Investigation of the role of the NLP and TDP1 chromatin associated proteins in transcription control in *Trypanosoma brucei*

Mani Shankar Narayanan

The Queen's College

2011

A thesis submitted for the degree of Doctor of Philosophy

to the

Department of Biochemistry,

University of Oxford

Supervisor: Dr Gloria Rudenko

Abstract

The African trypanosome *Trypanosoma brucei* evades the immune system of the mammalian host by periodically switching its surface coat which is made up of Variant Surface Glycoprotein (VSG). *T. brucei* shows monoallelic expression of one *VSG* out of a repertoire of ~1200 genes with the active *VSG* gene expressed from one of ~15 telomeric expression sites (ESs). The mechanism behind the monoallelic exclusion of ESs is unclear.

NLP was identified as a novel and essential AT-hook protein binding transcriptionally silent simple sequence repeats in *T. brucei*. I depleted NLP using RNAi in various *T. brucei* reporter lines containing an *eGFP* in different transcriptionally silent areas of the genome and monitored derepression of eGFP using flow cytometry. After NLP knock-down, I observed 45-65 fold derepression of silent ESs, and up to 5 fold derepression of other transcriptionally silent areas. Using chromatin immunoprecipitation (ChIP) I found an enrichment of NLP in certain non-transcribed loci including the rDNA spacers. I also found that blocking NLP synthesis results in a rapid fall in levels of the active *VSG* transcript. Lastly, I discovered using tandem affinity purification (TAP) followed by mass spectrometry that NLP is a part of a novel TbISWI complex in *T. brucei*, which also includes two previously unidentified protein partners.

The high mobility group B (HMGB) protein family constitutes a major abundant class of non-histone chromatin associated DNA-binding proteins which play a role in chromatin architecture in a wide range of eukaryotes. In *T. brucei*, the HMGB protein TDP1, which contains two HMG boxes and one DEK C terminal DNA-binding domain, was first identified as binding to *VSG* ES promoter oligomer sequences. I report that TDP1 is an essential nuclear protein enriched in the nucleolus and expression site body, and is involved in facilitating transcription. Blocking TDP1 synthesis using RNAi mediated knock-down results in approximately 40-90% reduction in transcription of RNA polymerase I transcribed genes. Using ChIP, I find that TDP1 is enriched in the rDNA and on the active *VSG* ES in bloodstream form *T. brucei*. Additionally, the relative proportion of TDP1 binding the procyclin promoter compared with the upstream spacer and downstream EP1 genes is greater in procyclic form compared with bloodstream form *T. brucei*. Lastly, I performed TAP experiments with TDP1 and found that TDP1 interacts with the core histones. These results indicate that TDP1 is an architectural chromatin protein important for transcription control in *T. brucei*.

Contents

Abstract 2

Acknowledgements 4

I. General Introduction 5
- a. Introduction to African trypanosomes 6
- b. Chromatin regulation in gene expression 14
- c. Monoallelic expression in other eukaryotes 17
- d. Chromatin regulation in *T. brucei* 21
- e. Regulation of ribosomal DNA 25
- f. Aims of this thesis 28

II. NLP is a novel transcription regulator in *T. brucei* 30
- a. Introduction 31
- b. Results 34
- c. Discussion 55

III. NLP is a member of a novel ISWI complex 61
- a. Introduction 62
- b. Results 68
- c. Discussion 81

IV. TDP1 is an architectural protein important for transcription control in *T. brucei* 88
- a. Introduction 89
- b. Results 92
- c. Discussion 113

V. Investigation of other potential *T. brucei* chromatin regulators 121
- a. Introduction 122
- b. Results 127
- c. Discussion 133

VI. Conclusions 139

VII. Materials and Methods 149

VIII. References 165

IX. Appendices 183
- a. Translation control in *T. brucei* 184
- b. List of primers used for qPCR analyses 197
- c. Published work 200

Acknowledgements

I am very grateful to my supervisor Dr. Gloria Rudenko for allowing me to work in her laboratory and for being a very supportive and positive mentor. Her supervision, scientific acumen and sincere dedication to her work and mine have all gone a long way in shaping me into a better researcher. I am also grateful to Prof. Jane Mellor for being a kind and helpful College Advisor and also for her help in our initial ChIP experiments. Many thanks also to The Queen's College simply for being the best college in Oxford.

I thank all the past and present members of the Rudenko lab, particularly Viola Denninger, Megan Lindsay Povelones, Manish Kushwaha, Nadina Wand and Alexandra Ling for all the support, friendship, and memories. I also owe a big thank you to all my friends, especially Siladitya Chatterjee, Upneet Singh and Vaibhav Rajan for the good times in between work. I'd like to mention a word in honour of my idols Charles Darwin, APJ Abdul Kalam and MS Dhoni, whose lives have been an inspiration to me and have always provided the 'bigger picture'.

A special thanks to my wife Suchismita whom I met in the first week of my D.Phil degree and has been a soul mate and best friend ever since. Thank you also to my sister Yamini for all the helpful career advice over many years. Finally, a very special thanks to my parents Shubha and Ravi whose dedication, stringent work ethic and constant love and belief in me throughout the years have brought me to this moment. I dedicate this thesis to my truly wonderful parents.

This work was funded by the Wellcome Trust, and I was supported by the Clarendon Fund and The Queen's College Graduate Award.

Chapter 1

General Introduction

a. Introduction to African trypanosomes

The African trypanosome *Trypanosoma brucei* is the causative agent of human sleeping sickness and the related disease *nagana* in livestock. Human sleeping sickness is caused by two morphologically identical subspecies of *T. brucei – T. b. gambiense* which occurs in Central and West Africa and causes a chronic form of the disease, and *T. b. rhodesiense* which occurs in Southern and East Africa and causes the acute and potentially fatal form of the disease (W. H. O, 2010). It is estimated that together these diseases affect about 70 million people in sub-Saharan Africa and are responsible for the deaths of 50,000 people a year (Beaglehole et al. 2004). The subspecies *T .b. brucei* is non-infective in humans, being transmitted from a reservoir of mammalian hosts including wild animals and domestic animals, especially cattle in which it causes *nagana* and can be fatal.

Life cycle of T. brucei

T. brucei sp. are obligate protozoan parasites which multiply extracellularly in the bloodstreams of their mammalian hosts and in the gut of their insect vector, which are tsetse flies belonging to the *Glossina* spp (Oberle et al. 2010). The life cycle of *T. brucei* sp. follows a pattern in which they tightly regulate their outermost protein coat which changes between life-cycle stages (Figure. 1.1). The parasite is injected as a metacyclic trypomastigote into the bloodstream of the mammalian host by an infected tsetse fly during a blood meal. These metacyclic cells are proliferative in the bloodstream, and morphologically appear long and slender. Here in the bloodstream, they are protected by a dense layer of variant surface glycoprotein (VSG). As parasitemia increases over the course of the infection, a morphologically stumpy and non-proliferative form of trypomastigotes appears (reviewed in Barry & McCulloch 2001, Matthews et al. 2004, Fenn & Matthews 2007). The stumpy form

is pre-adapted to life in the tsetse fly. When ingested during a blood meal, these stumpy form cells differentiate in the midgut of the insect into procyclic trypomastigotes. These have a procyclin coat, containing EP (Glu-Pro) or GPEET (Gly-Pro-Glu-Glu-Thr) repeats (Bütikofer et al. 1997; Vassella et al. 2003). On reaching a critical stage the parasites differentiate into epimastigotes and migrate to the salivary gland. Here they are covered by an alanine-rich protein called BARP (Urwyler et al. 2007). Finally epimastigotes differentiate back into the metacyclic form which expresses VSG.

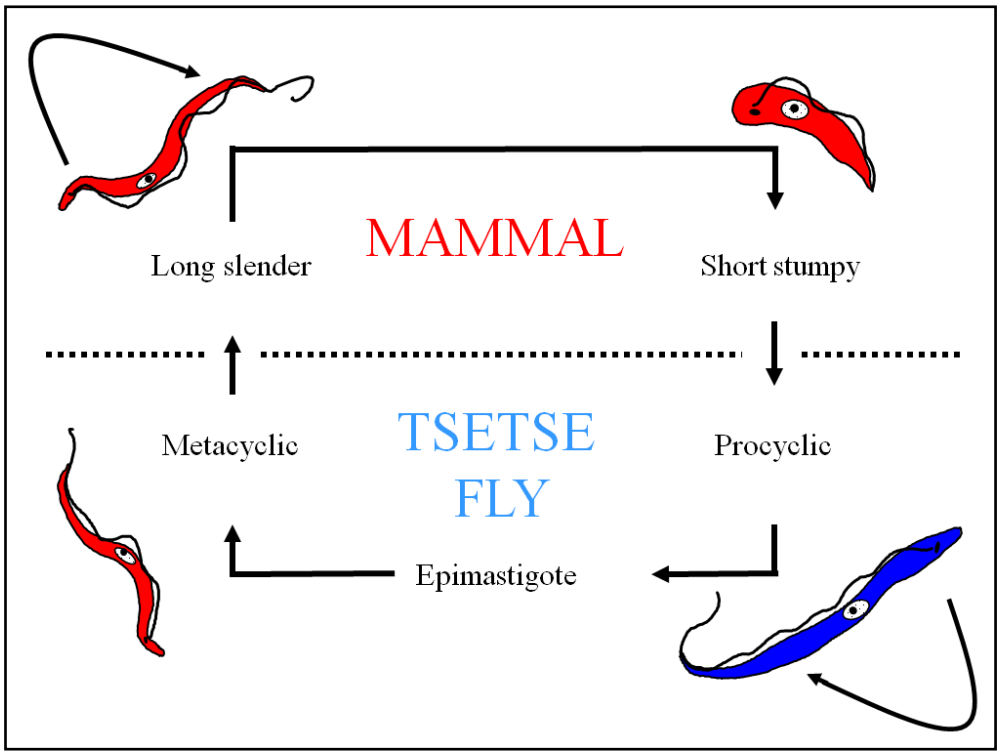


Figure 1.1 Schematic representation of the *Trypanosoma brucei* life cycle. Long slender forms expressing VSG (shown in red) are proliferative in the mammalian bloodstream. Non-proliferative short stumpy forms are pre-adapted to life in the tsetse fly, and on reaching the fly midgut adapt to the procyclic form where they are covered in procyclin (shown in blue). Procyclic *T. brucei* migrate to the salivary gland where they differentiate to the metacyclic form. They are injected with the fly bite into the mammalian bloodstream where they differentiate back into the bloodstream long and slender form. Curved arrows represent stages of cell replication.

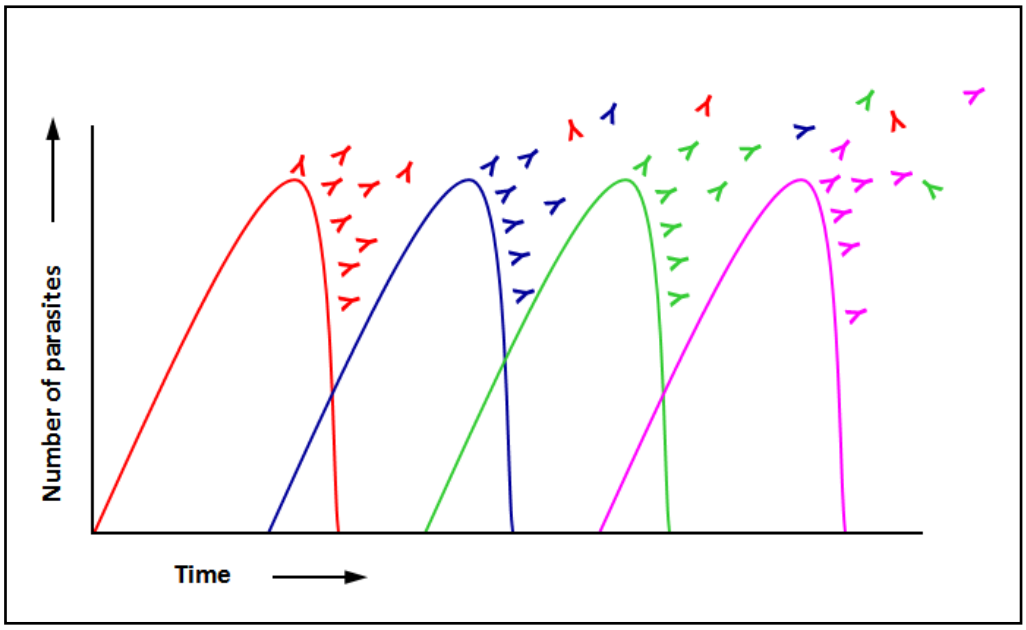


Figure 1.2 Antigenic variation in T. *brucei*: Schematic representation of fluctuation of the number of parasites over time in the bloodstream of an infected mammal. As the immune system of the host recognises one type of VSG (shown in red) and is able to mount an immune response, parasites are uncovered which have switched to a new VSG (shown blue). These switch variants are thus able to prolong the infection (Taylor et al. 2006).

The glycoprotein surface coats of T. brucei

Bloodstream form *T. brucei* faces a formidable challenge – being extracellular, they are completely exposed to the immune system of their mammalian host, and in particular to complement mediated lysis. To evade this, they express a single VSG on their surface from a repertoire of ~1200 VSG genes (Berriman et al. 2005). When the immune system of the mammalian host has raised an antibody response against the predominant VSG, this variant is effectively killed. However VSG switch variants which have switched to new VSGs not recognized by host antibodies invariably arise. Thereby the parasite is able to avoid eradication by periodically varying its outermost antigenic surface coat (Doyle 1977). This results in characteristic waves of parasitemias during the course of an infection, in which the number of parasites rises and falls through time (Figure. 1.2). Another function of the VSG is

to prevent immune recognition of other surface proteins which are invariant. The VSGs form a dense coat with ~10^7 molecules per cell (Jackson et al. 1985) which acts as a physical barrier shielding underlying invariant surface proteins. VSG is the most abundant protein in the bloodstream form cell, comprising ~10% of total cellular protein (Böhme & Cross 2002).

The active VSG is expressed from one of ~15 telomeric bloodstream form VSG expression site (ES) transcription units (Hertz-Fowler et al. 2008). Only one of these *VSG* ESs is kept active at a time in bloodstream form *T. brucei*, with the rest maintained in an inactive state, though a low level of activity which progressively diminishes towards the telomeres is seen (Vanhamme et al. 2000) (Figure 1.3). The mechanisms operating behind the strict monoallelic restriction of *VSG* ESs in bloodstream form *T. brucei* are not clearly understood.

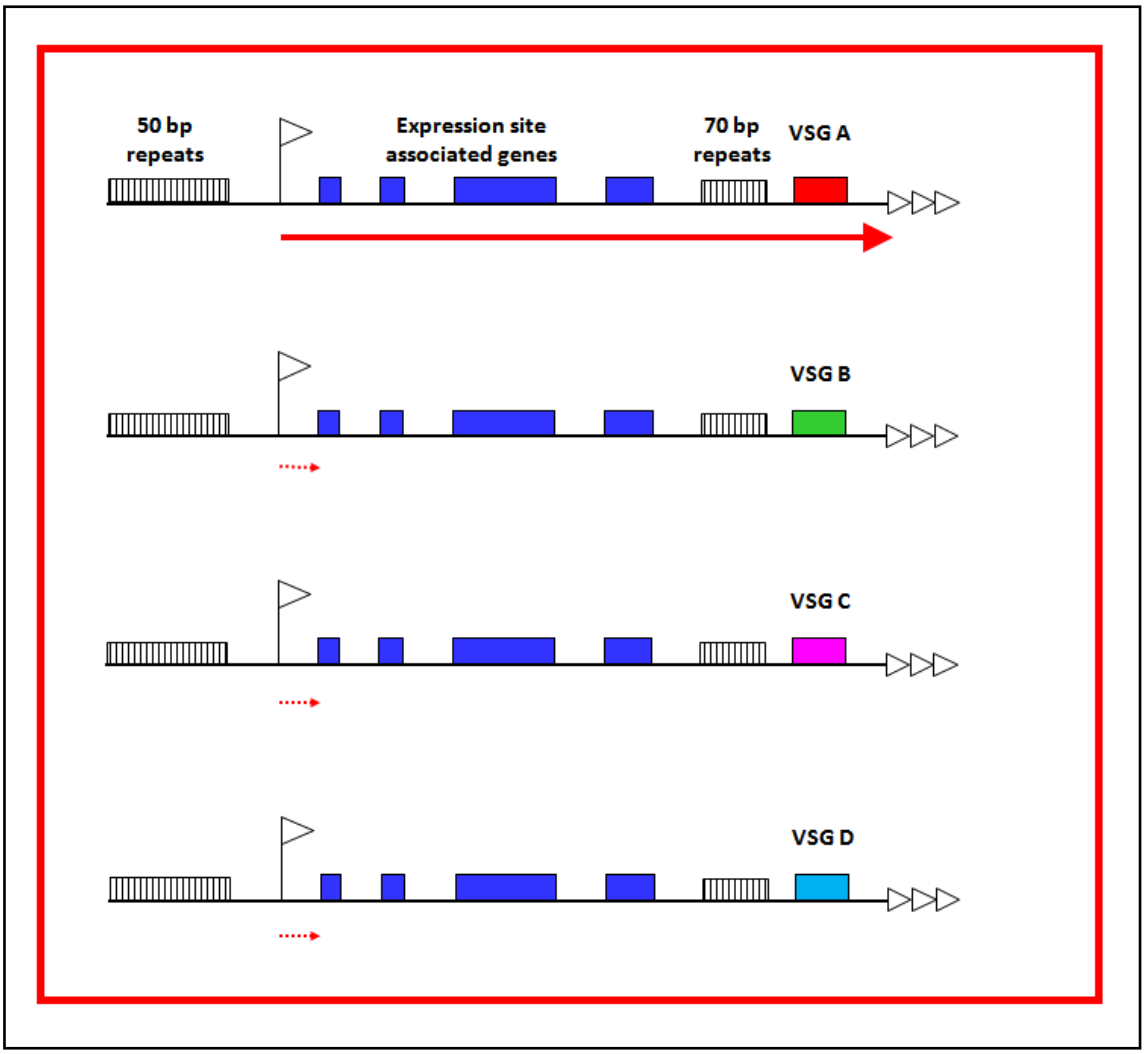


Figure 1.3 Bloodstream form *T. brucei* maintains expression from a single *VSG* ES at a time while all other ESs are kept silent. The topmost panel shows an active *VSG* ES expressing the active *VSG* (arbitrarily named *VSG A)*. Lower panels show inactive *VSG* ES with their *VSG* genes (*VSG B, C* and *D*). The ES promoter is represented by a white flag, the expression site associated genes (ESAGS) by blue boxes and the *VSG* genes by other coloured boxes. The 50 bp simple sequence repeats upstream of the ES promoter and the 70 bp repeats upstream of the *VSG* gene characteristic of all *T. brucei* expression sites are shown by striped boxes. Triangles represent the telomeres of the chromosomes. The thick red arrow represents the direction of transcription of the active *VSG* ES. Small broken red arrows represent the low levels of transcription that occur at all other *VSG* ESs. The outer red box represents a single bloodstream form *T. brucei* cell expressing the active VSG (VSG A) on its surface.

Procyclin, previously called the procyclic acidic repetitive protein (PARP) is one of the most abundant proteins of procyclic *T. brucei* (Bütikofer et al. 1997). The identification

of procyclin lagged nearly two decades behind the identification of its bloodstream form counterpart VSG as common procedures such as Coomassie staining used to identify proteins failed to detect it due to its unusual properties including surface charge (Roditi & Pearson 1990). Depending on the strain of *T. brucei*, procyclin is encoded by six or seven genes with extensive tandem EP repeats and one or two genes containing procyclin GPEET repeats (Bütikofer et al. 1997; Vassella et al. 2003). It is thought that these repeat domains make the procyclin protein resistant to protease digestion, and protect the parasite from digestive enzymes in the tsetse gut (Acosta-Serrano et al. 2001). GPEET is the major form of procyclin protein in the *T. b. brucei* 427 strain (Bütikofer et al. 1997).

Polycistronic transcription in T. brucei

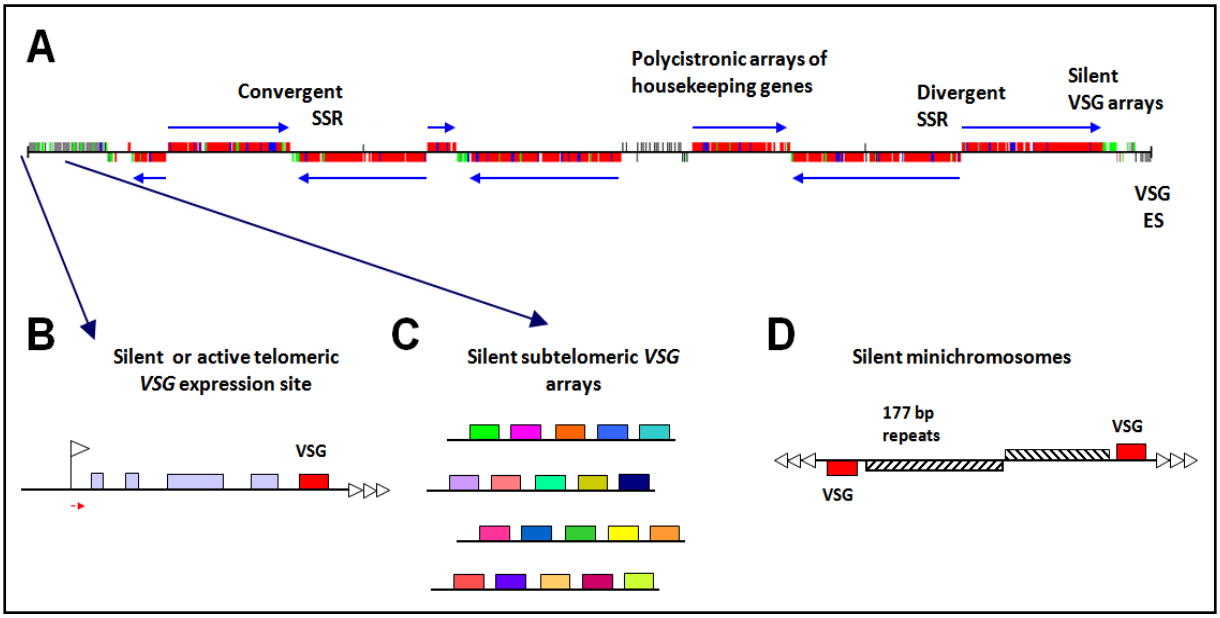


Figure 1.4 Most of the *T. brucei* genome is organized in polycistronic transcription units. A schematic of a typical large chromosome is shown in **A**. Open reading frames are shown in red, and polycistronic transcription units are indicated with blue arrows. **B)** The telomeric and **C)** sub-telomeric regions have been expanded below to show the VSG expression site and arrays of silent VSG genes respectively. **D)** Schematic representation of a typical minichromosome, of which there are ~100 in each cell. These are made up of

palindromes of arrays of 177 bp repeat sequences (striped boxes) with VSG genes located at the telomeres (adapted from Rudenko 2010).

There is evidence that the ancestor of *T. brucei* branched from the eukaryotic evolutionary tree relatively early (Embley & Martin 2006). As a result, *T. brucei* exhibits certain molecular features which are atypical of eukaryotes. First, there is very little control of RNA polymerase II mediated transcription, with most of the genome transcribed constitutively. Levels of mRNA in the cell are predominantly regulated post-transcriptionally, mainly at the level of RNA stability. The genes within *T. brucei* chromosomes (Figure 1.4) are arranged in long clusters of genes which are transcribed polycistronically and constitutively sometimes covering over 150 kb (Clayton 2002). This is superficially reminiscent of bacterial operons. However, unlike in bacteria, the transcription of these genes is neither co-regulated nor are adjacent genes necessarily involved in the same biological pathway (Berriman et al. 2005). Individual mRNAs are cleaved from precursor transcripts by the *trans*-splicing of a spliced leader (SL) transcript to the 5' end of the mRNA as well as polyadenylation of the 3' end (Liang et al. 2003). *T. brucei* has up to 200 SL RNA transcription units arranged in arrays transcribed by RNA polymerase II (Schimanski et al. 2005) (see Figure 2.8B in Chapter 2).

Apart from the SL promoter, no RNA polymerase II promoter has yet been definitively identified. The areas between the polycistronic transcription units are called divergent or convergent strand switch regions (SSRs) (Figure 1.4) (Wright et al. 2010; Siegel et al. 2009). Promoters are undoubtedly present between polycistronic transcription units which diverge away from each other, but mapping them so far has not been possible for technical reasons.

The unusual role of RNA polymerase I in T. brucei

Another unusual feature of trypanosomes is that RNA polymerase I does not exclusively transcribe ribosomal DNA (rDNA) as is the case in other eukaryotes. RNA polymerase I also transcribes the only protein coding loci that are known to be transcriptionally regulated in trypanosomes – the *VSG* expression sites (ESs) and the procyclin loci (Gunzl et al. 2003). RNA polymerase I transcription units are transcribed at a roughly 10-fold higher rate than RNA polymerase II transcription units (Biebinger et al. 1996), This could be advantageous for the expression of a single copy gene like the active VSG, which must be expressed at very high levels (Böhme & Cross 2002).

The architecture of these *T. brucei* RNA polymerase I promoters (rDNA, procyclin and *VSG* ES) is not identical. The rDNA and procyclin promoters both comprise four structural domains which appear to be functionally conserved with that of the *S. cerevisiae* rDNA promoter. The essential domain I constitutes the core promoter which marks the transcription initiation site. The upstream domains II and III form a bipartite element which facilitates transcription factor binding. Domain IV is located even further upstream and is involved in chromatin remodelling (Gunzl et al., 2007).

However, the structure of the *VSG* ES promoter is considerably different. The *VSG* ES consists of two small domains both of which are essential for promoter activity and together form the core promoter. The ES promoter lacks an identifiable transcription factor binding element and upstream chromatin remodelling domain (Gunzl et al., 2007), indicating that this promoter could be more sensitive to local chromatin structure. This has been at least partially confirmed by experiments which have shown that the ES promoter but not the rDNA promoter is silenced when inserted into a transcriptionally silent telomere-proximal location (Horn & Cross 1995) and into the silent rDNA spacer and VSG Basic Copy arrays (Horn &

Cross 1997). These differences in RNA polymerase I promoter behaviour are observed in procyclic *T. brucei*, though not in bloodstream form *T. brucei* (Rudenko et al. 1995; Horn & Cross 1997).

b. Chromatin and regulation of gene expression

The eukaryotic genome, if stretched out end to end, far exceeds that of the typical eukaryotic cell. It is therefore essential to pack DNA in an organized way, and to unravel parts of it to enable enzymes to access it for repair, replication, and transcription. DNA is packaged as nucleosomes, which are made up of 146 bp of DNA wound around a protein octamer core comprised of eight histone proteins – two molecules each of histone H2A, H2B, H3 and H4. These core histones are among the most highly conserved proteins in eukaryotes. Another histone, H1, is usually associated with the DNA 'linker' region between two adjacent nucleosomes, and is less conserved than the core histones. All other chromatin proteins are broadly classified as non-histone proteins, and include various proteins such as scaffold proteins, transcription factors and chromatin remodelling proteins.

The histone code

All histones, in particular histones H3 and H4, are subject to post-translational modifications (PTMs) by the covalent linkage of extra moieties to amino acids on their N-terminal tails. At least 60 different modified histone residues have been identified by mass spectroscopy or by using specific antibodies, and there are likely to be additional modifications discovered (Kouzarides 2007). Of the major classes of covalent PTMs, some are known to be associated with active or decondensed chromatin while others are associated

with repressed or condensed chromatin. Histone acetylation involves neutralization of the positive charge on the free ε amino group of lysine by the negative CH_3CO group. This change in charge allows the site to act as a binding site for other proteins (Lewin 2004) and occurs primarily during DNA replication (Verreault 2000) or during gene activation (Pokholok et al. 2005). A comparison of patterns of histone acetylation at different loci in the mammalian genome shows the presence of conserved acetylation islands which are regions rich in acetylated lysine residues occurring on the 9^{th} and 14^{th} amino acid positions of the highly conserved H3 protein (denoted as H3K9acK14ac) (Mellor et al. 2008). Many of these modified areas correspond to known transcriptional regulatory elements such as the enhancers and promoters of active genes (Kouzarides 2007; Mellor et al. 2008).

Histone methylation involves the addition of a methyl group to an amino acid side chain (usually lysine, and less frequently arginine). However unlike acetylation, one to three methylations can occur on an amino acid residue, increasing the complexity of this PTM. Each of the three methylated forms is capable of specific interactions with different domains on other proteins in different contexts (Mellor et al. 2008). While histone acetylation occurs mainly at the regulatory elements such as the promoters situated at the beginning of the gene, histone methylation can occur throughout transcribed regions (Pokholok et al. 2005; Mikkelsen et al. 2007).

Some of the other known PTMs include ubiquitylation of lysine residues on H2A and H2B usually correlated with repressed and activated transcription respectively (Wang et al. 2004; Berger 2007) and phosphorylation on serine and threonine residues of histones where it has been associated with transcription activation and DNA repair (Berger 2007; Kouzarides 2007). An important feature of all histone PTMs is that they are removable. Histone

deacetylases (HDACs) are responsible for removing acetyl groups and are associated with transcriptional repression (Richards et al. 2002).

Histone variants

Eukaryotes contain variants of all histones (with the exception of H4) which co-exist with core histones within the cell (Lewin 2004). The H2A variant H2A.Bbd is believed to be associated with transcriptionally active euchromatic regions of the genome and possibly operates by destabilizing the nucleosome and locally unfolding the chromatin (González-Romero et al. 2008). The histone variants H2A.Z and H3.3 carry their own modification patterns, and are exchanged within chromosomal domains (Goldberg et al. 2007). Histone H3.3 is enriched in the dosage compensated male X chromosome in *Drosophila*) (Ahmed & Brickner 2007). Another interesting example of an H3 variant is CENP-A, which is found in nucleosomes at the centromeres of chromosomes, where it is possibly involved in bringing together sister chromatids during replication in yeast (Black & Bassett 2008; Ekwall 2004).

DNA modifications

In addition to modifying histones, eukaryotes also modify DNA sequences as a means of epigenetically regulating genes. The most common DNA modification is methylation, which occurs almost exclusively on the cytosine residues of CpG di-nucleotides (Bird 2002). The functions of DNA methylation appear to be quite diverse. It has been reported that highly methylated CpG islands (areas of the genome rich in CpG di-nucleotides) are correlated with transcription repression (Goldberg et al. 2007), while promoters of highly active genes appear to be excluded from DNA methylation.

c. Monoallelic expression in eukaryotes

Monoallelic expression of a single gene from a multigene family is seen in a variety of eukaryotic systems, but remains a poorly understood phenomenon. In this situation all genes except for one of a multi-gene family are kept inactive, and different organisms do this by various means. *T. brucei* expresses only a single type of VSG molecule on its surface at a given time, necessitating a stringent system of monoallelic control of this large gene family. What are other examples of this?

Mammalian olfactory system

Olfactory receptors (ORs) in mammals are expressed in a mutually exclusive manner in olfactory sensory neurons (OSNs). The OR gene family is one of the largest mammalian gene families described, comprising ~1000 different genes in mice and scattered over 51 loci across 17 chromosomes (Godfrey et al. 2004). Only a single allele of an OR is expressed in a given cell, such that within an individual organism the proportion of OSNs expressing the maternal or paternal alleles is approximately equal and these coexist mosaically. The expression of these OR genes is distributed over a few million OSNs which are located in ~1800 different glomeruli within the olfactory bulb (Mombaerts 2001). Serizawa et al. (2000) were able to use yeast artificial chromosomes containing a mouse OR and its promoter to show that without the OR upstream region the OR genes were not expressed. They were subsequently able to identify a 2 kb region or 'H enhancer' (marked H in Figure 1.5), about 75 kb upstream of the OR gene cluster that shared homology with the human genome as well (Serizawa et al. 2003). They proposed that the regulation of OR expression occurred through interaction with the H enhancer which functioned as a *cis*-acting locus control region (LCR)

in this case. In addition, it is postulated that the functional OR protein in some way prevents further activation of other OR genes (Figure 1.5) (Serizawa et al. 2004).

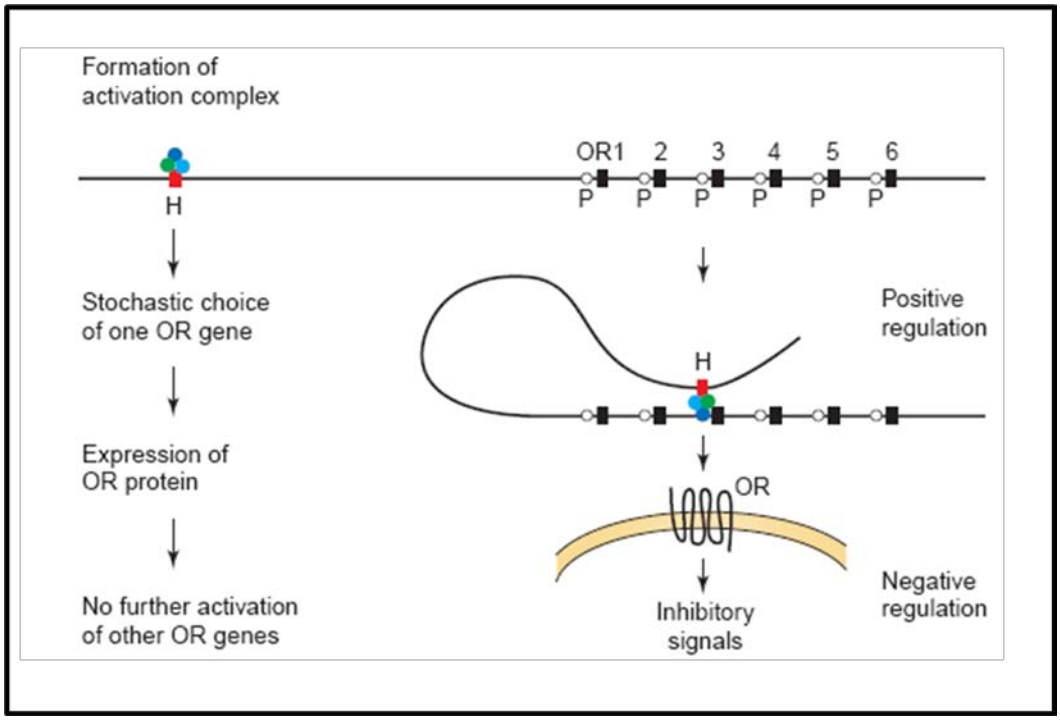


Figure 1.5 Involvement of the 'H enhancer' locus control region (LCR) in regulation of OR gene expression: The H enhancer (the homologous region about 75 kb upstream of the OR cluster and denoted by H) stochastically selects one of the promoters (P) randomly resulting in transcriptional activation of that OR. Expression of the functional OR protein subsequently blocks further OR activation in that OSN (from Serizawa et al., 2004).
Reprinted from Trends in Genetics, 20(12), Serizawa, S., Miyamichi, K. & Sakano, H., One neuron–one receptor rule in the mouse olfactory system, p. 650, © 2004, with permission from Elsevier. http://dx.doi.org/10.1016/j.tig.2004.09.006

Yeast mating type switch

Silencing is best understood in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, and mating type switching is one of the most studied examples within these organisms. Haploid budding yeast can have one of two mating types, determined by the sequence present at the active mating-type locus (MAT). The two yeast mating types are usually represented by 'MATa' and 'MATα' (Nasmyth 1982). In addition to the active MAT locus (which can contain either MAT type a or α), there are additional silent MAT loci

containing each mating type, denoted by HMRa and HMLα respectively. These silent MAT loci are separated from each other by an 11 kb gap, and from the active MAT loci by at least 15 kb (Grewal 2000) (Figure 1.6). Yeast can be propagated as either haploids or diploids, and two haploid cells of different mating types can fuse to produce a diploid cell (which can subsequently sporulate meiotically). Switching between the two yeast mating types occurs if MATa in the active MAT cassette is replaced by HMRα, or if the MATα in the active MAT is replaced by HMRa (Nasmyth 1982).

It has been found using deletion analysis that the regions on either side of HMR and HML are essential for keeping these regions silent (Laurenson & Rine 1992). These regulatory regions, which are located ~1 kb from the two HM loci, contain an enhancer region. Point mutations within these enhancer regions did not affect silencing (Huang 2002). It was found that the introduction of mutations in a group of silent information regulator (Sir) proteins resulted in silencing of the HMR cassettes being either reduced or abolished (Nasmyth 1982). Silencing at both HMR loci requires the Sir complex which comprises Sir1p, Sir2p, Sir3p and Sir4p, none of which are essential for survival but which are required for mating type switching. Hecht et al. (1995) were able to show a direct link between chromatin, the Sir complex, and silencing when they found that the introduction of mutations in the N-terminal tails of histones H3 and H4 lead to cessation of silencing at the two HM loci. They also demonstrated that the N-terminal tails of histones within these regions interacted with the Sir3 and Sir4 proteins. In addition to this role in silencing, there is evidence that the Sir complex is also involved in telomeric heterochromatin (Hecht et al. 1995). It is currently believed that the Sir complex does not mediate silencing on its own, but initiates a pathway that results in silencing further downstream. It is also evident that there are other factors which are involved in the inheritance of yeast mating types, though their mechanisms are as yet poorly understood (Huang 2002).

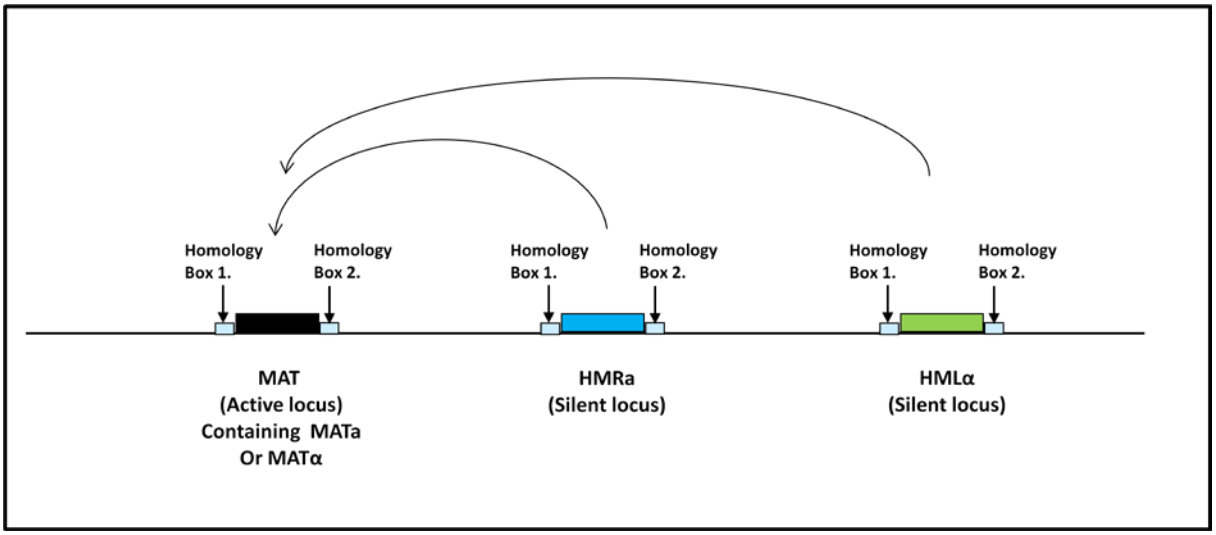


Figure 1.6 Schematic representation of the yeast mating-type region: Of the three yeast mating type cassettes HMRa and HMLα are kept silent, while MAT is active. MAT can contain either of the two mating types; MATa or MAT α. Haploid cells express only one of the two mating types, and can fuse with another haploid cell expressing the opposite mating type to form a diploid cell (compiled from Nasmyth, 1982 and Grewal, 2000).

Monoallelic exclusion in Plasmodium

The protozoan parasite *Plasmodium falciparum*, which causes malaria in humans, infects mammalian erythrocytes during a part of its lifecycle. In order to prevent clearance of infected erythrocytes (IEs) by the spleen, the parasites express a variable protein on the surface of the erythrocyte called erythrocyte membrane protein 1 (PfEMP1). This parasite derived protein mediates binding of the erythrocyte to several host factors expressed on endothelial cells enabling IEs to sequester into tissues and organs (Montgomery et al. 2007). However, as PfEMP1 is directly exposed to the host's immune system, the parasites switch between variants during infection in order to avoid antibody mediated clearance. PfEMP1 is encoded by ~60 *var* genes located mainly at the sub-telomeric regions of chromosomes of which one is expressed in a mutually exclusive manner (Scherf et al. 2008). There is evidence that a 'counting mechanism' involving the *var* intron operates to keep only one *var* gene active at a time. When the *var* intron is disrupted, the *var* promoter is constitutively expressed

and evades silencing by the counting mechanism. Interestingly this has no effect on the expression of other *var* genes. Therefore, the *var* intron appears to be essential for silencing a given *var* gene but not for mutually exclusive silencing of *var* (Frank et al. 2006). There is evidence for epigenetic regulation of *var* genes in *P. falciparum* by way of histone modifications and specific sub-nuclear positioning of heterochromatinous regions (reviewed in Scherf et al. 2008 and in Cui & Miao 2010) which clearly indicates that there are as yet unknown factors waiting to be discovered.

d. Chromatin regulation in *T. brucei*

The involvement of chromatin regulation in transcription control in *T. brucei* was disputed until recently. This was primarily due to what appeared to be conflicting results from various studies. It was first reported that while the active *VSG* ES gene chromatin was preferentially digested by single-strand-specific endonucleases, there was no detectable difference in the nucleosome pattern of active or silent ESs when treated with micrococcal nuclease (Greaves & Borst 1987). A subsequent study which investigated chromatin structure around the ES promoter using a marker gene inserted downstream of the ES promoter detected a nuclease hypersensitive site within the core promoter, in both active and silent ESs, with no differences in nuclease sensitivity between active and inactive states (Navarro & Cross 1998). These findings led to the conclusion that chromatin remodelling was presumably not involved in regulating the active ES in bloodstream form *T. brucei*.

This view was partly revised after another study presented different results on ES regulation between bloodstream and procyclic *T. brucei*. Here, an exogenous T7 promoter was used to probe for chromatin state. A cassette containing the reporter gene luciferase flanked by an upstream T7 RNA polymerase (T7 RNAP) promoter was integrated into

several loci in trypanosomes which constitutively expressed T7 RNAP. Luciferase expression was monitored to investigate transcriptional accessibility of chromatin in these regions *in vivo*. While there was no observable silencing of the T7 RNAP mediated transcription in bloodstream form *T. brucei*, transcription was repressed along the length of the entire ES in procyclic cells. It was proposed that two independent transcriptional mechanisms operate on the ES depending on the lifecycle stage. Chromatin remodelling was proposed to be correlated with silencing all ESs in the procyclic form. In contrast, a nucleation site recruitment-dependent activation of the active ES was thought to be involved in bloodstream form *T. brucei* (Navarro et al. 1999). This conclusion that different silencing mechanisms operate in the two life cycle stages was compatible with earlier studies, which had shown that if the endogenous ES promoter was replaced by an rDNA promoter, transcription was derepressed in procyclic but not in bloodstream form *T. brucei* (Rudenko et al. 1994; Rudenko et al. 1995).

DNA modifications

There is evidence for the role of DNA modification in transcription control. A novel DNA base β-D-glucopyranosyloxymethyluracil (base J) was discovered serendipitously when certain restriction enzymes failed to cut at sites in silent telomeric *VSG* genes (Gommers-Ampt et al. 1993; Borst & Sabatini 2008; Cliffe et al. 2010). Base J was eventually identified as a glucosylated dUTP which is present at a fraction of the sites in silent ESs, and to date remains the only example of a hyper-modified DNA base in eukaryotes. Base J has been subsequently found in other kinetoplastids, including those which do not undergo antigenic variation, indicating that at least ancestrally it could have had a more general role in transcription regulation (Borst & van Leeuwen 1997). Base J is synthesised by the

hydroxylation of thymidine by two thymidine hydroxylase (TH) enzymes in *T. brucei*, named JBP1 and JBP2 (DiPaolo et al. 2005; Cliffe et al. 2009), prior to the resultant intermediate being glucosylated (Borst & Sabatini 2008). JBP1 appears to mainly stimulate J-synthesis at chromosome-internal positions and JBP2 mediates J-deposition at the telomeric regions (Cliffe et al. 2010).

Interestingly, J is absent in procyclic *T. brucei* (but not in the insect stages of other trypanosomatids). In bloodstream form *T. brucei* over 50% of base J is located in the silent telomeric repeats and in the silent ESs as well as other silent repeat regions such as the 177 bp repeats of the minichromosomes, as well as regions flanking RNA polymerase II polycistronic transcription units (Borst & Sabatini 2008; Cliffe et al. 2010). These findings could suggest that J is involved in gene silencing. However, even after a knock down of JBP1 producing a 20-fold decrease in J levels, there is no derepression of silent telomeric VSGs (Cross et al. 2002). Additionally, null-mutants of both JBP1 and JBP2 (which have no J) also show no derepression of the silent VSGs (Cliffe et al. 2009). It is possible that J is involved in preventing homologous recombination between repeat sequences (Borst & Sabatini 2008) or in transcription termination and initiation (Cliffe et al. 2010). However its absence from procyclic *T. brucei* shows that it is not essential in this life-cycle stage.

Epigenetic regulation in trypanosomes

The importance of chromatin regulation in *T. brucei* became more clear after the discovery of various chromatin-associated proteins which play critical roles in ES control. Additionally, the nucleosome distribution pattern across the *T. brucei* genome was investigated using more modern techniques. Using chromatin immunoprecipitation (ChIP) of core histones and micrococcal nuclease digestion (MNase) of genomic DNA, it was shown

that nucleosomes are greatly depleted on the active ES and around the promoter regions of actively transcribed loci including the rDNA units in bloodstream form *T. brucei* (Stanne & Rudenko 2010; Figueiredo & Cross 2010). Additionally, comprehensive genome-wide analyses of several recently identified histone variants in *T. brucei* have shown that these are enriched up to 300-fold at putative RNA polymerase II transcription start sites (TSSs) and transcription termination sites (TTSs) (Siegel et al. 2009).This indicates that *T. brucei* uses histone variants to regulate transcription.

Chromatin remodelling proteins involved in the stringent control of ESs in *T. brucei* have been identified. TbISWI, a member of the pan-eukaryotic SWI2/SNF2 family of chromatin remodelers (Mohrmann & Verrijzer 2005) (discussed in detail in Chapter 3), was shown to be involved in maintaining transcriptional inactivity of the silent ESs, as its depletion led to derepression of ES promoter proximal regions (Hughes et al. 2007). A non-essential histone methyltransferase called DOT1B has also been shown to be required for complete ES silencing. In DOT1B knock-out cells, a ten-fold upregulation of transcription of silent *VSG* transcripts is reported. In addition, there is delayed silencing of the previously active ES during ES switching (Figueiredo et al. 2008). Landeira et al. (2009) showed that partial depletion of the cohesin complex subunit SCC1 led to an increase in the frequency of switching events resulting in cells expressing VSG from a new ES. Silent ES promoters show a low level of transcriptional activity in both bloodstream and procyclic *T. brucei* but this does not continue far into the ES (Vanhamme et al. 2000). A telomere binding protein, RAP1, was recently found to be essential for maintaining transcriptional silence of the distal or telomeric end of the ES. RAP1 depletion led to a derepression of transcription from the ES promoter to the telomeric ends of ESs, though this derepressed transcription was still approximately 70-100 folds lower than that observed from an active ES (Yang et al. 2009). Lastly, the depletion of TbSpt16, a sub-unit of the FACT (facilitator of transcription)

complex, triggers a cell cycle arrest and a 20-30 fold derepression of silent ESs specifically in the G2/M cell cycle stage (Denninger et al. 2010).

From these studies, it is evident that ES control is a multi-layered phenomenon, with multiple factors controlling the level of activity of different areas of the ~50 kb length of ESs (Pays et al. 2001). A model proposed by Yang et al. (2009) suggests that there are two types of silencing gradients occurring on ESs – one which is mediated by players like TbISWI and operates around the region of the ES promoter, and another which is imposed by RAP1 and occurs as a telomeric gradient increasing towards the telomeric ends.

e. Regulation of ribosomal DNA

Ribosomal RNA (rRNA) synthesis is mediated by RNA polymerase I in all eukaryotes investigated. As rRNA needs to be expressed at very high levels, eukaryotes have multiple copies of the ribosomal DNA (rDNA) genes in their genomes. However, of the 150-200 rDNA copies in yeast or of the ~400 copies in humans, only about 50% of the repeats are transcribed at a time in actively growing cells (Birch & Zomerdijk 2008). This is probably a way of regulating rRNA synthesis (Russell & Zomerdijk 2005). It is known that active and silent rRNA genes exist in distinct epigenetic states in mammalian cells. Cross-linking with psoralen, which intercalates specifically with double-stranded DNA in a non-nucleosomal conformation, has been used to show that only about 50% of rDNA genes are accessible, while the rest are associated with heterochromatin (McStay & Grummt 2008). Additionally, chromatin immunoprecipitation (ChIP) followed by the digestion of DNA with methylation-sensitive restriction enzymes has shown there are differences in DNA methylation patterns between active and silent rRNA genes, with silent genes being enriched in methylated CpG dinucleotides and active genes being hypo-methylated (McStay & Grummt 2008). ChIP has

also been used to show the differential distribution of modified histones in these regions, with hypomethylated active genes associated with acetylated histones, as well as the H3K4me3 modification. In contrast, silent genes are associated with methylated histones (Santoro & Grummt 2005; Berger 2007; Schlesinger et al. 2009).

The precise mechanisms by which cells maintain this distinction between the epigenetic state of rDNA repeats is unclear, although the players involved are emerging. A working model of how rRNA genes may be silenced is shown in Figure 1.7. Briefly, TTF-I (transcription termination factor I) is a multi-function protein which binds to mammalian rDNA promoter proximal terminator elements. Here, apart from influencing nucleosome positioning, it interacts with TIP5 (TTF-I-interaction protein 5) which has a bromodomain motif that interacts with histone deacetylases. The interaction of TTF-I and TIP5 is key in recruitment of NoRC (nucleolar-remodelling complex) to the rDNA promoter. In turn the NoRC complex, which comprises of the ATPase SNF2h (also known as ISWI and discussed in detail in Chapter 3) and TIP5 itself, recruits various histone modifiers to rDNA resulting in silencing of the locus (Birch & Zomerdijk 2008; McStay et al. 2008).

The importance of silenced rDNA is highlighted by the fact that the nucleolus is surrounded by heterochromatin. This suggests that heterochromatin is required for maintaining the architecture of the nucleolus, and indeed it has been shown that loss of rDNA silencing correlates with rDNA instability and nucleolar disintegration (McStay & Grummt 2008). Similarly, the exact mechanism by which maintenance of activity of rRNA genes occurs is not known, although there is a growing body of evidence which suggests that DNA-binding proteins containing HMG (high mobility group) box domain are involved in facilitating rDNA transcription. Active rDNA repeats have certain distinguishing features. In yeast, a HMG box containing protein called Hmo1 localizes specifically along the length of active rDNA repeats, and has a role in defining the boundaries of a region near the

transcription start site where the preinitiation complex assembles (Hall et al. 2006; Kasahara et al. 2011). In mammals, another HMG box containing protein, UBF (upstream binding factor) binds throughout the intergenic spacers and rRNA coding region of active repeats and possibly recruits RNA polymerase I transcription machinery (Birch & Zomerdijk 2008; McStay & Grummt 2008) and is involved in the physical assembly of the nucleolus (Mais et al. 2005). Key aspects of Hmo1 and UBF are discussed in greater detail in Chapter 4.

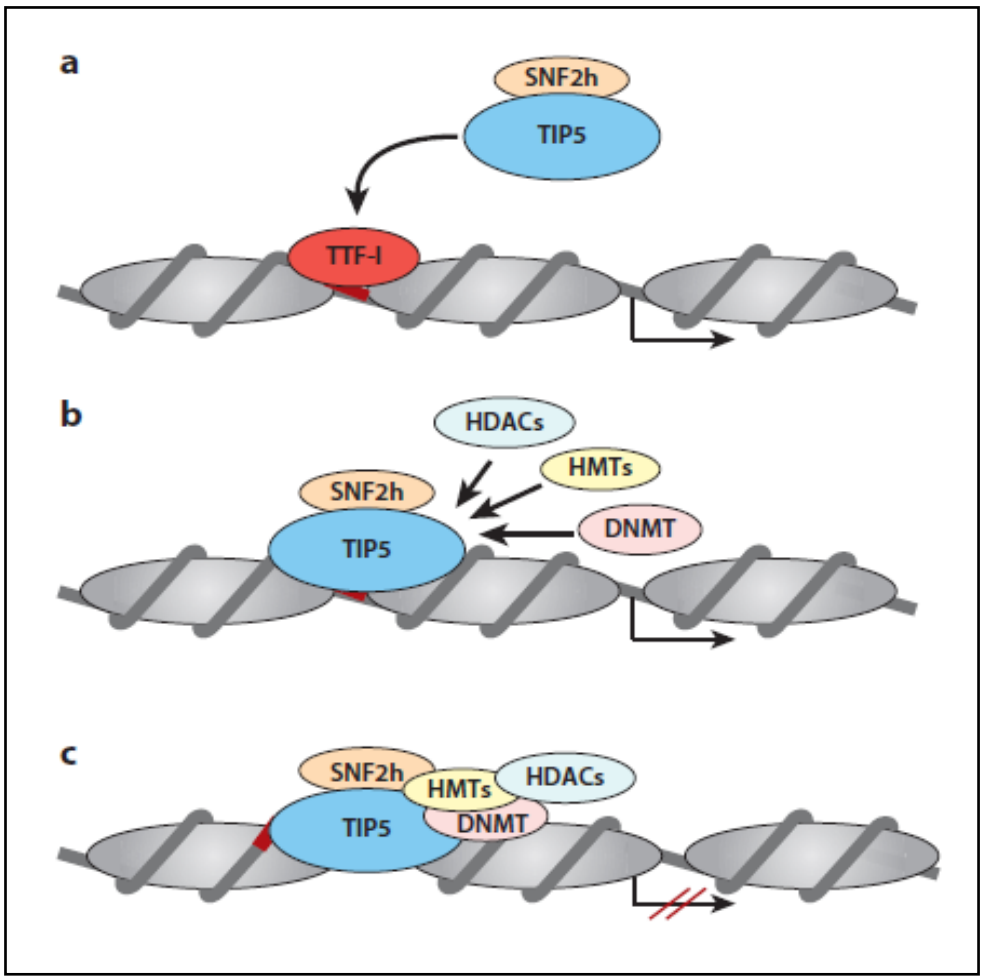


Figure 1.7 Schematic representation of silencing of rDNA genes. TTF-1, which is bound to the rDNA promoter, recruits NoRC which is an ISWI-containing complex (a). NoRC then facilitates the deacetylation of histones H3 and H4 and methylation of H3K9, H3K20 and H3K27 by interacting with various histone deacetylases (HDACs), histone methyltransferases (HMTs), and DNA methyltransferases (DNMTs) (b). These histone marks possibly provide a signal for ISWI to shift the promoter nucleosome further downstream into a position which is

unfavourable for transcription initiation (c). Alternatively, SNF2h action could expose CpG at position -133 to methylation, which impairs assembly of the transcription machinery (adapted from McStay & Grummt 2008).

f. Aims of this thesis

The data presented in this thesis is divided into four chapters, each designed with an aim to better understand transcription control and regulation of chromatin in *T. brucei*.

Aim 1: To investigate the role of the DNA-binding protein NLP in bloodstream form *T. brucei*.

NLP had originally been identified by affinity chromatography as binding transcriptionally silent repetitive minichromosome DNA in *T. brucei* (K. Ersfeld, unpublished results). I wanted to know how trypanosomes are able to specifically express VSG from a single ES, while maintaining the rest of their VSG repertoire in an inactive state. I investigated if the depletion of NLP led to a perturbation of the transcriptional states of *VSG* ESs in bloodstream form *T. brucei*. I found that NLP is essential for maintaining normal levels of transcription in both active and silent *VSG* ESs. Our findings are reported in Chapter 2. A part of this Chapter has been published recently (Narayanan et al. 2010).

Aim 2: To examine if NLP interacts with other chromatin proteins and if so to characterize those interactions.

Within the last four years a number of proteins have been shown to be involved in transcription control in *T. brucei*. These results indicated that chromatin regulation plays an important role in *VSG ES* regulation, and is a multi-layered process. I investigated if NLP interacts with other chromatin proteins in order to gain a better understanding of its role. Interestingly, I discovered that NLP is a member of a novel TbISWI complex along with two

other novel proteins which I have temporarily named FYRN1 and TbRCC1. Our preliminary results show that both FYRN1 and TbRCC1 are also chromatin-associated proteins. In addition, I have evidence that NLP is part of more than one complex in bloodstream form *T. brucei*. These results are reported in Chapter 3.

Aim 3: To investigate the functions of the HMG box containing protein TDP1 in *T. brucei*.

The trypanosome DNA-binding protein (TDP1) is predicted to contain two HMG boxes, which indicates that it functions as a chromatin regulating protein in *T. brucei*. I report that TDP1 is an essential nuclear protein which could be involved in facilitating transcription. I found that blocking TDP1 synthesis results in a reduction in transcription of RNA polymerase I transcribed genes. I also found that TDP1 is significantly enriched in the nucleolus and strikingly also in the active *VSG* ES in bloodstream form *T. brucei*. These results indicating that TDP1 is an architectural chromatin protein important for transcription control in *T. brucei* are presented in Chapter 4.

Aim 4: To discover novel transcription regulators in bloodstream *T. brucei*

While blocking the synthesis of proteins such as TbISWI, NLP and TbRAP1 (to name just a few) led to the derepression of silent *VSG* ESs in bloodstream form *T. brucei*, this level of transcriptional activation remained 10-100 fold lower than the level of transcription in a fully active *VSG* ES. I identified proteins in the *T. brucei* genome database with BLAST searches using chromatin remodelers in other eukaryotes. I subsequently screened these candidate proteins for involvement in *VSG* ES transcription control. I found that inducing RNAi against two of the proteins (which I provisionally named Brom3 and Saga1) resulted in *VSG* ES derepression as well as in a striking cell cycle arrest. These results are presented in Chapter 5.

Chapter 2

NLP is a novel transcription regulator in *T. brucei*

(Parts of this Chapter are published in *Nucleic Acids Research* (Narayanan et al. 2010). See Appendix C)

A. Introduction

The nucleoplasmin family

The nucleoplasmin (NPM) family includes a variety of chromatin associated proteins which are found in all branches of the eukaryotic kingdom. They are named after *Xenopus laevis* nucleoplasmin, which was one of the first histone chaperones to be identified (Laskey et al. 1978; Laskey & Earnshaw 1980). *In vitro*, nucleoplasmin binds to the core histones and translocates them to DNA where they are assembled into nucleosomes (Laskey et al. 1980). A characteristic feature of this family is that they are acidic proteins containing a polyglutamic acid rich C-terminal tract (Eirín-López et al. 2006). This is perhaps not surprising considering histones are amongst the most basic proteins found in cells.

The nucleoplasmin family is divided into four categories based on protein sequences: these are NPM1, NPM2, NPM3, and the invertebrate nucleoplasmin proteins which have been typically described as 'nucleoplasmin-like' (Frehlick et al. 2007). NPM1 and the nucleoplasmin-like proteins appear to have a monophyletic origin in their protein topologies, while NPM2 and NPM3 appear to be polyphyletic in origin due to independent differentiation in mammals compared to other vertebrates (Eirín-López et al. 2006).

NPM1 or nucleophosmin is functionally diverse, and has been found to have roles in cell proliferation and differentiation. It is subjected to a variety of post-translational modifications such as phosphorylation, acetylation, SUMOylation and ubiquitination which regulate these functions. It is one of the best characterized nucleoplasmins in humans, as mutations leading to an alteration of its function have been reported in a number of human cancers (Grisendi et al. 2006; Okuwaki 2008). NPM2, unlike NPM1 and NPM3 which are expressed ubiquitously, is expressed mainly in oocytes, where it has crucial roles in the condensation and re-organization of sperm chromatin (Tamada et al., 2006; Okuwaki, 2008).

NPM2 includes the family's founding member *X. laevis* nucleoplasmin. This protein has been shown to act as a reservoir for H2A-H2B dimers during high rates of DNA replication, and in the assembly of nucleosomes as well as interacting with other basic non-histone chromatin proteins (Ramos et al. 2005; Taneva et al. 2009).

The NPM3 group is currently not as well-characterised as the other categories of metazoan nucleoplasmins. In humans, it is homologous with the N-terminus of NPM1, with which it binds to form a complex which mainly localizes to the nucleolus. Here, it appears to be indirectly involved with modulating transcription of rDNA. NPM3 lacks the nucleic acid binding domains of NPM1, and by binding to it, could be preventing NPM1 from carrying out ribosomal biogenesis (Huang et al. 2005). When it was investigated if NPM3 could modulate histone chaperone activity of NPM1 in this manner, it was found that NPM3 was a potent inhibitor of NMP1-mediated histone assembly activity. Its over-expression resulted in an increase in transcription of a reporter gene, presumably due to reduced histone assembly and a resultant 'opening up' of chromatin at that location. It was also reported that NPM3, unlike NPM1, had no histone assembly activity on its own, indicating that the unique C-terminus of NPM1 is essential for this function (Gadad et al. 2010). Together, these results indicate that a balance between NPM3 and NPM1 is important in maintaining homeostasis. This is a particularly important finding, as it has been reported that NPM1 is over-expressed in human tumorigenesis (Grisendi et al. 2006)

The final category of nucleoplasmins which includes "nucleoplasmin-like" proteins (NLPs) from invertebrates (Frehlick et al. 2007) is poorly characterized. Presently, the only known NLPs are those reported in the molluscans *Spisula solidissima* (Herlands & Maul 1994), *Mytilus californianus* (Rice et al. 1995) and the best characterised nucleoplasmin of this category – *Drosophila* nucleoplasmin-like protein (dNLP) (Ito et al. 1996). Little is known about these proteins other than that they share the fundamental nucleoplasmin

property of being highly acidic, as well as having an ability to decondense sperm chromatin. The crystal structure of dNLP has been solved, and it is reported that it is similar to *Xenopus* nucleoplasmin in that each dNLP molecule has an eight-stranded, β-barrelled topology (Namboodiri et al. 2003; Taneva et al. 2009). This degree of conservation is not surprising, as the core histones they interact with are highly conserved across eukaryotes.

The identification of NLP in T. brucei

NLP was first discovered binding to DNA using nuclear purification followed by DNA-affinity chromatography of salt-extracted nuclear proteins and mass spectrometry (K. Ersfeld, unpublished results). The bait DNA used in these experiments was the *T. brucei* 177 bp repeat sequence found predominantly on the minichromosomes (Wickstead et al. 2004). This study was able to unambiguously detect several proteins including NLP as binding to the bait DNA. Another protein that was detected was the previously identified trypanosome DNA binding protein (TDP1) (Erondu & Donelson 1992). In addition, a protein which has subsequently been confirmed as being TbISWI (Hughes et al. 2007) was identified. NLP is likely to be a DNA-binding protein due to the presence of a nucleoplasmin motif and an AT-hook domain, which are motifs also found in other chromatin-associated proteins.

Since NLP was originally found binding transcriptionally silent DNA, I investigated if it was involved in the downregulation and repression of transcription in these regions, and more specifically if it was involved in silencing telomeric *VSG* expression sites (*VSG* ESs) in the bloodstream form parasite. Here, I report that NLP is a novel nucleoplasmin-like DNA-binding protein which is essential in *T. brucei*, and is involved in maintaining appropriate levels of transcription from *VSG* ESs in both their active and their silent states.

B. Results

Sequence analysis

T. brucei NLP (accession number Tb927.10.5450) is a novel protein with little sequence similarity with non-kinetoplastid proteins, though it appears to be syntenic in *Leishmania* as well as *T. cruzi*. Database searches showed that it contains a nucleoplasmin-like domain ($e = 2.3 \times 10^{-3}$) and an AT-hook domain ($e = 9.8 \times 10^{-3}$) (MyHits; ISB-SIB). Additionally, a characteristic glutamic acid-rich region ($e = 3.8 \times 10^{-9}$) extending approximately from residues 600-670 partially overlaps the nucleoplasmin-like domain (Figure 2.1A) (PROSITE). Lastly, two coiled-coil regions extend between residues 429-458 and 569-691 (SMART, EMBL).

Using ClustalW (Larkin et al. 2007) I aligned NLP with characterised nucleoplasmins from a variety of higher eukaryotes, and were able to detect certain conserved or similar amino acids in the NLP nucleoplasmin-like domain as predicted by Hidden Markov Models (PFAM) (Figure 2.1B). Conserved residues include a leucine residue at position 557 of NLP, glycine residues at positions 560, 562 and 569 and several glutamic acid residues further in the glu-rich region which is typical of the nucleoplasmin domain, and is required for its translocation through the nuclear pore in *Xenopus* (Vancurova et al. 1997). This polyglutamate region is one of the features shared by all nucleoplasmins and is also found in *T. brucei* NLP (Figure 2.1C). The fact that NLP is approximately 800 amino acids larger than the nucleoplasmins of higher eukaryotes (with the nucleoplasmin-like region of NLP itself about 50-100 amino acids larger) and the relatively low degree of conservation (low ‘e’ value) could indicate that NLP possibly has overlapping functions.

AT-hook motifs get their name from the AT-rich regions of the minor groove of DNA that they are known to bind to, and are small motifs with a typical glycine-arginine-proline

(GRP) tri-peptide sequence which alone is necessary and sufficient to bind DNA (Reeves & Nissen 1990; Aravind & Landsman 1998). These motifs almost always occur in known or predicted DNA-binding proteins that also have other known DNA-interacting domains (Aravind & Landsman 1998). The small size of this motif makes statistically significant identification of its presence difficult. I manually scanned the putative NLP AT-hook, and were able to identify features characteristic of the motif – including the GRP core, typical lysine and arginine residues preceding the core, and other polar residues immediately downstream of GRP including a characteristic lysine residue in the fourth downstream site (Figure 2.1D) (Aravind & Landsman 1998).

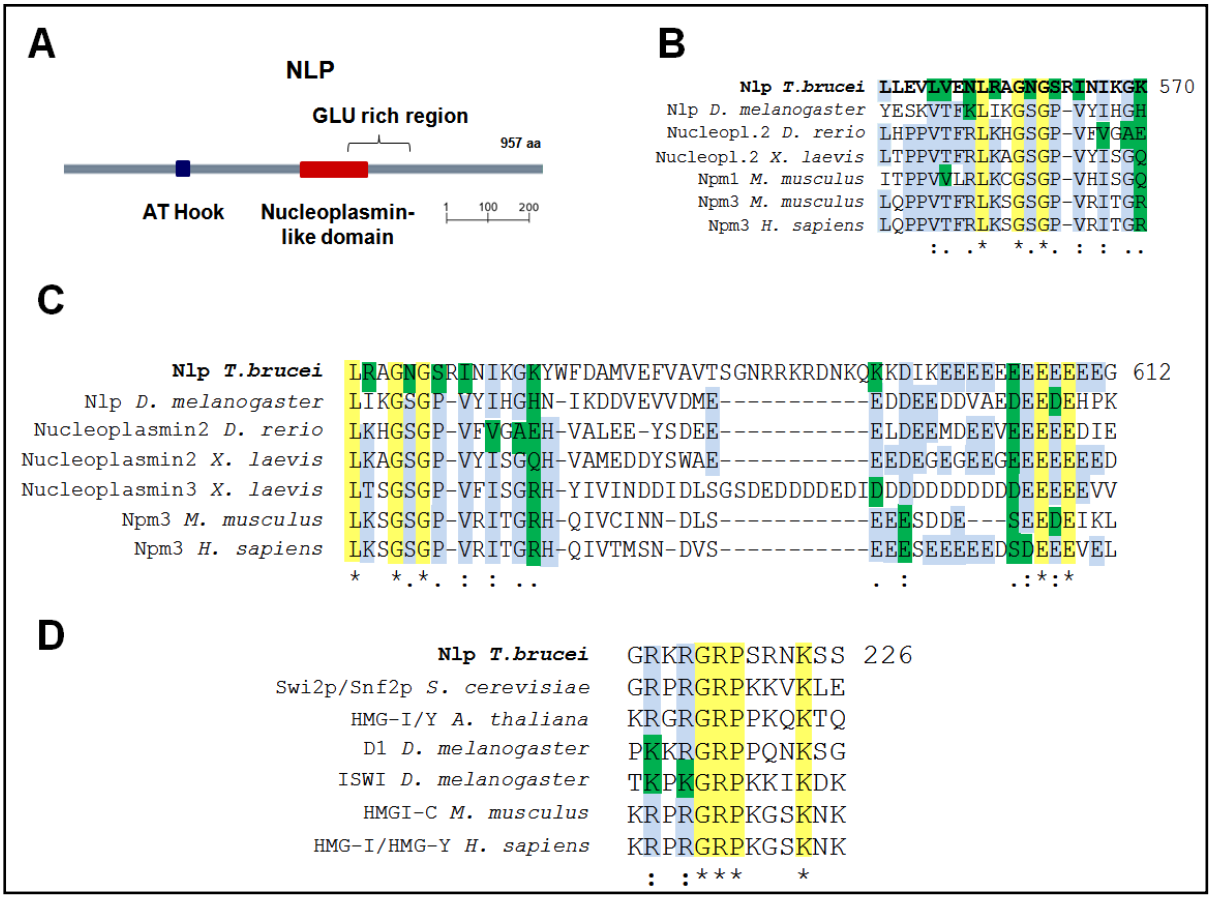


Figure 2.1 NLP is a novel nucleoplasmin-like protein in *T. brucei*. **A)** Schematic representation of the protein motifs of NLP. NLP is 957 amino acids and has an AT-hook motif extending from residues 214-226 and a nucleoplasmin domain between residues 497-

647. A highly acidic glutamate-rich region extends from residues 600-671. **B)** The core nucleoplasmin domain of NLP compared with nucleoplasmins of other eukaryotes using ClustalW alignment. Fully conserved amino acids are shown with yellow blocks and asterisks. Amino acids conserved in 5-6 sequences are shown in blue. Amino acids which are changed across sequences but share similar properties are shown in green. Double dots represent conservative changes and single dots semi-conservative changes between sequences. **C)** The core nucleoplasmin domain is followed by a polyglutamate tract in nucleoplasmins in a variety of eukaryotes including *T. brucei*. **D)** The AT-hook motif of NLP compared with AT-hook motifs in other chromatin proteins in other eukaryotes. Accession numbers of all protein sequences shown are listed in the 'Materials and Methods' chapter. Figure modified from Narayanan et al. 2010.

NLP is an essential nucleoplasmin-like protein in bloodstream form T. brucei

Given that NLP was identified as a DNA-binding protein binding to the transcriptionally inactive minichromosomes, I tested if it was involved in transcription control of the ES in bloodstream form *T. brucei*. Induction of RNAi against NLP using a tetracycline-inducible system in the two *T. brucei* clones T3-NLP1 and T3-NLP2 led to a rapid growth arrest within 24 hours (Figure 2.2A). This result was confirmed by repeating the experiment in the bloodstream form *T. brucei* 90-13 line (Wirtz et al. 1999) (Figure 2.2B). By performing Western blot analysis using an anti-NLP antibody (recombinant protein expressed by A. Fullbrook and antibody made by Eurogentec), I identified a ~120 kDa protein in both bloodstream and procyclic form *T. brucei* (Figure 2.2C). This is slightly larger than the expected size for NLP (predicted to be 107 kDa). The relatively minor difference in size could however be the consequence of post-transcriptional modifications or retarded protein electrophoretic mobility due to specific SDS-PAGE running conditions. I also observed by Western blot analysis that NLP is expressed at approximately equivalent levels in both life stages both at the protein level (Figure 2.2C) as well as at the transcript level (Figure 4.2B).

Western blot analysis also confirmed the knock-down of a protein band approximately corresponding in size to that expected for NLP, and slightly smaller than a

band known to be HA-tagged NLP (see section immediately below) (Figure 2.2D). NLP protein levels were reduced to 55% of normal levels by 48 hours and to <50% after 72 hours of *NLP* RNAi. To check if the growth arrest was related to a cell cycle arrest, I examined RNAi induced cells at various time points using the DNA stain DAPI. I did not see significant accumulation of cells at any particular cell cycle phase, although within 24 hours induction of RNAi I noted an approximately 10% increase in the number of cells with an abnormal karyotype (Figure 2.2E).

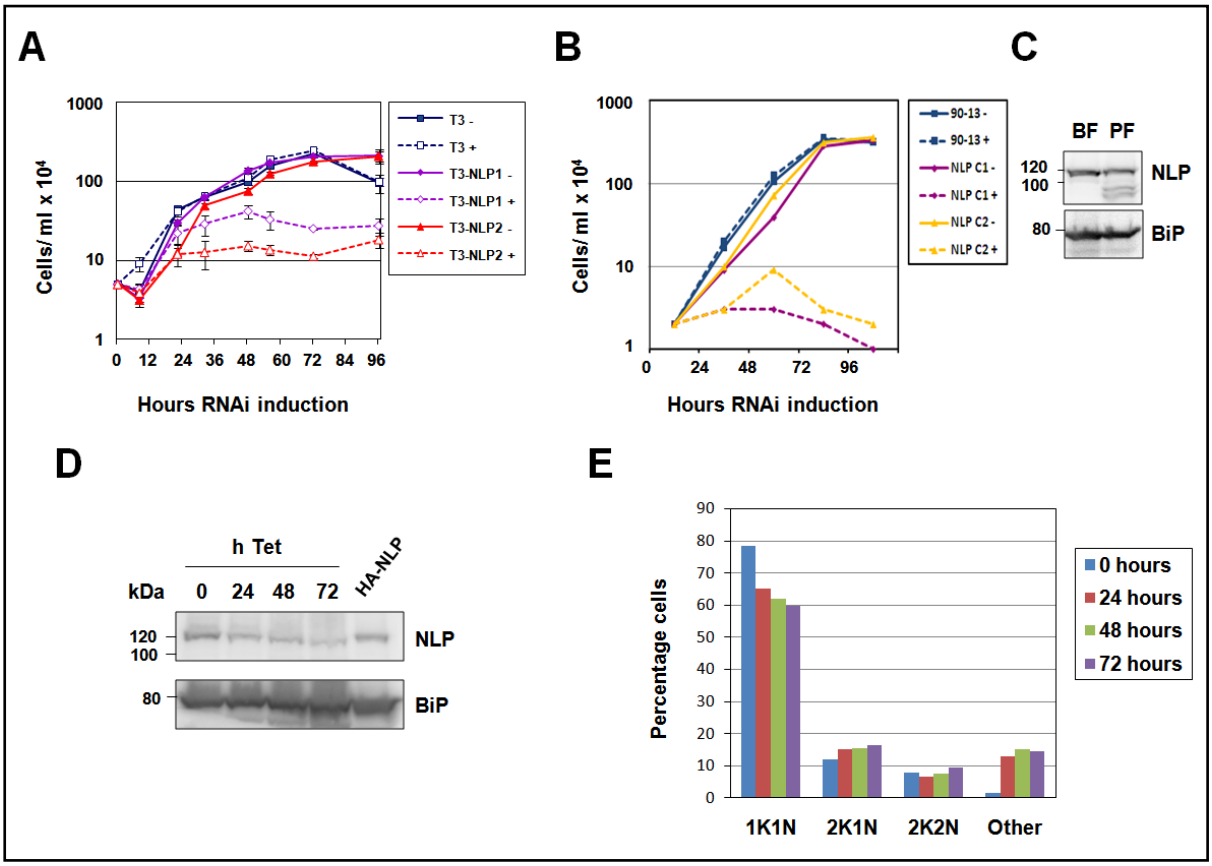


Figure 2.2 NLP is an essential protein in bloodstream form *T. brucei*. **A)** Growth kinetics of bloodstream form *T. brucei* after the induction of *NLP* RNAi. The + denotes cell lines in which tetracycline was added to induce RNAi. Growth of the two clones T3-NLP1 and T3-NLP2, and the parental cell line T3-SM lacking the *NLP* RNAi construct is plotted. Error bars denote the standard deviation from three independent repeats. **B)** A growth curve as seen in panel **A** using the bloodstream 90-13 line. Growth kinetics of the two *NLP* RNAi clones NLP C1 and NLP C2 and the parental 90-13 line lacking the *NLP* RNAi construct are plotted.

C) Western blot analysis comparing NLP protein levels in bloodstream form and procyclic form *T. brucei*. Lysate from 1×10^7 bloodstream or procyclic cell equivalents was loaded in each lane and reacted with an anti-NLP antibody. The blot was re-probed with an anti-BiP antibody to demonstrate equal loading. **D)** Western blot analysis monitoring NLP protein levels after RNAi induction. Lysates from 1×10^7 cell equivalents were prepared at each time point (indicated as hours after tetracycline addition) after RNAi induction in the T3-NLP1 cell line. An equivalent amount of lysate from a cell line expressing HA-epitope tagged NLP was also included to provide an additional size marker. The same blot was then probed with anti-BiP antibody to demonstrate equal loading between lanes. **E)** Microscopic analysis of DAPI stained bloodstream form cells was performed at various time points after the induction of *NLP* RNAi (indicated in hours). Cells were categorised according to the number of nuclei (N) or kinetoplasts (K) present and the percentage of cells in the different categories was plotted. Figure modified from Narayanan et al. 2010.

I also made *T. brucei* cell lines which had an endogenous allele of NLP epitope-tagged with a fragment of haemagglutinin (HA), for use in immunofluorescence microscopy and chromatin immunoprecipitation experiments (Figure 2.3A). Cells with one NLP allele tagged with the HA-epitope grew as well as the parental *T. brucei* GPI(VO2+) cell line. I knocked out the second NLP allele in these lines, and thereby made the cells dependent on the one HA-tagged copy of NLP. However, this did not lead to an observable change in cell growth (Figure 2.3B), indicating that addition of the HA-epitope to the carboxy terminus of NLP does not observably compromise NLP protein function. Immunofluorescence analysis of the HA-tagged NLP *T. brucei* lines showed that NLP is a nuclear protein with a relatively even distribution in the nucleus throughout the cell-cycle (Figure 2.3C).

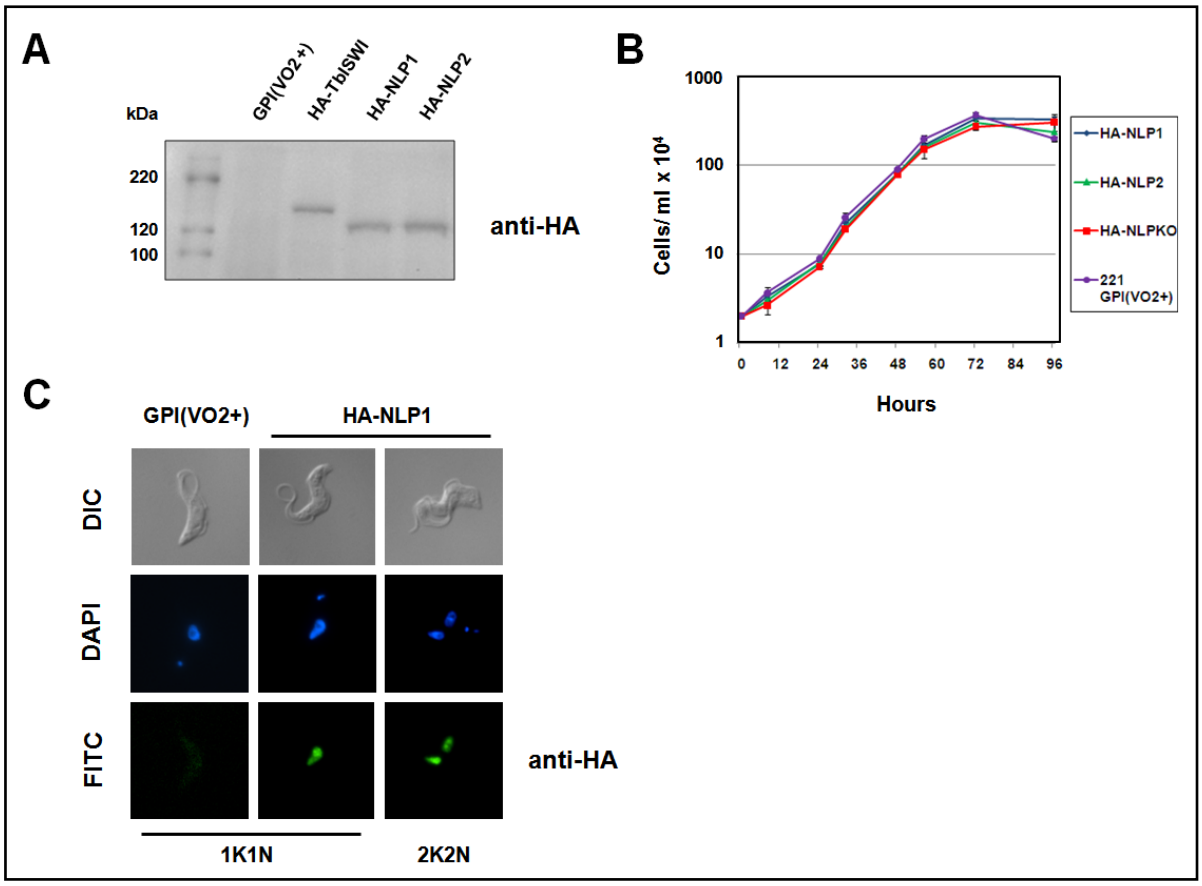


Figure 2.3 NLP is a nuclear protein. **A)** Western blot analysis of the HA-epitope tagged NLP in the bloodstream form 221GPI(VO2+) line using an α-HA antibody. Lysates from the *T. brucei* clones HA-NLP1 and HA-NLP2 and lysate from the parental 221GPI(VO2+) line are compared. Lysate from an HA-TbISWI expressing cell line (Stanne et al. 2011) was also loaded as a positive control for HA antibody specificity. Lysates from 1×10^7 cell equivalents were loaded into each lane. **B)** Graphical representation of growth kinetics of bloodstream form *T. brucei* expressing the HA-NLP protein. HA-NLP expressing clones HA-NLP1 and HA-NLP2 are compared to the parental GPI(VO2+) cells and the HA-NLPKO cell line which was derived by knocking out the second endogenous copy of NLP. **C)** Panels show parental GPI(VO2+) and HA-NLP1 cells visualized by immunofluorescence microscopy, after treatment with α-HA primary antibody and Alexa-488 secondary antibody through either differential interference contrast (DIC), DAPI, or FITC channels. NLP localization was compared at different cell cycle stages. Figure modified from Narayanan et al. 2010.

Depletion of NLP by RNAi results in derepression of silent VSG ESs as well as ES promoters inserted into other transcriptionally silent regions of the T. brucei genome

Since NLP protein had originally been identified binding to the transcriptionally inactive *T. brucei* 177 bp repeats, I investigated if it plays a role in maintaining the

transcriptionally silent state of such inactive genomic regions. I first investigated if it had a role in regulation of *VSG* ESs by using the bloodstream form T3-SM reporter cell (Hughes et al. 2007). This line has an *eGFP* gene inserted downstream of the expression site promoter (EP) of a silent *VSG221* ES, as well as the tetR and T7 RNA polymerase genes necessary for tetracycline inducible expression inherited from its parent, the 'single-marker' cell line (Wirtz et al. 1999). The T3-SM line stably expresses VSG T3 using selection with the drug blasticidin (Figure 2.4A). I induced RNAi against NLP in the T3-NLP1 line and monitored eGFP expression by flow cytometry. I observed a 45-65 fold derepression of the silent *VSG221* site within 72 hours induction of RNAi (Figures 2.4B). In contrast, I did not see ES derepression on inducing RNAi against other essential DNA-binding proteins such as TDP1 (Figure 4.3). This indicates that the increased *eGFP* expression from the silent ESs was not simply an artefact due to cell death or a stress-response caused by unavailability of an essential protein. I also induced RNAi against NLP in the 121-SM bloodstream form reporter line. This line is analogous to the T3-SM line, however activation of the *VSG121* ES (instead of the *VSGT3* ES) can be maintained using blasticidin selection (Hughes et al. 2007) (Figure 2.5A). I observed a comparable 19-52 fold derepression of the *eGFP* gene in this line (Figure 2.5B).

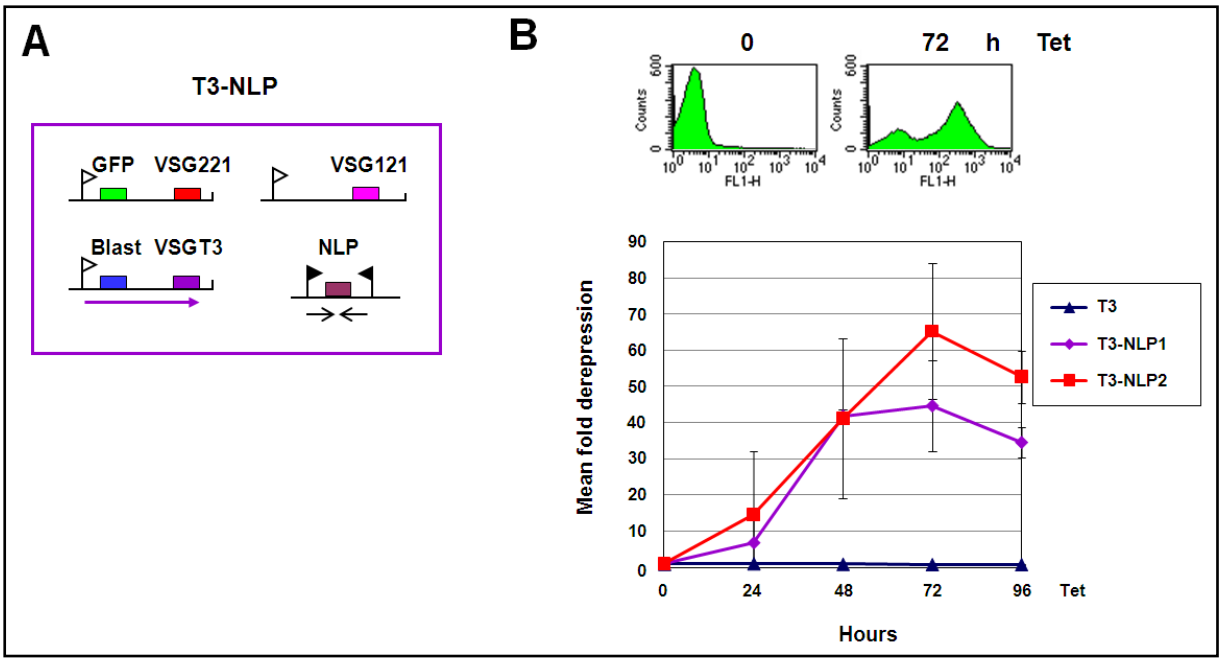


Figure 2.4 RNAi mediated depletion of NLP leads to a 45-65 fold derepression of silent telomeric ESs. **A)** Schematic representation of the bloodstream form T3-NLP line. Cells expressing *VSG T3* are selected for using blasticidin selection. An *eGFP* gene is inserted into the silent *VSG221* ES. RNAi can be induced against NLP by activating the tetracycline inducible T7 promoters shown by black flags. **B)** Representative flow cytometry traces in the FL1 channel at 0 and 72 hours after RNAi induction, and below, a graphical representation of the same data showing fold derepression at each time point (calculated by dividing mean fluorescence at a given time point with that of the 0 hour time point). Error bars represent standard deviation from three independent repeats. Figure modified from Narayanan et al. 2010.

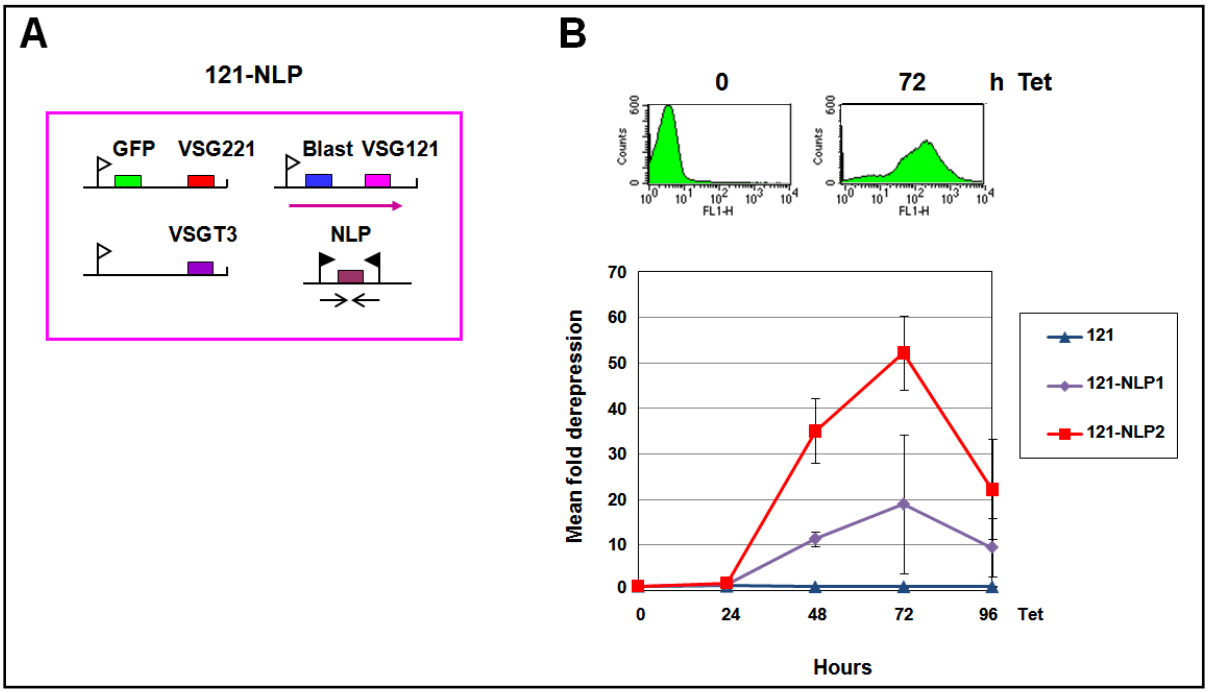


Figure 2.5 RNAi mediated depletion of NLP leads to 19-52 fold derepression of silent telomeric ESs in the 121-NLP line. **A)** Schematic representation of the bloodstream form 121-NLP line. Cells expressing *VSG 121* are selected for by blasticidin resistance. An *eGFP* gene is inserted into the silent *VSG221* ES. RNAi can be induced against NLP by induction of the tetracycline inducible T7 promoters shown by black flags. **B)** Representative flow cytometry traces in the FL1 channel at 0 and 72 hours after RNAi induction, and below, a graphical representation of the same data showing fold derepression at each time point (calculated by dividing mean fluorescence at a given time point with that of the 0 hour time point). Error bars represent standard deviation from three independent repeats. Figure modified from a supplementary figure in Narayanan et al. 2010.

Judging by the relatively even distribution of NLP throughout the nucleus in the immunofluorescence experiments, I hypothesised that NLP could play a role in maintaining a transcriptionally silent state in genomic areas other than just the telomeric *VSG* ESs. As has been discussed in the previous chapter, of the three *T. brucei* RNA polymerase I promoters investigated, the core ES promoter is the most sensitive to its chromosomal context. For example, an exogenous ES promoter is more comprehensively silenced than a comparable rDNA promoter, when inserted in the transcriptionally silent VSG Basic Copy arrays (Horn et al. 1997). This observation allowed us to use the ES promoter as a "probe" for repressed chromatin regions (Urményi & Van der Ploeg 1995). Using this approach, Manish Kushwaha in the Rudenko lab prepared a series of reporter cell lines which had a construct carrying an ES promoter and flanking *eGFP* gene, or just an *eGFP* gene without an ES promoter. These constructs were inserted into transcriptionally silent areas in the *T. brucei* genome such as the minichromosomes, the VSG Basic Copy arrays and the procyclin locus in bloodstream form cells (Stanne et al. 2011).

Given that NLP had initially been found binding to minichromosomes, I investigated the role of NLP in silencing these region, which are comprised primarily of arrays of 177-bp simple repeat sequences (Wickstead et al. 2004). I used the first of these reporter lines, which contains a construct with an ES core promoter upstream of *eGFP* inserted into the

T. brucei minichromosomes (the MCEP reporter line) (Stanne et al. 2011). A low basal rate of transcription from the ES promoter inserted into a transcriptionally silent minichromosome is seen even in parental cells which do not have an RNAi construct inserted into them (also seen in uninduced *NLP* RNAi cells in the flow cytometry trace for time point 0 in Figure 2.6A). Induction of *NLP* RNAi resulted in 5-6 fold derepression of *eGFP*, indicating that NLP plays a role in silencing the minichromosomal 177 bp repeat regions. Similarly, I induced *NLP* RNAi in a reporter line identical to MCEP except that it lacked the ES promoter upstream of the *eGFP* gene (MCe), and tested for derepression by flow cytometry. Depletion of *NLP* in this cell line did not result in an observable increase in *eGFP* transcription (data not shown). This indicates that although blocking the synthesis of NLP appears to facilitate transcription from the ES promoter, it does not lead to a significant increase in fortuitous initiation of transcription from cryptic promoters at minichromosomes.

Horn et al. (1997) had shown that the ES promoter is particularly repressed if integrated into the VSG Basic Copy arrays of bloodstream form *T. brucei*. The bloodstream form *T. brucei* line VBEP contains an ES promoter upstream of *eGFP* integrated upstream of the silent *VSG118* Basic Copy gene (Figure 2.6B) (Stanne et al. 2011). Similar to the MCEP line, the depletion of NLP by RNAi induction led to nearly 5 fold derepression of *eGFP* within 72 hours (Figure 2.6B). The VBe line is a comparable reporter line which contains a similar construct integrated into the VSG Basic Copy array but without an ES promoter upstream of *eGFP*. As expected there was no observable increase of *eGFP* transcription after induction of *NLP* RNAi. Therefore, NLP appears to be involved in maintaining the repressive chromatin environment silencing the exogenous ES promoter inserted into the VSG Basic Copy arrays. Its depletion however, does not cause an observable increase in fortuitous initiation of transcription from cryptic promoters in these arrays (data not shown). I did not observe derepression of eGFP after induction of *TDP1* RNAi in the MCEP and VBEP cell

lines, indicating that the observed derepression of *eGFP* after the induction of *NLP* RNAi is not simply an artefact of cell lethality (Figures 4.4 and 4.5).

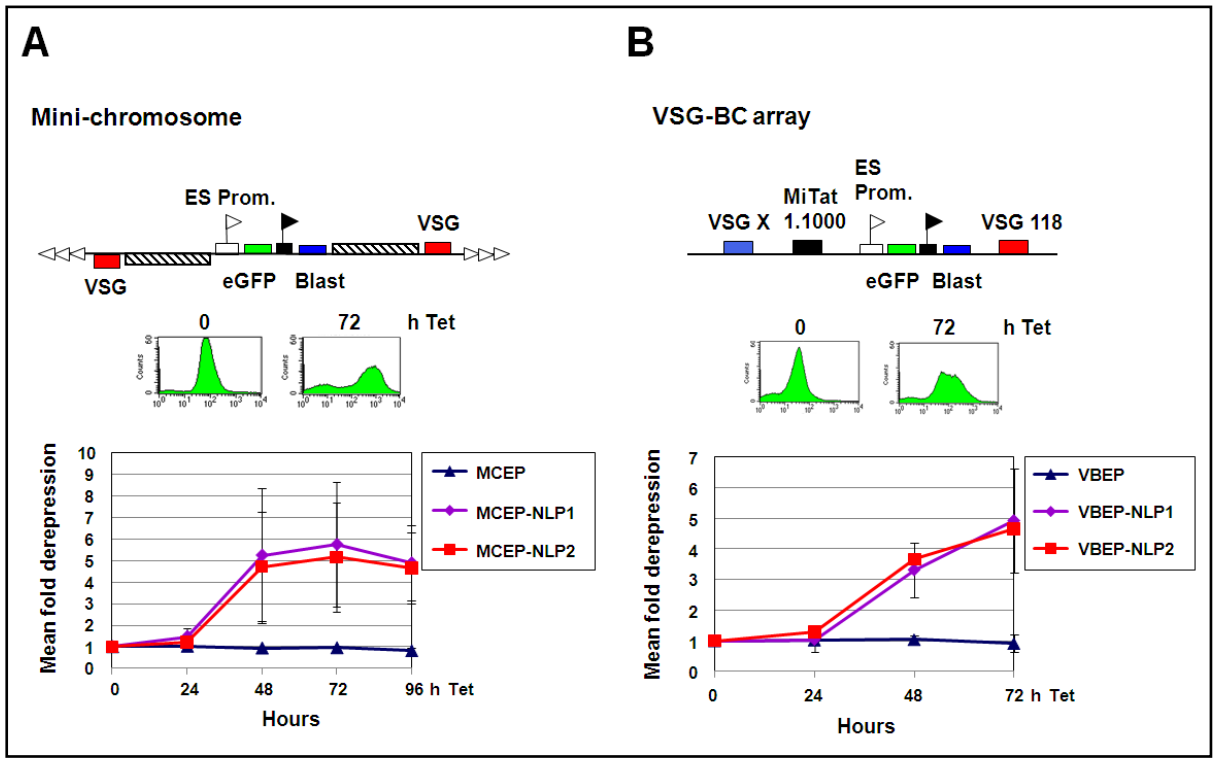


Figure 2.6 NLP is involved in silencing an ES promoter inserted into transcriptionally silent areas of the *T. brucei* genome. **A)** The upper panel shows a schematic representation of the reporter construct integrated into the 177 bp minichromosome repeats in bloodstream form *T. brucei*. The MCEP line (Mini-chromosome) contains a reporter construct with an ES promoter upstream of an *eGFP* gene inserted into the silent minichromosome sequence repeats. The 177 bp repeats are shown with striped boxes, telomeres with triangles, *VSG* genes with red boxes, the blasticidin resistance gene with a blue box and the *eGFP* gene with a green box. The ES promoter is shown with a white flag and the rDNA promoter with a black flag. Middle panels show representative traces from flow cytometry experiments at the 0 and 72 hour time points after the induction of *NLP* RNAi (h Tet). The lower panel shows quantification of FL1 values from 3 independent experiments expressed as mean fold derepression at each time point. **B)** Data from the VBEP line (VSG-BC array) reported in the same format as in **A**. VBEG has the reporter construct inserted into the inactive VSG Basic Copy array. In both cases cells capable of *NLP* RNAi are compared with the parental MCEP or VBEP lines respectively. Figure modified from Narayanan et al. 2010.

Finally, I investigated the role of NLP in silencing procyclin genes, which are largely maintained in a transcriptionally inactive state in bloodstream form *T. brucei* (Wickstead et

al. 2004). I used the EP1 cell line, which has a reporter construct containing an EP1 promoter and an *eGFP* gene integrated behind a silent EP1 promoter in bloodstream form *T. brucei* (Stanne et al. 2011). Procyclin promoters in bloodstream form *T. brucei* are not completely silenced but transcribe very low levels of mRNA, which explains the observed *eGFP* fluorescence in the flow cytometry trace at time point 0 in Figure 2.7. I find that depleting *NLP* by induction of RNAi leads to an approximately 4 fold increase in transcription of *eGFP* driven by the previously silent EP1 promoter (Figure 2.7), which is comparable to the derepression seen in the MCEP and VBEP reporter lines. To summarize, blocking the synthesis of NLP had the most pronounced effect on derepression of silent *VSG* ESs in bloodstream form *T. brucei.* However, NLP also appears to play a role in establishing a repressive transcriptional environment for the ES promoter regardless of its genomic location. In our bloodstream form *T. brucei* reporter cell lines where the ES promoter was inserted into other transcriptionally silent genomic locations such as the minichromosomes and the VSG Basic Copy arrays I observed a smaller but significant 5 fold derepression of *eGFP*. Additionally, I show that the inactive EP1 promoter in the silent procyclin transcription units is also derepressed about 4 fold after depletion of NLP.

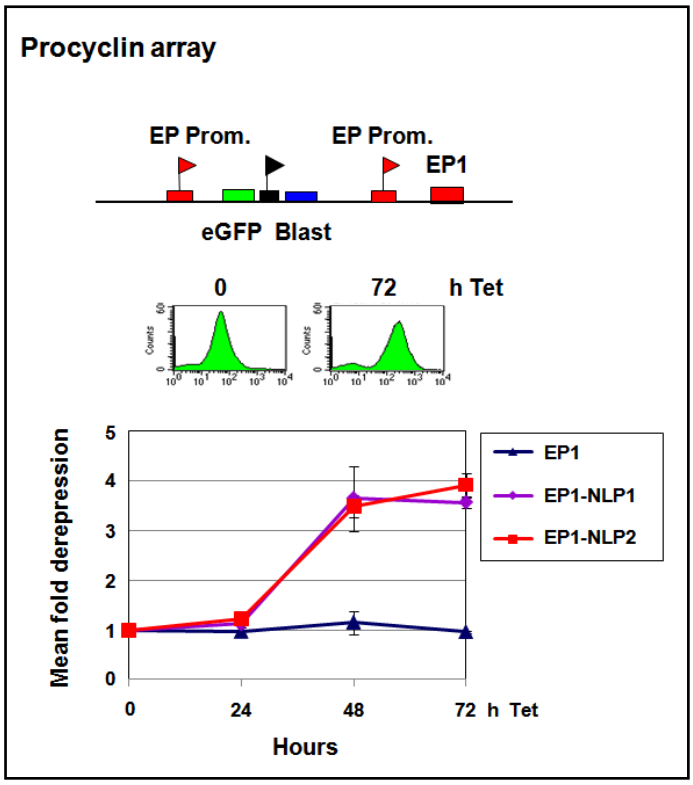


Figure 2.7 Depletion of NLP by RNAi leads to ~4 fold derepression of an endogenous procyclin EP1 promoter in bloodstream form *T. brucei*. Schematic representation of the reporter construct containing an *eGFP* gene downstream of an EP1 promoter integrated into the procyclin locus in the EP1 cell line. Representative flow cytometry traces at 0 and 72 hours after the induction of *NLP* RNAi with tetracycline (h Tet) are shown below. The same data is quantified below. EP1-NLP1 and EP1-NLP2 lines are compared to the parental EP1 reporter line. Error bars represent standard deviation from three independent repeats. Figure modified from Narayanan et al. 2010.

NLP is more abundant in certain non-transcribed regions of the T. brucei genome

Since our results thus far strongly implicated NLP as a DNA-binding protein, I further investigated if there was a differential distribution of NLP within different areas of the *T. brucei* genome. I performed chromatin immunoprecipitation (ChIP) on the bloodstream form HA-NLP1 cell line in which the carboxy terminus on one NLP allele had been epitope-tagged with the HA epitope (Figures 2.3A-C).

I cross-linked chromatin using formaldehyde, and then sheared it into roughly 200 bp fragments by subjecting it to sonication. I next used an antibody against the HA-tag to immunoprecipitate HA-NLP cross-linked to DNA. As a positive control, I also simultaneously performed a ChIP experiment with an antibody against histone H3, or as a negative control performed ChIP without using any antibody. The ChIP procedure was also carried out simultaneously on the parental *T. brucei* 221GPI(VO2+) cells which lack HA-NLP to provide an additional negative control for specificity of the anti-HA antibody.

I first investigated the distribution of NLP on the 177 bp minichromosome repeat sequences, as NLP had been found binding there, and since our MCEP reporter cell line revealed that NLP could play a role in this region. Slot blot analysis confirmed that NLP does indeed bind to the 177 bp repeats (Figure 2.8A). Significant levels of NLP were also found binding the 50 bp repeats which extend as large arrays upstream of all known *T. brucei* ESs (Figure 2.8A). Quantitative PCR (qPCR) was subsequently performed to determine the presence of NLP in various other genomic regions. qPCR analysis was performed for each genomic location using a primer pair that produced a single amplicon of ~200 bp. Input chromatin (diluted to 10% of total input) was analysed as well as immunoprecipitated material from the HA-NLP1 and parental lines treated with anti-HA or no antibody. A complete list of primer pairs used in these experiments is shown in Appendix B. Schematic representations of the various transcription units investigated are shown in Figure 2.8B-C and 2.8E-F. I found NLP at the spliced leader (SL) genes transcribed by RNA polymerase II (Aksoy et al. 1990). Comparable levels were found at the intergenic regions, SL promoters, and the SL genes themselves (Figure 2.8D). I also found relatively high amounts of NLP at the 5S rRNA transcription units, which are transcribed by RNA polymerase III, both in the intergenic spacers and within the genes themselves (Figure 2.8D).

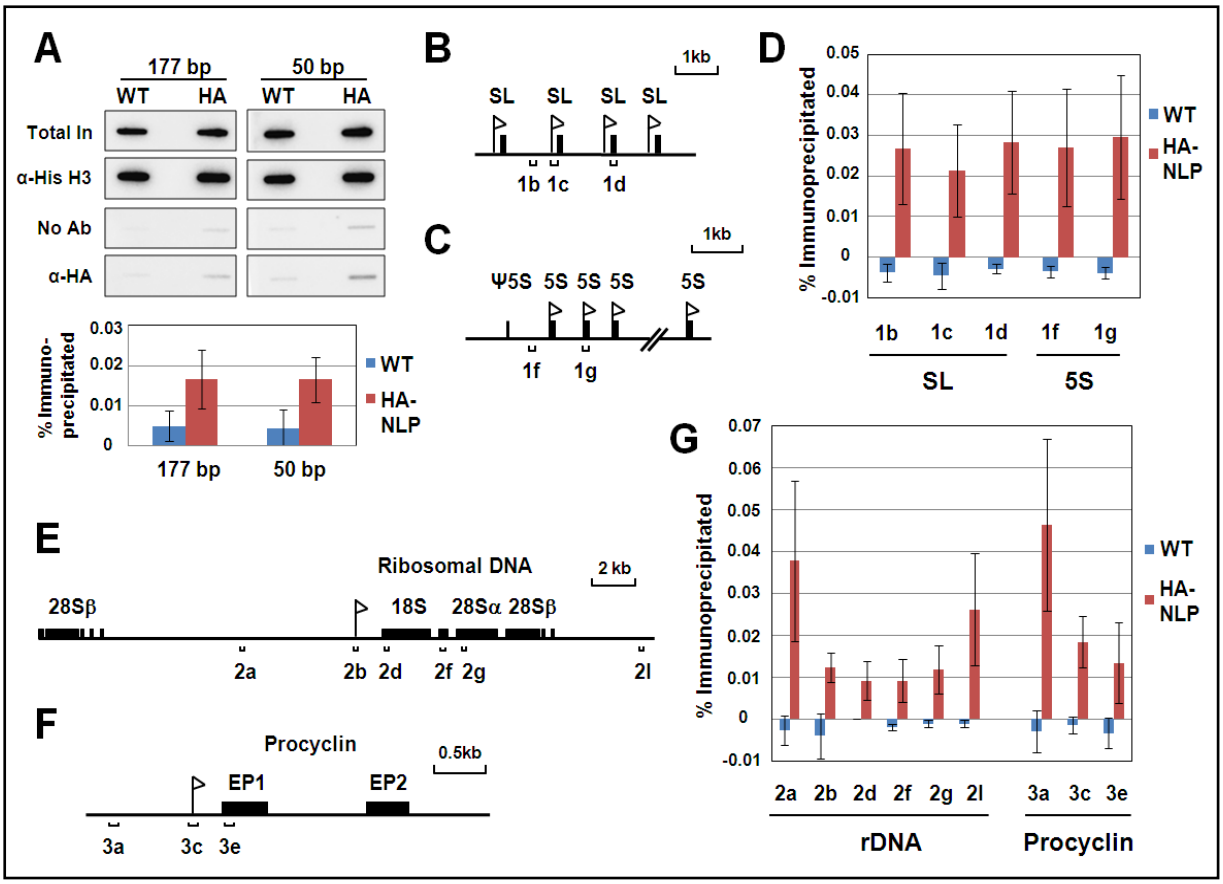


Figure 2.8 NLP is enriched in various transcriptionally silent areas in *T. brucei* including the simple sequence repeats and intergenic spacer loci. The bloodstream form HA-NLP line and its parental line (WT) were used in all experiments. Immunoprecipitation was performed with an anti-HA antibody, and all values are shown as % immunoprecipitated after subtraction of the values obtained for an identical ChIP experiment performed without the antibody. Primer pairs used in qPCRs are shown in schematic figures with lettered brackets. Corresponding graphs show % immunoprecipitated values obtained in qPCRs with those primer pairs. Error bars indicate standard error from three independent repeat ChIP experiments. **A)** ChIP followed by slot-blotting and probing with radio-labelled 177 bp or 50 bp probes. ChIP material from WT and HA-NLP cells is compared. Material from the total immunoprecipitation sample (Total In), anti-histone H3 (α-His H3), anti-HA (α-HA) and the no antibody control (No Ab) are shown. **B-C and E-F)** Schematic representations of the spliced leader RNA transcription units, the 5S rRNA transcription units, the rDNA locus, and the procyclin EP1 locus respectively. Flags represent promoters and black boxes show transcribed genes. **D and G)** Quantification of ChIP results by qPCR using the WT and HA-NLP lines. Figure modified from Narayanan et al. 2010.

I also investigated RNA polymerase I transcribed areas such as the rDNA and procyclin loci, as well as the *VSG* ESs (Gunzl et al. 2003). I first investigated the rDNA

locus, and found relatively low levels of NLP within the rDNA transcription unit itself (primers 2b, 2d, 2f and 2g in Figures 2.8E, Figure 2.8G). However, NLP binding was enriched in the adjacent non-transcribed rDNA spacer both upstream as well as downstream of the rDNA transcription unit (primers 2a and 2l; Figures 2.8E, Figure 2.8G). Levels of NLP were over 3 fold higher in the intergenic region using primer pair 2a or nearly 3 fold using primer pair 2l compared with levels in the 18S or 28S rDNA transcription unit itself. Similarly, in the procyclin locus, approximately 3 fold higher levels of NLP were found upstream of the procyclin *EP1* and *EP2* transcription units (primer 3a in Figures 2.8F, Figure 2.8G) compared to within the *EP1* gene itself (primer 3c and 3e). To summarize, NLP appears to be present at higher levels in the flanking transcriptionally silent regions compared to within the transcription units themselves in RNA polymerase I transcribed areas in the *T. brucei* genome.

Since our results using the T3-SM line had shown that the knock down of NLP led to the derepression of silent ESs, I next investigated binding of NLP to ES sequences (Figure 2.9A). ESs are highly similar to each other in sequence (Hertz-Fowler et al. 2008), and therefore primers can be expected to bind most if not all ESs. In these ChIP experiments I used the HA-NLP1 line which has exogenous marker genes inserted in both the active and in a silent ES, allowing us to distinguish between the two. The HA-NLP1 line contains a single copy neomycin resistance gene immediately downstream of the promoter of the active *VSGVO2* ES, and a construct containing a puromycin resistance gene and an *eGFP* gene downstream of the silent *VSG221* ES promoter (Figure 2.9B). The *VSG221* gene itself is an additional single copy sequence present within the silent *VSG221* ES, while the *VSGVO2* gene in the *VSGVO2* ES can be distinguished from a *VO2* gene elsewhere in the genome using our primer pair 4m (Figure 2.9B), which produces an amplicon overlapping the distal end of the *VO2* gene and the adjacent untranslated region (UTR). However, all other ES

primers used ("All ESs" in Figure 2.8) can be expected to recognise the active as well as most if not all of the silent ESs. I was able to find a small amount of NLP binding within the ES promoter region. However I could not distinguish between active and silent ESs with our primers. qPCR analysis using the single copy sequences specific for either the active or silent ES showed NLP binding to both silent and active ESs (Figure 2.9C).

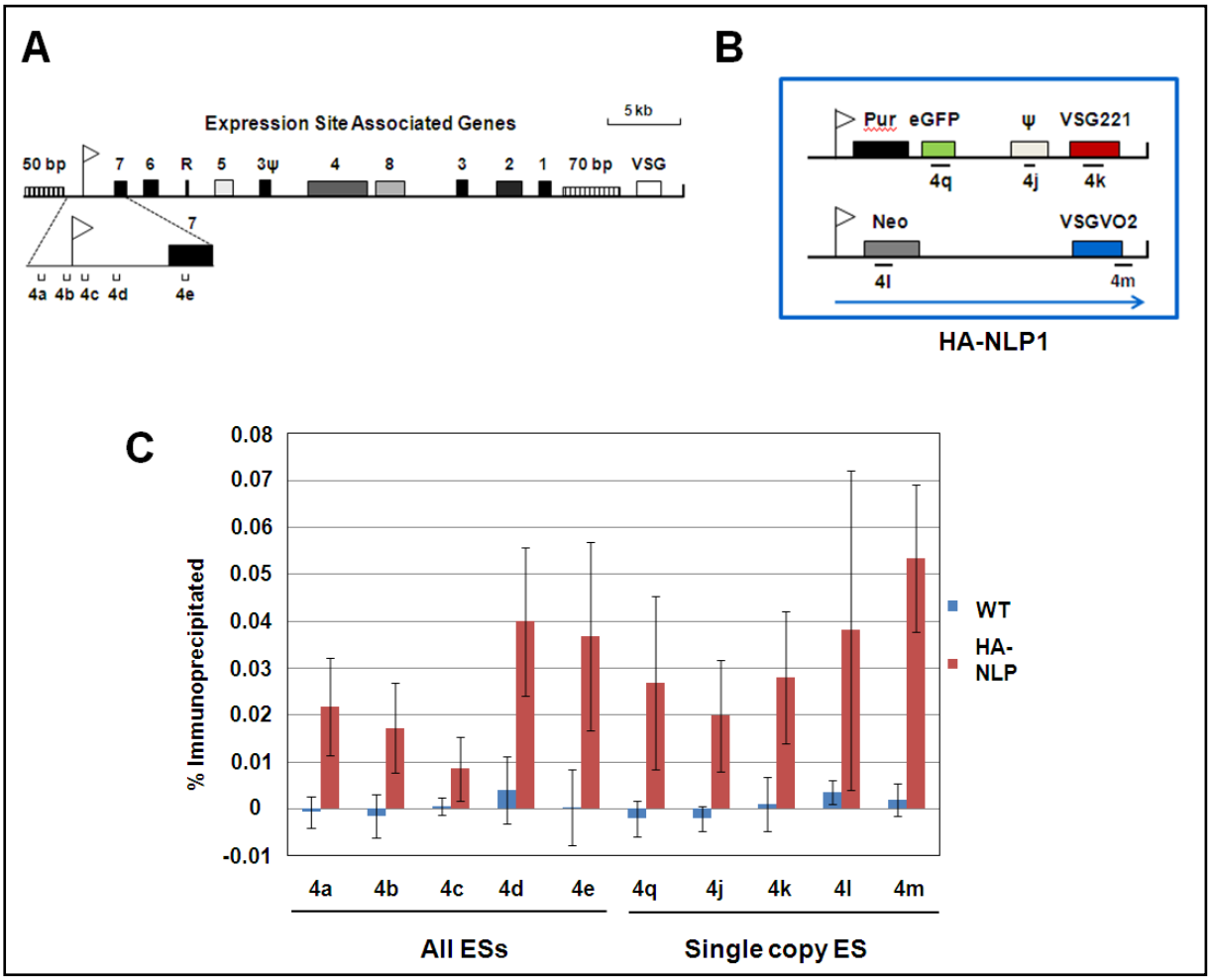


Figure 2.9 NLP binds to both active and silent *VSG* ESs but is comparatively depleted at ES promoters. **A)** Schematic representation of a typical *VSG* ES. The promoter is denoted by a flag and the various ES associated genes with boxes. Primers used in qPCRs are shown by lettered brackets, and can be expected to recognize most if not all ESs in the *T. brucei* repertoire. **B)** Schematic representation of the bloodstream form HA-NLP1 line used in ChIP experiments. The cells are selected for maintenance of activation the *VSGVO2* ES using neomycin selection. An *eGFP* gene is present in the silent *VSG221* ES. Primer pairs shown here are specific to these regions, and allow us to distinguish between active and silent ESs. **C)** Quantification by qPCR of ChIP at the above loci. Those primer pairs which recognize all ESs and those which recognize only a single ES are indicated below. Error bars show the

standard deviation from three independent ChIP experiments. Figure modified from Narayanan et al. 2010.

Depletion of NLP leads to reduced transcription of the active VSG ES

I was able to detect 45-65 fold derepression of the reporter *eGFP* gene which is located immediately downstream of the ES promoter after RNAi-mediated depletion of NLP. Hence I wanted to know if the derepressed state led to fully processive transcription extending throughout the ES down to the telomere end including the *VSG* gene. I thus monitored levels of silent *VSG* transcripts by performing reverse transcriptase (RT) qPCR on RNA samples extracted from T3-NLP1 cells at various time points after the induction of RNAi. I first confirmed our earlier Western blot analysis and flow cytometry results by establishing that there was a significant decrease in levels of *NLP* and increase in levels of *eGFP* transcript within 24 hours of RNAi induction. *NLP* transcript was reduced down to ~25% of normal levels by 48 hours, while the *eGFP* transcript increased nearly 20 fold after 72 hours of RNAi induction against NLP (Figure 2.10).

I next determined if the striking derepression of the *VSG221* ES promoter observed by monitoring *eGFP* transcript was limited to the promoter proximal region, or extended to the telomere end of the silent ES where the *VSG* gene is located. I therefore performed qPCR analysis on six *VSG* transcripts (*VSG 13, VSG 121, VSG 224, VSG 800, VSG bR2* and *VSG JS1* (Aitcheson et al. 2005; Hertz-Fowler et al. 2008)) which are genes which are all present in various silent ESs. I found no significant increase in the *VSG* transcripts at any time point. However, the level of *VSG* transcripts from silent ESs was markedly lower after the induction of *NLP* RNAi for 48 hours compared with uninduced cells. There is some increase in levels of these transcripts at 72 hours, but this is likely to be a consequence of the emergence of revertants somehow escaping a lethal RNAi phenotype. This is indicated in the levels of *NLP*

transcript, which increase by about 2 fold from the level at 48 hours to that at 72 hours after RNAi induction.

Strikingly, I observed from these experiments that the level of the active *VSG T3* transcript itself fell drastically to 4% of normal levels after 48 hours of induction of *NLP* RNAi. I also monitored levels of transcript from the blasticidin resistance gene which is also present in the active *VSGT3* ES. Levels of this transcript also fell to approximately 50% after 48 hours and subsequently to about 30% of original levels after 72 hours of *NLP* RNAi. This indicates that NLP plays an important role in regulating appropriate levels of transcription, not only from silent ESs but also from the active ES itself. NLP appears to have a role in facilitating fully processive transcription of the ES. I see that levels of expression of *VSG T3* which is located at the telomeric end of the ES, decrease more dramatically after induction of *NLP* RNAi than do levels of expression of the blasticidin resistance gene, which is located closer to the ES promoter. The experiment was performed both in the presence or absence of blasticidin selection, which maintains selection for activation of the *VSGT3* ES, without seeing significant differences (data not shown). The observed lack of processive transcription also provides an explanation as to why transcripts from silent *VSG* ESs also appear reduced after 48 hours compared to levels in uninduced cells.

I also investigated transcripts derived from the RNA polymerase I transcribed EP1 procyclin genes. These increased ~5 fold after the induction of a block in NLP synthesis (Figure 2.10), as expected from the flow cytometry experiment which showed derepression of the EP1 procyclin promoter (see Figure 2.7) after induction of RNAi. All these experiments were performed in triplicate, and transcript levels are presented after normalisation to the amount of actin transcript present at each time point. The level of actin transcript quantified at each time point after the induction of *NLP* RNAi was not found to change significantly,

and is also shown as a reference point (Figure 2.10). Levels of transcripts from two other RNA polymerase II transcribed genes namely the spliced leader and the RNA polymerase I large subunit also did not show any observable change and are shown in Figure 2.10.

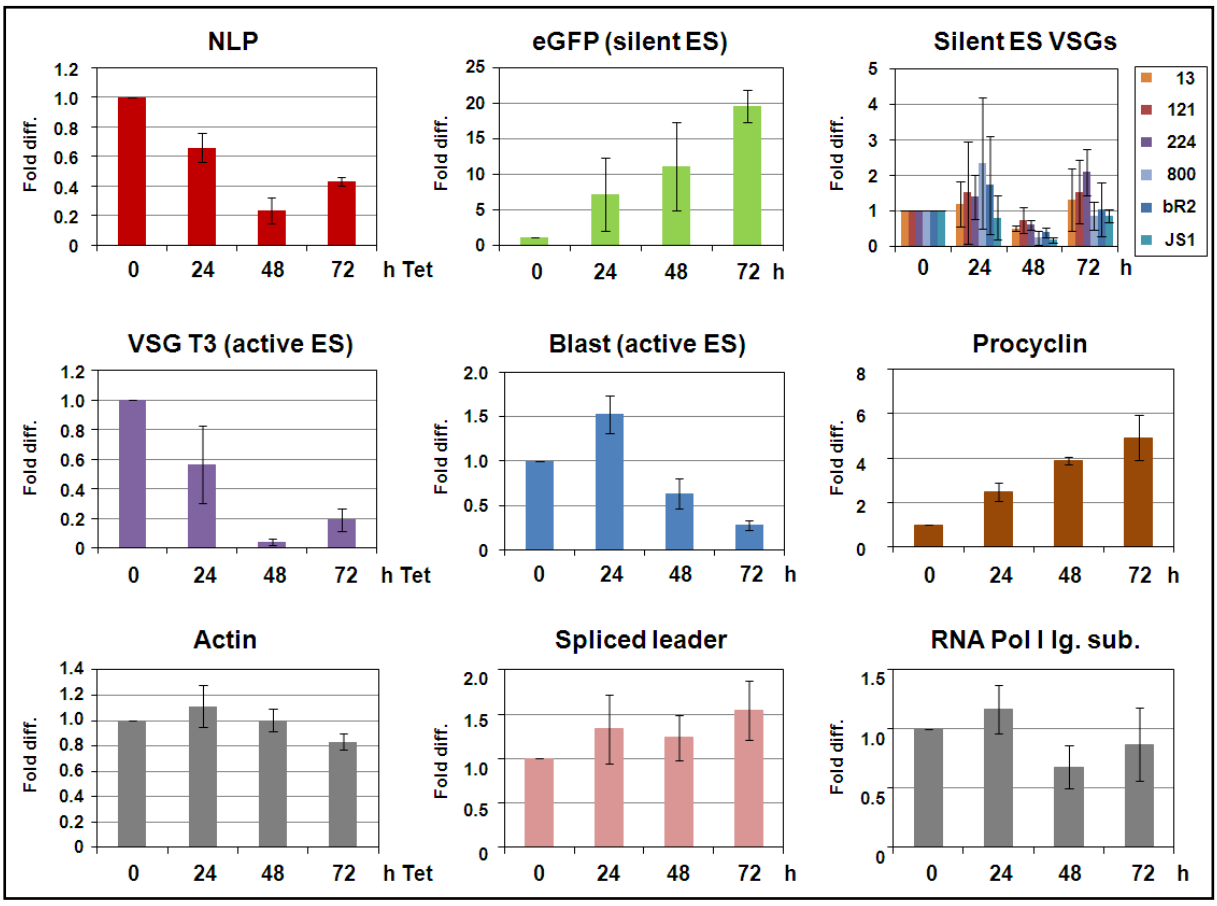


Figure 2.10 Quantification of various transcripts after depletion of NLP in the T3-NLP line for the time indicated in hours. RNA was extracted from cells in which RNAi had been induced at each of the given time points (h Tet) and cDNA was prepared. qPCR was then performed to monitor changes in the levels of various transcripts. All data shown have been normalized to the level of actin transcript at each time point, and is plotted as fold difference between induced and uninduced time points. The level of actin transcript (to which other data has been normalized) is shown in this figure, with the exception of data for the procyclin panel which was normalized to an independent quantification of actin (data not shown). Error bars show standard deviation from three independent repeat experiments. The transcripts quantified in this figure include those of *NLP*, *eGFP* (located in the silent *VSG221* ES), a selection of six silent VSGs located at telomeric ESs (*VSGs 13*, *121*, *224*, *800*, *bR2*, *JS1*), *VSG T3* (in the active ES), blasticidin resistance gene (also in the active ES), procyclin (EP1) gene, actin, spliced leader and RNA polymerase I large subunit. Figure adapted from Narayanan et al. 2010.

An identical experiment was subsequently performed using the analogous bloodstream form *T. brucei* cell line 121-SM which expresses *VSG 121* rather than *VSG T3* from the active ES. I had already observed that the level of derepression in these cells was lower than that in the T3-SM line when measured by flow cytometry (Figure 2.4), and had hypothesized that this was due to a lower degree of NLP knock-down in this cell line. I found this to be the case when I analysed NLP transcript levels by qPCR (Figure 2.11). Additionally, levels of *eGFP* transcript showed a correspondingly smaller increase while the levels of the active *VSG T3* transcript showed a correspondingly smaller decrease (Figure 2.11).

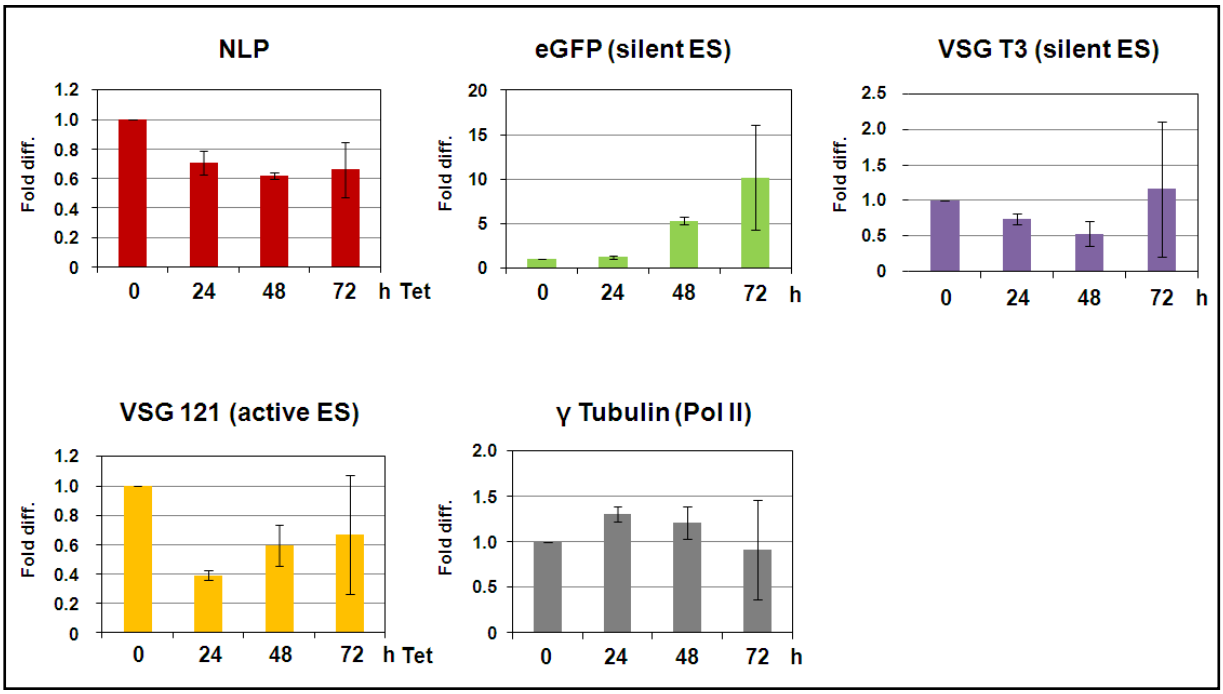


Figure 2.11 Quantification of various transcripts after depletion of NLP in the 121-NLP line for the time indicated in hours. RNA was extracted from cells in which RNAi had been induced at each of the given time points (h Tet) and cDNA was prepared. qPCR was then performed to monitor changes in the levels of various transcripts. All data shown have been normalized to the level of γ-tubulin transcript at each time point, and is plotted as fold difference between induced and uninduced time points. The values of γ-tubulin to which other data was normalized is also shown in this figure. Error bars show standard deviation from three independent repeat experiments. The transcripts shown in this figure include those of *NLP*, *eGFP* (located in the silent *VSG221* ES), *VSG T3* (in a silent ES), *VSG 121* (in the active ES), and γ-tubulin. Figure adapted from Narayanan et al. 2010.

C. Discussion

I have identified NLP as a novel DNA-binding protein in *T. brucei*. Analysis of its protein sequence identified certain key residues in common with the histone chaperone nucleoplasmins of higher eukaryotes ranging from *Xenopus* to humans, leading us to classify it as a novel nucleoplasmin-like protein. As expected from its identification as a DNA-binding protein and from its similarity to nucleoplasmins, NLP localizes to the nucleus of cells, is associated with chromatin, and is enriched in certain transcriptionally silent regions. The depletion of NLP in bloodstream form cells leads to a rapid growth arrest and about 45-65 fold derepression of silent ESs, and ~5 fold derepression of ES promoters inserted into other transcriptionally silent areas of the genome. Although NLP depletion does not lead to complete activation of silent ESs, I report a striking decrease in transcription of the *VSG* gene in the active ES. This suggests that NLP is involved in two levels of ES control: in maintaining transcriptional processivity of the ES to the telomere end where the *VSG* gene is located, and in downregulating transcription at the promoter proximal region to basal levels.

NLP is a nucleoplasmin-like protein

NLP does not appear to be closely related to any known chromatin remodelling protein in other eukaryotes, and is the latest addition to a small but growing list of proteins involved in ES control in *T. brucei*. Its most distinctive feature is a nucleoplasmin motif, which is a DNA-binding motif that derives its name from the vertebrate DNA binding protein of the same name. However, the nucleoplasmin domain of NLP on its own is similar in size to the typical vertebrate nucleoplasmin protein. Vertebrate nucleoplasmins are small (~150-300 amino acids long) acidic proteins which function as histone chaperones (Frehlick et al. 2007; Eitoku et al. 2008). The first nucleoplasmin to be identified was found in *Xenopus*

oocytes as a highly abundant protein which was capable of assembling nucleosome cores in vitro from purified histones and DNA (Earnshaw et al. 1980). Its main function in *Xenopus* oocytes appears to be to mediate sperm chromatin decondensation, in which it binds and removes sperm basic proteins replacing them with histones H2A and H2B (Philpott et al. 1991; Laskey et al. 1993). Additionally, de-phosphorylated nucleoplasmin is also important for regulating the chromatin condensation that occurs during apoptosis (Lu et al. 2005). Little is known about the non-vertebrate nucleoplasmins, which are usually categorized as 'nucleoplasmin-like' (Frehlick et al. 2007). The *Drosophila* nucleoplasmin-like protein (dNLP) remains the best characterized of this class, and is known to bind to core histones, and function in assembling regularly spaced nucleosomal arrays (Ito et al. 1996).

It is therefore not immediately clear if NLP could be included in a category with typical nucleoplasmins such as those present in the vertebrates. Typical nucleoplasmins are small phosphoproteins which sometimes function as pentamers (Taneva et al. 2009). However, NLP at nearly 1000 amino acids is more than five times larger than a typical nucleoplasmin. Database searches show that nucleoplasmins are seldom larger than 150-280 amino acids, which is comparable to the region within NLP identified as a nucleoplasmin-like domain (150 amino acids long). It is possible that NLP represents an evolutionarily early form of nucleoplasmin. Additionally, NLP contains an AT-hook domain which is not found in typical vertebrate nucleoplasmins. AT-hook motifs are small (approximately eleven amino acid) DNA binding motifs centred around a glycine-arginine-proline (GRP) tripeptide, which alone is necessary and sufficient to mediate DNA binding (Reeves & Nissen 1990). The AT-hook motif typically functions as an auxiliary protein motif capable of binding to the minor groove of DNA, and is most often found in proteins also containing other DNA-binding motifs. The AT-hook motif has also been found in a wide variety of different DNA binding proteins including transcription regulators and co-factors of different DNA binding proteins

(Aravind & Landsman 1998). *T. brucei* NLP does not appear to be a vertebrate-like nucleoplasmin in the strict sense of the word, but the observed homology to nucleoplasmins indicates that it could be functioning as a histone chaperone in *T. brucei*. Given the manner in which dNLP has been classified (Ito et al. 1996), it is perhaps most appropriate to describe NLP as being nucleoplasmin-like. Further experiments determining the exact function of each identified domain will help clarify this issue.

The role of NLP in transcription control in T. brucei

NLP plays a key role in ES control. Blocking its synthesis leads to 45-65 fold derepression of silent ESs, and approximately 5 fold derepression of other transcriptionally silent areas of the bloodstream form *T. brucei* genome including the minichromosomes, the VSG Basic Copy arrays and the procyclin locus. I also noted that blocking NLP synthesis resulted in a dramatic decrease in levels of transcription from the active ES. Levels of transcript of the active *VSG*, *VSG T3*, fell to just 4% of basal levels within 72 hours induction of RNAi. Using chromatin immunoprecipitation, I was able to determine that NLP binds preferentially to certain silent areas around RNA polymerase I transcription units within the *T. brucei* genome. I was unable to find evidence that NLP preferentially binds either the silent ESs or the active ES. This fits with our observation that NLP is involved in transcription regulation in both kinds of ES. Nonetheless, in spite of the striking degree of derepression of ES promoters observed after NLP knock-down, and the fact that transcription from the active ES decreased; I did not find an observable increase in ES switching. However the drastic reduction in processive transcription of the active ES observed after induction of RNAi against NLP could be linked to the lack of change in VSG switching. Silent ESs might not be able to upregulate fully processive transcription extending as far as the telomere end where the *VSG* genes are located.

What is the role of NLP in transcription control in *T. brucei* in general? NLP was first discovered binding DNA by DNA-affinity chromatography of salt-extracted nuclear proteins from procyclic *T. brucei* extracts (K. Ersfeld, unpublished results). I found the levels of both NLP transcript as well as protein to be roughly equivalent in both *T. brucei* life-cycle stages. These indicate that NLP is likely to function as a general transcription regulator in both bloodstream form as well as procyclic form *T. brucei*. I also show that in addition to binding both silent and active ESs in bloodstream form cells, NLP also binds to non-transcribed loci such as the minichromosomes, VSG Basic Copy array and procyclin locus. NLP appears to play a role in regulating transcription from all of these regions. These results suggest that NLP is a general transcription regulator, and is not uniquely involved in ES control. To an extent, I have been impaired by the lack of a suitable reporter line with an *eGFP* gene inserted into a silent RNA polymerase II transcribed region, comparable to the various reporter lines used in this study which had a *eGFP* gene downstream of silent RNA polymerase I promoters such as the silent *VSG* ES and the procyclin EP1 promoters. I have found NLP binding to RNA polymerase II transcribed regions of the genome, but are unable to speculate on its role here. Our transcript analysis experiments indicate that levels of mature transcripts of RNA polymerase II transcribed genes such as actin, γ-gamma tubulin, SL, and RNA polymerase I large subunit do not change significantly upon depletion of NLP. However these transcripts may have long half-lives.

It has recently been shown that nucleosomes are highly depleted in areas of the *T. brucei* genome which are actively transcribed by RNA polymerase I such as the rDNA locus and the active ES (Figueiredo & Cross 2010; Stanne & Rudenko 2010). In contrast, there does not appear to be significant transcriptional regulation of the RNA polymerase II transcribed arrays of housekeeping genes (Clayton 2002). It is possible that NLP functions as a chromatin remodeler, and is therefore important first for maintaining fully processive

transcription of the active ES, and second for maintaining an appropriate silent state in inactive ESs. These could be achieved by modulating nucleosome presence as required in these locations. Our results so far indicate that NLP plays a role in facilitating as well as inhibiting transcription. There is precedent for this, as the ISWI complexes in other eukaryotes provide an example of a chromatin factor involved in both transcription repression and activation within a single cell. In *Saccharomyces cerevisiae* the Isw1a complex (consisting of Isw1p and Ioc3p) is involved in transcription silencing, while the Isw1b complex (consisting of Isw1p together with Ioc2p and Ioc4p) is involved in transcription elongation (Mellor & Morillon 2004).

Conclusion

The involvement of various epigenetic mechanisms in *T. brucei* to ensure monoallelic expression of VSGs has become clear only recently after the discovery that the active ES in bloodstream form *T. brucei* is greatly depleted in nucleosomes (Figueiredo & Cross 2010; Stanne & Rudenko 2010). In addition, various chromatin proteins such as TbISWI, TbRAP1, DOT1B, TbSpt16 etc are involved in regulating ES transcription (Hughes et al. 2007; Yang et al. 2009; Figueiredo et al. 2008; Denninger et al. 2010). From these findings, it is becoming clear that many of these proteins share some common functions and properties. *VSG* ESs extend over tens of kilobases (Rudenko 2010). It is likely that multiple silencing gradients operate simultaneously on silent ESs in *T. brucei* (Horn 2009; Yang et al. 2009). A relatively short distance repressive gradient extends up from the *T. brucei* ES telomere, and involves the sirtuin protein SIR2 (Glover & Horn 2006; Alsford et al. 2007). SIR2 is involved in repression in the immediate proximity of the telomere repeats, but does not appear to play a role in the long distance repressive effect operating over many tens of kilobases, necessary for control of the ES promoter (Alsford et al. 2007). In contrast, the telomere binding protein

RAP1 establishes a silencing gradient operating over much longer distances including the ES promoter, and which increases in strength closer to the ES telomere (Yang et al. 2009). Blocking RAP1 synthesis results in derepression of silent ESs, and an increase in transcripts from *VSGs* located at silent ES telomeres. In contrast, while blocking TbSpt16 also results in derepression of silent ESs, it does not lead to an increase in *VSG* transcripts from silent ESs (Denninger et al. 2010).

Depletion of NLP results in the derepression of a silent ES promoter and the downregulation of the active ES, but not to a significant increase in VSG switching. An explanation for this could be that NLP is involved in establishing or maintaining repressed chromatin in the immediate vicinity of the ES promoter. This hypothesis is strengthened by the argument that ES promoters inserted in other genomic areas such as the transcriptionally silent minichromosomes and VSG Basic Copy arrays are relatively silenced in the presence of NLP, but derepressed after its depletion. Once NLP is depleted, the region immediately downstream of the ES promoter may be derepressed, but this transcription may not continue as far as the telomere end of the ESs where *VSGs* are located. This could either be the consequence of a second repressive gradient impeding the elongating RNA polymerases. Alternatively NLP itself could be important for establishing the chromatin environment necessary for the elongating RNA polymerase I transcription complex, as NLP is important for maintaining fully processive transcription from the active ES.

One of the future challenges will be to determine precisely how NLP interacts with other chromatin remodelling proteins and histone modification enzymes in maintaining ES control. Our preliminary findings have shown that NLP indeed interacts with other chromatin remodelling proteins and is part of a multi-protein complex. These results are presented and discussed in the following chapter.

Chapter 3

NLP is a member of a novel ISWI complex

A. Introduction

ATP-dependent chromatin remodelers

Nucleosomes function at the primary level of DNA compaction into chromatin. Nucleosomal structures act as a barrier to various transcription regulatory factors and therefore at times need to be removed or displaced to allow access to DNA. This remodelling occurs by two main mechanisms: firstly, by the action of chromatin modifying enzymes which introduce covalent modifications to histone tails or the histone core (Kouzarides 2007). Secondly, chromatin remodelers can non-covalently alter DNA-histone interactions (Kwon & Wagner 2007). Such non-covalent changes include nucleosome sliding, nucleosome removal, the replacement of a histone with a histone variant, and the creation of new conformational states in which nucleosomes remain, but DNA becomes more accessible. Modifications which require an expenditure of energy in the form of ATP are performed by a class of ATP-dependent remodelers. These have so far been classified into four major sub-families of Swi2/Snf2 chromatin remodelling complexes on the basis of a distinct Swi2/Snf2 ATPase and other signature domains: the INO80, Mi-2, SWI/SNF, and ISWI complexes (Mohrmann & Verrijzer 2005).

The two best studied ATP-dependent chromatin remodelling complexes are the SWI/SNF (switching defective/sucrose non-fermenting) and ISWI (imitation SWI) complexes. Interestingly, they have opposing functions. SWI/SNF is involved in altering nucleosomal positions and compositions, which serves to expose regions of DNA to binding by transcription activators and silencers. The ISWI complexes on the other hand are involved in the regular spacing of nucleosomes, which is generally associated with higher order chromatin condensation and therefore with gene silencing (Narlikar 2010).

The protein domains of ISWI

ISWI was first identified as the ATPase-containing subunit of the novel nucleosome-remodelling factor (NURF) complex using an *in vitro* assay for nucleosome-remodelling factors in *Drosophila* embryo extracts (Tsukiyama et al. 1995; Tsukiyama & Wu 1995). Since its identification, ISWI homologs have been found to be highly conserved across eukaryotes ranging from yeast to metazoa. The protein domains found in a typical ISWI are the DEXD ATPase domain, a helicase domain, a HAND domain, a SANT domain and a SLIDE domain (Figure 3.1A). The DEXD ATPase domain is also known as the SNF2_N domain and together with the helicase domain it comprises the Swi2/snf2 ATPase domain class, which is the common feature of all ATP-dependent chromatin remodelers (Mohrmann & Verrijzer 2005). Interestingly, the helicase domain does not appear to confer helicase activity to any these proteins, although both SWI/SNF and ISWI have the ability to translocate on DNA, which is another property of helicases (Racki & Narlikar 2008). The SANT domain was identified on the basis of its similarity with a canonical DNA-binding domain called Myb, but has been shown to be functionally divergent and not involved with interaction with DNA (Pinskaya et al. 2009). Instead it binds unmodified histone tails (Yadon & Tsukiyama 2011). The SANT domain is also thought to be a part of a paired domain with SLIDE (SANT-like ISWI domain). The SLIDE domain resembles SANT but has considerable sequence insertions. In *S. cerevisiae* ISWI Isw1, the SANT and SLIDE domains comprise a nucleosome recognition module, and are required for nucleosome positioning (Pinskaya et al. 2009). There is also evidence from yeast Isw2 that the three C-terminal domains including SANT, SLIDE and HAND associate with extra-nucleosomal DNA, with the HAND domain associating with DNA at the entry/exit site of nucleosomes. Therefore, it is proposed that the C-terminal domain of ISWI anchors the complex to linker DNA near

nucleosomes and facilitates movement along DNA bound to nucleosomes (Dang & Bartholomew 2007).

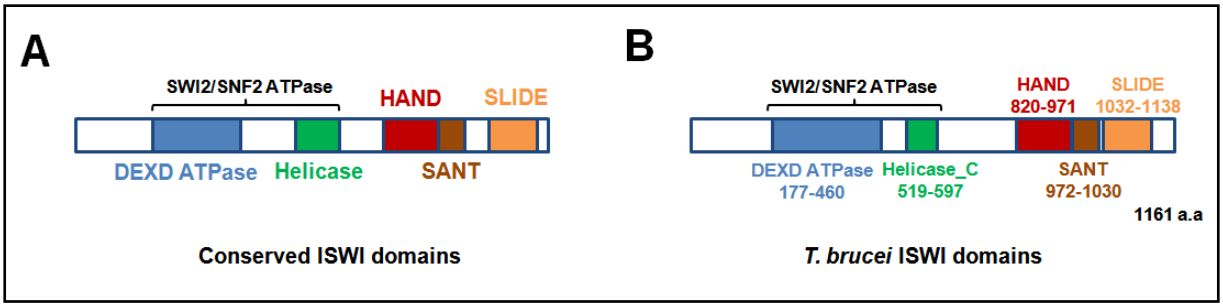


Figure 3.1 ISWIs are remarkably conserved across eukaryotes. **A)** The general domain structure of a typical ISWI. All ATP-dependent chromatin remodelers including SWI/SNF, ISWI, INO80 and Mi-2 contain the catalytic DEXD ATPase and helicase domains which together form a type of SWI2/SNF2 ATPase. ISWIs are distinguished by their archetypal C-terminal domain comprising the SANT, SLIDE and HAND domains, which together form a nucleosome recognition module. **B)** The predicted domains of TbISWI using HHpred (a remote technology and secondary structure based remodelling tool) and InterProScan. Panel **A** adapted from Mohrmann & Verrijzer 2005; Yadon & Tsukiyama 2011. Panel **B** adapted from Hughes et al. 2007; Kushwaha 2010.

ISWI complexes

Although ISWI can facilitate chromatin assembly in vitro by sliding nucleosomes on DNA, its chromatin remodelling ability is quite limited on its own (Längst & P. B. Becker 2001). Instead, ISWI functions in conjunction with various other proteins in complexes. The partners of ISWI therefore determine the specific role of the complex. Complexes which contain ISWI as one of their subunits can be classified into two major categories. The first is the ACF (ATP-utilizing chromatin assembly and remodelling factor) and CHRAC (chromatin-accessibility complex) subfamily. The second is the NURF (nucleosome remodelling factor) subfamily. Both subfamilies are named after the three complexes of the same names (Bao & Shen 2007) present in *Drosophila*, in which they were first identified

(Muchardt & Yaniv 1999) but have since been discovered to be conserved in various eukaryotes including yeast and humans (Bao & Shen 2007).

The ACF category of ISWI complexes comprises ISWI and its interacting partner ACF1 in *Drosophila* and in humans, and the ACF1 homolog Itc1 or the unique subunits Ioc2-4 in *S. cerevisiae*. The CHRAC complexes are similar to the ACF complex, but differ in having additional ISWI interacting partners which are CHRAC proteins or CHRAC homologs (Figure 3.2A-C). The ACF and CHRAC complexes are involved in mediating both transcription activation and repression, and there is evidence that the CHRAC complex is involved in DNA replication in *Drosophila* (Varga-Weisz et al. 1997; Corona & Tamkun 2004). The NURF complex is absent in yeast, *Xenopus*, and *Arabidopsis*, but is present in *C. elegans* and in humans (Figure 3.2A-C). Apart from containing ISWI, the NURF complex includes a large bromodomain containing subunit called NURF301 in *Drosophila*, and other subunits including a pyrophosphatase and a WD40 repeat containing protein (see Chapter 5). NURF has been shown to be essential for transcription activation and expression of homeotic genes. In addition, it interacts with a number of transcription factors including the GAGA factor (Badenhorst et al. 2002). In humans, the ISWI homologs SNF2H and SNF2L are members of the above complexes and of additional complexes such as RSF, which is involved in maintenance of chromosomal centromeres, and NoRC, which is implicated in the repression of ribosomal RNA genes (LeRoy et al. 2000; Yadon & Tsukiyama 2011).

The *T. brucei* ISWI homolog TbISWI has been shown to be involved in the repression of the RNA polymerase I transcribed *VSG* ESs in both bloodstream and procyclic form *T. brucei* (Hughes et al. 2007). Additionally, TbISWI is highly enriched at a number of strand switch regions which demarcates the boundaries between RNA polymerase II transcribed transcription units in *T. brucei* (Stanne et al. 2011). The various archetypal ISWI domains found in other eukaryotes are remarkably conserved in TbISWI (Figure 3.1B) (Kushwaha

2010), though functional studies have not yet been completed. It is also not known if TbISWI forms a complex with other proteins in *T. brucei* (Figure 3.2D).

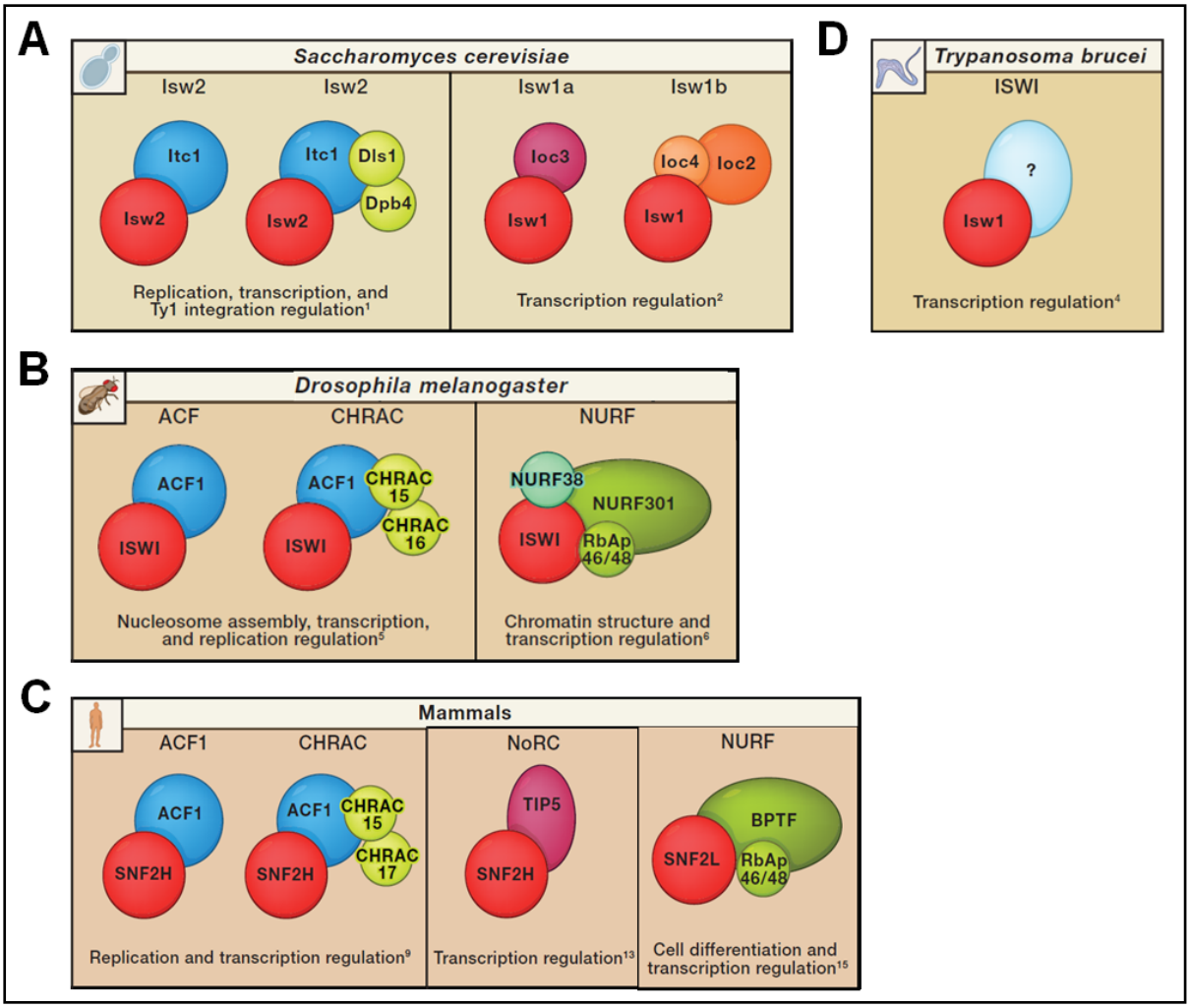


Figure 3.2 Pictorial depictions of representative ISWI-containing complexes from various eukaryotes. The interacting partners of ISWI determine the role that the complex plays. The ATPase ISWI is represented with red spheres, ACF and its homologs are shown with blue spheres, the NURF large subunit with green ellipses, and other partners with other coloured shapes. **A)** *S. cerevisiae* ISWI homolog Isw2 is part of the ACF/CHRAC complexes which are involved in DNA replication and repressive transcription regulation. The homolog Isw1 forms ACF/CHRAC-related complexes with its partners Ioc2-4. These complexes are involved in both transcription repression and activation. **B)** ISWI is a member of three ISWI complexes in *Drosophila* – the ACF, CHRAC, and the NURF complexes, which are involved in nucleosome assembly, gene silencing, and maintaining higher-order chromatin structure. **C)** The human ISWI homologs SNF2H and SNF2L are members of multiple complexes including ACF, CHRAC, NURF, and NoRC. ACF and CHRAC are linked with DNA replication and transcription repression, NURF with cellular differentiation, while NoRC is a repressor of transcription of ribosomal DNA. **D)** *T. brucei* ISWI is involved in transcription

downregulation. Its interacting partners are yet to be discovered. Adapted from Yadon & Tsukiyama (2011).

Reprinted from Cell, 144(3), Yadon, A. N. & Tsukiyama, T., Snapshot: Chromatin Remodeling: ISWI, p.453, © 2011, with permission from Elsevier. http://dx.doi.org/10.1016/j.cell.2011.01.019

I had previously found a novel nucleoplasmin-like protein (NLP) to be involved in *VSG* ES regulation in bloodstream form *T. brucei* (see previous chapter and Narayanan et al. 2010). Results from our own experiments as well as from others had shown that there are several proteins such as NLP, TbISWI, TbRAP1 and TbSpt16 (to name just a few) which are involved in *VSG* ES control (Hughes et al. 2007; Yang et al. 2009; Denninger et al. 2010). I therefore investigated if NLP interacted with other proteins to form a functional complex(es). I tagged NLP with the PTP-tag and tandem affinity purified (TAP) (Schimanski et al. 2005) it from bloodstream form *T. brucei* to find its interacting partners. I discovered that NLP interacts with TbISWI, and with two other proteins which were previously unidentified providing evidence for a novel ISWI complex in *T. brucei*.

B. Results

NLP is part of a complex in T. brucei

I had previously determined that NLP plays a critical role in *VSG* ES regulation. I had shown that NLP is an essential protein in bloodstream form *T. brucei*, and that after blocking its synthesis I observed a striking 45-65 fold derepression of a reporter *eGFP* gene in a silent *VSG* ES. Additionally, I found that depletion of NLP via RNAi resulted in a dramatic reduction of transcription from the active *VSG* ES, with *VSG* transcript levels falling to approximately 4% of normal levels within 48 hours of *NLP* RNAi. I next attempted to find if NLP interacted with any specific factors by performing tandem affinity purification (TAP) (Schimanski et al. 2005).

I tagged one of the NLP alleles with the PTP-tag in wild-type procyclic form *T. brucei* in the two *T. brucei* clones TAP-NLP1 and TAP-NLP2. I subsequently knocked-out the second NLP allele in the *T. brucei* clone TAP-NLP1, generating the clones KO/TAP-NLP1 and KO/TAP-NLP2. These results were confirmed by Western blot analyses. First I used the anti-ProtA antibody, which is expected to be specific for a section of the PTP tag on the tagged-NLP. Subsequently I used an anti-NLP antibody, which is expected to recognize tagged as well as untagged-NLP (Figure 3.3A). When reacted with the anti-NLP antibody the *T. brucei* KO/TAP-NLP clones showed only a single band representing tagged-NLP. In contrast, the TAP-NLP clones showed two distinct bands, one of which represented tagged-NLP and the other wild-type NLP. I next performed immunofluorescence microscopy on epitope-tagged and wild type cells using the anti-ProtA antibody specific to the PTP-tag to show that TAP-NLP localized to the nucleus (Figure 4.12B). I monitored the growth of the two KO/TAP-NLP clones, and found that the growth rate was similar in both TAP-NLP

clones, as well as wild-type parental cells (data not shown). These experiments indicated that PTP-tagged NLP is a fully functional protein.

I next purified TAP-NLP under non-denaturing conditions from our KO/TAP-NLP1 clone as previously described (Schimanski et al. 2005). Using Western blot analyses with an anti-NLP antibody I was able to follow the purification procedure over the different stages and determine that the process had proceeded as expected (Figure 3.3B). I next separated proteins in the purified extract under denaturing conditions in an SDS-PAGE gel and visualized them using the protein stain Coomassie blue (Figure 3.3C). Individual bands numbered 1-9 were excised separately and identified via liquid chromatography-tandem mass-spectrometry (MS). All results returned with a significant score for each band are listed in Table 3.1.

In our TAP experiment I obtained highly significant scores for three proteins other than NLP itself: Tb927.2.1810 (TbISWI), and two previously unidentified proteins Tb11.01.2100 and Tb927.7.1060. I temporarily named these two novel proteins TbRCC1 and FYRN1 respectively (see following sections). Apart from these proteins, I also identified α-tubulin, β-tubulin and translation elongation factor 1-alpha (EF1-α) with significant albeit considerably lower scores. α- and β-tubulin have previously been reported to be frequent contaminants in such protein affinity purification experiments (Schimanski et al. 2005). EF1-α is also assumed to be a contaminant given its comparatively low score. In addition, it would be expected to have a cytoplasmic sub-cellular location rather than the nuclear localization of NLP.

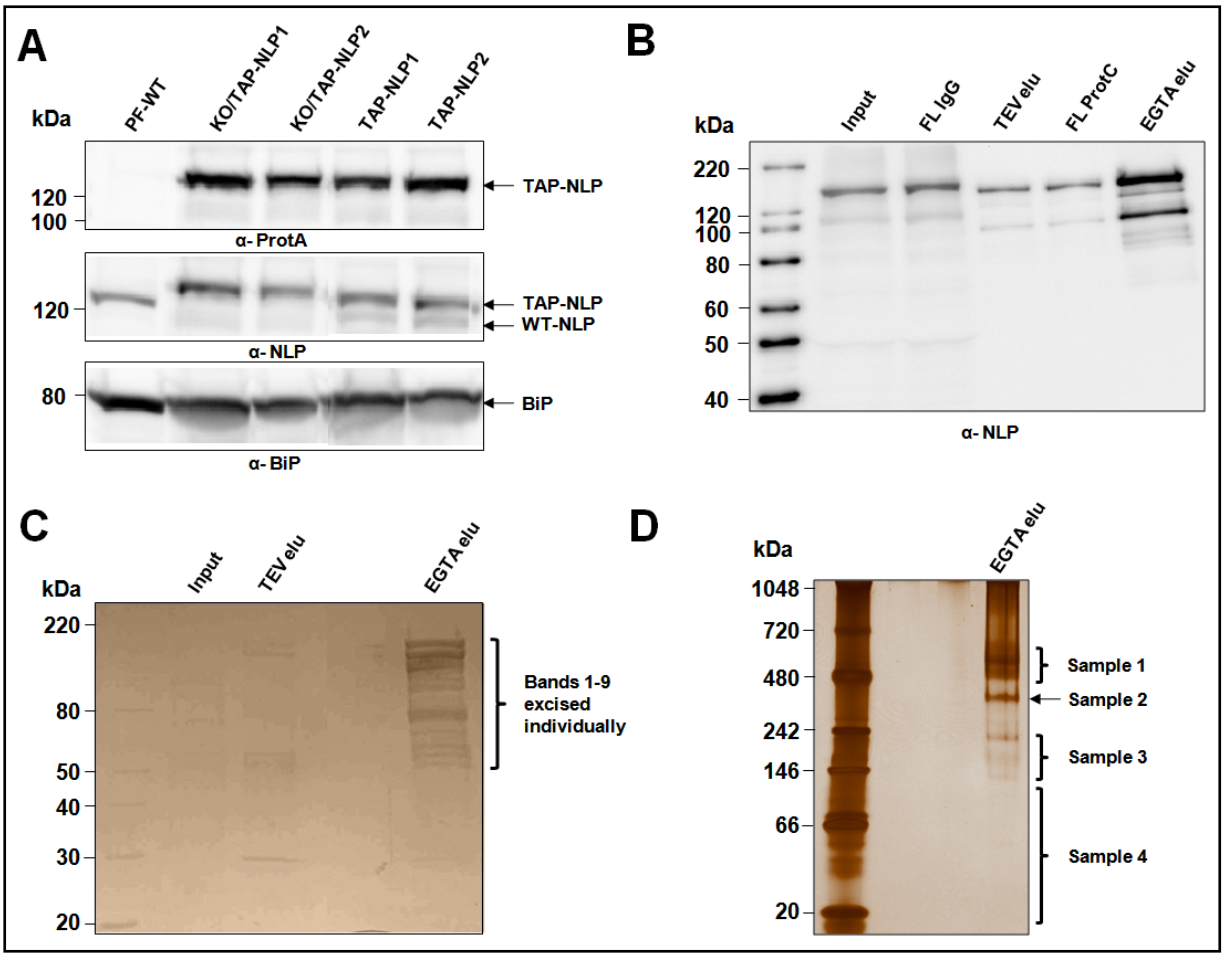


Figure 3.3 NLP is part of a complex in *T. brucei*. **A)** Western blot analyses of total protein lysates collected from wild-type procyclic form *T. brucei* (PF-WT), two clones with one allele of NLP tagged with PTP (TAP-NLP1-2), and two clones with one allele of NLP tagged with PTP and the other allele knocked-out (KO/TAP-NLP1-2). The blot shown in the upper panel has been reacted with an anti-ProtA antibody which recognizes the PTP tag. The middle panel shows the same blot re-probed with an anti-NLP antibody while the lower panel shows the same blot probed with an anti-BiP antibody to demonstrate equal loading. Lysates from 1×10^7 cell equivalents were loaded onto each lane. **B)** Western blot analysis of NLP during different stages of the TAP-NLP purification procedure. The input material, flow through after the IgG column (FL IgG, 0.002%), TEV protease eluate (TEV elu, 5%) and flow through after the ProtC column (FL ProtC 5%) were loaded and compared to the final eluate (EGTA elu, 100%) on a 12% SDS-polyacrylamide gel and blotted onto nitrocellulose membrane. The percent values refer to the relative volumes of each fraction analysed. An anti-NLP antibody was used to monitor for NLP. **C)** Final purified material (EGTA elu) was separated under denaturing conditions in a 4-15% SDS-PAGE gel. Input material and TEV protease eluate (TEV elu) were also analysed for comparison. The gel was stained with Coomassie blue and visualized bands (indicated 1-9) were excised individually and sent for analysis by mass spectrometry (MS). **D)** Final purified material (EGTA elu) was separated under non-denaturing conditions in a 4-15% polyacrylamide gel. The gel was silver stained and visualized bands were excised in 4 separate sections (samples 1-4) and sent for analysis

by MS. Size markers in kiloDaltons (kDa) are indicated on the left and the position of bands of interest are indicated by arrowheads or brackets on the right.

Band number	GeneID	Score	Name	Total peptide matches (Unique peptides)
1	Tb927.2.1810	3298	TbISWI	115 (95 unique)
	Tb927.10.5450	780	NLP	24 (24 unique)
2	Tb927.10.5450	3309	NLP	120 (55 unique)
	Tb927.2.1810	2369	TbISWI	86 (79 unique)
3	Tb927.10.5450	4526	NLP	151 (58 unique)
	Tb927.2.1810	2041	TbISWI	70 (65 unique)
4	Tb927.10.5450	2138	NLP	81 (44 unique)
	Tb927.2.1810	1962	TbISWI	71 (65 unique)
5	Tb927.10.5450	2152	NLP	90 (47 unique)
	Tb927.2.1810	1414	TbISWI	46 (45 unique)
	Tb11.01.2100	175	TbRCC1	9 (8 unique)
6	Tb11.01.2100	3694	TbRCC1	154 (46 unique)
	Tb927.2.1810	2060	TbISWI	74 (66 unique)
	Tb927.10.5450	875	NLP	32 (24 unique)
7	Tb927.2.1810	2243	TbISWI	79 (60 unique)
	Tb927.10.5450	1047	NLP	31 (26 unique)
	Tb11.01.2100	800	TbRCC1	25 (23 unique)
	Tb927.7.1060	598	FYRN1	14 (14 unique)
8	Tb927.7.1060	2422	FYRN1	67 (26 unique)
	Tb927.2.1810	1047	TbISWI	44 (42 unique)
	Tb927.1.2330	925	β-tubulin	24 (21 unique)
	Tb11.01.2100	611	TbRCC1	21 (20 unique)
	Tb927.10.5450	546	NLP	19 (18 unique)
9	Tb927.1.2340	1479	α-tubulin	54 (20 unique)
	Tb927.2.1810	1263	TbISWI	54 (47 unique)
	Tb927.1.2330	1041	β-tubulin	46 (28 unique)
	Tb927.10.5450	665	NLP	17 (17 unique)

	Tb11.01.2100	585	TbRCC1	20 (17 unique)
	Tb927.7.1060	543	FYRN1	11 (11 unique)
	Tb927.10.2100	253	Elongation factor-1α	10 (10 unique)

Table 3.1 The peptides identified from each of the excised bands (Band number) by mass-spectrometry (*T. brucei* sequence data base accession numbers listed under GeneID). All matches found with a significant score (Score) (where score $\geq$ 80) are shown. The total number of peptides is shown in the column peptide matches and the number of unique peptides found is indicated in brackets.

I next aimed to determine if NLP was part of more than one complex in *T. brucei*, and if so, in what combinations it was associated with its partners. I therefore repeated the entire TAP procedure, but this time separated our proteins in the final purified extract under non-denaturing conditions (Figure 3.3D). Resulting bands were silver stained, excised into 4 samples, and identified via MS (Table 3.2). Using the data from the MS results in combination with information provided by the size markers in the native gel, I was able to gain an insight into the composition of the NLP complex. I found that the first sample, which appeared to contain a complex of approximately 500-600 kDa in size as judged by the molecular ladder on our native gel, consisted predominantly of NLP, TbISWI and TbRCC1 in a 1:1:1 stoichiometric ratio. Sample 2 also consisted of the same three proteins, but this time in a 1:2:2 ratio of NLP, TbISWI and TbRCC1 respectively, while samples 3 consisted of NLP and TbRCC1 in a 4:1 ratio and sample 4 of TbRCC1 and NLP in a 6:1 ratio. Given that the predicted molecular sizes of NLP, TbISWI and TbRCC1 are ~107, ~133 and ~70 kDa respectively, the result from sample 1 indicates that this complex is likely to comprise of two molecules each of all three proteins. It is possible that the different proportions in which the same three proteins were found in all the other samples represent various sub-complexes,

though this would be technically difficult to differentiate from dissociation events occurring in the complex during our purification procedure.

Interestingly, I observed the remaining protein I had seen in our first TAP experiment, FYRN1, in all our samples but with scores below our threshold score of 80. It appears as if the association of this protein with our NLP complex is less stable compared to that of the other proteins, and therefore is lost from the complex during our experimental procedures.

Sample number	Size (kDa)	GeneID	Score	Name	Peptide matches (Unique peptides)
1	~500-600	Tb927.10.5450 Tb927.2.1810 Tb11.01.2100	1285 1078 922	NLP TbISWI RCC1	24 (19 unique) 34 (27 unique) 19 (15 unique)
2	~380-400	Tb927.2.1810 Tb11.01.2100 Tb927.10.5450	1009 618 525	TbISWI RCC1 NLP	35 (25 unique) 17 (13 unique) 13 (10 unique)
3	~120-220	Tb927.10.5450 Tb11.01.2100	1252 148	NLP RCC1	21 (18 unique) 7 (4 unique)
4	~20-120	Tb11.01.2100 Tb927.10.5450	226 92	RCC1 NLP	10 (4 unique) 6 (1 unique)

Table 3.2 The peptides identified from each of the excised bands (Sample number) by mass-spectrometry (accession numbers of the *T. brucei* sequence database listed under GeneID). All matches found with a significant score (Score) (where score $\geq$ 80) are shown. The number of peptides found in each case is shown in the last column (Peptide matches) and the number of unique peptides identified is indicated within brackets.

The interacting partners of NLP are also likely to be chromatin-associated proteins

From our TAP experiments I was able to conclude that NLP interacted with at least three other proteins in *T. brucei*. One of these proteins was TbISWI, an established repressor of RNA polymerase I transcribed *VSG* ESs which also appears to be involved in regulating RNA polymerase II transcription (Hughes et al. 2007; Stanne et al. 2011). I examined the two remaining putative NLP partner proteins for the presence of protein motifs using various protein domain detection packages (PROSITE, Pfam, MyHits.ISB). I found that Tb927.7.1060 (FYRN1) is a 494 amino acid long protein which contains a single FYRC (phenylalanine/tyrosine rich C-terminal) domain between amino acids 74-162 near the N-terminus ($e = 1.5 \times 10^{-2}$) (Figure 3.4A). FYRC domains have an unknown function, but have been reported to occur in several chromatin-associated proteins such as methyltransferases and the trithorax group of proteins (InterPro, EBI-EMBL). I provisionally named this protein FYRN1 (F/Y rich N-terminal 1).

The third protein, Tb11.01.2100 (TbRCC1), is a 646 amino acid long protein which contains four RCC1 (regulator of chromosome condensation 1) domains between amino acid positions 112-167, 216-282, 282-336 and 385-438 ($e = 5.8 \times 10^{-3}$, $e = 5.8 \times 10^{-3}$, $e = 1.1 \times 10^{-4}$, and $e = 5.1 \times 10^{-3}$ respectively) (Figure 3.4B). RCC1 domains are named after the domains of the human cell cycle regulator protein RCC1, which contains seven homologous repeats of 51-68 amino acids each (Hadjebi et al. 2008). Proteins with RCC1 and RCC1-like domains (RLDs) have been found in virtually all eukaryotes, some prokaryotes and in even in some viruses. These proteins are involved in a variety of functions including guanine nucleotide exchange, and in facilitating inter-protein interactions (Hadjebi et al. 2008). I provisionally named this protein TbRCC1.

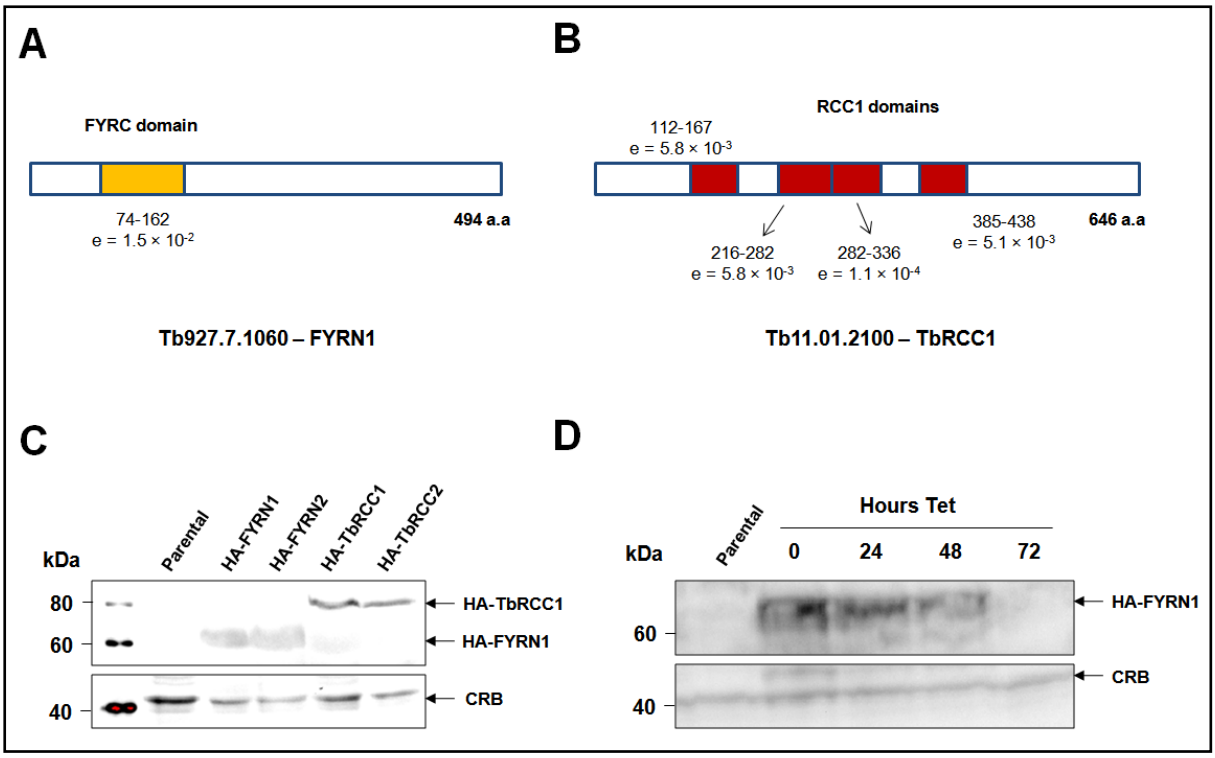


Figure 3.4 The interacting partners of NLP are also likely to be chromatin-associated proteins. **A)** Schematic representation of FYRN1 (Tb927.7.1060). FYRN1 is a 494 amino acid long protein with a single FYRC domain extending from positions 74-162 indicated with a yellow rectangle. **B)** Schematic representation of TbRCC1 (Tb11.01.2100), a 646 amino acid protein with four RCC1 domains indicated with red rectangles. Protein motifs were identified using PROSITE and MyHits.ISB. The 'e' values are shown beside each domain in the two schematic figures. **C)** Bloodstream form *T. brucei* 90-13 lines expressing HA-tagged copies of FYRN1 and TbRCC1. Protein lysate from parental 90-13 cells and the derivative clones HA-FYRN1, HA-FYRN2, HA-TbRCC1 and HA-TbRCC2 were subjected to Western blot analysis using an anti-HA antibody (upper panel). The lower panel shows an unknown cross-reactive band (CRB) which is recognized by the anti-HA antibody, and is used in this case to demonstrate equal loading. Lysates from 1×10^7 cell equivalents were loaded in each lane. **D)** The bloodstream form *T. brucei* clone HA-FYRN1 was transfected with a *FYRN1* RNAi construct. Protein lysate was collected at various time points (indicated in hours) after induction of *FYRN1* RNAi (Hours Tet). Lysates from 1×10^7 cell equivalents were loaded in each lane. Western blot analysis was performed using the anti-HA antibody. The signal from CRB demonstrates equal loading of lanes.

I next epitope-tagged the C-terminus of both FYRN1 and TbRCC1 in bloodstream form *T. brucei* 90-13 (Wirtz et al. 1999) with the HA-epitope (Oberholzer et al. 2006). Using Western blot analysis with an anti-HA antibody, I was first able to confirm that these proteins

were expressed in *T. brucei*, and second that I had successfully epitope-tagged them in the clones HA-FYRN1, HA-FYRN2, HA-TbRCC1 and HA-TbRCC2 (Figure 3.4C). In the HA-FYRN clones I found a band of approximately 65 kDa (predicted size of FYRN1 53.73 kDa). In the two HA-TbRCC clones I identified a band of approximately 80 kDa (predicted size 70.39 kDa).

I next determined if FYRN1 and TbRCC1 are essential in *T. brucei*. I transfected the clones HA-FYRN1 and HA-TbRCC1 with constructs allowing RNAi against *FYRN1* and *TbRCC1* respectively. These clones also allowed us to determine if these proteins were indeed knocked-down after RNAi induction, as I could monitor levels of the HA-tagged protein by Western blot analysis using the anti-HA antibody. In the case of the *FYRN* RNAi clones HA-FYRN.C1 and HA-FYRN.C2, I did not observe any change in growth for at least 96 hours after the induction of RNAi (data not shown). This was despite FYRN1 being knocked-down to virtually undetectable levels after 72 hours of RNAi (Figure 3.4D). Similarly, I did not see an effect on growth after inducing RNAi against RCC1 for 96 hours (data not shown). However, in this case I was not able to confirm knock-down of TbRCC1 protein in these cells. Other members of the Rudenko lab are presently raising antibodies against both these proteins. Since these antibodies can be expected to recognize protein generated from both alleles as opposed to only from the epitope-tagged copy, they would be useful in determining more accurately how much depletion is observed after the induction of RNAi and in establishing if both these proteins are truly non-essential in bloodstream form *T. brucei*.

I also MYC-tagged FYRN1 and TbRCC1 in both bloodstream and procyclic form *T. brucei* which already expressed HA-tagged NLP (data not shown). These lines have since been successfully been used by other members of the Rudenko lab in co-immunoprecipitation (co-IP) experiments using antibodies against the HA and MYC epitopes. These results

confirm our TAP results that NLP binds to TbISWI, FYRN1 and RCC1 in both bloodstream and procyclic form *T. brucei* (T. Stanne, A. Ling, unpublished results).

The simultaneous depletion of NLP and TbISWI by RNAi does not lead to higher levels of derepression of silent VSG ESs in bloodstream form T. brucei

The depletion of both TbISWI as well as NLP leads to over 50-fold derepression of silent *VSG* ESs in bloodstream form *T. brucei* (Hughes et al. 2007; Narayanan et al. 2010). Preliminary results from other members of the Rudenko lab have indicated that the induction of RNAi against FYRN1 as well as against TbRCC1 also results in derepression of silent *VSG* ESs (A. Ling, unpublished results). I next performed RNAi against both TbISWI and NLP simultaneously in our bloodstream form T3-SM reporter line (described in Figure 2.4A) (Hughes et al., 2007) to see if this resulted in greater ES derepression than that observed after performing *TbISWI* RNAi or *NLP* RNAi individually. I transfected a tetracycline-inducible RNAi construct containing both *TbISWI* as well as *NLP* fragments into the T3-SM cell line to give us the two clones T3-DB1 and T3-DB2. I first verified that I were indeed able to knock-down both of these proteins in these clones. I compared protein levels by Western blot analysis from samples isolated at various time points from RNAi induced cultures with samples from uninduced cells and were able to confirm depletion of both proteins following RNAi induction (Figure 3.5A). I also monitored eGFP as a marker for ES derepression using flow cytometry in our two clones after induction of RNAi. I did not observe greater derepression after simultaneously knocking-down both TbISWI and NLP compared with knocking-down either protein individually (Figure 3.5B).

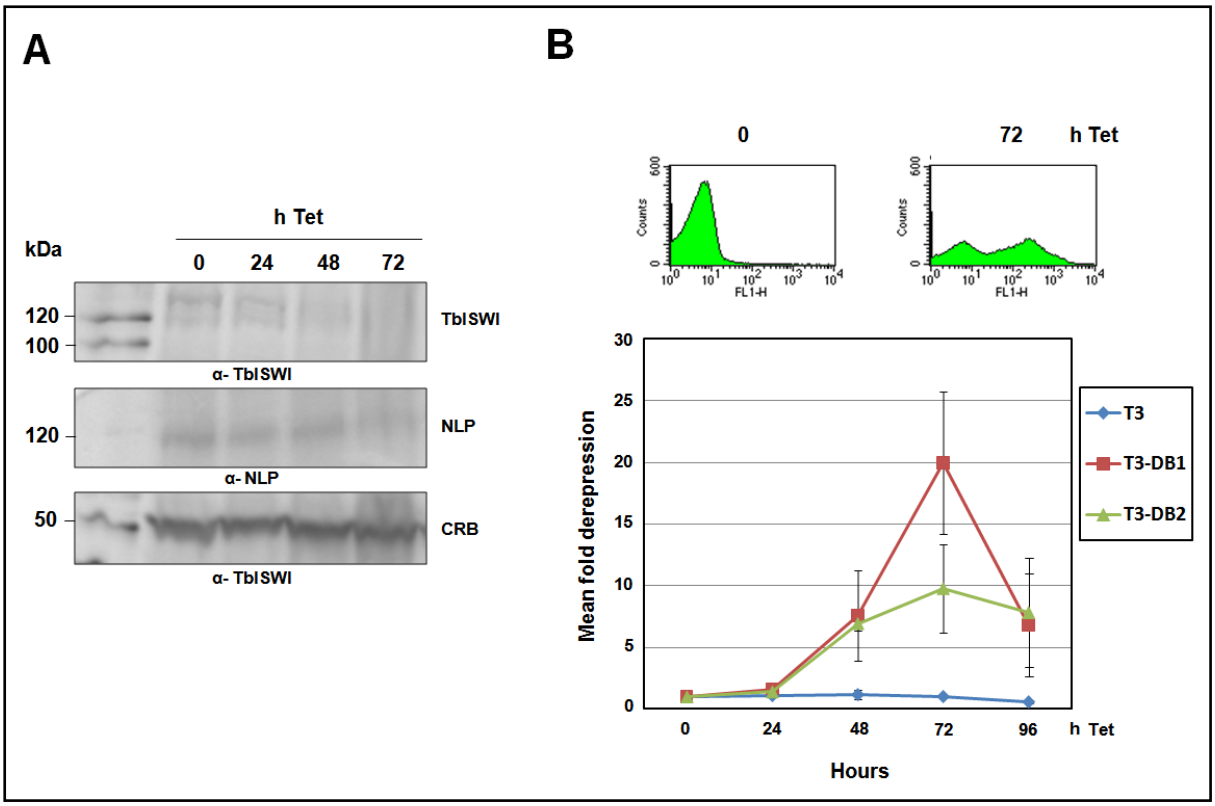


Figure 3.5 Depleting the cell simultaneously of both TbISWI and NLP does not result in greater derepression of silent *VSG* ESs compared with depleting the cell individually of either one. **A)** Western blot analysis of total protein lysate isolated from the bloodstream form *T. brucei* clone T3-DB1 after the simultaneous induction of *TbISWI* and *NLP* RNAi for the number of hours shown (h Tet). In the middle panel the blot has been probed with the anti-NLP antibody, while in the top and bottom panels the blot was probed with an anti-TbISWI antibody. The cross reactive band (CRB) demonstrates equal loading of lanes. Lysates from 1 × 10^7 cell equivalents were loaded in each lane. **B)** Representative flow cytometry histograms in the FL1 channel at 0 and 72 hours after RNAi induction, and below, a graphical representation of the same data showing fold derepression at each time point (calculated by dividing mean fluorescence at a given time point with that of the 0 hour time point) for the two clones T3-DB1 and T3-DB2. Data from identical experiments using the parental T3-SM (T3) line is shown for comparison. Error bars represent standard deviation from three independent repeats.

The distribution of NLP binding the strand-switch regions is different from that of TbISWI

Stanne et al. (2011) have recently reported that TbISWI is enriched in the strand-switch regions (SSRs) which mark the boundaries of RNA polymerase II transcribed units. I investigated if NLP is also particularly abundant in these areas. I performed qPCR on our

HA-NLP ChIP material (see Chapter 2) and determined NLP abundance on one 'divergent' and one 'convergent' SSR analysed by Stanne et al. (2011) (Figure 3.6A). In both SSRs, I selected primer pairs which hybridised with the genes flanking the SSRs (b and f), primer pairs which hybridised with genes approximately at the mid-point of the flanking transcription unit (a and g), and primer pairs within the SSRs themselves. I did not observe an enrichment of NLP within the SSR regions as reported by Stanne et al. (2011) for TbISWI. Instead, I found relatively minor enrichment of NLP on the genes flanking both the convergent as well as the divergent SSR (Figure 3.6B). In the divergent SSR, I found NLP to be between 1-3.5 fold enriched in the two genes flanking the SSR compared with the SSR itself. NLP was about 4-5 fold enriched in this region compared with the genes at the midpoints of the flanking polycistronic units. This pattern was repeated (though comparatively less pronounced) in a convergent SSR. Here NLP levels were about 1.5-3 fold enriched on the genes flanking the SSR compared to within the SSR. NLP levels binding here were not significantly different to levels at genes approximately in the middle of the flanking transcription units. In general, unlike TbISWI, NLP does not appear to be significantly enriched at the SSRs compared with other areas of the *T. brucei* genome.

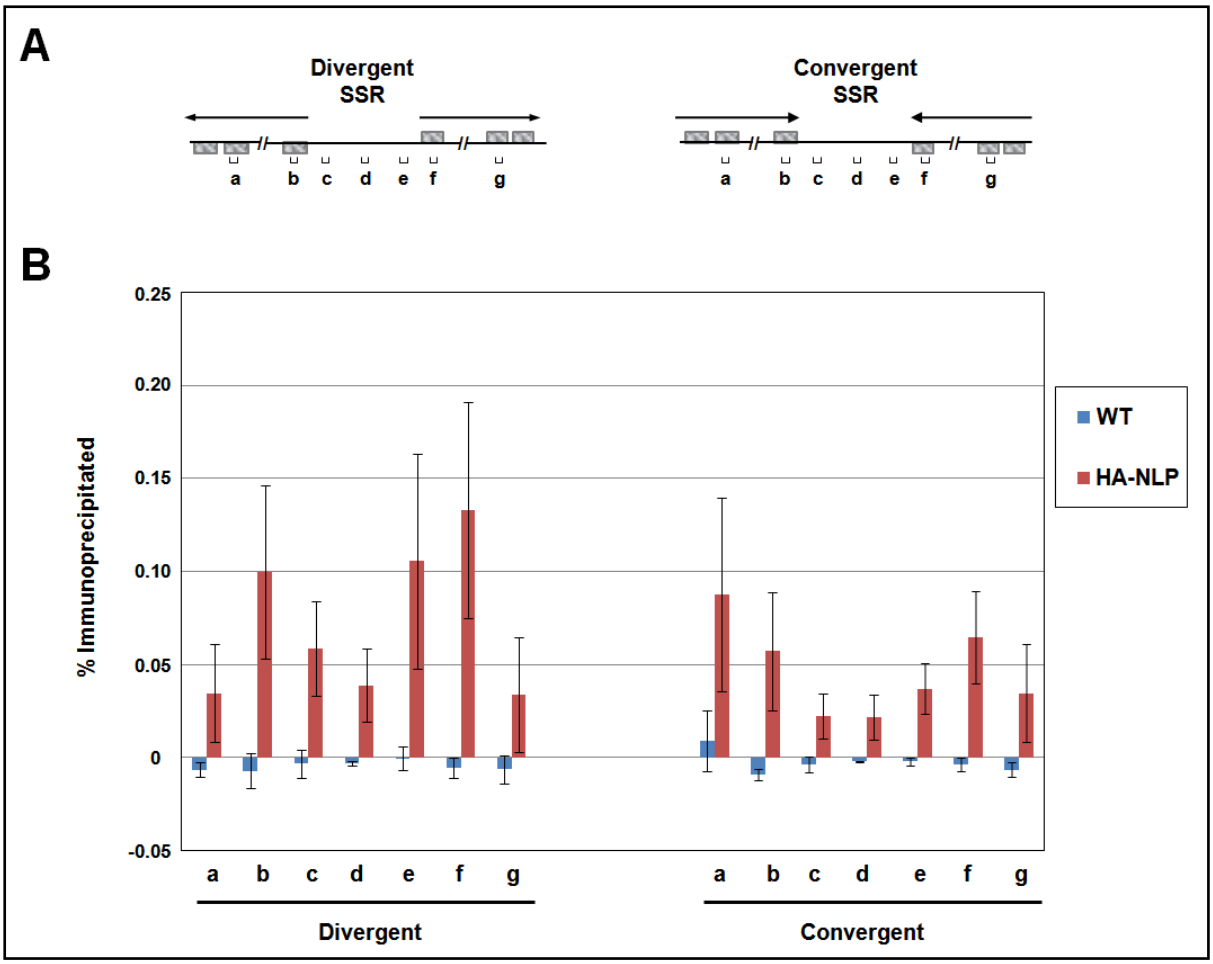


Figure 3.6 Unlike TbISWI, NLP is not significantly enriched in the strand switch regions (SSR) between RNA polymerase II-transcribed polycistronic transcription units. **A)** Schematic representation of a 'divergent' and a 'convergent' SSR. Boxes represent genes and arrows indicate the direction of transcription. Lettered brackets indicate primer pairs used for qPCR analysis of ChIP material in the next panel. **B)** qPCR analysis of ChIP material from the bloodstream form HA-NLP and the parental (WT) *T. brucei* lines. Immunoprecipitation was performed with an anti-HA antibody, and all values shown as % immunoprecipitated are shown after subtraction of the values obtained from an identical ChIP experiment using the same cell line but performed without the addition of antibody. Error bars indicate standard deviation from three independent repeat experiments.

C. Discussion

I have discovered that NLP is a member of an ISWI complex in *T. brucei*. I also find that in addition to TbISWI, NLP interacts with two novel proteins which I have provisionally named FYRN1 and TbRCC1 according to their respective protein domains. Our preliminary analyses indicate that similar to TbISWI and NLP, both FYRN1 and TbRCC1 are also likely to be chromatin-associated proteins. Additionally, I have evidence that NLP is likely to be a component of more than one complex in *T. brucei*, although the precise composition of these complexes is as yet unclear.

NLP is a member of a novel TbISWI complex

I were unable to identify possible interacting partners of TbISWI through BLAST searches of the *T. brucei* sequence databases using known ISWI partners in other eukaryotes (such as Itc1 in *S. cerevisiae*). An exception to this was *D. melanogaster* NURF301, which showed significant scores of $e = 2.5 \times 10^{-5}$, $e = 3.5 \times 10^{-5}$ and $e = 7.6 \times 10^{-5}$ for the bromodomain-containing proteins Tb11.01.1830, Tb927.10.7420 and Tb927.10.8150 (see Brom1-3 in Chapter 5). Similarly, I was unable to identify likely interacting partners of NLP using database searches. I was unable to find any *T. brucei* homologs of proteins such as N1 (which works in partnership with nucleoplasmin in *Xenopus* (Laskey et al. 1993)) with high scores. Using tandem affinity purification (TAP) (Schimanski et al. 2005) of NLP I made the unexpected discovery that it interacts with TbISWI.

In addition to interacting with TbISWI, NLP also interacts with two previously unidentified proteins. One of these, which I named TbRCC1, contains four RCC1 motifs, which are domains which were originally identified in the human cell cycle regulator protein

RCC1 which binds chromatin and acts as a guanine exchange factor (Hadjebi et al. 2008). The second interacting protein, which I provisionally named FYRN1, contains a single FYRC (phenylalanine/tyrosine rich C-terminal) domain at its N-terminus. These domains have an unknown function, but have been found in a variety of chromatin-associated proteins such as methyltransferases and the trithorax group of proteins (InterPro, EBI-EMBL). These results indicate that I have discovered an unusual TbISWI complex, where none of the TbISWI partners including NLP are typical ISWI partners, although all three bear known chromatin-associated protein motifs.

I next attempted to obtain independent verification for our TAP experiments as well as confirm that NLP interacted with each of the three putative partner proteins using co-immunoprecipitation experiments. I also investigated if the three putative TbISWI partner proteins also interacted with each other directly. I therefore generated both bloodstream and procyclic form *T. brucei* lines, each of which expressed an HA- and a MYC- epitope tagged copy of two of the four putative TbISWI complex proteins, such that all possible combinations of interactions could be tested (M. Narayanan, T. Stanne unpublished results). Using antibodies against the HA and MYC epitopes, other members of the Rudenko lab have been able to establish that NLP indeed binds TbISWI, TbRCC1 and FYRN1, and that these three proteins also bind with each other (T. Stanne, A. Ling, unpublished results). Moreover, a parallel TAP experiment in which TbISWI was purified instead of NLP also resulted in the identification of TbRCC1 and FYRN1 in addition to NLP (T. Stanne, unpublished results).

An important point to consider is that while TbISWI and NLP are both essential in *T. brucei*, so far I have not found that FYRN1 and TbRCC1 are essential, as inducing RNAi against them did not lead to a significant decrease in growth. It is possible that their functions are performed redundantly by other as yet undiscovered proteins. This could be verified by

genetically knocking-out FYRN1 and TbRCC1. If this is not lethal, one could investigate the newly changed composition of the NLP and TbISWI complexes.

In any case, our results provide strong evidence that TbISWI forms a novel complex in *T. brucei* which is unlike any other ISWI complex previously reported in other eukaryotes (Corona & Tamkun 2004; Yadon & Tsukiyama 2011). Further investigation of the *T. brucei* ISWI complex will provide valuable insight into *T. brucei* chromatin remodelers.

The role of the TbISWI complex in VSG ES control

It had been previously reported that the RNAi-mediated depletion of TbISWI led to a 30-60 fold derepression of silent *VSG* ESs in bloodstream form *T. brucei*, and a 10-17 fold derepression of the same loci in procyclic form *T. brucei*. Similarly, I found that knock down of NLP using RNAi resulted in a 45-65 fold derepression of silent *VSG* ESs in bloodstream form cells (see Chapter 2; Narayanan et al. 2010). However, in both cases the derepressed *VSG* ESs were transcriptionally activated to only 10% of the level of a fully active *VSG* ES as monitored by both flow cytometry and by qPCR analysis of *VSG* gene transcripts. This implies that *VSG* ES silencing involves multiple layers of repressors which form obstacles for fully active transcription, with the depletion of TbISWI or NLP removing only a subset of the layers. I therefore investigated if I could see an 'additive' effect on ES derepression when I simultaneously knocked-down both TbISWI and NLP in our bloodstream form *T. brucei* reporter line. This would indicate if TbISWI and NLP were part of different silencing pathways. However, I did not find that depleting both proteins simultaneously led to greater derepression than depleting either one individually. This result suggests that NLP and TbISWI are part of the same complex, at least as far as silencing *VSG* ESs is concerned. This result is in agreement with ChIP experiments investigating NLP (see Chapter 2; Narayanan et

al. 2010) and TbISWI (Stanne et al. 2011) distribution over *VSG* ESs, where both proteins appear to be present in approximately equal amounts.

However, despite similarities in distribution over ESs the roles of TbISWI and NLP appear to differ at the active versus the silent *VSG* ES. While depletion of TbISWI does not seem to result in a change in levels of transcription from the active *VSG* ES (Stanne et al. 2011), the depletion of NLP leads to a dramatic reduction in transcripts from this site (see Chapter 2; Narayanan et al. 2010). In addition, while induction of *TbISWI* RNAi leads to increased transcription of silent telomeric ES-located *VSG*s, induction of *NLP* RNAi does not. In fact, *NLP* RNAi actually leads to a slight reduction in the basal level of transcription of the silent telomeric ES-located *VSG*s. These results imply that NLP plays an additional role in ES regulation. In addition to being important for transcription downregulation of silent *VSG* ESs, NLP also appears to be essential for facilitating fully processive transcription of ESs down to the telomere end containing the *VSG* gene. It is possible that NLP acts in combination with its non-TbISWI partners or even independently of the TbISWI complex at active *VSG* ESs.

Could there be a complex containing NLP but not TbISWI? In our second TAP experiment I separated our purified material under non-denaturing conditions. I identified only NLP and TbRCC1 in our samples 3 and 4, suggesting that these two proteins together could be involved in control of the active *VSG* ES. Investigating this is however technically difficult. It would however be very informative to perform TAP experiments with TbISWI, FYRN1 and TbRCC1 independently. One could investigate via native gels if there is evidence for multiple complexes being present, and if so, in what combinations. Presently, other members of the Rudenko lab are presently investigating the binding distribution of FYRN1 and TbRCC1 by ChIP (A. Ling, unpublished results). These results would provide insight into the composition of the NLP complex at these loci.

An important point to remember is that in our TAP experiments I have used procyclic form *T. brucei*, while much of the data I report relating to *VSG* ES control has been obtained using bloodstream form cells. The TAP experiments were performed with procyclic *T. brucei* as extremely high numbers of cells are required for TAP experiments (approximately 3.5×10^{10} cells). These large amounts of cells are feasibly only cultivated *in vitro* using procyclic form cells. Therefore, it could be informative to repeat our TAP experiments in bloodstream form cells by cultivating them *in vivo* in rats, from which I would be able to obtain the requisite numbers of cells. However so far the co-IP data generated by other members of the Rudenko lab has indicated that the four TbISWI complex proteins in question form equivalent complexes in bloodstream form *T. brucei* as well.

The role of the ISWI complex in the regulation of RNA polymerase II transcription

Stanne et al. (2011) have recently found evidence that TbISWI is enriched at strand switch regions (SSRs), which form the boundaries between RNA polymerase II transcription units. SSRs, which are enriched in modified histones and histone variants, are the presumed sites of RNA polymerase II transcription initiation and termination (Siegel et al. 2009). I therefore investigated using our HA-NLP ChIP material (see Chapter 2) if NLP was also enriched in these regions. However, I did not find evidence that this was the case. It therefore appears that NLP is not a part of the TbISWI complex at the SSRs, though it remains to be seen if FYRN1 or TbRCC1 are enriched at these loci. It would be particularly interesting to investigate the genomic distribution of FYRN1. FYRN1 contains the F/Y rich C terminal domain, which has an unknown function but is present in trithorax proteins (InterPro, EBI-EMBL). There is evidence in *Drosophila* that a trithorax protein Zeste is involved in histone replacement at the boundaries of cis-regulatory elements, and it is proposed that this happens

in conjunction with a SWI/SNF nucleosome remodeler which evicts nucleosomes from these locations (Mito et al. 2007).

In our second TAP experiment I attempted to determine if NLP is present in more than one complex. This indicated that while NLP, TbISWI and TbRCC1 are present in a large complex, NLP and TbRCC1 are present in a smaller complex, and FYRN1 may be absent from both. I identified insignificant amounts of FYRN1 across all our samples in this experiment. It is possible that the interaction between NLP and FYRN1 is relatively weak compared with the interaction between TbISWI and FYRN1. Alternatively, NLP and FYRN1 might both interact with TbISWI but not with each other directly. In addition TbISWI and FYRN1 together, but not NLP, could play a role in regulating transcription at the SSRs. This too will become more clear after the ChIP experiments currently underway are completed.

Interestingly, it has been reported that a bromodomain containing protein is highly enriched at the SSRs in *T. brucei* and is implicated in the acetylation of histone H4 at these locations (Siegel et al. 2009). Bromodomains, which are frequently found on histone acetyltransferases, and are capable of recognizing acetylated lysines in histone tails, are a distinguishing feature of many SWI2/SNF2 ATPases. These domains are absent from ISWIs, which instead often form a complex with proteins which contain bromodomains, such as NURF301 in *Drosophila* (Mohrmann & Verrijzer 2005). The bromodomain protein reported in Siegel et al. (2009) was not picked up in the TbISWI TAP experiment (T. Stanne, unpublished results) or in our NLP TAP experiment (this thesis). There is however an intriguing correlation, in that the striking abundance of this bromodomain protein as well as TbISWI at the SSRs indicate that they could be involved in the same pathway. I have so far conducted some preliminary experiments on this bromodomain-containing protein (see Chapter 5), and it would be highly informative to investigate this protein further by attempting to find its interacting partners using TAP.

Conclusions

In conclusion, I have discovered that NLP is a part of a novel TbISWI-containing complex in *T. brucei*. This complex includes two previously unidentified proteins FYRN1 and TbRCC1, both of which are implicated in chromatin regulation on the basis of their protein motifs. Additionally, our data suggests that NLP is a member of at least two different complexes in *T. brucei* – one which contains TbISWI and TbRCC1 and is involved in *VSG* ES silencing, and another which includes TbRCC1, and which I speculate is involved in maintaining transcription of the active *VSG* ES. I also speculate that TbISWI is a member of at least one other complex which does not contain NLP, and that this complex is functionally active at the boundaries of polycistronic RNA polymerase II transcribed transcription units in *T. brucei*.

It was earlier unclear as to whether transcription was epigenetically regulated in *T. brucei* (Rudenko 2010). In recent years, while a number of proteins have been shown to be important for regulating transcription of the *VSG* ESs, this is one of the first studies which has identified a chromatin remodelling complex and thereby linked two proteins previously shown individually to be involved in *VSG* ES transcription control. I anticipate that further investigating the precise interactions of these proteins within the TbISWI and NLP complexes is a crucial step towards our eventual goal of understanding the web of interactions amongst the various players involved in transcription control and chromatin remodelling in *T. brucei*.

Chapter 4

TDP1 is an architectural chromatin protein important for transcription control in *T. brucei*

A. Introduction

The High Mobility Group (HMG) family

The high mobility group (HMG) proteins are named in relation to their electrophoretic mobility in polyacrylamide gels. HMG proteins were first discovered in mammalian cells over 40 years ago and since then have been identified in all eukaryotes (Bustin 2001; Stros et al. 2007). They comprise one of the classes of non-histone DNA binding proteins which together with DNA form chromatin. HMG proteins introduce structural changes in chromatin which are involved in regulating processes such as replication, repair and transcription, all of which require unwinding or condensation of DNA. Several HMG proteins are also known to interact with nucleosomes, making them after the H1 linker histone family the largest and most abundant group of architectural proteins in metazoans (Stros 2010).

The HMG proteins are sub-divided into three super-families, with each super-family having a characteristic sequence motif. The characteristic feature of the nucleosome binding domain protein (HMGN) super-family is the presence of a nucleosome binding domain (NBD). This is a 30 amino acid domain which is capable of binding specifically to the nucleosomal core particles, making it the only known protein domain with this ability (Gerlitz et al. 2009). The second super-family includes proteins which contain AT-hook domains (HMGA) as their main chromatin-binding motif. AT-hooks are small nine amino acid long motifs which preferentially bind to the minor groove of DNA. Proteins within this category have between one to about 15 copies of the motif, and frequently also contain a different DNA-binding domain (Aravind & Landsman 1998; Bustin 2001). Functions of both HMGN and HMGA proteins are regulated by numerous post-translation modifications that affect their binding to DNA or to other chromatin proteins (Gerlitz et al. 2009).

The third super-family is characterised by the presence of the HMG box domain and is hence called HMGB. These proteins are abundant in the nucleus and usually contain one to two HMG boxes, called HMG box A and HMG box B, which differ slightly from each other both in terms of structure and DNA binding and bending activities (Stros 2010). Another feature of some HMGB proteins in metazoans is an acidic C-terminal tail, but this element is absent in yeast (Stros et al. 2007). Several studies have shown that HMGB proteins have higher affinity for non-canonical DNA structures such as single-stranded DNA or cruciform DNA compared with normal linear DNA (Stros 2010). HMGB proteins also bind with high affinity to DNA that is already sharply bent, such as linker DNA at the entry and exit of nucleosomes; and thus is considered a structural protein of chromatin (Calogero et al. 1999). Damaged DNA is also a high-affinity substrate for HMGB proteins, and there is evidence emerging that HMGB proteins along with their interacting protein partners are involved in DNA repair (Liu et al. 2011). A well characterised example of an HMGB protein is the human testis determining gene SRY, which is a crucial transcription activator of another gene involved in gonad differentiation (Werner et al. 1995). Another example is the *S. cerevisiae* HMGB protein Hmo1, which contains two HMG boxes and is involved in transcription by both RNA polymerase I and II (Kasahara et al. 2011). The *T. brucei* trypanosome DNA-binding protein 1 (TDP1), falls into this last category of HMG proteins.

Characterization of TDP1

TDP1 was first discovered by Erondu & Donelson (1992) binding to VSG expression site promoter oligomer sequences. A *T. b. rhodesiense* expression cDNA library was screened for recombinant clones which expressed a fusion protein capable of binding to a specific double-stranded DNA sequence. This was a 25-mer oligonucleotide which included

an octamer sequence motif found upstream of some *VSG* genes. TDP1 was the only protein they could identify which bound to their DNA sequence.

As has been described in the introduction to Chapter 2, TDP1 was subsequently found by DNA-affinity chromatography as binding to the *T. brucei* 177 bp repeat sequence found predominantly on the minichromosomes along with TbISWI and NLP (K. Ersfeld, unpublished results). In the light of this data, I speculated that TDP1 was involved in chromatin regulation in *T. brucei*, and decided to further investigate its function. So far, I have found that TDP1 is an essential nuclear protein, and is particularly abundant in the nucleolus and expression site body (ESB) in bloodstream form *T. brucei*. I also find that blocking TDP1 synthesis results in a 40-90% reduction in transcription of genes transcribed by RNA polymerase I. Using chromatin immunoprecipitation (ChIP), I find that TDP1 appears to be enriched on the active VSG expression site in bloodstream form *T. brucei*, as well as at the promoter regions of certain RNA polymerase I, II and III transcribed loci of both bloodstream form and insect form *T. brucei*. I also performed tandem affinity purification (TAP) of TDP1 followed by mass spectrometry, and found that TDP1 interacts with the core histones. These results indicate that TDP1 is an architectural chromatin protein important for transcription control in *T. brucei*.

B. Results

Sequence analysis

The high mobility group B (HMGB) protein family constitutes a major abundant class of non-histone chromatin associated DNA-binding proteins which play a role in chromatin architecture in a wide range of eukaryotes. In *T. brucei*, the HMGB protein TDP1 was first identified as binding to *VSG* expression site promoter oligomer sequences (Erondu & Donelson 1992) and thus far remains the only HMGB protein identified in *T. brucei*.

TDP1 is a 271 amino acid long protein with a predicted size of 30.87 kDa, and contains two HMG boxes ($e = 9.7 \times 10^{-33}$ and $e = 5.2 \times 10^{-21}$) and one DEK C ($e = 1.8 \times 10^{-5}$**)** terminal DNA-binding domain. DEK C domains are named after the C-terminal domain of the human chromatin protein DEK. The C-terminus of DEK is structurally homologous to a member of a transcription factor family called E2F/DP, and is capable of binding dsDNA (Devany et al. 2004). Interestingly, TDP1 has a long N-terminal region upstream of its two HMG box motifs which is not present in most other HMGB members. However, this unusual feature is also present in its orthologues in *T. cruzi* (Tc00.1047053507951.114) and *Leishmania major* (LmjF29.0850), and in all three species it is within this region in which the DEK C domain is present. The DEK C domain extends from amino acids 13-63, while the HMG boxes extend from amino acids 118-186 and 206-270 (Figure 4.1A). Another interesting feature of TDP1 is that it completely lacks an acidic C-terminal tail, which is characteristic of many HMGB proteins in metazoans (Stros et al. 2007).

TDP1 is an essential protein in bloodstream form T. brucei

I first investigated if TDP1 was an essential protein in bloodstream form *T. brucei*. I blocked the synthesis of TDP1 in the *T. brucei* T3-SM reporter cell line (Hughes et al. 2007) by inducing RNAi against it in the two clones T3-TDP1 and T3-TDP2. I found that this resulted in a rapid growth arrest within 12 hours of induction, while the parental T3 line remained unaffected after the addition of tetracycline (Figure 4.1B). I also confirmed these results in the bloodstream form *T. brucei* 90-13 cell line (Wirtz et al. 1999). Here too I found a similar growth arrest following *TDP1* RNAi in our two clones TDP C1 and TDP C2, while the parental 90-13 cell line remained unaffected (Figure 4.1C). I confirmed that this growth arrest corresponded with a decrease in TDP1 protein levels by probing a Western blot containing protein lysate extracted from cells where *TDP1* RNAi had been induced with an anti-TDP1 antibody (kind gift of K. Ersfeld, University of Hull). I found that TDP1 protein levels are knocked down to 78% initial levels within 24 hours induction of *TDP1* RNAi, to 69% by 48 hours and to 57% by 72 hours (Figure 4.1D). These results indicate that even a partial block of TDP1 synthesis is sufficient to cause a rapid growth arrest. However, staining *TDP1* RNAi induced cells with the DNA stain DAPI and examining them under a light microscope showed that this clear growth arrest was not linked with any particular stage of the cell cycle (Figure 4.1E).

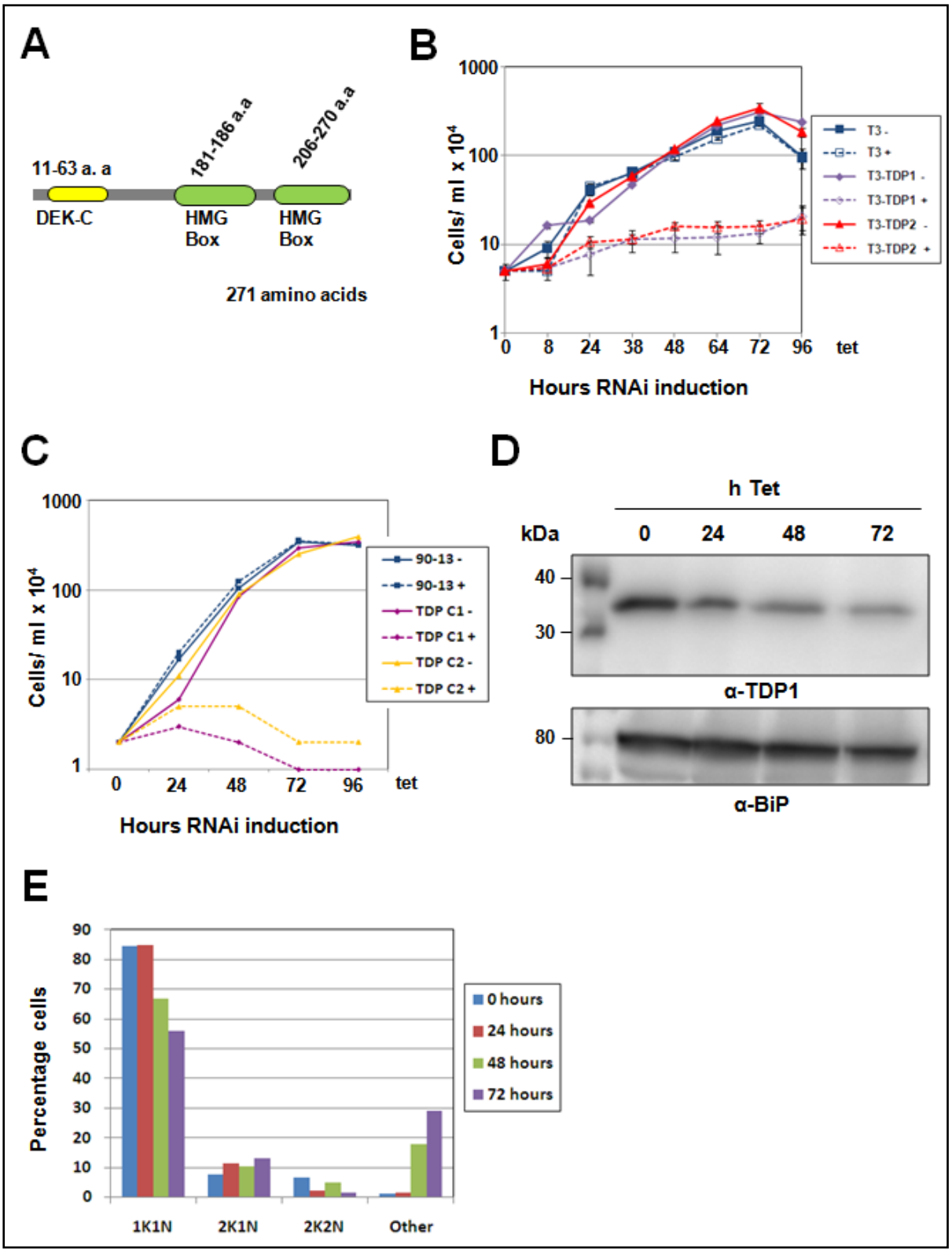


Figure 4.1 TDP1 is an essential HMG box containing protein in *T. brucei*. **A)** Schematic representation of TDP1 and its predicted protein domains. HMG box and DEK C DNA binding motifs are shown by elliptical boxes and the amino acid positions they are found in are shown above. **B)** Growth kinetics of bloodstream form *T. brucei* after the induction of *TDP1* RNAi. The + denotes cell lines in which tetracycline was added to induce RNAi.

Growth of the two clones T3-TDP1 and T3-TDP2, and the parental cell line T3-SM lacking the *TDP1* RNAi construct is plotted. Error bars indicate the standard deviation from three independent repeats. **C)** A growth curve as seen in panel **B** replicated in the bloodstream *T. brucei* 90-13 line. The growth kinetics of two *TDP1* RNAi clones TDP C1 and TDP C2 and the parental 90-13 line lacking the *TDP1* RNAi construct are plotted. **D)** Western blot analysis monitoring TDP1 protein levels after *TDP1* RNAi induction. Lysates from 1×10^6 cell equivalents were prepared at each time point (indicated as hours after tetracycline addition) after RNAi induction in the T3-TDP1 cell line. The same blot was then probed with anti-BiP antibody to demonstrate equal loading between lanes. **E)** Microscopic analysis of DAPI stained bloodstream form cells was performed at various time points after the induction of *TDP1* RNAi. Cells were categorised according to the number of nuclei (N) or kinetoplasts (K) present and the percentage of cells in the different categories was plotted.

It had been reported earlier that *TDP1* transcript levels were downregulated in cells differentiating from the slender bloodstream form to the stumpy bloodstream form, and then upregulated subsequently as cells differentiated into the procyclic form (Kabani et al. 2009). I investigated if I could detect any differences in the levels of *TDP1* transcript between bloodstream form and procyclic cells. cDNA was prepared from RNA isolated from both life cycle stages, and using qPCR, I first compared transcript levels of the major surface proteins of each life cycle stage. As expected, I found approximately 95-fold higher levels of *VSG* transcript in the bloodstream form cells compared with procyclic form cells, and approximately 11-fold higher levels of *EP1* transcript in the procyclic form cells (Figure 4.2A). However, I found levels of *TDP1* transcript to be roughly equivalent in both life cycle stages. I also compared the transcript levels of *NLP* and actin and were not able to detect a significant difference between the two *T. brucei* life forms (Figure 4.2B). In addition, I also found using Western blot analysis with the anti-TDP1 antibody that TDP1 protein levels were also similar in bloodstream and procyclic form cells (Figure 4.2C).

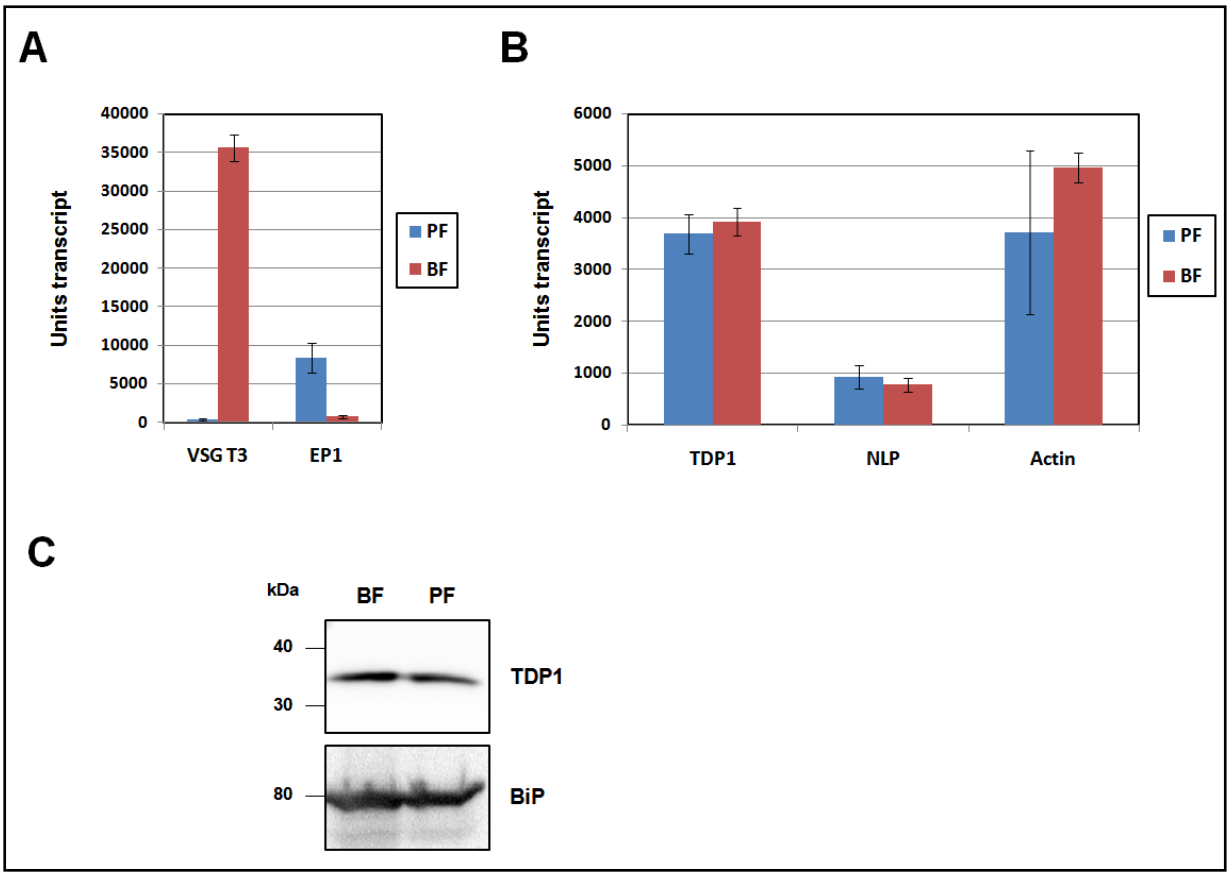


Figure 4.2 *TDP1* transcript and protein are both expressed at roughly equivalent levels in bloodstream form and procyclic *T. brucei*. RNA was isolated from both wild-type procyclic cells and the bloodstream form *T. brucei* T3-SM cell line and was used to prepare cDNA. The levels of various transcripts were then compared by qPCR. **A)** A comparison of the levels of transcript of the major surface proteins *EP1* and *VSG* in procyclic (PF) and bloodstream form (BF) *T. brucei*. **B)** A comparison of the transcript levels of *TDP1*, *NLP* and actin in bloodstream form and procyclic form *T. brucei*. Transcript levels are shown in arbitrary units. Error bars represent the standard deviation from three independent repeat experiments. **C)** Western blot analysis comparing TDP1 protein levels in bloodstream form and procyclic form *T. brucei*. Lysate from 1×10^7 bloodstream or procyclic cell equivalents was loaded in each lane and the blot was probed with an anti-TDP1 antibody. The blot was re-probed with an anti-BiP antibody to demonstrate equal loading.

Depletion of TDP1 does not lead to derepression of VSG ESs or ES promoters inserted into transcriptionally silent areas of the genome

HMGB proteins such as human HMGB1 have been shown to be involved in transcription repression (El Gazzar et al. 2009). Therefore, using our bloodstream form *T.*

brucei T3 reporter line (Hughes et al. 2007) I investigated if blocking the synthesis of TDP1 resulted in the derepression of the silent *VSG221* ES, as occurs after TbISWI and NLP knock-down. I was unable to observe any derepression of the *eGFP* gene in the silent *VSG221* ES after inducing RNAi mediated depletion of TDP1 (Figure 4.3).

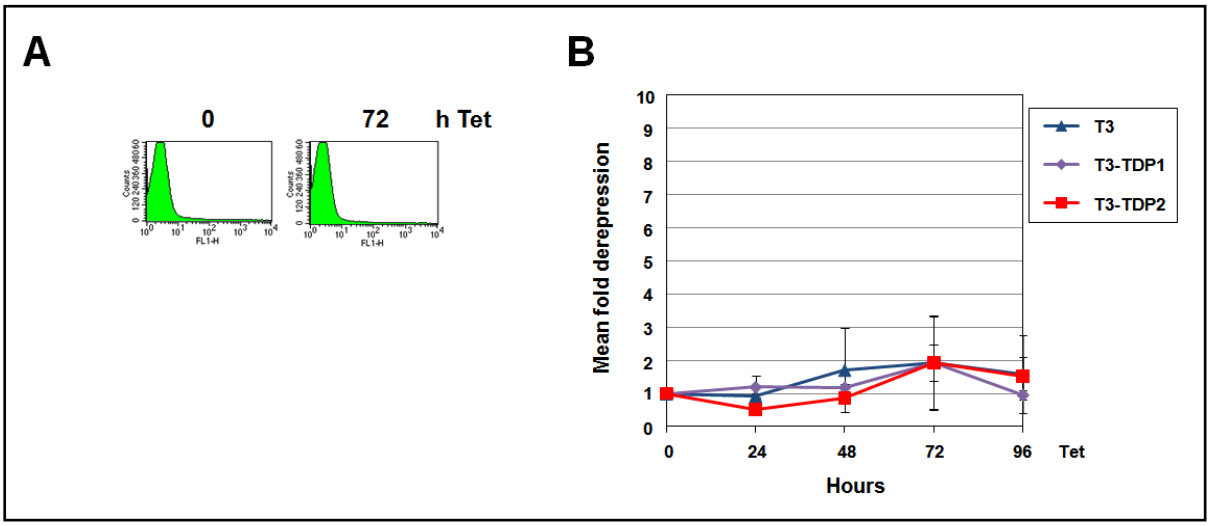


Figure 4.3 Blocking the synthesis of TDP1 using RNAi does not result in the derepression of the silent *VSG221* ES in bloodstream form *T. brucei*. **A)** The panels show representative flow cytometry traces in the FL1 channel at 0 and 72 hours after the addition of tetracycline to induce *TDP1* RNAi. **B)** The graph shows the mean fold change in the FL1 channel quantification over time in the two *T. brucei* clones T3-TDP1 and T3-TDP2. The parental T3-SM line (T3) is shown for comparison. Error bars represent the standard deviation from three independent repeat experiments.

As TDP1 had been found to bind to the 177 bp simple sequence repeats found in the minichromosomes (K. Ersfeld, unpublished results), I further investigated if TDP1 was involved in silencing or maintaining the transcriptionally silent states of the minichromosomes and the VSG Basic Copy arrays. I used the bloodstream form *T. brucei* reporter cell lines MCEP and VBEP which have a construct containing an ES promoter upstream of an *eGFP* gene integrated into these silent genomic locations. This allowed us to

screen for ES derepression by flow cytometry (Narayanan et al. 2010). In addition, I used the corresponding *T. brucei* reporter lines MCe and VBe which are identical to MCEP and VBEP other than that they lack the ES promoter upstream of the *eGFP* gene (Stanne et al. 2011). I did not find observable derepression of the reporter *eGFP* gene after blocking the synthesis of TDP1 via RNAi in any of these reporter lines, indicating that TDP1 does not play a role in transcription repression (Figure 4.4 and 4.5).

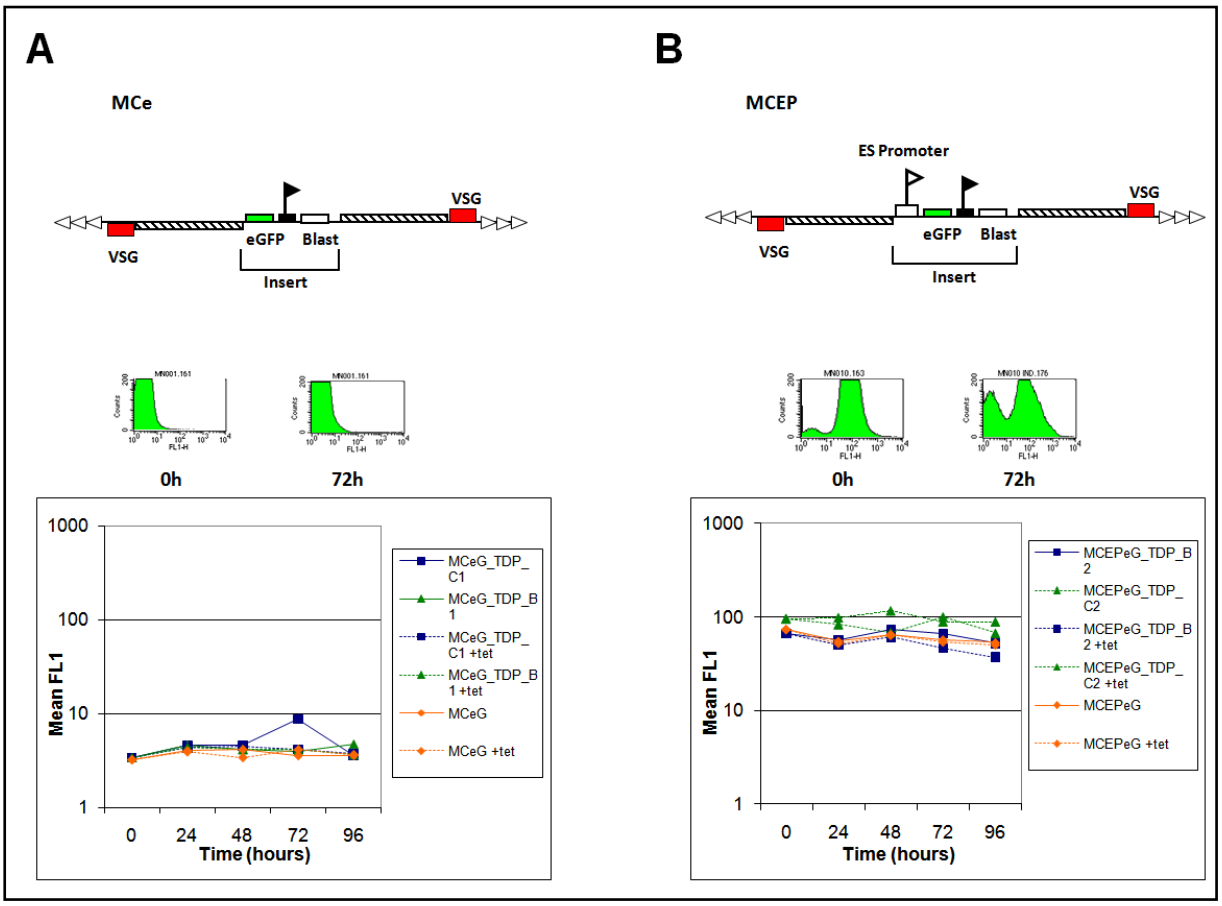


Figure 4.4 Monitoring eGFP derepression after blocking synthesis of TDP1 by RNAi in the bloodstream form *T. brucei* minichromosome reporter cell lines MCe and MCEP. Schematic representations of the two lines are shown at the top. Minichromosome simple sequence repeats are shown with striped boxes and telomeres with triangles. The *eGFP* gene is shown with a green box, the blasticidin resistance gene with a white box, and *VSG* genes with red boxes. rDNA promoters are shown with black flags and the ES promoter with a white flag. The reporter construct that has been integrated into the minichromosomes in each cell line is shown with a bracket. *eGFP* gene expression levels were measured by flow cytometry after the induction of *TDP1* RNAi. The *T. brucei* reporter lines were either the **(A)** *T. brucei* MCe

line lacking the ES promoter or **(B)** the *T. brucei* MCEP line which has an ES promoter upstream of the *eGFP* gene. Top panels show representative flow cytometry traces at 0 or 72 hours after *TDP1* RNAi induction. Lower panels show mean fluorescence from the FL1 channel in induced (+tet) or uninduced cells. Data from parental induced and uninduced MCe and MCEP cells (orange lines) are also shown for comparison.

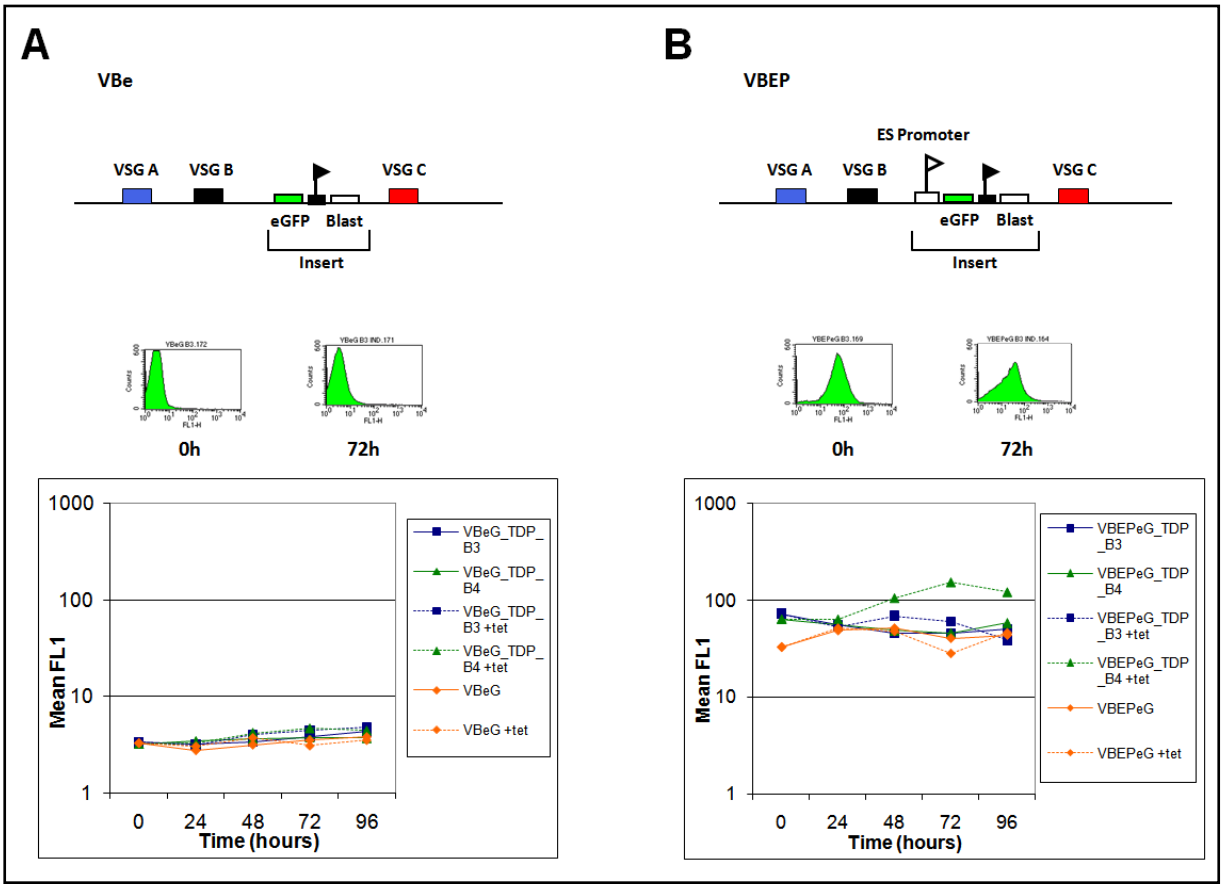


Figure 4.5 Monitoring eGFP derepression after induction of *TDP1* RNAi in the bloodstream form *T. brucei* VSG Basic Copy array reporter cell lines VBe and VBEP. Schematic representations of the two lines are shown at the top. The *eGFP* gene is shown with a green box, the blasticidin resistance gene with a white box, and *VSG* genes with other coloured boxes. rDNA promoters are shown with black flags and the ES promoter with a white flag. The reporter construct that has been integrated into the VSG Basic Copy array in each cell line is shown with a bracket. *eGFP* gene expression levels were measured by flow cytometry for *TDP1* RNAi clones derived from both the **(A)** *T. brucei* VBe line lacking the ES promoter and **(B)** the *T. brucei* VBEP line which has an ES promoter upstream of the *eGFP* gene. Top panels show representative flow cytometry traces at 0 or 72 hours after induction of *TDP1* RNAi. Lower panels show mean fluorescence from the FL1 channel in induced (+tet) and uninduced cells. Data from parental induced and uninduced *T. brucei* VBe and VBEP cells (orange lines) are also shown for comparison.

TDP1 is enriched in the nucleolus and in an additional T. brucei sub-nuclear compartment

I epitope tagged the C-terminus of TDP1 with haemagglutinin (HA) epitope tag in two isogenic bloodstream form *T. brucei* cell lines 221GPI(VO2+) and 221GPI(221+). These lines are identical, except that *T. brucei* 221GPI(VO2+) is maintained under neomycin selection to select for activation of a *VSG ES* expressing *VSG VO2*. *T. brucei* 221GPI(221+) is maintained under puromycin selection to select for activation of a *VSG 221* expressing ES (Sheader et al. 2004). Western blot analysis with an anti-HA antibody confirmed that TDP1 was tagged with HA in the two *T. brucei VSG VO2* expressing clones VO2-HAT1 and VO2-HAT2 as well as the *VSG 221* expressing clones 221-HAT1 and 221-HAT2 (Figure 4.6A-B). Similarly, I also epitope tagged TDP1 with HA in wild-type procyclic form *T. brucei* to generate the clones PF HAT1-4 (Figure 4.6C).

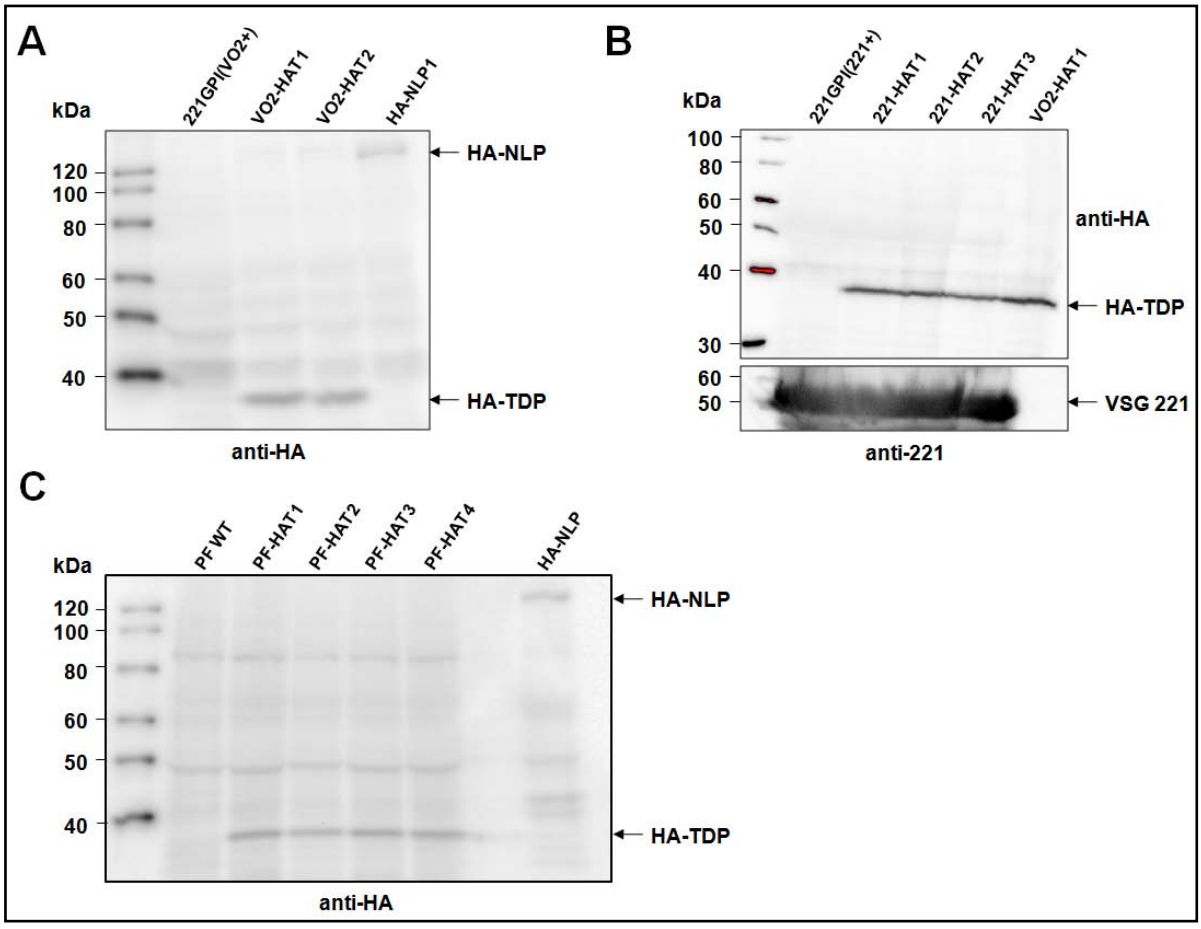


Figure 4.6 Epitope tagging of TDP1 with haemagglutinin (HA) epitope tag. Western blot analysis using an anti-HA antibody shows a band of the expected size in epitope tagged cells but not in parental cells. **A)** HA-tagging of TDP1 in the bloodstream form *T. brucei* 221GPI(VO2+) line. Lysates from the clones VO2-HAT1 and VO2-HAT2 and lysate from the parental *T. brucei* 221GPI(VO2+) line are compared. Lysate from the *T. brucei* HA-NLP1 line was also loaded as a positive control. **B)** Similarly, lysates from the parental *T. brucei* 221GPI(221+) line and its derivative clones 221-HAT1-3 are compared. The blot was re-probed with an antibody against VSG 221 (lower panel). Lysate from the *T. brucei* VO2-HAT1 line was also loaded as a positive control for the anti-HA antibody and as a negative control for the anti-221 antibody. **C)** Lysates from parental wild type cells (PF-WT) and the derivative clones PF-HAT1-4 with HA-tagged TDP1 are compared. Lysate from the HA-NLP1 line was also loaded as a positive control. 1×10^6 bloodstream or procyclic form *T. brucei* cell equivalents were loaded in each lane.

Based on its predicted motifs, I hypothesized that that TDP1 was a DNA-binding nuclear protein. I therefore performed immunofluorescence microscopy on our *T. brucei* VO2-HAT1 epitope tagged clone using an antibody specific to the HA-tag to determine the sub-cellular localization of TDP1. Microscopic analysis confirmed that TDP1 is not only a nuclear protein but is found throughout most of the nucleus, with a striking sub-nuclear distribution. I observed that in approximately 80% of the cells TDP1 appeared to be highly enriched in the nucleolus, and that in roughly a third of these (~20% of all cells) there was an additional and smaller area of enrichment close by. I obtained an antibody against an unknown nucleolar marker L1C6 (Devaux et al. 2007) (kind gift of K. Gull) and used it for microscopic analysis. Merged microscopy images clearly showed that TDP1 was enriched within the nucleolus as determined using the L1C6 marker in virtually all cells examined (Figure 4.7). As before, I could see an additional, smaller spot of TDP1 enrichment (shown by arrows in Figure 4.7) next to the nucleolus in ~20% of all cells examined. I speculated that this additional spot comprises the expression site body (ESB), which houses the active *VSG* ES in bloodstream form *T. brucei* (Navarro & Gull 2001).

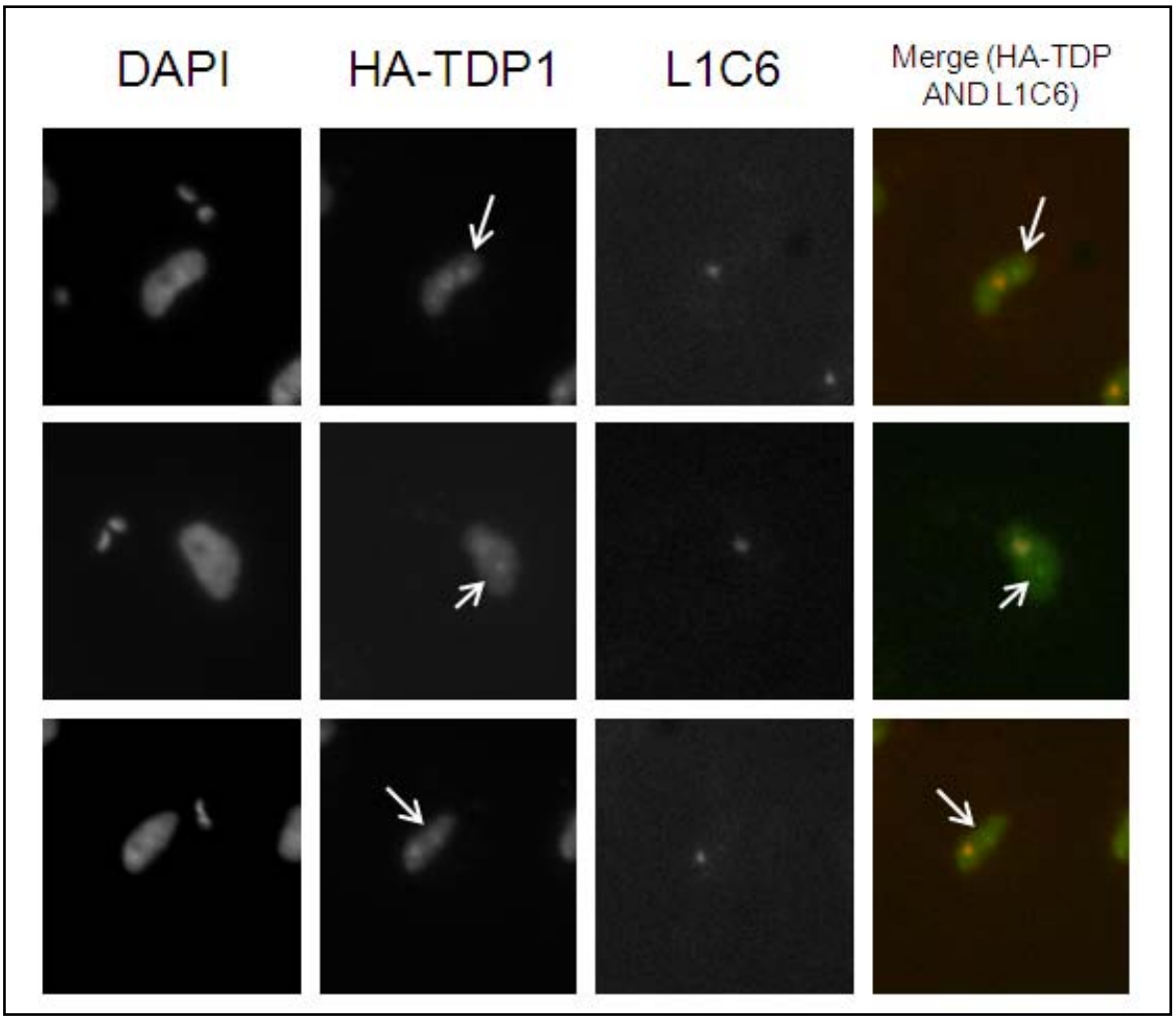


Figure 4.7 TDP1 is a nuclear protein enriched in the nucleolus as well as in an additional sub-nuclear compartment. Panels show bloodstream form cells stained with the DNA stain DAPI, the FITC channel after immunofluorescence (IF) using a rabbit polyclonal anti-HA antibody and an anti-rabbit Alexa-488 secondary antibody (HA-TDP1), or the TRITC channel after IF using a mouse monoclonal anti-L1C6 nucleolar marker antibody and an anti-mouse Alexa-594 secondary antibody (L1C6). The rightmost panels show merged images of the HA-TDP1 and L1C6 images on the left. Arrows indicate the additional area of TDP1 enrichment in the nucleus that does not correspond with the nucleolus.

TDP1 is enriched in RNA polymerase I transcribed regions such as the rDNA and the active VSG ES

I next performed chromatin immunoprecipitation (ChIP) to further examine the striking sub-nuclear localization pattern I had seen by immunofluorescence microscopy. ChIP was performed on the *T. brucei* clone VO2-HAT1 using an anti-HA antibody, as well as an

antibody against histone H3 as a positive control. As a negative control for the specificity of the anti-HA antibody, simultaneous ChIP experiments using the same antibodies were also performed on the parental 221GPI(VO2+) line. As an additional negative control, identical ChIP experiments were also simultaneously executed on both cell lines without any antibody. Identical ChIP experiments were also carried out on the *T. brucei* clone 221-HAT11 and its parental cell line 221GPI(221+). Immunoprecipitated (IP) material was subsequently slot-blotted. I hybridised these with radio-labelled probes specific for the 177 bp minichromosome repeat sequence and the 50 bp repeat sequence found upstream of all known *VSG* ESs. Both these areas are transcriptionally inactive in *T. brucei*. I found a small amount of TDP1 binding to these regions at approximately 2-10 fold background levels (percent IP = ~0.04 on the 50 bp repeats and ~0.02% on the 177 bp repeats) (Figure 4.8A-B). I next analysed the binding of TDP1 to various other areas of the *T. brucei* genome using quantitative PCR (qPCR). I analysed regions transcribed by RNA polymerase I, II or III. I first analysed the spliced leader (SL) genes transcribed by RNA polymerase II and the 5S rRNA transcription units transcribed by RNA polymerase III. Schematic diagrams of these regions and the primer pairs used in the qPCRs are shown in Figure 4.8C-D. I found considerable enrichment of TDP1 at these loci (36-175 fold background levels, percent IP = ~0.25), but were unable to find a significant difference in TDP1 enrichment between genic and intergenic regions (Figure 4.8E).

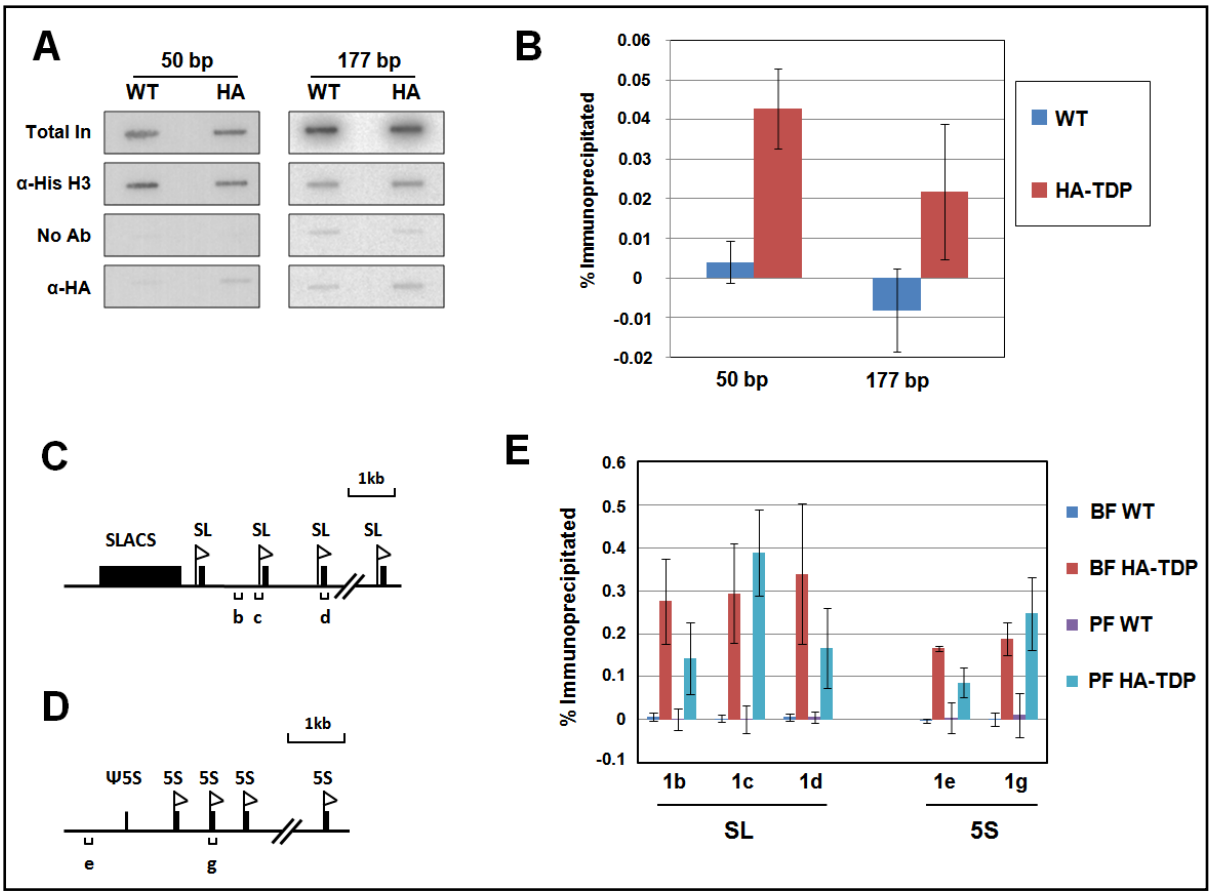


Figure 4.8 Chromatin immunoprecipitation (ChIP) in bloodstream form *T. brucei* indicates that TDP1 levels are approximately 5-10 folds lower on the simple sequence repeats compared with the transcribed regions. **A)** Representative slot blots of ChIP material hybridized with probes specific for the 50 bp sequence repeats found upstream of all known *VSG* ESs or the 177 bp sequence repeats found on minichromosomes. Panels show ChIP material from the parental cell line lacking HA-TDP (WT) as well as from HA-TDP cells (HA). Total input (Total In) is compared with ChIP performed using an anti-histone antibody (α-His H3), ChIP performed without antibody (No Ab), and ChIP performed with an anti-HA antibody (α-HA). **B)** Quantification of data from three independent ChIP and slot blot experiments with standard deviation shown with error bars. ChIP values for WT and HA-TDP ChIPs are presented after subtraction of the 'α-HA' values in **A** by the 'No Ab' values. **C)** Schematic diagram representing the spliced leader (SL) RNA transcription units transcribed by RNA polymerase II. Genes are shown by black boxes, promoters by flags and primer pair amplicons analysed by quantitative PCR (qPCR) are shown by lettered brackets. **D)** Schematic representation of the 5S rRNA transcription units transcribed by RNA polymerase III. **E)** Quantification of ChIP results from both bloodstream form (BF WT and BF HA-TDP) and procyclic form (PF WT and PF HA-TDP) *T. brucei* analysing the regions shown in **C** and **D** by qPCR. As in **B**, ChIP values shown for WT and HA-TDP ChIPs are presented after subtraction of values obtained when no antibody was used.

Our immunofluorescence microscopy experiments indicated that TDP1 was enriched in the nucleolus and in the ESB, which houses the active ES in bloodstream form *T. brucei*. RNA polymerase I transcribes both these regions as well as the procyclin locus in procyclic *T. brucei*. I therefore investigated the distribution of TDP1 on rDNA by qPCR. Strikingly, TDP1 was highly enriched in the area immediately downstream of the rDNA promoter (primer pair c in Figure 4.9A) in both bloodstream form as well as procyclic *T. brucei*. Enrichment was approximately 3 to 9 fold compared with the intergenic spacer (primer pairs a and l) and by 1.3 to 2.5 fold compared with the transcription unit itself (primer pairs d and g) (Figure 4.9B). The rDNA promoter (primer pair b) showed 2 to 3 fold enrichment of TDP1 compared with the upstream spacer region (primer pair a). Sequences immediately downstream of the promoter accounted for 0.6-0.8% of the total material immunoprecipitated which is ~2-8 folds higher than that observed for RNA polymerase II and III transcription units, supporting our microscopy data that TDP1 is enriched in the nucleolus. I next investigated TDP1 distribution over the procyclin transcription unit in both bloodstream and procyclic form *T. brucei*. While TDP1 was found to bind to both the intergenic spacer region (primer pair a) and the transcription units themselves (primer pairs c, d, e in Figure 4.9C) in both lifecycle stages, I found that TDP1 was particularly depleted (by about 6-7 fold) around the promoter region (primer pair c) specifically in bloodstream form *T. brucei*. TDP1 was much more evenly distributed within the procyclin transcription unit in procyclic *T. brucei*, with no significant variance in TDP1 binding on the promoter region (Figure 4.9D).

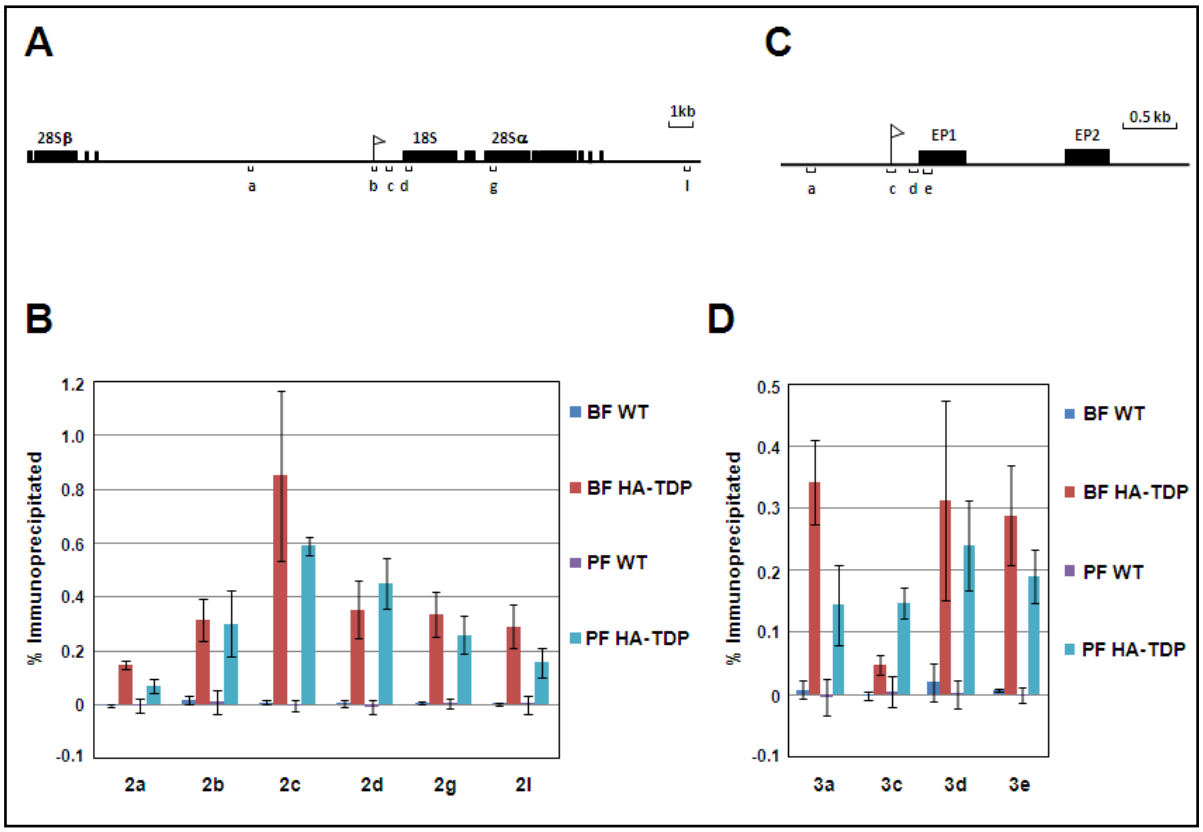


Figure 4.9 ChIP in bloodstream form and procyclic form *T. brucei* reveals that TDP1 is enriched in regions actively transcribed by RNA polymerase I. ChIP experiments were performed on HA-TDP1 expressing bloodstream (BF HA-TDP) and procyclic form (PF HA-TDP) *T. brucei* as well as parental bloodstream (BF WT) and procyclic form (PF WT) *T. brucei* lacking HA-TDP1. **A)** Schematic diagram representing the genomic region around the ribosomal DNA (rDNA) transcription units. Promoters are denoted by flags and genes by boxes. Regions analysed by qPCR are shown with lettered brackets. **B)** The graph below shows quantification of TDP1 ChIP as analyzed by qPCR in both life-cycle stages. **C)** Schematic representing part of the EP procyclin locus and regions analysed by qPCR. **D)** Graphical representation of qPCR analysis of the EP1 locus. Standard deviation from three independent experiments is represented with error bars.

Next, I examined TDP1 distribution on the *VSG* ES in bloodstream form *T. brucei*. I found that TDP1 was more abundant immediately downstream of the ES promoter as was the case with the rDNA (Figure 4.10B). However this distribution is less obvious as the primer pairs in question (primer pairs a, c, d in Figure 4.10A) recognise multiple ESs, most of which would be silent. ES sequences are highly similar to each other (Hertz-Fowler et al. 2008) and

normally cannot be distinguished from each other by qPCR. However, the isogenic bloodstream form *T. brucei* cell lines 221GPI(VO2+) and 221GPI(221+) (Sheader et al. 2004) provide a convenient way for allowing us to distinguish between active and silent ESs using the single copy drug markers inserted within them (Figure 4.10C). When I compared relative TDP1 abundance using ChIP material from the *T. brucei* VO2-HAT1 and 221-HAT1 lines, I discovered that TDP1 was 3.7 to 5.4 fold enriched on the active ES immediately downstream of the promoter, and 1.8 to 14.7 fold enriched at the telomere end of the active ES on the respective *VSG* gene (Figure 4.10D). These findings strongly support our immunofluorescence microscopy results indicating that TDP1 is enriched in the ESB.

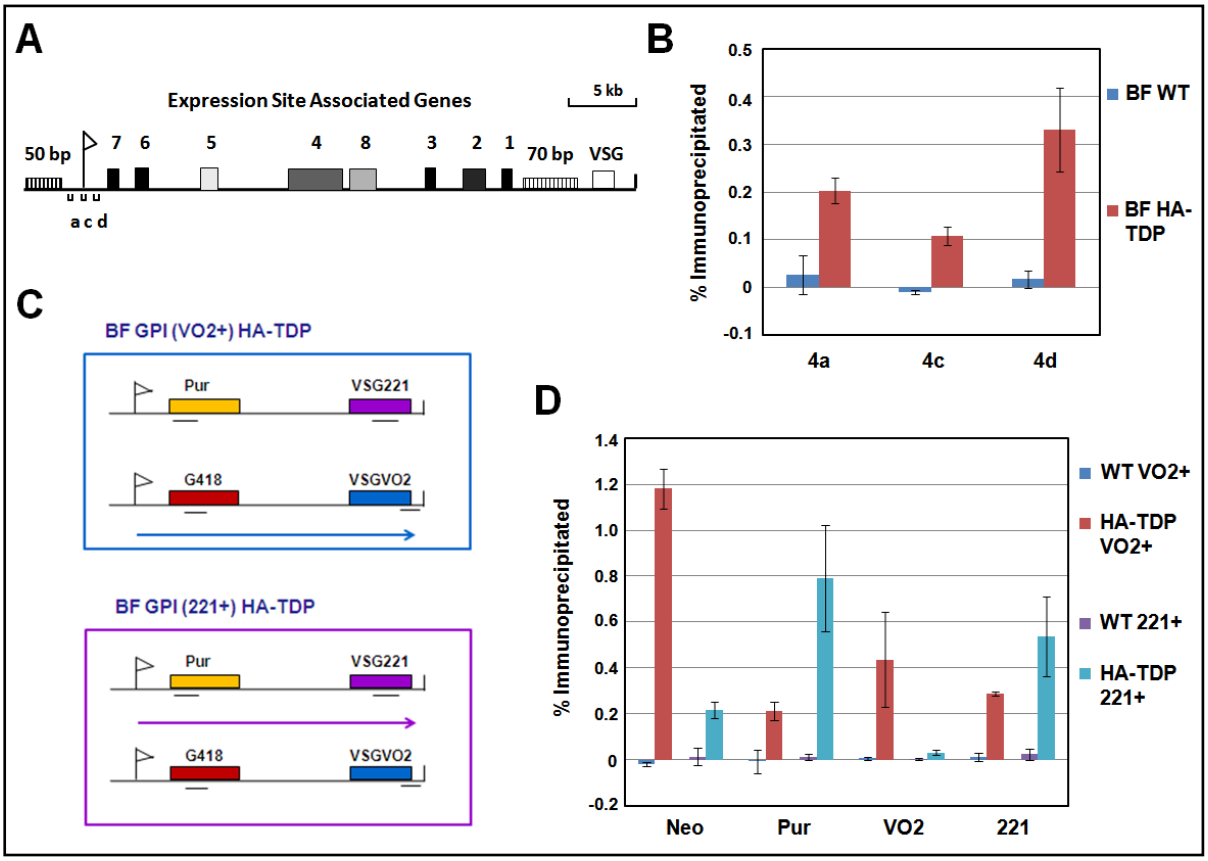


Figure 4.10 TDP1 is enriched on the active ES in bloodstream form *T. brucei*. **A)** Schematic representation of a typical *VSG* ES. Flag indicates the ES promoter and boxes represent genes. The 50 bp and 70 bp repeats (striped boxes) are also shown. Lettered brackets represent regions analysed by qPCR. Primer pairs a, c and d can be expected to recognize

sequences in most if not all ESs. **B)** Graphical representation of qPCR analysis of TDP1 binding around the ES promoter in *VSG* ESs. **C)** Schematic diagram representing two bloodstream form *T. brucei* HA-TDP lines used in ChIP experiments. BF221GPI(VO2+) HA-TDP expresses VSG VO2 in the active ES, while BFGPI(221+) HA-TDP expresses VSG 221, allowing distinction between active and silent ESs in qPCR experiments. Arrows indicate direction of transcription from the active ES. **D)** qPCR data comparing TDP1 distribution between parental BF221GPI(VO2+) (WT VO2+) and BFGPI(221+) (WT 221+) cells lacking HA-TDP1 as well as *T. brucei* clones HA-TDP VO2+ and HA-TDP 221+ which express HA-TDP1. Standard deviation from three independent repeat experiments is represented with error bars.

Depletion of TDP1 leads to a 40-90% reduction in transcription of RNA polymerase I transcription units

Our ChIP data showed that TDP1 was enriched around actively transcribed RNA polymerase I promoters including those of the rDNA and the active ES. I had earlier observed in flow-cytometry experiments that blocking the synthesis of TDP1 using RNAi did not lead to derepression of a silent reporter gene inserted into a variety of non-transcribed areas of the *T. brucei* genome including silent *VSG* ESs. Hence, I hypothesized that TDP1 was involved in promoting transcription rather than in repressing it. To investigate this I induced *TDP1* RNAi in the T3-TDP1 cell line (Figures 4.3 and 4.11A) and isolated total RNA at different time points which I used for the preparation of cDNA. I next designed a series of primers to monitor levels of nascent rRNA transcripts by qPCR. I designed primers which would only bind to precursor transcripts (White et al. 1986) rather than stable rRNA transcripts. This would allow us to measure changes in the rate of rDNA transcription (Figure 4.11B). I first monitored *TDP1* transcript levels, and as expected, I observed a rapid drop after the induction of RNAi. The level of *TDP1* transcript fell to ~20% normal levels within 12 hours of induction of *TDP* RNAi (Figure 4.11C). I next monitored changes in nascent rRNA levels. I observed an overall decrease in levels of nascent transcripts from the rDNA locus, which ranged from a decline of approximately 40% immediately downstream of the rDNA promoter

(primer pair a in Figure 4.11B and D) to a decrease of nearly 90% towards the end of the rDNA transcription unit by 48 hours induction of RNAi (primer pair d in Figure 4.11B and D). I also monitored the levels of *VSG* transcript from the active *VSGT3* ES as well as of a silent *VSG* ES containing *VSG 121*. Levels of *VSG T3* transcript fell to ~40% normal levels after 48 hours induction of RNAi while *VSG 121* transcript levels declined to ~30% normal levels within 24 hours of RNAi induction (Figure 4.10 D). These results indicate a reduction in the fully processive transcription of RNA polymerase I transcribed loci after depletion of TDP1.

I next monitored transcript levels of the RNA polymerase II transcribed genes actin, γ-tubulin and the RNA polymerase I large subunit (Figure 4.11D). The levels of these transcripts remained relatively stable after *TDP1* RNAi. These results suggest that TDP1 is a transcription facilitator involved specifically in RNA polymerase I transcription.

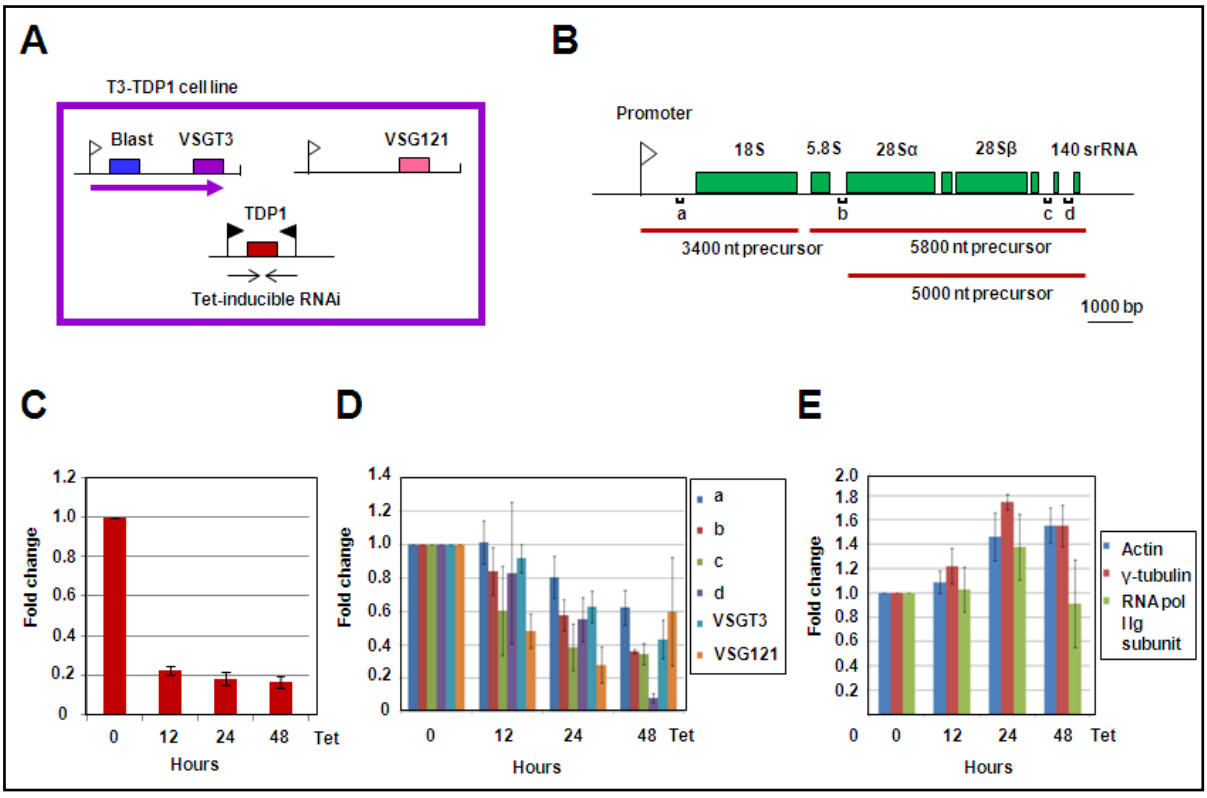
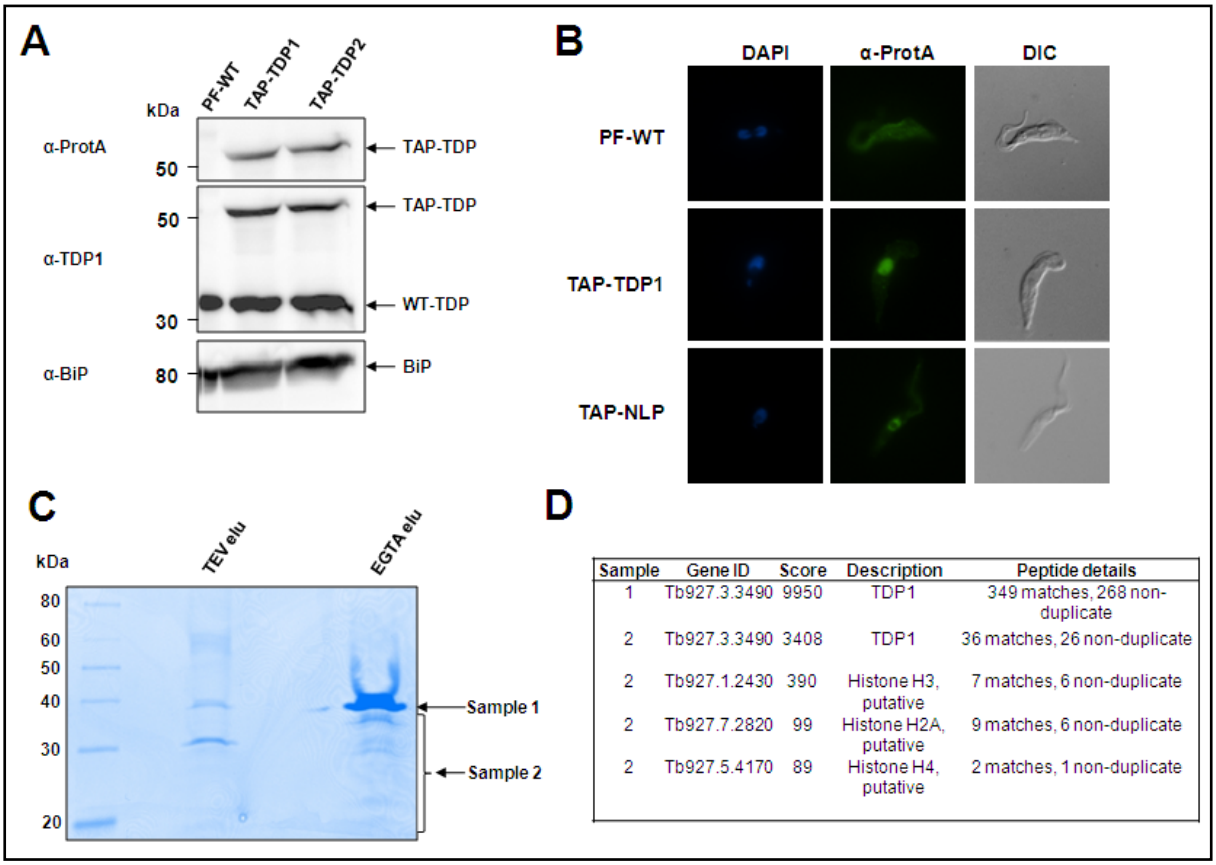


Figure 4.11 Transcript analyses in the *T. brucei* T3-TDP1 cell line after induction of RNAi against TDP1. cDNA was made from RNA collected at different time points and quantitative

RT-PCR was performed to monitor changes in transcript levels from various loci. **A)** The T3 reporter line stably expresses VSG T3 in the active ES which is under blasticidin selection. RNAi against TDP1 is induced in the *T. brucei* clone T3-TDP1 by the addition of tetracycline. **B)** Schematic representation of the rDNA locus. The various rRNA precursor transcripts are denoted by red lines and regions analysed by qPCR are shown with lettered brackets. **C)** Relative levels of TDP1 transcript after induction of RNAi. Time in hours indicates time points after induction at which RNA was isolated. **D)** Relative levels of the nascent transcripts as quantified using primer pairs shown with lettered brackets in **B**. Levels of transcript of active *VSG T3* and silent *VSG 121* were also monitored. Data shown in panels **C** and **D** has been normalized against the values for actin transcript shown in panel **E** of this figure. **E)** Analysis of levels of mRNA transcripts of the following RNA polymerase II transcribed genes after induction of RNAi against TDP1: actin, γ-tubulin and RNA polymerase I large subunit. Data shown in this panel is not normalized to levels of any other transcript. Relative levels of all transcripts were calculated by dividing the level of transcript at any given time point by the levels in uninduced (0 hour) samples. Results are the mean of triplicate RNAi induction experiments, with standard deviation indicated with error bars.

TDP1 interacts with core histones



Sample	Gene ID	Score	Description	Peptide details
1	Tb927.3.3490	9950	TDP1	349 matches, 268 non-duplicate
2	Tb927.3.3490	3408	TDP1	36 matches, 26 non-duplicate
2	Tb927.1.2430	390	Histone H3, putative	7 matches, 6 non-duplicate
2	Tb927.7.2820	99	Histone H2A, putative	9 matches, 6 non-duplicate
2	Tb927.5.4170	89	Histone H4, putative	2 matches, 1 non-duplicate

Figure 4.12 TDP1 interacts with the core histones. **A)** Western blot analysis showing TDP1 epitope tagged with PTP in the *T. brucei* clones TAP-TDP1 and TAP-TDP2. For comparison,

lysate from parental wild-type procyclic cells (PF-WT) is also shown. **B)** Panels show procyclic form *T. brucei* visualized using the DNA stain DAPI, the FITC channel after immunofluorescence (IF) using a rabbit polyclonal anti-ProtA antibody and an anti-rabbit Alexa-488 secondary antibody (α-ProtA), or visualized by differential image contrast (DIC). Immunofluorescence analysis was performed on PF-WT cells (top panels), TAP-TDP1 cells (middle panels) as well as on TAP-NLP cells (lower panels). **C)** TDP1 was tandem affinity purified (TAP) and purified material was separated by SDS-PAGE. Two bands were excised as indicated and sent for identification by mass spectrometry (MS)**. D)** The bands visualized in **C** were identified by MS. Matches with significant scores are shown**.**

I next investigated if TDP1 functions as part of a complex. To identify its partners, I epitope tagged TDP1 with a PTP-tag in procyclic form *T. brucei* to generate the two clones TAP-TDP1 and TAP-TDP2 (Figure 4.12A). I verified that the tagged TDP1 protein still localized to the nucleus. I found that TAP-TDP was abundant in the nucleus including the nucleolus, unlike the control proteins TAP-NLP which appeared to be excluded from the nucleolus (Figure 4.12B). I next tandem affinity purified TDP1 from crude extracts under native, non-denaturing conditions as previously described (Schimanski et al. 2005). The peptide eluate was separated by gel electrophoresis under denaturing conditions, and stained with Coomassie blue (Figure 4.12C). I excised the stained bands from the gel and identified the proteins by liquid chromatography-tandem mass-spectrometry (MS). A list of the peptides I obtained with significant scores is shown in Figure 4.12D. I found that TDP1 interacts with the core histones, though the relatively small number of peptides picked up suggests that this interaction could be transient. Indeed, this is also suggested by our ChIP data on the *VSG* ESs and the rDNA locus, where I see a striking inverse distribution of TDP1 compared to the distribution of histone H3. Stanne & Rudenko (2010) had shown that H3 is depleted in actively transcribed regions such as the rDNA locus and the active *VSG* ES. I repeated their experiments when I performed histone H3 ChIP simultaneously while performing TDP1 ChIP as a positive control and observed a striking inverse pattern in the distribution of TDP1 (Figures 4.9B and 4.10D) and histone H3 (Figure 4.13A-B).

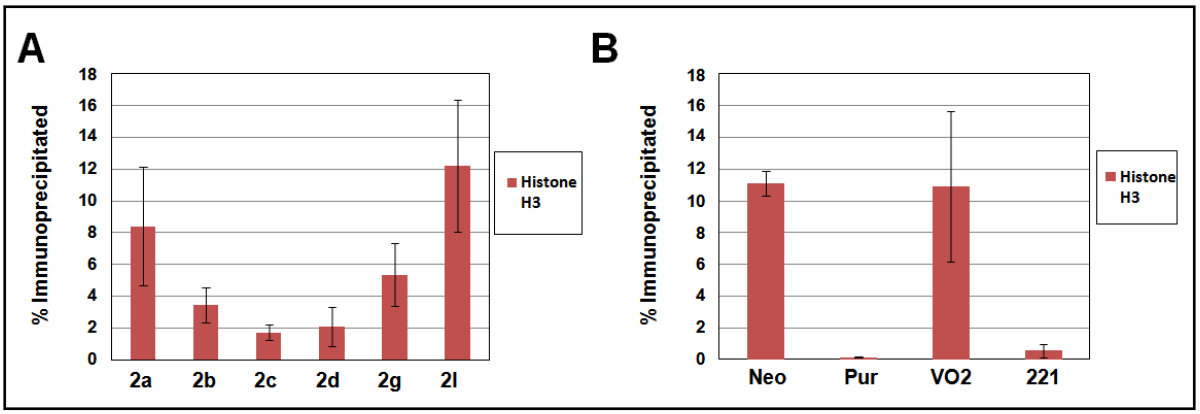


Figure 4.13 Histone H3 distribution has an inverse correlation with TDP1 distribution in the rDNA locus and active *VSG* ES. **A)** Quantification of ChIP of histone H3 on the rDNA locus as analyzed by qPCR. Primer pairs are specific to regions shown in the schematic in Figure 4.9A. H3 ChIP was performed cells at the same time as the HA-epitope tagged TDP1 ChIP on PF HA-TDP shown in Figure 4.9. **B)** Quantification of ChIP of histone H3 on the actively transcribed *VSG221* ES (in bloodstream form *T. brucei* clone HA-TDP 221+) as analyzed by qPCR. Primer pairs are specific to regions shown in the schematic in Figure 4.10C. H3 ChIP was performed on BF HA-TDP 221+ cells at the same time as the HA-TDP1 ChIP shown in Figure 4.10. Error bars show standard deviation from three independent experiments.

C. Discussion

TDP1 is an essential protein and the presence of two HMG boxes implicates a role in chromatin remodelling. I investigated if TDP1 is involved in the repression of silent ESs in bloodstream form *T. brucei*, as TDP1 had been shown to bind to VSG expression site promoter oligomer sequences (Erondu & Donelson 1992). I also investigated if TDP1 was involved in preventing transcription of the silent minichromosomes and the VSG Basic Copy arrays. However, I did not observe derepression of the reporter *eGFP* gene inserted in any of the transcriptionally inactive areas of the *T. brucei* genome in our different reporter cell lines after performing *TDP1* RNAi. Based on the striking sub-nuclear localization pattern of TDP1 which suggested that it was enriched in the nucleolus and the ESB, I investigated its genomic distribution using ChIP experiments. I found that TDP1 was abundant immediately downstream of actively transcribed RNA polymerase I promoters including the rDNA promoter and the active ES promoter. Interestingly, when I blocked synthesis of TDP1 via RNAi, I observed a 40-90% decrease in levels of RNA polymerase I transcribed precursor RNA compared to normal levels, suggesting that TDP1 plays an important role in maintaining transcription from RNA polymerase I promoters. Lastly, I performed tandem affinity purification (TAP) on TDP1 followed by mass spectrometry and found that TDP1 interacts with the core histones. These results indicate that TDP1 is an architectural chromatin protein important for transcription control in *T. brucei*.

The role of TDP1 in control of rDNA transcription

TDP1 plays a role in maintaining transcriptional activity in RNA polymerase I transcribed areas such as the rDNA transcription units and the active ES. Using immunofluorescence microscopy experiments I found that TDP1 was a nuclear protein

binding throughout the nucleus, with a striking enrichment of TDP1 in the nucleolus. These results were confirmed with ChIP experiments where TDP1 distribution within the rDNA was determined. I found that TDP1 was significantly enriched on the promoter proximal areas compared to the intergenic spacers in both bloodstream and procyclic form *T. brucei*. Together, these results provided a strong indication that TDP1 was involved in transcription control. After blocking the synthesis of TDP1 by RNAi, I observed a decrease in levels of precursor RNA transcribed from the rDNA transcription units. This decrease was more pronounced at the downstream end of the transcription unit, which showed a 90% decline in levels. In contrast, the promoter proximal region of the transcription unit showed only a 10% decline in transcript levels compared to original levels after TDP1 knock down. This provides a possible explanation for why *TDP1* RNAi is lethal in *T. brucei*, and also suggests that TDP1 plays a role in facilitating fully processive transcription of the rDNA unit.

HMG boxes do not bind DNA in a sequence specific manner. Using ChIP experiments I found that the enrichment of TDP1 at regions such as the 50 bp and 177 bp sequence repeats was only marginally above background levels, with IP values of approximately 0.04% compared to IP values of over 1% in other areas of the genome such as the active *VSG* ES. This is not surprising as TDP1 as an HMGB protein, is not expected to bind to DNA with sequence specificity, and is likely to be recruited to certain areas by other factors.

In mammals, the upstream binding factor (UBF), an architectural protein with 5-6 HMG box domains, has been shown to bind upstream and core rDNA promoter sequences, where it is involved in formation of the pre-initiation complex (Stefanovsky et al. 2001). UBF also binds throughout the rDNA transcription unit, although to a lesser extent in the intergenic spacers (O'Sullivan et al. 2002). Studies have also shown that under certain conditions UBF regulates the elongation phase of RNA polymerase I transcription (Moss et

al. 2006), and that UBF depletion leads to a decrease in the number of active rDNA transcription units as well as in a decrease in their transcriptional output (Sanij et al. 2008).

Given these similarities, is TDP1 a UBF-like protein in *T. brucei*? Apart from having only two HMG boxes compared to the 5-6 present in UBF, TDP1 also does not contain the typically metazoan HMGB feature of an acidic C-terminal tail characteristic of UBF (Stros et al. 2007). In addition, TDP1 contains the additional DNA binding domain, DEK C. It has been proposed that UBF molecules forming a dimer, could via the HMG boxes allow DNA to wrap around the complex in a structure that would resemble a canonical nucleosome (Stefanovsky et al. 2001). It is possible to investigate if TDP1 is capable of forming homo-multimers (and possibly functioning like UBF) by extracting TDP1 from *T. brucei* nuclei under non-denaturing conditions. In this situation a scenario can be imagined whereby nucleosome-like structures formed by TDP1 form in regions that are depleted of nucleosomes like the rDNA locus. This could facilitate transcription in these areas by providing an alternative mechanism to assemble and disassemble nucleosome-like structures more rapidly than conventional nucleosomes.

There is also evidence that UBF is involved in the organization of nuclear organiser regions (NORs), which are clusters of rDNA on different chromosomes together forming the nucleolus. Introducing *Xenopus* rDNA sequences with high affinity for UBF on human chromosomes resulted in the formation of NOR-like structures called pseudo-NORs at these loci (Mais et al. 2005). It would be interesting to block TDP1 synthesis by RNAi and observe changes in the nucleolar morphology by immunofluorescence microscopy using our nucleolar marker L1C6 (Devaux et al. 2007).

The role of TDP1 in regulation of the expression site body

Immunofluorescence studies have shown that RNA polymerase I is present in the nucleolus, and also at an additional site in bloodstream form *T. brucei*. It was shown by tagging the active *VSG* ES with GFP that this site was the expression site body (ESB) (Navarro & Gull, 2001). Our immunofluorescence microscopy and ChIP data indicate that TDP1 is present throughout the nucleus, but is particularly abundant in the nucleolus, and in a second nuclear compartment which is likely to be the ESB. This is supported by our ChIP experiments which showed TDP1 to be highly enriched on the active ES (IP values of 0.79 – 1.18% downstream of the active ES promoter compared to IP values of ~0.2% at the silent ESs). Very little is known about the higher order architecture of the *T. brucei* nucleus, particularly of the ESB. To the best of our knowledge, TDP1 is the first protein identified (apart from RNA polymerase I subunits) to localize to the ESB. It has been shown that the *S. cerevisiae* HMGB protein Hmo1, which is enriched in actively transcribed rDNA, but is not associated with the RNA polymerase III transcribed 5S gene (Kasahara et al. 2007) remains bound to rDNA even in the absence of RNA polymerase I (Merz et al. 2008). Possibly TDP1 is the functional equivalent of Hmo1. It remains to be seen if TDP1 is recruited to the active ES by RNA polymerase I or is able to distinguish between active and silent *VSG* ESs by some other mechanism.

Our transcript analyses showed that the transcript levels of both the active *VSG T3* and the silent *VSG 121* decreased by ~60-70% after induction of *TDP1* RNAi. It is known that silent *VSG* ESs are not entirely transcriptionally silent, but maintain a low level of activity which progressively diminishes towards the telomeres. This could possibly be due to a lack of transcription elongation machinery (Vanhamme et al. 2000), or due to a repressive gradient operating from the telomere end (Yang et al. 2009). Our observations are compatible with a situation in which TDP1 is important for maintaining transcription from all ESs.

However, it is only the active ES which is highly enriched for TDP1, helping in the maintenance of higher levels of transcription than the other ESs.

In this regard future experiments will follow the pattern of TDP1 localization over the course of the *T. brucei* life-cycle in conjunction with the ESB. I will probe cells simultaneously with antibodies against TDP1 and a suitable RNA polymerase I subunit such as the large subunit. This will allow us to verify if TDP1 indeed localizes to the ESB. Secondly, I can investigate if TDP1 is present in the ESB at all cell-cycle stages, or is only located here transiently, for example during the replication and division of the ESB present in post-mitotic cells (Navarro & Gull 2001). It would also be informative to select for activation of a new *VSG* ES using drug selection pressure and monitoring the ESB for changes.

The interaction between TDP1 and the histones

Our TAP protein affinity purification experiment to find the interaction partners of TDP1 showed that TDP1 interacts with the core histones H2A, H3, and H4. Although the numbers of unique peptides picked up was small (between one and six), the histone identification was statistically relevant due to the lengths of individual peptides identified. This suggests that the interaction between TDP1 and the core histones is transient, and possibly occurs only during events such as transcriptional activation or silencing of a region, or during DNA replication or repair. Our histone H3 and TDP1 ChIP experiments show that the two proteins have a striking inverse distribution, with TDP1 being abundant in areas such as the rDNA transcription units and the active ES which are highly depleted of nucleosomes (Figueiredo & Cross, 2010; Stanne & Rudenko, 2010).

There is abundant evidence from other systems that HMGB proteins bind to internucleosomal linker DNA, which is also bound by the linker histone H1. The inverse distribution between these two proteins is likely to be competitive. This is not surprising as that both H1 and the HMGB proteins share some features like binding to DNA without sequence or structure specificity. In addition, both proteins are highly abundant (Zlatanova & van Holde 1998), although in vertebrates H1 binds to DNA with up to 17-fold more affinity than vertebrate HMGB protein HMG1 (Hill & Reeves 1997). The conventionally prevailing view holds that H1 has a silencing effect on transcription while HMGB proteins are activating, though there are several documented exceptions to this rule (Zlatanova & van Holde 1998). Interestingly, An et al. (1998) have evidence that *Xenopus* HMG1 protects linker DNA from nuclease digestion on one side of the core nucleosome particle, while the opposite side of the particle is protected by H1. Results from other members of our laboratory indicate that the distribution of *T. brucei* H1 is similar to that of the core histones, in that it is depleted from actively transcribed regions. Additionally, the depletion of H1 by RNAi in *T. brucei* results in a roughly 10-fold derepression of the silent *VSG221* ES as monitored by flow cytometry in our T3-SM line (M. Povelones, unpublished results), while similar depletion of TDP1 does not lead to observable ES derepression. TDP1 is enriched in areas depleted of nucleosomes and H1. It is possible that one of the functions of TDP1 is to protect 'naked' DNA in these regions.

What is the relationship between TDP1 and the core histones? In *S. cerevisiae*, actively transcribed rDNA units are largely devoid of histones, and are instead associated with Hmo1 (Merz et al. 2008). Recently, Kasahara et al. (2011) have presented evidence that Hmo1 occupies a region upstream of the rDNA promoter called the intervening region (IVR) which is depleted of nucleosomes. The Hmo1 binding this region and a downstream nucleosome overlapping the transcription start site (TSS) (+1 nucleosome) together

determine the 5'- and 3'-boundaries of a zone which is available for the assembly of the pre-initiation complex. Therefore, Hmo1 occupies a site that is usually occupied by a nucleosome upstream of the TSS (-1 nucleosome) in other promoters. However, it is unclear if the relative depletion of nucleosomes from actively transcribed areas is directly caused by the HMG proteins or if it is HMG proteins which are relatively excluded from transcriptionally silent areas. It would be extremely informative in this regard to conduct experiments using the enzyme micrococcal nuclease (MNase), which digests away linker DNA allowing estimation of nucleosome occupancy. These experiments could be performed on our *TDP* RNAi cells and compared with control uninduced cells. I can envisage two scenarios. In one scenario in our *TDP* RNAi cells I would observe a decreased sensitivity to MNase. This would indicate deposition of nucleosomes on actively transcribed loci such as the rDNA due to the absence of TDP1 in the region, and could explain our transcript analysis experiments which reported decreased transcription. Alternatively, I might observe increased sensitivity to MNase, which could indicate that TDP1 does not exclude nucleosomes from forming on actively transcribed areas. This would indicate that the decreased transcription I observe from these regions is not due to transcriptional inactivation of these regions by an increased presence of nucleosomes but possibly due to the consequences of the lack of TDP1 itself, which acts as a transcription factor.

Is TDP1 a homolog of S. cerevisiae Hmo1 or vertebrate UBF?

S. cerevisiae does not encode a homolog of mammalian UBF, which contains six HMG boxes. Instead, Hmo1, which has a single HMG box, is considered to be the functional UBF equivalent (Xiao et al. 2010). TDP1 on the other hand, has two HMG boxes and a DEK C domain. In addition, some HMGB proteins interact with the basic tail of H1 through their

acidic C-terminal tail (Cato et al. 2008). TDP1 lacks this feature, but instead has a long N-terminal tract which is not found in typical HMGB proteins. Despite these differences, I find some similarities that TDP1 shares with UBF and Hmo1, in that all are enriched in the rDNA transcription units and facilitate RNA polymerase I transcription. Additionally, I have not been able to find other essential HMG box containing proteins in *T. brucei* other than TDP1 (see Chapter 5). This indicates that although TDP1, UBF and Hmo1 are likely to be functional homologs, the three proteins are not closely related to each other.

Conclusion

It is known that in eukaryotes ranging from yeast to vertebrates only about 50% of ~200 copies of the rDNA transcription units are active at any given time (Conconi et al. 1989; Sanij et al. 2008). A key issue in understanding RNA polymerase I transcriptional control is finding out how and why cells regulate the number of active rDNA units. Although it is clear that HMGB proteins like Hmo1 and UBF play essential roles in maintaining the active state of rDNA, the precise molecular mechanisms are unknown due to complications arising from difficulties in distinguishing the active and silent rDNA states. *T. brucei* provides an alternative model system for studying control of RNA polymerase I transcription due to the single copy active *VSG* ES locus transcribed. TDP1 is the first protein found to be selectively enriched at the active *VSG* ES, and I show that it plays a critical role in regulating transcription of the rDNA locus as well as the active ES. Future experiments investigating TDP1 involvement in expression of the active *VSG* ES will additionally provide insight into understanding monoallelic control of expression in *T. brucei*. This knowledge would be of high importance in understanding the pathogenicity of *T. brucei*.

Chapter 5

Investigation of other potential *T. brucei* chromatin remodelers

A. Introduction

The identification of T. brucei homologs of characterized chromatin proteins

As discussed in the previous chapters, the various molecular players involved in the stringent monoallelic control of VSG expression sites are gradually being found. Of these, the proteins TbISWI, NLP and TDP1 were found by DNA-affinity chromatography using the transcriptionally inactive 177 bp minichromosome simple sequence repeats as bait (K. Ersfeld, unpublished data; Hughes et al. 2007; Narayanan et al. 2010). Alternatively, ES regulatory proteins have also been found by conducting queries of trypanosome databases using sequences of known chromatin-associated proteins from other organisms.

For example, TbTRF, the functional homologue of mammalian TRF2 – a member of the telomere complex important for maintaining chromosome end integrity, was discovered by homology searches of the trypanosome database (Li et al. 2005). TbTRF was subsequently used as bait in yeast two-hybrid screens to find other structural telomere-associated proteins, and was found to interact with over 10 candidates. One of these was a conserved hypothetical protein which was postulated on the basis of sequence similarity to be the trypanosome ortholog to RAP1, another mammalian telomere protein. In *T. brucei*, RAP1 is involved in maintaining a silencing gradient along silent telomeric *VSG* ESs (Yang et al. 2009).

The histone methyltransferase DOT1B was also first identified on the basis of sequence homology (Horn 2007), and has subsequently been shown to be required for complete silencing of *VSG* ESs (Figueiredo et al. 2008). Similarly, the cohesion complex components SMC1, SMC3, and SCC1 were all identified on the basis of homologies with well-characterized SMC proteins of yeast and metazoa (Gluenz et al. 2008). It has since been reported that the partial depletion of SCC1 in bloodstream form *T. brucei* leads to an

increased rate of VSG switching, probably due to the premature separation of sister chromatids of the active *VSG* ES (Landeira et al. 2009).

The FACT (facilitates chromatin transcription) complex subunit TbSpt16 is another chromatin regulator that was discovered on the basis of sequence homology. TbSpt16 has been shown to be essential for *T. brucei*. Its depletion leads to a striking G2/early M cell-cycle arrest and the derepression of silent *VSG* ESs in bloodstream and insect form *T. brucei* (Denninger et al. 2010).

Querying sequence databases appears to be a robust way for discovering transcription regulators in *T. brucei*, despite the fact that trypanosomes are evolutionarily highly divergent from well characterised model eukaryotes including the yeasts (Sogin et al. 1989). I therefore identified various *T. brucei* candidate proteins and screened these proteins for involvement in transcription control using our bloodstream form T3-SM reporter cell line which has an *eGFP* gene inserted into the silent *VSG221* ES (Hughes et al. 2007). I blocked synthesis of these proteins using RNAi-mediated depletion and monitored for essentiality by tracking growth kinetics, and for derepression of silent ESs by flow cytometry. I attempted to find *T. brucei* homologues to the following protein classes: members of the RSC and SAGA complexes, and the HMG box proteins.

The RSC Complex

The *Saccharomyces cerevisiae* RSC (remodels the structure of chromatin) complex is a member of the SWI/SNF family and contains the ATPase subunit Sth1 and the core proteins Rsc8, Rsc6, Sfh1, Arp7, Arp9 (Cairns et al. 1996), Rsc4 (Sengupta et al. 2001), Rsc1 and Rsc2 (Bungard et al. 2004). Rsc4 and Rsc1 are two of the DNA-interacting proteins

of this complex and have two bromodomain motifs each (Pfam; Sengupta et al. 2001; Bungard et al. 2004). Bromodomains are motifs which recognize acetylated lysine residues such as those on the N-terminal tails on histones (Zeng & Zhou 2002), which are commonly found in many chromatin associated proteins including Gcn5, a major component of the SAGA complex (see below) (Daniel et al. 2007).

I therefore decided to search for novel *T. brucei* proteins which contained bromodomain motifs. I performed two-way BLAST searches of the trypanosome database with the yeast proteins Rsc1 and Rsc4 and were able to identify at least three proteins (Tb927.10.8150, Tb927.10.7420 and Tb11.01.1830) with bromodomain motifs with the scores of $e = 3.9 \times 10^{-3}$, $e = 2.6 \times 10^{-4}$ and $e = 1.3 \times 10^{-1}$ respectively. Additionally, I identified the same three proteins by querying the trypanosome databases with NURF301, a bromodomain containing subunit of the NURF (nucleosome remodelling factor) complex in *Drosophila* (Badenhorst et al. 2002). I provisionally named these Brom1, Brom2 and Brom3.

The SAGA and TFIID complexes

The SAGA (Spt-Ada-Gcn5-acetyltransferase) complex in *S. cerevisiae* is a histone acetyltransferase complex involved in recruitment of the primary transcription machinery. It is a multifunctional complex, and SAGA is annotated to have over 20 subunits in the yeast databases. SAGA appears to have two chief enzymatic functions through its components Upb8 and Gcn5, namely to deubiquitinate and to acetylate histones (Rodríguez-Navarro 2009), though it has a modular nature which allows this ~1.8 MDa complex to carry out other roles during transcription including histone methylation and phosphorylation. Functionally, SAGA comprises at least three types of proteins – histone acetyltransferase (HAT) comprising Gcn5; the Spt group comprising Spt3, Spt7 and Spt20/Ada5; and the TATA

binding protein-associated factors (TAFs) comprising Taf5, Taf6, Taf9, Taf10 and Taf12 (Daniel et al. 2007). Many of the TAFs were originally identified as members of the TFIID complex in *S. cerevisiae* which comprises the TATA binding protein (TBP) and the TAFs. TFIID is a transcription factor complex required for RNA polymerase II transcription, as it is involved in forming the pre-initiation complex (Tansey & Herr 1997).

A subset of the TAFs including Taf5, Taf6, Taf9, Taf10 and Taf12 are therefore a part of both the SAGA and TFIID complexes. I therefore used Taf5 in order to query trypanosome sequence databases. I performed two-way BLAST searches using the *S. cerevisiae* Taf5 sequence, and obtained high matches with two conserved hypothetical proteins Tb927.8.4210 and Tb927.10.10170 with e values of 2.8×10^{-32} and 1.1×10^{-33} respectively. These proteins similar to Taf5, contain WD40 repeats (MYHITS.ISB) which are domains also found in other proteins involved in a number of nuclear functions including transcription regulation and cell-cycle control (Smith et al. 1999). I temporarily named these two proteins Saga1 and Saga2.

HMG proteins in T. brucei

As discussed in Chapter 4, the high mobility group (HMG) family of DNA-binding proteins is one of the largest and most abundant groups of architectural non-nucleosomal proteins in metazoa (Stros 2010). The *T. brucei* HMG box containing protein TDP1 has been previously identified (Erondu & Donelson 1992). I interrogated the trypanosome database for other proteins containing an HMG box, and found 5 proteins (Table 5.1). Of these, TDP1 was the only previously identified protein, while the other four were hypothetical conserved. Of these four HMG box containing candidates, two candidates (accession numbers Tb927.10.8950 and Tb927.10.8950) are extremely small proteins (predicted size ~14 kDa) and are predicted to localize to the kinetoplast on the basis of structure and sequence. I

therefore only investigated Tb09.160.3560 and Tb927.10.8960, which I provisionally named HM1 and HM2 in our screen for proteins likely to be involved in transcription control in *T. brucei*.

Gene ID	Description	Predicted size
Tb927.3.3490	TDP1	30.87 kDa
Tb09.160.3560	hypothetical protein, conserved	52.17 kDa
Tb927.10.8960	hypothetical protein, conserved	26.87 kDa
Tb927.10.8970	kinetoplast DNA-associated protein, putative	14.27 kDa
Tb927.10.8950	kinetoplast DNA-associated protein, putative	14.18 kDa

Table 5.1 Proteins annotated with HMG domains in *T. brucei* in GeneDB. TDP1 was the only previously identified protein. Tb927.10.8970 and Tb927.10.8950 are indicated to be kinetoplast proteins and were excluded from this study.

B. Results

Sequence analysis

I selected seven proteins from the trypanosome genome database on the basis of their predicted protein motifs as candidate regulators of *VSG* ES regulation in bloodstream form *T. brucei*. Table 5.2 summarizes the candidate proteins I selected and their salient features. All of our candidate proteins were predicted to have chromatin-associated motifs such as bromodomains, WD40 repeats, or HMB boxes. In addition, Tb927.10.7420 (Brom2) and Tb927.10.8960 (HM2) were predicted to have a glutamic acid-rich region and a lysine-rich region respectively. These domains are therefore likely to be negatively or positively charged respectively, and could function as ancillary histone or DNA-binding domains functioning in conjunction with other major motifs within the protein.

Name	**Gene ID**	**Protein Motif(s)**	**Score(s) (e value)**	**Position(s)/Total size (a. a)**
Brom1	Tb927.10.8150	Bromodomain	9.2×10^{-22}	34-115/301
Brom2	Tb927.10.7420	Bromodomain Glutamic acid-rich profile	3.8×10^{-17} 7.1×10^{-02}	15-98/224 137-217
Brom3	Tb11.01.1830	Bromodomain	1.4×10^{-15}	63-136/221
Saga1	Tb927.8.4210	WD40 repeats	7.7×10^{-11} 4.3×10^{-17} 2×10^{-13} 3.8×10^{-14} 3.6×10^{-13} 2.7×10^{-16} 1.1×10^{-15} 1.3×10^{-14}	85-123/419 127-166 170-208 212-250 254-292 296-334 338-376 380-418

Saga2	Tb927.10.10170	WD40 repeats	5.9×10^{-15} 5.1×10^{-13} 5.4×10^{-16} 1.2×10^{-12} 9.7×10^{-07} 2×10^{-08}	131-169/444 173-211 215-253 257-295 299-335 404-441
HM1	Tb09.160.3560	HMG box	0.2*	87-144/451
HM2	Tb927.10.8960	HMG box Lysine-rich profile	7.0×10^{-6} 4.9×10^{-04}	162-221/237 154-237

Table 5.2 List of candidate proteins screened for involvement in *VSG* ES regulation in bloodstream form *T. brucei*. Protein motifs were analysed using MYHITS.ISB. The total predicted length of the protein in amino acids (a. a) and the position of the predicted motifs is shown in the rightmost column. Asterisk indicates a non-significant score shown by MYHITS.ISB on the present date, although the same motif was found with a significant score in August 2009.

Screening for essentiality of the candidate proteins in bloodstream form T. brucei

I used the bloodstream form *T. brucei* T3-SM reporter cell line (Hughes et al. 2007) to screen whether our candidate proteins were essential or involved in *VSG* ES regulation. The T3-SM reporter cell line (described in Chapter 2, Figure 2.4), expresses *VSG T3* from the active *VSG* ES which is selected for using blasticidin selection. It also contains *eGFP* inserted into a silent *VSG221* ES, thereby allowing monitoring for depression of the silent ES by flow cytometry. RNAi constructs containing PCR amplified fragments from our candidate genes were transfected into this cell line. I induced the RNAi mediated depletion of each of the proteins in the resultant RNAi lines and found that Brom1, Brom3 and Saga1 were all essential in bloodstream form *T. brucei* (Figure 5.1 A-C). In each of these cases, a growth arrest was seen within 8-24 hours of RNAi induction. However, I did not see any effect on

the growth of cells after the induction of RNAi against Brom2, Saga2, HM1 or HM2 (data not shown).

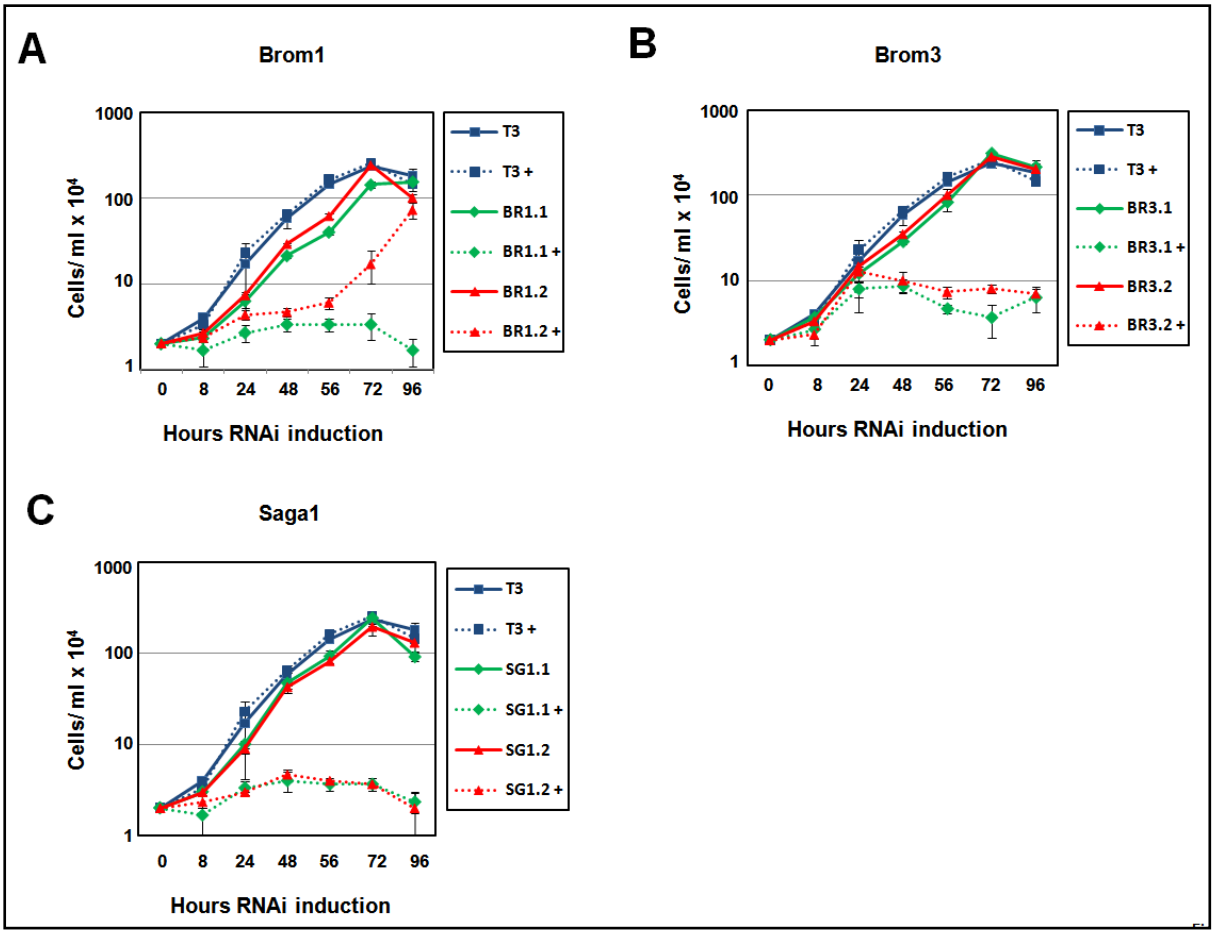


Figure 5.1 Brom1, Brom3, and Saga1 are essential for growth in bloodstream form *T. brucei*. **A)** Induction of tetracycline inducible RNAi against Brom1 in bloodstream form *T. brucei* clones T3-BR1.1 and T3-BR1.2. **B)** Induction of tetracycline inducible RNAi against Brom3 in the clones T3-BR3.1 and T3-BR3.2. **C)** Induction of tetracycline inducible RNAi against Saga1 in the clones T3-SG1.1 and T3-SG1.2. In all panels, cell lines that have been grown in the presence (+) (dotted lines) or absence of tetracycline (solid lines) are compared. The parental *T. brucei* T3 cell line (in blue) is shown for comparison. The values plotted are the mean of three different experiments with the standard deviation indicated with error bars.

Monitoring for involvement in cell-cycle progression

I next investigated if the cell-cycle arrest I observed after blocking the synthesis of Brom1, Brom3 and Saga1 occurred at any specific stage of the cell-cycle. Progression

through the cell cycle was monitored by counting the number of nuclei and kinetoplasts in tetracycline treated and untreated cells. Cells were stained with DAPI at various time points after the induction of RNAi, and subjected to microscopic analysis. A minimum of 300 cells were graded for each cell line at each time point.

The induction of RNAi against Brom1 in the *T. brucei* clone T3-BR1.1 did not lead to an accumulation of cells at any particular stage of the cell-cycle (Figure 5.2A). The cells also appeared morphologically normal, though there was a small increase (~5%) in cells which showed an abnormal karyotype with three or more nuclei. However, induction of RNAi against Brom3 produced a dramatic effect. There was a rapid accumulation of cells in the 2K1N stage, increasing from under 8% at 0 hours to over 11% after 24 hours induction and nearly 22% after 48 hours induction of RNAi. By 72 hours, nearly a third of the cells were stalled in the 2K1N stage (Figure 5.2B). There was also a smaller increase in the percentage of 2K2N cells observed, increasing marginally from ~8% at 0 hours to ~13% after 72 hours induction of RNAi.

I also observed an intriguing cell-cycle arrest linked with growth reduction after the induction of *Saga1* RNAi. There was a sharp decline in 1K1N cells, which accounted for ~85% of all cells at 0 hours but for only ~26% by 72 hours after RNAi induction. This decrease was accompanied by an increase in 2K1N and 2K2N cells, which went up from ~7% to ~26% each by 72 hours (Figure 5.2C). There was also a smaller increase in the number of cells with an abnormal karyotype, which increased from being undetectable at 0 hours to accounting for about 20% of all cells after 72 hours of RNAi induction.

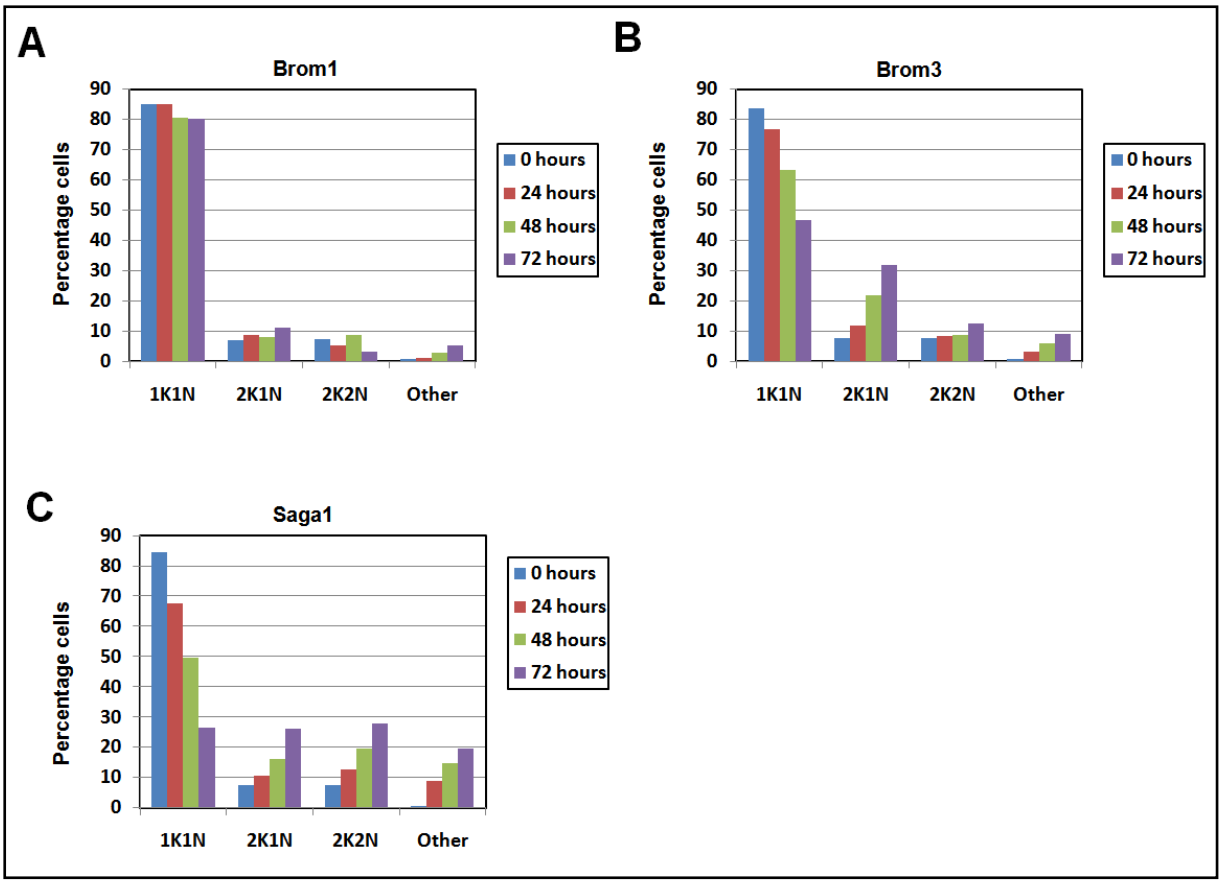


Figure 5.2 Analysis of the *T. brucei* cell-cycle after the induction of RNAi against **(A)** Brom1, **(B)** Brom3 or **(C)** Saga1 was performed using microscopic analysis of DAPI stained cells. Cell cycle stage of the cells (n>300) was categorized according to the numbers of nuclei (N) or kinetoplasts (K) present. Cells with more than 2 nuclei and/or more than 2 kinetoplasts are represented in the category 'other'. The percentage of cells in the different categories at various time points (in hours) after induction of RNAi using tetracycline is shown. The depletion of Brom3 and Saga1 but not Brom1 leads to a stage-specific cell-cycle arrest in bloodstream form *T. brucei*.

Monitoring for involvement in ES control

I next investigated if any of our candidate proteins were involved in ES repression. In addition, I also investigated the four non-essential candidate proteins for their involvement in silencing ESs. I induced RNAi against each of the candidate proteins and monitored for derepression of the *eGFP* reporter gene present in the silent *VSG221* ES in our T3-SM line

(Hughes et al. 2007). I was unable to detect derepression of *eGFP* after inducing RNAi against Brom1, Saga2, HM1 or HM2 by RNAi (data not shown).

Induction of *Brom3* RNAi led to a ~5 fold derepression of the silent *VSG221* ES after 48 hours of adding tetracycline (Figure 5.3A). Similarly, the induction of RNAi against *Saga1* led to 5-7 fold ES derepression (Figure 5.3B). As a positive control, I also induced RNAi against NLP in the T3-NLP1 cell line (Figure 2.4), which as expected resulted in over 50 fold ES derepression (data not shown).

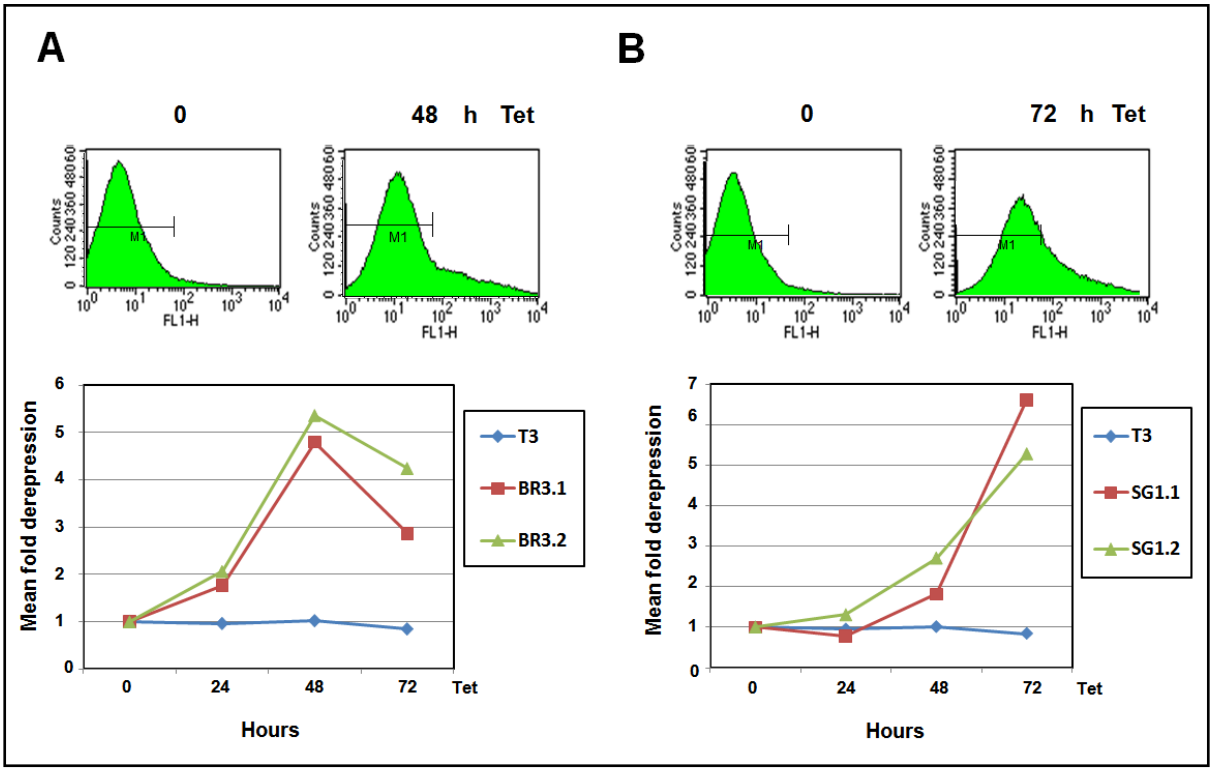


Figure 5.3 Blocking synthesis of Brom3 or Saga1 in bloodstream form *T. brucei* results in derepression of the silent *VSG221* ES. In **A** and **B** upper panels are representative flow-cytometry traces in the FL1 channel after the induction of RNAi for the time shown. Below is a graph showing the fold derepression at each time point calculated by dividing the mean fluorescence of a tetracycline induced culture with an uninduced culture at the same time point. M1 denoted in the flow-cytometry traces is the gate created to include only those cells which are part of the main peak observed. 1×10^4 cells were analysed at each time point. **A)** *Brom3* RNAi in the two clones BR3.1 and BR3.2. **B)** *Saga1* RNAi in the two clones SG1.1 and SG1.2. The parental T3-SM line (T3) is shown for comparison.

Discussion

Bloodstream form *T. brucei* avoids immune recognition by periodically switching its outermost VSG coat which is monoallelically expressed from one of ~15 telomeric *VSG* ESs (Berriman et al. 2005). Several chromatin associated proteins have been identified to be involved in controlling monoallelic expression of the ESs, but there appear to be additional layers of control mediated by as yet unidentified proteins (Rudenko 2010).

I identified seven putative chromatin-associated proteins in the *T. brucei* sequence database GeneDB on the basis of similarity to known chromatin regulatory proteins in other organisms or on the basis of being annotated with the DNA-binding domain HMG box. I investigated if these proteins were essential in bloodstream form *T. brucei*, and if these proteins were involved in cell-cycle progression and in *VSG* ES control.

The HMG proteins HM1 and HM2

The high mobility group B (HMGB) protein family constitutes an abundant class of non-histone chromatin associated DNA-binding proteins which play a role in chromatin architecture in a wide range of eukaryotes. I have found that TDP1, a previously identified HMGB protein in *T. brucei*, is involved in transcription control of the active *VSG* ES in bloodstream form cells. I tried to find additional HMGB proteins, and identified a short list of candidates. Of these, I investigated the functions of Tb09.160.3560 and Tb927.10.8960 which I provisionally named HM1 and HM2 as they were the only proteins that appeared likely to be involved in transcription control. However, on performing RNAi against these proteins I did not see any observable decrease in growth rate, or any derepression of the silent *VSG221* ES in our T3-SM reporter line. Although I have not yet verified if I was actually able

to successfully knock down the transcripts of these genes, our data is in agreement with the genome-wide RNAi targeting and sequencing (RIT-seq) phenotyping study completed recently (Alsford et al. 2011). Here these proteins did not appear to be essential in either bloodstream form or procyclic form *T. brucei*.

The RSC homologs Brom1, Brom2, and Brom3

I identified Tb927.10.8150, Tb927.10.7420 and Tb11.01.1830 (Brom1, Brom2 and Brom3) in homology searches performed against the *T. brucei* database using the *S. cerevisiae* RSC complex subunits Rsc1 and Rsc4 (Sengupta et al. 2001; Bungard et al. 2004), and *Drosophila* NURF301 (Badenhorst et al. 2002). I found that Brom1 and Brom3 (but not Brom2) are essential in bloodstream form *T. brucei*, and induction of RNAi led to a growth arrest in each case within 8 hours. These results are in contrast with the data of Alsford et al. (2011) who report that RNAi against Brom1, Brom2 and Brom3 does not result in a loss of fitness in either *T. brucei* life-cycle stage. However it has not been confirmed if the transcripts of these genes were knocked-down in this study. Interestingly, the ortholog of Brom2 in *T. cruzi*, TcBDF2 (Tc.00.1047053506553.20) has recently been shown to be a nuclear protein in different life-cycle stages, and is associated with acetylated histones H2A and H4 (Villanova et al. 2009). I have not verified if this is also the case in *T. brucei*.

The growth arrest I observed after the induction of *Brom1* RNAi was not specific to any particular stage of the cell-cycle, which suggests that Brom1 is not involved in replication or division. Depletion of Brom1 also did not result in detectable ES derepression. The putative bromodomain predicted on Brom1 ($e = 9.2 \times 10^{22}$) indicates that it is a histone-interacting protein. It is therefore possible that Brom1 functions as a general essential transcription factor that is not involved in ES control. This could be verified by monitoring

the levels of genes from various RNA polymerase I, II and III transcription units after blocking the synthesis of Brom1. There is abundant evidence that bromodomains are found in nearly all known histone acetyltransferases (reviewed in Zeng & Zhou, (2002)) and hence could be responsible for transcription activation.

I observed in the case of *Brom3* RNAi a specific cell-cycle arrest, with approximately 30% of the cells stalled in the G2/M stage, and over 10% stalled in the early M stage after 72 hours of RNAi induction. I also found that the induction of *Brom3* RNAi results in an approximately 5 fold derepression of the silent *VSG221* ES in our T3-SM bloodstream form reporter cell line (Hughes et al. 2007). This is significantly lower than the derepression observed in the same reporter line after depleting TbISWI (30-60 folds) or NLP (45-65 folds). It is possible that the observed ES derepression is due to a switch event whereby only those cells that have switched to the *VSG221* ES show a derepression phenotype, rather than a situation where most if not all cells in the population upregulate all silent ESs. Interestingly, since induction of *Brom3* RNAi also appears to affect cell-cycle progression, it is possible that depletion of Brom3 leads to increased ES switching. There is precedent for such a phenotype, as the partial depletion of the essential cohesin subunit SCC1 leads to ~10 fold derepression of a silent ES and an ~20 fold increase in VSG switching frequency (Landeira et al. 2009). It would be interesting to see if blocking Brom3 synthesis also leads to an increased rate of VSG switching. However, another possibility is that the ES derepression I observe is due to a relatively minor relaxation of the silencing mechanism present on the silent ESs. In our flow-cytometry traces I observe that the peak shift is relatively minor, making it unlikely that some cells are expressing very high levels of eGFP from an activated ES.

The RSC complex remodels nucleosomes in other organisms and it is known that a bromodomain is present in the yeast RSC complex SWI/SNF subunit Sth1 (Mohrmann & Verrijzer 2005; Cairns et al. 1996). Bromodomains are a distinguishing feature of the

SWI2/SNF2 class of ATPase subunits, but are absent in the ISWIs. However, ISWI complexes generally contain a non-ATPase bromodomain harbouring subunit, such as for example ACF1 and NURF301, in the ISWI complexes ACF and NURF respectively (Mohrmann & Verrijzer 2005). In affinity purification experiments of TbISWI (T. Stanne, unpublished results) or the TbISWI partner NLP (see Chapter 3) I have not found evidence that the TbISWI complex contains a bromodomain protein. However, it has recently been shown that both Brom3 (Siegel et al. 2009) as well as TbISWI (Stanne et al. 2011) are highly enriched in RNA polymerase II transcription start sites in *T. brucei*. It is possible that depleting Brom3 has an indirect effect through its partners in the equivalent of a *T. brucei* RSC or NURF-like ISWI complexes through removing some histones around silent ESs resulting in derepression.

The SAGA homologs Saga1 and Saga2

Tb927.8.4210 and Tb927.10.10170 (Saga1 and Saga2) were identified through similarities to *S. cerevisiae* Taf5, a subunit of both the SAGA and the TFIID complexes (Tansey & Herr, 1997; Daniel & Grant, 2007). *T. brucei* Saga1 and Saga2 showed high levels of homology to Taf5 (e values of 2.8×10^{-32} and 1.1×10^{-33} respectively). Saga2 was not essential in bloodstream form *T. brucei*, nor did it appear to be involved in the repression of silent *VSG* ESs. However, after the induction of *Saga1* RNAi I found a rapid growth arrest in bloodstream form cells within 8 hours of RNAi induction. These results are in accordance with those of Alsford et al. (2011). They found that depletion of Saga2 did not result in loss of fitness in either bloodstream form or procyclic form *T. brucei*, but there was a significant loss of fitness in bloodstream form *T. brucei* after induction of *Saga1* RNAi.

Additionally, I observed that the growth arrest observed after the induction of *Saga1* RNAi was coupled to a peculiar cell-cycle block. There was a sharp decline in the number of 1K1N cells from about 85% total in uninduced cultures to only about 26% total within 72 hours of induction of RNAi. By this time, cells with a 2K1N or 2K2N configuration accounted for over 50% total cells. I also observed a 5-7 fold derepression of the silent *VSG221* ES at the same time point. *S. cerevisiae* Taf5 (which was used in homology searches to find Saga1) is associated with large multi-member complexes. It would be informative to determine if Saga1 is also part of a *T. brucei* complex using a technique such as tandem affinity purification (TAP), as this would help us understand if Saga1 is also part of SAGA and TFIID-like complexes in *T. brucei*.

Conclusion

How similar are our putative chromatin remodelers Brom1, Brom3 and Saga1 to RSC and SAGA complex subunits in other eukaryotes? The two-way BLAST searches and the protein domains identified on our candidate proteins show that our *T. brucei* proteins were highly similar in terms of sequence identity to the *S. cerevisiae* proteins used to find them. However, the relatively high scores obtained in these searches are largely due to the presence of domains such as the bromodomain motif and the WD40 repeats. The regions of protein extending outside these domains shared little identity. Therefore although our candidate proteins could be functional homologs of the *S. cerevisiae* proteins, I have no clear evidence that they are also orthologs. The next step would be to obtain further functional data on these proteins such as determining their sub-nuclear localization pattern and life-cycle stage expression profiles, as well as finding interacting partners.

T. brucei has a number of putative as well as identified genes coding for histone modifying proteins such as histone acetyltransferases, deacetylases and methyltransferases. Of these, some histone deacetylases have been found to be essential, and are involved in cell-cycle progression (Ingram & David Horn 2002). Certain acetyltransferases (Kawahara et al. 2008) and deacetylases (Wang et al. 2010) have been shown to be important for *VSG* ES control. Further studies would help us determine if Brom1 and Brom3, as putative acetyl-lysine binding proteins, and Saga1, as a member of a putative *T. brucei* acetyltransferase SAGA-like complex, are involved with any of these previously characterised histone modifying proteins in regulating transcription in *T. brucei*.

Chapter 6

Conclusions

The goal of this thesis

RNA polymerase II transcription is predominantly constitutive in *T. brucei* and is regulated mainly post-transcriptionally. Hence as expected, *T. brucei* displays a relative paucity of histone post-translational modifications and chromatin-modulating enzymes related to controlling RNA polymerase II transcription (Figueiredo et al. 2009). In addition, unusually for a eukaryote *T. brucei* uses RNA polymerase I to transcribe a subset of protein coding genes in addition to the *rRNA* genes. These genes encode the major surface proteins, VSG and procyclin in bloodstream and procyclic form *T. brucei* respectively. Bloodstream form *T. brucei* expresses a single VSG at a time from a repertoire of ~1200 *VSG* genes and pseudogenes, from one of ~15 telomeric *VSG* expression sites (Berriman et al. 2005; Hertz-Fowler et al. 2008). It was long disputed if the mechanism operating behind the strict monoallelic expression of VSG was epigenetically regulated (Rudenko 2010).

However, recently a wealth of information concerning the epigenetic regulation of transcription in *T. brucei* has been obtained. These discoveries can be grouped into three categories. First, chromatin remodelers such as TbISWI and the FACT (facilitates chromatin transcription) complex subunit TbSpt16 have been shown to play crucial roles in maintaining appropriate levels of RNA polymerase I transcription. This is particularly the case for the *VSG* ESs, in bloodstream and procyclic form *T. brucei* (Hughes et al. 2007; Denninger et al. 2010). Second, histone modifying enzymes such as the histone deacetylase SIR2rp1 and the histone acetyltransferase HAT1 have been shown to be involved in silencing telomeres (Alsford et al. 2007; Kawahara et al. 2008). The histone methyltransferase DOT1B is involved in the repression of the entire length of the telomeric *VSG* ESs (Figueiredo et al. 2008), while the histone deacetylase DAC3 represses *VSG* ES promoters (Wang et al. 2010). Third, it appears that histone variants play a role in modulating RNA polymerase II transcription. The variants H2AZ and H2BV are highly enriched at transcription start sites

(TSSs) while the variants H3V and H4V are enriched at transcription termination sites (TTSs) (Siegel et al. 2009).

Despite the progress made, it is still not known how these multiple regulatory layers interact with each other in modulating RNA polymerase I transcription in *T. brucei*. It is also not known how the active *VSG* ES is marked by epigenetic factors, allowing monoallelic expression. This thesis attempts to improve our understanding of these areas.

The epigenetic landscape of VSG ES control

I investigated the role of NLP, a novel nucleoplasmin-like protein, in *VSG* ES regulation in bloodstream form *T. brucei*. I found that NLP plays a role in the repression of transcriptionally inactive areas such as the silent *VSG* ESs. Intriguingly, NLP is also involved in maintaining transcriptional activity from the active *VSG* ES. I also investigated TDP1, an HMG box domain containing DNA-binding protein. I found that TDP1 is significantly enriched in the nucleolus in both bloodstream and procyclic form *T. brucei*. In addition, TDP1 is significantly enriched on the active *VSG* ES in bloodstream form *T. brucei*, making it the first non-RNA polymerase I subunit protein to be found enriched here. Blocking TDP1 synthesis results in a 40-90% reduction in levels of RNA polymerase I transcription including of the rRNA and *VSGs* located in ESs.

The regulation of *VSG* ESs in bloodstream form *T. brucei* is a multi-layered process. How do our results fit with the available literature? So far, a number of proteins such as TbISWI, TbSpt16, and DAC3 have been shown to be involved in repression of the promoter-proximal regions of inactive *VSG* ESs (Hughes et al. 2007; Denninger et al. 2010; Wang et al. 2010). TbRAP1 has been shown to create a silencing gradient extending from the telomere to

the ES promoter (Yang et al. 2009). I have now shown that blocking NLP synthesis also results in the derepression of silent *VSG* ESs. The extent of derepression I observed is, along with TbISWI, the highest reported so far. The significance of this observation became more clear when I found that NLP was an interacting partner of TbISWI. Judging by the similar levels of ES binding observed using ChIP experiments, both proteins could be part of a complex which silences *VSG* ESs. Therefore, the mechanism of downregulation of transcription from silent ESs is gradually becoming more clear – factors such as the TbISWI and the FACT complexes repress promoter activity, while proteins such as TbRAP1 and Sir2rp1 are part of a repressive gradient extending from the downstream telomere end of the ESs.

However, how transcriptional activity from the active ES is maintained remains mysterious. It is known that the active *VSG* ES is transcribed in an extra-nucleolar RNA polymerase I transcribed body called the expression site body (ESB). The ESB appears to be a distinct sub-nuclear compartment. It is not solely comprised of RNA polymerase I components, as DNase digestion of nuclei does not result in its degradation (Navarro & Gull 2001). I have now identified TDP1 as the first protein which is found to be preferentially enriched on the active *VSG* ESs. It is known that the active *VSG* ES is highly depleted of nucleosomes (Figueiredo & Cross 2010; Stanne & Rudenko 2010). It has been proposed that the HMG box domains of proteins like mammalian UBF can bind DNA forming structures resembling canonical nucleosomes (Stefanovsky et al. 2001). TDP1 could behave in a similar way, thereby providing a model for the physical structure of the ESB. The depletion of TDP1 by induction of RNAi results in an approximately 60% decline in levels of transcript for the active *VSG*. It is not yet clear if this decline in *VSG* transcript is due to a primary effect of TDP1 acting as a transcription factor, or if it is due to a secondary effect arising from structural damage to the ESB. I have planned experiments to monitor the changes in the ESB

and nucleolus using an antibody against an RNA polymerase I subunit after the induction of *TDP1* RNAi.

However it is clear that TDP1 does not function alone at the active ES. So far, only TbSpt16 (the large subunit of the FACT complex) has been shown to be important for maintaining transcription from the active ES. Blocking synthesis of TbSpt16 resulted in a sharp decline in levels of transcript from the active *VSG* to less than 1% normal levels after 48 hours induction of *TbSpt16* RNAi. This is an intriguing result, especially as TbSpt16 is significantly depleted on active *VSG* ESs compared with silent ESs (Denninger et al. 2010), suggesting that an exclusion mechanism operates in the ESB. Our results show that NLP is also important for maintaining transcriptional activity from the active ES. Blocking NLP synthesis also resulted in a decline in levels of transcript from the active *VSG* to less than 4% normal levels within 48 hours induction of RNAi. In contrast to the FACT complex, the distribution of NLP on the active and silent *VSG* ESs appears approximately equal. Our TAP experiments suggest that NLP is a partner of another protein, TbRCC1, in a complex which lacks TbISWI. It is possible that this NLP complex along with the FACT complex facilitates transcriptional processivity in the active ES.

Figueiredo et al. (2008) have shown that bloodstream form *T. brucei* lacking the histone methyltransferase DOT1B take significantly longer to switch from activation of one ES to another. TheseΔDOT1B cel ls simultaneously express two VSGs from two different ESs. Moreover, this state of activation of two active ESs is stably inherited in dividing cells. These authors also found no evidence that theseΔDOT1B cells contained an additional ESB, indicating that the two active ESs localize to the same ESB. Chaves et al. (1999) have also shown that it is possible to obtain cells expressing VSGs from two ESs by forcing simultaneous activation using two different drug markers on the ESs. Here too, it appeared that the two activated ESs were physically in close proximity to one another. Interestingly,

neither study has reported a simultaneous high level of expression of both VSGs. In the first study, the number of molecules of the second VSG ranged from 5-90% of the number of VSG molecules in a wild-type cell, while the number of molecules of the first VSG was comparable to the number of VSG molecules in a wild-type cell. In the second study, cells did not stably express two VSGs but instead appeared to rapidly switch (minimally once every six generations) between the two ESs.

What prevents the simultaneous maximal expression of two *VSG* ESs? It has been proposed that an unidentified regulation mechanism prevents two VSG Es from being fully transcribed, but allows one ES to be fully active while the second ES is partially active (Figueiredo et al. 2008). It is possible that the amount of TDP1 binding to an ES functions as the deciding factor as to how much transcription occurs at that ES. This suggestion is supported by our observation that I find a smaller amount of TDP1 binding to the silent ESs, and that the low basal levels of 'silent' VSG transcript reduce even further after induction of *TDP1* RNAi. Perhaps high levels of TDP1 are recruited to the ESB, and TDP1 is a limiting factor for ESs that are outside the ESB. The basal amounts of TDP1 on the silent ESs could be indicative of a 'pre-active' state. Perhaps when inactive ESs are in close proximity to the ESB, as indicated in the experiments of Chaves et al. (1999), they are partially activated.

Examination of levels of TDP1 on the two active ESs of ΔDOT1B cells using ChIP will provide an indication if this is the case. It would answer if TDP1 is also partially enriched on the partially active ES. It would also be interesting to induce RNAi against *NLP* (or *TbSpt16*) in ΔDOT1B double expresser cells. I can envisage two likely scenarios in this case. In the first scenario, the transcript levels of both *VSGs* would decline, indicating that NLP (or TbSpt16) is important for fully processive transcription from *VSG* ESs in general. In the alternative scenario, the transcript levels of only the fully active *VSG* would decline, thereby resulting in a relative increase in the proportion of transcripts from the partially active

VSG. This would indicate that NLP (or TbSpt16) is somehow specifically able to facilitate transcription from the active *VSG* ES.

The link between VSG ES regulation and cell cycle control

In Chapter 5, I searched for homologues of known chromatin proteins from other eukaryotes in *T. brucei*. I first identified Brom3 by interrogating the *T. brucei* sequence database with the *S. cerevisiae* RSC SWI/SNF-containing chromatin remodelling complex subunits Rsc1 and Rsc4 (Sengupta et al. 2001; Bungard et al. 2004), and with the *D. melanogaster* ISWI complex subunit NURF301 (Yadon & Toshio Tsukiyama 2011). Brom3 is predicted to contain a bromodomain, which is a motif that recognizes acetylated lysine residues such as those on the N-terminal tails on histones (Zeng & Zhou 2002). I also found Saga1, by searching for *T. brucei* homologues of Taf5, which is a component of the *S. cerevisiae* SAGA acetyltransferase complex (Daniel et al. 2007). I found that both Brom3 and Saga1 were essential in bloodstream form *T. brucei*. Strikingly, inducing RNAi against either of these proteins led to arrest in the G2/M phase of the cell cycle. In both cases, this arrest was coupled with an approximately 5-10 fold derepression of the silent *VSG* ES in our T3-SM reporter bloodstream form *T. brucei* line (Hughes et al. 2007). This is relatively minor compared with the 45-65 fold derepression observed after inducing RNAi against NLP.

What do we know about the role of chromatin-modifying factors in regulating cell cycle progression in *T. brucei*? Cell cycle regulation appears to be divergent in *T. brucei* compared with other model eukaryotes. Regulation varies between lifecycle stages, and mitosis and cytokinesis appear to be controlled by separate mechanisms. Additionally, many important cell cycle regulators found in other organisms are not conserved in *T. brucei*, which appears to utilize different checkpoints when progressing through the cell cycle (Hammarton

2007). For example, blocking the synthesis of VSG in bloodstream form *T. brucei* leads to a specific pre-cytokinesis arrest (Sheader et al. 2005). There is at least one other example of a protein which is required for cells to pass through the G2/M checkpoint. Depleting the FACT complex subunit TbSpt16 in *T. brucei*, results in an accumulation of cells in the G2/early M phase. This arrest is accompanied by 16-25 fold derepression of the silent *VSG* ES in the T3-SM reporter line, specifically in those cells stalled in the G2/early M stage (Denninger et al. 2010). This is not a general consequence of the cell cycle arrest, as blocking synthesis of the mitotic cyclin Cyc6 also results in a G2/M block in *T. brucei* (Hammarton et al. 2003), which however is not accompanied by derepression of silent *VSG* ESs (V. Denninger, unpublished results).

It appears that there is a link between *VSG* ES regulation and cell cycle progression. Evidence points in the direction of this link being histone post-translational modifications. The histone deacetylase DAC4 is required for normal cell cycle progression and its depletion leads to an accumulation of cells in the late G2 phase (Ingram & David Horn 2002), while DAC3 is involved in repression of *VSG* ES promoters (Wang et al. 2010). Depletion of the histone acetyltransferases HAT1 and HAT2 leads to the arrest of cells post-mitotically, while HAT1 also silences telomeres (Kawahara et al. 2008). The methyltransferases DOT1A and DOT1B are responsible for dimethylating and trimethylating H3K76 respectively. Normally, the dimethylated state is only observed in mitotic cells, while the trimethylated state is seen throughout the cell cycle. The depletion of DOT1B leads to a dimethylated H3K76 state throughout the cell cycle, the derepression of silent *VSG* ESs, and an increase in time needed to switch between *VSG* ESs (Figueiredo et al. 2008). The depletion of DOT1A however disrupts a mitotic checkpoint resulting in an accumulation of haploid cells (Janzen et al. 2006).

It remains to be seen if any of the proteins I have investigated are capable of modifying histones, or if they are associated with histone modifying protein partners. It is possible that the G2/M arrest seen after inducing RNAi against Brom3, Saga1 and TbSpt16 are a consequence of these proteins or their interacting partners being involved in modulating transcription of cell cycle regulators (such as Cyc6) via histone PTMs. Further studies investigating the similarities and differences of the *T. brucei* histone code compared with other model eukaryotes will improve our understanding on this subject.

Future directions

An important point which needs consideration in all these experiments, particularly in those utilizing procedures such as ChIP and TAP is that they represent an average 'snapshot' of the events in the cell at a given time point, rather than a dynamic picture. It is possible that I have failed to detect transient or relatively weak interactions and patterns. Further studies will address this issue.

DNA adenine methyltransferase identification (DamID) is a powerful method for mapping the potential binding sites of a protein. Here, the protein of interest is tagged with *Escherichia coli* DNA adenine methyltransferase. When this is expressed *in vivo*, the adenines surrounding the native binding sites of the protein are preferentially methylated. These can then be mapped using sensitivity of genomic DNA to certain methylation-specific restriction enzymes and by microarray hybridization (Massie & I. G. Mills 2008). ChIP-sequencing (ChIP-seq) allows large-scale identification of binding sites of the chromatin protein in question within the genome, via the sequencing of ChIP material. This has an advantage over standard ChIP assays which utilize qPCR as a means of investigating protein binding patterns over a candidate region. For example, it was found that only ~4% of

oestrogen receptor-α binding sites were located within 1 kb of a TSS, and instead were mostly found in regions that had not been previously associated with transcription regulation (Massie & I. G. Mills 2008). This would provide indication for novel regulatory regions, and whether loci such as the *VSG* ESs are regulated by mechanisms such as DNA-looping.

It has been reported that out of the ~7500 protein-coding genes in *T. brucei*, 1187 are essential in bloodstream form *T. brucei* and nearly 50% of these are currently annotated as 'hypothetical' or 'conserved hypothetical' in the *T. brucei* sequence databases (Alsford et al. 2011). It may be possible to infer functions for a subset of these proteins via homology comparisons with chromatin-proteins from other eukaryotes. However, *T. brucei* branched off early in the tree of eukaryotic evolution (Embley & Martin 2006), and is one of the few organisms which utilizes RNA polymerase I to express protein-coding *VSG* genes. It is therefore likely that some of these proteins will be novel components of trypanosome-specific machinery, such as that involved in monoallelic *VSG* expression. The main challenge in the future will be to develop a working model for how the recently reported ES-regulating proteins and the 'to-be' identified proteins function together to help unravel this web of interactions.

Chapter 7

Materials and Methods

1. *Trypanosoma brucei* cell lines

Bloodstream form T. brucei

T. brucei subsp. *brucei* strain Lister 427 was used for all experiments. Cells were cultured at 37 °C with 5% CO_2 in HMI-9 medium (Hirumi & Hirumi 1989) supplemented with 20% heat-inactivated foetal bovine serum.

Bloodstream form *T. brucei* S16 (Wirtz et al. 1999) and its derivatives were used as reporter cell lines to monitor for derepression of *eGFP* inserted in various genomic locations following knock down of candidate proteins in screening experiments. *T. brucei* S16 contains genes encoding T7 RNA polymerase and tetracycline repressor (TetR). The MC^{177} (Sheader et al. 2004) NLP hygro construct (prepared by M. Wand) was integrated into a minichromosome of the *T. brucei* T3-SM line, which has an active *VSGT3* expression site maintained on blasticidin selection and a reporter *eGFP* gene (Clontech) inserted into the silent *VSG221* expression site (ES) (Hughes et al. 2007) to generate the *T. brucei* T3-NLP1 and T3-NLP2 clones. The *T. brucei* 121-SM is an analogous cell line which expresses VSG 121 instead of VSG T3, and its derivative clones 121-NLP1 and 121-NLP2 were similarly obtained by transfecting them with the MC^{177} NLP hygro construct. The MC^{177} NLP-ISWI double RNAi construct was transfected into the T3-SM line to generate the clones T3-DB1 and T3-DB2.

Similarly, the MC^{177} TDP1 hygro construct (prepared by M. Wand) was transfected into the *T. brucei* T3-SM line to generate the T3-TDP1 and T3-TDP2 clones. The MC^{177} Brom1, Brom2, Brom3, HM1, HM2, Saga1, and Saga2 hygro constructs were also transfected into the T3-SM reporter lines to generate the various RNAi cell lines reported in this study.

VSG RNAi was performed in the *T. brucei* 90-13 line (Wirtz et al. 1999) expressing VSG 221 as described in Sheader et al., (2005). The MC^{177} eIF2K2 RNAi construct was prepared and transfected into the *T. brucei* S16 line to generate the S16 TbeIF2K2 RNAi clone (N. Wand, unpublished results).

The MC^{177} construct (Sheader et al. 2004) can be selected for using hygromycin. The MC^{177} construct was modified such that it contained a phleomycin resistance gene instead of hygromycin. This modified construct was used to prepare the MC^{177} NLP phleo construct, which was transfected into the *T. brucei* 90-13 line to generate the clones NLP C1 and NLP C2. Similarly, the MC^{177} TDP1 phleo construct was made and transfected into the *T. brucei* 90-13 line to generate the clones TDP C1 and TDP C2.

The *T. brucei* MCEP reporter line had a construct containing an *eGFP* gene and an upstream ES promoter inserted into the minichromosomes, the VBEP line had the construct inserted in a VSG Basic Copy array and the EP1 line had the construct in the procyclin locus. These lines are all derived from *T. brucei* S16.221, a derivative of the S16 single marker cell line. S16.221 was obtained by transfecting S16 with the pHNES221Pur1.6 construct (T. Isobe, unpublished) which results in the integration of a puromycin resistance gene downstream of the active *VSG221* ES promoter. The *T. brucei* MCe and VBe lines are identical to the MCEP and VBEP lines, only the former two lines lack the ES promoter upstream of the reporter *eGFP* gene. These five *T. brucei* reporter lines were made by M. Kushwaha (Stanne et al. 2011). These reporter lines were transfected with the MC^{177} NLP hygro and MC^{177} TDP1 hygro constructs to generate the various MCEP, MCe, VBEP, VBe and EP1 clones used in this study.

T. brucei 221GPI(VO2+) has an active *VSGVO2* ES wherein the activated state is selected for by G418. This cell line also has an *eGFP* gene and a puromycin resistance gene

in the silent *VSG221* ES. The *T. brucei* 221GPI(221+) strain is an isogenic line with an active *VSG221* ES which can be selected for using puromycin (Sheader et al. 2004). 221GPI(VO2+) was transfected with the pMOTagHA-NLP and pBSphleoNLPKO constructs in order to epitope-tag or to genetically knock out an endogenous NLP allele respectively, thereby generating the HA-NLP lines and the HA-NLPKO lines. Similarly, the pMOTagHA-TDP construct was transfected into the *T. brucei* 221GPI(VO2+) and 221GPI (221+) cell lines in order to epitope tag TDP1 with HA, generating the various HA-TDP VO2+ and HA-TDP 221+ clones.

T. brucei HA-NLP was also transfected with the pMOTagMYC-FYRN1 and the pMOTagMYC-TbRCC1 constructs to MYC epitope-tag the FYRN1 and TbRCC1 proteins respectively. This transfection generated the *T. brucei* HA-NLP/MYC-FYRN1 and HA-NLP/MYC-TbRCC1 lines used in co-immunoprecipitation studies.

The *T. brucei* 90-13 line (Wirtz et al. 1999) was also transfected with the pMOTagHA-FYRN1 and the pMOTagHA-TbRCC1 constructs to generate *T. brucei* HA-FYRN1-2 and HA-TbRCC1-2. HA-FYRN1 and HA-TbRCC1 were transfected with the MC^{177} FYRN1 phleo and MC^{177} TbRCC1 phleo constructs (these two RNAi constructs prepared by S. Wiesler) to generate the *T. brucei FYRN* RNAi clones HA-FYRN.C1 and HA-FYRN.C2 and the *T. brucei TbRCC1* RNAi clones HA-TbRCC.C1 and HA-TbRCC.C2 clones respectively.

RNAi was induced in cells with 750 ng/ml tetracycline for as many hours as specified.

Procyclic form T. brucei

Wild-type procyclic *T. brucei* 427 (AmWt) was maintained at 27 °C in SDM-79 medium (Brun et al. 1979) supplemented with 10% FCS and 5 mg/L haemin. These cells were transfected with the above described pMOTagHA-NLP or pMOTag-HA-TDP constructs in order to epitope tag the corresponding proteins with HA. The second allele of the PF-HA-NLP lines was subsequently knocked-out via transfection with the pBSphleoNLPKO construct to ensure functionality of the epitope-tagged copy. *T. brucei* PF-HA-NLP was also transfected with the pMOTagMYC-FYRN1 and the pMOTagMYC-TbRCC1 constructs to generate the *T. brucei* PF-HA-NLP/MYC-FYRN1 and PF-HA-NLP/MYC-TbRCC1 clones.

For tandem affinity purification (TAP), NLP was epitope tagged at the C-terminus with a protein C epitope/TEV protease cleavage site/protein A (PTP) epitope (Schimanski et al. 2005) through transfection of the pC-PTP-hygroNLP construct. To ensure functionality of the NLP-PTP protein, cell lines were made where the second allele of NLP was knocked out using the pBSphleoNLPKO vector. TDP1 was similarly tagged at the C-terminus with the PTP epitope using the pC-PTP-hygroTDP1 construct. Functionality of the PTP tagged NLP was established by knocking out the endogenous NLP allele as described above.

2. Constructs and sequences

DNA constructs

The various constructs prepared during this study are listed in Table 7.1.

Table 7.1 List of constructs made during this study. The RNAi construct targeted to the minichromosomes is as described in Wickstead et al. 2002. pMOTag hygromycin constructs used to HA-tag or MYC-tag proteins of interest as described in Oberholzer et al. 2006. The pMOTag blasticidin vectors used to HA-tag or MYC-tag proteins of interest are adapted from the pMOTag hygromycin vector (A. Fullbrook, unpublished results). The pC-PTP epitope-

tagging vector as described in Schimanski et al. 2005.			
Construct name	**Relevant gene**	**Primer Sequences**	**Construct description**
MC[177] NLP phleo RNAi	NLP	5' tttgctcgagcgatgcttgtgtccg 5' gacaggatccttgaggcggttactg	Construct targeted to the minichromosomes (Phleomycin selection)
MC[177] TDP1 phleo RNAi	TDP	5' gcggctatagatgacattgttg 5' ctcctcgtagaccttcctctcc	Construct targeted to the minichromosomes (Phleomycin selection)
MC[177] NLP hygro RNAi	NLP	5' tttgctcgagcgatgcttgtgtccg 5' gacaggatccttgaggcggttactg	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] TDP1 hygro RNAi	TDP1	5' gcggctatagatgacattgttg 5' ctcctcgtagaccttcctctcc	Minichromosome-targeting RNAi vector (Hygromycin selection)
MC[177] NLP-TbISWI double RNAi	NLP and TbISWI	<u>NLP:</u> 5' tttgctcgagcgatgcttgtgtccg 5' gacaggatccttgaggcggttactg <u>TbISWI:</u> 5' tatctagaaccggtgcagttttatacgg 5' tactcgaggacgctgccactagtgatga	Construct targeted to the minichromosomes (Hygromycin selection). The NLP fragment was first integrated into the construct. The TbISWI construct was later inserted using blunt-end ligation.
MC[177] HM1 hygro RNAi	HM1	5' catctcgagatgccgctgtttgtgcagttgg 5' gtaggatcccatggcccctatttcacgcgcg	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] HM2 hygro RNAi	HM2	5' catctcgagtgttgcgttttacggttgcgcta 5' gtaggatccttcgtttggtgggcgccttctt	Construct targeted to the minichromosomes (Hygromycin selection)

MC[177] Brom1 hygro RNAi	Brom1	5’ gagctcgaggtgaagctgctatggactcg 5’ gatggatccctgtggaacgttttgctctt	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] Brom2 hygro RNAi	Brom2	5’ gagctcgagcatttgcgtatccacgtctg 5’ gatggatcctcatcgtcctcctcatcact	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] Brom3 hygro RNAi	Brom3	5’ gagctcgaggtcctcacaggggtcaaact 5’ gagggatccattgctctcaggtgcttcac	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] Saga1 hygro RNAi	Saga1	5’ gagctcgagttcttgctgcgatactaccc 5’ gatggatccacaattcttgccctgtttca	Construct targeted to the minichromosomes (Hygromycin selection)
MC[177] Saga2 hygro RNAi	Saga2	5’ gagctcgagaaaggagaatgcgcagctt 5’ gatggatccttcaaggtcccaacactt	Construct targeted to the minichromosomes (Hygromycin selection)
pBSphleo-NLP KO	NLP	<u>Upstream Fragment:</u> 5’ tagtctagagtgattacattccctactttc 5’ agggatccttatccgcctggtacttccct <u>Downstream Fragment:</u> 5’ tagaagctttgggtgttaggaaaccttcaa 5’ tagctcgagtcaccaagcaaatcaatacga	Knock-out construct that replaces targeted gene with Phleomycin gene
pMOTagHA-NLP	NLP	5’ gttggtacctctcctgaacacgggactttcac 5’ caactcgagcaactccagccttagcttgttc	HA epitope-tagging construct (Hygromycin selection). Integrates using a single-crossover.
pMOTagHA-TDP1	TDP1	5’ catggtaccgcggctatagatgacattgttg 5’ catctcgagcttgggtagtgctgccatttccctc	HA epitope-tagging construct (Hygromycin selection). Integrates using a single-crossover.
pMOTagHA-	FYRN1	<u>Upstream Fragment:</u> 5’ atcaggtaccatgcgaagcgacaaatgctg 5’ agatctcgagcgatgagtctgggttgctgt	HA epitope-tagging construct (Blasticidin selection).

FYRN1		<u>Downstream Fragment:</u> 5’ atcaggatccgggcacttaggctcagttgg 5’ atttgcgtctttgtcttgcggagctccaat	Integrates using a double-crossover.
pMOTagMYC-FYRN1	FYRN1	As above	MYC epitope-tagging construct (Blasticidin selection). Integrates using a double-crossover.
pMOTagHA-TbRCC1	TbRCC1	<u>Upstream Fragment:</u> 5’ agctggtaccacagtaccttggtccatcagt 5’ gaaactatgtcgtcggaaaactcgagtacc <u>Downstream Fragment:</u> 5’ctgaggatccaatggtaggtgatacacaaaagc 5’ tatatcccgttatgtatatcctgagctccaat	HA epitope-tagging construct (Blasticidin selection). Integrates using a double-crossover.
pMOTagMYC-TbRCC1	TbRCC1	As above	MYC epitope-tagging construct (Blasticidin selection). Integrates using a double-crossover.
pC-PTP-hygro eIF2α	eIF2α	5’ attgggccccgtgccatcggtaaggtcatc 5’ttccagcggccgcccgtccgtgtcttcgccttcatc	PTP-epitope tagging construct (Hygromycin selection). Integrates using a single-crossover.
pC-PTP-hygroNLP	NLP	<u>Upstream Fragment:</u> 5’ caagggccctcaccaccagtctcaaccagata 5’ tcccacttttcctcgagtttgatgcg <u>Downstream Fragment:</u> 5’caagggcccgcacgcatcaaactcgaggaaaagtgg 5’ttccagcggccgcaactccagccttagcttgttctg	PTP-epitope tagging construct (Hygromycin selection). Integrates using a double-crossover.
pC-PTP-hygroTDP1	TDP1	5’gatgggcccgaggagctgaaggcgaaaaaccccggg 5’ttacagcggccgcttgggtagtgctgccatttccct	PTP-epitope tagging construct (Hygromycin selection). Integrates using a single-crossover.

Sequences

The *T. brucei* genes (accession numbers indicated in brackets): NLP (Tb927.10.5450), TDP1 (Tb927.3.3490), TbISWI (Tb927.2.1810), TbRCC1 (Tb11.01.2100), FYRC1 (Tb927.7.1060), Brom1 (Tb927.10.8150), Brom2 (Tb927.10.7420), Brom3 (Tb11.01.1830), Saga1 (Tb927.8.4210), Saga2 (Tb927.10.10170), HM1 (Tb927.10.8950), HM2 (Tb927.10.8950), TbeIF2K2 (Tb927.4.2500), and TbeIF2a (Tb927.3.2900) were investigated in this study.

NLP was compared with *Arabidopsis thaliana* HMG-I/Y (accession number Y10836.1), *Saccharomyces cerevisiae* Swi2p/Snf2p (accession number NM_001183709.1), *Drosophila melanogaster* Nucleoplasmin (Nlp) (accession number NP_524557), ISWI (accession number NM_078995.1) and D1 (accession number NM_079562.2), *Dario rerio* Nucleoplasmin2 (accession number NM_001123007.2), *Xenopus laevis* Nucleoplasmin2 (accession number BC157168.1) and Nucleoplasmin3 (accession number NM_001016456.2), *Mus musculus* Nucleoplasmin1 (accession number BC090843.1), Nucleoplasmin3 (accession number NM_008723.1) and HMGI-C (accession number NM_010441.2), and *Homo sapiens* Nucleoplasmin3 (accession number NM_006993.2) and HMG-I/HMG-Y (accession number NM_145899.2) using ClustalW (Larkin et al. 2007).

3. *T. brucei* transfections

Transfections of bloodstream form *T. brucei* were performed essentially as described by Blundell et al. (1996). 5-10 μg of sterile DNA, isolated using phenol:chloroform extraction and ethanol precipitation, was electroporated into approximately 3×10^7 cells using an AMAXA nucleofector according to the manufacturer's instructions. After about 16 hours

of recovery in HMI-9 medium at 37 °C / 5% CO_2, appropriate drug selection was applied to the cells, and the cultures were diluted in 24 well plates and re-incubated. The plates were checked after about 4-6 days, when typically untransfected cells had died and the transformants had grown out.

Procyclic transfections were carried out using a protocol modified from Blundell et al. (1996). 2.5×10^7 cells were spun down and resuspended in 500 µl cytomix (David Horn 2001). 5 µg of the purified plasmid DNA was added to the cells in 0.2 cm gap cuvettes (BIO-RAD) and electroporated with a BTX ECM 830 machine using three 100 µs pulses at 1.5 kV, capacitance 25 µF with 200 µs intervals between pulses. The electroporated cells were allowed to recover for 16-20 hours before the appropriate drug selection was applied. Cells were subsequently diluted serially and plated out over 96 well plates. These were monitored over the next 7-10 days to check for death of parental cells and the emergence of transformants.

4. Protein analysis

Protein lysates were made by centrifuging cells, washing in 1X PSG, and dissolving the pellet in boiling hot 1X Laemmli buffer at an end concentration of 10^5 cells/µl. This was incubated at 100 °C for 10 minutes prior to loading on 7%, 10%, or 12% SDS-polyacrylamide gels. Gels were blotted on to Hybond-P (Amersham), and probed with rabbit polyclonal or mouse monoclonal antibodies raised against NLP (prepared by A. Fullbrook. As described in Narayanan et al. (2010), BiP (kind gift of Jay Bangs, University of Wisconsin) (Bangs et al. 1993) HA tag (ab9110 and ab1424, AbCam), EP1 (Cedarlane), TDP1 (kind gift of Klaus Ersfeld, University of Hull), MYC tag (clone 4A6, 05-724, Upstate), TbK2 and eIF2α (kind gift of Beatriz Castilho, Universidade Federal de São Paulo),

eIF4E1-4 and eIF4A (kind gifts of Mark Carrington, University of Cambridge), VSG 221 (kind gift of Jay Bangs, University of Wisconsin), or protein-A (ProtA) (Sigma). The appropriate anti-rabbit IgG or anti-mouse IgG secondary antibodies conjugated to horseradish peroxidase were used to detect bound antibodies and the blots were visualized with Western Lightning Plus ECL (PerkinElmer) or ECL Plus (Amersham).

2-dimensional gel electrophoresis was performed essentially as described in Portman et al. (2009). Proteins were separated in the first dimension on pH 3–11 NL IPG strips (GE Healthcare) and in the second dimension by SDS-PAGE-10% acrylamide: bisacrylamide 37.5:1. Gels were blotted on to Hybond-P (Amersham) and probed with a rabbit polyclonal antibody against TbeIF2a.

5. Dephosphorylation of proteins

Bloodstream form *T. brucei* cells were centrifuged at different time points and resuspended at an end concentration of 2.5 x 10^5 cells ml^{-1} in RIPA(-EDTA) buffer (50 mM Tris-Cl pH 7.4, 1% NP-40, 150 mM NaCl). 10X Lambda Phosphatase Buffer and the supplied 10X $MnCl_2$ (New England Biolabs) were added. 1 unit of Lambda Phosphatase was added per 2 x 10^4 cells and incubated at 30 °C for 10 minutes. The reaction was then stopped by addition of boiling Laemmli buffer (to a final concentration of 1X) and was placed at 90 °C for 5 minutes. Lysates containing 1.25 x 10^7 cells were loaded per well and analysed in a 7% SDS-PAGE gel for immunoblot analysis.

6. Immunofluorescence microscopy

Sub-nuclear localization of epitope-tagged proteins was performed by centrifuging and washing cells in 1X PSG, and fixing with an end concentration of 1% paraformaldehyde. Fixed cells were allowed to settle on positively charged glass slides (SuperFrost Plus Gold). Cells were washed, permeabilized with 0.1% NP-40, and reacted with a 1:100 dilution of a anti-HA rabbit polyclonal antibody (ab9110, AbCam), or a 1:1000 dilution of an anti-ProtA rabbit polyclonal antibody (Sigma), or a 1:100 dilution of a monoclonal mouse antibody against the uncharacterized nucleolar marker L1C6 (Devaux et al. 2007) (kind gift of Keith Gull). Cells were washed twice in 1X PSG and treated with the relevant anti-rabbit or anti-mouse secondary antibody coupled to Alexa-488 or Alexa-594 (Molecular Probes) and mounted in Vectashield containing 40 ng/ml DAPI (Vector Laboratories). Microscopy was performed using a Zeiss Axioplan 2 microscope, and images were taken using a Roper Scientific (Trenton, NJ) Cool Snap HQ camera and analysed with METAMORPH 6.20 imaging software (Universal Imaging, Downington, PA). Alternatively, microscopy was performed on a Zeiss Imager.M1 microscope and images taken with an AxioCam MRm camera and analysed with AxioVision40 V4.8.10 software.

7. Nucleic acid analysis

Total *T. brucei* RNA was isolated using RNeasy RNA isolation kits (Qiagen). RNA was separated in formaldehyde agarose gels (Sambrook & Russel 2001), and transferred to Hybond-XL membranes (GE Healthcare). Northern blots were hybridised with radio-labelled probes using Amersham Ready-To-Go DNA Labelling Beads.

8. Cell cytometry

Trypanosome cultures were centrifuged at 1200 × g for 8 minutes at different time points, washed in 1X PSG, and resuspended in 1X PSG to an end concentration of ~2 × 10^6 cells ml^{-1} prior to flow cytometry analysis. Flow cytometry was carried out using Becton Dickinson FACSCalibur and CellQuest (BD) software. The mean of all events was calculated using Histogram Stats in CellQuest software. The amount of derepression was expressed as fold derepression, and calculated by dividing mean FL1-H values of RNAi induced cultures at each time point by the mean FL1-H values of uninduced cultures.

9. Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as described in Stanne & Rudenko (2010) with minor modifications. Briefly, bloodstream form cells were fixed for 1 hour in 1% formaldehyde at room temperature. Glycine was added to a final concentration of 125 mM to stop the reaction. Chromatin was sheared by sonication (BioRuptor, Diagenode) to generate DNA of size 200-500 bp, and the extract was clarified by spinning down at 15,000 x g at 4 °C. The solution was pre-cleared by incubating with protein A-Sepharose CL-4B beads (GE Healthcare) for 90 minutes at 4 °C before incubating with the relevant antibody (or 'no antibody' as control) for 16 hours. Antibodies used included a mouse monoclonal antibody against the HA epitope (ab1424, AbCam) and anti-histone H3 (ab1791, AbCam).

Each experiment was performed in triplicate, and further analysis of ChIP material was performed by quantitative PCR. ChIP material was analysed either by quantitative PCR or using slot blots (see below). ChIP material was analysed by qPCR using primers optimised for producing a single amplicon of the correct size as described in Stanne & Rudenko (2010)

or in Appendix B. Quantification of histone H3 and HA-epitope tagged ChIP material was performed by subtracting values obtained from material immunoprecipitated with values obtained from the corresponding 'no antibody' control using the same cell line.

10. Slot blot analysis

ChIP material was analysed by slot-blotting on to Hybond-XL membranes (GE Healthcare) and hybridising with radio-labelled probes prior to imaging with a Phosphorimager (BIO-RAD) as described in Stanne & Rudenko (2010).

11. Transcription analysis (qPCR of cDNA)

The bloodstream-form *T. brucei* T3-NLP1, 121-NLP1, T3-TDP1 cell lines were all used for quantitative reverse-transcriptase PCR (qRT-PCR). RNAi against the given protein was induced, and total RNA was isolated at various time points using the RNeasy kit (Qiagen). RNA was DNase-treated using the TURBO DNA-free kit (Ambion) and cDNA was synthesized using the Omniscript RT kit (Qiagen) with random hexamer primers (Promega). qPCR was performed on a LightCycler® 480 Real-Time PCR System using LightCycler® 480 SYBR Green I Master Mix (Roche) or on a 7500 Fast Real-Time PCR System (Applied Biosystems) using Brilliant II SYBR Green (Agilent Technologies) in triplicate. The reaction conditions for each primer pair were individually optimized to obtain a single product of the correct size. The sequences of the primers used are detailed in Appendix B. Control reactions without reverse transcriptase (RT) were performed using DNase-treated RNA for each time point. Transcript levels were normalized to levels for γ-tubulin or actin transcripts as indicated in individual figures and plotted as fold increase with

respect to the 0 hour time point after subtraction of the no-RT controls. Three independent experiments were performed, and standard deviation was calculated.

12. Tandem Affinity Purification

Tandem affinity purification (TAP) was performed essentially as described in Schimanski et al. (2005). In brief, approximately 3.5×10^{10} procyclic form *T. brucei* cells were grown to late mid-log stage and harvested by centrifugation at 2700 × g at room temperature. Collected cells were washed first in ice-cold wash solution (100 mM NaCl, 3 mM $MgCl_2$, 20 mM Tris-Cl (pH 7.5)) and then in ice-cold transcription buffer (150 mM sucrose, 20 mM L-glutamic acid (mono potassium salt), 20 mM HEPES-KOH (pH 7.7), 3 mM $MgCl_2$) and the packed cell volume (PCV) was determined. The cells were resuspended in 1.5 volumes PCV of transcription buffer supplemented with 1 mM DTT, 10 µg/ml leupeptin , 10 µg/ml aprotinin and lysed by douncing on ice until >70% of cells were broken as determined by visual examination under a light microscope. This extract was shock-frozen in liquid nitrogen and then mixed with extraction buffer (150 mM sucrose, 1500 mM KCl, 20 mM HEPES-KOH (pH 7.7), 3 mM $MgCl_2$). This mix was centrifuged at high speed at 2 °C and supernatant was transferred to a chilled micro-tube. In parallel, IgG Sepharose Fast Flow bead suspension (GE Healthcare) was equilibrated in a Poly-Prep chromatography column (BIO-RAD) with freshly prepared PA-150 buffer (150 mM KCl, 20 mM Tris-Cl (pH 7.7), 3 mM $MgCl_2$, 0.5 mM DTT, 0.1% (v/v) Tween 20). One Complete Mini protease inhibitor tablet (Roche) was dissolved in 1 ml PA-150 buffer, and 0.5 ml of this solution was added to 7.5 ml of extract and then supplemented with Tween 20 to a final concentration of 0.1%. This mixture was added to the column containing the IgG Sepharose beads and incubated in an end-over-end rotator for four hours at 4 °C. A sample was removed as ‘Input’, and the

solution was allowed to pass through the column. A sample of the flow-through was removed as ‘FL_IgG’. The beads were washed twice with PA-150 and then twice with freshly prepared TEV protease buffer (150 mM KCl, 20 mM Tris-Cl (pH 7.7), 3 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 0.1% (v/v) Tween 20). 2 ml TEV protease buffer was added to the beads and they were then incubated with 400 units AcTEV protease (Invitrogen) and rotated for 19 hours at 4 °C. Anti-ProtC matrix (Roche) was then equilibrated in a new Poly-Prep column with freshly prepared PC-150 (150 mM KCl, 20 mM Tris-Cl (pH 7.7), 3 mM $MgCl_2$, 0.5 mM DTT, 1 mM $CaCl_2$, 0.1% (v/v) Tween 20) and TEV protease eluate was collected. A sample was removed as ‘TEV elu’. 7.5 µl of $CaCl_2$ and 0.5 ml of the previously prepared Complete Mini protease cocktail were added to 6 ml of the TEV eluate and incubated at 4 °C for 2 hours in the rotor. Flow-through was collected and a sample was removed as ‘FL_ProtC’. The column was washed with PC-150 and finally TAP material was eluted by the addition of freshly prepared elution buffer (5 mM Tris-Cl (pH 7.7), 5 mM EDTA, 10 mM EGTA). The eluate was concentrated using a vacuum concentrator, and then StrataClean resin (Stratagene) according to the manufacturer’s instructions.

Collected samples monitoring progression of the experiment were examined by SDS-PAGE followed by Western blot analysis. Concentrated final eluate was analysed by separation either under denaturing conditions in a 4-15% SDS-PAGE gel or under non-denaturing conditions in a 4-15% polyacrylamide gel. Gels were stained either with Imperial Protein Stain (Thermo Scientific) or with Pierce® Silver Stain for Mass Spectrometry (Thermo Scientific). Visualized bands were excised as specified and sent for analysis by liquid chromatography-tandem mass-spectrometry (Central Proteomics Facility, University of Oxford).

Chapter 8

References

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. Proceedings of the National Academy of Sciences of the United States of America. 98, 1513-8.

Ahmed, S., and Brickner, J. H. (2007). Regulation and epigenetic control of transcription at the nuclear periphery. Trends in Genetics. 23, 396-402.

Aitcheson, N., Talbot, S., Shapiro, J., Hughes, K., Adkin, C., Butt, T., Sheader, K., and Rudenko, Gloria (2005). VSG switching in Trypanosoma brucei: antigenic variation analysed using RNAi in the absence of immune selection. Molecular Microbiology 57, 1608-22.

Aksoy, S., Williams, S., Chang, S., and Richards, F. F. (1990). SLACS retrotransposon from Trypanosoma brucei gambiense is similar to mammalian LINEs. Nucleic Acids Research. 18, 785-92.

Alsford, S., Kawahara, T., Isamah, C., and Horn, David (2007). A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. Molecular Microbiology. 63, 724-36.

Alsford, S., Turner, D. J., Obado, S. O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., and Horn, David (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Research. 21(6):915-24.

An, W., Holde, K. van, and Zlatanova, J. (1998). The non-histone chromatin protein HMG1 protects linker DNA on the side opposite to that protected by linker histones. The Journal of Biological Chemistry. 273, 26289-91.

Aravind, L., and Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Research. 26, 4413-21.

Badenhorst, P., Voas, M., Rebay, I., and Wu, Carl (2002). Biological functions of the ISWI chromatin remodeling complex NURF. Genes and Development. 16, 3186-98.

Bangs, J D, Uyetake, L., Brickman, M. J., Balber, A. E., and Boothroyd, J. C. (1993). Molecular cloning and cellular localization of a BiP homologue in Trypanosoma brucei. Divergent ER retention signals in a lower eukaryote. Journal of Cell Science. 105 (Pt 4), 1101-13.

Bao, Y., and Shen, X. (2007). SnapShot: chromatin remodeling complexes. Cell. 129, 632.

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

Beaglehole, R., Irwin, A., and Prentice, T. (2004). World Health Report. Available online at: http://www.who.int/whr/2004/en/report04_en.pdf

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. Nature. 447, 407-12.

Bernales, S., Papa, F. R., and Walter, P. (2006). Intracellular signaling by the unfolded protein response. Annual Review of Cell and Developmental Biology. 22, 487-508.

Berriman, M. et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science. 309, 416-22.

Biebinger, S., Rettenmaier, S., Flaspohler, J., Hartmann, C., Peña-Diaz, J., Wirtz, L. E., Hotz, H. R., Barry, J D, and Clayton, C (1996). The PARP promoter of Trypanosoma brucei is developmentally regulated in a chromosomal context. Nucleic Acids Research. 24, 1202-11.

Birch, J. L., and Zomerdijk, J. C. B. M. (2008). Structure and function of ribosomal RNA gene chromatin. Biochemical Society Transactions. 36, 619-24.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes and Development. 16, 6-21.

Black, B. E., and Bassett, E. A (2008). The histone variant CENP-A and centromere specification. Current Opinion in Cell Biology. 20, 91-100.

Blundell, P. A., Rudenko, G, and Borst, P (1996). Targeting of exogenous DNA into Trypanosoma brucei requires a high degree of homology between donor and target DNA. Molecular and Biochemical Parasitology. 76, 215-29.

Borst, P, and Leeuwen, F. van (1997). beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. Molecular and Biochemical Parasitology. 90, 1-8.

Borst, Piet, and Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. Annual Review of Microbiology. 62, 235-51.

Brun, R, and Schönenberger, M. (1979). Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Acta Tropica. 36, 289-92.

Bungard, D., Reed, M., and Winter, E. (2004). RSC1 and RSC2 are required for expression of mid-late sporulation-specific genes in Saccharomyces cerevisiae. Eukaryotic Cell. 3, 910-8.

Bustin, M. (2001). Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends in Biochemical Sciences. 26, 152-153.

Böhme, U., and Cross, G. A. M. (2002). Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in Trypanosoma brucei. Journal of Cell Science. 115, 805-16.

Bütikofer, P., Ruepp, S., Boschung, M., and Roditi, I (1997). "GPEET" procyclin is the major surface protein of procyclic culture forms of Trypanosoma brucei brucei strain 427. The Biochemical Journal. 326 (Pt 2, 415-23.

Cairns, B R et al. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell. 87, 1249-60.

Calogero, S., Grassi, F., Aguzzi, a, Voigtländer, T., Ferrier, P., Ferrari, S., and Bianchi, M. E. (1999). The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. Nature Genetics. 22, 276-80.

Cato, L., Stott, K., Watson, M., and Thomas, Jean O (2008). The interaction of HMGB1 and linker histones occurs through their acidic and basic tails. Journal of Molecular Biology 384, 1262-72.

Chaves, I, Rudenko, G, Dirks-Mulder, A, Cross, M, and Borst, P (1999). Control of variant surface glycoprotein gene-expression sites in Trypanosoma brucei. The EMBO Journal. 18, 4846-55.

Clayton, C E (1999). Genetic manipulation of kinetoplastida. Parasitology Today. 15, 372-8.

Clayton, Christine E (2002). Life without transcriptional control? From fly to man and back again. The EMBO Journal. 21, 1881-1888.

Cliffe, L. J., Kieft, Rudo, Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., and Sabatini, R. (2009). JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Research. 37, 1452-62.

Cliffe, L. J., Siegel, T Nicolai, Marshall, M., Cross, G. A. M., and Sabatini, R. (2010). Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. Nucleic Acids Research. 38, 3923-35.

Conconi, A., Widmer, R. M., Koller, T., and Sogo, J. M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell. 57, 753-61.

Corona, D. F. V., and Tamkun, J. W. (2004). Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochimica et Biophysica Acta. 1677, 113-9.

Cross, Mike, Kieft, Rudo, Sabatini, R., Dirks-Mulder, Anita, Chaves, Inês, and Borst, Piet (2002). J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. Molecular Microbiology. 46, 37-47.

Cui, L., and Miao, J. (2010). Chromatin-mediated Epigenetic Regulation in the Malaria Parasite Plasmodium falciparum. Eukaryotic Cell. 9, 1138-1149.

Dang, W., and Bartholomew, B. (2007). Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. Molecular and Cellular Biology. 27, 8306-17.

Daniel, J. A, and Grant, P. A (2007). Multi-tasking on chromatin with the SAGA coactivator complexes. Mutation Research. 618, 135-48.

Denninger, V., Fullbrook, A., Bessat, M., Ersfeld, K., and Rudenko, Gloria (2010). The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in Trypanosoma brucei. Molecular Microbiology. 78(2):459-74.

Devany, M., Kotharu, N. P., and Matsuo, H. (2004). Solution NMR structure of the C-terminal domain of the human protein DEK. Protein science: a publication of the Protein Society. 13, 2252-9.

Devaux, S., Kelly, S., Lecordier, L., Wickstead, B., Perez-Morga, D., Pays, Etienne, Vanhamme, Luc, and Gull, Keith (2007). Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. Molecular Biology of the Cell. 18, 1293-301.

DiPaolo, C., Kieft, Rudo, Cross, Mike, and Sabatini, R. (2005). Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Molecular Cell. 17, 441-51.

Doyle, J. J. (1977). Antigenic variation in the salivarian trypanosomes. Adv. Exp. Med. Biol. 93, 31-363.

Earnshaw, William C, Honda, Barry M, Laskey, Ronald A, and Thomas, J O (1980). Assembly of nucleosomes: the reaction involving X. laevis nucleoplasmin. Cell. 21, 373-83.

Eirín-López, José M, Frehlick, L. J., and Ausió, J. (2006). Long-term evolution and functional diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear chaperones. Genetics. 173, 1835-50.

Eitoku, M., Sato, L., Senda, T., and Horikoshi, M. (2008). Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. Cellular and Molecular Life Sciences. 65, 414-44.

Ekwall, K. (2004). The roles of histone modifications and small RNA in centromere function. Chromosome Research : an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology. 12, 535-42.

El Gazzar, M., Yoza, B. K., Chen, X., Garcia, B. a, Young, N. L., and McCall, C. E. (2009). Chromatin-specific remodeling by HMGB1 and linker histone H1 silences proinflammatory genes during endotoxin tolerance. Molecular and Cellular Biology. 29, 1959-71.

Embley, T. M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. Nature. 440, 623-30.

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. Molecular and Biochemical Parasitology. 51, 111-8.

Fenn, K., and Matthews, K. R. (2007). The cell biology of Trypanosoma brucei differentiation. Current Opinion in Microbiology. 10, 539-46.

Figueiredo, Luisa M, and Cross, G. A. M. (2010). Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. Eukaryotic Cell. 9, 148-54.

Figueiredo, Luisa M, Cross, G. A. M., and Janzen, C. J. (2009). Epigenetic regulation in African trypanosomes: a new kid on the block. Nature Reviews, Microbiology. 7, 504-13.

Figueiredo, Luisa M, Janzen, C. J., and Cross, G. A. M. (2008). A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biology. 6, e161.

Frank, M., Dzikowski, R., Costantini, D., Amulic, B., Berdougo, E., and Deitsch, K. (2006). Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum. The Journal of Biological Chemistry. 281, 9942-52.

Frehlick, L. J., Eirín-López, José María, and Ausió, J. (2007). New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. BioEssays. 29, 49-59.

Fujimura, K., Kano, F., and Murata, M. (2008). Identification of PCBP2, a facilitator of IRES-mediated translation, as a novel constituent of stress granules and processing bodies. RNA. 14, 425-31.

Gadad, S. S., Shandilya, J., Kishore, a H., and Kundu, T. K. (2010). NPM3, a member of the nucleophosmin/nucleoplasmin family, enhances activator-dependent transcription. Biochemistry. 49, 1355-7.

Gerlitz, G., Hock, R., Ueda, T., and Bustin, M. (2009). The dynamics of HMG protein – chromatin interactions in living cells. Biochem. Cell Biol. 87, 127-137.

Glover, L., and Horn, David (2006). Repression of polymerase I-mediated gene expression at Trypanosoma brucei telomeres. EMBO Reports. 7, 93-9.

Gluenz, E., Sharma, R., Carrington, M., and Gull, Keith (2008). Functional characterization of cohesin subunit SCC1 in Trypanosoma brucei and dissection of mutant phenotypes in two life cycle stages. Molecular Microbiology. 69, 666-80.

Godfrey, P. a, Malnic, B., and Buck, L. B. (2004). The mouse olfactory receptor gene family. Proceedings of the National Academy of Sciences of the United States of America. 101, 2156-61.

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell. 128, 635-8.

Gommers-Ampt, J. H., Leeuwen, F. V., and Vliegenthart, J. F. G. (1993). A Novel Modified Base Present in the DNA of the Parasitic Protozoan T . brucei. Cell. 75, 1129-1136.

González-Romero, R., Méndez, J., Ausió, J., and Eirín-López, José M (2008). Quickly evolving histones, nucleosome stability and chromatin folding: all about histone H2A.Bbd. Gene. 413, 1-7.

Greaves, D. R., and Borst, P (1987). Trypanosoma brucei variant-specific glycoprotein gene chromatin is sensitive to single-strand-specific endonuclease digestion. Journal of Molecular Biology. 197, 471-83.

Grewal, S. I. (2000). Transcriptional silencing in fission yeast. Journal of Cellular Physiology. 184, 311-8.

Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P. P. (2006). Nucleophosmin and cancer. Nature Reviews, Cancer. 6, 493-505.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T., and Lee, M. G. S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryotic Cell. 2, 542.

Gunzl et al., 2007 in “Trypanosomes, After the Genome” (ed. D. Barry, R. McCulloch, J. Mottram and A. Acosta-Serrano), (Horizon Bioscience, Wymondham, UK) (2007). Parasitology. 134.

Hadjebi, O., Casas-Terradellas, E., Garcia-Gonzalo, F. R., and Rosa, J. L. (2008). The RCC1 superfamily: from genes, to function, to disease. Biochimica et Biophysica Acta. 1783, 1467-79.

Hall, D., Wade, J., and Struhl, K. (2006). An HMG protein, Hmo1, associates with promoters of many ribosomal protein genes and throughout the rRNA gene locus in Saccharomyces cerevisiae. Molecular and Cellular Biology. 26, 3672-3679.

Hammarton, T. C. (2007). Cell cycle regulation in Trypanosoma brucei. Molecular and Biochemical Parasitology. 153, 1-8.

Hammarton, T. C., Clark, J., Douglas, F., Boshart, M., and Mottram, J. C. (2003). Stage-specific differences in cell cycle control in Trypanosoma brucei revealed by RNA interference of a mitotic cyclin. The Journal of Biological Chemistry. 278, 22877-86.

Hecht, A, Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell. 80, 583-92.

Herlands, L., and Maul, G. G. (1994). Characterization of a major nucleoplasmin-like germinal vesicle protein which is rapidly phosphorylated before germinal vesicle breakdown in Spisula solidissima. Developmental Biology. 161, 530-7.

Hertz-Fowler, C. et al. (2008). Telomeric expression sites are highly conserved in Trypanosoma brucei. PloS One. 3, e3527.

Hill, D. A, and Reeves, R. (1997). Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA. Nucleic Acids Research. 25, 3523-31.

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. The Journal of Parasitology. 75, 985-9.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nature reviews. Molecular Cell Biology. 6, 318-27.

Horn, D, and Cross, G.A.M. (1995). A developmentally regulated position effect at a telomeric locus in Trypanosoma brucei. Cell. 83, 555-61.

Horn, D, and Cross, G.A.M. (1997). Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. The EMBO Journal. 16, 7422-31.

Horn, David (2009). Antigenic variation: extending the reach of telomeric silencing. Current Biology. 19, R496–R498.

Horn, David (2001). Cytomix. Available at: http://homepages.lshtm.ac.uk/davidhorn/Horn Lab - Protocols.htm.

Horn, David (2007). Introducing histone modification in trypanosomes. Trends in parasitology 23, 239-42.

Huang, N., Negi, S., Szebeni, A., and Olson, M. O. J. (2005). Protein NPM3 interacts with the multifunctional nucleolar protein B23/nucleophosmin and inhibits ribosome biogenesis. The Journal of Biological Chemistry. 280, 5496-502.

Huang, Y. (2002). Transcriptional silencing in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Nucleic Acids Research. 30, 1465-82.

Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K., and Rudenko, Gloria (2007). A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. The EMBO Journal. 26, 2400-10.

Ingram, A. K., and Horn, David (2002). Histone deacetylases in Trypanosoma brucei: two are essential and another is required for normal cell cycle progression. Molecular Microbiology. 45, 89-97.

Ito, T., Tyler, J. K., Bulger, M., Kobayashi, R., and Kadonaga, J. T. (1996). ATP-facilitated chromatin assembly with a nucleoplasmin-like protein from Drosophila melanogaster. The Journal of Biological Chemistry. 271, 25041-8.

Jackson, D. G., Owen, M. J., and Voorheis, H. P. (1985). A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from Trypanosoma brucei. The Biochemical Journal. 230, 195-202.

Jackson, R. J. (2005). Alternative mechanisms of initiating translation of mammalian mRNAs. Biochemical Society Transactions. 33, 1231-41.

Janzen, C. J., Hake, S. B., Lowell, Joanna E, and Cross, George A M (2006). Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in Trypanosoma brucei. Molecular Cell. 23, 497-507.

Kabani, S., Fenn, K., Ross, A., Ivens, A., Smith, T. K., Ghazal, P., and Matthews, Keith (2009). Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. BMC Genomics. 10, 427.

Kasahara, K., Ohtsuki, K., Ki, S., Aoyama, K., Takahashi, H., Kobayashi, T., Shirahige, K., and Kokubo, T. (2007). Assembly of regulatory factors on rRNA and ribosomal protein genes in Saccharomyces cerevisiae. Molecular and Cellular Biology. 27, 6686-705.

Kasahara, K., Ohyama, Y., and Kokubo, T. (2011). Hmo1 directs pre-initiation complex assembly to an appropriate site on its target gene promoters by masking a nucleosome-free region. Nucleic Acids Research. 1, 1-15.

Kaufman, R. J. (1999). Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. Genes and Development. 13, 1211.

Kawahara, T., Siegel, T Nicolai, Ingram, A. K., Alsford, S., Cross, G. a M., and Horn, David (2008). Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Molecular Microbiology. 69, 1054-68.

Koumenis, C., Naczki, C., Koritzinsky, M., Rastani, S., Diehl, A., Sonenberg, N., Koromilas, A., and Wouters, B. G. (2002). Regulation of protein synthesis by hypoxia via

activation of the endoplasmic reticulum kinase PERK and phosphorylation of the translation initiation factor eIF2alpha. Molecular and Cellular Biology. 22, 7405.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell. 128, 693-705.

Kramer, S., Queiroz, R., Ellis, L., Webb, H., Hoheisel, J. D., Clayton, Christine, and Carrington, M. (2008). Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169. Journal of Cell Science. 121, 3002-14.

Kushwaha, M. (2010). TbISWI and its role in transcriptional control in Trypanosoma brucei. D.Phil thesis submitted to the Department of Biochemistry, University of Oxford, UK.

Kwon, C. S., and Wagner, D. (2007). Unwinding chromatin for development and growth: a few genes at a time. Trends in Genetics. 23, 403-12.

Landeira, D., Bart, J.-M., Van Tyne, D., and Navarro, Miguel (2009). Cohesin regulates VSG monoallelic expression in trypanosomes. The Journal of Cell Biology. 186, 243-54.

Larkin, M. A. et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England). 23, 2947-8.

Laskey, R A, and Earnshaw, W C (1980). Nucleosome assembly. Nature. 286(5775), 763-767.

Laskey, R A, Honda, B M, Mills, A. D., and T, F. J. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature. 275(5679), 416-420.

Laskey, R A, Mills, A. D., Philpott, A., Leno, G. H., Dilworth, S. M., and Dingwall, C. (1993). The role of nucleoplasmin in chromatin assembly and disassembly. Philosophical transactions of the Royal Society of London. Series B, Biological Sciences. 339, 263-9; discussion 268-9.

Laurenson, P., and Rine, J. (1992). Silencers, silencing, and heritable transcriptional states. Microbiological Reviews. 56, 543-60.

LeRoy, G., Loyola, A, Lane, W. S., and Reinberg, D. (2000). Purification and characterization of a human factor that assembles and remodels chromatin. The Journal of Biological Chemistry. 275, 14787-90.

Lewin, B. (2004). Genes VIII (Pearson Education, Inc).

Li, B., Espinal, A., and Cross, G.A.M. (2005). Trypanosome telomeres are protected by a homologue of mammalian TRF2. Molecular and Cellular Biology. 25, 5011.

Liang, X.-hai, Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis Splicing in Trypanosomatids : Mechanism , Factors , and Regulation. Eukaryotic Cell. 2, 830-840.

Liu, Y., Prasad, R., and Wilson, S. H. (2011). HMGB1: roles in base excision repair and related function. Biochimica et Biophysica Acta. 1799, 119-30.

Lu, Z., Zhang, C., and Zhai, Z. (2005). Nucleoplasmin regulates chromatin condensation during apoptosis. Proceedings of the National Academy of Sciences of the United States of America. 102, 2778-83.

Längst, G., and Becker, P. B. (2001). Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. Journal of Cell Science. 114, 2561-8.

Mais, C., Wright, J. E., Prieto, J.-L., Raggett, S. L., and McStay, B. (2005). UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. Genes and Development. 19, 50-64.

Massie, C. E., and Mills, I. G. (2008). ChIPping away at gene regulation. EMBO Reports. 9, 337-43.

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

McStay, B., and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. Annual Review of Cell and Developmental Biology. 24, 131-57.

Mellor, J., Dudek, P., and Clynes, D. (2008). A glimpse into the epigenetic landscape of gene regulation. Current Opinion in Genetics & Development. 18, 116-22.

Mellor, J., and Morillon, A. (2004). ISWI complexes in Saccharomyces cerevisiae. Biochimica et Biophysica Acta. 1677, 100-12.

Merz, K., Hondele, M., Goetze, H., Gmelch, K., Stoeckl, U., and Griesenbeck, J. (2008). Actively transcribed rRNA genes in S. cerevisiae are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes and Development. 22, 1190-204.

Mikkelsen, T. S. et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 448, 553-560.

Mito, Y., Henikoff, J. G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science. 315, 1408-11.

Mohrmann, L., and Verrijzer, C. P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochimica et Biophysica Acta. 1681, 59-73.

Mombaerts, P. (2001). How smell develops. Nature Neuroscience. 4 Suppl, 1192-8.

Montgomery, J., Mphande, F. A, Berriman, M., Pain, A., Rogerson, S. J., Taylor, T. E., Molyneux, M. E., and Craig, A. (2007). Differential var gene expression in the organs of patients dying of falciparum malaria. Molecular Microbiology. 65, 959-67.

Moraes, M. C. S. et al. (2007). Novel membrane-bound eIF2alpha kinase in the flagellar pocket of Trypanosoma brucei. Eukaryotic Cell. 6, 1979-91.

Moss, T, Stefanovsky, V, Langlois, F., and Gagnon-Kugler, T. (2006). A new paradigm for the regulation of the mammalian ribosomal RNA genes. Biochemical Society Transactions. 34, 1079-81.

Muchardt, C., and Yaniv, M. (1999). ATP-dependent chromatin remodelling: SWI/SNF and Co. are on the job. Journal of Molecular Biology. 293, 187-98.

Namboodiri, V. M. H., Dutta, S., Akey, I. V., Head, J. F., and Akey, C. W. (2003). The crystal structure of Drosophila NLP-core provides insight into pentamer formation and histone binding. Structure. 11, 175-86.

Narayanan, M. S., Kushwaha, M., Ersfeld, K., Fullbrook, A., Stanne, T. M., and Rudenko, Gloria (2010). NLP is a novel transcription regulator involved in VSG expression site control in Trypanosoma brucei. Nucleic Acids Research. 39, 2018-2031.

Narlikar, G. J. (2010). A proposal for kinetic proof reading by ISWI family chromatin remodeling motors. Current Opinion in Chemical Biology. 14, 660-5.

Nasmyth, K. A (1982). Molecular genetics of yeast mating type. Annual Review of Genetics. 16, 439-500.

Navarro, M, and Cross, G.A.M. (1998). In situ analysis of a variant surface glycoprotein expression-site promoter region in Trypanosoma brucei. Molecular and Biochemical Parasitology. 94, 53-66.

Navarro, M, Cross, G.A.M., and Wirtz, E. (1999). Trypanosoma brucei variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. The EMBO Journal. 18, 2265-72.

Navarro, M, and Gull, K (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. Nature. 414, 759-63.

Oberholzer, M., Morand, S., Kunz, S., and Seebeck, T. (2006). A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. Molecular and Biochemical Parasitology. 145, 117-20.

Oberle, M., Balmer, O., Brun, Reto, and Roditi, Isabel (2010). Bottlenecks and the maintenance of minor genotypes during the life cycle of Trypanosoma brucei. PLoS Pathogens. 6, e1001023.

Okuwaki, M. (2008). The structure and functions of NPM1/Nucleophsmin/B23, a multifunctional nucleolar acidic protein. Journal of Biochemistry. 143, 441-8.

O'Sullivan, A. C., Sullivan, G. J., and McStay, B. (2002). UBF binding in vivo is not restricted to regulatory sequences within the vertebrate ribosomal DNA repeat. Molecular and Cellular Biology. 22, 657-68.

Pays, E, Lips, S., Nolan, D., Vanhamme, L, and Pérez-Morga, D. (2001). The VSG expression sites of Trypanosoma brucei: multipurpose tools for the adaptation of the parasite to mammalian hosts. Molecular and Biochemical Parasitology. 114, 1-16.

Philpott, A., Leno, G. H., and Laskey, R A (1991). Sperm decondensation in Xenopus egg cytoplasm is mediated by nucleoplasmin. Cell. 65, 569-78.

Pinskaya, M., Nair, A., Clynes, D., Morillon, A., and Mellor, J. (2009). Nucleosome remodeling and transcriptional repression are distinct functions of Isw1 in Saccharomyces cerevisiae. Molecular and Cellular Biology. 29, 2419-30.

Pokholok, D. K. et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell. 122, 517-27.

Portman, N., Lacomble, S., Thomas, B., McKean, P. G., and Gull, Keith (2009). Combining RNA interference mutants and comparative proteomics to identify protein components and dependences in a eukaryotic flagellum. The Journal of Biological Chemistry. 284, 5610-9.

Racki, L. R., and Narlikar, G. J. (2008). ATP-dependent chromatin remodeling enzymes: two heads are not better, just different. Current Opinion in Genetics & Development. 18, 137-44.

Ramos, I., Prado, A., Finn, R. M., Muga, A., and Ausió, J. (2005). Nucleoplasmin-mediated unfolding of chromatin involves the displacement of linker-associated chromatin proteins. Biochemistry. 44, 8274-81.

Reeves, R., and Nissen, M. S. (1990). The AT-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. Journal of Biological Chemistry. 265, 8573-82.

Rice, P., Garduño, R., Itoh, T., Katagiri, C., and Ausio, J. (1995). Nucleoplasmin-mediated decondensation of Mytilus sperm chromatin. Identification and partial characterization of a nucleoplasmin-like protein with sperm-nuclei decondensing activity in Mytilus californianus. Biochemistry. 34, 7563-8.

Richards, E. J., and Elgin, S. C. R. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell. 108, 489-500.

Roditi, Isabel, and Pearson, T. W. (1990). The procyclin coat of African trypanosomes (or the not-so-naked trypanosome). Parasitology Today. 6, 79-82.

Rodríguez-Navarro, S. (2009). Insights into SAGA function during gene expression. EMBO Reports. 10, 843-50.

Rudenko, G, Blundell, P. A, Dirks-Mulder, A, Kieft, R, and Borst, P (1995). A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in T. brucei. Cell. 83, 547-53.

Rudenko, G, Blundell, P. A, Taylor, M. C., Kieft, R, and Borst, P (1994). VSG gene expression site control in insect form Trypanosoma brucei. The EMBO Journal. 13, 5470-82.

Rudenko, Gloria (2010). Epigenetics and transcriptional control in African trypanosomes. Essays in Biochemistry. 48, 201-19.

Russell, J., and Zomerdijk, J. C. B. M. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. Trends in Biochemical Sciences. 30, 87-96.

Sambrook, J., and Russel, D. (2001). Molecular Cloning: A Laboratory Manual (3rd ed.) (Cold Spring Harbor Laboratory Press).

Sanij, E. et al. (2008). UBF levels determine the number of active ribosomal RNA genes in mammals. The Journal of Cell Biology. 183, 1259-74.

Santoro, R., and Grummt, I. (2005). Epigenetic Mechanism of rRNA Gene Silencing : Temporal Order of NoRC-Mediated Histone Modification , Chromatin remodelling, and DNA methylation. Molecular and Cellular Biology. 25, 2539-2546.

Scherf, A., Lopez-Rubio, J. J., and Riviere, L. (2008). Antigenic variation in Plasmodium falciparum. Annual Review of Microbiology. 62, 445-70.

Schimanski, B., Nguyen, T. N., and Gunzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Molecular and Cellular Biology. 25, 7303.

Schlesinger, S., Selig, S., Bergman, Y., and Cedar, H. (2009). Allelic inactivation of rDNA loci. Genes and Development. 23, 2437-47.

Sengupta, S. M., VanKanegan, M., Persinger, J., Logie, C., Cairns, Bradley R, Peterson, C. L., and Bartholomew, B. (2001). The interactions of yeast SWI/SNF and RSC with the nucleosome before and after chromatin remodeling. The Journal of Biological Chemistry. 276, 12636-44.

Serizawa, S. et al. (2000). Mutually exclusive expression of odorant receptor transgenes. Nature Neuroscience. 3, 687-93.

Serizawa, S., Miyamichi, K., Nakatani, Hiroko, Suzuki, Misao, Saito, M., Yoshihara, Y. and Sakano, Hitoshi (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science. 302, 2088-94.

Serizawa, S., Miyamichi, K., and Sakano, Hitoshi (2004). One neuron-one receptor rule in the mouse olfactory system. Trends in Genetics. 20, 648-53.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, Keith, and Rudenko, Gloria (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. Proceedings of the National Academy of Sciences of the United States of America. 102, 8716-21.

Sheader, K., Vruchte, D. te, and Rudenko, Gloria (2004). Bloodstream form-specific up-regulation of silent vsg expression sites and procyclin in Trypanosoma brucei after inhibition of DNA synthesis or DNA damage. The Journal of Biological Chemistry. 279, 13363-74.

Siegel, T.N., Hekstra, D. R., Kemp, L. E., Figueiredo, L.M., Lowell, J.E., Fenyo, D., Wang, X., Dewell, S., and Cross, G.A.M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes and Development. 23, 1063.

Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999). The WD repeat: a common architecture for diverse functions. Trends in Biochemical Sciences. 24, 181-5.

Smith, T. K. et al. (2009). Blocking variant surface glycoprotein synthesis in Trypanosoma brucei triggers a general arrest in translation initiation. PloS One. 4, e7532.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., and Peattie, D. A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from Giardia lamblia. Science. 243, 75-7.

Stanne, T. M., Kushwaha, M., Wand, M., Taylor, J. E., and Rudenko, Gloria (2011). TbISWI regulates multiple Pol I transcribed loci and is present at Pol II transcription boundaries in Trypanosoma brucei. Eukaryotic Cell. 10(7):964-76.

Stanne, T. M., and Rudenko, Gloria (2010). Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. Eukaryotic Cell. 9, 136-47.

Stefanovsky, V. Y., Pelletier, G., Bazett-Jones, D. P., Crane-Robinson, C., and Moss, T (2001). DNA looping in the RNA polymerase I enhancesome is the result of non-cooperative in-phase bending by two UBF molecules. Nucleic Acids Research. 29, 3241-7.

Stros, M, Launholt, D., and Grasser, K. D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cellular and Molecular Life Sciences. 64, 2590-606.

Stros, Michal (2010). HMGB proteins: interactions with DNA and chromatin. Biochimica et Biophysica Acta. 1799, 101-13.

Tamada, H., Thuan, N. V., Reed, P., Nelson, D., Katoku-Kikyo, N., Wudel, J., Wakayama, T., and Kikyo, N. (2006). Chromatin decondensation and nuclear reprogramming by nucleoplasmin. Molecular and Cellular Biology. 26, 1259.

Taneva, S. G., Bañuelos, S., Falces, J., Arregi, I., Muga, A., Konarev, P. V., Svergun, D. I., Velázquez-Campoy, A., and Urbaneja, M. a (2009). A mechanism for histone chaperoning activity of nucleoplasmin: thermodynamic and structural models. Journal of Molecular Biology. 393, 448-63.

Tansey, W. P., and Herr, W. (1997). TAFs: guilt by association? Cell. 88, 729-32.

Taylor, J.E. & Rudenko, Gloria (2006). Switching trypanosome coats: what's in the wardrobe? Trends in Genetics. 22(11), pp.614-20.

Trypanosomes, After the Genome (ed. D. Barry, R. McCulloch, J. Mottram and A. Acosta-Serrano), (Horizon Bioscience, Wymondham, UK) (2007). Parasitology 134.

Tsukiyama, T, Daniel, C., Tamkun, J., and Wu, C (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell. 83, 1021-6.

Tsukiyama, T, and Wu, C (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell. 83, 1011-20.

Urményi, T. P., and Van der Ploeg, L. H. (1995). PARP promoter-mediated activation of a VSG expression site promoter in insect form Trypanosoma brucei. Nucleic Acids Research. 23, 1010-8.

Urwyler, S., Studer, E., Renggli, C. K., and Roditi, Isabel (2007). A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. Molecular Microbiology. 63, 218-28.

Vancurova, I., Vancura, A, Lou, W., and Paine, P. L. (1997). A domain distinct from nucleoplasmin's nuclear localization sequence influences its transport. Biochemical and Biophysical Research Communications. 235, 19-25.

Vanhamme, L, Poelvoorde, P., Pays, A, Tebabi, P., Van Xong, H., and Pays, E (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. Molecular Microbiology. 36, 328-40.

Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P. B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature. 388, 598-602.

Vassella, Erik, Bu, P., Engstler, M., Jelk, J., and Roditi, Isabel (2003). Procyclin Null Mutants of Trypanosoma brucei Express Free Glycosylphosphatidylinositols on Their Surface. Molecular Biology of the Cell. 14, 1308 -1318.

Verreault, A. (2000). De novo nucleosome assembly: new pieces in an old puzzle. Genes and Development. 14, 1430-1438.

Villanova, G. V., Nardelli, S. C., Cribb, P., Magdaleno, A., Silber, A. M., Motta, M. C. M., Schenkman, S., and Serra, E. (2009). Trypanosoma cruzi bromodomain factor 2 (BDF2) binds to acetylated histones and is accumulated after UV irradiation. International Journal for Parasitology. 39, 665-73.

W. H. O (2010). African trypanosomiasis (sleeping sickness) Fact sheet number 259. Available at: http://www.who.int/mediacentre/factsheets/fs259/en/index.html.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R., and Zhang, Y. (2004). Role of histone H2A ubiquitination in polycomb silencing. Nature. 431, 873-8.

Wang, Q.-P., Kawahara, T., and Horn, David (2010). Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Molecular Microbiology. 77, 1237-1245.

Werner, M. H., Huth, J. R., Gronenborn, A M., and Clore, G. M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell. 81, 705-14.

White, T. C., Rudenko, Gloria, and Borst, Piet (1986). Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Research. 14, 9471-89.

Wickstead, B., Ersfeld, K., and Gull, Keith (2002). Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of Trypanosoma brucei. Molecular and Biochemical Parasitology. 125, 211-6.

Wickstead, B., Ersfeld, K., and Gull, Keith (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome Research. 14, 1014.

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. Molecular and Biochemical Parasitology. 99, 89-101.

Wright, J. R., Siegel, T Nicolai, and Cross, G. A. M. (2010). Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in Trypanosoma brucei. Molecular and Biochemical Parasitology. 172, 141-4.

Xiao, L., Williams, A. M., and Grove, A. (2010). The C-terminal domain of yeast high mobility group protein HMO1 mediates lateral protein accretion and in-phase DNA bending. Biochemistry. 49, 4051-9.

Yadon, A. N., and Tsukiyama, Toshio (2011). SnapShot: Chromatin Remodeling: ISWI. Cell. 144, 453-453.e1.

Yang, X., Figueiredo, Luisa M, Espinal, A., Okubo, E., and Li, B. (2009). RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. Cell. 137, 99-109.

Zeng, L., and Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Letters. 513, 124-8.

Zlatanova, J., and Holde, K. van (1998). Linker histones versus HMG1/2: a struggle for dominance? BioEssays. 20, 584-8.

Appendix A

Investigation of the translation arrest triggered by the induction of a VSG synthesis block

A. Introduction

Blocking VSG synthesis through induction of RNAi against *VSG* in *T. brucei* leads to a global translation arrest (Sheader et al. 2005). I investigated the mechanism of this arrest by attempting to determine if there was an eIF2α-phosphorylation mediated stress-activated regulatory pathway. Additionally, I attempted to determine if post-translational modifications were induced in other translation initiation factors after RNAi induction.

Translation control in T. brucei

VSG synthesis is sensed in *T. brucei*. When VSG synthesis is blocked using tetracycline-inducible RNAi (Sheader et al. 2005) a rapid and specific precytokinesis stage cell-cycle arrest is induced. These stalled cells have the same volume as wild-type precytokinesis cells. Blocking VSG synthesis triggers a novel translation arrest (Smith et al. 2009), which results in a reduction in total protein synthesis to less than 1% normal. These data suggest that *T. brucei* is able to detect the absence of VSG synthesis, and responds by stalling cell division and arresting protein synthesis. I investigated the mechanism behind the *VSG* RNAi induced translation arrest. Data from Smith et al. 2009 indicated that this translation arrest occurred at the level of translation initiation rather than elongation.

The accumulation of unfolded proteins in the endoplasmic reticulum is known to cause ER stress in many eukaryotes (Kaufman 1999). ER stress in turn triggers a translation arrest. This mechanism operates by the action of PERK kinase, an ER trans-membrane protein which is activated during ER stress (Bernales et al. 2006). Subsequently PERK inactivates eIF2α, a translation initiation factor, by phosphorylating it (Fig AA.1). eIF2α-phosphorylation mediated translation arrest is a highly conserved stress-activated pathway (Clayton 1999). Moraes et al. (2007) have reported a novel trypanosome membrane-bound

eIF2α kinase located in the flagellar pocket (TbeIF2K2) which is topologically similar to PERK kinase. The localization of this trans-membrane eIF2α kinase in the flagellar pocket suggested a possible mechanism for how trypanosomes could signal depletion of the VSG coat to the translation machinery.

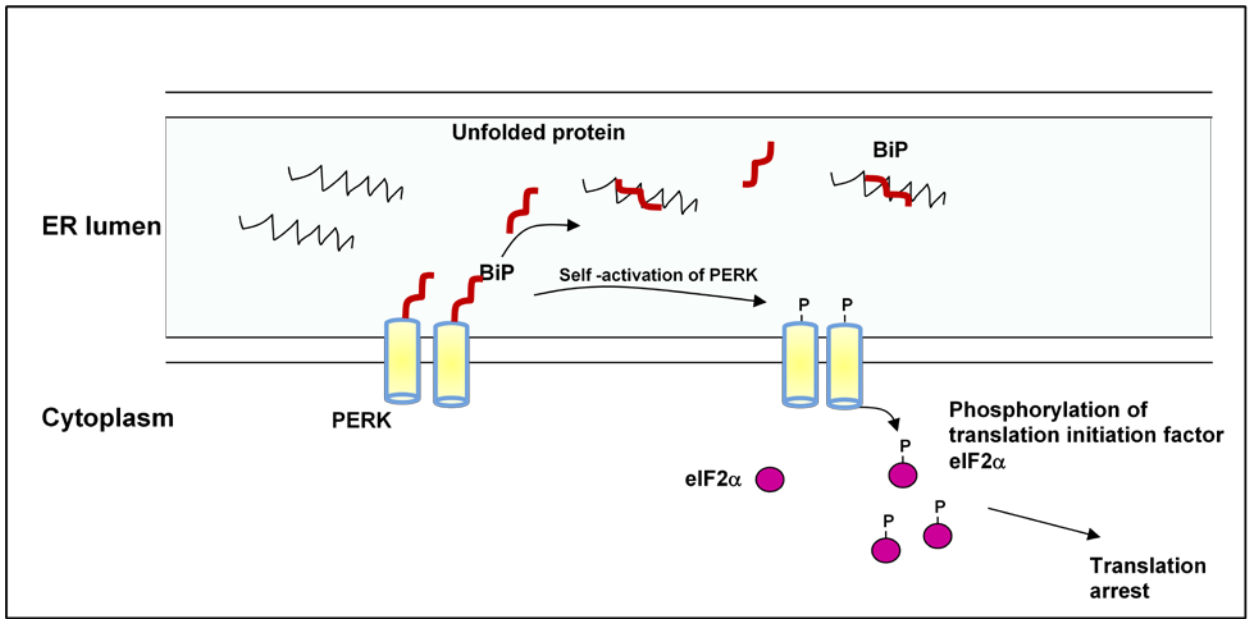


Figure AA.1 Schematic representation of ER stress-associated translation arrest. Accumulation of unfolded protein in the lumen of the ER causes ER stress. Upon accumulation of unfolded protein, BiP, a chaperone protein, which in non-stressful conditions is bound to the ER transmembrane eIF2α kinase – PERK kinase – disassociates itself from PERK and binds to the unfolded protein in the lumen of the ER. PERK kinase subsequently dimerizes and phosphorylates itself, leading to its activation. Activated PERK kinase then phosphorylates the translation initiation factor eIF2α, which triggers a global translation arrest.

Translation initiation

An overview of translation initiation is provided by Jackson (2005) and by Holcik & Sonenberg (2005).We investigated a few different components of the translation initiation process (Fig. AA.2) in order to gain insight in to where the pathway was likely to be disrupted in the translation arrest observed in our cells.

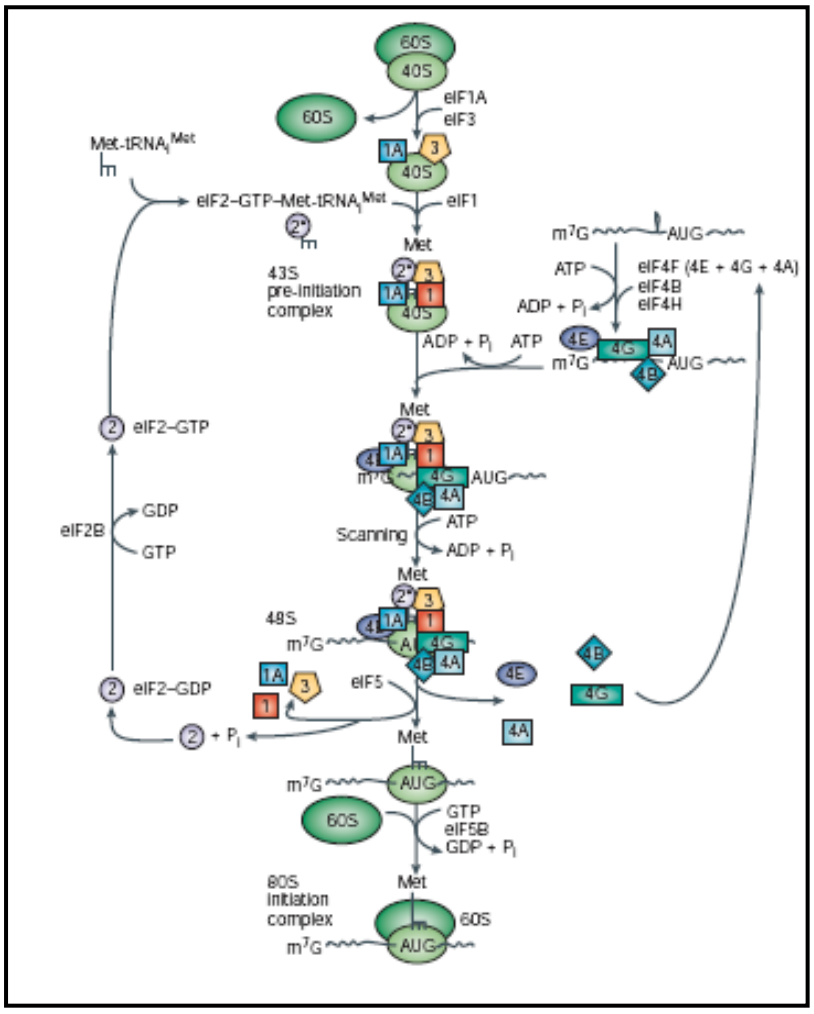


Figure AA.2 Schematic diagram of translation initiation. The initiation process involves the recognition and recruitment of the mRNA by the eIF2-GTP-Met-tRNA$_i$ complex to form the 43S pre-initiation complex, along with the 40S subunit which associates with eIF3 and eIF1A. The cap on the 5' end of the mRNA is recognised by the cap-binding complex which includes among others, translation factors eIF4A and eIF4E. Subsequently the 40S subunit scans the mRNA in the 5'-3' direction until it finds the initiation codon, after which it is joined by the 60S subunit to form the 80S initiation complex. Finally, eIF5 facilitates the release of the initiation factors which are recycled for further rounds of initiation (adapted from Holcik & Sonenberg (2005).

The aims of this project

AA.1. To investigate the role of TbeIF2α in the translation arrest observed after *VSG* RNAi induction. Determine if TbeIF2K2 is responsible for mediating the inactivation of TbeIF2α after induction of *VSG* RNAi through inhibiting its synthesis with post-translational modifications.

AA.2. Investigate if post-translational modifications occur in other translation initiation factors after induction of *VSG* RNAi.

AA.3. Investigate if the translation arrest observed after the induction of *VSG* RNAi shares similarities with translation blocks observed in other stress reactions (i.e. such as after heat shock). This would allow us to focus on components previously identified other responses to better understand our phenotype.

B. Results

Evidence for changes in eIF2α

As eIF2α-phosphorylation mediated translation control is a highly conserved stress-activated regulatory pathway, I searched for evidence for eIF2α modification after the induction of *VSG* RNAi by Western blotting (α-eIF2α and TBeIF2K2 antibodies kindly given by B. Castilho). I also performed 2-dimensional gel electrophoresis followed by Western blotting analysis of eIF2α on lysate from uninduced *T. brucei* cells to determine if this method was sensitive enough to differentiate eIF2α from its phosphorylated form (Fig AA.3). No significant changes in sizes of eIF2α could be detected by Western blot analyses in *VSG* RNAi induced cells. While it may have been possible to separate the two forms of eIF2α on 2D gels, further optimization was not carried out. I also epitope tagged eIF2α with PTP in order to purify it in the event that I found evidence for post-translational modifications, but ultimately did not proceed with this.

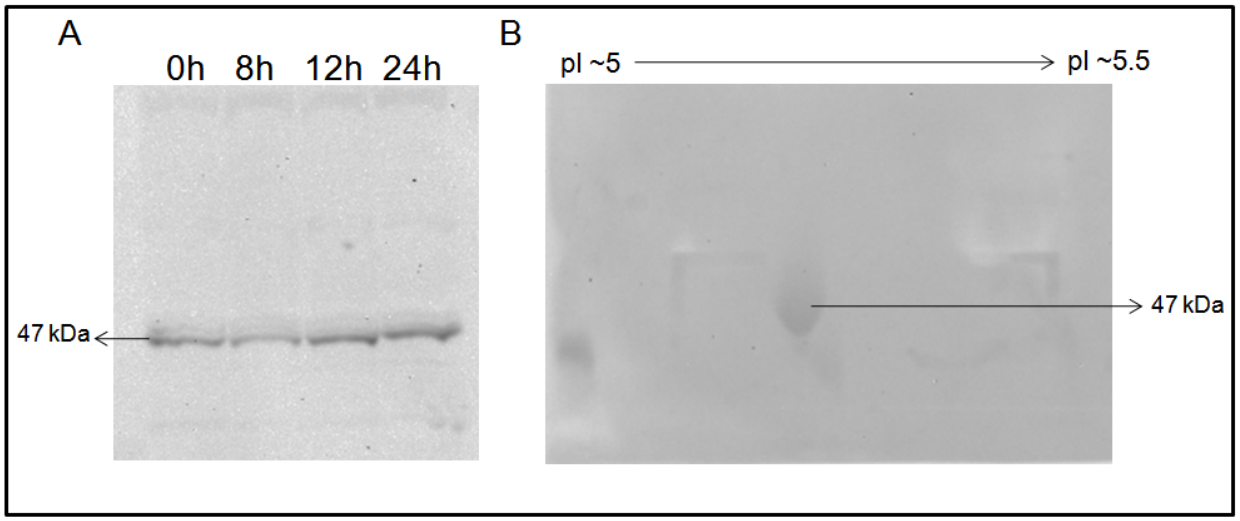


Fig AA.3 Western blot analysis of eIF2α after induction of *VSG* RNAi. A) No clear evidence for increase in eIF2α levels was found after induction of *VSG* RNAi. Protein lysates from 2×10^7 cells were loaded in each well. **B)** 2-Dimensional gel electrophoresis and Western blot analysis of levels of eIF2α 24 hours after *VSG* RNAi induction. There is no clear evidence for the presence of modified forms of eIF2α. This lack of observable change could

be due to the small difference in size and pI between eIF2α and its phosphorylated form. 300µg of protein was loaded on a pI 5-6 IPG strip.

Evidence for post-translational changes in TbeIF2K2

I conducted similar experiments on TBeIF2K2, a transmembrane kinase that phosphorylates eIF2α and which has similarities with PERK, an ER transmembrane eIF2Kα kinase in many eukaryotes. Since the activated form of PERK is hyperphosphorylated (Koumenis et al. 2002), I investigated if this modification occurred with TBeIF2K2 as well. I determined if there was an increase in the amount of the activated form of TBeIF2K2 in cells where *VSG* RNAi had been induced. Western blot analysis indicated the presence of two forms of the TbeIF2K2 protein, and phosphatase experiments confirmed the same observation (Figs AA.4 and AA.5). However, there were no obvious signs of change in either form of the protein after the induction of *VSG* RNAi.

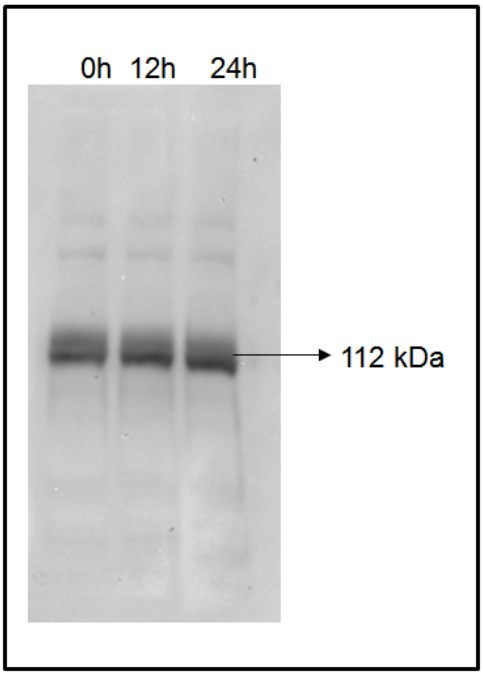


Fig AA.4 Western blot analysis detecting TbeIF2K2 after induction of *VSG* RNAi. No clear evidence for post-translational modification of TbeIF2K2 was seen after *VSG* RNAi. The fuzzy appearance of the bands suggested the possibility that the upper portion of the

bands was comprised of the phosphorylated form of the protein. Protein lysates from extracted at 0, 12, and 24 hours after the induction of RNAi against *VSG* were analysed. Protein lysate from 2×10^7 cells were loaded in each well.

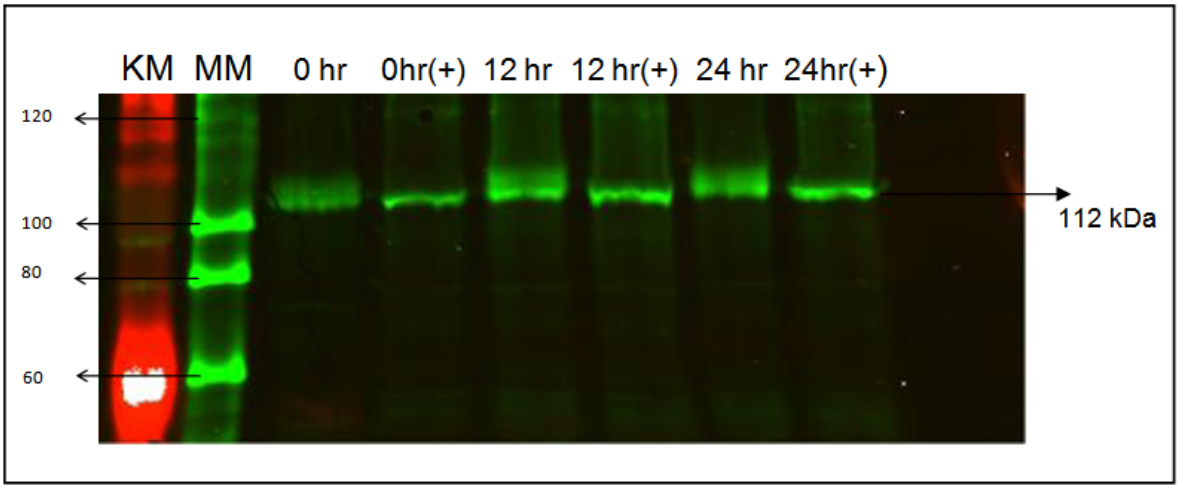


Fig AA.5 No changes in phosphorylation of TbeIF2K2 were observed after the induction of *VSG* RNAi. Protein lysates from uninduced *T. brucei* or cells after 12 hours following the induction of *VSG* RNAi were incubated with lambda phosphatase (lanes with a (+) denote lysates that had been incubated with the enzyme), and electrophoresed on an SDS-PAGE gel. The gels were blotted and analysed using anti-TbeIF2K2 antibodies. Infrared imaging was used in order to quantify levels of the phosphorylated form after the induction of *VSG* RNAi. No significant changes were seen. Protein lysates from 1.25 x 10^7 cells were loaded in each well. KM and MM indicate the molecular markers loaded on the left, with sizes of various markers (kDa) shown on the left.

Generation of TbeIF2K2 knock-out lines

RNAi was induced against TBeIF2K2 (RNAi construct prepared and transfected into *T. brucei* by N. Wand) and was not found to have any effect on *T. brucei* growth (Fig AA.6). Northern and Western blot analysis confirmed that both transcript and protein levels of TbeIF2K2 were reduced after induction of RNAi (Figs AA.7 and AA.8). Unexpectedly, multiple attempts to knock-out the TBeIF2K2 genes failed, despite the evidence indicating that it is not an essential protein. Possibly, very low residual levels of expression of this protein remaining after the induction of RNAi, are adequate to rescue the cells.

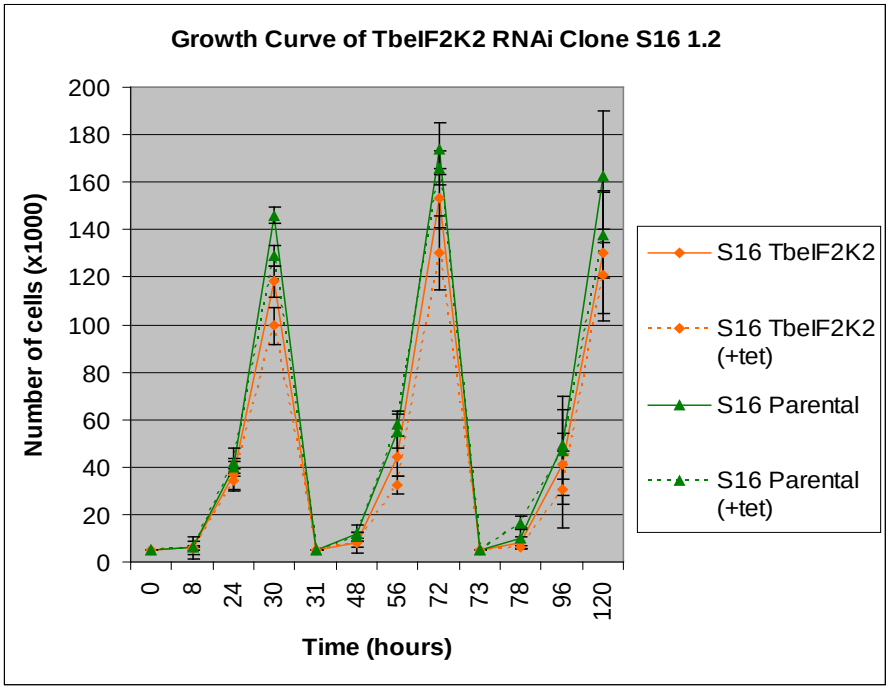


Fig AA.6 Growth kinetics of *T. brucei* after induction of TbeIF2K2 RNAi. No significant growth effect is seen in cells where TbeIF2K2 RNAi is induced compared with uninduced cells or cells from the parental line. This suggests that TbeIF2K2 is a non-essential protein. Cells were successively diluted to 5×10^4 cells/ml after cell densities of over 1×10^6 cells/ml were reached. Plotted values represent the mean of triplicate experiments.

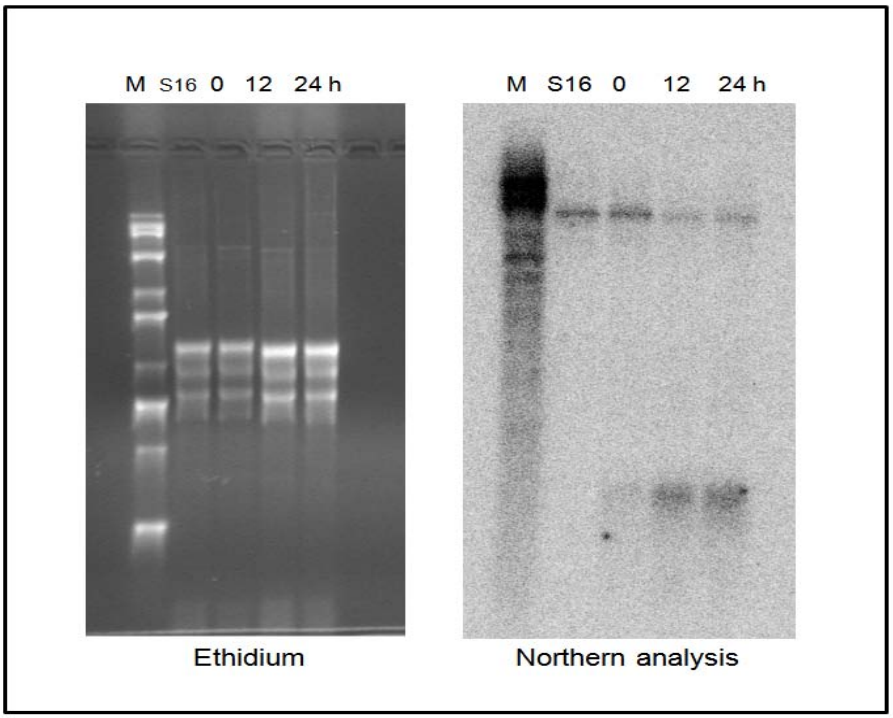


Fig AA.7 Northern blot analysis of TbeIF2K2 transcript after the induction of RNAi shows a rapid reduction in the amount of the transcript within 12 hours. S16 denotes the

parental *T. brucei* cell line (not containing the RNA construct). The other lanes show RNA isolated from cells 0, 12, or 24 hours after the induction of RNAi against TbeIF2K2. The left panel shows an ethidium stained gel. The right panel shows the blot hybridised with a probe for TbeIF2K2.

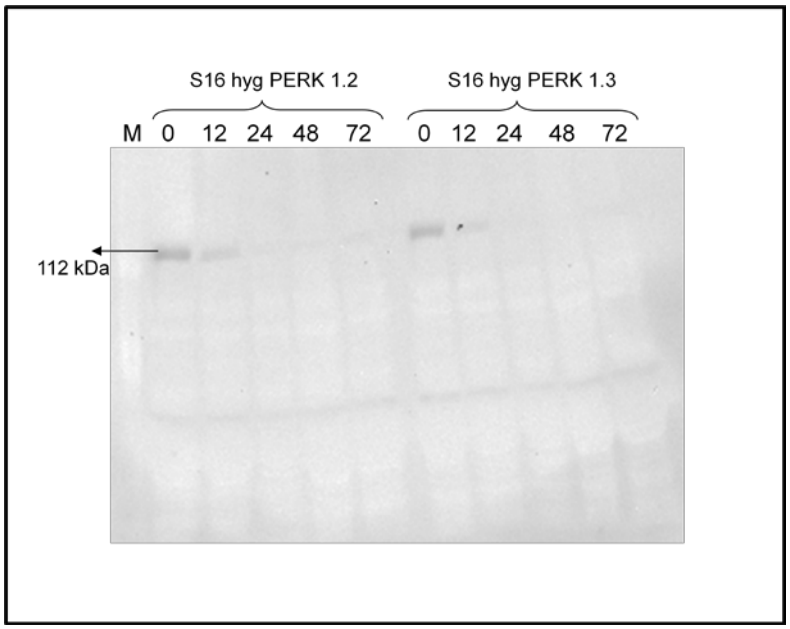


Fig AA.8 Western blot analysis of TbeIF2K2 after the induction of RNAi shows a rapid knock-down of the protein within 12 hours with the protein almost undetectable by 24 hours. S16 hyg PERK 1.2 and 1.3 are two clones on which RNAi was performed. Protein lysates from extracted at 0, 12, 24, 48 and 72 hours after the induction of RNAi against TbeIF2K2 were analysed. Protein lysates from 2×10^7 cells were loaded in each lane. The blot was probed with an antibody against TbeIF2K2 showing the depletion of an ~112 kDa band. M indicates the molecular marker loaded on the left.

Evidence for post-translational modifications in other translation initiation factors

Next, I investigated if there were any changes observed in levels or migration behaviour of several other translation initiation factors (eIF4a, eIF4E1, eIF4E2, eIF4E3 and eIF4E4) in cells where *VSG* RNAi had been induced (antibodies kind gift of M. Carrington). None of these translation factors showed any obvious changes with the exception of eIF4E1. This protein appeared to increase with time after the induction of *VSG* RNAi, however as other bands also increased this observation was not pursued (Figs AA.9 and AA.10).

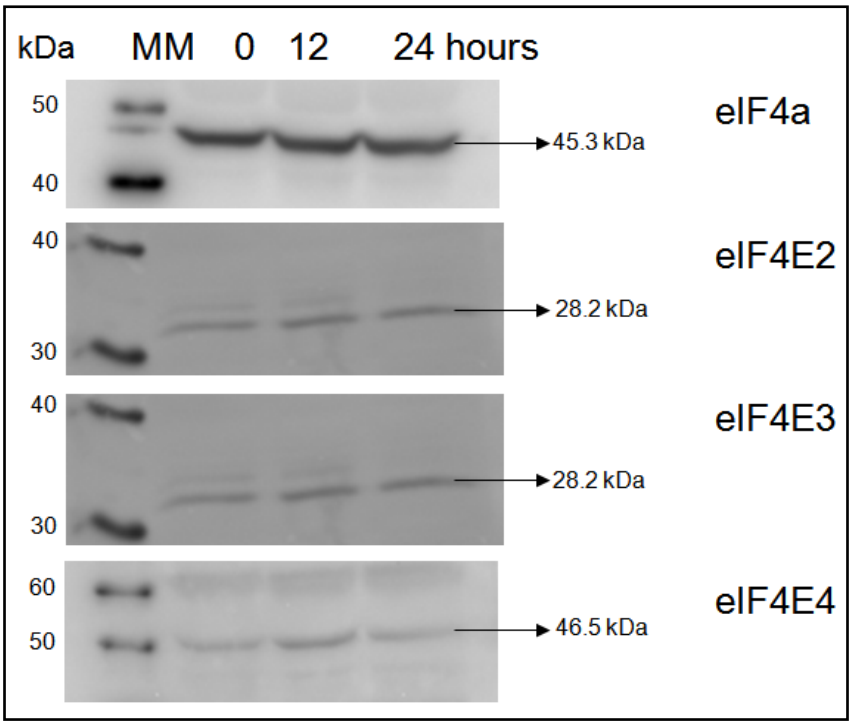


Fig AA.9 Western blot analysis of translation initiation factors eIF4a, eIF4E2, eIF4E3 and eIF4E4 show no significant changes in the migrations of any of these translation initiation factors within 24 hours after the induction of *VSG* RNAi. Protein lysates from extracted at 0, 12, and 24 hours after the induction of RNAi against *VSG* were analysed. Protein lysates from 2×10^7 cells were loaded in each lane. The blots were probed with antibodies against eIF4a, eIF4E2, eIF4E3 or eIF4E4. MM indicates the molecular marker loaded on the left, with sizes of various markers (kDa) shown on the left.

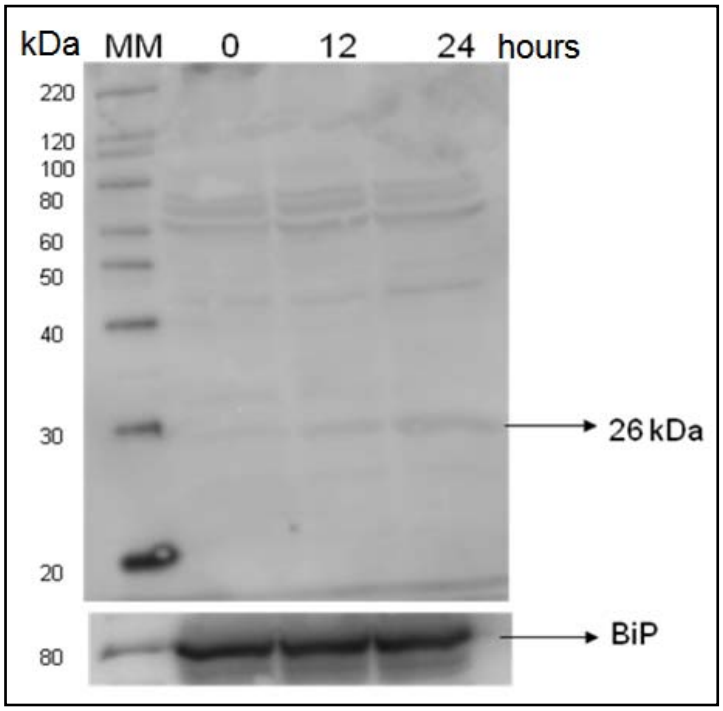


Fig AA.10 Western blot analysis of eIF4E1 after the induction of *VSG* RNAi. A slight increase was seen over time after RNAi induction in a 26 kDa band which corresponds to the

size of the protein according to database matches, while the loading control BiP at 80 kDa showed consistency between lanes. However, other faint bands in the ~40-100 kDa region also show increases over time for reasons that are unclear. Protein lysates from extracted at 0, 12, and 24 hours after the induction of RNAi against *VSG* were analysed. Protein lysates from 2×10^7 cells were loaded in each well. MM indicates the molecular marker loaded on the left, with sizes of various markers (kDa) shown on the left.

Visualization of Stress Granule and P-body markers

Kramer et al. (2008) showed the induction of stress granules (SG) and processing bodies (P-bodies) (Fujimura et al. 2008) in bloodstream form *T. brucei* after heat shock. These results were obtained using YFP-tagged SG and P-body markers – PAB2 and DHH1 respectively. I obtained these constructs and transfected them into our *T. brucei* cells capable of *VSG* RNAi cells in order to monitor if a similar stress pathway was operating in our cells after induction of *VSG* RNAi. However I was unable to visibly induce SGs (Fig AA.11) or P-bodies (data not shown) by blocking VSG synthesis using RNAi in our cells. Unfortunately, these results were inconclusive as I was unable to visualize SGs or P-bodies even after heat shock as demonstrated by Kramer et al. (2008). Possibly this is a consequence of mutations in our midi-prepped and subsequently transfected plasmids, which could have compromised proper YFP expression. I did not sequence the constructs to test this hypothesis.

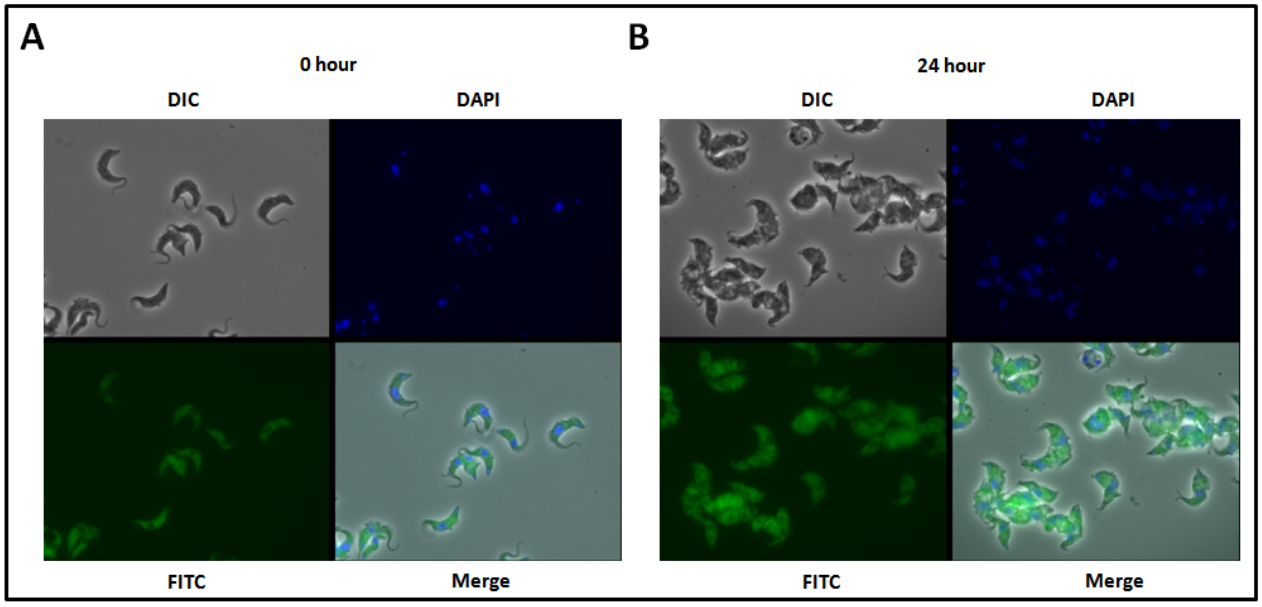


Fig AA.11 Stress granules were not visibly induced following induction of *VSG* RNAi in bloodstream form *T. brucei* cells. A) Control *T. brucei* cells with PAB2 epitope tagged with YFP (YFP-PAB2) in which *VSG* RNAi had not been induced. Clockwise, the panels show *T. brucei* using differential image contrast (DIC), the DNA stain DAPI, a merge of the other three channels, and the FITC channel. **B)** *T. brucei* cells in which RNAi against *VSG* had been induced for 24 hours.

C. Discussion

Our aim was to better understand the molecular mechanisms behind the global translation arrest reported by Smith et al. (2009) after the induction of *VSG* RNAi. Smith et al. (2009) had shown that the translation arrest was probably occurring at the level of translation initiation, and I attempted to understand the mechanism behind this by investigating the role of different translation initiation factors.

I first examined the role of TbeIF2α and one of its activators TbeIF2K2 in this translation arrest, as eIF2α-phosphorylation mediated translation control is a highly conserved stress-activated regulatory pathway (Clayton 1999). TbeIF2K2 was selected due to its similarity with another eIF2α phosphorylator – PERK kinase. Both are transmembrane kinases. TbeIF2K2 is localized on the cell membrane of the flagellar pocket, while PERK is

localized on the membrane of the ER. PERK is known to be involved in stress-induced translation arrest following the unfolded protein response (UPR). In UPR, there is a build-up of unfolded proteins in the lumen of the ER, causing ER stress. PERK activation during UPR results in phosphorylating eIF2α, thereby leading to translation arrest (Bernales et al. 2006; Moraes et al. 2007). I hypothesized that by depleting VSG on the cell surface, I were causing a similar stress which could be detected by TbeIF2K2, then triggers an eIF2α-phosphorylation mediated translation arrest.

However I did not find evidence that eIF2α or TbeIF2K2 was involved in the translation arrest reported after induction of *VSG* RNAi. I subsequently looked for post-translational modifications of other translation initiation factors including eIF4A and the subunits of eIF4E (which are important components of the mRNA cap-binding complex (Holcik & Sonenberg 2005)) after induction of *VSG* RNAi. I did not find evidence suggesting that these were involved.

I subsequently attempted to investigate if our observed translation arrest was related to one found by Kramer et al. (2008), which was triggered by a heat-shock to bloodstream form *T. brucei*. Kramer et al. (2008) also found that this heat-shock resulted in the production of stress-granules and P-bodies (Fujimura et al., 2008). I was however unable to proceed with these experiments, as I were unable to visibly induce SGs or P-bodies by heat shock, as described by Kramer et al. (2008).

Appendix B

List of primers used for qPCR analysis

Primers used in qPCR experiments reported in this thesis and not appearing in Stanne & Rudenko (2010):

Gene/Location	Primer name	Sequence (5'-3')
Gamma Tubulin	gammaTub_491s gammaTub_583as	aatcaccgtatttggacacaag gtttgagtgttcgctgaaagag
Actin	Actin_1031s Actin_1091as	gttccatcctctcatcacta tcgtattcactcttcgttatc
RNA Pol I large subunit	PolI_349s PolI_486as	acttatatgccagacgaagtgg tgtgaatctctgggaagttgtc
NLP	Spot4_216s Spot4_290as	acgaagtcctgaagataaga ataatttccgagaggtagtt
eGFP	eGFP_184s eGFP_324as	gtgaccaccctgacctac cttgtagttgccgtcgtc
Blasticidin	Blasticidin_213s Blasticidin_337a	caacctgacttgtatcgtc catcactgtccttcactatg
Spliced Leader	SL_in_35s SL_in_116as	ggtatgagaagctcccagta ccagattcccgcagtat
Procyclin	Procyclin_8s Procyclin_102as	ctcgttccctttatctactc cttgtcttctggtccttc
VSG T3	VSGT3_675s VSGT3_904as	aaacgacgcaacctctatc ctatgtcttctgggcttga
VSG 121	VSG121_953s VSG121_1022as	tacagaaggaaggcaaatac aatatcttgtcgaggttctt
VSG bR2	VSGbR2_190s VSGbR2_252as	atcctagaacttaacatgacag atccttgtctctgaattgat
VSG 224	VSG224_77s VSG224_171as	cagatgacaacaaggaagac aagtcctgacgtgtctattg
VSG 1.8	VSG1.8_753s VSG1.8_834as	gcagcaaacaataacggact tgacgactgaaagagggtgt
VSG JS1	VSGJS1_70s VSGJS1_234as	actgtaagggtgttgcttct ttgtgctgttaggctcttct
VSG 800	VSG800_99s VSG800_295as	gaaggtctgggaacctctat ggctgtaatatgctcgtaga
VSG 13	VSG13_136s VSG13_305as	gacctttacacaagtaagtacg agttctctccgacaaatatc
TDP1	TDPqPCR_33s TDPqPCR_113as	acagatattgaagagaccgtaa ctccatttggtatttagattcaagc
Convergent area a	C2UP_551s C2UP_687as	atgccctaacgtaccactac tacatgcaggaggagtgtat
Convergent area b	Conv2_UP_428s Conv2_UP_536as	acagaatacagttctttcactc aaatcaatcaaatacacttca
Convergent area c	Conv2_SSR_423s Conv2_SSR_549as	atcagtttgtgtaatgtgtgta tttatttggttgtttgagaa
Convergent	Conv2_SSR_1741s	tatgggttatgtgtgaagat

area d	Conv2_SSR_1848as	gtagaggttggaaggtttat
Convergent area e	Conv2_SSR_3262s Conv2_SSR_3335as	agtgcgttatggtattctgt ggacacctctgtattatgact
Convergent area f	Conv2_DN_78s Conv2_DN_221as	gtgtcttacgtgatggagt ccttgtttctcttgacatt
Convergent area g	C2DN_1779s C2DN_1875as	aacattggaacaacaactgta ctcagcaccttctcgactat
Divergent area a	As above	as above
Divergent area b	DiveUP_124s DiveUP_225as	gtccgttcccatgtctactac agaagaaacgcaacacataag
Divergent area c	DiveSSR_525s DiveSSR_716s	cacctacttactgattgtcat taattttcattccctctatct
Divergent area d	DiveSSR_961s DiveSSR_1028s	cacagaactatatgcttttaca tagaggttaatcagaaatgc
Divergent area e	DiveSSR_1609s DiveSSR_1756as	tgttggactatggaatcgttat tatgctacacacggcaagtta
Divergent area f	DiveDN_922s DiveDN_1038	atgctcacataaatggaagtaa accggatatgaataaggtatc
Divergent area g	D1DN_2313s D1DN_2433as	actttccatccacaacaatac aatacgcttcccacttacag
18S precursor	18S_up_377s 18S_up_265as	ccatgctctctcgtgtgtgta ttcctcaaggcgtcactctatc
28S precursor	28Sa_up_136s 28Sa_up_52as	aaagaggcggcggatagtg acgaaagaagcacaagcacata
70S precursor	70srRNA_up_182s 70srRNA_up_68as	ttgaagggaatgcaaaagtgta aactggaagagacggaggtaaa
140S precursor	140srRNA_up_89s 140srRNA_1as	ttgtgtttctatgtgtgtgtgtaag cgtttggagagggacaaaatat
Alpha-tubulin precursor	aTub_up_331s aTub_up_205as	gaggaggtgggaagggtatatg gaaggcgtgttgatgagttgta
Alpha-tubulin precursor	aTub_up_443as aTub_up_333as	ctgtcattttgttcttttgtgttta taccttcccacctcctctac

Appendix C

Published work

Nucleic Acids Research Advance Access published November 13, 2010

Nucleic Acids Research, 2010, 1–14
doi:10.1093/nar/gkq950

NLP is a novel transcription regulator involved in *VSG* expression site control in *Trypanosoma brucei*

Mani Shankar Narayanan[1,2], Manish Kushwaha[1,2], Klaus Ersfeld[3], Alexander Fullbrook[1], Tara M. Stanne[1] and Gloria Rudenko[1,*]

[1]Division of Cell and Molecular Biology, Sir Alexander Fleming Building, Imperial College London, South Kensington, London SW7 2AZ, [2]Department of Biochemistry, University of Oxford, Oxford OX1 3QU and [3]Department of Biological Sciences and Hull York Medical School, University of Hull, Cottingham Road, Hull HU6 7RX, UK

Received June 16, 2010; Revised and Accepted September 29, 2010

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

ABSTRACT

***Trypanosoma brucei* mono-allelically expresses one of approximately 1500 variant surface glycoprotein (VSG) genes while multiplying in the mammalian bloodstream. The active *VSG* is transcribed by RNA polymerase I in one of approximately 15 telomeric *VSG* expression sites (ESs). *T. brucei* is unusual in controlling gene expression predominantly post-transcriptionally, and how ESs are mono-allelically controlled remains a mystery. Here we identify a novel transcription regulator, which resembles a nucleoplasmin-like protein (NLP) with an AT-hook motif. NLP is key for ES control in bloodstream form *T. brucei*, as NLP knockdown results in 45- to 65-fold derepression of the silent *VSG221* ES. NLP is also involved in repression of transcription in the inactive *VSG* Basic Copy arrays, minichromosomes and procyclin loci. NLP is shown to be enriched on the 177- and 50-bp simple sequence repeats, the non-transcribed regions around rDNA and procyclin, and both active and silent ESs. Blocking NLP synthesis leads to downregulation of the active ES, indicating that NLP plays a role in regulating appropriate levels of transcription of ESs in both their active and silent state. Discovery of the unusual transcription regulator NLP provides new insight into the factors that are critical for ES control.**

INTRODUCTION

An intriguing puzzle in transcriptional control is how organisms control the expression of very large numbers of highly similar genes in a mono-allelic fashion. Examples include regulation of the mammalian olfactory receptor family, in which an individual neuron has more than 1200 olfactory receptor genes, of which only one is activated at a time (1). This mono-allelic expression has been proposed to involve the interaction of the active receptor gene promoter with an activating H-enhancer element, although whether this event regulates the 'counting' observed remains controversial (2,3). Parasitic protozoa like the malaria parasite *Plasmodium falciparum* also mono-allelically express only one out of approximately 50 *VAR* genes encoding the VAR proteins present on the surface of the parasitized erythrocyte (4). Activation of a *VAR* promoter alone is sufficient for silencing endogenous *VAR* transcription (5,6), while a role for the *VAR* introns in mutually exclusive *VAR* expression remains controversial (4,7).

Similarly, the African trypanosome *Trypanosoma brucei* contains more than 1500 variant surface glycoprotein (VSG) genes of which one is mono-allelically expressed (8–10). *T. brucei* is the causative agent of African trypanosomiasis in humans, which is endemic to sub-Saharan Africa and is transmitted by tsetse flies. Central to its success as an extracellular pathogen is a protective VSG coat which is antigenically varied during the course of an infection. Key for antigenic variation to work, is the ability of an individual trypanosome to ensure that only one *VSG* is expressed at a time from one of approximately 15 telomeric ESs (11,12). A stringent control operates on the active ES. Selection for simultaneous full activation of a second ES is not tolerated, and leads to trypanosomes which appear to be alternately switching between the two ESs (13). In insect form *T. brucei* all ESs are to a great extent transcriptionally downregulated, although not to the degree observed in silent ESs in bloodstream form *T. brucei* (14,15). It is still mysterious how the 'counting' machinery behind the mono-allelic transcription of *VSG* ESs in bloodstream form *T. brucei* is mediated.

Initially, it was not apparent that changes in chromatin structure played a significant role in ES regulation in bloodstream *T. brucei*. For example, experiments using exogenous T7 RNA polymerase as a probe showed a likely

*To whom correspondence should be addressed. Tel: +44 207 594 8137; Fax: +44 207 584 2056; Email: gloria.rudenko@imperial.ac.uk

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

role for chromatin in silencing ESs in insect form, but not in bloodstream form *T. brucei* where the mono-allelic exclusion of ESs operates (16). This viewpoint has recently been revisited. Active ESs in bloodstream form *T. brucei* have recently been shown to be highly depleted of nucleosomes compared with silent ESs (17,18). In addition, a number of proteins involved in epigenetic regulation have recently been shown to be instrumental for controlling ESs in bloodstream form *T. brucei*. These include *T. brucei* ISWI (member of the SNF2 superfamily of chromatin remodeling proteins) (19), the telomere binding protein RAP1 (20) and the histone modification enzyme DOT1B (21).

Here, we identify a novel protein with nucleoplasmin-like features (nucleoplasmin-like protein—NLP) as playing a key role in regulation of ESs in bloodstream form *T. brucei*. Knockdown of NLP results in 45- to 65-fold derepression of the silent *VSG221* ES. NLP is also implicated in silencing other transcriptionally inactive areas of the genome in bloodstream form *T. brucei* including the silent *VSG* Basic Copy arrays, minichromosomes and procyclin loci. We also find that after blocking NLP synthesis fully processive transcription of the active ES is compromised. NLP therefore plays a role in regulating appropriate transcription of ESs in both their active and inactive states.

MATERIALS AND METHODS

Trypanosome strains and culturing

Bloodstream form *T. brucei brucei* 427 was used for all experiments. Cells were cultured at 37°C with 5% CO_2 in HMI-9 medium supplemented with 20% fetal bovine serum (22). The *T. brucei* T3-NLP1 and *T. brucei* T3-NLP2 clones are cell lines which have the MC^{177}NLP RNAi construct integrated into minichromosomes in bloodstream form *T. brucei* T3-SM (19). These cells have an active *VSGT3* expression site (ES), and *eGFP* in the silent *VSG221* ES (19). Maintaining these cells on blasticidin selection allows the propagation of a population of trypanosomes homogeneous for expression of VSGT3. The *T. brucei* 121-NLP1 and *T. brucei* 121-NLP2 clones are analogous cell lines, but express VSG121 instead of VSGT3.

The *T. brucei* MCEP reporter cell line has a construct containing *eGFP* (Clontech) behind an ES promoter that is inserted into a minichromosome. The *T. brucei* VBEP reporter cell line contains *eGFP* behind an ES promoter that is inserted into a *VSG* Basic Copy array. The *T. brucei* EP1eGFP cell line has a construct containing *eGFP* inserted behind an endogenous EP1 procyclin promoter. These reporter cell lines were created in a derivative of the *T. brucei* 'single marker' cell line (23) where the pHNES221Pur1.6 construct containing the puromycin resistance gene is inserted immediately behind the active VSG221 ES promoter. The MC^{177}NLP RNAi construct was inserted into these different reporter lines, and derepression of eGFP was monitored after the induction of *NLP* RNAi using tetracycline.

Epitope tagging of NLP with the haemagglutinin (HA) epitope was performed using *T. brucei* 221GPI(VO2+) which has an active VO2 ES containing a neomycin resistance gene, and *eGFP* and a puromycin resistance gene in the silent *VSG221* ES (24). This cell line was transfected with the pMOTagHA–NLP construct to produce *T. brucei* HA–NLP (25). Knocking out the second *NLP* allele was performed by transfecting with the pBSphleoNLPKO construct to produce *T. brucei* HA–NLPKO. As knocking out the second *NLP* allele did not lead to an observable effect on growth, this indicates that the HA epitope tagged copy of NLP is fully functional.

Sequence analysis

Trypanosoma brucei NLP (accession number Tb927.10.5450) protein domains were identified using MyHits (ISB-SIB) and PROSITE. *T. brucei* NLP was compared with *Saccharomyces cerevisiae* Swi2p/Snf2p (accession number NM_001183709.1), *Arabidopsis thaliana* HMG-I/Y (accession number Y10836.1), *Drosophila melanogaster* Nucleoplasmin (Nlp) (accession number NP_524557), ISWI (accession number NM_078995.1) and D1 (accession number NM_079562.2), *Dario rerio* Nucleoplasmin2 (accession number NM_001123007.2), *Xenopus laevis* Nucleoplasmin2 (accession number BC157168.1) and Nucleoplasmin3 (accession number NM_001016456.2), *Mus musculus* Nucleoplasmin1 (accession number BC090843.1), Nucleoplasmin3 (accession number NM_008723.1) and HMGI-C (accession number NM_010441.2), and *Homo sapiens* Nucleoplasmin3 (accession number NM_006993.2) and HMG-I/HMG-Y (accession number NM_145899.2) using ClustalW (26). *T. brucei* TDP1 (accession number Tb927.3.3490) had earlier been identified as binding the ES promoter by (27).

DNA constructs

The MC^{177}NLP RNAi construct has an 835-bp fragment (positions 1522–2356) of the NLP open reading frame, which was amplified using primers 5′- tttgctcgagCGATG CTTGTGTCCG and 5′- gacaggatccTTGAGGCGGTTA CTG and inserted between the opposing T7 promoters of the $p2T7^{Ti}$-177 construct (28). The MCEP reporter construct has the ES promoter from rDES1 (14) upstream of *eGFP* with flanking tubulin RNA processing signals (29) and a blasticidin resistance gene. This construct can be inserted into *T. brucei* minichromosomes using the 177-bp repeat containing sequence from the $p2T7^{Ti}$-177 construct (28). The VBEP reporter construct has the ES promoter upstream of *eGFP* as described above, targeted into the VSG118 Basic Copy array. The *VSG118* Basic Copy array targeting fragment was PCR amplified from *T. brucei* 427 genomic DNA using primers 5′-tatggatccCT TAAGTGAGCCAACACACCAG and 5′-ctggaattcCTG ACGAACCTCTAGTTACCG. The EP1 reporter construct has *eGFP* as described above targeted behind an endogenous EP1 procyclin promoter. The EP1 procyclin targeting fragment was PCR amplified from *T. brucei* genomic DNA using primers: 5′-TCAGTGGTTGTGTG GTAG and 5′-TACTTCTTGGTGTAATTGAAGT.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

The pMOTagHA–NLP construct for HA-epitope tagging NLP was derived from the pMOTag4H vector (25). A 2163-bp fragment (positions 708–2871) of the NLP open reading frame was cloned using primers 5′-gttggtaccTCTCCTGAACACGGGACTTTCAC and 5′-caactcgagCAACTCcAG-CCTTAGCTTGTTC. The pBSphleoNLPKO vector allowed the genetic knockout of an *NLP* allele. A 492-bp fragment immediately upstream of the NLP open reading frame was amplified using primers 5′-tagtctagaGTGATTACATTCCCTACT TTC and 5′-agggatccTTATCCGCCTGGTACTTCCCT. Similarly a 645-bp downstream targeting fragment was amplified using primers: 5′-tagaagcttTGGGTGTTAGG AAACCTTCAA and 5′-tagctcgagTCACCAAGCAAAT CAATACGA. Construct integration into the genome allowed the replacement of NLP sequences with a phleomycin resistance cassette.

The His-tagged NLP fusion protein used for immunization of rabbits to produce a polyclonal anti-NLP antibody was made by amplifying a 597-bp fragment (positions 880–1476) of the NLP open reading frame using primers: 5′-catggatccGTCACCACCAGTCTCAACCAG and 5′-catctcgagGAGATCGCTGTAGTACACAGG and inserting this fragment into the pRSETA vector (Invitrogen) for expression in *Escherichia coli*.

Analysis of nucleic acids and protein

Protein lysates were made by washing cells, and dissolving the pellet in boiling hot SDS gel loading buffer at a concentration of 10^5 cells μl^{-1} (30). Protein lysates were incubated at 100°C for 10 min prior to loading on 10% SDS polyacrylamide gels. Gels were blotted on to Hybond-P membrane (Amersham) and probed with rabbit polyclonal antibodies raised against NLP, BiP (gift of Jay Bangs, University of Wisconsin, USA) or HA epitope tag (ab9110 AbCam). Polyclonal anti-NLP antibody was raised in rabbits immunised against His-tagged NLP fusion protein (positions 880–1476 of the NLP open reading frame). Recombinant fusion protein was expressed in BL21(DE3)pLysS cells [OneShot BL21(DE3)pLysS, Invitrogen] and then purified using denaturing conditions with the ProBond purification system (Invitrogen) according to manufacturer's instructions. Antisera were produced by Eurogentec (Belgium) using standard immunization protocols.

Subnuclear localisation of NLP was determined using immunofluorescence microscopy. Cells were washed, fixed with paraformaldehyde (end concentration 2%) and then allowed to settle on glass slides. Fixed cells were permeabilized with 0.1% NP-40, and then reacted with a rabbit polyclonal anti-HA antibody (ab9110, AbCam) and an anti-rabbit secondary antibody coupled to Alexa-488 (Molecular Probes). Microscopy was performed using a Zeiss Axioplan 2 microscope, and images were taken using a Roper Scientific (Trenton, NJ) Cool Snap HQ camera, and analysed with Metamorph 6.20 imaging software (Universal Imaging, Downington, PA).

Transcript levels were quantitated using quantitiative RT–PCR. RNA was isolated using RNeasy kits (Qiagen). RNA was DNase treated using Turbo DNA Free kit (Ambion) and cDNA synthesised using Omniscript Reverse Transcriptase (Qiagen) and random hexamer primers (Promega). Quantitative PCR (qPCR) was performed as described in ref. (18) on a Roche LightCycler480 Real-Time PCR system using primers optimized for producing a single amplicon of the correct size and listed in Supplementary Table S1 or ref. (18). Transcript levels at each time point were normalized to levels of actin transcript at that time point for experiments using the VSGT3 expressors, and γ-tubulin for experiments using the VSG121 expressors, and then plotted as fold increase with respect to the 0 h time point.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as described in (18) with minor modifications. Bloodstream form cells were fixed for 1 h in 1% formaldehyde at room temperature. Glycine was added to a final concentration of 125 mM to stop the reaction. Chromatin was sheared by sonication (BioRuptor, Diagenode) to produce DNA in the range of 200–500 bp, and the extract was clarified by centrifugation at 15 000*g* at 4°C. The solution was pre-cleared by incubation with protein A-Sepharose CL-4B beads (GE Healthcare) for 90 min at 4°C before incubation with the relevant antibody (or with 'no antibody' as negative control) for 16 h. Antibodies used include a mouse monoclonal antibody against the HA epitope (antibody ab1424, AbCam) and anti-histone H3 (ab1791, AbCam). Each experiment was performed in triplicate and further analysis of ChIP material was performed by quantitative PCR (qPCR). ChIP material was analysed by qPCR using primers optimized for producing a single amplicon of the correct size and are listed in (18) or in Supplementary Table S1. Final quantitation of HA-epitope tagged ChIP material in the *T. brucei* HA–NLP or wild-type cell lines was obtained after subtracting values derived from material immunoprecipitated with the corresponding 'no antibody' control experiments from both of these cell lines. ChIP material was slot blotted on to Hybond-XL membranes (GE Healthcare) and hybridized with radio-labelled probes as described previously (18).

Flow cytometry

RNAi against NLP was induced in *T. brucei* by the addition of 750 ng ml^{-1} tetracycline. Cells were centrifuged and washed at specified time points. Flow cytometry was carried out using a Becton-Dickinson FACSCalibur and CellQuest (BD) software. The amount of derepression was expressed as fold derepression, and calculated by dividing mean FL1-H values of RNAi-induced cultures at each time point by the mean FL1-H values of uninduced cultures.

RESULTS

NLP is a nucleoplasmin-like protein which is essential in bloodstream form *T. brucei*

We initially identified *T. brucei* NLP in DNA affinity chromatography experiments as binding the transcriptionally silent 177-bp simple sequence repeats, which comprise the

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

bulk of the *T. brucei* minichromosomes (accession number Tb927.10.5450) (V. Tilston and K.E., unpublished data). NLP shows little sequence similarity with proteins in non-Kinetoplastid protozoa, but contains a nucleoplasmin-like domain ($e = 2.3 \times 10^{-3}$) as well as a DNA binding AT-hook motif ($e = 9.8 \times 10^{-3}$) (MyHits; ISB-SIB) (Figure 1A). A characteristic glutamate-rich region ($e = 3.8 \times 10^{-9}$) extends approximately from residues 600–670, partially overlapping the nucleoplasmin-like domain (PROSITE). Additionally, two coiled-coil regions extend between residues 429–458 and 586–691 (SMART, EMBL).

Nucleoplasmin was first identified in *Xenopus* oocytes as a highly abundant nuclear protein and has since been found in other higher eukaryotes (31,32). We aligned NLP from *T. brucei* with previously characterized nucleoplasmins from a number of higher eukaryotes using ClustalW (Figure 1B). We were able to detect certain conserved or similar amino acids in the *T. brucei* NLP nucleoplasmin-like domain as predicted using Hidden Markov Models (Pfam). These residues include a leucine residue at position 557, glycine residues at positions 560, 562 and 569 and several glutamate residues in the downstream glutamate-rich region, which is typical of the nucleoplasmin domain (Figure 1B). This polyglutamate region is one of the features shared by all nucleoplasmins, and is also found in *T. brucei* NLP (Supplementary Figure S1A). AT-hook domains are small DNA-binding motifs containing a central glycine–arginine–proline (GRP) tri-peptide sequence (33,34). In *T. brucei* NLP, the AT-hook motif has in addition to the GRP core, typical lysine and arginine residues preceding the core as well as other polar residues immediately downstream of GRP, including a characteristic lysine residue in the fourth position downstream (Supplementary Figure S1B) (34).

Western blot analysis using an anti-NLP antibody identified a ~120 kDA protein in both bloodstream and procyclic form *T. brucei* (Figure 1C). This is slightly larger than the expected size for NLP (predicted to be 107 kDa) but the minor difference in size could be the consequence of post-translational modifications. NLP is expressed at equivalent levels in both life-cycle stages. NLP is an essential protein in bloodstream form *T. brucei*, as induction of

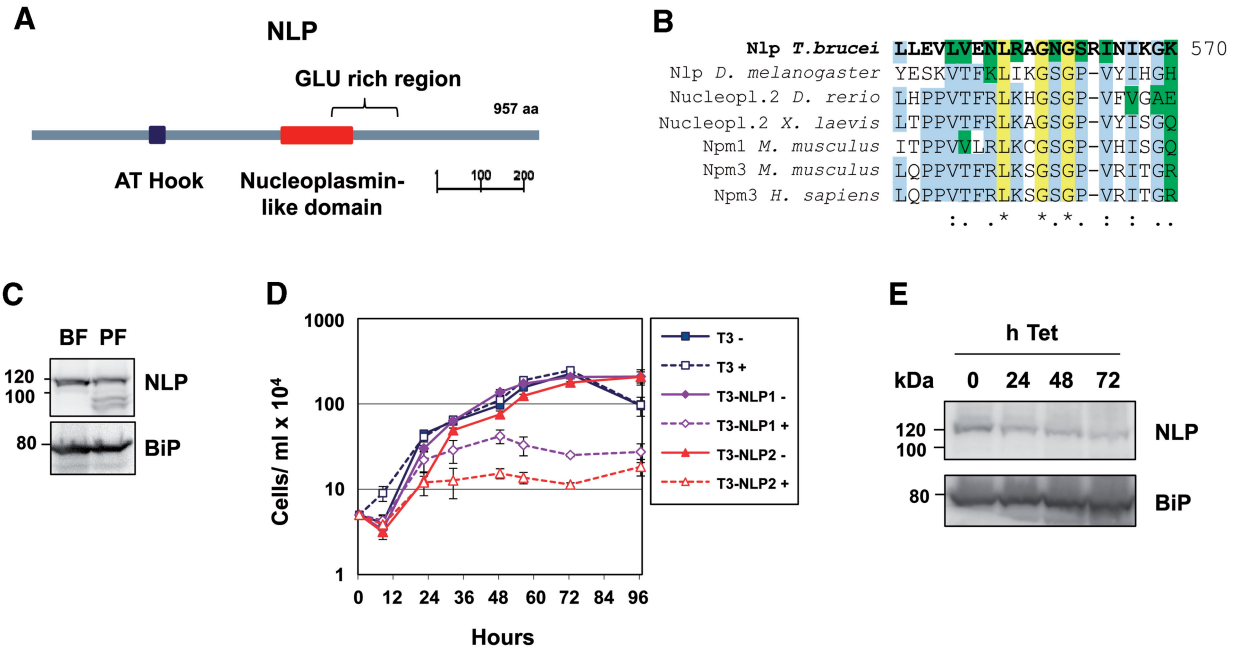


Figure 1. NLP is a nucleoplasmin-like DNA-binding protein. (**A**) Schematic of NLP with protein domains identified using PROSITE and MyHits (ISB-SIB). The NLP protein is 957 amino acids with an AT-hook domain extending from residues 214 to 226 represented with a dark blue rectangle, and a nucleoplasmin-like domain extending from residues 497 to 647 represented with a red rectangle. A glutamate (GLU)-rich region overlaps the nucleoplasmin-like domain and extends from residues 600 to 671. (**B**) Alignment using ClustalW of *T. brucei* NLP with nucleoplasmins from other eukaryotes. *T. brucei* NLP is compared with: *Drosophila melanogaster* Nucleoplasmin (Nlp), zebra fish (*Dario rerio*) Nucleoplasmin2, *Xenopus laevis* Nucleoplasmin2, *Mus musculus* Nucleoplasmin1 (Npm1) and Nucleoplasmin3 (Npm3) and *Homo sapiens* Nucleoplasmin3 (Npm3). Sequence accession numbers are listed in the 'Materials and Methods' section. A conserved region characteristic of nucleoplasmins is shown. Particularly conserved amino acids are indicated with yellow blocks and asterisks. Amino acids that are conserved in five to six sequences are highlighted with light blue blocks, and amino acids which are changed but have similar properties are indicated in green. Double dots indicate conservative and single dots indicate semi-conservative amino acid changes across all sequences as defined by ClustalW (26). (**C**) NLP is expressed at approximately equal levels in both bloodstream form (BF) and procyclic form (PF) *T. brucei*. Lysates from 1×10^7 cells were analysed on SDS–PAGE gels and the blots were probed with an antibody against NLP, or the loading control BiP. Size markers in kiloDaltons are indicated on the left. (**D**) Blocking synthesis of NLP by the induction of tetracycline inducible RNA interference in bloodstream form *T. brucei* T3-NLP1 and T3-NLP2 leads to growth arrest within 24 h. Cell lines that have been grown in the presence (+) or absence (−) of tetracycline are compared. The parental *T. brucei* T3 cell line is shown for comparison. (**E**) Reduction in NLP protein as confirmed by Western blot analysis using an anti-NLP antibody (NLP). Protein lysates were isolated from *T. brucei* T3NLP1 after the induction of NLP RNAi with tetracycline (Tet) for the time in hours (h) indicated above. The blot was also probed with an anti-BiP antibody as a loading control. Size markers in kiloDaltons are shown on the left.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

NLP RNAi in the two *T. brucei* clones T3-NLP1 and T3-NLP2 leads to an abrupt growth arrest within 24 h of the addition of tetracycline (Figure 1D). Western blot analysis confirmed the knockdown of the NLP protein to 55% of normal levels by 48 h, and to <50% after 72 h induction of *NLP* RNAi (Figure 1E). Although a clear growth arrest was observed after the induction of *NLP* RNAi, this was not correlated with a significant accumulation of cells at any particular stage of the trypanosome cell cycle (results not shown).

We next made *T. brucei* cell lines which had an endogenous allele of NLP tagged with the HA epitope (Figure 2A). Cells with one *NLP* allele tagged with the HA-epitope grew as well as the parental *T. brucei* GPI(VO2+) cell line. Knocking out the second *NLP* allele in these lines, and thereby making the cells dependent on the one HA-tagged copy of NLP did not lead to an observable change in cell growth (result not shown). This indicates that addition of the HA-epitope to the carboxy terminus of NLP does not observably compromise NLP function. Microscopic analysis of *T. brucei* lines with a copy of HA-tagged NLP showed that NLP is a nuclear protein with a relatively even distribution through the nucleus throughout the cell cycle (Figure 2B).

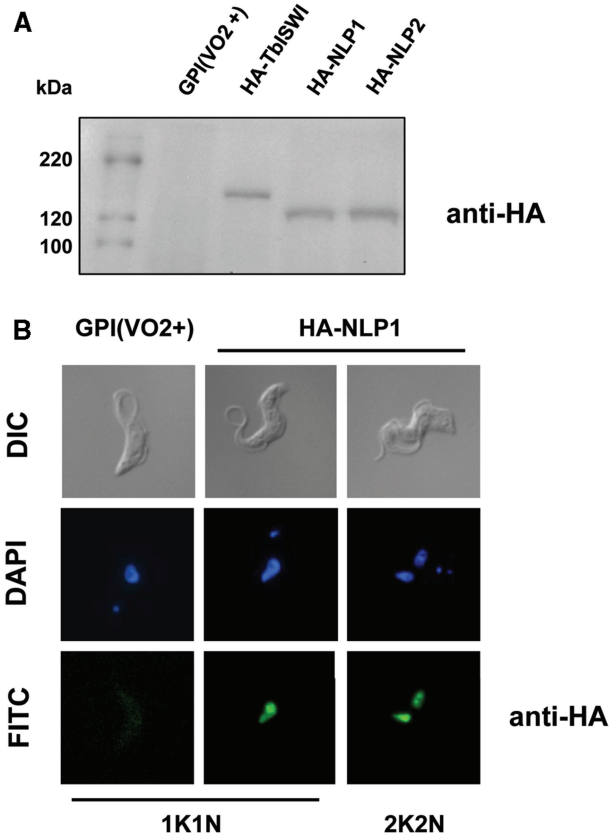


Figure 2. NLP is a nuclear protein. (**A**) Bloodstream form *T. brucei* cell lines expressing HA-epitope tagged NLP were generated (*T. brucei* HA–NLP1 and HA–NLP2). Protein lysates from these cell lines were compared with those from the parental *T. brucei* GPI(VO2+) cell line and *T. brucei* HA–TbISWI as a control for HA antibody specificity. (**B**) HA-epitope tagged NLP localizes to the nucleus. Panels show trypanosomes visualized using differential interference contrast (DIC), the DNA stain DAPI, or the FITC channel after immunofluorescence using a rabbit polyclonal anti-HA antibody and an anti-rabbit Alexa-488 secondary antibody. Cells containing either one kinetoplast (K) and one nucleus (N) (1K1N) or precytokinesis cells containing two kinetoplasts and two nuclei (2K2N) are shown.

Blocking synthesis of NLP results in derepression of silent ESs as well as other transcriptionally silent regions

As NLP had originally been identified as a DNA-binding protein binding the transcriptionally inactive *T. brucei* 177-bp repeats, we investigated if it plays a role in silencing inactive genomic regions. We first investigated its role in maintaining the repressed status of the silent *VSG* ESs. The bloodstream form *T. brucei* T3-SM cell line has a construct containing *eGFP* inserted immediately downstream of the promoter of the silent *VSG221* ES allowing monitoring for ES derepression with flow cytometry (19) (Figure 3A). In this cell line a blasticidin resistance gene has been inserted into the VSGT3 ES, allowing the maintenance of a population homogeneous for expression of VSGT3 in the presence of blasticidin selection. We induced RNAi against *NLP* using tetracycline, and followed ES derepression by flow cytometry.

Induction of *NLP* RNAi resulted in a striking 45- to 65-fold derepression of the silent *VSG221* ES after 72 h of induction of *NLP* RNAi (Figure 3B). In contrast, ES derepression was not observed when we induced RNAi against the essential DNA binding protein TDP1 (27) (Supplementary Figure S2), indicating that increased expression of *eGFP* was not simply a stress response caused by the absence of an essential protein or an artefact as a result of RNAi induced cell death. We also induced RNAi against *NLP* in the the *T. brucei* 121-SM cell line (19). This is an analogous bloodstream form *T. brucei* reporter cell line, with the *VSG121* ES active rather than the *VSGT3* ES. Here we observed 19- to 52-fold derepression of *eGFP* (Supplementary Figure S3).

Given the abundance of NLP in the nucleus, it seemed unlikely that its biological role was restricted to ES silencing. We were therefore next interested in determining the role of NLP in silencing other transcriptionally inactive areas of the *T. brucei* genome. In *T. brucei,* both rDNA promoters and ES promoters are transcribed by RNA polymerase I (35). Of the two promoters, transcription from the core ES promoter appears to be more sensitive to its chromosomal context. For example, an exogenous ES promoter was shown to be more comprehensively silenced than a comparable rDNA promoter when inserted in the transcriptionally silent *VSG* Basic Copy arrays (36). This observation allows the use of the ES promoter as a 'probe' for repressed chromatin regions (37).

As *T. brucei* NLP had originally been identified as a protein binding the 177-bp repeats, we investigated its role in silencing the *T. brucei* minichromosomes, which are comprised primarily of arrays of these simple sequence repeats (38). We therefore integrated a construct containing an ES core promoter upstream of *eGFP* into *T. brucei* minichromosomes (*T. brucei* MCEP in

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

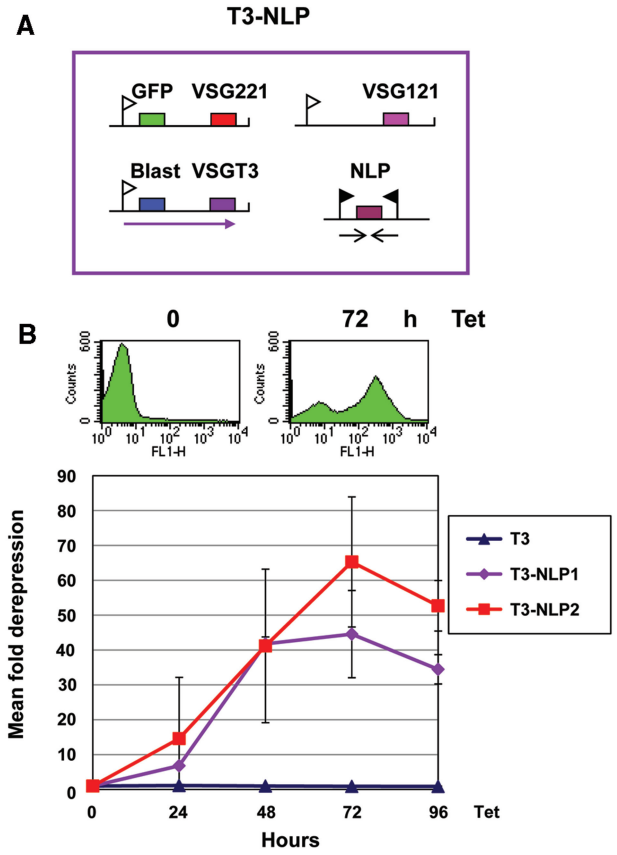


Figure 3. Blocking synthesis of NLP in bloodstream form *T. brucei* results in derepression of the silent *VSG221* ES. (**A**) Schematic of the *T. brucei* T3-NLP line. The large box indicates a trypanosome which expresses VSGT3 from the active *VSGT3* ES (transcription indicated with an arrow). A blasticidin resistance gene has been inserted immediately downstream of the active *VSGVT3* ES promoter (indicated with a white flag), allowing maintenance of a population homogeneous for VSGT3 expression through selection on blasticidin. The silent *VSG221* ES has *eGFP* inserted immediately behind the promoter. ES derepression can be monitored after the induction of *NLP* RNAi from tetracycline inducible T7 promoters (black flags). (**B**) The silent *VSG221* ES is 45–65-fold derepressed after the induction of *NLP* RNAi with tetracycline (Tet) for the time indicated in hours (h). *T. brucei* T3-NLP1 and *T. brucei* T3-NLP2 are compared with the parental *T. brucei* VSGT3 expressor (T3). Above are representative flow-cytometry traces in the FL1 channel after the induction of *NLP* RNAi for 0 or 72 h. Below is a graph showing the fold derepression at each time point calculated by dividing the mean fluorescence of a tetracycline induced culture with an uninduced culture at the same time point. Error bars show the standard deviation between three experiments.

Figure 4A). The blasticidin resistance gene used for selection in this construct is driven by an rDNA promoter. Induction of *NLP* RNAi resulted in 5–6-fold derepression of *eGFP*, indicating that NLP plays a role in silencing the minichromosomal 177-bp repeat regions. Although the degree of derepression is ~10-fold higher in the *T. brucei* T3-SM line compared with the MCEP line, the FL1-H signal after 72 h induction of *NLP* RNAi is comparable. This is explained by the low basal level of transcription from the ES promoter inserted into a transcriptionally silent minichromosome (as seen in the FACS trace for time point 0 in Figure 4A). An equivalent *T. brucei* reporter line which contained a construct without an ES promoter upstream of *eGFP* inserted into minichromosomes was also tested. Induction of *NLP* RNAi in this second cell line did not result in a significant increase in *eGFP* transcription (data not shown). This indicates that although blocking the synthesis of NLP appears to facilitate transcription from the ES promoter, it does not lead to a significant increase in fortuitous initiation of transcription from cryptic promoters at minichromosomes.

As it had earlier been shown that the ES promoter is particularly repressed if integrated into the *VSG* Basic Copy arrays of bloodstream form *T. brucei* (36), we next investigated the role of NLP in silencing this area of the genome. The VBEP construct, containing an ES promoter upstream of *eGFP*, was integrated upstream of the silent *VSG118* Basic Copy gene (Figure 4B). As seen after integrating an exogenous ES promoter in the minichromosomes, there is a low rate of transcription from an ES promoter integrated into the *VSG* Basic Copy arrays. Blocking synthesis of NLP led to a ~5-fold derepression of *eGFP* by 72 h (Figure 4B). As in the minichromosome targeting experiments, a comparable reporter line containing a construct without an ES promoter upstream of *eGFP* inserted upstream of the *VSG118* Basic Copy gene did not show significantly increased transcription of *eGFP* after the induction of *NLP* RNAi (result not shown). Therefore, although NLP appears to play a role in establishing the repressive chromatin environment silencing the exogenous ES promoter in the *VSG* Basic Copy arrays, depleting its synthesis does not result in a significant increase in fortuitous initiation of transcription from cryptic promoters in this genomic region. Our observed derepression of *eGFP* is not simply the consequence of a lethal phenotype, as induction of RNAi against the essential TDP protein in the *T. brucei* MCEP and VBEP cell lines did not lead to an increase in *eGFP* transcription (27) (results not shown).

Lastly, the role of NLP was investigated in silencing procyclin genes, which are to a great extent transcriptionally inactive in bloodstream form *T. brucei* (39,40). We used a cell line where a reporter construct was integrated behind the silent EP1 promoter in bloodstream form *T. brucei*. Although procyclin protein is not detectably expressed in bloodstream form *T. brucei,* there is a small amount of transcription from procyclin promoters. This explains the observed *eGFP* fluorescence in the FACS trace at time point 0 in Figure 4C. Blocking synthesis of NLP leads to a ~4-fold increase in transcription of *eGFP* driven by the silenced EP1 promoter (Figure 4C). In summary, blocking the synthesis of NLP had the most spectacular effect on derepressing silent *VSG* ESs in bloodstream form *T. brucei*. However, it also appears to play a role in establishing a repressive transcriptional environment for the ES promoter in bloodstream form *T. brucei* at different genomic regions, including the transcriptionally silent *T. brucei* minichromosomes, the *VSG* Basic Copy arrays and the inactive *EP1* transcription units.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

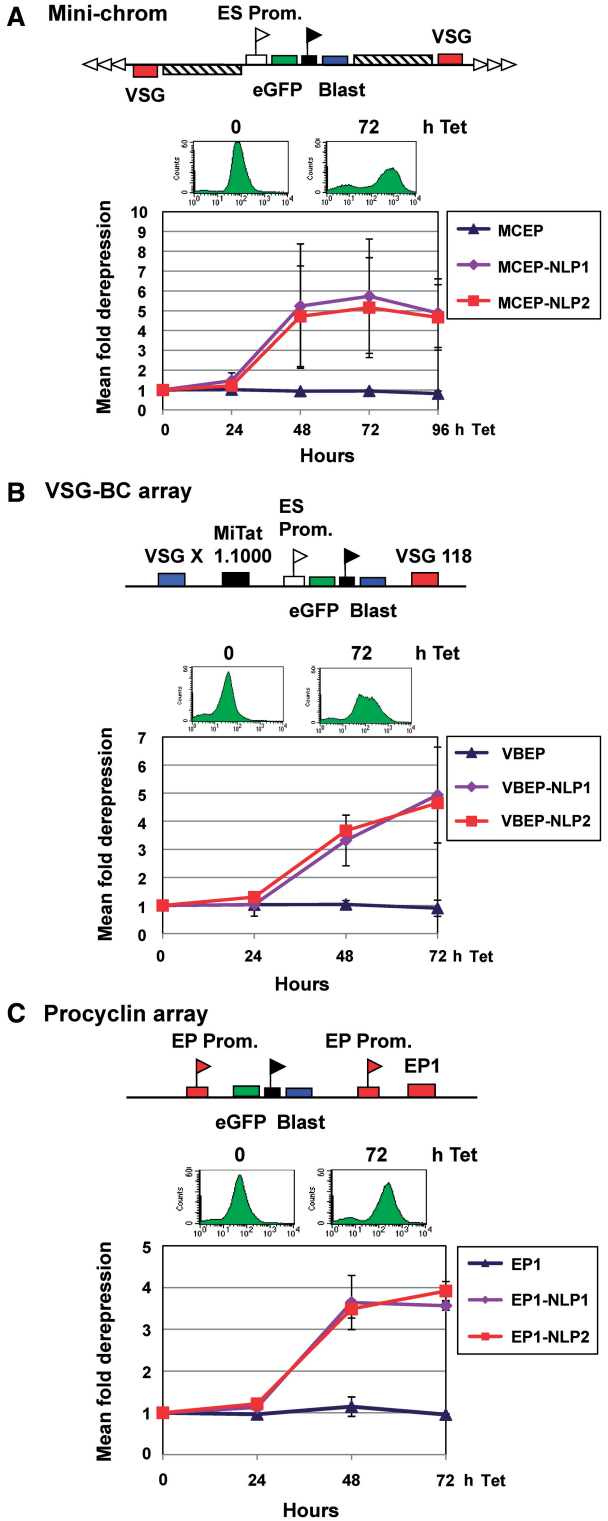


Figure 4. NLP plays a role in silencing transcriptionally inactive regions of the bloodstream form *T. brucei* genome. (**A**) Blocking synthesis of NLP leads to derepression of an ES promoter located in a transcriptionally silent minichromosome (Mini-chrom). A construct containing an ES promoter (white flag) was inserted into a *T. brucei* minichromosome using characteristic 177-bp repeat sequences which make up the bulk of the minichromosomes (diagonally hatched box). Silent *VSG*s are located at the telomeres, which are indicated with horizontal triangles. The blasticidin resistance gene (Blast) used for selection is driven by an rDNA promoter (black flag). The *T. brucei* MCEP-NLP1 and MCEP-NLP2 lines were monitored for derepression of the ES promoter using flow cytometry after the induction of *NLP* RNAi with tetracycline (Tet) for the time indicated in hours (h). Analysis of the parental *T. brucei* MCEP line is shown for comparison. Characteristic FACS traces are shown above. The results are the mean of triplicate *NLP* RNAi experiments with the standard deviation indicated with error bars. (**B**) Blocking synthesis of NLP leads to derepression of an ES promoter inserted into a transcriptionally silent *VSG* Basic Copy (BC) array. The experiment was performed as in (A) where a similar construct containing an ES promoter was inserted into the *VSG118* Basic Copy array, with *VSGX, MiTat1.100* and *VSG118* indicated with coloured boxes. Derepression of *eGFP* was monitored by flow cytometry. Above are representative flow cytometry traces after the induction of *NLP* RNAi with tetracycline. The *T. brucei* VBEP-NLP1 and *T. brucei* VBEP-NLP2 cell lines are compared with the parental *T. brucei* VBEP line. (**C**) Induction of *NLP* RNAi results in derepression of an endogenous procyclin EP1 promoter in bloodstream form *T. brucei*. The EP1 promoters (including the duplicated EP1 promoter from the construct) are indicated with red flags. Similar to panels (A) and (B), derepression of *eGFP* was monitored by flow cytometry after the induction of *NLP* RNAi with tetracycline for the time indicated. The *T. brucei* EP1-NLP1 and EP1-NLP2 lines are compared with the parental *T. brucei* EP1 cell line.

Binding of NLP is enriched in certain non-transcribed regions of the *T. brucei* genome

We next attempted to determine where NLP binds in the *T. brucei* genome using ChIP experiments. For this purpose, we used bloodstream form *T. brucei* cell lines where the carboxy terminus of one NLP allele had been epitope tagged with the HA epitope (Figure 2A). We cross-linked chromatin using formaldehyde in bloodstream form *T. brucei,* and used an antibody against the HA-tag to immunoprecipitate NLP-HA cross-linked to DNA. As a positive control, we also performed the ChIP experiment with an antibody to histone H3, or as a negative control performed the ChIP procedure without using an antibody (18). The ChIP procedure was also carried out using the parental *T. brucei* 221GPI(VO2+) cells to provide an additional negative control for specificity of the anti-HA antibody.

As NLP had originally been identified binding the 177-bp repeats, we first investigated its distribution on these sequences. Slot blot analysis showed that NLP indeed binds the 177-bp repeats (Figure 5A). In addition, NLP was found binding the 50-bp repeats which extend as large arrays upstream of all known *T. brucei* ESs, at levels that were ~3- to 4-fold background, with non-overlapping error bars indicating standard deviation (Figure 5A). We also analysed the presence of NLP in various other genomic locations. Quantitative PCR (qPCR) was performed on the input chromatin, as well as immunoprecipitated material from the *T. brucei* NLP–HA and parental cell lines to detect the presence of NLP. Schematics of the different transcription units investigated are shown in Figure 5. NLP was found at the spliced leader (SL) genes transcribed by RNA polymerase II (Figure 5B and D) (41) as well as at the 5S rRNA transcription units transcribed by RNA polymerase III (Figure 5C and D). Here also, we found relatively high amounts of NLP binding both in the intergenic spacers and within the genes themselves.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

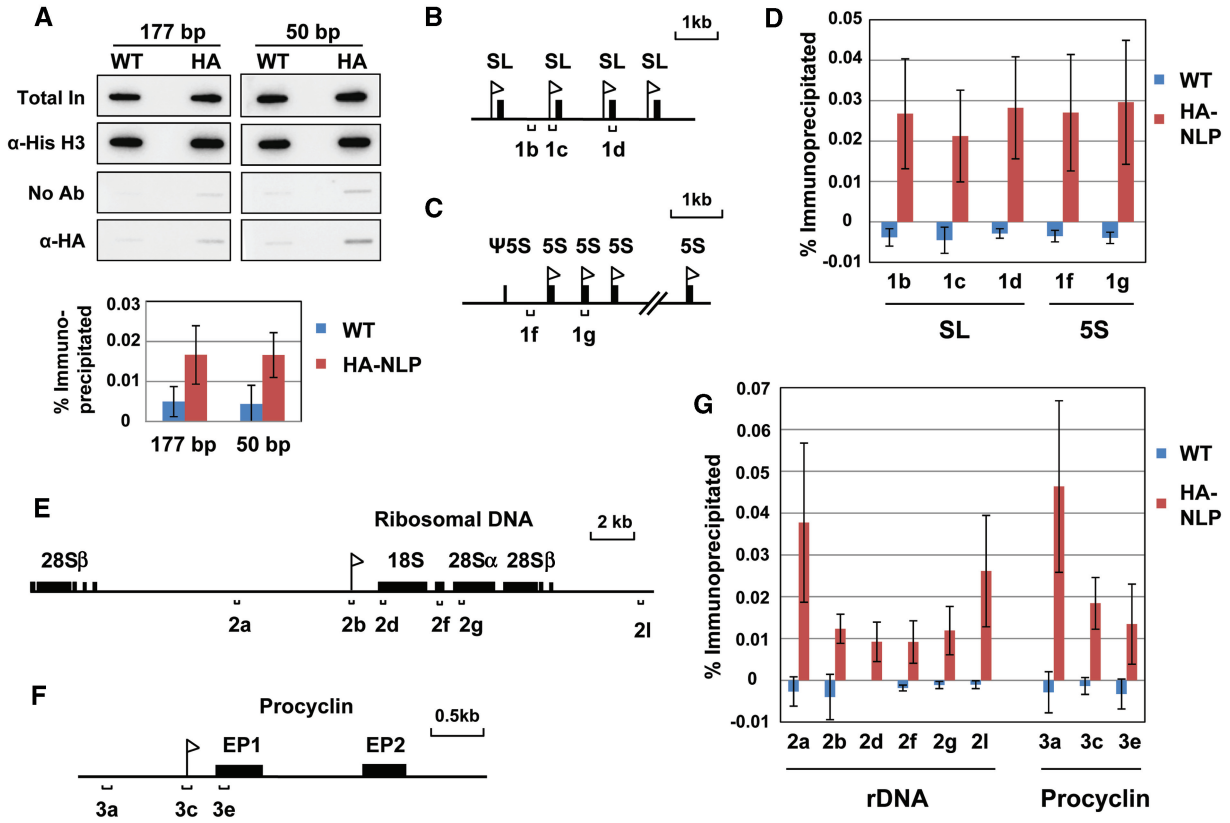


Figure 5. ChIP experiments show enrichment of NLP in simple sequence repeats as well as the non-transcribed regions around the rDNA and procyclin loci. The bloodstream form *T. brucei* HA–NLP cell line containing one allele of *NLP* tagged with the HA epitope (HA–NLP) (red bars), or the parental wild-type cell line (WT) (blue bars) was used. In all of these experiments immunoprecipitation with an anti-HA antibody was performed, and the relative amount immunoprecipitated is shown after subtraction of the values quantitated in a corresponding ChIP experiment performed in the absence of antibody. Primer pairs used for quantitative PCR (qPCR) in these ChIP experiments are indicated with numbered brackets, and are described in (18) or Supplementary Table S1. All graphs show the mean of triplicate ChIP experiments, with standard deviation indicated with error bars. (**A**) NLP is enriched on the 177-bp repeats comprising *T. brucei* minichromosomes, as well as the 50-bp repeats present upstream of all known ESs. Representative slot blots of immunoprecipitated chromatin hybridized with probes for the relevant repeat region are shown. Material from the parental cell line (WT) is compared with that from the HA–NLP line (HA). Total input (Total In) is compared with material precipitated either with an anti-histone H3 antibody (α-His H3), no antibody (No Ab) or an anti-HA epitope antibody (α-HA). A graph showing the amounts immunoprecipitated (after subtraction of the no antibody control) is indicated below the slot blot. Error bars indicate the standard deviation from three independent experiments. (**B**) Schematic diagram showing the spliced leader (SL) RNA transcription units transcribed by RNA polymerase II. The relevant genes are indicated with black boxes, with the SL promoters indicated with flags. (**C**) Diagram of the 5S rRNA locus transcribed by RNA polymerase III. The 5S genes and pseudo-gene (ψ) are indicated with black boxes and the 5S promoters with flags. (**D**) Quantitation of ChIP results using the *T. brucei* HA–NLP and parental wild-type (WT) cell lines. Immunoprecipitation with an anti-HA antibody was performed, and the relative amount of SL DNA or 5S rDNA immunoprecipitated is shown after subtraction of the values quantitated in a corresponding ChIP experiment using no antibody. (**E**) Diagram of a typical Pol I transcribed ribosomal DNA (rDNA) locus with the rDNA promoter indicated with a flag. (**F**) Schematic of the Pol I transcribed procyclin EP locus with two tandem EP genes indicated with black boxes. The EP promoter is indicated with a flag. (**G**) Quantitation of ChIP results as in D). Immunoprecipitation with an anti-HA antibody was performed, and the relative amount of rDNA or procyclin DNA immunoprecipitated is shown.

RNA polymerase I transcribes procyclin and *VSG* genes as well as the rDNA in *T. brucei* (35). First NLP distribution within the rDNA was determined. Relatively low levels of NLP were found within the rDNA transcription unit (primers 2b, 2d, 2f and 2g in Figure 5E and G). However, levels of NLP were enriched in the non-transcribed rDNA spacer either upstream or downstream of a given rDNA transcription unit (primers 2a and 2l; Figure 5E and G). Compared with levels of NLP in the 18S or 28S rDNA, levels of NLP were over 3-fold higher using primer pair 2a and nearly 3-fold using primer pair 2l. An approximately 3-fold higher level of NLP was also found upstream of the procyclin *EP1* and *EP2* transcription units (primer 3a in Figure 5F and G) compared to within the *EP1* gene itself (primer 3c and 3e). In summary, in a variety of *T. brucei* transcription units, higher levels of NLP are found binding the adjacent transcriptionally silent regions compared to within the transcription units themselves.

As downregulation of NLP leads to striking levels of derepression of silent ESs, we next investigated its binding to ES sequences (Figure 6). As ESs are highly similar to each other in sequence (12), primers can be expected to bind most if not all ESs. To circumvent this problem, we performed ChIP in a bloodstream form *T. brucei* line that contained a single copy neomycin

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

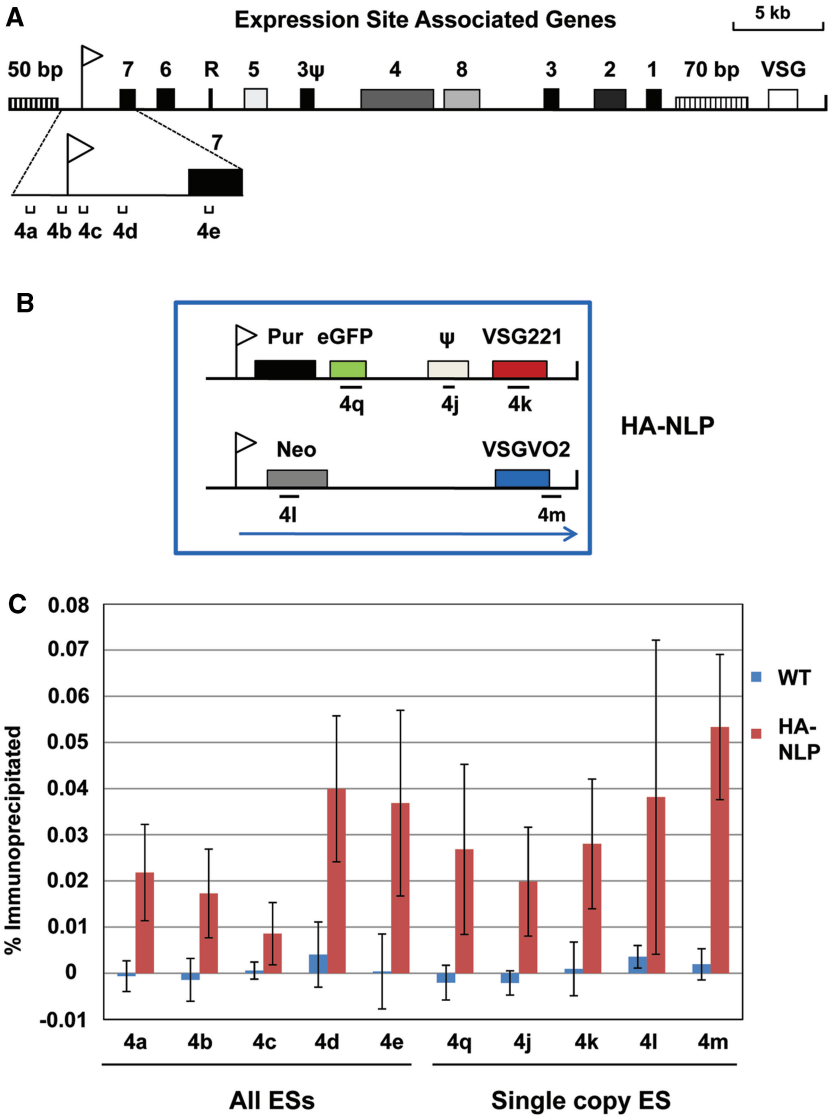


Figure 6. NLP is relatively depleted at ES promoters but binds both active and silent ESs. ChIP experiments were performed with the *T. brucei* HA–NLP and parental wild-type (WT) cell lines. (**A**) A schematic of a typical ES is shown above with the ES promoter shown with a flag, and characteristic Expression site associated genes (*ESAG*s) indicated with filled boxes. The immediate area around the ES promoter is expanded below. Characteristic 50-bp repeat arrays flanking the ES promoter, or 70-bp repeat sequences present upstream of the telomeric *VSG* are indicated with vertically striped boxes. Primers used are as described in (18). As ESs are multi-copy and highly similar, these primers can be expected to recognise most if not all ESs. (**B**) Schematic of the *T. brucei* HA–NLP cell line used for ChIP. Transcription of the active *VSGVO2* ES from the ES promoter (flag) is indicated with an arrow. The single copy neomycin resistance gene is indicated (Neo). The silent *VSG221* ES contains a construct containing a puromycin resistance gene (Pur) and *eGFP*. Other single copy genes are the single copy *VSG* pseudogene (ψ), as well as *VSG221*. These single copy genes are indicated with filled boxes. Relevant primers are indicated with numbered brackets. Primers used are as described in (18) or in Supplementary Table S1. (**C**) Quantification of the immunoprecipitated material using an anti-HA antibody to detect the HA-epitope tagged NLP using either the *T. brucei* HA–NLP cell line (HA–NLP) (red bars) or the wild-type cell line (WT) (blue bars) after subtraction of the no antibody control as described in Figure 5. Results are the mean of triplicate ChIP experiments, with standard deviation indicated with error bars.

resistance gene immediately downstream of the promoter of the active VSGVO2 ES, as well as a construct containing a puromycin resistance gene and *eGFP* behind the silent *VSG221* ES promoter (Figure 6B). The *VSG* pseudogene (ψ) and the *VSG221* gene are additional single copy sequences present within the silent *VSG221* ES. All of the common ES primers used ('All ESs' in Figure 6C) can be expected to recognise the active as well as most if not all of the silent ESs. Although we were able to find a low amount of NLP binding within the ES promoter region (primer 4c in Figure 6C), we could not distinguish between active and silent ESs with this primer set. qPCR analysis using single copy sequences specific for either the active or silent ES showed NLP binding to both silent and active ESs along the length of the telomere.

The levels of NLP immunoprecipitated in our experiments are low, but comparable to those observed for some chromatin remodeling proteins in other systems, for example ISWI in *Drosophila* (42,43). Typically, in our

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

experiments HA–NLP immunoprecipitation was ten to 30-fold enriched over background levels obtained after immunoprecipitation with an anti-HA antibody in wild-type cell line. The relative inefficiency of NLP ChIP compared with the histone H3 ChIP control (results not shown) (18), is presumably due to the relatively low affinity of NLP for DNA, as one might expect for a nucleoplasmin-like protein compared with a histone or a specific transcription factor.

Depletion of NLP leads to reduced transcription of the active ES

As induction of *NLP* RNAi resulted in 45- to 65-fold derepression of an *eGFP* gene present immediately downstream of the inactive *VSG221* ES promoter, we next wanted to know if this derepression led to fully processive ES transcription resulting in full activation of silent ESs or increased *VSG* ES switching. Total RNA was isolated from bloodstream form *T. brucei* NLP1 after the induction of *NLP* RNAi for different time points, cDNA was prepared and levels of various transcripts were monitored by qPCR. As expected, induction of *NLP* RNAi led to a reduction in *NLP* transcript down to ~25% normal levels by 48 h (Figure 7). Also as expected from the flow cytometry data which showed an increase in eGFP protein (Figure 3), *eGFP* transcript from the derepressed *VSG221* promoter increased almost 20-fold after the induction of *NLP* RNAi. However, despite the striking derepression of the *VSG221* ES promoter observed after the induction of *NLP* RNAi, there was no significant increase in the *VSG* transcripts themselves from a selection of six silent *VSG* ESs. *VSG* transcripts monitored include those from the *VSG13, VSG121, VSG224, VSG800, VSGbR2* and *VSGJS1* ESs (12,44). *VSG* transcripts from silent ESs were particularly low after the induction of *NLP* RNAi for 48 h. The subsequent increase in these transcripts at 72 h could be a consequence of increases in levels of NLP, possibly as a result of the emergence of revertants escaping a lethal RNAi phenotype.

A striking observation from these experiments is that the level of the active *VSGT3* transcript itself fell drastically to 4% normal levels after 48 h induction of *NLP* RNAi. Levels of transcript from the blasticidin resistance gene present in the active *VSGT3* ES also fell almost

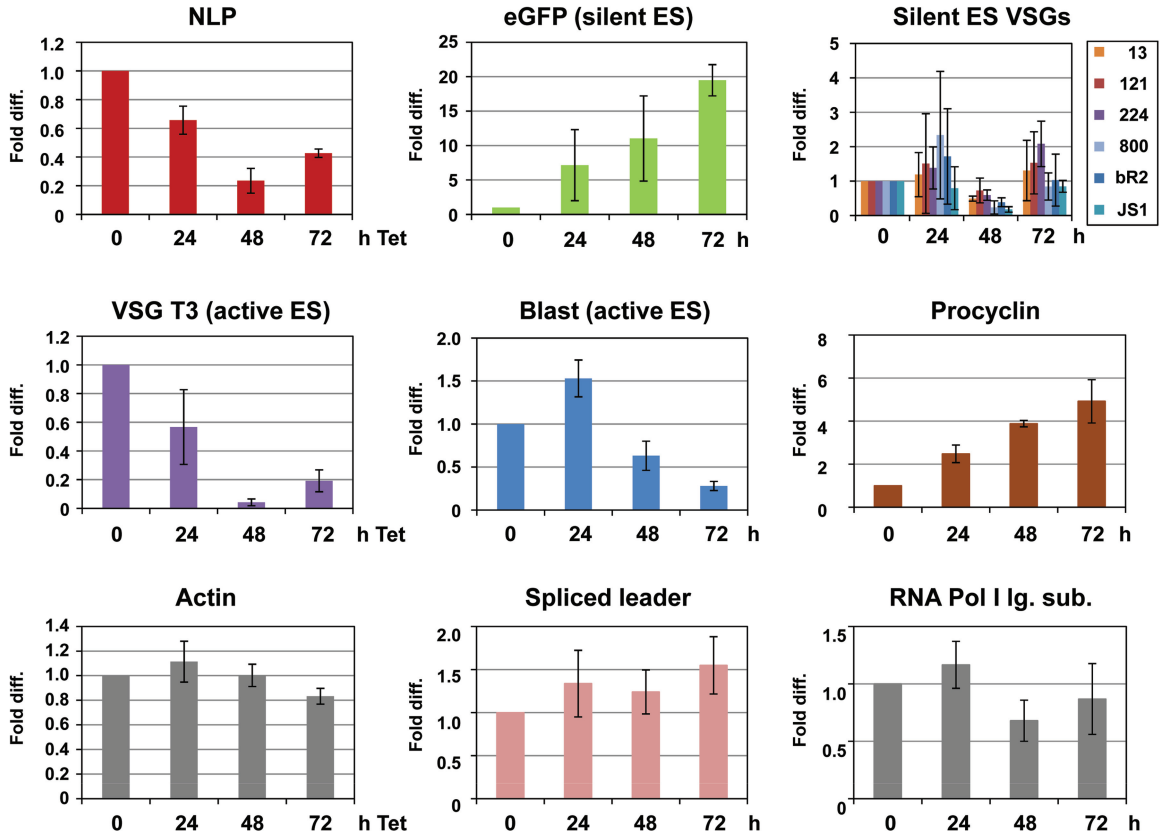


Figure 7. Quantitation of transcripts in the *T. brucei* T3-NLP1 cell line after the induction of *NLP* RNAi with tetracycline (Tet) for the time indicated in hours (h). RNA was collected from cells at different time points, cDNA was prepared and qPCR was performed to monitor changes in transcript levels. Data was normalized to the level of actin transcript at each time point, and is plotted as the fold difference between the induced and uninduced time points. Relative levels of transcript after the induction of *NLP* RNAi are shown for: *NLP*, *eGFP* present in the silent *VSG221* ES, a representative number of silent ES-located *VSGs*, *VSGT3* and the blastidin resistance gene (Blast) present in the active *VSGT3* ES, EP1 procyclin, spliced leader and the genes encoding actin and the large subunit of RNA polymerase I. The silent *VSG*s are all located at the telomeres of silent ESs, and are: *VSG13*, *VSG121*, *VSG224*, *VSG800*, *VSGbR2* and *VSGJS1* (12,44). All data shown are the mean of triplicate RNAi induction experiments with standard deviation represented with error bars. The data was normalized against the values for actin transcript shown in this figure, with the exception of the quantitation of the EP1 procyclin transcript, which was part of a separate triplicate experiment and normalized to an independent quantitation of actin.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

3-fold after the induction of *NLP* RNAi for 72 h. These results indicate that NLP plays an important role in facilitating fully processive transcription from the active *VSG* ES as well as silencing the silent ESs. This lack of processive transcription from the active ES in the face of an NLP synthesis block could explain why transcripts from silent *VSG* ESs also appear reduced after 48 h. Levels of *VSG* transcripts from the active and silent *VSG*s were comparable irrespective of whether the cells had been maintained on blasticidin selection to select for continued expression of VSGT3. These results were also confirmed using an analogous bloodstream form *T. brucei* cell line expressing *VSG121* rather than *VSGT3* from the active ES, even though the degree of NLP knockdown was less pronounced in this cell line (Supplementary Figure S4).

Transcripts derived from the Pol I transcribed EP1 procyclin genes increased ~5-fold after the induction of a block in NLP synthesis (Figure 7), in agreement with the observed derepression of the EP1 procyclin promoter (see Figure 4C). In these experiments performed in triplicate, transcript amounts are presented as normalized to the amount of actin transcript present at each time point. As a reference point, the amount of actin transcript quantitated at each time point after the induction of *NLP* RNAi is also shown (Figure 7). Quantitated levels of the spliced leader or the transcripts encoding actin or the RNA polymerase I large subunit did not change significantly after NLP knockdown (Figure 7).

DISCUSSION

Here we describe a novel transcription regulatory protein NLP, which is essential, and has a nucleoplasmin-like domain and an AT-hook motif. *T. brucei* NLP plays a key role in ES control, as blocking its synthesis leads to 45- to 65-fold derepression of silent ESs. NLP also plays a role in silencing other transcriptionally inactive areas in bloodstream form *T. brucei* including the minichromosomes, the *VSG* Basic Copy arrays and the procyclin locus. Using ChIP analysis, NLP was found to bind simple sequence repeats as well as certain silent areas around Pol I transcription units in the *T. brucei* genome. However, despite the observed derepression of silent ESs after the induction of *NLP* RNAi, we did not find evidence that NLP preferentially binds silent ESs. Also, in spite of the striking degree of derepression of ES promoters observed after NLP knockdown, we did not find fully processive transcription of silent ESs or an observable increase in ES switching. Blocking synthesis of NLP resulted in a drastic reduction in processive transcription of the active ES. As one needs to have full activation of silent ESs in order to see an increase in *VSG* ES switching, the observed lack of change in *VSG* switching could be a consequence of the lack of processive transcription in the derepressed silent (as well as active) *VSG* ESs.

It has only recently become apparent that ES regulation in bloodstream form *T. brucei* involves chromatin remodeling. It is now clear that the active ES in bloodstream form *T. brucei* is radically depleted in nucleosomes compared with the silent ESs (17,18). Therefore chromatin remodeling is highly likely to be involved in the conversion of an ES from its silent to active state. The first chromatin remodeler identified to be involved in ES silencing in *T. brucei* was TbISWI, which is a member of the SNF2 superfamily of ATP-dependent chromatin remodeling complexes (19). Knockdown of TbISWI results in 30–60-fold derepression of silent ESs. In addition, the telomere binding protein TbRAP1 is involved in a gradient of silencing operating at and increasing in strength towards the ES chromosome end (20). Blocking synthesis of TbRAP1 results in an increase in ES switching as well as ES promoter derepression. Lastly, the histone methyltransferase DOT1B appears to be essential for mono-allelic expression of VSG (21). Inhibition of DOT1B synthesis results in silent ESs becoming partially derepressed, with the full switch from one ES to another occurring at a slower rate than average.

NLP does not closely resemble known chromatin remodeling proteins in other eukaryotes, and is a new addition to the small list of proteins shown to be involved in ES control in *T. brucei*. One of its distinctive features is a nucleoplasmin motif, which is similar in size to a typical nucleoplasmin protein. Nucleoplasmins are small (~150–300 amino acids) acidic proteins which function as histone chaperones (45,46). Nucleoplasmin was originally identified in *Xenopus* oocytes as a highly abundant protein which can assemble nucleosome cores *in vitro* from purified histones and DNA (31). Nucleoplasmin mediates sperm chromatin decondensation in *Xenopus* where it binds and removes sperm basic proteins and replaces them with histones H2A and H2B (47,48). In addition, nucleoplasmin has been shown to be important for regulating the chromatin condensation that occurs during apoptosis (49). In *Drosophila* a nucleoplasmin like protein (dNLP) binds to core histones, and functions in assembling regularly spaced nucleosomal arrays in an ATP-dependent fashion (50).

Is NLP a nucleoplasmin? Nucleoplasmins are typically small phosphoproteins which can function as pentamers (51). NLP at almost 1000 amino acids is more than five times bigger than a typical nucleoplasmin. In addition, it contains a non-typical AT-hook domain. AT-hook motifs are small (~11 amino acid) DNA binding motifs centred around a GRP tripeptide, which is necessary and sufficient to mediate DNA binding (33). This motif typically functions as an auxiliary protein motif, which in conjunction with other DNA binding motifs in the same protein facilitates binding to DNA (34). The AT-hook motif has also been found in a very wide variety of different DNA binding proteins including transcription regulators and co-factors of different DNA binding proteins (34). *T. brucei* NLP does not appear to be a nucleoplasmin in the strict sense of the word, but the homology to this type of protein could indicate that it interacts with histones in trypanosomes. Further experiments will be required to address this issue.

What is the role of NLP in the mono-allelic exclusion of ESs in bloodstream *T. brucei*? NLP was originally identified in DNA affinity chromatography experiments using procyclic form *T. brucei* extracts, and is not

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

specific to bloodstream form trypanosomes. Compatible with being a general transcription regulator, levels of NLP protein appear to be equivalent in both procyclic and bloodstream form *T. brucei* life-cycle stages. In bloodstream form trypanosomes, in addition to its importance for the establishment of ES silencing, NLP also plays a role in silencing other transcriptionally inactive regions including the minichromosomes, *VSG* Basic Copy arrays and the procyclin loci. This indicates that NLP is a general transcription regulator, and is not uniquely involved in ES control. This is consistent with its localisation in ChIP experiments to the rDNA non-transcribed spacer, and the region upstream of the Pol I transcribed procyclin loci. In the ES itself, NLP appears to bind both silent and active ESs indicating that it could play a functional role in both locations.

Compatible with this observation, knockdown of NLP synthesis resulted in reduced transcription from the active Pol I transcribed ES itself. This phenomenon was not observed with the Pol II transcribed actin, tubulin or RNA polymerase I large subunit transcripts. It is possible that chromatin remodeling is required to maintain transcription of the regulated ESs, which are transcribed in trypanosomes at a very high rate. In contrast, there does not appear to be significant transcriptional regulation of the Pol II transcribed arrays of housekeeping genes. NLP could have chromatin remodeling activity which is important for maintaining fully processive transcription of the active ES. This could, for example, operate through facilitating removal of histones including histone H3, which has been shown to be particularly depleted from the active ES (17,18). These different activities of NLP argue that it plays a role in both facilitating and inhibiting transcription. There is precedent for this, as the ISWI complexes in other eukaryotes provide a good example of a chromatin factor involved in both transcription repression and activation within a single cell. In *Saccharomyces cerevisiae* the Isw1a complex (consisting of Isw1p and Ioc3p) is involved in transcription silencing, while the Isw1b complex (consisting of Isw1p together with Ioc2p and Ioc4p) is involved in transcription elongation (52).

A challenge will come in determining how NLP interacts with other chromatin remodeling proteins and histone modification enzymes including ISWI, RAP1 and DOT1B in maintaining ES control. It is likely that multiple silencing gradients operate simultaneously on silent ESs in *T. brucei* (20,53). A relatively short distance repressive gradient extends up from the *T. brucei* ES telomere, and involves the sirtuin protein SIR2 (54,55). SIR2 is involved in repression operating in the immediate proximity of the telomere repeats, but does not appear to play a role in the long distance repressive effect operating over many tens of kilobases, necessary for control of the ES promoter (55). In contrast, the telomere binding protein RAP1 establishes a silencing gradient operating over much longer distances including the ES promoter, and increasing in intensity closer to the ES telomere (20). Blocking RAP1 synthesis results in derepression of silent ESs, and an increase in transcripts from *VSG*s located at silent ES telomeres.

Blocking NLP synthesis results in derepression of a silent ES promoter but no observable increase in VSG switching. One possibility is that NLP is involved in establishing or maintaining repressed chromatin in the immediate vicinity of the ES promoter. Blocking NLP synthesis could possibly result in promoter derepression, but as the derepressed chromatin is relatively localized, the Pol I does not transcribe in a processive fashion down to the telomere end. This could be the consequence of a second repressive gradient (possibly mediated by RAP1) impeding the elongating RNA polymerases. A complication with interpreting our findings however is the fact that transcription of the active ES itself is impeded by NLP knockdown. This implies that NLP could be involved in establishing the chromatin environment necessary for the elongating Pol I transcription complex. A challenge for the future will be in unravelling the interplay of the multiple regulatory factors implicated in ES control.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Tomoko Isobe, Matthew Wand and Andrew Voak for making DNA constructs and performing initial experiments. They are very grateful to Prof. Jane Mellor (Department of Biochemistry, University of Oxford) for help with the ChIP experiments and allowing use of the BioRuptor sonicator. They are grateful to Prof. Thomas Seebeck (University of Bern) for providing epitope tagging constructs. They thank Prof. Jay Bangs (University of Wisconsin) for providing anti-BiP antibody. They are grateful to Megan Lindsay Povelones, Viola Denninger, Nadina Vasileva Wand, Simone Wiesler, Robin Chisman, Marcin Dembek and Siobhan Durran for comments on the article.

FUNDING

Wellcome Trust; G.R. is a Wellcome Senior Fellow in the Basic Biomedical Sciences; Clarendon Fund and the Queen's College Graduate Scholarship, Oxford funding to M.S.N.; Inlaks Foundation and the Clarendon Fund funding to M.K. Funding for open access charge: The Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Serizawa,S., Miyamichi,K. and Sakano,H. (2004) One neuron-one receptor rule in the mouse olfactory system. *Trends Genet.*, **20**, 648–653.
2. Lomvardas,S., Barnea,G., Pisapia,D.J., Mendelsohn,M., Kirkland,J. and Axel,R. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell*, **126**, 403–413.
3. Fuss,S.H., Omura,M. and Mombaerts,P. (2007) Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, **130**, 373–384.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

4. Scherf,A., Lopez-Rubio,J.J. and Riviere,L. (2008) Antigenic variation in Plasmodium falciparum. *Ann. Rev. Microbiol.*, **62**, 445–470.
5. Voss,T.S., Healer,J., Marty,A.J., Duffy,M.F., Thompson,J.K., Beeson,J.G., Reeder,J.C., Crabb,B.S. and Cowman,A.F. (2006) A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature*, **439**, 1004–1008.
6. Voss,T.S., Tonkin,C.J., Marty,A.J., Thompson,J.K., Healer,J., Crabb,B.S. and Cowman,A.F. (2007) Alterations in local chromatin environment are involved in silencing and activation of subtelomeric var genes in Plasmodium falciparum. *Mol. Microbiol.*, **66**, 139–150.
7. Frank,M. and Deitsch,K. (2006) Activation, silencing and mutually exclusive expression within the var gene family of Plasmodium falciparum. *Int. J. Parasitol.*, **36**, 975–985.
8. Horn,D. and Barry,J.D. (2005) The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Res.*, **13**, 525–533.
9. Taylor,J.E. and Rudenko,G. (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet.*, **22**, 614–620.
10. Morrison,L.J., Marcello,L. and McCulloch,R. (2009) Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell Microbiol.*, **11**, 1724–1734.
11. Berriman,M., Hall,N., Sheader,K., Bringaud,F., Tiwari,B., Isobe,T., Bowman,S., Corton,C., Clark,L., Cross,G.A. *et al.* (2002) The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **122**, 131–140.
12. Hertz-Fowler,C., Figueiredo,L.M., Quail,M.A., Becker,M., Jackson,A., Bason,N., Brooks,K., Churcher,C., Fahkro,S., Goodhead,I. *et al.* (2008) Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One*, **3**, e3527.
13. Chaves,I., Rudenko,G., Dirks-Mulder,A., Cross,M. and Borst,P. (1999) Control of variant surface glycoprotein gene-expression sites in Trypanosoma brucei. *EMBO J.*, **18**, 4846–4855.
14. Rudenko,G., Blundell,P.A., Taylor,M.C., Kieft,R. and Borst,P. (1994) VSG gene expression site control in insect form Trypanosoma brucei. *EMBO J.*, **13**, 5470–5482.
15. Vanhamme,L., Poelvoorde,P., Pays,A., Tebabi,P., Van Xong,H. and Pays,E. (2000) Differential RNA elongation controls the variant surface glycoprotein gene expression sites of Trypanosoma brucei. *Mol. Microbiol.*, **36**, 328–340.
16. Navarro,M., Cross,G.A. and Wirtz,E. (1999) Trypanosoma brucei variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *EMBO J.*, **18**, 2265–2272.
17. Figueiredo,L.M. and Cross,G.A. (2010) Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes. *Eukaryot. Cell*, **9**, 148–154.
18. Stanne,T.M. and Rudenko,G. (2010) Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. *Eukaryot. Cell*, **9**, 136–147.
19. Hughes,K., Wand,M., Foulston,L., Young,R., Harley,K., Terry,S., Ersfeld,K. and Rudenko,G. (2007) A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *EMBO J.*, **26**, 2400–2410.
20. Yang,X., Figueiredo,L.M., Espinal,A., Okubo,E. and Li,B. (2009) RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. *Cell*, **137**, 99–109.
21. Figueiredo,L.M., Janzen,C.J. and Cross,G.A. (2008) A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biol.*, **6**, e161.
22. 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.
23. 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.
24. Sheader,K., te Vruchte,D. and Rudenko,G. (2004) Bloodstream form-specific up-regulation of silent vsg expression sites and procyclin in Trypanosoma brucei after inhibition of DNA synthesis or DNA damage. *J. Biol. Chem.*, **279**, 13363–13374.
25. Oberholzer,M., Morand,S., Kunz,S. and Seebeck,T. (2006) A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. *Mol. Biochem. Parasitol.*, **145**, 117–120.
26. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
27. 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.
28. Wickstead,B., Ersfeld,K. and Gull,K. (2002) Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **125**, 211–216.
29. Ligtenberg,M.J., Bitter,W., Kieft,R., Steverding,D., Janssen,H., Calafat,J. and Borst,P. (1994) Reconstitution of a surface transferrin binding complex in insect form Trypanosoma brucei. *EMBO J.*, **13**, 2565–2573.
30. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbour Press, New York, USA.
31. Earnshaw,W.C., Honda,B.M., Laskey,R.A. and Thomas,J.O. (1980) Assembly of nucleosomes: the reaction involving X. laevis nucleoplasmin. *Cell*, **21**, 373–383.
32. Eirin-Lopez,J.M., Frehlick,L.J. and Ausio,J. (2006) Long-term evolution and functional diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Genetics*, **173**, 1835–1850.
33. Reeves,R. and Nissen,M.S. (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.*, **265**, 8573–8582.
34. Aravind,L. and Landsman,D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.*, **26**, 4413–4421.
35. 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.
36. Horn,D. and Cross,G.A. (1997) Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. *EMBO J.*, **16**, 7422–7431.
37. Urmenyi,T.P. and Van der Ploeg,L.H. (1995) PARP promoter-mediated activation of a VSG expression site promoter in insect form Trypanosoma brucei. *Nucleic Acids Res.*, **23**, 1010–1018.
38. Wickstead,B., Ersfeld,K. and Gull,K. (2004) The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. *Genome Res.*, **14**, 1014–1024.
39. Rudenko,G., Bishop,D., Gottesdiener,K. and Van der Ploeg,L.H. (1989) Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form Trypanosoma brucei. *EMBO J.*, **8**, 4259–4263.
40. 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.
41. Aksoy,S., Williams,S., Chang,S. and Richards,F.F. (1990) SLACS retrotransposon from Trypanosoma brucei gambiense is similar to mammalian LINEs. *Nucleic Acids Res.*, **18**, 785–792.
42. Song,H., Spichiger-Haeusermann,C. and Basler,K. (2009) The ISWI-containing NURF complex regulates the output of the canonical Wingless pathway. *EMBO Rep.*, **10**, 1140–1146.
43. Liu,Y.I., Chang,M.V., Li,H.E., Barolo,S., Chang,J.L., Blauwkamp,T.A. and Cadigan,K.M. (2008) The chromatin remodelers ISWI and ACF1 directly repress Wingless transcriptional targets. *Dev. Biol.*, **323**, 41–52.
44. Aitcheson,N., Talbot,S., Shapiro,J., Hughes,K., Adkin,C., Butt,T., Sheader,K. and Rudenko,G. (2005) VSG switching in Trypanosoma brucei: antigenic variation analysed using RNAi in the absence of immune selection. *Mol. Microbiol.*, **57**, 1608–1622.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011

45. Eitoku,M., Sato,L., Senda,T. and Horikoshi,M. (2008) Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell. Mol. Life Sci.*, **65**, 414–444.
46. Frehlick,L.J., Eirin-Lopez,J.M. and Ausio,J. (2007) New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays*, **29**, 49–59.
47. Philpott,A., Leno,G.H. and Laskey,R.A. (1991) Sperm decondensation in Xenopus egg cytoplasm is mediated by nucleoplasmin. *Cell*, **65**, 569–578.
48. Laskey,R.A., Mills,A.D., Philpott,A., Leno,G.H., Dilworth,S.M. and Dingwall,C. (1993) The role of nucleoplasmin in chromatin assembly and disassembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **339**, 263–269; discussion 268–269.
49. Lu,Z., Zhang,C. and Zhai,Z. (2005) Nucleoplasmin regulates chromatin condensation during apoptosis. *Proc. Natl Acad. Sci. USA*, **102**, 2778–2783.
50. Ito,T., Tyler,J.K., Bulger,M., Kobayashi,R. and Kadonaga,J.T. (1996) ATP-facilitated chromatin assembly with a nucleoplasmin-like protein from Drosophila melanogaster. *J. Biol. Chem.*, **271**, 25041–25048.
51. Taneva,S.G., Banuelos,S., Falces,J., Arregi,I., Muga,A., Konarev,P.V., Svergun,D.I., Velazquez-Campoy,A. and Urbaneja,M.A. (2009) A mechanism for histone chaperoning activity of nucleoplasmin: thermodynamic and structural models. *J. Mol. Biol.*, **393**, 448–463.
52. Mellor,J. and Morillon,A. (2004) ISWI complexes in Saccharomyces cerevisiae. *Biochim. Biophys. Acta*, **1677**, 100–112.
53. Horn,D. (2009) Antigenic variation: extending the reach of telomeric silencing. *Curr. Biol.*, **19**, R496–498.
54. Glover,L. and Horn,D. (2006) Repression of polymerase I-mediated gene expression at Trypanosoma brucei telomeres. *EMBO Rep.*, **7**, 93–99.
55. Alsford,S., Kawahara,T., Isamah,C. and Horn,D. (2007) A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation. *Mol. Microbiol.*, **63**, 724–736.

Downloaded from nar.oxfordjournals.org at Imperial College London on January 25, 2011