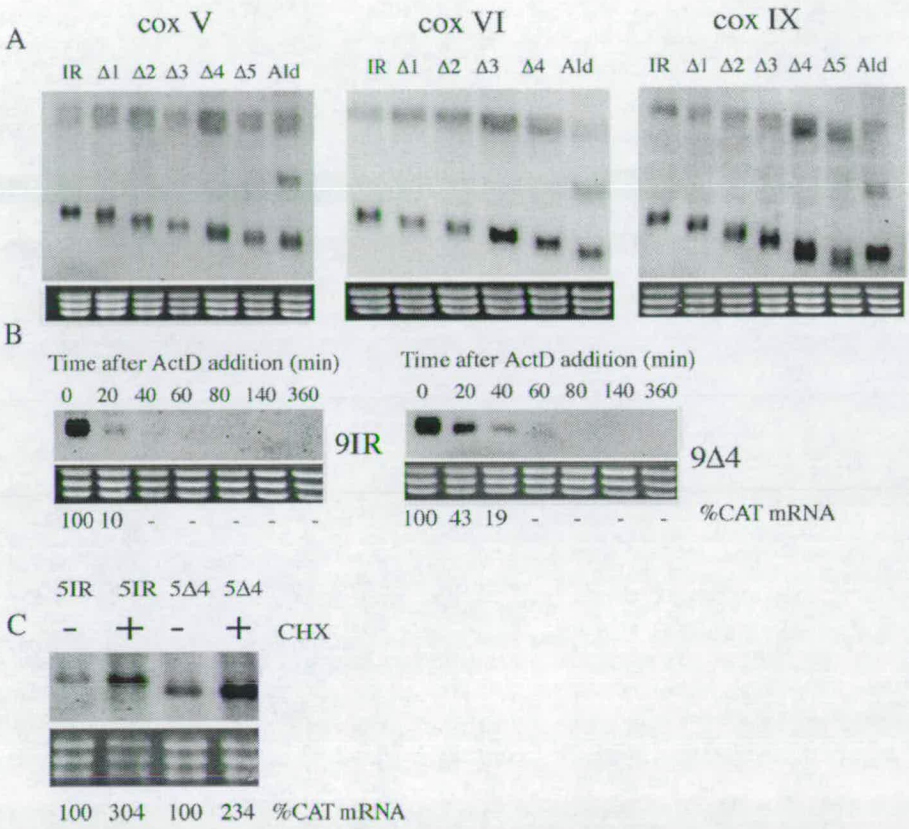

Figure 6. (**A**) Northern blots of CAT mRNA provided with successive deletions of the *COX V*, *COX VI* or *COX IX* intergenic regions. In each case the rRNA region of the ethidium stained gel is provided to demonstrate relative loading. (**B**) Northern blots of CAT mRNA derived from cell lines in which the CAT gene is provided with the *COX IX* intergenic region (9IR) or a deletion (*COX IX* Δ4, '9Δ4') which results in altered mRNA abundance. All cell lines were exposed to actinomycin D at $t = 0$ and RNA isolated at time points thereafter in order to follow transcript decay. The relative abundance of CAT mRNA is denoted beneath each lane. (**C**) Northern blot of CAT mRNA derived from cells transfected with constructs comprising either the intact intergenic region for *COX V* or a deleted derivative (*COX V*Δ4; '5Δ4'). In each case cells were either untreated (−) or treated (+) with cycloheximide, with mRNA being isolated 4 h after the addition of drug. CAT mRNA abundance is denoted beneath each lane.

Identification of shared control elements with the EP procyclin genes

To analyse in more detail those regions identified as contributing to *COX* regulation we focused on the *COX IX* Δ4 and *COX V* Δ2 regions, each of which had a distinct effect on CAT gene expression. Initially we examined the *COX IX*Δ 4 region, which affected both CAT mRNA abundance and stability. To determine if this element was able to operate in isolation, this domain was inserted into the aldolase 3′-UTR immediately after the CAT gene stop codon. In this context, CAT protein expression was observed to be reduced to 50% of the aldolase control, with CAT mRNA being decreased to the same extent (Figure 7A and B). This established that this region could operate to down-regulate mRNA abundance in a context independent manner.

In contrast to the *COX IX* Δ4 region, our reporter assays demonstrated that the *COX V* Δ2 element operated exclusively at the level of translation. Moreover it gave the largest single effect of any of the mapped regions in *COX V*, *VI* or *IX* demonstrating the presence of a strong regulatory element within that domain. This prompted us to analyse this region for recognizable elements that may contribute to this regulatory role. Significantly, this revealed that this region contained a sequence motif similar to that of the 26mer regulatory element identified previously in the EP procyclin 3′-UTR (14,15) (Figure 8A and Supplementary Figure 2). The 26mer element is a region comprising oligoU sequence interrupted by a spacer region which has the potential to form a stem–loop structure containing a U-rich bulge (14). Interestingly, the element in *COX V* Δ2 could be folded into a very similar structure to that predicted for the 26mer in the EP procyclin 3′-UTR, with changes in one side of the stem from the 26mer sequence being matched by compensatory changes in the corresponding base pair partners on the other side of the stem (Figure 8A). Since structural mapping does not predict extensive stable base pair interactions in the EP procyclin 26mer element *in vitro* (42), we also analysed the predicted S-fold structures for both EP procyclin and *COX V* 3′-UTRs (43). This algorithm predicted that each element was in a single stranded bulge. Thus, although different methods assign different predicted structures, the folding of the EP procyclin 26mer and the related *COX* V element were similar in each case.

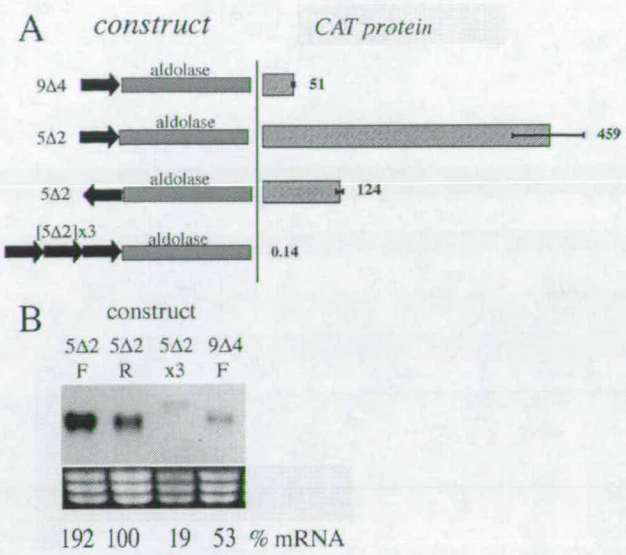

Figure 7. (**A**) CAT activity generated by constructs in which either the *COX IX*Δ4 ('9Δ4') or *COX V*Δ2 ('5Δ2') regions are placed in front of the aldolase 3′-UTR, with 5Δ2 being either in forward or reverse orientation, or in multiple copies. Values are expressed as a percentage of the protein derived from a construct with the aldolase 3′-UTR alone. (**B**) Northern blot of CAT mRNA derived from the cell lines in which either the *COX IX*Δ4 ('9Δ4') or *COX V*Δ2 ('5Δ2') regions are placed in front of the aldolase 3′-UTR, with 5Δ2 being either in forward (F) or reverse (R) orientation, or in multiple copies (x3). mRNA quantification is expressed as a percentage of a cell line with the aldolase 3′-UTR alone, these being normalized to relative loading as determined by rRNA levels.

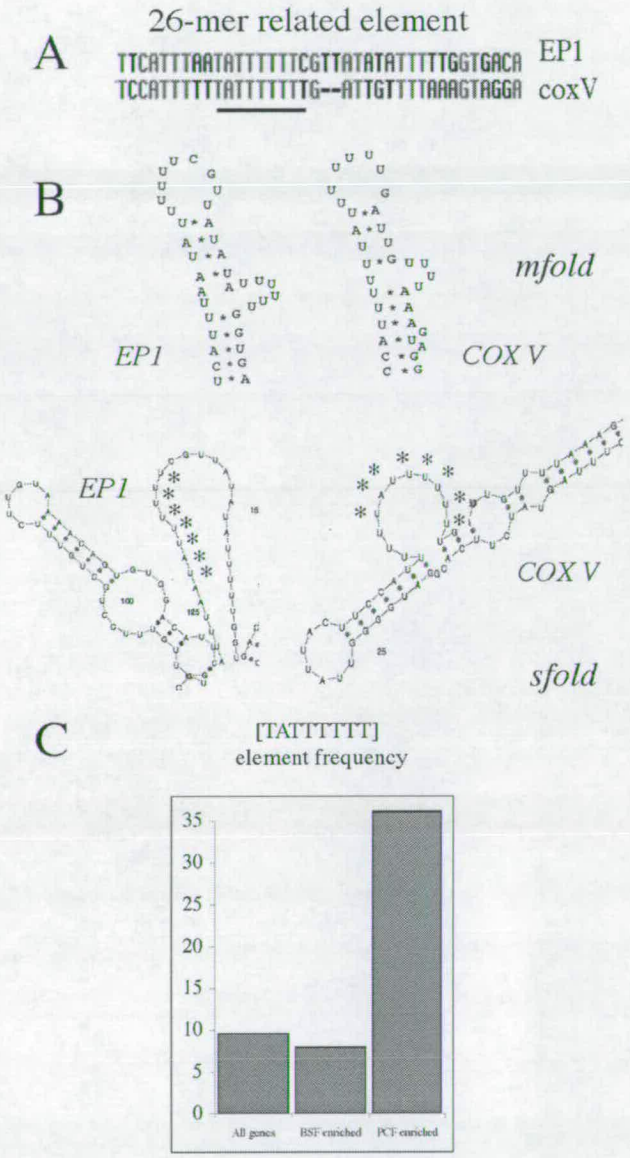

Figure 8. (**A**) Alignment of the EP procyclin 26mer element with the corresponding region in the *COX V* 3′-UTR (the [TATTTTTT] element is underlined). (**B**) Structure prediction for the 26mer core element in EP procyclin and *COX V* 3′-UTRs. In the upper structures, an m-fold prediction (47) for the 26mer element in the procyclin 3′-UTR (14) is used to model the related region in the *COX V* 3′-UTR. The lower structures represent the centroid consensus derived by s-fold structure prediction (43) covering the same region. Asterisks indicate the UAUUUUUU RNA sequence of the core element. (**C**) Frequency (expressed as a percentage of genes within each group) of the 26mer core element [TATTTTTT] in the 300 nt downstream from all genes in the trypanosome genome, transcripts enriched in bloodstream forms (BSF-enriched), or transcripts enriched in procyclic forms (PCF-enriched). Expression data were derived from microarray data kindly provided by Christine Clayton, University of Heidelberg.

In the EP procyclin 3′-UTR, the 26mer operates in a context-specific manner, with distinct effects depending upon the adjacent 3′-UTR or coding sequence (16). To analyse whether the *COX V* Δ2 region exhibited the same characteristics, this region was placed immediately after the CAT stop codon in the aldolase 3′-UTR (Figure 7A and B). In this context the element conferred a 450% increase in CAT protein and 192% increase in CAT mRNA whereas the same sequence in inverted orientation had no significant effect on CAT expression. In contrast to this, when three copies of the *COX V* Δ2 region were inserted, CAT protein was reduced to 0.14% of the intact aldolase 3′-UTR and the CAT mRNA was barely detectable. This was not due to the activation of an alternative, perhaps less efficient, polyadenylation site because the transcript size detected on northern blots matched that of the intact aldolase 3′-UTR taking into account the additional size provided by the inserted element (Figure 7B).

We conclude that the major regulatory element in the *COX V* intergenic region is related structurally and functionally to the 26mer regulatory element mapped in the EP procyclic 3′-UTR. However, it effects on CAT gene expression appear to be strongly context dependent.

Oligonucleotide counting as a tool to identify regulatory sequences in developmentally expressed genes

We sought a bioinformatics approach to search for conserved regulatory sequences either among many procyclic-specific genes, or the more restricted *COX* gene subset. In order to achieve this we employed an approach that searches for oligonucleotides that are statistically overrepresented in a co-regulated gene set when compared to their frequency in

the intergenic region of other genes (3). This approach has been used successfully to identify nuclear-encoded mitochondrial genes in yeast that are regulated in response to catabolite repression (4). In order to derive oligonucleotide frequency tables for a large cohort of trypanosome genes that do not show co-regulation, we downloaded and analysed the first 300 nt after the stop codon of genes predicted on chromosome I and II of the *T.brucei* genome to provide a training set comprising 883 genes. This length of 3′-UTR was chosen as representative of the average length of 3′-UTRs for experimentally characterized trypanosome mRNAs and closely matches the 348 nt median 3′-UTR length determined recently by a predictive algorithm (44). The resulting frequency tables were then screened with the 300 nt downstream of those genes up-regulated in procyclic forms as determined by microarray analysis (kindly provided by Stefanie Brems, Joerg Hoheisel and Christine Clayton, University of Heildelberg and publicly available at http://www.zmbh.uni-heidelberg.de/Clayton/default.shtml) in order to identify sequence elements overrepresented in this group. This identified an overrepresented octamer oligonucleotide set, of which the most statistically significant was the sequence TATTTTTT. Interestingly, this oligonucleotide sequence comprises the core of the 26mer element identified in EP and GPEET procyclin, as well as in PGKB, PPDE (45) and *COX V* (this study). Scanning of the transcripts up-regulated in procyclic forms found that 56/179 (31%) of procyclic-enriched transcripts contained precisely this element (Figure 8C and Supplementary Table 1), whereas only 8% of bloodstream-enriched transcripts (23/287 transcripts) contained the sequence. Moreover, analysis of the frequency of the TATTTTTT element in the 3′-UTR of all genes predicted in the genome (10 765 genes analysed) revealed that only 9.6% of all putative genes contained this sequence (Figure 8C). This emphasizes that when transcripts known to be up-regulated in procyclic forms are considered, there is significant overrepresentation of the 26mer core element.

Not all *COX* 3′-UTRs analysed contained a U-rich element in regions mapped as contributing to gene expression. Moreover, not all deletions of regions containing U-rich elements resulted in changes to gene expression. Therefore, we also applied the oligonucleotide counting approach to search for motifs specifically overrepresented in the *COX* subunit 3′-UTRs. In this case, the trypanosome oligonucleotide frequency dataset was screened using the sequence of the seven 3′-UTRs of the nuclear-encoded *COX* subunits. In addition, the search set was expanded by inclusion of cytochrome oxidase subunits identified in *T.congolense*, a closely related trypanosomatid where respiratory activity is also stage-regulated. All seven cytochrome oxidase subunits were identified by TBLASTX searching of that genome with the *T.brucei* protein sequences isolated in this study. The resulting search identified a consensus sequence UAG (G) UA (G/U) which was present in 6/7 *COX* genes analysed whether derived from *T.brucei* or *T.congolense* (Figure 9A). Interestingly, a copy of this element is in nt 1–47 of the *COX VI* intergenic region (UAGUAGUAG) where deletion results in a 3-fold increase in CAT activity in bloodstream forms and a related sequence UAaGUAUAUA is in the *COX IX*Δ4 nt 181–257 region, whose deletion results in a 4-fold increase in CAT protein and 2.6-fold increase in CAT mRNA

A

Gene	organism	1	2	3	4	5	6	7
COX V	BRUCEI	U	A	G		U	A	U
COX V	BRUCEI	U	A	G		U	A	U
COXVI	BRUCEI	U	A	G		U	A	G
COXVI	BRUCEI	U	A	G		U	A	U
COX VII	BRUCEI	U	A	G		U	A	G
COX VII	BRUCEI	U	A	G		U	A	G
COX VII	BRUCEI	U	A	G	G	U	A	G
COX VIII	BRUCEI	U	A	G		U	A	G
COX VIII	BRUCEI	U	A	G	G	U	A	G
COX IX	BRUCEI	U	A	G	G	U	A	U
COX X	BRUCEI	U	A	G	G	U	A	G
COX X	BRUCEI	U	A	G		U	A	G
COX V	CONGO	U	A	G		U	A	U
COX V	CONGO	U	A	G		U	A	G
COX VI	CONGO	U	A	G		U	A	G
COX VI	CONGO	U	A	G	G	U	A	U
COX VII	CONGO	U	A	G	G	U	A	G
COX VIII	CONGO	U	A	G		U	A	U
COX IX	CONGO	U	A	G		U	A	G
COX IX	CONGO	U	A	G	G	U	A	U
COX X	CONGO	U	A	G	G	U	A	G
COX X	CONGO	U	A	G		U	A	U
CONSENSUS		**U**	**A**	**G**	**(G)**	**U**	**A**	**(G/U)**

B

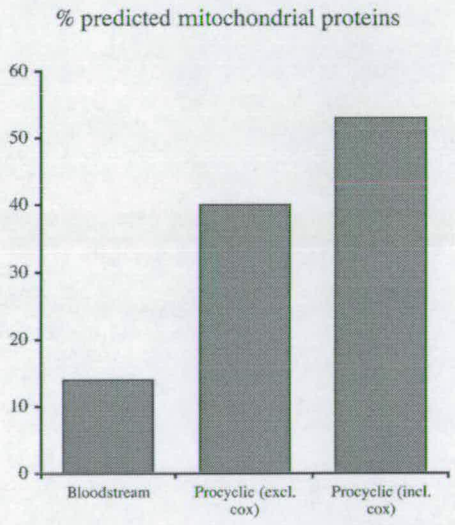

Figure 9. (**A**) The 3′-UTR of *COX* transcripts conserve the oligonucleotide consensus UAG (G) UA (G/U). The 3′-UTR of *COX* subunits identified in *T.brucei* and by bioinformatic interrogation of the incomplete *T.congolense* genome were analysed for overrepresented oligonucleotide sequences. The occurrence of variants of the identified consensus is shown for each transcript, some genes containing multiple representatives. (**B**) The frequency of predicted mitochondrially located proteins containing the conserved oligonucleotide in either bloodstream or procyclic-enriched transcripts. Mitochondrial location was determined by a combination of PSORT II and manual analysis.

abundance. Analysis of transcripts enriched in bloodstream or procyclic forms also indicated that this element was overrepresented in procyclic forms, with 18% (32/179) of transcripts containing the element compared to 9.7% of bloodstream-enriched transcripts (28/287). Moreover, even excluding the *COX* subunits, 40% of transcripts with the element were predicted to be mitochondrial on the basis of PSORTII or manual analysis, this including the respiratory chain components rieske iron–sulphur protein, cytochrome

C1 and a subunit of the F1 ATPase. This contrasts with the bloodstream-enriched transcripts of which only 14% were predicted to encode proteins with a mitochondrial location (Figure 9B).

We conclude that in addition to the 26mer element identified in *COX V*, additional elements can be recognized among the *COX* gene family which may be conserved among a wider subset of procyclic-enriched transcripts, particularly those associated with mitochondrial function.

DISCUSSION

To date, widely conserved regulatory signals governing stage-specific gene expression in trypanosomes have eluded detection, preventing the prediction of gene expression by bioinformatic approaches. Here we have characterized the signals that contribute to the expression of a developmentally regulated multi-subunit enzyme complex in order to recognize conserved sequence elements likely to control stage-specific gene expression. Consistent with previous analyses, this has revealed the importance of both mRNA stability and translational efficiency in the control of developmentally regulated nuclear gene expression. However, one identified element in the cytochrome oxidase subunit V mRNA was closely related in sequence, predicted secondary structure and function to the 26mer element identified previously as important in the control of the procyclic form surface antigens, EP and GPEET procyclin. Moreover, the core of this conserved element was found by an oligonucleotide frequency analysis approach to be conserved in a large cohort of transcripts that are up-regulated in the procyclic form of the life cycle. This predicts that the 26mer core element is an important regulatory element for a large set of stage-regulated pol I and pol II transcribed genes. This facilitates bioinformatic prediction of gene expression profiles in *T.brucei*.

The 26mer element was identified in the EP procyclin gene 3′-UTR regions and demonstrated to have a role in mRNA stability and translational control (14,15). Similarly the PGKB and PPDE 3′-UTRs, which contain the U-rich core of the 26mer sequence contribute to gene expression via stage-regulated mRNA stability (45). In the *COX V* 3′-UTR, however, the 26mer core element demonstrated a role only at the translational level, at least when in the context of its complete 3′-UTR. This is not necessarily surprising: mutational analysis of the 26mer element derived from the EP procyclin gene revealed that very limited mutation resulted in the loss of effect on mRNA stability and yet retained the effects on translation (16). Thus, subtle differences in sequence appear to have quite important consequences in the action of the element. This is further reinforced by analysis of the function of the element when inserted into an unrelated intergenic region downstream from the aldolase gene. In this case, incorporation of the element resulted in a surprising increase in CAT expression, although three copies of the element almost completely abolished CAT mRNA and protein. Similar to this, the EP procyclin 26mer element has also been observed to generate both increases and decreases of CAT gene expression when placed in to an unrelated 3′-UTR. Thus, the 26mer element exhibits context-specific effects on gene expression.

We exploited an oligonucleotide counting approach to identify overrepresented motifs among procyclic-enriched transcripts. This revealed that the core element of the 26mer sequence (UAUUUUUU at the RNA level) was significantly overrepresented among the 3′-UTRs of regulated genes, being present in the 3′-UTR of one-third of procyclic form enriched transcripts, contrasting with 9.6% of all 3′-UTRs and 8% of 3′-UTRs for bloodstream-enriched transcripts. In fact, the frequency of such U-rich elements is even higher since the top five overrepresented octamer oligonucleotide sequences identified in the procyclic-enriched mRNA population were entirely comprised of AU-rich sequence, these being recognizable in 44% of procyclic-enriched transcripts. Interestingly, when the overall AU content of intergenic regions derived from all trypanosome genes, bloodstream-enriched or procyclic-enriched transcripts is compared it is remarkably consistent (59, 60 and 58%, respectively) reinforcing the overrepresentation of the AU-rich core sequence as a discrete element regulating developmental gene expression. In contrast to procyclic-enriched transcripts, the 3′-UTR of bloodstream-enriched transcripts did not contain any clearly overrepresented oligonucleotide sequence. These genes may, therefore, have more cryptic signals or be recognized by their absence of the 26mer element (and, possibly, other contributing sequences). Indeed the presence of the 26mer core element in 8% of these transcripts may represent short mRNAs erroneously included by the arbitrary 300 nt 3′-UTR cut-off used in our analysis.

The regulation of AU-rich sequences in mammalian cells can be governed by a number of RNA binding proteins including HuR. Consistent with this, ectopic expression of human HuR results in the stabilization of AU-rich containing transcripts in *T.brucei* bloodstream forms (45). More recently, AU-rich mRNA destabilization has been found to be effected in mammalian cells by the activity of the RNA silencing pathway, involving microRNAs and components of the RNAi machinery (34). Although there is to date no evidence for microRNA regulation of gene expression in trypanosomes, we investigated whether regulation via RNAi could contribute to the regulated expression of *COX* transcripts. Our analysis revealed that *COX* transcripts were of equivalent abundance in wild-type cells and in cells ablated for the Argonaute component of the RNAi machinery in trypanosomes. This demonstrates that regulated expression of these genes was not effected by the RNA silencing machinery, matching the capacity of Argonaute null mutants to complete differentiation events with normal kinetics (35).

Although the RNAi machinery had no effect on *COX* gene expression, cycloheximide treatment generated a significant and stage-specific increase in the abundance of *COX* mRNA but not constitutively expressed transcripts. As has been proposed previously in the analysis of several regulated genes in *T.brucei*, this may reflect the expression of labile negative regulatory factors which destabilize developmentally regulated mRNAs (39,40). More recently, however it has been recognized that cycloheximide can prevent the trafficking of mRNAs to cytoplasmic ribonuclear granules, which act as sites of mRNA degradation (46). Inhibition of this pathway may therefore provide an alternative explanation for the enhanced abundance of *COX* transcripts in blood-

stream forms upon cycloheximide treatment. Interestingly, and consistent with observations for EP procyclin (41), the cycloheximide mediated enhancement of CAT transcripts harbouring the *COX V* intergenic region was not mediated by sequences contained within the 3′-UTR. This invokes, therefore, either a role for sequence elements in the intergenic region downstream of the polyadenylation site, or a more complex response linked to translation of the gene.

Oligonucleotide counting was originally applied to the identification of signals controlling catabolite repression in yeast, resulting in the recognition of regulated nuclear-encoded mitochondrial mRNAs (4). This identified the consensus CYUGUAA—UA, which was subsequently confirmed to be a generic sequence involved in regulation of nuclear-encoded mitochondrial transcripts by the yeast Puf3 mRNA regulator (recognition sequence UGUR—UA) (2). Our analysis of a subset of nuclear-encoded cytochrome oxidase genes in both *T.brucei* and *T.congolense* revealed that these contained a significantly overrepresented oligonucleotide sequence (UAG [G] UA [G/T]) that is similar to (but distinct from) the yeast element. Interestingly, deletion of some regions containing this element, or close relatives of it, also caused an increase in CAT expression in bloodstream forms suggesting that this element may be a contributor to *COX* gene regulation. Other similar sequences were also present in regions whose deletions did not effect CAT gene expression, however, suggesting context may also be important. Nonetheless, analysis of regulated transcripts annotated via microarray indicated significant overrepresentation of the element in procyclic-enriched mRNAs when compared to bloodstream-enriched transcripts. Moreover, 40% of the non-*COX* transcripts (50% of all transcripts) containing the element were predicted mitochondrial proteins, several of which contained multiple copies of the element (see supplementary Table 2). This indicates this element is likely to be a contributor to stage-regulation of a subset of mRNAs in trypanosomes, particularly those encoding mitochondrial components.

We initiated our study as a route to search for common sequences among a tightly co-regulated gene family, the developmentally regulated components of the cytochrome oxidase complex. The co-regulation of the cohort of stage-regulated nuclear-encoded mitochondrial proteins is an excellent example of a post-transcriptional operon. Our findings illustrate that even within a clearly related and co-ordinately regulated gene set, the signals controlling expression are complex. Nonetheless, by a combination of bioinformatic analysis and detailed experimental validation we have uncovered the general importance of one element, the 26mer and related sequences, among procyclic-specific genes, as well as a more gene-specific motif which may contribute to the regulation of a much more restricted set of mRNAs, including those of the respiratory chain. The identification of these elements provides the necessary tools to isolate the regulatory protein factors that control the expression of this mRNA cohort in a procyclic-specific manner.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are indebted to Professor Christine Clayton, Stephanie Brems and Joerg Hoheisel for permission to use microarray data. We also thank Professor Stephen Hajduk and members of his laboratory for discussions throughout this work. M.M. was supported by a doctoral training account from the BBSRC. Work in Keith Matthews' laboratory is supported by a programme grant from the Wellcome Trust. Funding to pay the Open Access publication charges for this article was provided by a programme grant to KM by the Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Moore,M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
2. Gerber,A.P., Herschlag,D. and Brown,P.O. (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol.*, **2**, E79.
3. van Helden,J., Andre,B. and Collado-Vides,J. (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.*, **281**, 827–842.
4. Jacobs Anderson,J.S. and Parker,R. (2000) Computational identification of *cis*-acting elements affecting post-transcriptional control of gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **28**, 1604–1617.
5. Vickerman,K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.*, **41**, 105–114.
6. Matthews,K.R. (2005) The developmental cell biology of *Trypanosoma brucei*. *J. Cell Sci.*, **118**, 283–290.
7. Clayton,C.E. (2002) Life without transcriptional control? From fly to man and back again *EMBO J.*, **21**, 1881–1888.
8. LeBowitz,J.H., Smith,H.Q., Rusche,L. and Beverley,S.M. (1993) Coupling of poly (A) site selection and *trans*-splicing in Leishmania. *Genes Dev.*, **7**, 996–1007.
9. Matthews,K.R., Tschudi,C. and Ullu,E. (1994) A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.*, **8**, 491–501.
10. Ehlers,B., Czichos,J. and Overath,P. (1987) RNA turnover in *Trypanosoma brucei*. *Mol. Cell. Biol.*, **7**, 1242–1249.
11. Jefferies,D., Tebabi,P. and Pays,E. (1991) Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression at the posttranscriptional level. *Mol. Cell. Biol.*, **11**, 338–343.
12. Roditi,I., Furger,A., Ruepp,S., Schurch,N. and Butikofer,P. (1998) Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **91**, 117–130.
13. Hehl,A., Vassella,E., Braun,R. and Roditi,I. (1994) A conserved stem–loop structure in the 3′ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA*, **91**, 370–374.
14. Furger,A., Schurch,N., Kurath,U. and Roditi,I. (1997) Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol. Cell. Biol.*, **17**, 4372–4380.
15. Hotz,H.R., Hartmann,C., Huober,K., Hug,M. and Clayton,C. (1997) Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3′-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res.*, **25**, 3017–3025.
16. Hotz,H.R., Biebinger,S., Flaspohler,J. and Clayton,C. (1998) PARP gene expression: control at many levels. *Mol. Biochem. Parasitol.*, **91**, 131–143.
17. Vassella,E., Den Abbeele,J.V., Butikofer,P., Renggli,C.K., Furger,A., Brun,R. and Roditi,I. (2000) A major surface glycoprotein of

Trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev.*, **14**, 615–626.
18. Gunzl,A., Bruderer,T., Laufer,G., Schimanski,B., Tu,L.C., Chung,H.M., Lee,P.T. and Lee,M.G. (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell*, **2**, 542–551.
19. Pays,E. (2005) Regulation of antigen gene expression in *Trypanosoma brucei*. *Trends Parasitol.*, **21**, 517–520.
20. Boucher,N., Wu,Y., Dumas,C., Dube,M., Sereno,D., Breton,M. and Papadopoulou,B. (2002) A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3′-untranslated region element. *J. Biol. Chem.*, **277**, 19511–19520.
21. Di Noia,J.M., D'Orso,I., Sanchez,D.O. and Frasch,A.C. (2000) AU-rich elements in the 3′-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J. Biol. Chem.*, **275**, 10218–10227.
22. Priest,J.W. and Hajduk,S.L. (1994) Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. *J. Bioenerg. Biomembr.*, **26**, 179–191.
23. Bringaud,F., Riviere,L. and Coustou,V. (2006) Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol. Biochem. Parasitol.*, **149**, 1–9.
24. Grossman,L.I. and Lomax,M.I. (1997) Nuclear genes for cytochrome *c* oxidase. *Biochim. Biophys. Acta*, **1352**, 174–192.
25. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416–422.
26. Speijer,D., Muijsers,A.O., Dekker,H., Dehaan,A., Breek,C.K.D., Albracht,S.P.J. and Benne,R. (1996) Purification and characterization of cytochrome-c-oxidase from the insect Trypanosomatid *Crithidia fasciculata*. *Mol. Biochem. Parasitol.*, **79**, 47–59.
27. Tasker,M., Timms,M., Hendriks,E. and Matthews,K. (2001) Cytochrome oxidase subunit VI of *Trypanosoma brucei* is imported without a cleaved presequence and is developmentally regulated at both RNA and protein levels. *Mol. Microbiol.*, **39**, 272–285.
28. Maslov,D.A., Zikova,A., Kyselova,I. and Lukes,J. (2002) A putative novel nuclear-encoded subunit of the cytochrome *c* oxidase complex in trypanosomatids. *Mol. Biochem. Parasitol.*, **125**, 113–125.
29. Biebinger,S., Wirtz,L.E., Lorenz,P. and Clayton,C. (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **85**, 99–112.
30. Bastin,P., Bagherzadeh,Z., Matthews,K.R. and Gull,K. (1996) A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **77**, 235–239.
31. Matthews,K. and Gull,K. (1998) Identification of stage specific and differentiation enriched transcripts during transformation of the African trypanosomes from its bloodstream to procyclic form. *Mol. Biochem. Parasitol.*, **95**, 81–95.
32. Erondu,N.E. and Donelson,J.E. (1992) Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. *Mol. Biochem. Parasitol.*, **51**, 111–118.
33. Clement,S.L. and Koslowsky,D.J. (2001) Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in *Trypanosoma brucei*. *Gene*, **272**, 209–218.
34. Jing,Q., Huang,S., Guth,S., Zarubin,T., Motoyama,A., Chen,J., Di Padova,F., Lin,S.C., Gram,H. and Han,J. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell*, **120**, 623–634.
35. Janzen,C.J., van Deursen,F., Shi,H., Cross,G.A., Matthews,K.R. and Ullu,E. (2006) Expression site silencing and life-cycle progression appear normal in Argonaute1-deficient *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **149**, 102–107.
36. Shi,H., Tschudi,C. and Ullu,E. (2006) Functional replacement of *Trypanosoma brucei* Argonaute by the human slicer Argonaute 2. *RNA*, **12**, 943–947.
37. Graham,S.V. and Barry,J.D. (1996) Polysomal, procyclin mRNAs accumulate in bloodstream forms of monomorphic and pleomorphic trypanosomes treated with protein synthesis inhibitors. *Mol. Biochem. Parasitol.*, **80**, 179–191.
38. Schurch,N., Furger,A., Kurath,U. and Roditi,I. (1997) Contributions of the procyclin 3′ untranslated region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **89**, 109–121.
39. Webb,H., Burns,R., Ellis,L., Kimblin,N. and Carrington,M. (2005) Developmentally regulated instability of the GPI-PLC mRNA is dependent on a short-lived protein factor. *Nucleic Acids Res.*, **33**, 1503–1512.
40. Gruszynski,A.E., van Deursen,F.J., Albareda,M.C., Best,A., Chaudhary,K., Cliffe,L.J., Del Rio,L., Dunn,J.D., Ellis,L., Evans,K.J. *et al.* (2006) Regulation of surface coat exchange by differentiating African trypanosomes. *Mol. Biochem. Parasitol.*, **147**, 211–223.
41. Fluck,C., Salomone,J.Y., Kurath,U. and Roditi,I. (2003) Cycloheximide-mediated accumulation of transcripts from a procyclin expression site depends on the intergenic region. *Mol. Biochem. Parasitol.*, **127**, 93–97.
42. Drozdz,M. and Clayton,C. (1999) Structure of a regulatory 3′ untranslated region from *Trypanosoma brucei*. *RNA*, **5**, 1632–1644.
43. Chan,C.Y., Lawrence,C.E. and Ding,Y. (2005) Structure clustering features on the Sfold Web server. *Bioinformatics*, **21**, 3926–3928.
44. Benz,C., Nilsson,D., Andersson,B., Clayton,C. and Guilbride,D.L. (2005) Messenger RNA processing sites in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **143**, 125–134.
45. Quijada,L., Guerra-Giraldez,C., Drozdz,M., Hartmann,C., Irmer,H., Ben-Dov,C., Cristodero,M., Ding,M. and Clayton,C. (2002) Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic Acids Res.*, **30**, 4414–4424.
46. Brengues,M., Teixeira,D. and Parker,R. (2005) Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science*, **310**, 486–489.
47. Zuker,M. and Jacobson,A.B. (1998) Using reliability information to annotate RNA secondary structures. *RNA*, **4**, 669–679.