2009

Regulation of Gene Expression in Trypanosoma brucei

Tim Nicolai Siegel

Regulation of Gene Expression in *Trypanosoma brucei*

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Tim Nicolai Siegel

June 2009

Regulation of Gene Expression in *Trypanosoma brucei*

Tim Nicolai Siegel, Ph.D.
The Rockefeller University 2009

The protozoan parasite *Trypanosoma brucei* is one of the most divergent well-studied eukaryotes and many discoveries of general interest have been made in *T. brucei*.

Atypical for a eukaryote, in *T. brucei* genes transcribed by RNA polymerase II (RNA pol II) are arranged in polycistronic transcription units (PTUs). mRNAs are separated post-transcriptionally by coupled splicing and polyadenylation reactions. During the splicing reaction a 39 nt 'sliced-leader' is added to every mRNA, a process termed *trans* splicing. The arrangement of generally unrelated genes in PTUs has lead to the assumption that little gene regulation occurs at the level of transcription initiation but that gene expression is regulated at the levels of mRNA maturation and stability. To better understand gene regulation in *T. brucei*, I performed a systematic analysis to correlate *trans*-splicing efficiency with specific DNA sequence motifs and observed large variations in *trans*-splicing efficiency depending on the DNA sequence upstream of the 3´ splice site.

With one exception, no RNA pol II promoter motif has been identified in *T. brucei*, and how transcription is initiated remains an enigma. The second part of my thesis was based on the assumption that, given the apparent lack of RNA pol II promoter motifs, RNA pol II transcription start sites (TSS) are marked by distinct chromatin structures that facilitate recruitment of the transcription machinery. *T. brucei* has four histone variants: H2AZ, H2BV, H3V and H4V. Using ChIP-seq to examine the genome-wide distribution of chromatin components, I showed that H2AZ, H2BV, the K10-Ac form of H4, and the bromodomain factor BDF3 are significantly enriched at probable RNA pol II TSSs and used this mark to identify more than 60 previously unanticipated TSS candidates. Co-IP experiments with tagged H2A, H2AZ, H2B, and H2BV indicated that less histone H3 and histone H4 co-immunoprecipitated with variant histones than with core histones suggesting that variant nucleosomes are less stable than canonical nucleosomes. Apparently unique to trypanosomes, additional histone variants H3V and H4V are enriched at probable RNA pol II transcription termination sites. These findings suggest that histone modifications and histone variants play crucial roles in transcription initiation and termination in trypanosomes and that destabilization of nucleosomes by histone variants is an evolutionarily ancient and general mechanism of transcription initiation, demonstrated in an organism in which general RNA pol II transcription factors have been elusive.

to B. M. S and A. W.

Acknowledgements

First and foremost, I want to thank Dr. George Cross for giving me the opportunity to work in his laboratory when I contacted him as an undergraduate with no laboratory experience in molecular biology. He has been a great mentor, providing freedom that allowed me to learn to think and work independently, yet always accessible and helpful when I needed guidance. From him I learned to design experimental strategies and to critically evaluate the published literature.

I am grateful to my faculty advisory committee consisting of Drs. Titia de Lange, David Allis and Michael Rout for their time, constant encouragement, helpful suggestions and constructive criticism. It has been a great honor to have leaders in their fields show interest in my work and share their experience. I also want to thank Dr. Stephen Beverley for traveling to New York and serving as the external member on my thesis committee.

I thank Dr. David Horn for a fruitful collaboration and warm hospitality during my stay at his laboratory at the London School of Hygiene and Tropical Medicine. Drs. Jeffrey deGrasse and Doeke Hekstra have been great collaborators and friends at The Rockefeller University.

I am thankful to the many great colleagues I had during my time in the Cross laboratory. Without their friendship, advice and encouragement I would not have succeeded in completing my work. Drs.

Kevin Tan, Simone Leal, Bibo Li, Joanna Lowell and especially Christian Janzen and Luisa Figueiredo have all been great mentors to me and I feel indebted for all they have done for me – too much to be listed on these pages. I thank Michael Scahill, Jenny Li and Dongmyung Ahn for being great friends and for all the little and big things they have helped me with. In addition, I thank Louise Kemp for working with me and assisting in generating the many cell lines needed for my project.

I want to thank the laboratories of Drs. David Allis, Michael Rout, Brian Chait and especially of Robert Darnell for sharing reagents, equipment and great technical advice. In particular, I want to acknowledge the help of Jeff Smith, Aldo Mele, John Fak and Dr. Ileana Cristea.

I am grateful to Drs. Alison North, Svetlana Mazel, Connie Zhao and Haiteng Deng and their staff for running wonderful resource centers and providing me with critical help at many stages of my project.

Drs. Scott Dewell, Xuning Wang and David Fenyo have been instrumental in my efforts to establish ChIP-seq technology for trypanosomes.

I want to thank Drs. Keith Gull, Arthur Günzl, Nina Papavasiliou, Michael Rout and Christian Janzen for providing antibodies.

Throughout my thesis work Umut Sahin and Doeke Hekstra have been great friends and constant sources of helpful ideas.

I am also thankful for all the other good friends I met at The Rockefeller University and outside and for the great time we spent together in New York. Additionally, I thank Andrea Procko for proofreading this thesis.

I thank the Deans Office for their financial support and personal guidance throughout my years at The Rockefeller University. The level of support and dedication to make life for us students as easy as possible was truly remarkable. I am thankful for additional financial support provided by the Boehringer Ingelheim Fonds.

Last but not least, I thank my family and my girlfriend Annette for their constant support. They have always been there for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3´SS	3´ splice sites
5´SS	5´ splice sites
AF	auxiliary factor
ASF1	anti-silencing function 1
ATP	adenosinetriphosphate
BB2	monoclonal antibody specific for the Ty1 peptide
BDF	bromodomain factor
BF	bloodstream forms
BLAT	blast like alignment tool
BSA	bovine albumin serum
CAF-1	chromatin assembly factor 1
CAF1	ccr4-associated factor 1
CENP-A	centromeric protein A
CF	cytoplasmic fraction
CFA	Complete Freund's Adjuvant
ChIP	chromatin immunoprecipitation
CHIP-seq	chromatin immunoprecipitation-sequencing
chromo	chromatin organization modifier
co-IP	co-immunoprecipitation
CPRG	chlorophenol red-β-D-galactopyranoside
DAPI	4,6-diamidino-2-phenylindole
DCO	DyeCycle Orange, Invitrogen
DMSO	dimethyl sulfoxide
Dot1	disruptor of telomeric silencing, a type of KMT
DSB	double stranded break
DTT	DL-1,4-dithiothreitol
Elp3	elongation protein 3, a type of KAT
ES	expression site
FACS	fluorescence activated cell sorting
Gcn5	general control nonrepressed, a type of KAT
GFP	green fluorescent protein
GPI	glycosylphosphatidylinisotol
H2A	histone H2A
H2AX	a variant form of H2A
H2AZ	a variant form of H2A
H2B	histone H2B
H2BV	histone H2B variant
H3	histone H3
H3.3	a variant form of H3
H3K4me3	H3 tri-methylated at lysine 4

H3K76me2	H3 di-methylated at lysine 76
H3K76me3	H3 tri-methylated at lysine 76
H3K79me2	H3 di-methylated at lysine 79
H3K79me3	H3 tri-methylated at lysine 79
H3V	histone H3 variant
H4	histone H4
H4K10-unmodified	H4 unmodified at lysine 10
H4K17me3	H4 tri-methylated at lysine 17
H4K4	H4 lysine 4
H4K4-unmodified	H4 unmodified at lysine 4
H4K4ac	H4 acetylated at lysine 4
H4S6	H4 serine 6
H4V	histone H4 variant
HAT1–3	trypanosome histone acetyltransferases 1-3
Hda1	histone deacetylase 1
HDAC	histone deacetylase
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
HF buffer	high fidelity buffer, NEB
HP1	heterochromatin protein 1
HPLC	high performance liquid chromatography
HTZ-1	a variant form of H2A, H2AZ ortholog in *S. cerevisiae*
IF	immunofluorescence
IFA	Incomplete Freund's Adjuvant
Ig	immunoglobuline
INO80	inositol-requiring protein 80
IP	immunoprecipitation
ISWI	imitation SWI2
KAT	lysine acetyltransferase
KDM	lysine demethylase
kDNA	kinetoplast DNA
KMT	lysine methyltransferase
MLL1	mixed-lineage leukaemia 1, a type of KMT
MOF	males-absent-on-the-first, a type of KAT
MYST	type of KAT named after the founding members of its group (MOZ, YBF2/SAS3, SAS2 and TIP60
nDNA	nuclear DNA
ORF	open reading frame
PARP	poly(A)-binding protein
PBS	phosphate-buffered saline
PCAF	P300/CBP-associated factor

PCR	polymerase chain reaction
PF	procyclic forms
PGK	phosphoglycerate kinase
PMSF	phenylmethylsulfonyl fluoride
RNA pol I	RNA polymerase I
RNA pol II	RNA polymerase II
RNA pol III	RNA polymerase III
poly(A)	polyadenine
poly(Y)	polypyrimidine
PRMT	arginine methyltransferase
PTM	post-translational modifications
PTU	polycistronic transcription unit
PVP	polyvinylpyrrolidone
RAD	RecA homolog
RFI	refractive index
RNA	ribonucleic acid
RNAi	RNA-interference
Rpd3	reduced potassium dependency 3, a type of HDAC
RSC	chromatin structure-remodeling complex
RT	room temperature
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sir2	silent information regulator 2, a type of HDAC
Sir2rp	Sir2-related protein
SL	spliced leader
SLRNA	spliced leader RNA
snRNP	small nuclear ribonucleoproteins
SR-rich domains	serine/arginine-rich domains
SSR	strand-switch region
SWI/SNF	mating type SWItching / Sucrose Non Fermentation
TAFII250	TATA box binding protein (TBP)-associated factor, 250kDa
TDB	trypanosome dilution buffer
Tritryp	term referring to *T. brucei*, *T. cruzi* and *L. major*
TSA	trichostatin A
TSS	transcription start site
UR	upstream region, the region upstream of the 5´UTR containing a poly(Y) tract and branch site
UTR	untranslated region
VSG	variant surface glycoprotein
WCL	whole cell lysate
WHO	world health organization

Chapter 1: General Introduction

Summary

Trypanosoma brucei is a protozoan parasite that branched early in evolution from the eukaryotic lineage (Fernandes et al., 1993). *T. brucei* causes African trypanosomiasis, commonly referred to as sleeping sickness in humans and nagana in livestock. Soon after the parasite was discovered more than hundred years ago by Surgeon-Captain David Bruce (Bruce, 1895), its great evolutionary distance from other organisms and its many unique solutions to 'biological problems' started to fascinate researchers. Decades of research have made *T. brucei* an important model organism that has led to discoveries like GPI-mediated anchoring of proteins (Ferguson et al., 1988), RNA-editing (Benne et al., 1986), and *trans* splicing (Parsons et al., 1984; De Lange et al., 1984). Yet, many aspects of trypanosome biology are still not well understood, including the regulation of gene expression – the topic of my Ph.D. thesis.

This introduction contains three parts – a general background about *T. brucei*, an overview of the different processes that are thought to contribute to the regulation of gene expression in *T. brucei* and a brief summary of our current knowledge about the role of chromatin in transcription regulation.

Trypanosoma brucei: general overview

African Trypanosomiasis

African trypanosomiasis, responsible for ~50,000 human deaths annually (WHO, 2004), is a vector-borne disease caused by different sub-species of trypanosomes. The two human infective sub-species are *Trypanosoma brucei gambiense*, found in western and central Africa and responsible for chronic infections, and *Trypanosoma brucei rhodesiense*, found in eastern and southern Africa and responsible for a rapidly progressive, acute form of the disease. A third subspecies, *Trypanosoma brucei brucei*, is pathogenic only to animals and is the subspecies most commonly used in laboratory settings.

Transmission of the disease is restricted to regions where its vector, the tsetse (*Glossina* sp.), is endemic and occurs almost exclusively in Sub-Saharan Africa. Parasite reservoirs are wild game animals in East Africa and infected humans and domestic cattle in West Africa. Upon transmission of the parasite from an infected tsetse to a human host, the parasite accumulates in the lymphatic system and the capillary beds. As the infection progresses, parasites cross the blood-brain barrier and migrate to the central nervous system where they lead to confusion, poor coordination and a disturbance of the sleep-wake cycle, thus the sleeping sickness. If left untreated, African trypanosomiasis is generally lethal. Despite decades of research, no vaccine has been developed, and

effective drugs only exist against the early stages of the disease before the parasite has crossed the blood-brain barrier.

The WHO estimates that in 1986 some 70 million people lived in areas where infection could take place, implying that the actual number of infected individual may be much higher than current data suggest. Therefore, a better understanding of the parasite's molecular biology, especially its mechanism of immune evasion, is of great medical importance.

Antigenic Variation in the Bloodstream Form Trypanosoma

Because *T. brucei* is an extracellular parasite, it has to evade the host immune system constantly. To do so, the parasite takes advantage of its capacity to change its surface coat. This infamous ability to undergo antigenic variation leads to characteristic remitting and relapsing waves of high parasitemia (see Figure 1.1) (Ross and Thomson, 1910), eventually exhausting the immune system of its mammalian host. Antigenic variation has long been one of the most intensively studied phenomena of *T. brucei*, but the details of its molecular regulation remain unsolved (Cross, 1996; Barry and McCulloch, 2001).

Each *T. brucei* cell has a surface coat (Figure 1.2) consisting of ~10 million identical copies of a specific variant surface glycoprotein (VSG) (Vickerman, 1969; Cross, 1975). The *T. brucei* genome contains several hundred different *VSG* genes (Van der Ploeg et al., 1982), but only one

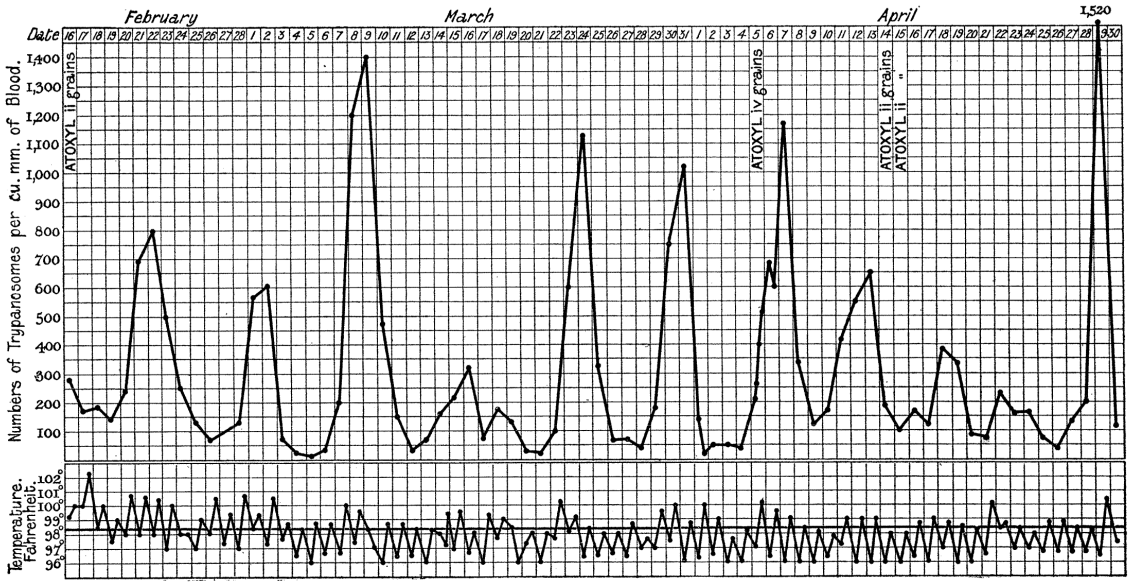


Figure 1.1. Sleeping sickness is characterized by a wave-like pattern of parasitemia

Levels of parasitemia (top panel) and body temperature (lower panel) are shown for a 26-year-old patient from Northumberland, UK, who was infected with trypanosomes in N. E. Rhodesia (today Zambia) in September 1909. The accurate determination of parasite counts over a span of ~80 days (x-axis) revealed a well-defined periodicity of alternating high and low levels of parasitemia (y-axis). Figure adapted from Ross and Thomson, 1910.

VSG gene is expressed at any time, from one specific transcription unit called an expression site (ES) (Johnson et al., 1987). Different *T. brucei* strains contain between 9 and 15 ESs, which are ~40–60 kb polycistronic transcription units (Hertz-Fowler et al., 2008) and are always located adjacent to a telomere (de Lange and Borst, 1982). Only one ES is active at any time while all others are repressed by an unknown mechanism. The restriction of ESs to sub-telomeric regions and the observation of a telomere position effect-like repression of

subtelomeric genes support the hypothesis that telomere-associated proteins and a distinct chromatin structure play a role in either ES activation or repression (Yang et al., 2009; Horn and Cross, 1995; Horn and Cross, 1997).

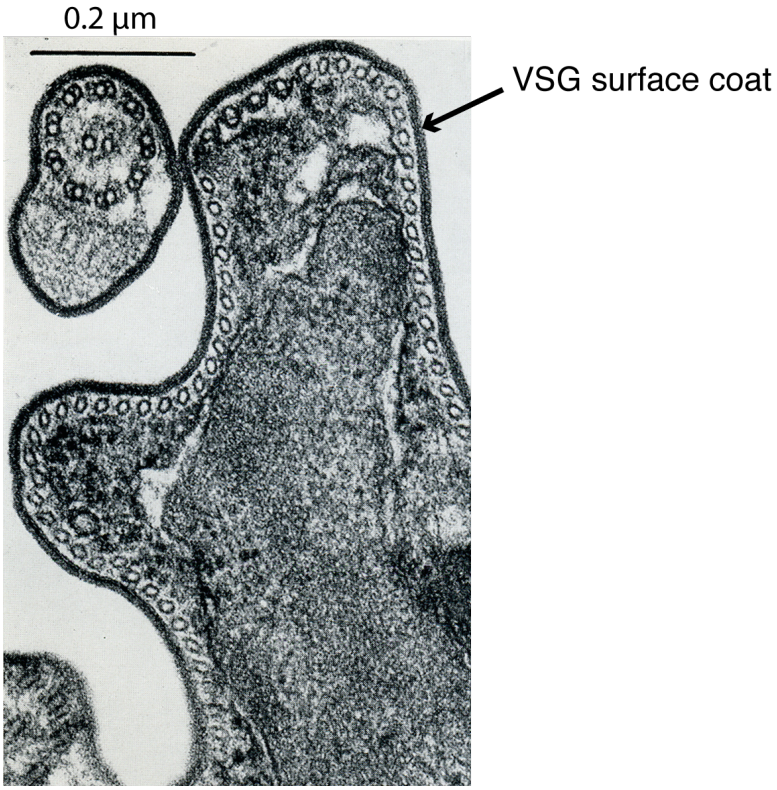


Figure 1.2. African trypanosomes are covered by a dense surface coat
Electron microscopy revealed a dense coat covering the outer surface of *T. brucei*. Image taken from Cross, 1975.

The *Trypanosoma brucei* life cycle

Like other vector-transmitted parasites, *T. brucei* exists in several life-cycle stages (Figure 1.3) (reviewed in Matthews et al., 2004).

With the bite of an infected tsetse, non-dividing metacyclic forms are injected into the bloodstream of the mammalian host. To establish a

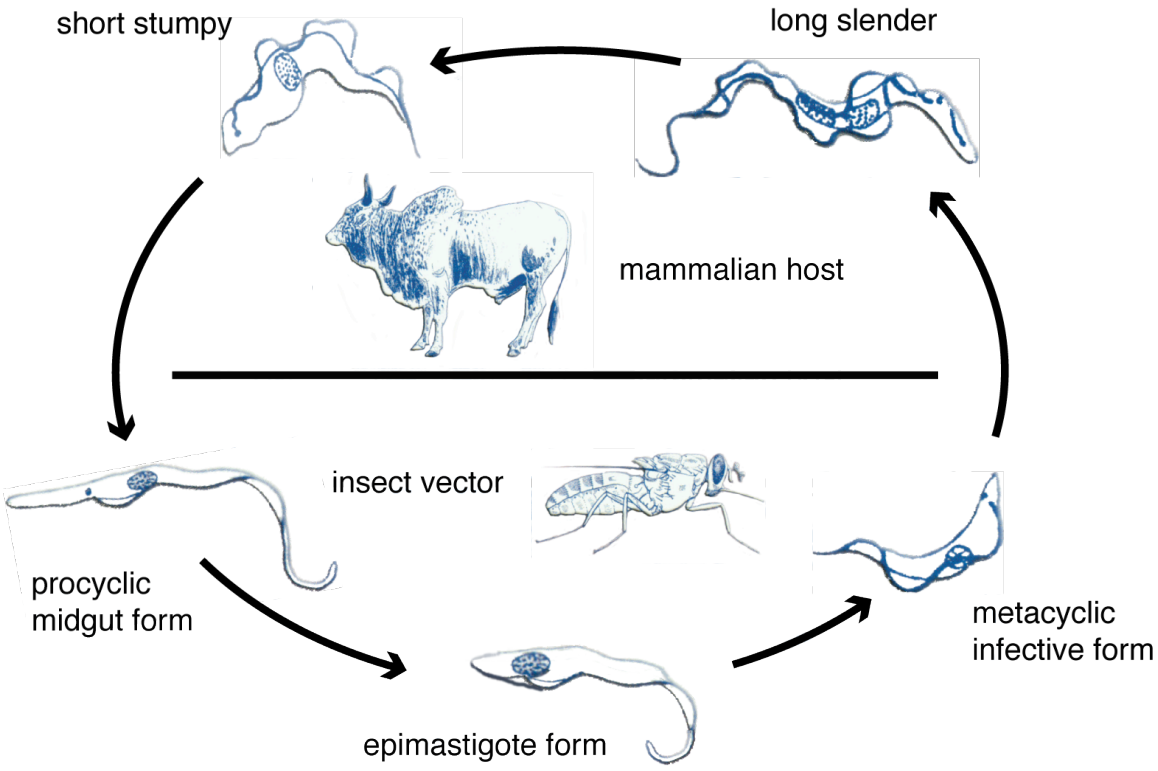


Figure 1.3. The life cycle of *T. brucei*

During a blood meal the tsetse injects the infective metacyclic form into the bloodstream of the mammalian host. Inside the bloodstream, the parasite differentiates into the prolific long slender form, a process that is accompanied by significant morphological changes and the expression of a characteristic surface coat. Over time a pleomorphic population develops containing some short stumpy bloodstream forms that are pre-adapted for uptake by the tsetse. Inside the vector, the parasite again undergoes significant morphological changes and differentiates into the parasite-specific procyclic form. After migrating to the salivary glands, the parasite differentiates into the dividing epimastigote form and further into the non-dividing, infective metacyclic form.

stable infection in the mammalian bloodstream, the parasites re-enter the cell cycle, begin utilizing a larger repertoire of surface antigens and undergo significant morphological changes. Eventually, the bloodstream population of parasites becomes pleomorphic, consisting of proliferating slender bloodstream forms and non-proliferating stumpy forms. Once stumpy forms are present, the population is pre-adapted for uptake by the insect vector. Upon uptake by the tsetse, stumpy forms differentiate into the so-called procyclic form. This differentiation is accompanied by a very rapid loss of the bloodstream form-specific VSG surface coat, which is replaced by insect form-specific surface antigens called procyclins. The life-cycle-specific replacement of surface antigens is followed by extensive morphological changes, induced by remodeling of the cytoskeleton, repositioning of the kinetoplast (the mitochondrial DNA of trypanosomes) and changes to the cell metabolism. Whereas bloodstream forms, living in the glucose rich environment of the bloodstream, solely depend on glycolysis to generate energy, procyclic forms utilize enzymes of the citric acid cycle and oxidative phosphorylation to generate adenosine triphosphate. Procyclic forms grow and replicate in the midgut of the tsetse until they pass through the midgut wall and on to the salivary glands. There they differentiate into dividing epimastigote forms and further to non-dividing, infective metacyclic forms. Metacyclic forms express a limited set of surface antigens and can, once passed on to a

mammalian host, differentiate into proliferative slender bloodstream forms.

Both the slender bloodstream form and procyclic form can be cultured and readily studied in the laboratory.

Regulation of gene expression in *T. brucei*

Genes are arranged in polycistronic transcription units

Transcription of protein-coding genes in *T. brucei* differs in two important aspects from other eukaryotes. First, transcription is polycistronic (Tschudi and Ullu, 1988) – arrays of numerous genes are transcribed in polycistronic transcription units (PTUs). Second, mRNAs are separated post-transcriptionally by coupled *trans*-splicing and polyadenylation reactions (LeBowitz et al., 1993; Matthews et al., 1994).

The organization of protein-encoding genes into polycistronic transcription units is reminiscent of operons in prokaryotes except that there is no evidence to suggest clustering of functionally related genes in *T. brucei*. While shorter PTUs have also been identified in nematodes (Spieth et al., 1993) and di-cistronic units have been reported for *Drosophila* (Brogna and Ashburner, 1997) and humans (Lee, 1991), the relevance of polycistronic transcription is not known. Within a PTU, genes are transcribed from the same strand, but transcription of two neighboring PTUs can either be convergent or divergent. The regions

between PTUs are referred to as strand-switch regions (SSRs) (see Figure 1.4).

Despite intensive work by numerous laboratories, no RNA polymerase II (RNA pol II) promoter for protein-encoding genes has been identified; thus, it is not known how long PTUs are or where they begin and end. Strand-specific nuclear run-on assays performed in *Leishmania* (Martinez-Calvillo et al., 2003), a genus related to *T. brucei*, have shown that RNA pol II transcription starts at SSRs between two transcriptionally divergent PTUs (divergent-SSRs) and ends at SSRs between two transcriptionally convergent PTUs (convergent-SSRs). Because 75% of all *L. major* genes can be found in the same genomic context in *T. brucei* (El-Sayed et al., 2005b), indicating a high degree of synteny, it is reasonable to hypothesize that divergent-SSRs in *T. brucei* are transcription start sites (Figure 1.4).

Tritryp genomes

The unusual organization of genes in PTUs was confirmed on a genome-wide scale in 2005 with the publication of the genome of three Trypanosomatid species – *T. brucei*, *Trypanosoma cruzi* (incomplete) and *Leishmania major*, collectively referred to as tritryps (Berriman et al., 2005; El-Sayed et al., 2005a; El-Sayed et al., 2005b; Ivens et al., 2005). Organization of genes in PTUs allows for very little regulation at the level of transcription initiation, yet, comparing cDNA levels from different life

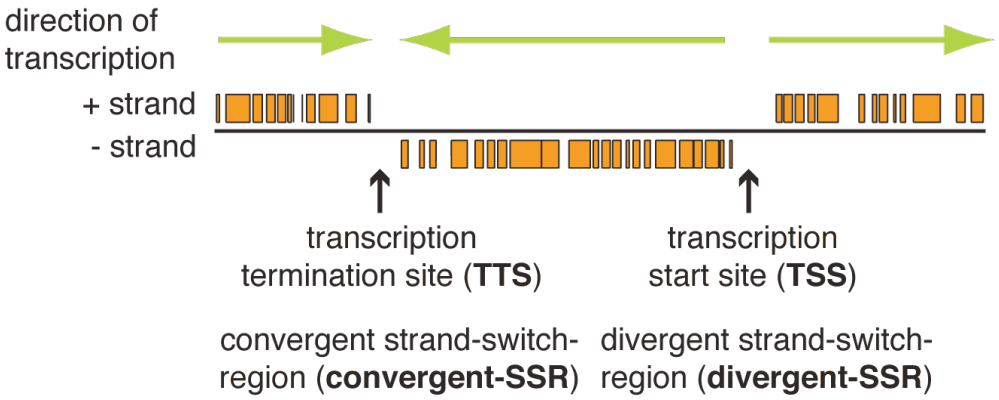


Figure 1.4. Genes are arranged in polycistronic transcription units
Divergent strand switch regions (divergent-SSRs) contain probable RNA pol II transcription start sites (TSSs). Convergent-SSRs contain probable RNA pol II transcription termination sites (TTSs). Orange boxes represent open reading frames and green arrows indicate the direction of transcription.

cycle stages, it was found that ~14% of genes are differentially expressed (El-Sayed et al., 2000). Therefore, it is generally assumed that regulation occurs post-transcriptionally at the levels of *trans* splicing/ polyadenylation, RNA export, RNA degradation, protein translation and protein stability (Clayton, 2002). Analysis of the tritryp genomes supported this assumption. For example, the *T. brucei* genome encodes 103 proteins with a CCCH-type zinc finger RNA-binding domain. Such RNA-binding proteins could very well play a role in RNA stability, RNA export, etc. Usually two such domains are required for RNA binding, but the fact that the tritryp genomes each encode for ~40 proteins containing a single RNA recognition motif compared to only 12 in *Schizosaccharomyces pombe* suggests the presence of unique molecular mechanisms of post-transcriptional control (Ivens et al., 2005).

As mentioned above, the tritryp genomes also revealed large-scale synteny among the three genera that has greatly facilitated the construction of genome maps. Synteny was particularly evident for a set of 6158 'core genes', genes containing orthologs in genomes of all three pathogens. Of those genes, 94 % were found in the same genetic context in *T. brucei* and *L. major*. The smaller number of chromosomes in *T. brucei* compared to *L. major* (see Table 1.1) is probably due to chromosome fusions because 20 of the 36 *L. major* chromosomes are almost completely syntenic within the *T. brucei* genome (El-Sayed et al., 2005b).

Table 1.1. General features of the tritryp genomes

	T. brucei	*T. cruzi*	*L. major*	*S. cerevisiae*	*D. melanogaster*	*H. sapiens*
Haploid genome size in Mb	35	55	33	12	165	3,000
Number of megabase-sized chromosomes (per haploid genome)	11	~28*	36	16	4	23
Number of intermediate-sized chromosomes	5–6	N/A	N/A	N/A	N/A	N/A
Number of mini-sized chromosomes	~100	N/A	N/A	N/A	N/A	N/A
Number of genes, excluding pseudogenes (per haploid genome)	8,164	~12,000*	8277	5,770	~14,000	20,000-25,000

*The *T. cruzi* genome has not been fully assembled and the number of genes can only be estimated.

The next sections summarize the current knowledge about two mechanisms implicated in post-transcriptional gene regulation – *trans* splicing/polyadenylation and RNA stability.

Trans splicing and polyadenylation

To generate mature mRNA, newly transcribed RNA is *trans*-spliced and polyadenylated. *Trans* splicing serves two functions: it dissects mRNAs from polycistronic primary transcripts and adds a 39-nt mini-exon, also called the spliced leader (SL), to form a cap structure for the mRNA (Figure 1.5). The unusual cap structure at the 5´ end of the spliced leader RNA is referred to as 'cap 4' because the 4 nt after the 7-methylguanosine are methylated (Perry et al., 1987; Freistadt et al., 1987; Freistadt et al., 1988; Bangs et al., 1992). Because the capped precursor spliced leader RNA (SLRNA) is transcribed independently from the polycistronic primary transcript, this process of joining two independently transcribed exons is referred to as *trans* splicing, in contrast to conventional *cis* splicing during which an internal RNA fragment is removed from one precursor transcript (Campbell et al., 1984; Milhausen et al., 1984). *Trans* and *cis* splicing, however, share remarkable similarities. Both require the same characteristic sequence motifs – a polypyrimidine tract [poly(Y) tract], a GT dinucleotide at the 5´ splice site (5´SS), an AG at the 3´ splice site (3´SS) and, possibly, exonic enhancer motifs (Huang and Van der Ploeg, 1991; Lopez-Estrano et al.,

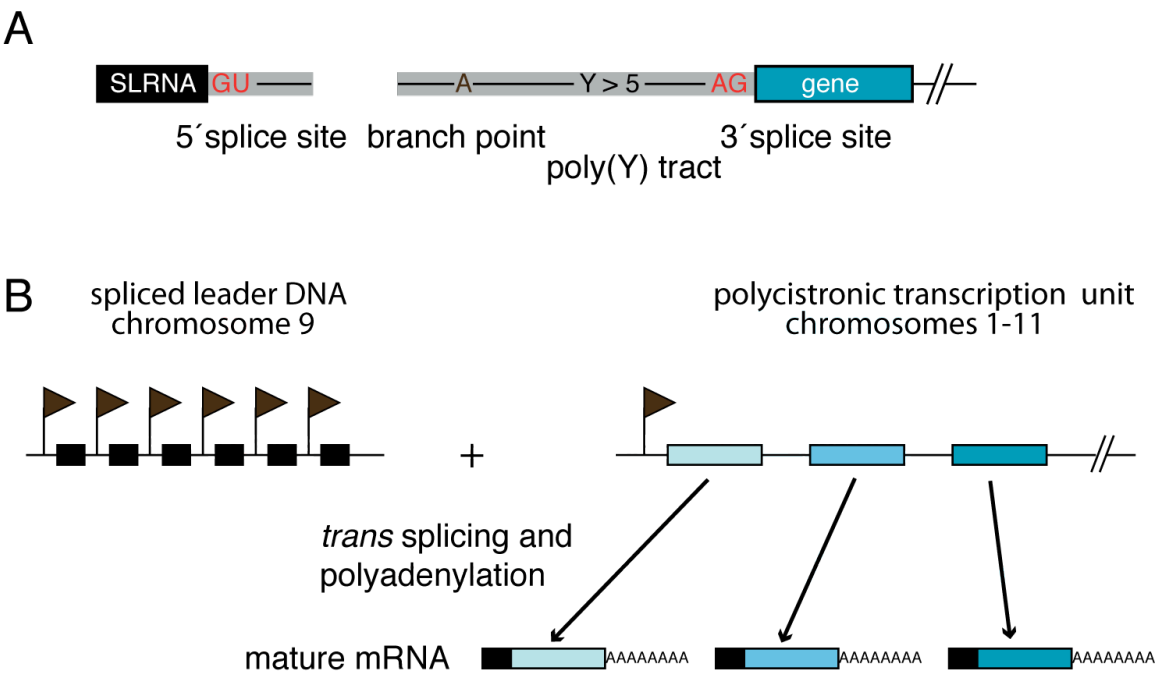


Figure 1.5. Polycistronic transcription units are separated by *trans*-splicing reactions

(A) Sequence elements required for *trans* splicing. Black and blue boxes mark a spliced leader and a protein-encoding gene, respectively that are joined by a *trans*-splicing reaction. Grey boxes indicate intronic sequences containing regulatory elements – 5´ splice site, branch point, polypyrimidine [poly(Y)] tract and 3´ splice site. Conserved nucleotides are marked in red.

(B) Spliced leader DNA and protein-encoding genes are located in separate genomic locations. Long fragments of polycistronic mRNA are separated post-transcriptionally by a coupled *trans*-splicing and polyadenylation reaction. During *trans* splicing and polyadenylation a 39-nt spliced-leader RNA and a polyadenine (polyA) tail of variable length are added to the 5´ and 3´ end of every mRNA molecule, respectively. Flags represent spliced leader promoters (left panel) and probable RNA pol II transcription start sites (right panel).

1998; Patzelt et al., 1989). Both *trans* and *cis* splicing follow the same general mechanism, which involves two consecutive catalytic *trans*-

esterification reactions. Most major components of the yeast and human spliceosomes are conserved in *T. brucei* (reviewed in Liang et al., 2003).

By inducing a series of block mutations, it has been shown that both a poly(Y) tract upstream of the splice site and sequence elements in the 5´ untranslated region (UTR) are important for efficient *trans* splicing (Huang and Van der Ploeg, 1991). Because *trans* splicing is a prerequisite for protein expression, differences in the poly(Y)-rich tract that lead to differences in *trans*-splicing efficiency can be reflected by differences in protein levels. It has been suggested that, by regulating *trans*-splicing efficiency, trypanosomes could fine tune gene expression post-transcriptionally (Huang and Van der Ploeg, 1991). However, no extensive systematic study has been performed to determine minimal and optimal sequence motifs required for efficient *trans* splicing. As part of my thesis work to elucidate factors involved in post-transcriptional gene regulation, I systematically measured the effect of poly(Y) sequence elements on *trans*-splicing efficiency. The goals of this study were to identify motifs required for efficient *trans* splicing and to test the hypothesis that inefficient *trans* splicing correlates with low levels of mRNA (see Chapter 6).

Polyadenylation refers to the addition of long stretches of adenines, the so-called poly(A) tail, to the 3´ end of the 3´ UTR. Poly(A) tails play an important role in conferring RNA stability as they provide a binding site for poly(A)-binding proteins (PABPs). PABPs are involved in preventing

decapping of the 5´ end of mRNA (reviewed in Wilusz et al., 2001), one of the early steps in mRNA degradation (see next section). Experiments in *T. brucei* show that depletion of CAF1 (Ccr4-associated factor 1), an enzyme capable of degrading poly(A) tails, leads to increased mRNA stability (Schwede et al., 2008). Because *trans* splicing and polyadenylation are coupled, the same sequence motifs could control increased stability of one gene and more efficient *trans* splicing of the neighboring downstream gene in the same PTU (Matthews et al., 1994).

Role of 3´ UTR in RNA stability

While the exact mechanisms of post-transcriptional gene regulation in *T. brucei* are not known, the importance of RNA stability/ degradation has been established recently (Haanstra et al., 2008).

In yeast and other eukaryotes, mRNA degradation can be initiated by shortening of the poly(A) tail mediated by poly(A)-specific exoribonucleases. Poly(A) tail shortening is followed by decapping of the 5´ end of the mRNA and exonuclease-mediated RNA digestion in the 5´ → 3´ or the 3´ → 5´ direction (reviewed in Wilusz and Wilusz, 2004). The rate of mRNA degradation can be greatly influenced by sequence motifs in the 3´ UTR, which function as binding sites for components of the mRNA degradation machinery or factors contributing to increased RNA stability (Zubiaga et al., 1995; Yang et al., 2003).

To reconcile the apparent lack of transcriptional regulation in *T. brucei* with the large number of differentially expressed proteins, it has been proposed that regulation may occur at the level of mRNA stability. Indeed, sequence elements of 16 bp and 26 bp conferring mRNA stability in procyclic forms (PF) were identified in the 3´ UTR of different procyclin mRNAs, which are highly abundant in PF (Hehl et al., 1994; Hotz et al., 1997; Furger et al., 1997). Incorporation of these sequence motifs into the 3´ UTR of a reporter gene (chloramphenicol acetyltransferase, CAT) led to developmentally regulated expression of the CAT gene with low expression in bloodstream forms (Hotz et al., 1997). Subsequently, regulatory elements were identified in the 3´ UTRs of other RNA pol II transcribed genes in tritryps (Hotz et al., 1995; Di Noia et al., 2000; Boucher et al., 2002; Mayho et al., 2006). In a more systematic approach to identify life-cycle specific stabilizing motifs, Mayho et al. analyzed the 3´ UTRs of seven components of the cytochrome oxidase complex (Mayho et al., 2006). This complex is absent from bloodstream forms (BF), where glycolysis provides sufficient energy, but is upregulated upon differentiation of cells to PF, where oxidative phosphorylation takes place (Priest and Hajduk, 1994). Given the necessity to co-upregulate all ten components upon differentiation, one would expect the presence of conserved regulatory motifs in all transcripts. The authors were able to identify such regulatory motifs, including a motif resembling the 26-mer originally identified in the 3´ UTR of the procyclin transcript. An

extension of this approach to a genome-wide analysis of UTR sequences, as it has been performed in yeast and mammalian cells (Shalgi et al., 2005; Xie et al., 2005), could provide a better understanding of the role of RNA stability in the regulation of gene expression. However, a genome-wide analysis is currently not possible for two reasons – exact UTR length has only been determined for very few genes and genome-wide RNA levels are not known.

Chromatin

Eukaryotic DNA is densely coated with proteins making up a DNA/ protein structure called chromatin. The fundamental unit of chromatin is the nucleosome; 147 bp of DNA wrapped around an octamer of core histones, two each of histones H2A, H2B, H3 and H4 (Figure 1.6) (Olins and Olins, 1974; Kornberg, 1974).

The wrapping of DNA around nucleosomes leads to a 5–10-fold compaction of DNA (Felsenfeld and Groudine, 2003). Even denser packaging of DNA is achieved by folding of chromatin into distinct higher order structures that allow, for example, the ~2 m of human DNA to fit into a nucleus ~20 μm in diameter. Given the expansion of genome size during the course of evolution from prokaryotes (~4 Mb) to eukaryotes and then from single cellular organisms (*S. pombe*: 13.8 Mb; *T. brucei*: 30 Mb) to multi-cellular organisms (*Drosophila*: 165 Mb, human: 3.0 Gb), it was originally assumed that the primary role of chromatin was to pack

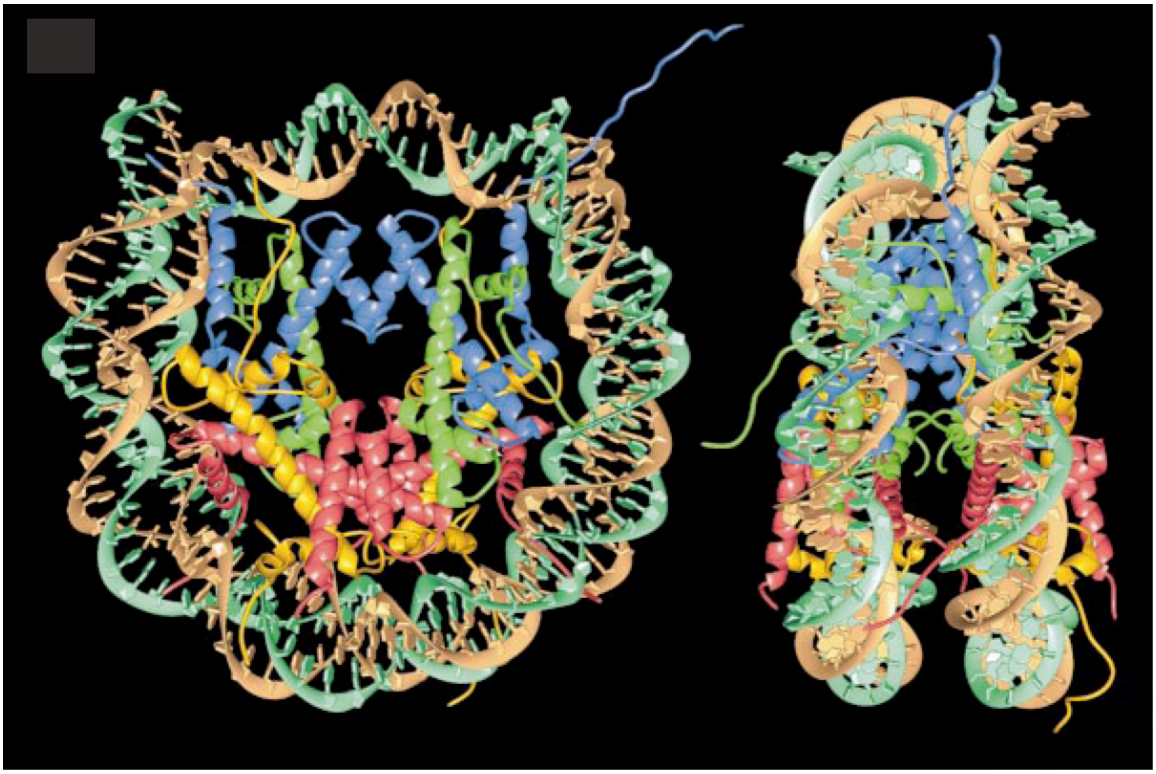


Figure 1.6. High-resolution crystal structure of a nucleosome
The image represents the first high-resolution (2.8 Å) crystal structure of a nucleosome. It shows 147 bp of DNA wound around an octamer of core histones, two each of histone H2A (yellow), H2B (red), H3 (blue) and H4 (green). The extruding N-terminal histone tails are highly flexible and their structure has not been solved. Figure adapted from Luger et al., 1997.

DNA densely into the nucleus. However, it was soon realized that chromatin plays vital roles in every process that requires access to DNA including replication, mitotic chromosome condensation, recombination, apoptosis, DNA repair and transcription. The general importance of histones probably led core histones to be among the most evolutionary conserved eukaryotic proteins. Histone H4 sequences from organisms as

divergent as humans and the most virulent malaria parasite, *Plasmodium falciparum,* differ in only one out of 102 amino acids.

Given the breadth of the chromatin field, in this introduction I will focus on factors thought to be involved in transcriptional activation in particular – histone variants, histone acetylation, enzymes responsible for addition and removal of histone acetylations and proteins binding to acetylated histones.

The role of chromatin in DNA transcription

Originally, different types of chromatin were identified by light microscopy studies of higher eukaryotic interphase nuclei (Heitz, 1928). All visibly condensed chromatin, darkly stained, has been called heterochromatin, and all other chromatin, much less condensed and more lightly stained, has been named euchromatin. Because gene transcription requires numerous factors of the transcriptional machinery to have direct access to DNA, coding sequences 'poised' for transcription should be found in euchromatin, the more 'open' chromatin conformation. Indeed, studies using nucleases to determine DNA accessibility have found gene regulatory sequences to be hypersensitive to DNase I, suggesting that they are relatively more accessible (Weintraub and Groudine, 1976; McGhee et al., 1981; Sabo et al., 2006).

The regulatory role of chromatin is mediated by two general, often overlapping pathways – rearrangement of chromatin conformation (*cis*-

mediated) and recruitment of chromatin-binding proteins (*trans*-mediated) (see Figure 1.7). *Cis*-mediated pathways lead to structural changes in the chromatin structure that may expose regulatory sequence elements of a promoter. Transcription factors could then bind the exposed DNA motifs and initiate assembly of the transcriptional machinery. In *trans*-mediated pathways, chromatin provides binding sites for transcriptional activators or repressors (reviewed in Kouzarides, 2007).

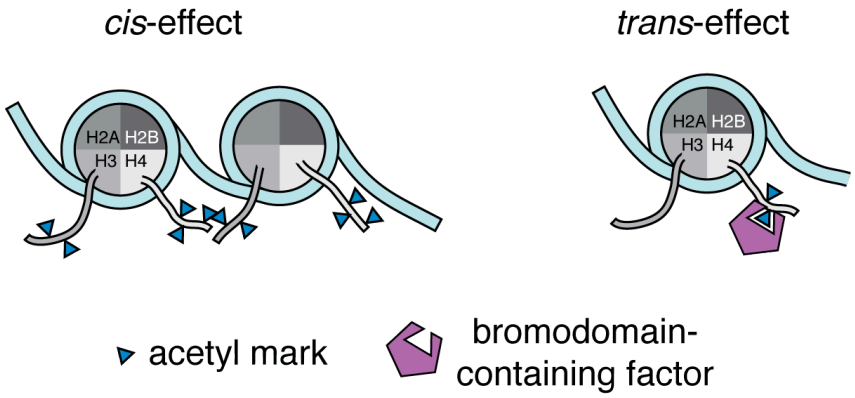


Figure 1.7. *Cis*- and *trans*-mediated effect of histone PTMs
Histone acetylations can lead to disruption of local chromatin condensation and thus contribute to exposure of regulator DNA motifs – *cis*-mediated pathway. Alternatively, histone acetylations and other histone PTMs can provide binding sites for effector proteins – *trans*-mediated pathway.

Several factors have been shown to contribute to an active chromatin conformation in *cis*-mediated pathways including the following:

DNA sequence can greatly influence its propensity to wrap around nucleosomes. Analysis of nucleosome-bound DNA revealed a strong ~10-bp periodicity of AA/TT/AT dinucleotides that oscillate in phase with each other but out of phase with GC dinucleotides and are thought to facilitate bending of DNA around nucleosomes (Satchwell et al., 1986; Widom, 2001). These findings suggest that DNA sequence can influence the positioning of nucleosomes along the genomic DNA and contributes to the depletion of nucleosomes in gene regulatory sequences. This hypothesis has been confirmed by genome-wide analyses of yeast DNA that indicate that gene regulatory sequences have a lower tendency to be associated with nucleosomes (Kaplan et al., 2009). It has long been known that removal of histones from nuclei leads to increased mRNA synthesis (Allfrey et al., 1963; Allfrey and Mirsky, 1962); thus, DNA sequence elements can influence transcription efficiency.

ATPase-driven *chromatin remodeling* complexes contribute to an open chromatin conformation in multiple ways (reviewed in Johnson et al., 2005; Cairns, 2007). First, sliding nucleosomes along DNA leads to exposure of DNA. Second, removal of the H2A-H2B dimers leads to destabilization of nucleosomes and exposure of DNA. Third, facilitating eviction of entire nucleosomes leads to

exposure of DNA. Based on differences in activity and composition, chromatin-remodeling complexes have been grouped in multiple families (SWI/SNF, ISWI, INO80 etc). The exact mode of function for these remodelers is unknown.

Histone variants may contribute to transcription activation by destabilizing nucleosomes at promoter proximal regions (see below: histone variants)

Post-translational histone modification, especially of the N-terminal tails of histones H3 and H4, can affect chromatin structure directly and lead to a more open chromatin conformation (see below: Molecular Mechanisms and Figure 1.7) (reviewed in Kouzarides, 2007).

Because the degree of chromatin compaction is highly dynamic and a particular gene may be found in open chromatin structure allowing active transcription during one cell-cycle stage and in a transcriptionally repressed chromatin state in another stage, chromatin can greatly influence the levels of gene expression (reviewed in Henikoff, 2008). However, chromatin can influence gene expression not only by making gene regulatory sequences accessible but also by recruiting binding proteins. These so called *trans*-mediated processes are

orchestrated by post-translational histone modifications that provide a dynamic binding platform for effector-proteins. Effector proteins bind specific histone modifications and 'interpret' the modification pattern by triggering appropriate downstream processes (see below: Molecular Mechanisms).

Histone variants

Most eukaryote genomes not only contain a set of four core histones but also variant forms of histones, which can differ from their canonical counterparts in as little as one amino acid (human H3.1 / H3.2) or they can contain more dramatic alterations like additional domains (macroH2A) (Pehrson and Fried, 1992). Numerous classes of histone variants of the H2A and H3 families have been identified in multiple organisms; in contrast, variant forms of H2B and H4 appear to be very rare (Bernstein and Hake, 2006). Tritryps are the only species known to contain histone variants of all four canonical histones (for alignments see Figure 1.8). Whereas canonical histones are often encoded by multigene arrays (e.g. H2A in *T. brucei* by 13 genes in one array), histone variants are always encoded by single genes.

Differences in the primary sequence allow histone variants to be marked with specific posttranslational modifications. For example, H2AX is a variant form of H2A that is characterized by a C-terminal SQ(E/D)Θ motif (Mannironi et al., 1989). The C-terminal serine of this motif is

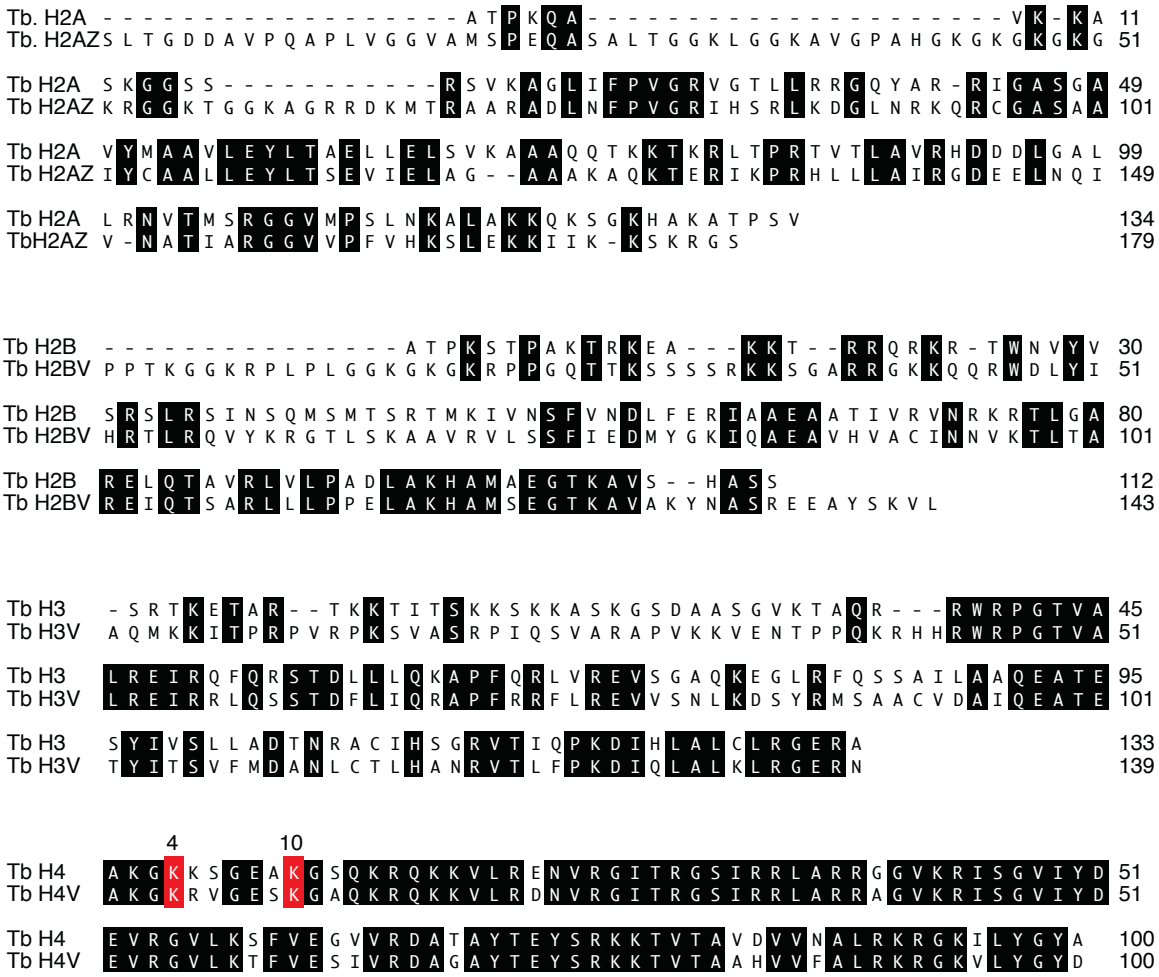


Figure 1.8. Pairwise alignment of *T. brucei* core and histone variants
Alignment of *T. brucei* histone variants with their canonical counterparts reveals low sequence identity. Sequence identity between H2A and H2AZ: 44%; H2B and H2BV: 37%; H3 and H3V: 45%; H4 and H4V: 85%. Residues H4K4 and H4K10 (red background) were analyzed as part of this thesis work.

phosphorylated in response to DNA damage (Rogakou et al., 1998). This relatively simple motif has evolved independently in multiple organisms underlining the importance of this variant (reviewed in Redon et al., 2002). No H2AX has been identified in the tritryp genomes.

Given that histone variants, unlike canonical histones, can be synthesized and deposited in a replication independent manner and that specialized chaperones aid proper deposition (reviewed in Henikoff and Ahmad, 2005), it is not surprising that histone variants are deposited in specific genomic locations where they fulfill specialized functions. One example of distinct localization is the centromeric protein A (CENP-A) (Guldner et al., 1984), a highly divergent variant form of H3 (Palmer et al., 1991; Palmer et al., 1987). CENP-A, whose presence has been confirmed in most eukaryotes but not in the tritryp genomes, is found exclusively in centromeric nucleosomes where it replaces H3 and plays an essential role in kinetochore formation and chromosome segregation (Blower and Karpen, 2001). Different lines of evidence even suggest that it is CENP-A positioning that defines the human centromere rather than DNA sequence motifs, making it a true epigenetic mark (Henikoff et al., 2001; Amor et al., 2004; Dawe and Henikoff, 2006).

Less well understood is the relationship between genomic location and function of H3.3 and H2AZ, variant forms of histone H3 and H2A, respectively. Both variant forms have been identified in multiple organisms. While several studies point to a link between H3.3 and transcription activation (Ahmad and Henikoff, 2002; Daury et al., 2006; Jin and Felsenfeld, 2006), findings with regard to H2AZ are somewhat contradictory. The observations that H2AZ was found only in the active macronucleus of *Tetrahymena* and was absent from the silent germ-line

micronucleus first pointed to a role for H2AZ in active transcription (Allis et al., 1980). In *Saccharomyces cerevisiae*, the H2AZ ortholog, HTZ-1, locates to actively transcribed genes in the vicinity of heterochromatin, and it has been suggested that it prevents spreading of heterochromatin (Meneghini et al., 2003). Immunofluorescence (IF) microscopy and chromatin immunoprecipitation-sequencing (ChIP-seq, see also Chapter 2: Materials and Methods) studies of mammalian H2AZ led to opposite findings. While in undifferentiated mouse cells H2AZ marks pericentric heterochromatin (Rangasamy et al., 2003), in human $CD4^+$ T cells, H2AZ is enriched at promoter regions of actively transcribed genes (Barski et al., 2007).

A plausible solution to this conundrum was provided by the recent finding that both H3.3 and H2AZ can co-exist in the same nucleosome and that the presence of these two variants in the same nucleosome rendered the nucleosome significantly less stable compared to nucleosomes containing only one variant (Jin and Felsenfeld, 2007). It, thus, appears plausible that both H2AZ and H3.3 act jointly to destabilize nucleosomes at gene regulatory elements. Destabilization could facilitate nucleosome eviction, expose DNA sequence and put genes in a state poised for active transcription. Nucleosomes containing only H2AZ but not H3.3 may serve a different role, which could explain the different distribution patterns for H2AZ.

As mentioned above, the tritryps are unique among eukaryotes as they contain one variant for each canonical histone, H2AZ, H2BV, H3V and H4V. In *T. brucei*, the degree of sequence identity between variants and their respective canonical counterparts is low compared to higher eukaryotes, ranging from 37% for H2BV to 85 % for H4V (Figure 1.8).

In *T. brucei,* the two histone variants H2AZ and H2BV, both essential for viability, have been shown to co-exist in the same nucleosome, but their function remains unknown (Lowell et al., 2005). In contrast to H2AZ and H2BV, H3V is not essential, and ChIP experiments followed by dot blot analysis revealed H3V enrichment in telomeric and sub-telomeric regions (Lowell and Cross, 2004). Even less is known about the function of H4V in *T. brucei,* but preliminary work indicates nuclear localization of the variant (unpublished data, Dorthe Hoeg and Pradeep Patnaik).

In Chapter 5, I describe my findings on the distribution of all four histone variants in the *T. brucei* genome and suggest a possible biological function for the H2AZ/H2BV dimer.

Histone post-translational modifications

Post-translational acetylation and methylation of N-terminal histone tails were first identified by Vincent Allfrey and coworkers more than 40 years ago (Allfrey et al., 1964). Using calf thymus nuclei, they also showed that acetylated histones had a much less inhibitory effect on

RNA transcription than unmodified histones, indicating a function in transcription activation. However, the exact role of histone acetylation in regulating gene expression has long remained contested. In the years that followed, numerous other histone post-translational modifications (PTMs), including phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, deimination and proline isomerization were identified (Kouzarides, 2007). Some of these protein modifications, like mono-ubiquitination, were originally identified on histones (Goldknopf et al., 1977).

The generation of a large number of highly specific antibodies to different histone PTMs and the development of new techniques like ChIP have allowed PTMs to be linked to specific chromatin states. For example, using ChIP, it was shown that DNase-hypersensitive regions in the genome — those in an open chromatin conformation — contained hyperacetylated histones (Hebbes et al., 1994).

Despite a high degree of evolutionary conservation of histone sequences, the primary sequences of tritryp histones differ significantly from those in other eukaryotes. These differences have prevented tritryp researchers from taking advantage of the continuously growing number of commercially available antibodies against histone PTMs. This lack of antibodies has complicated the identification and study of histone modifications such that it was only a few years ago that, with the use of mass spectroscopy and Edman degradation, the first histone PTMs were

identified in the tritryps (Janzen et al., 2006a; Mandava et al., 2007; da Cunha et al., 2006). How the total number of *T. brucei* histone PTMs will compare to those of other eukaryotes is still not clear. While a similar number of histone methylations and acetylations have been identified in regions thoroughly analyzed, data are still lacking for histone H3 and other types of histone modifications (Mandava et al., 2007). Past analyses also revealed some notably different modification patterns. For example, there is a cluster of six acetylated lysines on the C-terminal tail of histone H2A that is absent from human histones (Figure 1.9), and unusually, the α-amido group of every N-terminal amino acid is modified in *T. brucei*. H2A is either acetylated or methylated. H2B and H4 are methylated, and H3 is acetylated. Acetylation of the α-amido group of the N-terminus blocks Edman degradation and is the reason for the lack of data available for the H3 histone tail.

Despite some similarity between the histone modification patterns in *T. brucei* and other eukaryotes, the extensive sequence differences in most cases hinder the identification of functionally homologous modifications. Possible exceptions may be *tb*H3K4me3, which probably corresponds to human H3K4me3, *tb*H3K76me3 (*h*H3K79me3) and *tb*H4K17/18me3 (*h*H4K20me3).

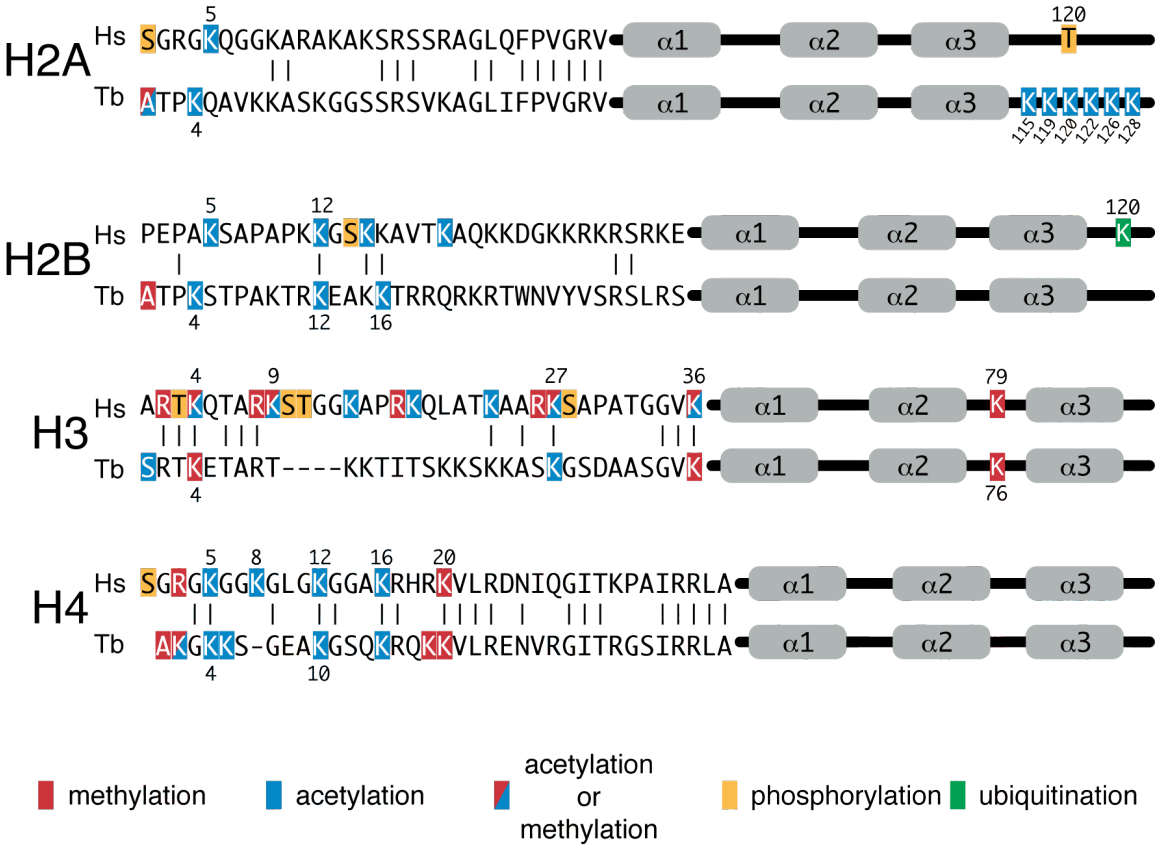


Figure 1.9. Comparison of histone modifications identified in humans and *T. brucei*

The degree of histone methylation and acetylation on the N-terminal tails of H2A, H2B and H4 is similar between *T. brucei* (Tb) and humans (Hs). In *T. brucei*, the α-amido group of every N-terminal amino acid is modified, a fact that thus far has prevented a thorough characterization of histone H3 modifications. The low degree of sequence homology between trypanosome histones and those of other eukaryotes makes it difficult to identify homologous modifications. Grey boxes indicate α-helices of the globular histone domain and are not drawn to scale.

Histone modifying enzymes

Correlative studies had long suggested a role of histone acetylation in transcription activation (Allfrey et al., 1963; Allfrey and Mirsky, 1962),

but it was the discovery that the enzyme Gcn5 (general control nonrepressed) possessed histone acetyltransferase activity that marked the breakthrough (Brownell et al., 1996). Gcn5 had previously been shown to function as a transcriptional co-activator (Georgakopoulos and Thireos, 1992); thus, a link between histone acetylation and transcription activation was established. At the same time, the first histone deacetylase (HDAC) was identified (Taunton et al., 1996). With a plethora of known PTMs and the continuing discovery of enzymes responsible for addition and removal of PTMs, the term 'histone code' was coined. It was proposed that this epigenetic code, established by addition of histone modifications and incorporation of histone variants, was responsible for regulating numerous cellular processes (Strahl and Allis, 2000). Much progress has been made to 'translate' the histone code, but the high number of PTMs (PTMs have been identified on more than 60 different sites, Kouzarides, 2007) and the large number of histone modifying enzymes (Table 1.2) have been a challenge.

Homology searches, using sequences of histone modifying-proteins from other eukaryotes as reference genes, indicate that *T. brucei* contains seven HDACs, five lysine acetyltransferases (KATs), six lysine methyltransferases (KMTs), four arginine methyltransferases (PRMTs), and four lysine demethylases (Table 1.2) (Gough et al., 2001; Ivens et al., 2005). Of the five putative KATs, two resemble elongation protein 3 (Elp3) histone acetyltransferase and three represent MYST-type KATs, which I

will refer to by their original name, histone acetyltransferase 1–3 (HAT1–3). MYST-type KATs, named after the founding members of its group (MOZ, YBF2/SAS3, SAS2 and TIP60), represent a large family of related histone acetyltransferases. These KATs have been intensively studied in yeast, *Drosophila* and humans and are implicated in transcription, DNA replication recombination and repair (Pillus, 2008).

Based on sequence homology to the three yeast histone deacetylases, Rpd3 (reduced potassium dependency 3), Hda1 (histone deacetylase 1) and Sir2 (silent information regulator 2), histone deacetylases are generally divided into three classes. Later with the identification of human HDAC11, a fourth class of histone deacetylases was added (Gao et al., 2002). Of the seven trypanosome HDACs, two share significant sequence similarity with class I HDACs (*tb*HDAC1–2), two with class II HDACs (*tb*HDAC3–4) and three with class III HDACs (*tb*Sir2rp1–3) (Ingram and Horn, 2002; Horn, 2008).

Even though the number of histone PTMs in *T. brucei* may be similar to those identified in other eukaryotes (Figure 1.9), the smaller number of histone modifying enzymes should facilitate linking of enzymes to respective PTMs. So far, histone-modifying activity has only been demonstrated for three enzymes in *T. brucei*, Dot1A (disruptor of telomeric silencing), Dot1B (Janzen et al., 2006b) and the histone deacetylase Sir2 (Garcia-Salcedo et al., 2003) (see Table 1.3).

Table 1.2. Histone-modifying enzymes

	T. brucei	*Human*
KATs	5 total HAT 1–3 / MYST type Elp3a–b	17 total
HDACs	7 total HDAC1–2 class I HDAC3–4 class II Sir2rp1–3 / class III	18 total
Lysine methyltransferases	6 total KMT1–3 Dot1A–C	24 total
Lysine demethylases	4 total KDM1–4	15 total
Arginine methyltransferases	4 total PRMT1–4*	27 total

The number of histone-modifying enzymes is based on a literature search, except for cases denoted with an asterisk (*), which are based on the domain search engine Superfamily (Gough et al., 2001).

Because Sir2 plays a central role in silencing telomere-proximal genes in yeast (Gottschling et al., 1990; Braunstein et al., 1993), a mechanism referred to as telomere position effect, much effort has been put into testing if *tb*Sir2 plays a role in silencing telomeric VSGs and other expression site associated genes. *T. brucei* contains three Sir2 paralogs, *tb*Sir2rp1, *tb*Sir2rp2 and *tb*Sir2rp3, but only *tb*Sir2rp1 localizes to the nucleus, whereas *tb*Sir2rp2 and *tb*Sir2rp3 localize to the kinetoplast (Garcia-Salcedo et al., 2003; Alsford et al., 2007). As is characteristic for yeast and human Sir2, *tb*Sir2rp1 possesses NAD-dependent histone deacetylase activity (Garcia-Salcedo et al., 2003). However, unlike in yeast where Sir2 deacetylates the N-terminal tail of H4 (Imai et al., 2000), in trypanosomes, deletion of the non-essential

Table 1.3. Summary of available functional data for histone-modifying enzymes in *T. brucei*

	Essential	Substrate	Function/Observations
HAT1	Yes	Unknown	Telomeric derepression
	(Kawahara et al., 2008)		
HAT2	Yes	H4K10	Possibly transcription activation
	(Kawahara et al., 2008) and Nicolai Siegel unpublished data		
HAT3	No	H4K4	80% of H4K4 is acetylated, possible role DNA repair
	(Siegel et al., 2008) and David Horn, personal communication		
Elp3a	No	Unknown	Two localized foci in nucleus
	David Horn, personal communication		
Elp3b	No	Unknown	Two localized foci in nucleus
	David Horn, personal communication		
HDAC1	Yes	Unknown	
	(Ingram and Horn, 2002)		
HDAC2	No	Unknown	Missing critical residue in acetyl-binding pocket
	(Ingram and Horn, 2002)		
HDAC3	Yes	Unknown	
	(Ingram and Horn, 2002)		
HDAC4	No	Unknown	Depletion leads to delay in G2/M phase
	(Ingram and Horn, 2002)		
Sir2rp1	No	H2A, H2B	ADP-ribosylation, role in DNA damage repair, telomeric derepression, deacetylases activity
	(Garcia-Salcedo et al., 2003; Alsford et al., 2007)		
Sir2rp2	No	Unknown	Localizes to mitochondrion
	(Alsford et al., 2007)		
Sir2rp3	No	Unknown	Localizes to mitochondrion
	(Alsford et al., 2007)		
Dot1A	Yes	H3K76	Causes K76 di-methylation, depletion leads to cell-cycle defect
	(Janzen et al., 2006b)		
Dot1B	No	H3K76	Causes K76 tri-methylation, depletion leads to defect in silencing VSGs
	(Janzen et al., 2006b; Figueiredo et al., 2008)		

Sir2rp1 did not have a detectable effect on acetylation levels of histone H4 and did not lead to derepression of ES-associated VSGs (Alsford et al., 2007; Mandava, 2007). Instead, *tb*Sir2rp1 showed NAD-dependent ADP-ribosyltransferase activity toward histones H2A and H2B. Levels of *tb*Sir2rp1 expression as well as histone H2A and H2B ADP-ribosylation were upregulated upon DNA damage (Garcia-Salcedo et al., 2003).

Dot1A and Dot1B are the only other enzymes with confirmed histone-modifying activity in *T. brucei*. Dot1A and Dot1B cause di-methylation and tri-methylation of H3K76, respectively, but only Dot1A is essential for viability and leads to cell cycle defects when downregulated. Deletion of Dot1B is accompanied by loss of tri-methylation and a delay in silencing VSG expression upon VSG switching, resulting in cells expressing two VSGs simultaneously (Figueiredo et al., 2008).

As part of my work to better understand the role of chromatin in transcription, I identified HAT2 as the enzyme responsible for H4K10 acetylation and HAT3 for H4K4 acetylation (see Chapter 4).

Molecular mechanisms of histone PTMs and the role of histone binding proteins

While the function of most histone modifications remains unknown, for some modifications, significant progress has been made

toward understanding the molecular mechanisms of how they exert biological functions.

As mentioned before, histone modifications can induce structural changes to chromatin conformation in a so-called *cis*-mediated manner. These changes in chromatin conformation are caused by modifications that change the physical property of histone tails, e.g. histone acetylation diminishes the net positive charge of histones. One model assumed that acetylation-mediated neutralization of histone tails would weaken the interaction between histones and negatively charged DNA, and thus facilitate binding of transcription factors to DNA. *In vitro* data, however, suggest that DNA-histone interactions are not weakened by histone acetylation (Mutskov et al., 1998). Instead it has been shown that acetylation of the histone H4 tail weakens inter-nucleosomal contacts, disrupts formation of chromatin fibers and, thus, contributes to a more open chromatin conformation favoring active transcription (Tse et al., 1998; Horn and Peterson, 2002; Luger et al., 1997; Waterborg, 2002). In the case of H4K16, it has been shown that acetylation of a single lysine residue can impede formation of chromatin fibers (Shogren-Knaak et al., 2006).

In an indirect, *trans*-mediated pathway, histone modifications serve as binding sites for effector proteins. Many effector proteins have been identified that bind to different histone modifications in a context dependent manner. Effector proteins are thus 'readers' of the histone

code that bind to a particular modification and trigger a distinct cellular event. These readers are characterized by the presence of specific domains such as the chromo (chromatin organization modifier) domain (Paro and Hogness, 1991), which specially binds to methylated histones. The list of chromo domain-containing proteins is long (Ruthenburg et al., 2007b) and includes well-characterized proteins like heterochromatin protein 1 (HP1), which specially binds to methylated H3K9 (Bannister et al., 2001; Lachner et al., 2001), and Polycomb, a component of a large complex in *Drosophila* that is important in development (Schwartz and Pirrotta, 2008). Binding of either protein exerts a repressive effect on transcription.

Another histone binding domain is the so-called bromodomain, a module binding to acetylated lysines (Tamkun et al., 1992; Dhalluin et al., 1999; Jacobson et al., 2000). In agreement with the long known correlation of hyperacetylation and actively transcribed genes, bromodomain-containing proteins include transcription factors (PCAF, Gcn5 and $TAF_{II}250$) and chromatin remodeling complexes (Swi/Snf, RSC). Therefore, once a site is acetylated and poised for transcription, it represents a binding site to recruit more histone acetyltransferases that can lead to additional acetylation. Thus, *cis*-mediated effects of histone acetylation that generate a more open chromatin conformation can be further enhanced by the ability of acetylated sites to act in *trans* and to provide a binding site for chromatin remodeling complexes.

Based on sequence homology, the trypanosome chromosome remodeler *Tb*ISWI contains a chromo domain (Ivens et al., 2005). Just like the chromo domain-containing proteins HP1 and Polycomb, *Tb*ISWI exerts a transcriptional repressive function, and depletion of the chromatin remodeler leads to derepression of silent expression sites (Hughes et al., 2007). Nothing is known about the molecular mechanisms of this repression. It will be interesting to see how a chromatin remodeler, a complex usually associated with activating gene transcription, exerts a repressive function in *T. brucei*.

Unlike the human histone acetyltransferases Gcn5 and PCAF that contain bromodomains, the trypanosome histone acetyltransferases, HAT1 and HAT2, contain a chromo domain and, thus, can be recruited to methylated sites, indicating a link between methylation and acetylation. A concerted role for a methyl and an acetyl mark has been described for other MYST-type KATs, for example, during activation of the Hox a9 gene in humans (Dou et al., 2005). In this case, one complex containing both MLL1 (mixed-lineage leukaemia 1) methyltransferase activity and MOF (males-absent-on-the-first) acetyltransferase activity leads to methylation of H3K4 and acetylation of H4K16. The concerted modification of different sites can provide highly specific docking sites for combinatorially generated chromatin regulatory complexes. Such complexes containing multiple histone-binding proteins (bromodomain-containing factor, chromo domain-containing factor, etc.) have been

identified in numerous organisms (reviewed in Wu et al., 2009). In addition, it has been suggested that it is the cooperative binding of a complex to multiple histone modifications that is of biological relevance rather than the relatively weak and often nonspecific interactions of an individual histone binding protein with a single histone modification (Ruthenburg et al., 2007b).

Another type of regulation through concerted histone modifications has been described in *S. cerevisiae*, where methylation of H3K4 is dependent on prior ubiquitination of H2B-K123 (Sun and Allis, 2002). Given the lack of a homologous ubiquitination site in trypanosome H2B, such a link is impossible. Recent findings by Veena Mandava indicate that *tb*H3K4me3 may be enriched in nucleosomes containing H2BV, suggesting some other form of interdependence (Mandava et al., 2008).

Besides the three chromo domain-containing proteins mentioned above, the *T. brucei* genome encodes six putative bromodomain-containing factors, most of which are relatively small (221–660 amino acids) and do not contain any other recognizable domains (see Figure 1.10). Neither chromo domain- nor bromodomain-containing factors have been further characterized in *T. brucei*.

Experimental approach

Despite a growing interest in trypanosome chromatin biology in recent years, very little is known about the role of individual histone

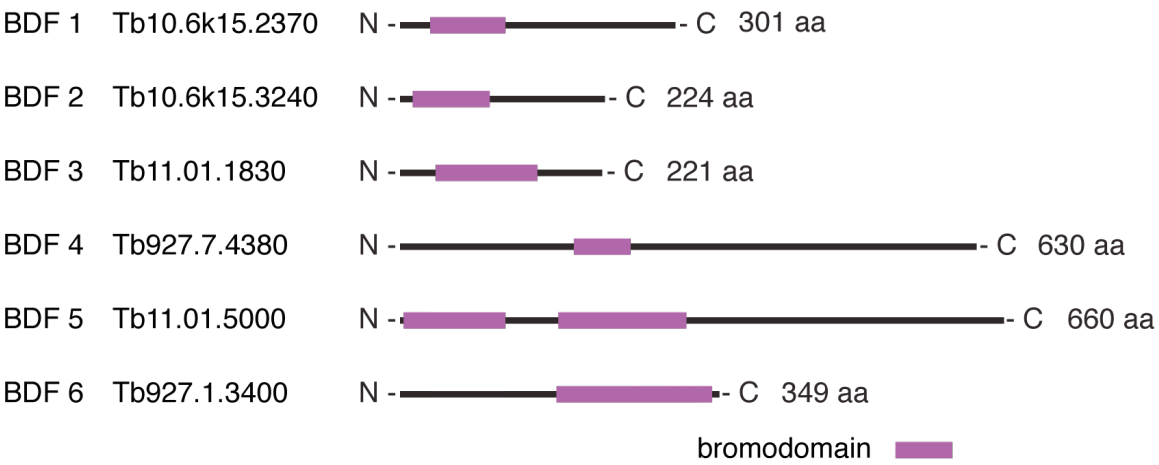


Figure 1.10. Bromodomain-containing factors in *T. brucei*
Based on 'superfamily' domain searches (http://supfam.mrc-lmb.cam.ac.uk, version 1.69), the *T. brucei* genome encodes for six putative bromodomain-containing factors. None of the six BDF contained any other recognizable domain besides the bromodomain.

PTMs and no molecular mechanisms of chromatin-mediated gene regulation have been worked out. Given the apparent lack of transcription regulation, little effort has been made to identify chromatin factors involved in transcription activation. Instead, research in *T. brucei* has almost exclusively focused on the identification of factors involved in mono-allelic VSG expression or, more specifically, in repression of silent expression sites. The relatively straightforward and generally successful approach taken by many laboratories involves genetic deletion or RNAi-mediated depletion of the factor of interest followed by an evaluation of the degree of derepression of silent expression sites (Alsford et al., 2007; Kawahara et al., 2008; Figueiredo et al., 2008; Hughes et al., 2007).

As part of my general focus on the regulation of gene expression, I decided to look at the role of chromatin in transcription activation. Given the prominent role of acetyl marks in transcription activation, I focused on two acetyl marks on the N-terminal tail of histone H4: H4K4 and H4K10. At the time I started my thesis work, these were the only histone acetylations identified on trypanosome histone H4 (Janzen et al., 2006a).

The most direct way to learn about the role of these specific histone acetyl marks would probably have been to delete the lysine of interest or to replace it with arginine or glutamine in order to mimic deacetylated and constitutively acetylated lysine, respectively, and look for a biological effect. These approaches have been applied successfully in yeast and confirmed the notion of a general redundancy of acetyl marks on the tails of histone H3 and H4 and a role of histone acetyl marks in regulating gene expression (Kayne et al., 1988; Megee et al., 1990). However, a characteristic feature of the trypanosome genome is the large number of gene families, especially for highly expressed genes like histones, which greatly complicates mutational analysis of histones (Berriman et al., 2005). Attempts to generate a trypanosome cell line in which the endogenous histone alleles can be replaced easily by a gene copy carrying the mutation of choice have failed due to the inability to replace a large multigene family with a single histone gene (Veena Mandava, unpublished data).

During my work I have focused, thus, on establishing a link between acetyl marks and the corresponding histone modifying enzymes. Furthermore, I have tried to deduce the possible function of modifications by determining their localization in the nucleus using IF microscopy and along the genome by ChIP-seq.

Chapter 2: Materials and Methods

Molecular Biology

Plasmid construction

Plasmids for trans-splicing analysis

All reporter plasmids used in these experiments are derivatives of pNS10, which contains luciferase and *lacZ* reporter genes and restriction sites that permit the insertion of alternative upstream regions (URs) (see Figure 6.1). pNS10 was constructed from pLew20 (Wirtz et al., 1998) as follows. First, pLew20 was digested with *Srt*I and *Stu*I to remove the *ble* cassette. Next, the construct was digested with *Sma*I and *Bsm*I to remove a 139-bp region between the procyclin promoter and the luciferase coding region to allow its replacement with a synthetic oligonucleotide that created sites for inserting alternative splice site motifs and eliminated all AG dinucleotides except for the wild-type splice site. The 3´ overhang generated by the *Bsm*I digest was removed by Klenow polymerase. Two complementary 90-nt oligonucleotides (5´-GGGAAAAAGCTTCAATTACACCAAAAAATAAAATTCACAAACTTGGAATTC CTTTGTGTTACATTCTTGAATGTCGCTCGCAATGACATT-3´) were annealed, and the double-stranded insert was ligated into the open vector.

To add *lacZ* as a second reporter, the *lacZ* sequence was PCR amplified from plasmid pSV-β-Galactosidase (Promega). 5´ and 3´ UTRs

were PCR amplified from the *ble* cassette of pLew20. The three fragments, the 5´ UTR, the *lacZ* coding region, and the 3´ UTR, were ligated, and the product was PCR amplified. The amplified *lacZ* cassette was then ligated into the single-reporter construct to yield the double-reporter construct pNS5.

Finally, two ATG sites were removed from the luciferase 5´ UTR of the double-reporter construct by PCR-based site-directed mutagenesis. The first of the sites had been inadvertently introduced into the 90-nt oligonucleotide above. The second was present in a small sequence upstream of the luciferase open reading frame (ORF) and was present in all previous luciferase containing constructs used in this and other laboratories, and created a short upstream ORF whose effect was not previously appreciated (see Results in Chapter 6). To remove the first ATG site, we used a forward primer (NS35 [5´-TCTCGTCCCGGGAAAA AGCTTC-3´]) containing the sequence surrounding the SmaI site of pNS5 and a reverse primer (NS24 [5´-CCATCCTCTAGAGGATAGAATGG-3´]) containing the sequence surrounding the XbaI site. In addition to these two outside primers, two inner primers were used to introduce the desired changes in nucleotide sequence. The inner primers are complementary and contain the following sequences: NS27, 5´-GAATGTCGCTCGCAcTGACATTacCATTCCGGTACTGTTGG-3´; NS28, 5´-CCAACAGTACCGGAATGgtAATGTCAgTGCGAGCGACATTC-3´. Lowercase letters represent introduced changes. Two separate PCRs were performed

with pNS5 as template and primer pairs NS23-NS28 and NS27-NS24. Next, the products from both PCRs were used as templates in a third PCR with outside primers NS23 and NS24. *Sma*I- and *Xba*I-digested amplicons were reintroduced into pNS5, replacing the original sequence between *Sma*I and *Xba*I to give pNS9. The second ATG was removed by a similar strategy with the outside primers NS35 and NS24 but with the inside primers NS36 (5´-CCTTTGTGTTACATTCTTGATCGCTCGCACTGA CATTACC-3´) and NS37 (5´-GGTAATGTCAGTGCGAGCGATCAAGAAT GTAACACAAAGG-3´) to generate pNS10.

To insert a UR of choice, two complementary synthetic oligonucleotides were annealed and ligated into the double-reporter construct pNS10 between the *Sma*I and *Hin*dIII sites.

pNS11 was generated by replacing the sequence between the *Sma*I and *Xba*I sites of pNS10 with the sequence between the *Sma*I and *Xba*I sites originally found in pLew20, which contains four cryptic AG sites. Finally, an ATG site was removed from the newly inserted sequence by PCR-based site-directed mutagenesis as described above for pNS10. pNS10/56 and pNS10/72 to pNS10/75 all contain the *T. brucei* α-tubulin UR and differ only in sequences of the luciferase 5´ UTR. Changes in 5´ UTRs were generated by site-directed mutagenesis as described above, with appropriate primers. pNS20/74 is a derivative of pNS10/74 in which we replaced the *lacZ* gene with a phleomycin resistance marker.

All oligonucleotides were purchased from Integrated DNA Technology and purified as recommended by the vendor. All constructs were verified by DNA sequencing.

Plasmids for tagging, deletion and knock down of genes

Tagging of bromodomain factors (BDFs) was performed by using a PCR-based tagging approach optimized in the Seebeck Lab (Oberholzer et al., 2006). The PCR product was cloned into pGEM-T Easy (Promega), digested with *Eco*RI, *Sac*I or *Sph*I (all NEB) and transfected into BF or PF cells. BDF1–5 were tagged with a C-terminal triple-HA-tag derived from pMOTag2H.

BDF3 and histone H4V deletion constructs were generated by PCR-amplification of resistance markers using primers with long overhangs corresponding to the 5´ UTR and 3´ UTR of the BDF3 or H4V genes.

The cell line for inducible RNAi-mediated depletion of BDF3 was generated as described previously (Alibu et al., 2005). A ~450-bp fragment of BDF3, chosen by RNAit software (Redmond et al., 2003), was PCR amplified, ligated into p2T7^{TAblue}, and the vector was transfected into a cell line containing a triple-HA-tagged BDF3. Efficient RNAi-mediated BDF3 depletion was monitored over 27 hours by western blotting (data not shown). Constructs for tagging, deletion and knock-down of proteins were generated by Louise Kemp.

Table 2.1. Plasmids used in this study

Plasmid	Description
pLew 20	*LUC* and *BLE* driven by GPEET promoter
pNS5	Like pLew20 but *BLE* is replaced by *lacZ* and AG sites are removed from 5´ UTR
pNS10	Like pNS5 but without ATG in 5´ UTR
pNS11	Like pNS10 but with native procyclin 5´ UTR and UR
DH1	Δ*hat3::PUR* deletion construct
DH2	Δ*hat3:: NEO* deletion construct
pMOTag2H	Construct for C-terminal triple HA-tagging, contains *PUR*
pMOTag3H	Construct for C-terminal triple HA-tagging, contains *NEO*
pMOTag4H	Construct for C-terminal triple HA-tagging, contains *HYG*
pMOTag5H	Construct for C-terminal triple HA-tagging, contains *BLE*
pMOTag2T	Construct for C-terminal TY1-tagging, contains *PUR*
ΔBDF2#1	Δ*bdf2::PUR*
ΔBDF2#2	Δ*bdf2::HYG*
ΔBDF3#1	Δ*bdf3::PUR*
ΔBDF3#2	Δ*bdf3::HYG*
ΔH4V#1	Δ*h4v::PUR*
ΔH4V#2	Δ*h4v::HYG*
pT7-^{GFP}HAT2	Ectopic, inducible expression of GFP-tagged HAT2
pT7-^{GFP}C-HAT2-A	Ectopic, inducible expression of GFP-tagged HAT mutated at the active site
(pHAT1RP^{SLRI})	Inducible, RNAi-mediated downregulation of HAT1
(pHAT2RP^{SLRI})	Inducible, RNAi-mediated downregulation of HAT2

Primer extension

T. brucei 29.13 cells were stably transfected with pNS20/74. RNA extractions and primer extensions were performed with an RNeasy Kit (Qiagen) and a Primer Extension Kit (Promega) by following essentially the provided protocols. Primer luc92/107 (AGCGATCAAGAATGTAACACA AAGG) anneals downstream of the 3´SS of luciferase on pNS20/74 and should yield products of 92 or 107 nt, depending on which AG is used as

the splice site. Primer luc108/123 TGGTAATGTCAGTGCGAGC GATCAAG) anneals further downstream than luc92/107 and should yield products of 108 or 123 nt. I also included primers α-tubulin130 (GTGCTTTGTTGTTGTTGTTAGTGGTGCT) and β-tubulin120 (GAAC GCAGACGATTTCGCGCATA). All primers were labeled with [γ-^{32}P]ATP (6,000 Ci/mmol).

Real-Time PCR

PCR primers were designed using Primer3 v. 0.4.0 (http:// frodo.wi.mit.edu, Rozen and Skaletsky, 2000) to amplify a region of 180–220 bp with an annealing temperature of ~62°C. The linear range of newly designed primers was tested on genomic DNA. All real-time PCR analyses were performed in duplicate at three different sample DNA concentrations to ensure linear range. The total reaction volume was 25 µl consisting of 12.5 µl 2× SYBR Green Mix (Applied Biosystems), 5 µl sample DNA, 1 µl primer 1 (10 µM), 1 µl primer 2 (10 µM) and 5.5 µl H_2O. The PCR was performed as follows: initial denaturation at 95°C for 10 min, subsequent denaturation at 95°C for 15 sec, annealing at 62°C for 15 sec, elongation at 72°C for 20 sec; for 40 cycles.

Cell-based methods

Trypanosome culture and transfection

All BF cell lines generated during my thesis work were derived from wild-type Lister 427 (MITat 1.2, clone 221a) or a derivative 'single marker' line, which expresses T7 RNA polymerase and the Tet repressor (Wirtz et al., 1999). Cells were cultured in HMI-9 medium (Hirumi and Hirumi, 1989) with appropriate drug selection (http://tryps.rockefeller.edu/trypsru2_genetics.html). Stable transfections were performed using a Nucleofector (Amaxa) as described previously (Scahill et al., 2008).

All PF cell lines were produced from wild-type Lister 427 or a derivative clone, 29.13, which expresses T7 RNA polymerase and the Tet repressor (Wirtz et al., 1999). Cells were grown in SDM-79 (Brun and Schonenberger, 1979) containing 10% fetal bovine serum and 0.25% hemin. Stable transfections were performed using a BTX electroporator as described previously (Wirtz et al., 1999). For transient expression studies 2×10^7 at ~8×10^6/ml cells were transfected using a BTX electroporator with 10 μg of DNA and incubated for 16 to 22 hours before harvest as previously described (Wirtz et al., 1998).

Table 2.2. PF cell lines used in this study

Strain Name	Genotype
427 PF	WT 427
PF 29.13	29.13; *TETR T7POL NEO HYG*
PF 29.13 pNS20/74	29.13; *LUC BLE*
PF HAT3 DKO	WT; Δ*hat3::NEO* | Δ*hat3::PUR*
PF BDF2-HA	WT; *BDF2::HA PUR*
PF BDF3-HA	WT; *BDF3::HA PUR*

Table 2.3. BF cell lines used in this study

Strain Name	Genotype
221 BF	WT
BF 'single marker' (SM)	SM; *TETR T7POL NEO*
SM HAT1 knockdown	SM; Δ*Hat1*^{Ti} *HYG*
SM HAT2 knockdown	SM; Δ*Hat2*^{Ti} *HYG*
BF HAT3 DKO	WT; Δ*hat3::NEO* | Δ*hat3::PUR*
HAT2-GFP	SM; *HAT2::GFP HYG*
C351A, HAT2-GFP,	SM; *HAT2C351A*^{Ti}*::GFP HYG*
SM BDF1-HA	SM; *BDF1::HA PUR*
SM BDF2-HA	SM; *BDF2::HA PUR*
SM BDF3-HA	SM; *BDF3::HA PUR*
SM BDF4-HA	SM; *BDF4::HA PUR*
SM BDF5-HA	SM; *BDF5::HA PUR*
BF BDF3-HA, SKO	WT; *BDF3::HA PUR* Δ*bdf3::NEO*
BF BDF3-Ty1, SKO	WT; *BDF3::Ty1 PUR* Δ*bdf3::NEO*
BDF2-HA in HAT2kd	SM; Δ*Hat2*^{Ti} *HYG BDF2::HA PUR*
BDF3-HA in HAT2kd	SM; Δ*Hat2*^{Ti} *HYG BDF3::HA PUR*
BF BDF2 SKO	WT; Δ*bdf2::NEO*
BF BDF3 SKO	WT; Δ*bdf3::NEO*
BFJEL28	SM; *Ty1::H3V BLE*
BFJEL22	SM; *Ty1::H4V BLE*
BFJEL43	SM; *Ty1::H2BV BSD* Δ*h2bv::PUR* | Δ*h2bv::HYG*
BFpFK3.4.5	SM; *Ty1::H2AZ BLE* Δ*h2az::PUR* | Δ*h2az::HYG*
SM H4V, DKO	SM; Δ*h4v::HYG* | Δ*h4v::PUR*
Ty1-H4V, DKO	SM; *Ty1::H4V BLE* Δ*h4v::HYG* | Δ*h4v::PUR*

Immunofluorescence microscopy

PF or BF *T. brucei* cells were suspended at 1 × 10^7 cells/ml in SDM-79 or HMI-9 containing 2% formaldehyde, incubated for 5 min at room temperature (RT) and washed three times with phosphate-buffered saline (PBS). The fixed cells were allowed to settle onto aminopropyltriethoxysilane-coated coverslips and permeabilized by immersion for 5 min in PBS containing 0.1% NP-40. After blocking by two rinses of 10 min with PBG (PBS containing 0.2% cold-fish gelatin (Sigma) and 0.5% bovine albumin serum (BSA, Sigma)), the coverslips were incubated with primary antibody for 1 hour. Subsequently, the cells were rinsed four times for 5 min with PBG, incubated with the corresponding secondary antibody for 1 hour, stained with DAPI (4,6-diamidino-2-phenylindole) at a final concentration of 1.0 ng/ml for 10 min and mounted in antifade mounting solution (Vectashield, Vecta Laboratories). Alternatively, a DAPI-containing mounting solution was used (Vectashield, Vecta Laboratories). Vertical stacks of 15–50 slices (0.1–0.2 mm steps) were captured using a DeltaVision microscope (Applied Precision). Deconvolution and pseudo-coloring was performed using softWoRxTM v3.5.1 software. For a more detailed description of quantitative image analysis and co-localization studies see Chapter 7.

Cell sorting

Unfixed PFs (~1.3 × 10^7 cells/ml) were incubated for 30 min at RT in SDM-79 containing 5 μM DyeCyclin Orange (Invitrogen). For comparison fixed cells were also sorted. To stain after fixation, 2 × 10^7 cells were washed twice with PBS, re-suspended in 200 μl of PBS and fixed by addition of 2 ml of ice-cold 70% ethanol while vortexing. Cells were stored overnight at 4°C, pelleted and resuspended in 0.5 ml of PBS containing 2 mM EDTA (ethylenediaminetetraacetic acid), 200 mg/ml RNAseA, 2.5 mg/ml propidium iodide and incubated at 37°C for 30 min. Cell sorting, based on relative DNA content, was performed in the Rockefeller University Flow Cytometry Resource Center using a FACSAria (BD Biosciences).

Luciferase and β-galactosidase assays

PF cells (2 × 10^7; ~8 × 10^6/ml) were transfected with 10 μg of DNA and incubated for 16 to 22 hours. Cells were pelleted at 700 × *g* at 4°C for 10 min, resuspended in 1 ml of cold PBS, transferred to 1.5-ml Eppendorf tubes, and centrifuged again at 8,000 × *g* at 4°C for 3 min. The supernatant was removed, and the cells were resuspended in 100 μl of Cell Culture Lysis Reagent (Promega). Ten microliters of the lysed cells was mixed with 45 μl of Promega luciferase assay buffer, and luciferase activity was measured immediately in a Turner TD-20e luminometer.

To measure β-galactosidase activity, the remaining 90 µl of lysed cells were centrifuged at 17,900 × *g* at 4°C for 4 min. Thirty microliters of the cell lysate supernatant was added to a reaction mixture containing 260 µl of 0.1 M sodium phosphate, 6.6 µl of 10× CPRG (Chlorophenol red-β-D-galactopyranoside, Roche), and 3 µl of 100× Mg^{2+} solution (0.1 M $MgCl_2$, 4.5 M β-mercaptoethanol) and incubated for 8 hours at 37°C. The reaction was terminated by adding 500 µl of 1 M Tris, and the absorbance was determined at 570 nm with a Novaspec II spectrophotometer.

All transfections were done in duplicate, and the values reported in Chapter 6 are the averages of these two measurements; the error bars represent the luciferase activities in the two transfections. All transfections were repeated at least once on a different date, and all trends could be reproduced. β-galactosidase values remained relatively constant throughout the experiments

Biochemical Methods

Antibody generation

Polyclonal antibodies specific for unmodified or acetylated histone H4K4 and for acetylated H4K10 were raised by immunizing rabbits with KLH-conjugated peptides AKGKKSGEAC, AKGK(ac)KSGEAC and KSGEAK(ac)GSQKC according to the following 77-day protocol (Sigma):

day 0, immunization with 200 μg peptide in Complete Freund's Adjuvant (CFA); day 14, boost with 100 μg peptide in Incomplete Freund's Adjuvant (IFA); day 28, boost with 100 μg peptide in IFA; day 42, boost with 100 μg peptide in IFA; day 49, first production bleed; day 56, boost with 100 μg peptide in IFA; day 63, second production bleed; day 70, boost with 100 μg peptide in IFA; day 77, third production bleed. A general modification-independent histone H4 antibody was derived fortuitously from another rabbit immunized with the KLH-conjugated peptide AKGKKSGEAC according to the following 118-day protocol (Covance): day 0, immunization with 500 μg peptide in CFA; day 21, boost with 500 μg peptide in IFA; day 42, boost with 250 μg peptide in IFA; day 52, test bleed; day 63, boost with 250 μg peptide in IFA; day 73, first production bleed; day 84, boost with 250 μg peptide in IFA; day 94, second production bleed; day 105, boost with 250 μg peptide in IFA; day 115, third production bleed; day 118, terminal bleed.

A polyclonal antibody specific for unmodified histone H4K10 was raised by immunizing chickens with the KLH-conjugated peptide SGEAKGSQKC according to the following 115-day protocol (Covance): day 0, immunization with 500 μg peptide in CFA; day 21, boost with 500 μg peptide in IFA; day 35, test bleed; day 41, boost with 250 μg peptide in IFA; day 55, first production bleed; day 62, boost with 250 μg peptide in IFA; day 76, second production bleed; day 83, boost with 250 μg peptide in IFA; day 97, third production bleed; day 104, boost with 250 μg

peptide in IFA; day 115, terminal bleed. The antibody was extracted from egg yolk using the EGGstract IgY Purification system (Promega) following the instructions provided.

Crude rabbit and chicken antisera were affinity purified using the corresponding peptides immobilized to SulfoLink coupling gels (Pierce) as described in Harlow and Lane, 1999). All peptides used for immunizations were HPLC purified.

Western blotting

A total of 2×10^6 cells were lysed in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.25% sodium deoxycholate 0.1% SDS (sodium dodecyl sulfate), 1% commercially available mammalian proteinase inhibitor cocktail (Sigma, P8340)), 200 µg/ml PMSF (Phenylmethylsulfonyl fluoride) and 4 µg/ml pepstatin. Lysates were separated on a 15% SDS-PAGE gel and transferred onto a nitrocellulose membrane. Membranes were blocked for 1 hour with 3% BSA. When using non-histone antibodies, membranes were blocked for 1 hour with 5% nonfat dry milk. To decrease background, 1% of nonfat dry milk was added to the secondary antibody if the membrane had been blocked with BSA. Primary antibodies were detected with horseradish-peroxidase-conjugated sheep anti-rabbit (Amersham-Pharmacia) or anti-chicken antibodies (Jackson Laboratories) and SuperSignal West Pico (Pierce) or

ECL Plus (Amersham-Pharmacia). Intensity was quantified using a Versadoc imaging system (Bio-Rad).

Preparation of mononucleosomes

For each immunoprecipitation (IP) 2×10^8 BF cells were pelleted at 3000 × *g* at 4°C for 10 min, washed in trypanosome dilution buffer (TDB, 5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM NaH_2PO_4, 20 mM glucose, pH 7.4), resuspended in 1 ml permeabilization buffer (100 mM KCl, 10 mM Tris, pH 8.0, 25 mM EDTA, 1 mM DTT (DL-1,4-dithiothreitol), 1% commercially available mammalian protease inhibitor cocktail), and digitonin was added to a final concentration of 40 µM before the cell lysate was incubated for 5 min at RT. Next, the cell suspension was washed three times with isotonic buffer (100 mM KCl, 10 mM Tris, pH 8.0, 10 mM $CaCl_2$, 5% glycerol, 1 mM DTT, 1% commercially available mammalian protease inhibitor cocktail), supplemented with 2 units of micrococcal nuclease (Sigma) and incubated for 10 min in a 25°C water bath. To stop the reaction, EGTA (Ethyleneglycol-O, O'-bis(2-aminoethyl)-N, N, N', N'-tetraacetic acid) was added to a final concentration of 10 mM. NaCl (200 mM–600 mM final concentration) and NP-40 (0.05% final concentration) was then added and the cell suspension placed on ice for 5 min. Following a 10 min centrifugation at 16,100 × *g* at 4°C, DNA was extracted from ~20 µl of the supernatant (Qiagen mini-elute Kit) and analyzed on an 1.5% agarose gel

to determine the percentage of mononucleosomes. Typical preparations yielded more than 95% mononucleosomes.

Co-immunoprecipitation

Co-IPs were performed with mononucleosomes from 2 × 10^8 cells. Fifty microliters of IgG-coated magnetic beads (Dynal) were washed three times with 1 ml blocking solution (0.5% BSA in PBS), resuspended in 250 µl blocking solution and incubated with 5 µg of antibody at 4°C with rotation overnight. Antibodies used were BB2 (a monoclonal mouse antibody specific for the Ty1 peptide (EVHTNQDPLD) (Bastin et al., 1996)) and α-FLAG (a polyclonal rabbit antibody specific for the FLAG peptide (DYKDDDDK) (Sigma, F7425). The beads were then washed three times with blocking solution, resuspended in 50 µl of blocking solution, added to the mononucleosome-containing solution and incubated at 4°C with rotation for 3 hours. Magnetic beads and bound material were transferred to new tubes and washed with wash buffer 1 (200 mM NaCl; 50 mM Tris; 0.05% NP-40) for 15 min at 4°C with rotation, with buffer 2 (300 mM NaCl; 50 mM Tris, pH 8.0; 0.05% NP-40) for 15 min at 4°C with rotation, and with TE for 1 min. To elute immunoprecipitated material, magnetic beads were boiled for 10 min in 15 µl 1× sample buffer (10% glycerol, 0.005% w/v bromophenol blue, 2% w/v SDS, 62.5 mM Tris, 1 mM DDT, 1% β-mercaptoethanol, 1% commercially available mammalian

protease inhibitor cocktail). IP efficiency with BB2 antibody was about five times higher than with α-FLAG.

Chromatin immunoprecipitation and analysis by high-throughput sequencing

Chromatin IP

For each chromatin IP, 1×10^8 cells (conc.: 1.0×10^6 BF or 10×10^6 PF) were suspended in 40 ml of their respective media. To cross-link cells, 4 ml formaldehyde solution (50 mM HEPES-KOH (Gibco), pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 11% formaldehyde (Fisher)) was added to the medium, thoroughly mixed and incubated for 20 min at RT. Cross-linking was stopped by adding 2.5 ml of 2.0 M glycine (Sigma). Cells were then pelleted at 4000 × *g* at 4°C for 20 min, washed with 30 ml PBS and lysed by resuspension in 10 ml lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1% commercially available mammalian protease inhibitor cocktail) followed by 10 min incubation at 4°C. Cells were pelleted again and resuspended in 10 ml lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% commercially available mammalian protease inhibitor cocktail) followed by a 10 min incubation at RT. Before sonication, the sample was centrifuged again and resuspended in 2 ml of lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-

Deoxycholate, 0.5% N-lauroylsarcosine, 1% commercially available mammalian protease inhibitor cocktail). Following cell lysis, the DNA was sonicated for 10 cycles (30 sec on / 30 sec off) using a Bioruptor (Wolf Laboratories Limited) before adding Triton X-100 to a final concentration of 1%. Cell debris was pelleted by centrifugation at 16,100 *g*, at 4°C for 10 min, the supernatant containing the sonicated DNA transferred to a new tube and the NaCl concentration adjusted to 350 mM. Hundred microliters of antibody-coupled magnetic beads (for beads / antibody coupling see section on co-IP) were added to the sonicated cell solution and incubated at 4°C with rotation overnight.

Magnetic beads and bound material were transferred to new tubes and washed eight times with RIPA buffer (50 mM HEPES-KOH, pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-Deoxycholate) and once with TE buffer plus 50 mM NaCl. To elute immunoprecipitated material from the magnetic beads, 200 µl of elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1.0% SDS) was added and the beads were incubated at 65°C for 30 min and vortexed frequently. Magnetic beads were pelleted by centrifugation at 16,100 × *g* for 1 min and the supernatant containing the immunoprecipitated chromatin transferred to new tubes and incubated at 65°C for 9 hours to reverse the cross-linking.

To remove RNA, 8 µl of RNaseA (10 mg/ml) was added to the sample, mixed by inverting and incubated at 37°C for 2 hours. Finally, 4 µl of Proteinase K (20 m/ml) was added to the sample, mixed by

inverting, incubated at 55°C for 2 hours and the DNA extracted using a gel extraction kit (Qiagen). Quality of DNA and IP efficiency were determined by real-time PCR.

The following antibodies were used: rabbit anti-H4K10ac and anti-histone H4 (Cross lab), BB2 (kind gift from Keith Gull (Bastin et al., 1996)), HA clone 12cA5, H3 (Abcam 1791).

DNA preparation for high-throughput sequencing

The immunoprecipitated DNA was blunt-ended using a Quick Blunting Kit (NEB, E1201S). The instructions provided by NEB were scaled up and 57 µl of immunoprecipitated DNA was mixed with 7.5 µl 10× blunting buffer, 7.5 µl 1 mM dNTP mix and 3 µl blunt enzyme mix, incubated for 45 min at RT and purified using a PCR purification kit (Qiagen).

DNA was polyadenylated by mixing 41 µl of blunted DNA with 1 µl of 10 mM dATP, 5 µl NEB buffer 2 and 3 µl Klenow Fragment 3´→5´ exo– (NEB, M0212S) followed by a 45 min incubation at 37°C and DNA purification using a Mini elute PCR purification kit (Qiagen).

Sequencing adapters were ligated to the polyadenylated DNA by mixing 16 µl of DNA with 2 µl ligase buffer (NEB), 1 µl of sequencing adapter mix (1:10–1:20 dilution of adapter mix provided by Illumina) and T4 ligase (NEB, M0202T) followed by a 6 hour incubation at RT. The adapter concentration was adjusted depending on the amount of DNA

immunoprecipitated. Following adapter ligation, the DNA was purified using a Mini elute PCR purification kit (Qiagen) and PCR amplified as follows, PCR mix: 10 µl of purified DNA was mixed with 1 µl of dNTPs (10 mM stock, Invitrogen), 0.5 µl Phusion hot start DNA polymerase (NEB, F-540S), 10 µl of 5× Phusion HF buffer, 1 µl of PCR primer 1.1 (Illumina), 1 µl of PCR primer 1.2 (Illumina), 1.5 µl DMSO and 25.5 µl of H_2O; PCR temperatures: denaturation 30 sec at 98°C, subsequent denaturation at 98°C for 10 sec, annealing 30 sec at 65°C, elongation for 30 min at 72°C and a final elongation for 5 min at 72°C. The number of cycles was adjusted based on the amount of immunoprecipitated DNA and ranged from 17 to 20 cycles. Ten percent of the amplified DNA was analyzed on a 1.5% agarose gel to determine the average size of the immunoprecipitated DNA. The PCR-amplified DNA was purified using a Mini elute PCR purification kit (Qiagen), eluted in 12 µl and the concentration determined using a NanoDrop (Thermo Scientific) and Quant-iT (Invitrogen, Q-33120). DNA concentrations ranged from 40 ng/ µl to 150 ng/µl, and were generally higher when measured with NanoDrop compared to Quant-iT because NanoDrop measures the total concentration of single and double-stranded DNA whereas Quant-iT yields concentration for double-stranded DNA only. High-throughput sequencing was performed by The Rockefeller University Genomics Resource Center using a Solexa sequencer (Illumina).

Tag alignment

The sequenced DNA tags (32 bp or 36 bp) were annotated based on the *T. brucei* genome (version 4) using the Blast Like Alignment Tool (BLAT) (Kent, 2002) allowing ≤ 2 mismatches. Default parameters were used except for: tileSize (sets the size of match that triggers an alignment) = 10 bp, stepSize (spacing between tiles) = 5, minScore (sets the minimum score and consists of matches minus mismatches minus a gap penalty) = 15. The software to perform DNA tag annotation was set up by Xuning Wang, The Rockefeller University Department of Information Technology.

The number of hits per nucleotide position was determined using custom PERL scripts and displayed using MATLAB (The MathWorks).

Peptide pull-down

For nine peptide pull-down experiments (e.g. three different salt concentrations and three different peptides) 5×10^8 (conc.: 15×10^6) PF cells were pelleted at 1800 × *g* at 4°C for 10 min, washed with 20 ml PBS and resuspended in 2 ml suspension buffer (8% PVP (polyvinylpyrrolidone), 0.05% Triton X-100, 5 mM DTT, 1% commercially available mammalian protease inhibitor cocktail) and immediately processed with a pre-cooled Polytron (PTA head 10, setting 6) with 1 min bouts for a total of 6 min of Polytron treatment. Between Polytron bouts the lysate was cooled on ice. Next, cells were underlayed with 1 ml 0.3 M

sucrose in suspension buffer, mixed and centrifuged at 11,000 × *g* at 4°C for 20 min. The pellet was washed with cold PBS, and repelleted by centrifugation at 16,100 × *g* at 4°C for 10 min. The pellet was resuspended in IP buffer (20mM HEPES, pH 7.5, 150 mM NaCl, 1mM EDTA, 0.2% NP 40, 1% commercially available mammalian protease inhibitor), the NaCl concentration increased to 200–400 mM and the suspension sonicated for 5 min at maximum strength. Next, Na-butyrate (final conc.: 50 mM) and DNase I (0.002% wt/vol, Sigma) were added to the sample and incubated for 10 min at RT. To remove cell debris, the sample was centrifuged at 16,100 × *g* and 4°C for 10 min and the supernatant split among nine tubes. The sample can be split earlier if different NaCl concentrations are being used during the IP. Next, 2 µg of biotinylated peptide was added to each sample, followed by an incubation for 2.5 hours with rotation at 4°C, addition of 40 µl streptavidin-coated magnetic beads (Dynal) and another incubation for 1.5 hours with rotation at 4°C. Finally, the samples were washed five times with IP buffer (same NaCl concentration as during the IP), immunoprecipitated material was eluted by boiling the beads for 10 min in 15 µl 1× sample buffer and analyzed by western blotting. The following peptides were used: AKGKKSGEAKGSQKRQKKVLK-biotin, AKGK(ac)KSGEAKGSQKRQKKVLK-biotin, AKGKKSGEAK(ac)GSQKRQKKVLK-biotin and AKGK(ac)KSGEAK(ac)GSQKRQKKVLK-biotin.

Reagents obtained from other laboratories or former laboratory members

Plasmids

Constructs for the deletion of HAT3 (DH1 and DH2), for RNAi-mediated depletion of HAT1 (pHAT1RP^{SLRI}) and HAT2 (pHAT1RP^{SLRI}), and for expression of GFP-tagged HAT2 were generated in the David Horn lab. pLew20 was generated by Elizabeth Wirtz. pMot-tagging constructs were obtained from the Seebeck laboratory via Jeff deGrasse (Chait laboratory).

Antibodies

Polyclonal antibodies specific to *T. brucei* enolase and *T. brucei* RNA polymerase I were kind gifts from Nina Papavasiliou and Arthur Günzl, respectively. Monoclonal antibodies specific to the Ty1 peptide tag (Bastin et al., 1996) and *T. brucei* α-tubulin were gifts of Keith Gull.

Peptides

All peptides were synthesized by the Rockefeller University Proteomics Resource Center.

Cell Lines

The following cell lines were obtained from Joanna Lowell: BFJEL28, BFJEL22, BFJEL43, BFpFK3.4.5

Chapter 3: Histone H4K4 is masked during G1

Summary

Despite their early divergence from other eukaryotes, trypanosomes contain an extensively acetylated N-terminal histone H4 tail (da Cunha *et al.*, 2006; Janzen *et al.*, 2006a). Given the prominent role post-translational histone acetylation plays in transcription activation, I decided to study the function of two of the originally identified acetyl marks: H4K4ac and H4K10ac. Because the large sequence divergence of *T. brucei* histones prohibits the use of commercially available antibodies, I generated four highly specific antibodies to acetylated and unmodified H4K4 and H4K10. Immunofluorescence microscopy and western blots with cells sorted based on DNA content revealed a strong enrichment of unmodified H4K4 in S phase and suggested G1/G0-specific masking of the site due to a non-covalently binding factor, possibly the bromodomain-containing factor 2 (BDF2). My findings pertaining to H4K10ac are presented in Chapter 5.

Introduction

Histone modifications have been studied intensively in several eukaryotes and are implicated in many cellular processes, but little is known about their function in *T. brucei*. To date, the only histone modification in *T. brucei* linked to a phenotype is methylation of H3K76 (Janzen et al., 2006b), most likely homologous to yeast and human H3K79me. Depletion of the two histone methyltransferases responsible for H3K76 methylation, *tb*Dot1A and *tb*Dot1B, leads to cell cycle irregularities and a defect in repressing silent expression sites, respectively (Figueiredo et al., 2008; Janzen et al., 2006b). Generally speaking, most chromatin-related research in *T. brucei* has focused on understanding the role of chromatin in antigenic variation and, more specifically, on identifying factors responsible for repression of silent expression sites and de-repression of active expression sites.

During my graduate work, I was specifically interested in general regulation of transcription, particularly in factors contributing to a chromatin conformation that permits active transcription: histone acetylation, histone acetyltransferases, acetyl-binding proteins and histone variants.

When I began my thesis work, the only post-translational acetylations identified on the N-terminal tail of histone H4 in *T. brucei* were acetylations of H4K4 and H4K10 (Janzen et al., 2006a). This pair of histone acetylations shows some resemblance to deposition-related

diacetylation found on newly synthesized histone H4 in *Tetrahymena* (H4K4/K11), *Drosophila* (H4K5/K12) and human cells (H4K5/K12) (Chicoine et al., 1986; Sobel et al., 1995). If H4K4ac and H4K10ac only played a role in histone deposition, one would expect both sites to be acetylated at similar frequency. However, Edman data indicated that, in *T. brucei,* 80% of H4K4 sites are acetylated compared to only 10% of H4K10 sites (Janzen et al., 2006a), suggesting that these two marks may have two independent functions.

In this chapter, I describe the use of custom-made antibodies to characterize H4K4 using IF microscopy, FACS (Fluorescence Activated Cell Sorting) sorting and western blotting.

Results

Generation of specific antibodies

It was first necessary to generate new antibodies to both the acetylated and unmodified sites of H4K4 and H4K10. A fifth antibody that recognized the N-terminal tail of histone H4 regardless of the acetylation state of H4K4 was fortuitously obtained after immunization with the same peptide that had been used to raise an antibody against unmodified H4K4. This antibody was therefore considered a general H4 antibody. After failing to obtain antisera specific for unmodified H4K10 peptides in rabbits, I succeeded in raising specific antisera in chickens. All other antibodies were generated in rabbits. Antibody specificity was

tested by pre-incubation with peptide competitors before western blotting (Figure 3.1 and Figure 3.2) or IF analysis (data not shown). The H4K4-unmodified, H4K4ac, H4K10-unmodified and H4K10ac antibodies showed affinity only for the peptides they were raised against and not for similar peptides bearing a different acetylation state. The general histone

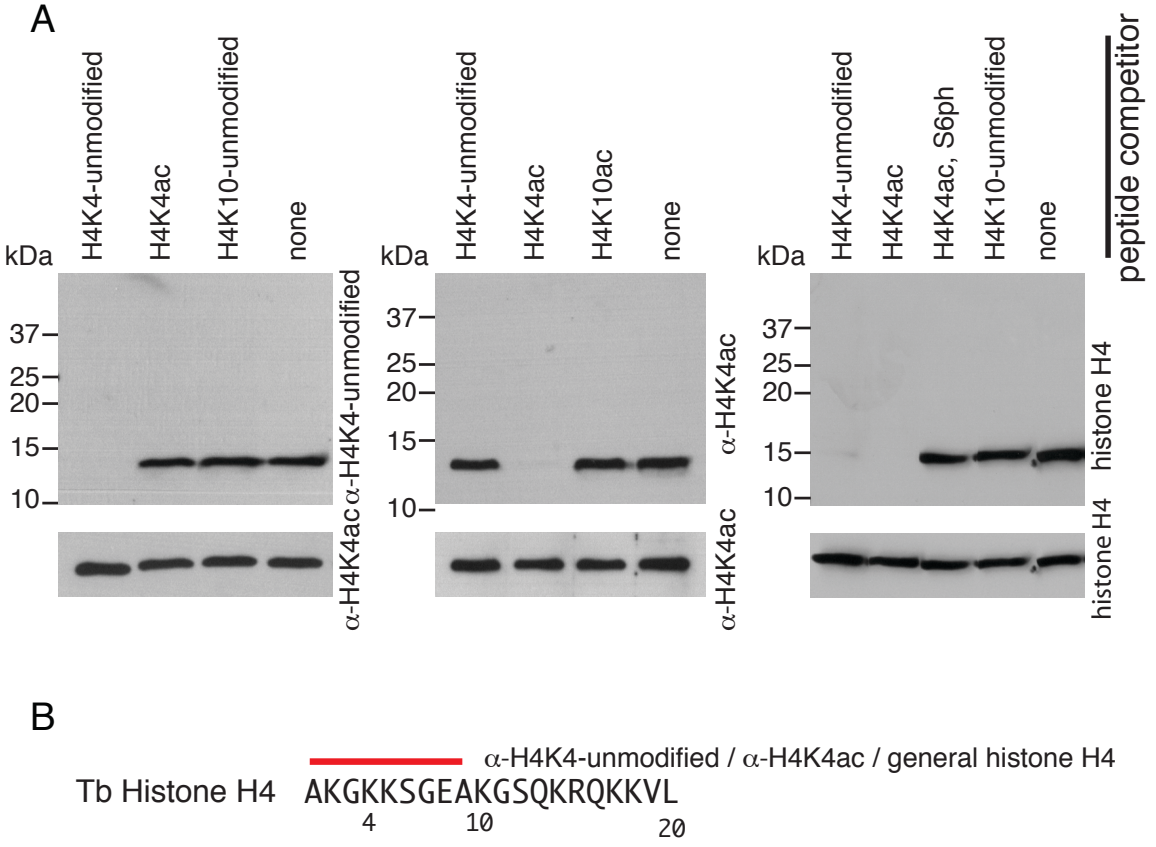


Figure 3.1. Peptide competition to evaluate antibody specificity for H4K4
(A) Whole cell lysates were analyzed by western blotting with antibodies that had been preincubated with different peptide competitors. Peptide competitors used are shown above. To confirm equal loading, membranes were stripped and reprobed with α-H4K4ac or α-histone H4 antibodies without peptide competitors.
(B) Sequence of N-terminal tail of trypanosome histone H4. The red bar marks the region used as antigen for rabbit immunization.

H4 antibody, on the other hand, showed high affinity for both the unmodified and the acetylated peptide. All western blots were done with whole cell lysates and no cross-reactivity of the antibodies with other histones was observed (Figure 3.1 and Figure 3.2).

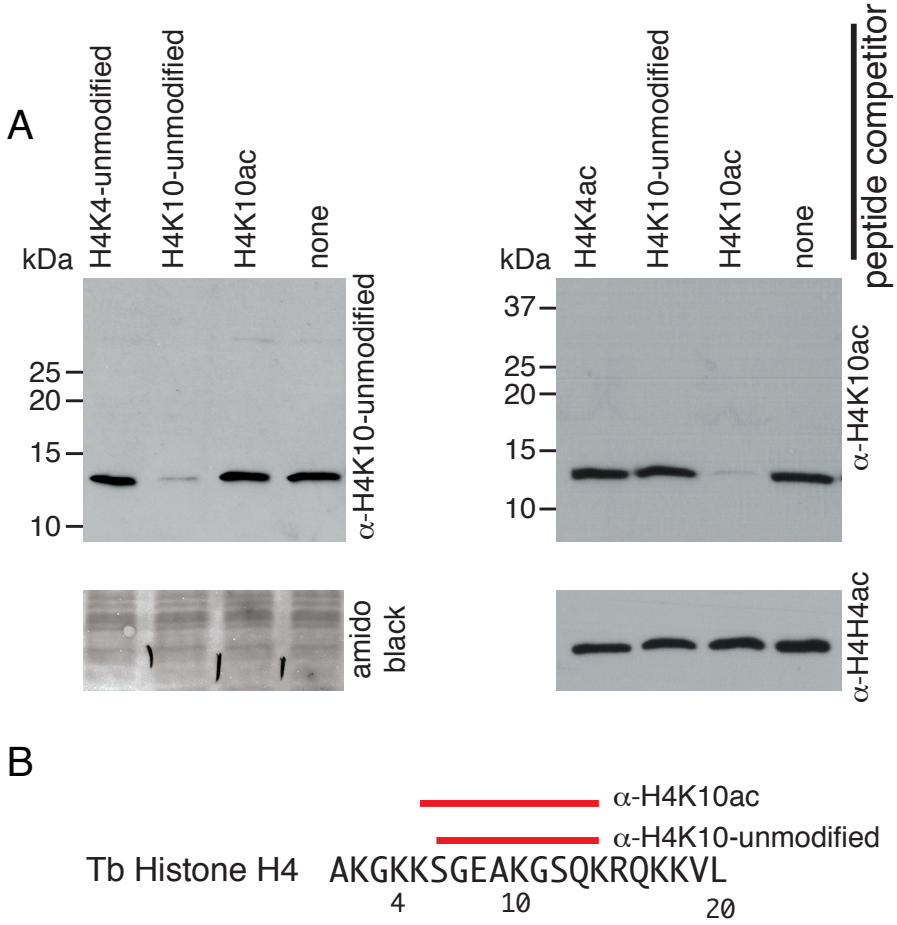


Figure 3.2. Peptide competition to evaluate antibody specificity for H4K10
(A) Whole cell lysates were analyzed by western blotting with antibodies that had been preincubated with different peptide competitors. Peptide competitors used are shown above. To confirm equal loading, membranes were stained with amido black or stripped and reprobed with an α-H4K4ac antibody without peptide competitors.
(B) Sequence of N-terminal tail of trypanosome histone H4. The red bar marks the regions used as antigen for rabbit and chicken immunization.

H4K4 epitope masking in cells during G1/G0

IF analysis using DeltaVision deconvolution microscopy and the antibody specific for H4K4ac revealed a punctate pattern throughout the nucleus, excluding only the nucleolus (Figures 3.3A–C).

Although antibodies to both the unmodified and acetylated H4K4 site worked in IF microscopy, neither antibody reacted with cells in G1/ G0 (Figures 3.3A–F). In contrast, the general H4 antibody bound throughout the cell cycle (Figures 3.3G–I), suggesting that the H4K4 site may have been specifically blocked in G1/G0, either by another covalent modification close to H4K4 or by a factor binding to that site. Sites for potential covalent modification are present at K2, K5, and S6, but no modifications have been detected at these sites in PF (Janzen *et al.*, 2006a). Only very minor (<10%) levels of acetylation have been detected at K2 and K5 in BF (Mandava *et al.*, 2007), ruling out the possibility that these modifications could completely block binding of an antibody to H4K4 throughout G1/G0. Neither study investigated the possible phosphorylation of H4S6. To test whether phosphorylation of H4S6 could explain my observations, I performed peptide competition experiments to test if my general histone H4 antibody could bind to a synthetic doubly modified phosphoacetylated peptide (K4ac, S6ph). The experiment demonstrated that the general H4 antibody did not bind the phosphorylated peptide (Figure 3.1A). Because IF studies had shown that

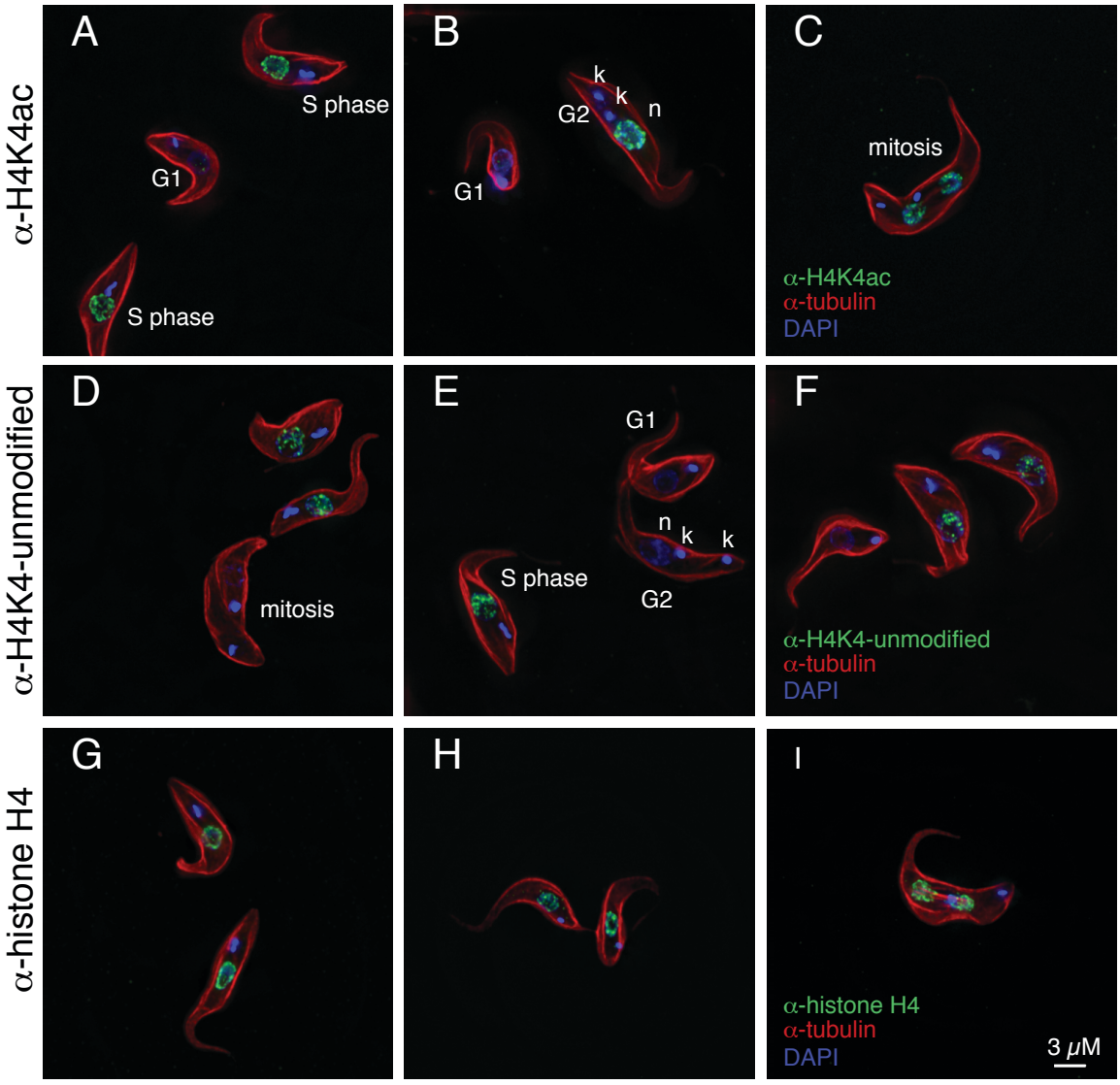


Figure 3.3. H4K4ac is not detectable in G1 cells by IF.

Shown are representative IF images of PF cells stained with:

(A–C) α-H4K4ac (green), α-tubulin (red) and DAPI (blue)

(D-F) α-H4K4-unmodified (green), α-tubulin (red) and DAPI (blue)

(G–I) α-histone H4 (green), α-tubulin (red) and DAPI (blue).

Cell containing one nucleus (n) and one elongated kinetoplast (k) are in S phase, cells containing one or two nuclei and two kinetoplasts are in in G2/M and cells containing one nucleus and one kinetoplast are in G1/G0.

the general H4 antibody bound to cells in G1/G0 (Figures 3.3G–I), I concluded that phosphorylation of H4S6 is not responsible for blocking the H4K4 site during G1/G0. I therefore investigated the possibility of blockage by non-covalent interactions.

If H4K4 were masked by a non-covalently interacting factor *in vivo*, I reasoned that analysis by SDS-PAGE would disrupt this interaction and the H4K4 site in G1/G0 cells would become accessible to antibodies. It is not trivial to synchronize the cell cycle in *T. brucei* (Forsythe et al., 2009; Chowdhury et al., 2008), so cells were sorted according to their DNA content (Figure 3.3A). To avoid potential problems reversing cross-links arising from fixation, I sorted unfixed DyeCyclin Orange-stained cells and then analyzed them by western blotting. Under these denaturing conditions, the H4K4ac antibody bound to H4 from cells in G1/G0 (Figure 3.4B), suggesting that masking of H4K4 in the IF analysis was most likely caused by some non-covalently-bound factor. Prior to western blot analysis, all fractions were FACS-analyzed again to confirm that the initial sort had generated homogenous cell populations (Figure 3.4A).

Bromodomain-containing factor 2 is upregulated in G1 cells

To identify the factor responsible for masking H4K4ac during G1/ G0 phase, I searched the *T. brucei* genome for proteins with putative acetyl-binding domains. The best-characterized acetyl-binding domain is the so-called bromodomain (Winston and Allis, 1999), and the *T. brucei*

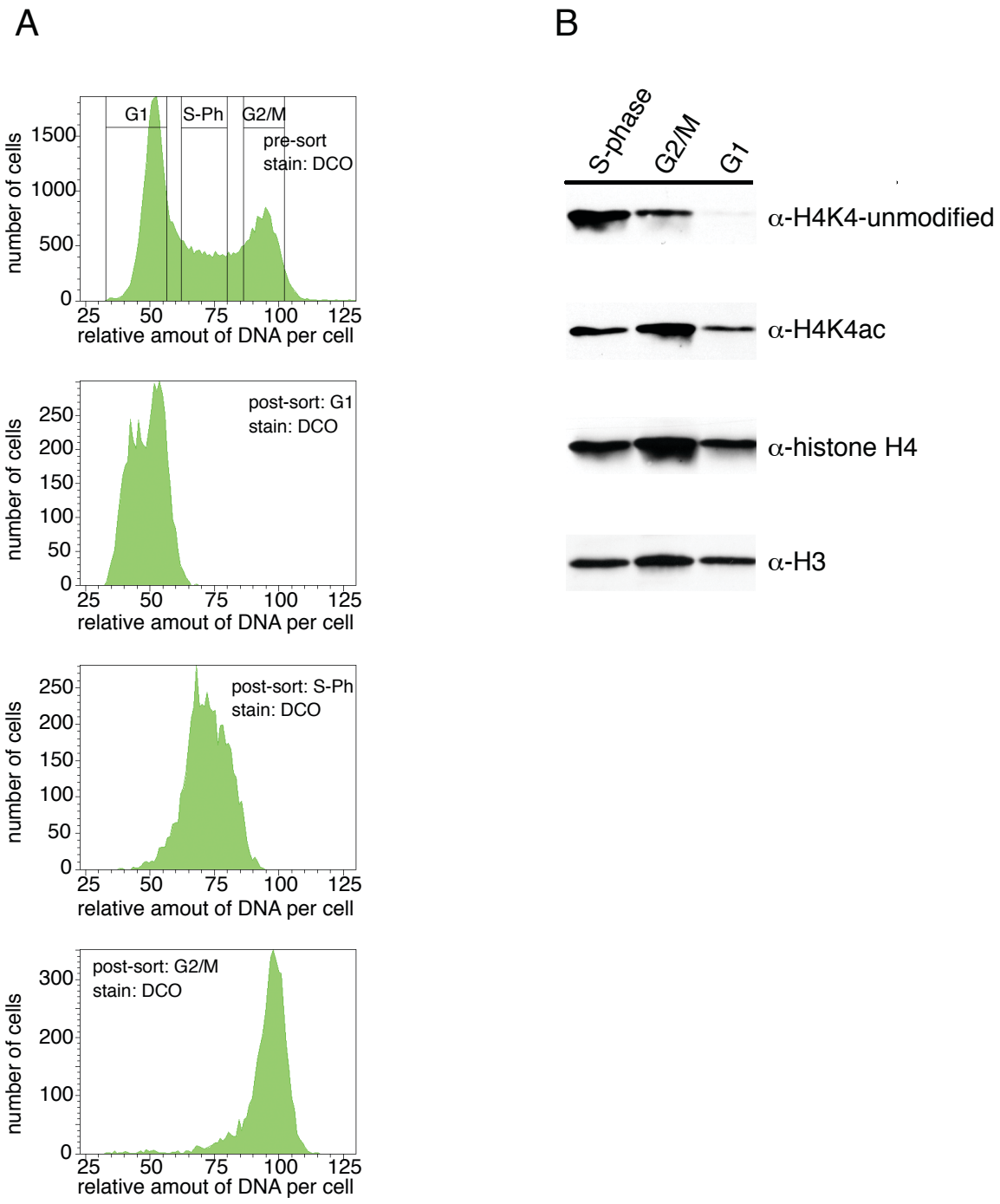


Figure 3.4. α-H4K4ac binds to cells in G1/G0 under denaturing conditions

Unfixed DCO-stained cells were sorted based on DNA content.

(A) Input cells (top panel) and sorted cells were analyzed by FACS.

(B) Cell lysates from sorted cells were analyzed by western blotting. The membrane was stripped repeatedly and reprobed with the antibodies listed.

genome encodes six putative BDFs. These trypanosome BDFs range in size from 221–660 amino acids and vary in sequence conservation from E-value 4.1 x 10^{-23} to 7.7 x 10^{-10} based on superfamily database v1.69 (Gough et al., 2001) (Figure 1.10).

Further analysis was done on the four most highly conserved candidates (BDF1–4) and on a single candidate with two potential bromodomains (BDF5), which were each tagged with an HA epitope, and their cellular localization determined by IF microscopy. BDFs 2, 3 and 5 localized to the nucleus, but only the expression of BDF2 appeared cell cycle regulated. Its expression was visibly increased in G1/G0 cells (Figure 3.5). This inverse correlation between H4K4ac and BDF2 staining indicated that BDF2 could be responsible for the G1/G0 phase-specific blocking of the H4K4ac mark. Further analyses, including peptide pull-downs or fluorescence anisotropy binding measurements, will be needed to establish a direct interaction between H4K4ac and BDF2.

Unmodified H4K4 is strongly enriched in S phase cells

In addition to demonstrating covalent blocking of H4K4ac during G1, western blot analysis of cells sorted for DNA content revealed a strong enrichment of unmodified H4K4 in cells in S phase (Figure 3.4B). This seemed to contradict the IF results, in which only cells with an elongated kinetoplast showed a signal with the unmodified H4K4 antibody (Figures 3.3D–F). Such cells had previously been classified as

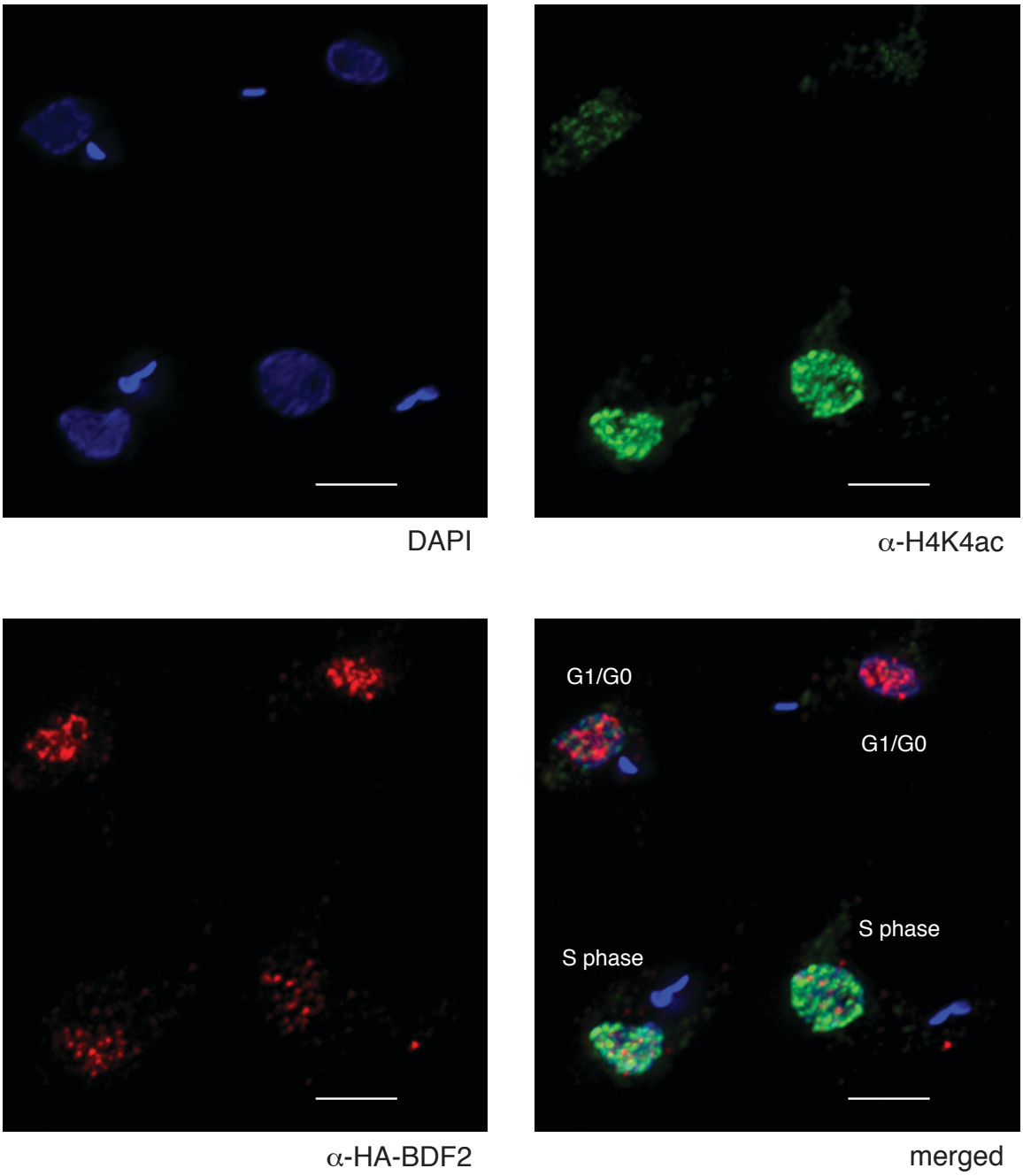


Figure 3.5. BDF2 expression is increased during G1/G0
Representative IF co-localization analysis of BDF2-HA (red), DAPI (blue) and H4K4ac (green). Bar = 3 μm.

cells in G2 (Woodward and Gull, 1990). To address this contradiction, I decided to reinvestigate the timing of kinetoplast division in relationship

to the nuclear cell cycle. Such an analysis appeared relevant considering that the cells used by Woodward and Gull had a population doubling time of 8.5 hours compared to 12.5 hours for the cells used here. I found that kinetoplast division, visualized by an elongated kinetoplast, is initiated during nuclear S phase (for a detailed description of quantitative DAPI imaging see Chapter 7). Therefore, the IF analysis is consistent with the western blot analysis of cells sorted for DNA content: the unmodified H4K4 site is strongly enriched in cells in S phase.

Discussion

Cell cycle dependent regulation of H4K4 acetylation

IF microscopy indicated that antibodies to both unmodified H4K4 or H4K4ac cannot bind to histones in cells in G1/G0. G1-specific blockage by nearby covalent modification seemed unlikely, based on studies that failed to identify such modifications. Instead, this observation can be attributed to epitope masking by a non-covalently binding factor. This finding suggests that epitope masking, in fixed cells that are generally used for IF analysis, can pose a serious problem when interpreting IF data. This is especially true for studies of histone modification in which antibodies are only available against the modified but not against the unmodified sites, which would otherwise provide a valuable control. Furthermore, my approach showed that cell cycle-

specific masking of an epitope could be revealed by taking advantage of SDS-PAGE after cell cycle-dependent cell sorting.

Currently, I can only speculate about the nature of the factor blocking the H4K4 site during G1, but correlative evidence points to BDF2. Of all five BDFs analyzed, only BDF2 showed a cell cycle regulated expression pattern. Furthermore, BDF2 was markedly upregulated during G1 phase, the same cell-cycle stage during which covalent blockage of H4K4ac was observed. However, to substantiate this correlation, analysis of direct binding interactions would be necessary.

Besides cell cycle-specific blockage of H4K4, IF analysis revealed a strong enrichment of the unmodified H4K4 residue in cells with an elongated kinetoplast. In a more detailed analysis, presented in Chapter 7, I showed that kinetoplast division, visualized by an elongated kinetoplast, occurs during nuclear S phase. This observation is in agreement with my western blot data that showed a strong enrichment of unmodified H4K4 signal for cells in S phase.

Chapter 4: HAT3 acetylates histone H4K4

Summary

Given the relatively small number of histone-modifying enzymes, *T. brucei* represents an excellent system in which to investigate the function of individual histone modifications and histone-modifying enzymes. Using cell lines depleted for different histone-modifying enzymes, I was able to show that HAT3 is responsible for acetylation of histone H4K4. In addition, treatment of cells with the protein synthesis inhibitor cycloheximide led to an immediate loss of unmodified H4K4. As HAT3 is located inside the nucleus, these findings suggest that newly synthesized histone H4 is imported rapidly into the nucleus, where H4K4 is acetylated.

Introduction

It has been proposed that histone PTMs constitute a 'histone-code' (Strahl and Allis, 2000) of epigenetic information for transcription regulation and chromatin structure. Deciphering this code in higher eukaryotes has been complicated by the large number of histone PTMs and histone-modifying enzymes and a high degree of redundancy. Research on lower eukaryotes with a more concise chromatin composition might more easily reveal the critical functions of individual histone PTMs.

One such eukaryote, *Tetrahymena thermophila*, is the subject of intense study. This ciliated protozoan contains two nuclei – a transcriptionally inactive germline micronucleus and a large polyploid somatic macronucleus that is transcriptionally active (Yao and Gorovsky, 1974; Gorovsky, 1973). *Tetrahymena* has been an invaluable model for understanding chromatin structure and function. Its transcriptionally active macronuclei served as a source for hyperacetylated histones and for the purification of the first histone acetyltransferase (Brownell et al., 1996).

To better understand the complex network of histone-modifying enzymes, histone PTMs, and factors binding to these modifications, it will be important to look towards other lower eukaryotes and to take advantage of their less complex histone-modifying machinery. The *T. brucei* genome contains five putative KATs (HAT1–3, Elp3a–b) and five

nuclear HDACs (HDAC 1–4, Sir2rp1) compared to 14 KATs and 18 HDACs in humans (see Chapter 1). Members of the David Horn laboratory (London School of Hygiene and Tropical Medicine, UK) have generated a set of cell lines that are either knock-out mutants of any non-essential KAT and HDAC, or that allow inducible RNAi-mediated depletion of any essential KAT and HDAC. Therefore, I decided to use the cell lines generated in the Horn laboratory in conjunction with my newly generated antibodies to identify enzymes responsible for acetylation and deacetylation of H4K4 and H4K10 (see Chapter 5).

Results

HAT3 acetylates histone H4K4

The roles of the non-essential HAT3, HDAC2, HDAC4 and Sir2rp1 genes were tested in homozygous null cell lines. All other identified histone-modifying enzymes (HAT1, HAT2, Elp3a, Elp3b, HDAC1 and HDAC3) were tested in cell lines that allowed inducible RNAi-mediated depletion of the respective enzyme (Alsford et al., 2007; Ingram and Horn, 2002). Western blot analysis of the various cell-lines indicated that HAT3 is responsible for H4K4 acetylation in BF and PF (Figures 4.1A and B). Eighty percent of H4K4 is normally acetylated in PF *T. brucei* (Janzen et al., 2006a). The residual H4K4ac signal in the HAT3$^{-/-}$ cells may be caused by the ability of the other HATs to inefficiently acetylate H4K4. Although I cannot exclude the possibility that my H4K4ac antibody

cross-reacted at lower affinity with other sites on H4, this seems unlikely, as I did not see cross-reactivity with unmodified H4K4, H4K10ac, or with any other histone (Figure 3.1A). As expected, the unmodified H4K4 signal was greatly increased in the HAT3$^{-/-}$ cell-lines (Figures 4.1A and B).

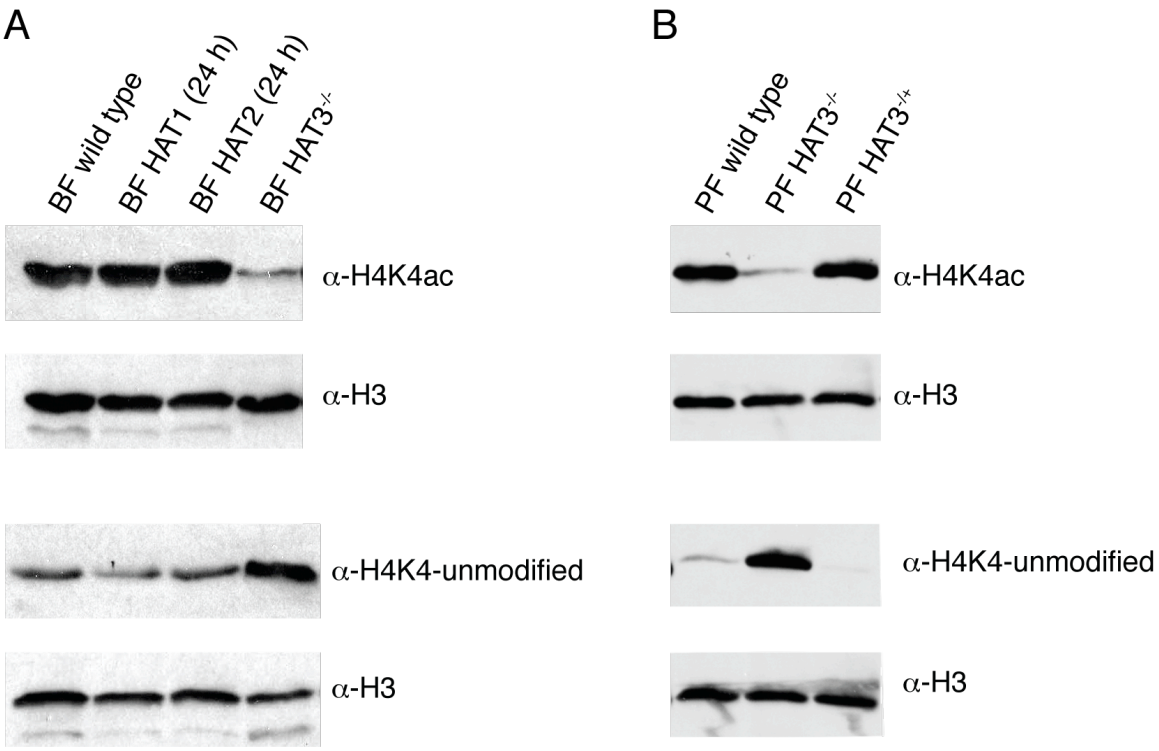


Figure 4.1. HAT3 acetylates H4K4 in BF and PF

(A) Whole cell lysates from BF deficient in HAT3 or depleted for HAT1 and HAT2 were analyzed by western blotting using antibodies specific for acetylated or unmodified H4K4. HAT1 and HAT2 specific RNAi was induced 24 h before harvest. Northern blotting to verify depletion of HAT1 and HAT2 transcripts was performed by T. Kawahara (Kawahara et al., 2008). All blots were stripped and reprobed with α-histone H3 (Abcam, 1791) to control for equal loading.

(B) Whole cell lysates from PF WT, PF HAT3$^{-/+}$ and PF HAT3$^{-/-}$ were analyzed by western blotting using antibodies specific for acetylated or unmodified H4K4.

Several sites on the C-terminal tail of histone H2A and the N-terminal tail of histone H4 can be acetylated (Janzen et al., 2006a; Mandava et al., 2007). To assess the specificity of HAT3 for histone H4K4, I purified histones from HAT3$^{-/-}$ cells and used Edman degradation and tandem mass spectroscopy to measure levels of acetylation on other sites known to exist in the acetylated form. Except for H4K4, no differences could be detected in the acetylation patterns of histones isolated from wild type and HAT3$^{-/-}$ cells.

Because HAT3 is found inside the nucleus (Kawahara et al., 2008) and unmodified H4K4 was strongly enriched in S phase (Figures 3.3D–F and Figure 3.4B), I speculated that – in contrast to other eukaryotes – newly synthesized histones in *T. brucei* remain unmodified at this site until they have been imported into the nucleus. To test this idea, I separated nuclear and cytosolic fractions (Rout and Field, 2001). No unmodified or acetylated H4 could be detected in the cytoplasmic fraction, suggesting that H4 is very rapidly imported into the nucleus (Figure 4.2). Purity of the fractions was confirmed with antibodies specific for cytoplasmic enolase and nuclear RNA pol I.

Unmodified H4K4 decreases rapidly when protein synthesis is inhibited

Depletion of an HDAC should lead to an accumulation of the modified lysine and a reduction in the levels of unmodified lysine.

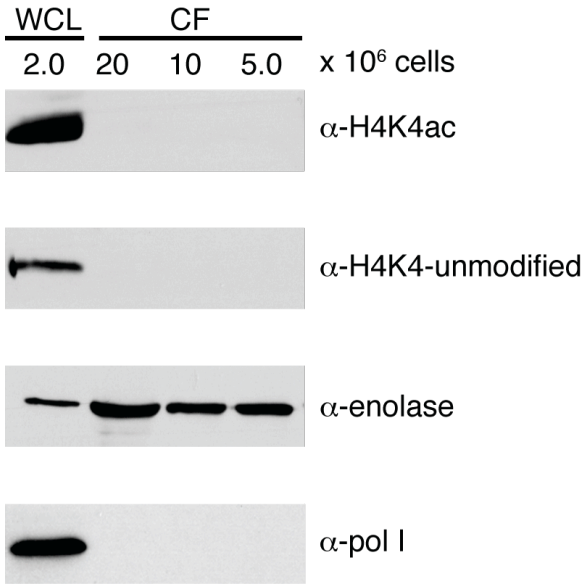


Figure 4.2. No histones can be detected in the cytosol
Whole-cell lysates (WCL) and cytoplasmic fractions (CF) from up to 2×10^7 cells were analyzed with antibodies specific for H4K4ac, H4K4-unmodified, enolase and RNA polymerase I.

However, in *T. brucei*, none of the knockout cell lines for HDAC2, HDAC4 or Sir2rp1 showed a decrease in the level of unmodified H4K4 (Figure 4.3A), suggesting that none of these enzymes are responsible for H4K4ac deacetylation. I therefore tested the effect of RNAi-mediated depletion of the essential HDAC1 and HDAC3, or treatment with the HDAC inhibitor Trichostatin A (TSA). All of these treatments were accompanied by a decrease in the unmodified H4K4 signal (Figure 4.3B). These results are unexpected because they imply that both essential HDACs are responsible for deacetylating H4K4ac. I suspected that the decrease in unmodified H4K4 was an indirect consequence of reduced cell growth. In

fact, as one would expect for the depletion of an essential enzyme, cell growth was slowed by depletion of HDAC1 or HDAC3 or inhibition by TSA (data not shown). As unmodified H4K4 is predominantly found in S phase, presumably representing newly synthesized histones, the loss of the unmodified H4K4 signal could be attributed to reduced cell growth and decreased histone synthesis, rather than a direct consequence of HDAC depletion.

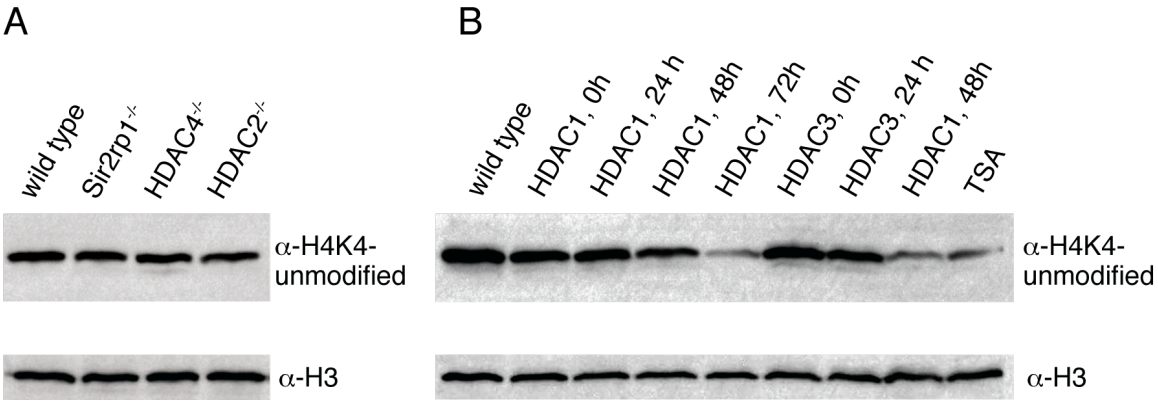


Figure 4.3. Deletion of HDACs had no specific effect on levels of unmodified H4K4

(A) Whole cell lysates from BF lacking both alleles of HDAC2, HDAC4 or Sir2rp1 were analyzed by western blotting using an antibody specific for unmodified H4K4. To confirm equal loading, the blot was stripped and reprobed with α-histone H3.

(B) Whole cell lysates from BF specifically depleted for HDAC1 or HDAC3 or treated with Trichostatin A (TSA) were analyzed by western blotting with an antibody specific for unmodified H4K4. HDAC1 and HDAC3 specific RNAi was induced 0, 24, 48, or 72 hours before harvest. Northern blotting to verify depletion of HAT1 and HAT2 transcripts was performed by Sam Alsford, London School of Hygiene and Tropical Medicine, UK, data not shown. To confirm equal loading, the blot was stripped and reprobed with α-histone H3.

To test this hypothesis, translation was inhibited by treating cells with cycloheximide (Hanas et al., 1975) over a period of 60 min and cell lysates were analyzed by western blotting (Figure 4.4A). Changes in unmodified H4K4 signal were quantified after normalization for differences in loading based on measurements with H3 antibody (Figure 4.4B). H3 was used instead of H4 to avoid complication arising from incomplete antibody stripping. Quantification revealed a remarkably rapid loss of the unmodified H4K4 signal. Previous studies have shown that untreated PF contain 20% unmodified H4K4 (Janzen et al., 2006a). I observed a 50% decrease to a total of 10% unmodified H4K4 when cells were harvested immediately after cycloheximide addition, meaning that cells were exposed to the inhibitor for ~4 min during centrifugation. The unmodified H4K4 signal declined to 1.3% of total H4 after 60 min of cycloheximide treatment. These results confirm that when protein synthesis is stopped, the levels of unmodified H4K4 decrease very rapidly. Therefore the effects observed upon RNAi-mediated depletion of HDAC1 and HDAC3 are most likely nonspecific. Such dramatic reduction of unmodified H4K4 levels further suggested that this site is very rapidly acetylated after histone synthesis.

Partition experiments detected no cytoplasmic accumulation of unmodified H4K4 in HAT3$^{-/-}$ cells (Figure 4.4C), which indicates that acetylation of H4K4 is not required for retention of H4 inside the nucleus.

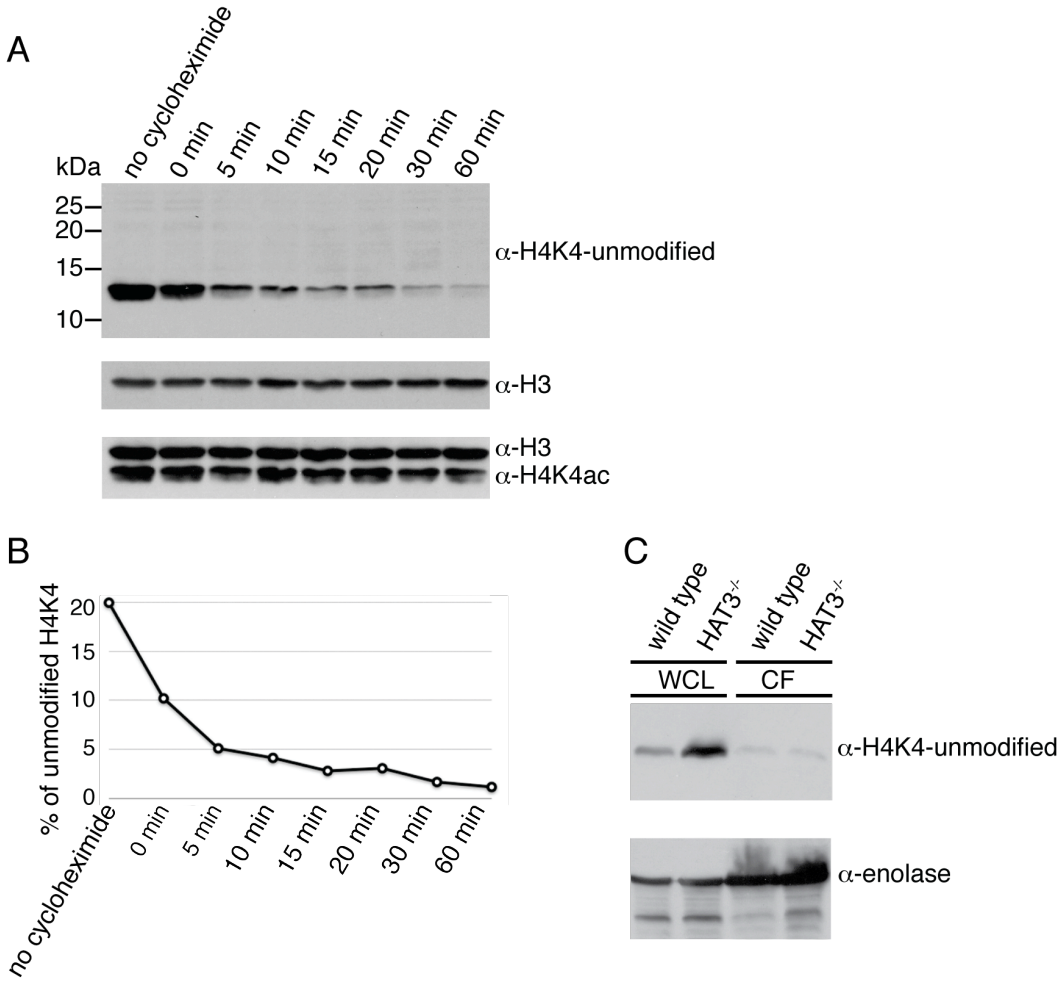


Figure 4.4. Inhibition of protein synthesis led to a rapid loss of unmodified H4K4 signal

(A) Western blot analysis of the unmodified H4K4 signal in cell lysates from control and cycloheximide-treated cells. Time points listed refer to start of cell harvest, which included an additional ~4 min of centrifugation before a gel-loading buffer was added and lysates were boiled. The blot was stripped and equal loading was confirmed with α-histone H3.

(B) Quantification of western blot signal was based on ECL plus (Amersham) luminescence as measured by VersaDoc Gel Imager (Bio-Rad). Percentages are based on published data (Janzen et al., 2006). Luminescence from histone H3 was used to normalize the H4K4-unmodified signal for differences in loading.

(C) Whole-cell lysates (WCL) and cytoplasmic fractions (CF) from 5×10^6 and 2.75×10^7 cells respectively, were analyzed with antibodies specific for unmodified H4K4 and enolase.

Discussion

Is *T. brucei* H4K4 functionally equivalent to H4K5 in other organisms?

In this chapter I showed that the large majority of H4K4 is acetylated by the non-essential HAT3. Thus far, no phenotype has been detected for HAT3$^{-/-}$ in BF or PF cultured in liquid medium or grown in rodents (GAMC, unpublished observations). Acetylation of H4K4 by a non-essential enzyme would be consistent with observations in other eukaryotes that H4K5 acetylation may not be essential but may be somehow evolutionarily advantageous. Are there other similarities that would suggest a functional equivalence between H4K4 in *T. brucei* and H4K5 in other organisms? The best-studied and most conserved role of H4K5 may be in deposition of newly synthesized histones (reviewed in Brownell and Allis, 1996). The majority of histone synthesis occurs during S phase (Wu and Bonner, 1981) and acetylation of H4 at K5 and K12 happens in the cytosol. This diacetylation is evolutionarily conserved in flies, humans, and in *Tetrahymena*, where the homologous residues K4 and K11 are used (Chicoine et al., 1986; Sobel et al., 1994; Sobel et al., 1995). Diacetylation of newly synthesized H4 is thought to occur by cytosolic KATs, traditionally defined as type-B KATs (Parthun et al., 1996; Brownell and Allis, 1996). The mechanistic link between cytosolic H4 acetylation and histone deposition stems from observations that the chromatin assembly factor 1 (CAF-1) selectively deposits acetylated

histone H3-H4 heterodimers onto newly replicated DNA (Verreault et al., 1996; Kaufman et al., 1995). Despite the conservation of these marks, *in vitro* experiments in yeast suggest that both are dispensable for nucleosome assembly onto newly replicated DNA (Ma et al., 1998).

As in other eukaryotes, histone synthesis in *T. brucei* reaches its peak during S phase (Ersfeld et al., 1996; Garcia-Salcedo et al., 1999), but no putative cytoplasmic HATs have been identified (Ivens et al., 2005). In agreement with these observations, unmodified H4K4 is strongly enriched in S phase, probably representing newly synthesized histones. It is difficult to speculate about histone deposition in *T. brucei* as many of the major components in human chromatin assembly, like the CAF1 complex, have no obvious homologues in trypanosomes. One protein that does seem to be conserved is ASF1 (anti-silencing function 1, Genbank accession no. Q4GZF6), a chaperone implicated in transport of the H3-H4 heterodimer into the nucleus (Tyler et al., 1999). In humans, ASF1 binds to the core region of the H3-H4 dimer and acetylation of the tails is not necessary for successful histone deposition (Natsume et al., 2007). Should ASF1 play a role in histone translocation in *T. brucei*, it would probably not require acetylation of the H4 tail.

Both microscopic (Taddei et al., 1999) and biochemical analysis (Jackson et al., 1976) in human cells indicate that H4 on newly assembled chromatin remains acetylated for 20–30 min before HDACs establish steady-state intermediate acetylation levels. Upon entry into

mitosis and the onset of chromatin condensation, histones become hypoacetylated, which is most clearly manifested in the loss of H4K5ac (Kruhlak et al., 2001). Substitutions of all N-terminal lysines of histone H4 with glutamines, to mimic the hyperacetylated state, display a marked delay in progression through G2 and M phases in yeast (Megee et al., 1990). It has been suggested that this block is caused by a defect in chromatin condensation. Alternatively, it has been proposed that deacetylated lysine plays a role in a specific checkpoint control mechanism, since the insertion of a single additional deacetylated lysine rescues the G2/M block (Megee et al., 1990). Further analysis indicated that the lysine-mediated G2/M block is controlled by the RAD9-dependent DNA damage checkpoint (Megee et al., 1995). Cell cycle control in *T. brucei* differs significantly from other eukaryotes, and different checkpoint mechanisms seem to control progression (reviewed in Hammarton, 2007). No RAD9-dependent checkpoint pathway has been identified in *T. brucei*, but the observation that depletion of the non-essential HDAC4 leads to a delay of G2/M phase suggests that a similar pathway may exist (Ingram and Horn, 2002).

None of the five nuclear HDACs identified in *T. brucei* had a specific effect on H4K4 acetylation, and levels of unmodified H4K4 were strongly diminished soon after inhibiting protein synthesis. These observations suggest that newly synthesized histones are the only source of unmodified H4K4 and that newly synthesized histones are rapidly

acetylated after import into the nucleus. Other, less likely, explanations for the decrease in unmodified H4K4 levels would be that inhibition of protein synthesis leads to an inhibition of HDACs or to strong upregulation of KATs.

I did not observe a decrease in H4K4ac after S phase, not even at the onset of mitosis. *T. brucei* undergoes a closed mitosis during which no visible chromatin condensation occurs (Ogbadoyi et al., 2000), so changes in acetylation may not be necessary from a structural point of view. The question of whether lack of chromatin condensation during mitosis is cause or consequence of constant histone acetylation may be addressed in the future using HAT3$^{-/-}$ cells.

Conclusions

I have shown that HAT3 is responsible for acetylation of H4K4 and the evidence suggests that H4K4ac may not be actively deacetylated. My findings further indicate that, in contrast to other eukaryotes, *T. brucei* may not diacetylate newly synthesized histones in the cytosol. Hyperacetylated chromatin is generally considered to be transcriptionally active and less densely packed, leaving DNA more accessible to interacting factors during transcription. The high level of H4K4 acetylation could serve to keep chromatin in an open conformation, which appears to be the general situation in *T. brucei* (Navarro et al., 1999).

Chapter 5: H4K10ac, BDF3 and four histone variants mark the boundaries of PTUs

Summary

Atypical for a eukaryote, genes transcribed by RNA pol II in *T. brucei* are arranged in PTUs. With one exception, no RNA pol II promoter motifs have been identified, and how transcription is initiated remains an enigma. A newly developed ChIP-seq protocol to examine the genome-wide distribution of chromatin components showed that histones H2AZ, H2BV, the K10-Ac form of H4, and BDF3 are significantly enriched at probable RNA pol II TSSs. Furthermore, these marks were used to identify more than 60 previously unanticipated TSS candidates. Co-IP experiments with tagged H2A, H2AZ, H2B, and H2BV indicated that less H3 and H4 co-immunoprecipitated with variant histones compared to core histones suggesting that nucleosomes containing H2AZ and H2BV are less stable than canonical nucleosomes. A classic promoter motif could not be identified but computational analysis revealed the presence of long guanine-runs at probable TSSs. Apparently unique to trypanosomes, additional histone variants H3V and H4V are enriched at probable RNA pol II TTSs. These findings suggest that destabilization of nucleosomes by histone variants is an evolutionarily ancient and general mechanism of transcription initiation, demonstrated in an organism in which general pol II transcription factors have been difficult to identify.

Introduction

RNA pol II promoters and associated transcription factors have been elusive in *T. brucei*, with the exception of the spliced-leader RNA promoter (Das et al., 2005; Schimanski et al., 2005). Hence, how most RNA pol II transcription is initiated in *T. brucei* remains a fundamental question. It has been suggested that RNA pol II transcription can be initiated from a site lacking an apparent promoter as long as transcription occurred at a nearby site (McAndrew et al., 1998). This led the authors to suggest that an 'open' chromatin structure generated by another polymerase at a nearby site, may be sufficient to allow RNA pol II transcription to initiate.

Genome-wide analyses in several different organisms indicate that the extent of chromatin-mediated DNA compaction varies and that regulatory sequences are generally located in less compact euchromatic regions, hypersensitive to DNAse treatment (Weintraub and Groudine, 1976; Shi et al., 2009; Boyle et al., 2008).

The dynamic nature of chromatin packing is mediated by multiple processes (see Chapter 1, reviewed in Henikoff, 2008). DNA sequence can affect nucleosome positioning, keeping regulatory DNA free of nucleosomes or facilitating nucleosome eviction. Nucleosome remodeling can lead to a high turnover of histones in actively transcribed genes, and replacement of canonical histones by histone variants can create less stable nucleosomes. Genome-wide analyses in yeast, human cells and

Drosophila have revealed that the histone variants H2AZ and H3.3 are enriched in proximity to RNA poll II promoters (Mito et al., 2005; Raisner et al., 2005; Barski et al., 2007). Furthermore, the structure of nucleosomes containing both histone variants H2AZ and H3.3 is less stable than of nucleosomes composed of canonical histones (Jin and Felsenfeld, 2007). This finding suggests that nucleosomes containing H2AZ/H3.3 are more easily evicted from promoters during the initiation of transcription than canonical nucleosomes. Finally, post-translational histone modification, especially of the N-terminal tails of histone H3 and H4, can influence nucleosome stability. For example, acetylation of the H4K16 has a negative affect on the formation of 30-nm chromatin fibers and generation of higher-order chromatin structures (Shogren-Knaak et al., 2006). Besides this structural role, acetylation of the H4 tail provides a binding platform for acetyl-binding proteins, including BDFs. The function of BDFs ranges from nucleosome remodeling to recruitment of histone-modifying enzymes or transcription factors.

Although the primary sequence of trypanosome core histones has diverged significantly from those of other eukaryotes, trypanosomes do contain an extensively acetylated H4 tail, several putative BDFs (Janzen et al., 2006a; Mandava et al., 2007; da Cunha et al., 2006; Ivens et al., 2005) and four histone variants (Lowell et al., 2005; Lowell and Cross, 2004). Given the lack of an obvious RNA pol II promoter sequence motif, the observation that divergent-SSRs contain somewhat elevated levels of

H4 acetylation in *T. cruzi* (Respuela et al., 2008), and the presence of four histone variants, I hypothesized that a specific chromatin structure at divergent-SSRs may compensate for the apparent lack of RNA pol II promoter sequence motifs.

To identify factors that contribute to the chromatin structure surrounding probable RNA pol II TSSs in *T. brucei*, I adopted ChIP-seq technology to determine the genome-wide distribution of H4K10ac, all four histone variants and one bromodomain protein (BDF3). The data provide strong evidence that at least three processes associated with the generation of open chromatin – distinctive DNA sequence, incorporation of histone variants, and histone tail modification – are involved in RNA pol II transcription initiation.

Results

Establishment of ChIP-seq for the genome-wide analysis of chromatin structure in *T. brucei*

In contrast to ChIP-on-chip, in which immunoprecipitated DNA is hybridized to oligonucleotide microarrays, the immunoprecipitated DNA in ChIP-seq is directly sequenced. Each DNA 'tag' is aligned to the genome and the number of tags aligned to each base pair is summed, giving the relative enrichment of a particular factor along the genome (Figure 5.1). Reproducibility and linear range of ChIP-seq are higher than for microarray-based systems (Barski et al., 2007), and resolution is

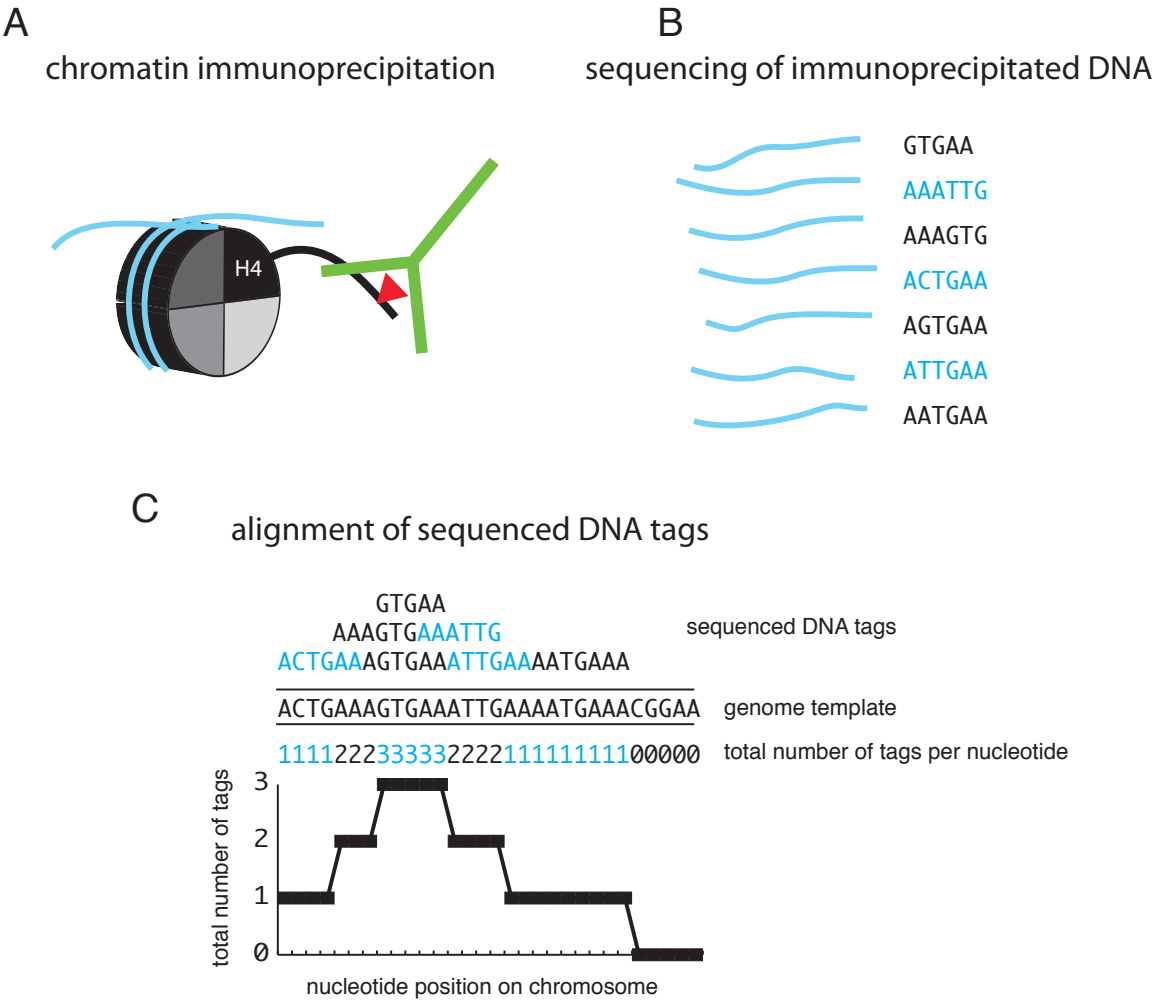


Figure 5.1. Schematic of ChIP-sequencing protocol

(A) Chromatin is immunoprecipitated with antibodies (green) specific for the chromatin mark of interest (red).

(B) Per ChIP-seq analysis, 3–5 million immunoprecipitated DNA tags are sequenced.

(C) The sequenced DNA tags are aligned onto a genome template, the total number of DNA sequence tags per nucleotide position is determined, and the number of tags per nucleotide position is displayed.

limited only by the size of the immunoprecipitated DNA fragments and not by the genomic spacing of the oligonucleotides used to create the microarray.

To ensure the robustness of the ChIP-seq approach, previously established ChIP protocols were optimized to obtain a protocol that was tolerated by antibodies specific for H3, H4, H4K10ac, HA- and Ty1-peptides (Lowell and Cross, 2004; Lee et al., 2006). Furthermore, sonication conditions were established that yielded highly reproducible DNA fragmentation (~300 bp fragments) and allowed omission of size-selection steps during preparation of the immunoprecipitated DNA for high-throughput sequencing (Illumina, 2008). Partly because the protocol did not include any DNA gel-extraction step, it can be applied to relatively few cells – 10^8 trypanosomes.

Repetitive elements in the genome template were not masked and hits were found using the BLAT-based algorithm (Kent, 2002) for each tag. To obtain consistent results, sequence tags with multiple hits were annotated to the lowest nucleotide position on the smallest-numbered chromosome that gave the best match. This annotation methodology produced artifactual peaks at the first occurrence of highly repetitive sequence motifs, even if the factor investigated was not enriched at a particular repeat region. This artifact was unavoidable because, although the *T. brucei* genome has been almost completely sequenced, the lengths of most repetitive regions are unknown. To identify and compensate for

such artifacts, sequencing was performed on sonicated genomic DNA, DNA from micrococcal nuclease-digested chromatin, and DNA immunoprecipitated with a custom-made control antibody specific for trypanosome H4 regardless of its modification state (see Chapter 3) and a commercial antibody to the C-terminal tail of H3 (Abcam 1791). Because DNA tags from sonicated genomic DNA should align with the genome uniformly, any observed peaks and gaps represent first and subsequent repeats of repetitive sequence elements, respectively, which is what was observed. Similarly, alignment of sequenced tags from micrococcal nuclease-treated DNA and ChIP-seq data for the H4 (Figure 5.2) and H3 antibodies revealed peaks and gaps at repetitive elements.

Histone H4K10ac is enriched at probable RNA pol II transcription start sites

Using ChIP-seq and custom-made antibodies specific to H4K10ac and H4, the genome-wide distribution of H4K10ac was determined. The generation and characterization of these antibodies are described in Chapter 3.

Peaks of H4K10ac enriched ~300-fold over background were distributed throughout the genome, with distinct twinned peaks occurring at every divergent-SSR and no enrichment at convergent-SSRs (Figure 5.2 and Figure 5.3), suggesting a link between H4K10 acetylation and transcription initiation. The distribution was remarkably similar

between two parasite life-cycle stages (bloodstream and procyclic forms), with the exception of one peak that was unique to each life-cycle stage (Figure 5.3, chromosomes 7 and 11, red asterisks).

I also observed 61 single H4K10ac-rich peaks at non-SSRs (Figure 5.2 and Figure 5.3), which I suggest are unanticipated TSSs. Many of these peaks occur immediately downstream of tRNA genes (see example in Figure 5.2, bottom panel). Most tRNA genes are located at convergent-SSRs but some are present at non-SSRs that appeared to be embedded within RNA pol II PTUs, which raised questions about the overlap of RNA pol III and RNA pol II transcription. I found that all but three tRNA genes that are not associated with convergent-SSRs are located upstream of H4K10ac-rich 'single' peaks, but not all 'single' peaks are associated with tRNA genes.

Two studies have reported RNA pol II transcription initiation outside of divergent-SSRs – upstream of the actin gene (Ben Amar et al., 1991) and in the HSP 70 locus (Lee, 1996). While the potential actin promoter region coincides with an H4K10ac-rich region (see green asterisk in Figure 5.3), the HSP70 promoter does not, suggesting that the HSP70 promoter was misassigned. This finding is in agreement with a previous study that failed to confirm promoter activity for the HSP70 promoter (Hausler and Clayton, 1996).

Finally, there appear to be higher levels of H4K10ac just upstream of the first PTU and downstream of the last PTU on each chromosome

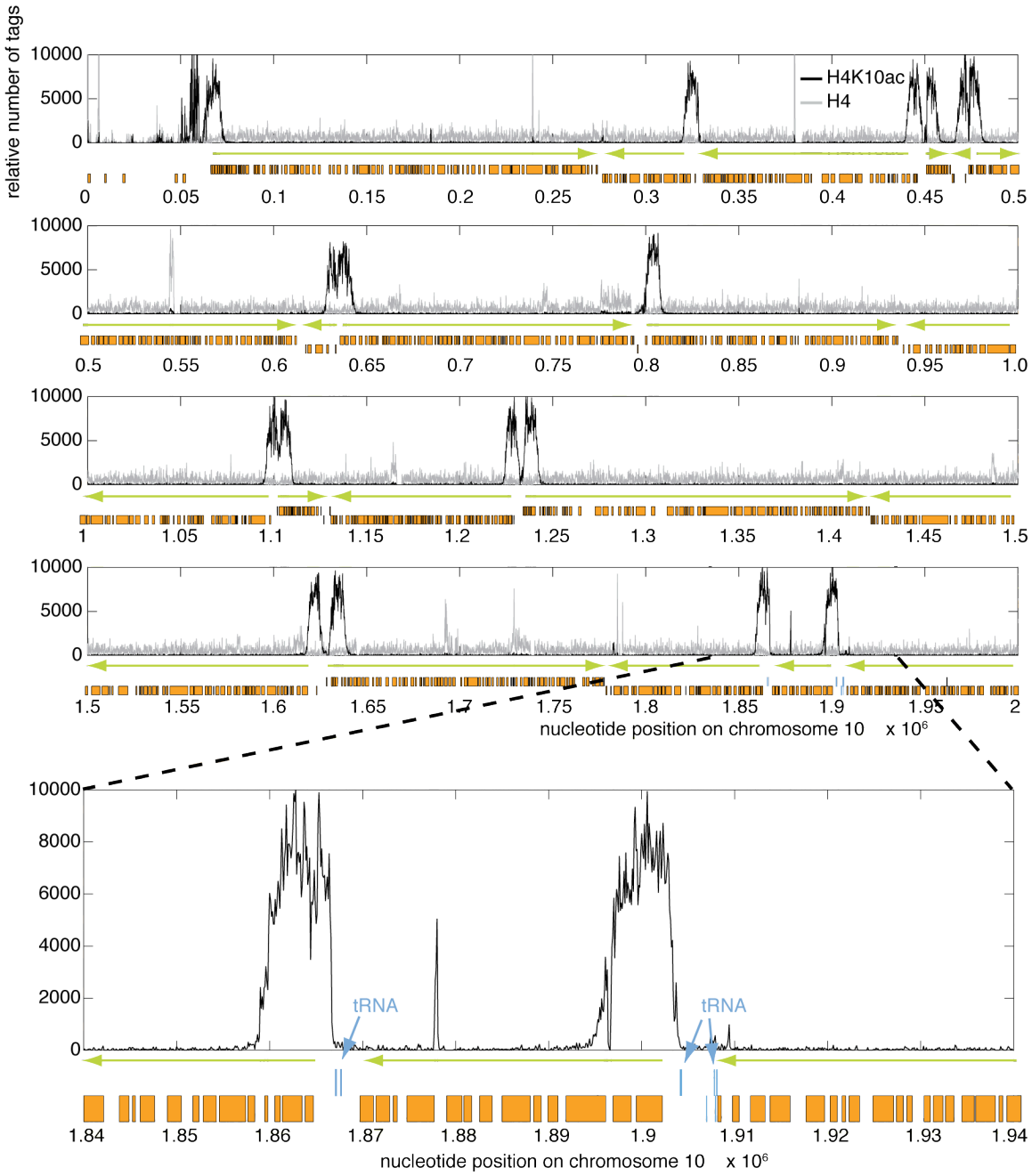


Figure 5.2. Histone H4K10ac is enriched at probable RNA pol II TSSs
Results of ChIP-seq analysis using cross-linked chromatin from BF cells and α-H4K10ac (black) or α-histone H4 (grey). The number of sequenced DNA tags is normalized between runs to account for differences in IP efficiency. For a representative region of chromosome 10, the relative number of tags is calculated for a window size of 100 bp. Orange boxes represent open reading frames and green arrows indicate direction of transcription. Graphs were generated using Matlab.

Figure 5.3. Genome-wide distribution of histone H4K10ac

Results of ChIP-seq analysis using antibodies against H4K10ac and cross-linked chromatin from BF and PF cells. The number of sequenced DNA tags is normalized between runs to account for differences in IP efficiency. Displayed is the genome-wide (chr 1–11) distribution of the relative number of sequenced DNA tags for BF cells (y-axis). For ChIP-seq using PF cells the results are shown for chr 7 and 11, see black boxes. The green asterisk marks the previously identified actin promoter (Ben Amar et al., 1991). The red asterisks mark life cycle-specific peaks. Window size: 100 bp. The x-axes indicate the nucleotide position along the chromosomes on a Mb scale. Graphs were generated using Matlab.


BF chr 1
BF chr 2
BF chr 3
BF chr 4
BF chr 5
BF chr 6
BF chr 7
PF chr 7
BF chr 8
BF ch 9
BF chr 10
BF chr 11
PF chr 11

(Figure 5.2 and Figure 5.3), but these H4K10ac-rich regions do not form distinct peaks. As for other repetitive sequences, tags were assigned to their first occurrence and therefore appear at only a few subtelomeric regions, and peak sizes cannot be corrected for the (unknown) lengths of the repeats.

The width of the H4K10-rich regions was uniform and surprisingly wide (full width at half maximum enrichment 5.6 ± 0.8 kb; total width ~8.7 kb) – well above the resolution limit of 250–300 bp of the ChIP-seq analysis. Based on the total number (207) of H4K10ac peaks, their average width, a genome size of 30 Mb and the fact that ~10 % of H4K10 sites are acetylated overall, it appears all H4 is acetylated at K10 within these peaks, which would explain their remarkably constant height. 207 peaks of an average width of ~8.7 kb which would span 1.8 Mb or 6% of the genome. If all H4K10 residues were acetylated at these sites, one would expect 6% of all histone H4 to be acetylated at K10. The fact that even more than 6% of H4 is acetylated at K10 (Edman data indicates acetylation of K10 for 10% of histone H4) may be due to incorporation of additional H4K10ac in subtelomeric regions.

HAT2 affects acetylation of histone H4K10

To identify a HAT responsible for H4K10 acetylation lysates from cell lines missing the non-essential HAT3, HDAC2, HDAC4 or Sir2rp1 and from cell lines depleted of HAT1, HAT2, Elp3a, Elp3b, HDAC1 or

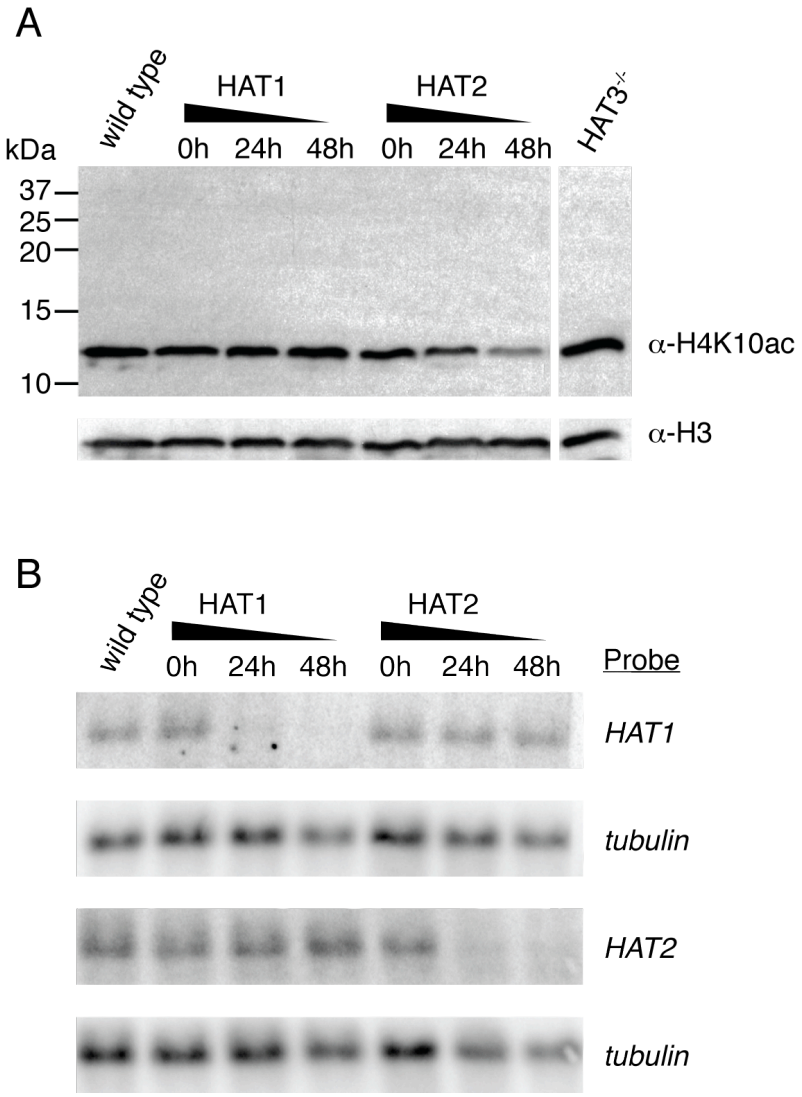


Figure 5.4. HAT2 affects acetylation of histone H4K10

(A) Whole cell lysates from BF deficient for HAT3 or specifically depleted for HAT1 and HAT3 were analyzed by western blotting with an antibody specific for H4K10ac. HAT1 and HAT2 specific RNAi was induced 0, 24, or 48 hours before harvest. To confirm equal loading, the blot was stripped and reprobed with α-histone H3.

(B) Depletion of HAT1 and HAT2 transcripts was monitored by northern blotting analysis using probes specific for *HAT1* and *HAT2* gene fragments. A tubulin probe was used as a loading control. HAT1 and HAT2 specific RNAi was induced 0, 24, or 48 hours before harvest.

HDAC3 were analyzed by western blotting using α-H4K10ac (Figure 5.4A). None of the non-essential histone modifying enzymes had an effect on H4K10 acetylation levels. In contrast, lysates from cells depleted of the essential HAT2 led to a marked decrease in H4K10ac signal (Figure 5.4A). Because the observed decrease of H4K10 acetylation was specific to depletion of HAT2 and not observed after depleting cells of the other essential histone acetyltransferase, HAT1, it appears that the loss of H4K10 acetylation was specifically caused by HAT2 depletion and not an unspecific reaction due to depletion of an essential enzyme. Efficient depletion of HAT1 and HAT2 transcripts was monitored by northern blotting (Figure 5.4B).

H2AZ and H2BV are enriched at probable RNA pol II transcription start sites

T. brucei H2AZ is essential for viability and dimerizes exclusively with H2BV, a variant form of H2B that is also essential for viability (Lowell et al., 2005). Thus, all *T. brucei* nucleosomes that contain H2AZ also contain H2BV. Although the function of H2AZ is not understood in *T. brucei*, it has been linked to transcriptional activation in other organisms, particularly when present in a nucleosome containing a variant of histone H3 (Jin and Felsenfeld, 2007). Using IF microscopy, H4K10ac showed a punctate nuclear distribution with no cell-cycle dependence (Figure 5.5A). This localization is similar to that observed

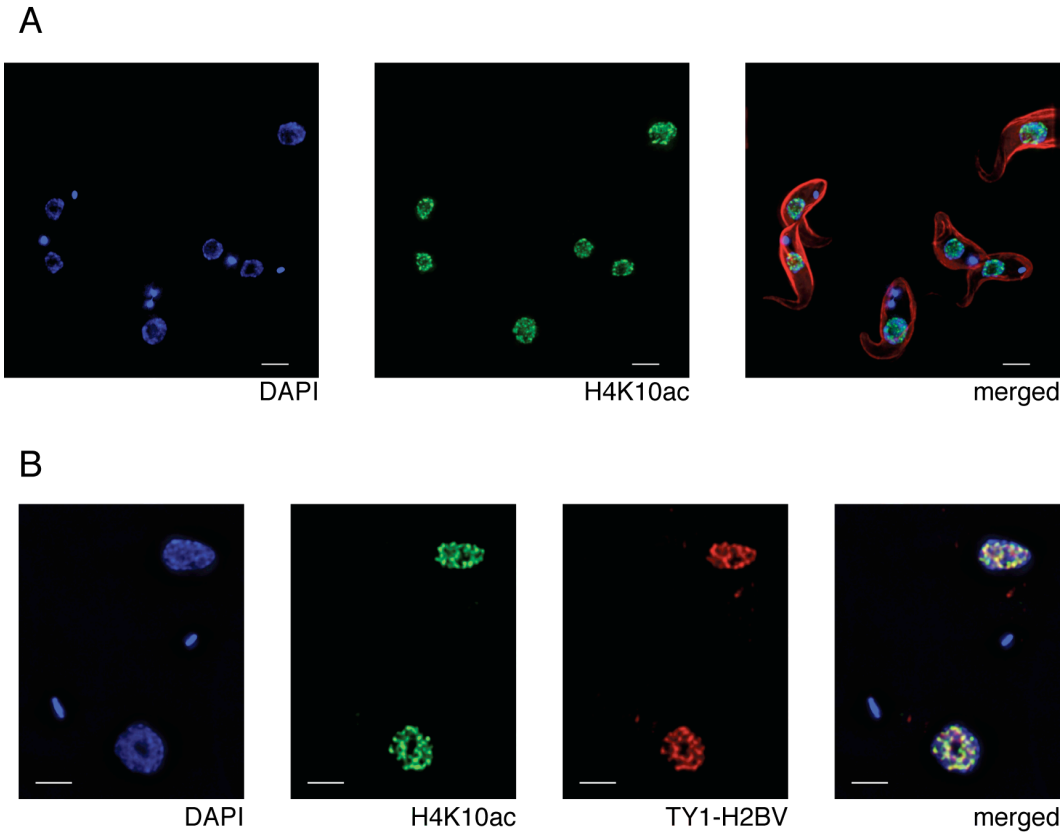


Figure 5.5. H4K10ac and H2BV co-localize and are not cell cycle regulated
(A) Representative IF analysis of H4K10ac, tubulin (red) and DAPI (blue). Bar = 3 μm.
(B) Representative IF co-localization analysis of H4K10ac (green),TY-H2BV (red) and DAPI (blue). Bar = 2 μm.

previously for H2AZ and H2BV, and significantly different from that of a general H4 antibody (Lowell et al., 2005 and Figure 5.7B). A Ty1-epitope-tagged version of H2BV also showed a high degree of co-localization with H4K10ac (Figure 5.5B). These observations raised the possibility that H2AZ/H2BV-containing nucleosomes might be enriched at probable RNA pol II TSSs.

To test this, ChIP-seq experiments were performed with N-terminally Ty1-tagged versions of H2AZ and H2BV. To assure functionality of the ectopically expressed and tagged versions of the variants, I used cell lines in which both endogenous alleles of each variant were deleted. No growth defect was detected, confirming functionality of the tagged versions. Sequence analysis of the DNA in the immunoprecipitated chromatin indicated that the genomic distribution of H2AZ and H2BV, including the width and intensity of the peaks, was almost indistinguishable from H4K10ac (Figures 5.6A and B). As was described for H4K10ac, two peaks of H2AZ/H2BV enrichment were observed at every divergent-SSR and a single peak of H2AZ/H2BV enrichment was present at every non-SSR H4K10ac-rich region. H2AZ and H2BV were also enriched upstream of the first PTU and downstream of the last PTU on every chromosome, as was H4K10ac (above).

In chickens, it has been shown that nucleosomes containing histone variants are destabilized relative to those containing core histones (Jin and Felsenfeld, 2007). To test if H2AZ/H2BV-containing nucleosomes of *T. brucei* are less stable than nucleosomes composed of canonical histones, co-immunoprecipitation experiments were performed. Mononucleosomes were prepared from four cell lines expressing either Ty1-tagged H2AZ or H2A or FLAG-tagged H2BV or H2B, as described previously (Lowell et al., 2005) except that the chromatin pellet was washed at increasing ionic strengths. The tagged

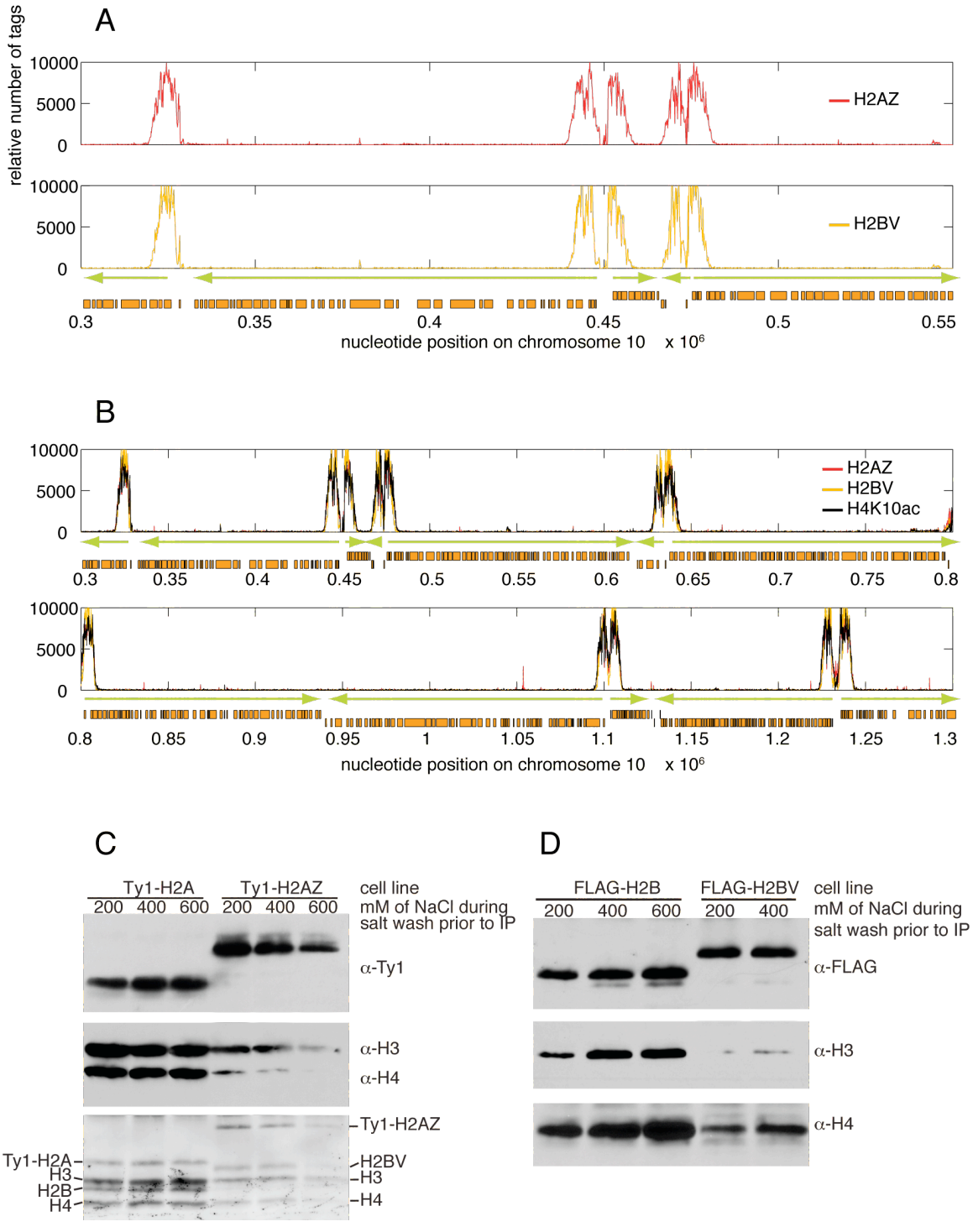
Figure 5.6. H2AZ and H2BV-containing nucleosomes are enriched at probable RNA pol II TSSs and are less stable than core nucleosomes

(A) Results of ChIP-seq analysis for H2AZ (red, top panel) and H2AZ (yellow, bottom panel). The number of sequenced DNA tags is normalized between runs to account for differences in IP efficiency. For a representative region of chromosome 10, the relative number of tags is calculated for a window size of 100 bp. Orange boxes represent open reading frames and green arrows indicate direction of transcription. Graphs were generated using Matlab.

(B) Overlaid results of ChIP-seq analyses for H2AZ (red), H2BV (yellow) and H4K10ac (black). Window size: 100 bp.

(C) Chromatin was prepared from cell lines containing Ty1-tagged H2A or H2AZ and equal amounts were washed with buffers containing increasing levels of NaCl. Chromatin was treated with micrococcal nuclease to generate mononucleosomes and immunoprecipitated with monoclonal antibody BB2. Co-immunoprecipitated histones were separated by SDS-PAGE and immunoblotted (top and middle panels) or stained with amido black (lower panel), to determine the levels of co-immunoprecipitated histones in the H2A-IP and H2AZ-IP.

(D) Chromatin was prepared as described in (B) except that I used cell lines containing FLAG-tagged H2B or H2BV. Mononucleosomes were immunoprecipitated with α-FLAG antibody and co-immunoprecipitated histones were separated by SDS-PAGE and immunoblotted to determine the levels of co-immunoprecipitated histones in the H2B-IP and H2BV-IP.


A
relative number of tags
10000
5000
0
H2AZ
H2BV
0.3
0.35
0.4
0.45
0.5
0.55
nucleotide position on chromosome 10 x 10^6
B
H2AZ
H2BV
H4K10ac
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.8
0.85
0.9
0.95
1
1.05
1.1
1.15
1.2
1.25
1.3
nucleotide position on chromosome 10 x 10^6
C
Ty1-H2A
Ty1-H2AZ
200 400 600 200 400 600
cell line
mM of NaCl during salt wash prior to IP
α-Ty1
α-H3
α-H4
Ty1-H2AZ
H2BV
H3
H4
Ty1-H2A
H3
H2B
H4
D
FLAG-H2B
FLAG-H2BV
200 400 600 200 400
cell line
mM of NaCl during salt wash prior to IP
α-FLAG
α-H3
α-H4

histones and histone variants were then immunoprecipitated and the extent of association with H3 and H4 was determined. A less stable nucleosome should lead to decreased co-immunoprecipitation; indeed, significantly less association of H4 and H3 was observed with tagged histone variants compared to tagged core histones (Figures 5.6C and D). Taken together, the ChIP-seq data and co-IP experiments demonstrate that probable RNA pol II TSSs are marked by distinct and less stable nucleosomes.

Bromodomain-containing factor 3 is enriched at probable RNA pol II TSSs

Acetylated histone N-termini can serve as binding platforms for effector proteins that interpret histone modification patterns and elicit downstream effects (Dhalluin et al., 1999). BDFs, which can bind to acetylated lysines (Winston and Allis, 1999), are examples of effector proteins. BDFs have been shown to tether HATs or nucleosome remodeling complexes to specific chromosomal sites and to function as transcriptional coactivators (Dhalluin et al., 1999; Hassan et al., 2002).

As described in Chapter 3, trypanosome BDF1-5 were HA-tagged and analyzed by IF microscopy. Given the prominent role BDFs play in recruiting transcription machineries, I decided to test whether any trypanosome BDF co-localizes with probable RNA pol II TSSs. Only BDF3 showed significant co-localization with H4K10ac (Figure 5.7A). To

substantiate the IF microscopy data, I performed ChIP-seq with a cell line containing an HA-tagged allele of BDF3, in which the second untagged allele was deleted to ensure functionality of the tagged protein. ChIP-seq demonstrated that BDF3 was highly enriched at the same sites as H4K10ac (Figure 5.7B), but there were distinct differences – whereas the signals clearly overlapped, BDF3 was concentrated towards the 5´ end of the H4K10-rich regions (Figure 5.7B and Figure 5.9B).

BDF3 is a 25-kDa protein containing no conserved domains other than the bromodomain. To test whether BDF3 is essential for viability, I conducted RNAi-mediated knock-down experiments. Reduction of BDF3 led to an immediate growth defect and most cells appeared to be dying after 48 hours, which is consistent with BDF3 being essential (data not shown). One interpretation of BDF3 and H4K10 co-localization would be that BDF3 binds to acetylated H4K10 and facilitates transcription initiation.

To test if BDF3 directly binds to H4K10ac, I performed *in vitro* binding assays using nuclear extracts from cell lines containing HA-tagged BDF3 and biotinylated peptides representing the unmodified H4 tail, and peptides acetylated at either K4 or K10, or acetylated at both K4 and K10. If BDF3 specifically binds to K10ac, I would expect BDF3 to specifically associate with peptides acetylated at K10. However, despite extensive protocol optimization, I did not observe any significant difference in BDF3 binding to the unmodified H4 tail and to any of the

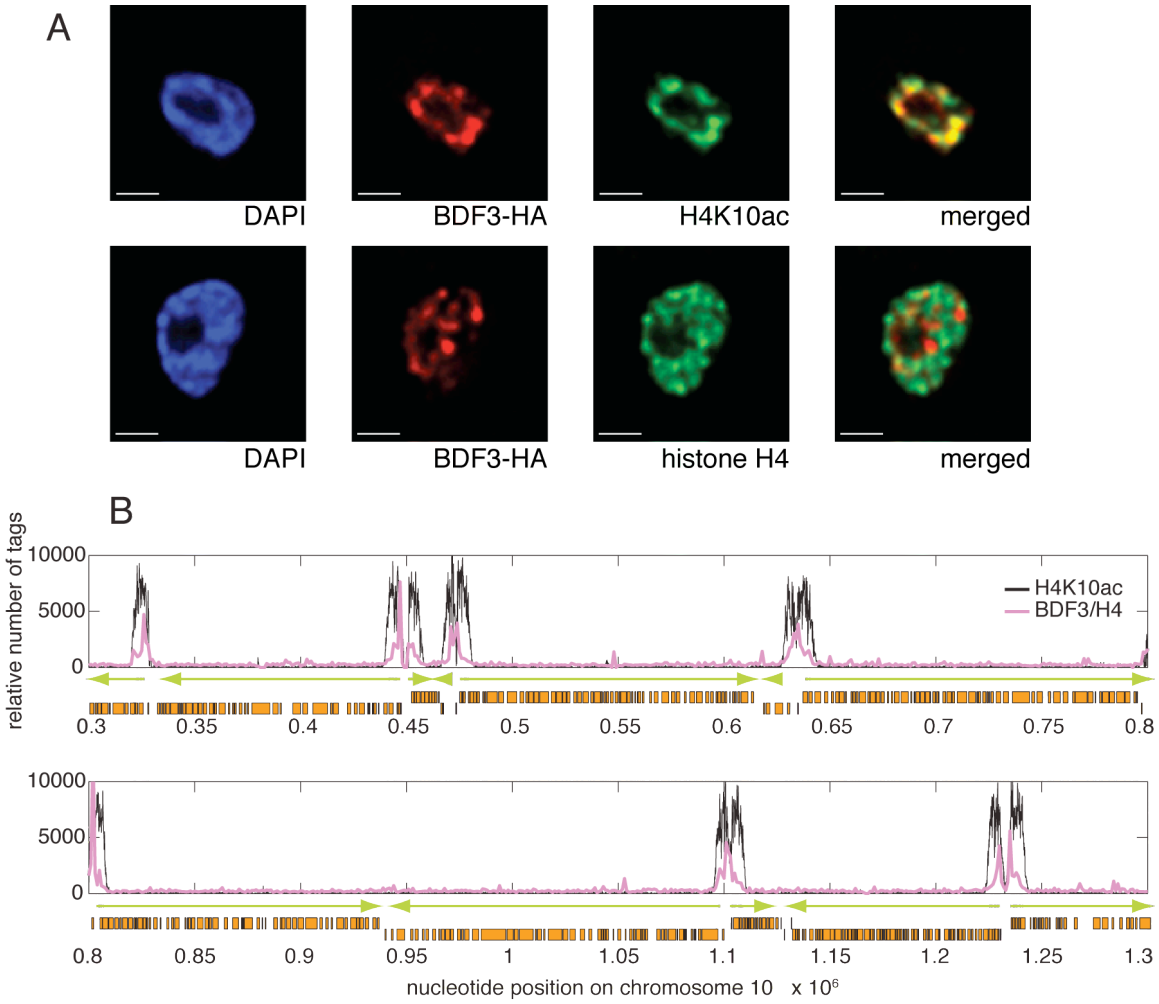


Figure 5.7. BDF3 co-localizes with H4K10ac and is enriched at probable RNA pol II TSSs

(A) A representative IF co-localization analysis of BDF3-HA (red), DAPI (blue) and H4K10ac (top panel, green) or H4 (bottom panel, green). Bar = 2 μm.
(B) Results of ChIP-seq analysis for BDF3 (magenta) and H4K10ac (black). The number of sequenced DNA tags is normalized between runs to account for differences in IP efficiency. Additionally, for BDF3 the number of tags is normalized based on the number of tags from an histone H4 ChIP-seq analysis to eliminate artifacts attributable to repetitive sequences. For a representative region of chromosome 10, the relative number of tags is calculated for a window size of 100 bp (H4K10ac) or 1000 bp (BDF3). Orange boxes represent open reading frames and green arrows indicate direction of transcription. Graphs were generated using Matlab.

three acetylated peptides. Even at 400 mM NaCl, BDF3 bound equally to each of the four peptides tested (data not shown). Others have reported that BDF binding is not always specific (Hassan et al., 2007), and it has been suggested that the cooperative engagement of several linked proteins, rather than the interaction between one mark and its binding partner, is of biological relevance (Ruthenburg et al., 2007b). Thus, given the limits of the *in-vitro* system, I cannot evaluate whether BDF3 binds directly to H4K10Ac.

H3 and H4 variants are enriched at probable RNA pol II transcription termination sites

In addition to H2AZ and H2BV, *T. brucei* contains variant forms of H3 and H4, whose functions are unknown. Past work, using IF microscopy and standard ChIP followed by dot-blot readout, revealed that H3V was enriched at telomeric repeats, but localization at additional sites could not be ruled out because of the limits of the technology employed (Lowell and Cross, 2004). No phenotype could be detected for a cell line lacking both alleles of H3V (Lowell and Cross, 2004). Despite H3V being non-essential under laboratory culture conditions, the conserved function of histone H3 variants in transcription regulation in other organisms suggests that H3V may have a subtle and undetected role in transcription regulation in *T. brucei.*

Because a role of H3V in transcription regulation would most likely lead to incorporation of H3V outside of telomeric and subtelomeric sites, I performed ChIP-seq using a cell line containing a Ty1-tagged copy of H3V. In agreement with previous results, I found that H3V was highly enriched at telomeric and subtelomeric regions of the genome (Figure 5.8A), but only ~65% of the immunoprecipitated DNA corresponded to telomeric and subtelomeric sequences. H3V was also enriched at convergent-SSRs and just upstream of all H4K10ac-rich regions not associated with SSRs (Figure 5.8A). Taken together, these data show that H3V is enriched at the presumed ends of RNA pol II PTUs.

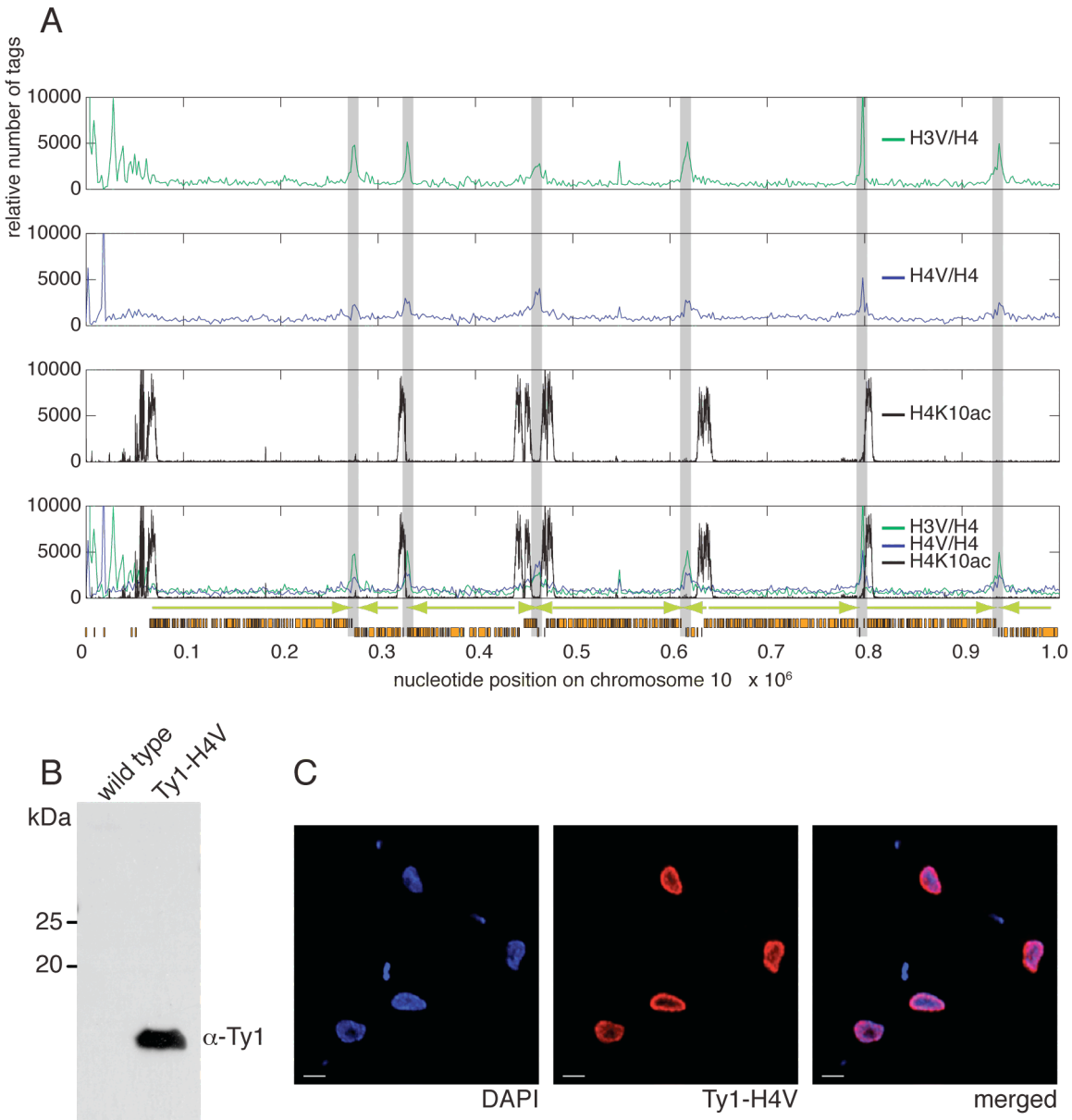
T. brucei H4V has not been characterized previously. It shares 85% sequence identity with *T. brucei* H4, whereas H2AZ, H2BV, and H3V share only 37–45% sequence identity with their core histone counterparts (Figure 1.8). To determine the cellular localization of H4V, Louise Kemp generated a cell line containing an ectopically expressed Ty1-tagged H4V and deleted both endogenous alleles (Figure 5.8B). IF microscopy indicated that Ty1-H4V localized to the nucleus and displayed no apparent cell-cycle regulation (Figure 5.8C). Using this cell line, I determined the genomic distribution of H4V. H4V was distributed broadly throughout the genome, but I observed peaks of enrichment at the presumed ends of RNA pol II PTUs, similar to those observed for H3V (Figure 5.8A). In contrast to H3V, H4V was much less enriched at telomeric or subtelomeric sites.

Figure 5.8. Histone variants H3V and H4V are enriched at probable RNA pol II TTSs

(A) Results of ChIP-seq analysis for H3V (1st panel, green), H4V (2nd panel, blue), H4K10ac (3rd panel, black) and an overlay of H4K10ac, H3V and H4V data (4th panel). The number of sequenced DNA tags is normalized between runs to account for differences in IP efficiency. Additionally, for H3V and H4V the number of tags is normalized based on the number of tags from an histone H4 ChIP-seq analysis to eliminate artifacts attributable to repetitive sequences. For a representative region of chromosome 10, the relative number of tags are calculated for a window size of 100 bp (H4K10ac) or 2000 bp (H3V and H4V). Orange boxes represent open reading frames, green arrows indicate direction of transcription and vertical grey boxes mark probable RNA pol II TTSs. Graphs were generated using Matlab.

(B) Western blot analysis indicates that Ty1-tagged H4V has an apparent molecular weight of 11 kDa.

(C) Representative IF analysis of histone H4V and DAPI (blue). Bar = 2 μm.


A
relative number of tags
10000
5000
0
H3V/H4
H4V/H4
H4K10ac
H3V/H4
H4V/H4
H4K10ac
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
nucleotide position on chromosome 10 x 10^6
B
wild type
Ty1-H4V
kDa
25
20
α-Ty1
C
DAPI
Ty1-H4V
merged

As for deletion of H3V, no growth defect was detected under culture conditions for a H4V knockout cell line (data not shown). Nevertheless, their genomic distribution at the presumed ends of PTUs strongly suggests a role for H3V and H4V in transcription regulation.

Analysis of DNA sequences at probable RNA pol II TSSs reveals long G-runs

Taken together, my data suggest that chromatin immediately upstream and downstream of PTUs has strikingly distinct features. Although it seems reasonable to speculate that DNA sequence could play a role in establishing these unique chromatin structures, no RNA pol II promoter sequence motif has been identified in *T. brucei*. In fact, previous findings led to the hypothesis that an open chromatin structure may be sufficient to initiate RNA pol II transcription (McAndrew et al., 1998). Nevertheless, data from the related organism *L. major* (Martinez-Calvillo et al., 2003), and the presence of two rather than one H4K10ac/H2AZ/ H2BV-rich peaks at divergent-SSRs, strongly suggest that RNA pol II transcription is unidirectional and that a specific DNA feature may determine directionality.

I took several approaches to identify DNA sequence motifs that were unique to H4K10ac-rich regions. With the help of Doeke Hekstra, The Rockefeller University, I also compared the sense strand to the complementary strand within H4K10ac-rich regions. The only motif

consistently identified was a G-rich stretch in the sense strand at the upstream end of every H4K10ac-rich region (Figure 5.9). Stretches of 9–15 consecutive guanines were enriched ~7-fold in these regions compared to other sites (Figure 5.9B). Specifically, within 91% of H4K10ac/H2AZ/H2BV-rich peaks (117/128), we found stretches of at least 10 guanines, allowing for up to 2 mismatches, between 1,000 and 5,000 bp upstream of the peak center (expected: 65%, $p < 10^{-12}$). In contrast to the G-rich upstream part of the H4K10ac-rich region, the central 3 kb of these regions was slightly AT-rich (56%) compared to the rest of the genome (50%) (Figure 5.9A). The results are modeled in Figure 5.9C.

Discussion

In this study I found that the beginning and end of RNA pol II PTUs are characterized by distinct chromatin modifications and that there may be more RNA pol II TSSs than previously assumed, and I also confirmed that probable RNA pol II TSSs do not contain classic sequence motifs for transcription-factor interactions.

RNA pol II transcription initiation and termination regions have distinct chromatin features

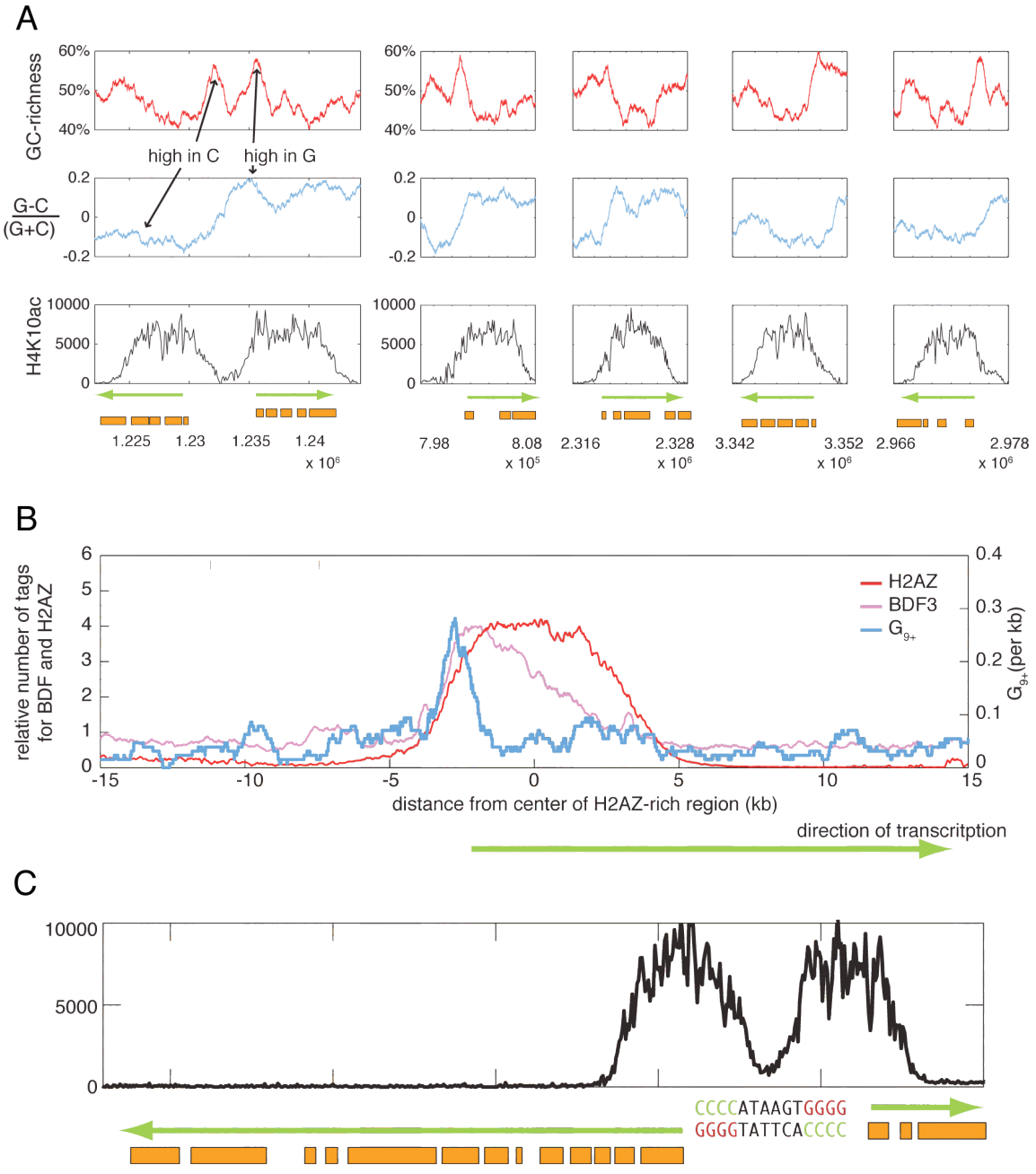
Four histone variants mark the beginning and end of RNA pol II PTUs in *T. brucei*. H2AZ and H2BV are enriched at probable RNA pol II

Figure 5.9. 5´ ends of probable RNA pol II TSSs are marked by runs of guanine

(A) Divergent-SSRs contain two peaks of high GC richness (red). The left peak indicates high levels of C in the upper strand and the right peak indicates high levels of G in the upper strand (blue). Orange boxes represent open reading frames and green arrows indicate direction of transcription. Numbers on x-axis indicate nucleotide position on chromosome 10.

(B) Density of stretches of G ($n \geq 9$) at TSSs (dark blue, right axis). Density shown is the mean over 38 putative TSSs at non-SSRs and 90 TSSs at divergent-SSRs using a 1,000-bp moving average (densities were statistically indistinguishable between the two types of TSSs, data not shown). For comparison, the mean shapes of H2AZ (red) and BDF3 (magenta) enrichment are shown at non-SSR TSSs (left axis). BDF3 data were scaled to have the same maximum as H2AZ data. TSS centers were determined based on Gaussian fits to the H2AZ data.

(C) Model indicating how regions upstream of pol II TSSs are enriched for guanine. Sequence and peak size are not drawn to scale.


A
GC-richness
60%
50%
40%
high in C
high in G
G-C
(G+C)
0.2
0
-0.2
H4K10ac
10000
5000
0
1.225
1.23
1.235
1.24
x 10^6
7.98
8.08
x 10^5
2.316
2.328
x 10^6
3.342
3.352
x 10^6
2.966
2.978
x 10^6
B
relative number of tags for BDF and H2AZ
H2AZ
BDF3
G9+
G9+(per kb)
distance from center of H2AZ-rich region (kb)
direction of transcritption
C
CCCCATAAGTGGGG
GGGGTATTCACCCC

TSSs. H3V and H4V are enriched at probable TTSs. I also see enrichment of H4K10ac and BDF3 at probable RNA pol II TSSs (summarized in Figure 5.10), where all H4K10 appears to be acetylated. It has been proposed before that an open chromatin structure may be sufficient for RNA pol II transcription initiation in *T. brucei* (McAndrew et al., 1998). In the following paragraphs I propose a model for the sequence of events leading to RNA pol II transcription initiation.

Methylation of histone H3 at K4 correlates with transcription activation (Ruthenburg et al., 2007a; Honda et al., 1975) and analyses of the genome-wide distribution of H3K4me3 have been performed in several organisms where H3K4me3 consistently localizes to promoter regions (Barski et al., 2007; Shilatifard, 2008). Currently no ChIP-grade H3K4me3 antibody exists for trypanosomes but, based on previous findings that H3K4me3 is enriched in H2BV containing nucleosomes (Mandava et al., 2008), it can be assumed that in *T. brucei* H3K4me3 is also enriched at RNA pol II TSSs. Furthermore, evidence exists that H3K4me3 is required for histone tail acetylation, another chromatin feature consistently associated with active chromatin (reviewed in Pillus, 2008). Because HAT2 contains a methyl-binding domain (Ivens et al., 2005) and is responsible for H4K10 acetylation, it appears plausible that it is targeted to TSSs by binding to a methyl mark, possibly to H3K4me3. In *T. brucei,* H3K4 tri-methylation may thus be required for H4K10

acetylation. The enzyme responsible for H3K4 methylation has not been identified in *T. brucei*.

The role of histone variants in transcription initiation is largely unknown, although nucleosomes containing H2AZ and H3.3 variants are significantly less stable than nucleosomes containing only canonical histones (Jin and Felsenfeld, 2007). These unstable nucleosomes are often enriched at RNA pol II TSSs, where they are likely to facilitate nucleosome eviction prior to transcription initiation (Mito et al., 2005; Guillemette et al., 2005; Zhang et al., 2005). In this study, I show that *T. brucei* histone variants H2AZ and H2BV are also enriched around probable RNA pol II TSSs, and nucleosomes containing H2AZ and H2BV are less stable than nucleosomes containing the corresponding core histones. These observations strongly support the hypothesis that the presence of H2AZ and H2BV at *T. brucei* RNA pol II TSSs contributes to a more open chromatin structure.

In yeast, the Swr1 chromatin remodeling complex is responsible for H2AZ incorporation. Targeting of this complex to the correct genomic location appears to be mediated by a member of the Swr1 complex, the bromodomain containing protein BDF1 (reviewed in Korber and Horz, 2004). Thus, yeast BDF1, may target the chromatin-remodeling complex to sites containing acetylated histones. Further evidence that H2AZ incorporation may depend on histone acetylation stems from the finding that H4K16ac is required for H2AZ incorporation in subtelomeric regions

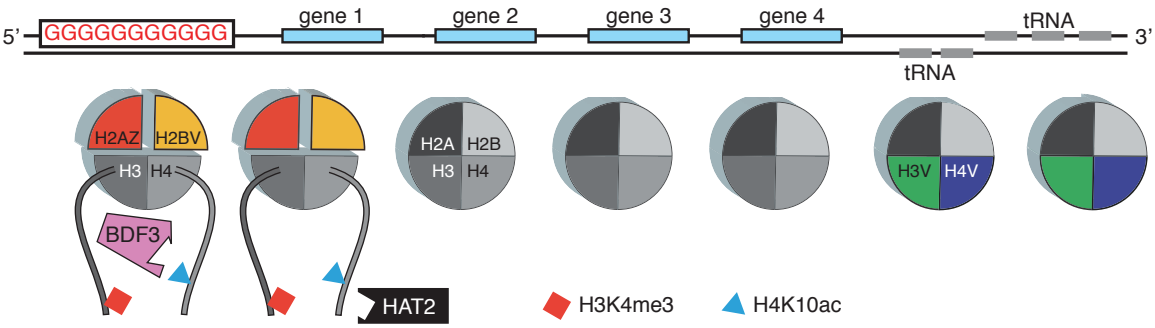


Figure 5.10. Model to illustrate a possible cascade of events leading to RNA pol II transcription initiation

Histone H3K4me3 enrichment at TSSs could provide a binding site for HAT2, which contains a chromo domain and affects H4K10 acetylation. H4K10ac is strongly enriched at TSSs and might serve as a binding site for BDF3. The latter is enriched at the 5´ end of TSSs, where it may play a role in recruiting RNA pol II transcription factors or the chromatin-remodeling complex involved in H2AZ and H2BV incorporation. These two histone variants are enriched at TSSs and form less stable nucleosomes that could facilitate binding of the transcriptional machinery. The G-tract at the 5´ end of TSSs could block anti-sense transcription and/or provide binding sites for transcription factors. H3V and H4V variants are enriched at TTSs, which often contain tRNA or other genes transcribed by RNA pol III.

(Shia et al., 2006). My data indicate that *T. brucei* BDF3 is localized to a relatively narrow region at TSSs compared to H2AZ. Therefore, it appears plausible that BDF3 plays a role in targeting a chromatin-remodeling complex to TSSs in *T. brucei*, a role similar to BDF1 in yeast. Because I was unable to demonstrate an interaction between H4K10ac and BDF3 *in vitro*, I cannot exclude the possibility that a different acetyl mark may serve as binding site for BDF3.

The *T. brucei* H3 and H4 variants mark the end of probable RNA pol II PTUs. This further emphasizes the notion that RNA pol II transcription is regulated by chromatin structure. However, the near complete acetylation of H4K10 at all probable RNA pol II TSSs suggests that this mark plays no role in regulating the level of transcription initiation: it may simply be required to turn on transcription. Trypanosomes appear to regulate RNA pol II-driven gene expression post-transcriptionally: the efficiency of *trans* splicing individual mRNAs differs significantly, as do mRNA stability, translation efficiency and protein stability (reviewed in Clayton, 2002).

Poly-guanine tracts could guide directional RNA pol II transcription

The presence of unstable non-canonical nucleosomes at probable RNA pol II TSSs leads to a more open chromatin structure that may be sufficient for transcription initiation. However, I also identified an enrichment of poly-guanine motifs at probable RNA pol II TSSs, a common feature of promoter regions (Huppert and Balasubramanian, 2007). The presence of short GC-rich tracts at SSRs was previously recognized in *Leishmania* (Martinez-Calvillo et al., 2003). Two lines of evidence placed TSSs on chromosome 1 quite precisely on either side of a GC-rich tract, but this SSR was very short compared to those in *T. brucei*. There are other significant differences in gene transcription

between *Leishmania* and *T. brucei*, so it is not possible to make unqualified extrapolations from one organism to the other. It will be interesting to determine the distribution of histone modifications and variants within the *Leishmania* genome.

The function of poly-guanine motifs in *T. brucei* remains to be determined. One possibility is that the G-rich DNA adopts a secondary structure, such as a G-quadruplex fold, that functions as transcription factor binding site. Transcription factors specifically binding to G-quadruplexes have been identified in humans (Etzioni et al., 2005). Alternatively, the poly-guanine motif could govern the unidirectionality of transcription in *T. brucei*. It has been shown that stretches of guanines located on the mRNA sense strand can hinder the progression of RNA pol II (Tornaletti et al., 2008). If an open chromatin structure were sufficient to initiate transcription in *T. brucei* and if transcription were to start within the H4K10ac-rich region and proceeded in both directions (divergent transcription), it would encounter the G-rich stretch only if it went the 'wrong' way (Figure 5.9C and Figure 5.10). The poly-guanine motif could thus hinder RNA pol II from proceeding and thereby promote unidirectional transcription. Divergent transcription initiation followed by termination of anti-sense transcription may be a common phenomenon. A recent study reports divergent transcription initiation from more than 40% of all protein-encoding gene promoters in mice. In mice, however,

transcription termination appears to correlate with the lack of H3K79me2 (Seila et al., 2008).

Novel RNA pol II transcription start sites

Whereas it has generally been assumed that RNA pol II transcription starts at divergent-SSRs, I also found many putative RNA pol II TSSs at non-SSRs, often downstream of tRNA genes. As tRNA genes can inhibit RNA pol II transcription in yeast (Hull et al., 1994), a similar mechanism for tRNA-mediated transcriptional silencing was proposed for *T. brucei* (Marchetti et al., 1998). It has been suggested that the subnuclear localization of the tRNA genes may be important for tRNA-mediated silencing (Kendall et al., 2000). More recently, it has been shown in yeast that TFIIIC, which is required for RNA pol III transcription complex assembly, tethers tRNA genes to distinct loci at the nuclear periphery (Noma et al., 2006). This could generate RNA pol III transcription clusters from which RNA pol II is largely excluded. TFIIIC binds to the B-box, one of two highly conserved promoter elements required for tRNA transcription in most eukaryotes, including *T. brucei* (Nakaar et al., 1994). As in yeast, the *T. brucei* B-box is only essential for tRNA transcription *in vivo*, not *in vitro* (Margottin et al., 1991; Nakaar et al., 1997), so its role in subnuclear clustering of tRNA may be conserved in trypanosomes, although a homologue of TFIIIC has not been identified in *T. brucei*. Such subnuclear clustering of tRNA genes into RNA pol III

transcription clusters could explain why I was unable to initiate RNA pol II transcription by placing long stretches of DNA sequence from divergent-SSRs into convergent-SSRs (data not shown).

If tRNA genes indeed interrupt RNA pol II transcription, reinitiation would be required downstream of the tRNA 'roadblock'. I found that all but three tRNA genes located outside of convergent-SSRs are surrounded on both sides by high levels of H3V and H4V and followed by high levels of H4K10ac/H2AZ/H2BV/BDF3, suggesting that tRNA indeed represents a boundary element and that RNA pol II transcription has to reinitiate downstream of this boundary. The three tRNA genes not located at convergent-SSRs and not next to a region of high H4K10ac are isolated and are not in clusters like most other tRNAs. Individual tRNA genes with modest TFIIIC binding are unable to function as boundary elements in yeast (Noma et al., 2006).

If all of the putative RNA pol II TSSs at non-SSRs represent reinitiation sites, and considering only some of these sites are located downstream of tRNA genes, what is the nature of the non-tRNA roadblocks? Could these regions contain unrecognized RNA genes that are also transcribed by RNA pol I or RNA pol III?

Conclusion

ChIP-seq technology offers new insight into the long-standing problem of transcription initiation in *T. brucei*, although a mechanistic

description of the role of the different chromatin components awaits detailed biochemical studies. The apparent lack of conserved RNA pol II promoter motifs makes trypanosomes an interesting model system in which to study the role of chromatin composition and modification in the control of transcription initiation and termination. Being evolutionarily divergent, *T. brucei* should also shed light on the evolution of chromatin-mediated epigenetic regulation in eukaryotes.

Chapter 6: Systematic study of sequence motifs for RNA *trans* splicing

Summary

mRNA maturation in *Trypanosoma brucei* depends upon *trans* splicing, and variations in *trans*-splicing efficiency could be an important step in controlling the levels of individual mRNAs. RNA splicing requires specific sequence elements including conserved 5´SS, branch points, poly(Y) tracts, 3´SS and, sometimes, intronic enhancer elements. To analyze sequence requirements for efficient *trans* splicing, in the poly(Y) tract and around the 3´SS, I constructed a luciferase-β-galactosidase double-reporter system. By testing ~90 sequences, I demonstrated that the optimum poly(Y) tract length is ~25 nt. Interspersing a purely uridine-containing poly(Y) tract with cytidine resulted in increased *trans*-splicing efficiency, whereas purines led to a large decrease. The position of the poly(Y) tract relative to the 3´SS is important – an AC dinucleotide at position -3 and -4 can lead to a 20-fold decrease in *trans* splicing. However, efficient *trans* splicing can be restored by inserting a second AG dinucleotide downstream, which does not function as splice site but may aid in recruitment of the splicing machinery.

Introduction

In eukaryotes, a central step in generating mature mRNA from pre-mRNA is the removal of introns and the joining of the two flanking exons, a process known as *cis* splicing (Reed, 2000; Wahl et al., 2009). For successful splicing, the intron has to be correctly identified. Several sequence elements are implicated in defining the two splice sites. The 5´ end of the intron is generally defined by a GT dinucleotide, whereas the 3´ end is marked by AG. Additional characteristics of the 3´SS are a branch point sequence followed by a poly(Y) tract. Enhancer regions, which may contribute to the assembly of the spliceosome or the identification of the correct 3´SS, have also been described (Reed, 2000). Splicing involves two *trans* esterification steps. During the first *trans* esterification, the conserved branch point adenosine forms a 2´-to-5´ phosphodiester bond with the 5´ end of the intron (Ruskin et al., 1984). This step depends on U2 snRNP binding to the branch point sequence, which in turn requires the help of the heterodimeric auxiliary factor U2AF, consisting of 65-kDa and 35-kDa subunits. The U2AF65 subunit has been shown to bind to the poly(Y) tract, whereas U2AF35 associates with the 3´SS AG dinucleotide (Zamore and Green, 1989; Wu et al., 1999; Zamore et al., 1992; Kielkopf et al., 2001). During the second *trans* esterification, the free hydroxyl of the upstream exon attacks the phosphate of the 3´SS to join the exons and release the lariat-shaped intron (reviewed in Wahl et al., 2009).

The mechanism for selecting the appropriate 3´SS AG dinucleotide from other cryptic AG sites for the second *trans* esterification reaction has not entirely been solved. Two models have been proposed. According to the scanning model, identification of the correct 3´SS AG occurs by a linear search mechanism, assuming that the spliceosome begins scanning at the branch point and selects the first downstream AG dinucleotide (Smith et al., 1993). This model is supported by research with HeLa cells, in which splicing is blocked when hairpin loops are inserted upstream of the AG splice site, possibly preventing movement of the spliceosome (Chen et al., 2000). With a bimolecular exon ligation assay, it was also shown that the substrate with the 5´-most AG is selected among different 3´ RNA substrates and that no poly(Y) tract was necessary for the second catalytic step (Anderson and Moore, 1997; Chen et al., 2000). In contrast to human cells, a simple scanning model cannot explain various findings on *S. cerevisiae*, where downstream AG sites can outcompete upstream AG sites if the upstream sites are located closer than 23 nt to the branch point sequence (Patterson and Guthrie, 1991; Chua and Reed, 2001). Moreover, hairpin loops upstream of an AG splice site had an enhancing effect, leading to increased usage of that splice site instead of inhibition (Deshler and Rossi, 1991).

The second model suggests a mechanism in which the correct 3´SS AG dinucleotide is identified by its distance from the branch point. Data supporting this model stem from findings obtained with yeast and

human cells. They show that the second *trans* esterification step occurs most efficiently when the 3´SS AG is located 19 to 23 nt downstream of the branch point (Chiara et al., 1997; Chua and Reed, 2001). In addition to the branch point to AG distance, the sequence of this region itself, especially the presence or absence of a poly(Y) tract, has been shown to strongly affect splicing efficiency in a number of organisms, including humans (Smith et al., 1993), *S. cerevisiae* (Patterson and Guthrie, 1991), and *T. brucei* (Huang and Van der Ploeg, 1991; Matthews et al., 1994).

T. brucei transcribes the majority of its genes as polycistronic units (Tschudi and Ullu, 1988). To generate mature mRNAs, a 39-nt miniexon, also called the spliced leader, must be *trans* spliced to the primary transcript, at appropriate points, from a capped precursor of ~140 nt (spliced leader RNA). *Trans* splicing serves two functions – it dissects mRNAs from polycistronic primary transcripts, and the spliced leader provides the cap structure for the mRNA (Perry et al., 1987; Freistadt et al., 1987; Freistadt et al., 1988). In contrast to *cis* splicing, *trans* splicing joins exons derived from two independently transcribed RNAs (Campbell et al., 1984; Milhausen et al., 1984). *Trans* splicing and *cis* splicing, however, share remarkable similarities: both require the same characteristic sequence motifs [GT at the 5´SS, an adenosine branch point, a poly(Y) tract, AG at the 3´SS, and possibly exonic enhancer motifs] (Huang and Van der Ploeg, 1991; Lopez-Estrano et al., 1998; Patzelt et al., 1989), both follow the same general mechanism (two

catalytic *trans* esterification reactions), and many of the major components of the yeast or human spliceosome are conserved in *T. brucei* (Liang et al., 2003). *Trans* splicing is a prerequisite for protein expression, and changes in the poly(Y) tract, leading to a difference in *trans*-splicing efficiency, should be reflected in differences in protein levels. Given these characteristics, *T. brucei* is an excellent model to evaluate the effects of intronic sequence motifs on splicing efficiency, as one can rely on measuring protein levels derived from a carefully designed reporter system. Understanding sequence requirements for *trans*-splicing efficiency may help to predict splice sites and their probable efficiency. A systematic series of experimental data could also help design specific bioinformatics tools that, by identifying true splice sites, can distinguish true genes from random ORFs and thereby assist in the annotation of the *T. brucei* genome (Gopal et al., 2003).

Previously, the importance of a poly(Y) tract for *trans* splicing in *T. brucei* has been demonstrated by insertion of block substitution mutations (Huang and Van der Ploeg, 1991; Matthews et al., 1994). Additionally, it has been observed that block substitution mutations in the 5´ UTR of β-tubulin can affect *trans* splicing, demonstrating a role for 5´ UTRs in *trans* splicing (Lopez-Estrano et al., 1998). However, no extensive systematic study has been performed to determine minimal and optimal intronic or exonic sequence motifs required for efficient *trans* splicing.

To systematically study sequence requirements for *trans* splicing in *T. brucei*, I constructed a luciferase-β-galactosidase double-reporter system that allows a large number of sequence motifs to be evaluated with great sensitivity and reproducibility in transiently transfected cells to define their effects on splicing efficiency. I tested ~90 constructs to explore the roles of the composition, length, and position of the poly(Y) tract and the length and composition of the spacer region separating the poly(Y) tract from the 3´SS. I identified a single-nucleotide change in the mRNA 5´ UTR that could compensate for the low splicing efficiency observed when the 3´SS AG dinucleotide is preceded by AC.

Results

Construction of a double-reporter system

The aim of this study was to systematically characterize sequence motifs upstream of the ORF that are necessary for mRNA *trans* splicing in *T. brucei*. The 'upstream region' (UR) is defined as the region between the branch point and the 5´ UTR of the mRNA (Figures 6.1A and B) and contains a poly(Y) tract, a spacer region, and the 3´SS. I constructed a reporter system in which I could switch the sequence of the UR by using the *T. brucei* procyclin promoter to drive a luciferase (*LUC*) reporter gene flanked by the procyclin 5´ UTR and the aldolase 3´ UTR plus a short extraneous sequence derived from the vector from which *LUC* was originally cloned. I was able to use a protein assay to measure the *trans*-

splicing efficiency of pre-mRNA by ensuring that luciferase activity would only appear if the mRNA was correctly spliced at the single available AG dinucleotide, all others having been deleted, upstream of the single essential AUG translation initiation codon. Luciferase activity can be assayed over a linear range of 5 orders of magnitude. The readout is sensitive enough to use transient transfection, which is the only practical approach to testing a large number of constructs in trypanosomes. To permit normalization for differences in transfection efficiency, I included a *lacZ* reporter in the plasmid as an internal control. *lacZ* was flanked by sites for splicing and polyadenylation and was inserted downstream of the luciferase reporter gene (Figure 6.1B).

To allow the insertion of a large number of alternative URs, *Hin*dIII and *Sma*I restriction sites were incorporated adjacent to the 3´SS AG dinucleotide and at the upstream end of the UR, respectively. Insertion of the procyclin UR into my reporter construct led to high levels of luciferase activity (Figure 6.1C, pNS10/54). Removal of an ATG site located in the 5´ UTR, which gave rise to a small ORF within the 5´ UTR, caused a sevenfold increase in luciferase activity (Figure 6.1C, compare pLew20 and pNS11). All subsequent experiments were performed with constructs without an ATG site in the 5´ UTR. Elimination of four AG dinucleotides from within the luciferase 5´ UTR, leaving the construct with only a single available 3´SS, avoided the possibility that changing the length and/or composition of the UR would lead to the selection of a

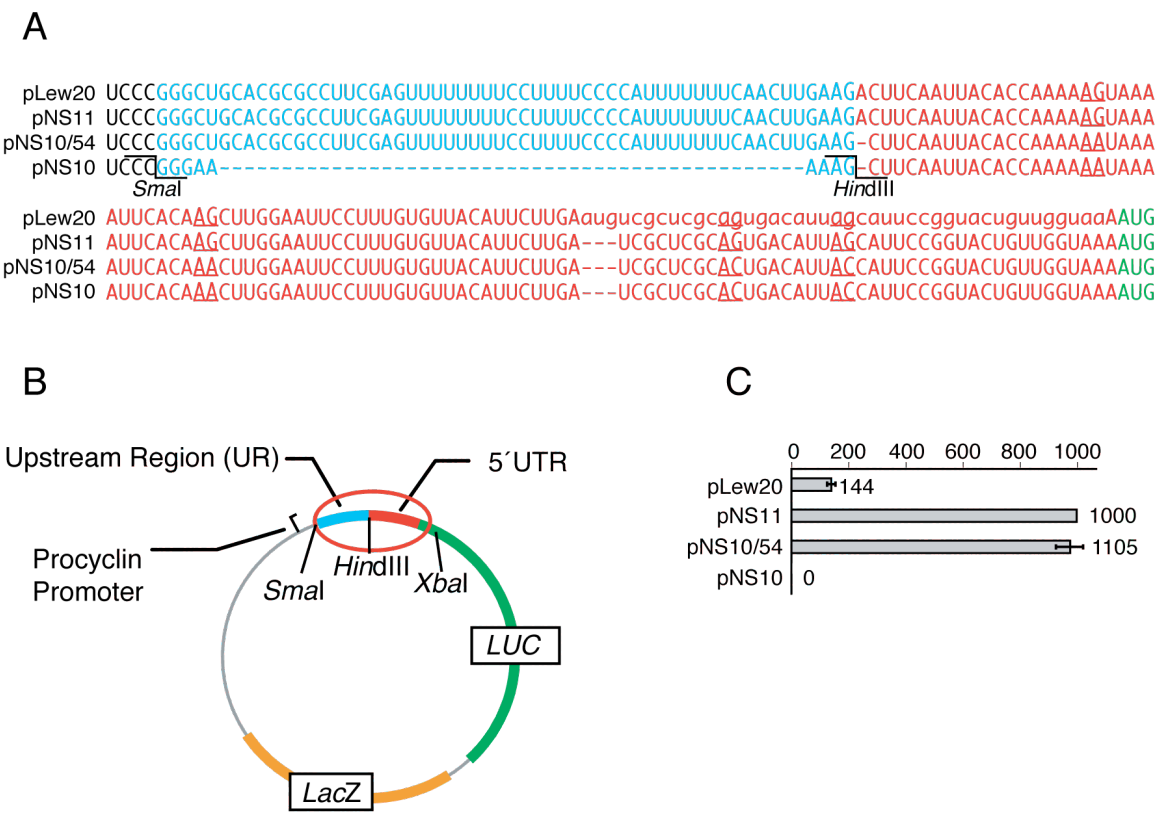


Figure 6.1. Design of a luciferase-β-galactosidase double-reporter system

(A) pLew20 is a construct containing a wild-type procyclin promoter, an upstream region UR (blue), and a 5´ UTR (red) preceding the luciferase ORF (*LUC*, green). Lowercase letters indicate a small ORF within the 5´ UTR that was carried over from the plasmid from which *LUC* was excised. pNS11 is a derivative of pLew20 in which the adventitious AUG was deleted from the 5´ UTR and a *lacZ* reporter gene was added downstream of *LUC*. In pNS10/54, all four cryptic AG sites were changed to AA or AC (underlined) and a *Hin*dIII site was introduced at the remaining AG splice site to facilitate the insertion of different URs. pNS10 lacks the UR.

(B) Organization of the pNS10/54 reporter plasmid, indicating key components and restriction sites.

(C) Luciferase activity in the basic reporter plasmids. All *trans*-splicing efficiency values shown are measurements of relative luciferase light units normalized to β-galactosidase activity and then to the pNS11 positive control, which was set to 1,000 in this and all subsequent experiments.

second cryptic splice site. This action did not decrease luciferase activity, demonstrating that those sites, in the wild-type procyclin UTR, had no essential role in regulating *trans* splicing (Figure 6.1C, pNS10/54).

Length and composition of poly(Y) tract

Previous reports on a wide variety of eukaryotes show the importance of a poly(Y) tract for efficient *cis* splicing, although *S. cerevisiae* can accurately remove introns that lack a poly(Y) tract, where a very conserved branch point sequence and splice site motif seem to provide the necessary recognition sites for the splicing machinery (Csank et al., 1990; Burge et al., 1999). In *T. brucei*, the available data suggest a minimal requirement of ~10 pyrimidines located between 10 and 40 nt upstream of the 3´SS (Huang and Van der Ploeg, 1991; Matthews et al., 1994). However, neither of these parameters, nor the optimum composition of the poly(Y) tract, has been systematically and precisely investigated. I tested poly(Y) tracts containing between 5 and 40 uridines (U) and observed a clear correlation between luciferase expression and poly(Y) tract length, in the context of two different spacer sequences between the poly(Y) tract and the 3´SS. Luciferase expression leveled out at ~25 U (Figures 6.2A and B).

To investigate the effect of poly(Y) tract composition on *trans* splicing, I sprinkled cytidine, adenosine, or guanosine within the poly(U) tract. Interspersion of a poly(U) tract with C's, generating a true poly(Y)

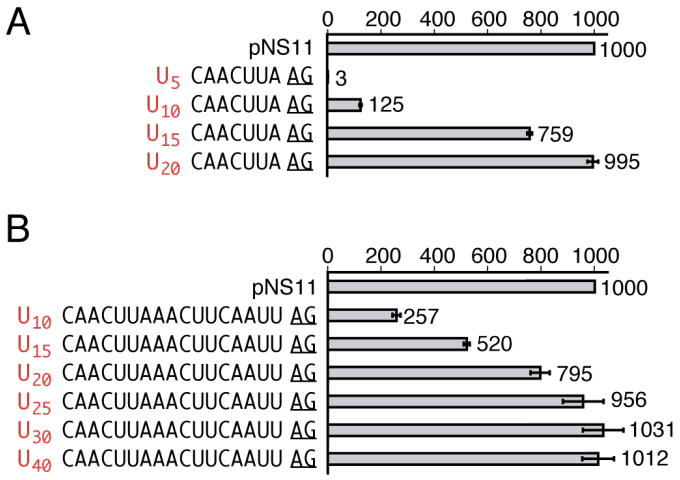


Figure 6.2. Effects of increasing poly(Y) tract length on *trans*-splicing efficiency.

Constructs containing URs with nonspecific 7-bp **(A)** or 17-bp **(B)** spacers and poly(Y) tracts of increasing length (red) show increasing *trans*-splicing efficiency which plateaus at ~25 uridines.

tract, caused a significant increase in *trans* splicing, but replacing a poly(U) tract with poly(C) eliminated *trans* splicing (Figure 6.3A). Interspersion of a poly(Y) tract with a single adenosine or guanosine caused a moderate decrease in *trans* splicing, but two or three consecutive purines had a severe effect (Figure 6.3B). These results suggest that continuity of the poly(Y) tract is important, rather than the presence of a minimum number of pyrimidines within an ~15-nucleotide window.

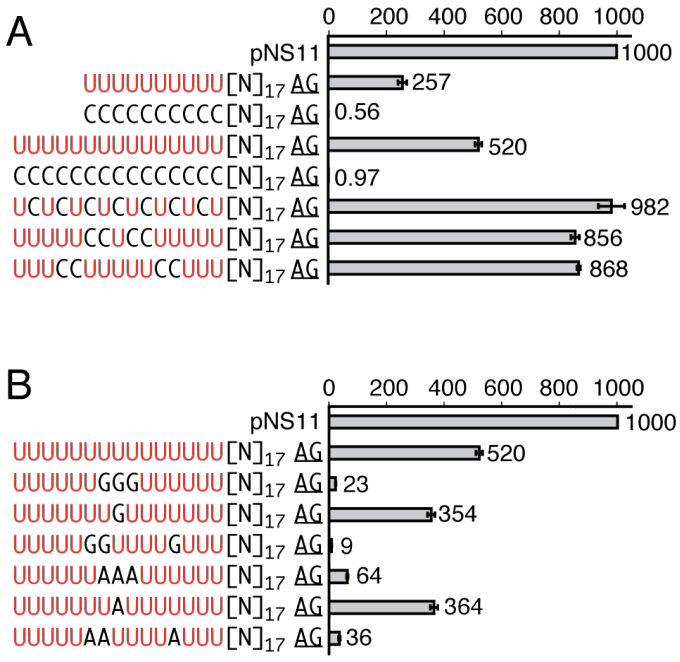


Figure 6.3. Effects of poly(Y) tract composition on *trans*-splicing efficiency
Constructs containing a 17-bp spacer were tested for the role of C in place of U (red) **(A)** and for the effect of A or G insertions into the poly(Y) tract **(B)**. $[N]_{17}$ = CAACUUAAACUUCAAUU.

Length and composition of spacer

It has been suggested that the position of the poly(Y) tract relative to the branch point and 3´SS may be important for efficient *trans* splicing (Metzenberg and Agabian, 1996). I therefore varied the distance between the poly(Y) tract and 3´SS in 3-nt increments while keeping all other parameters constant. My data show a clear correlation between spacer length and *trans*-splicing efficiency, suggesting an optimum spacer length of ~20 nt, which corresponds to a branch point to 3´SS distance of ~39 nt (Figure 6.4A). Subsequently, I looked for effects of spacer

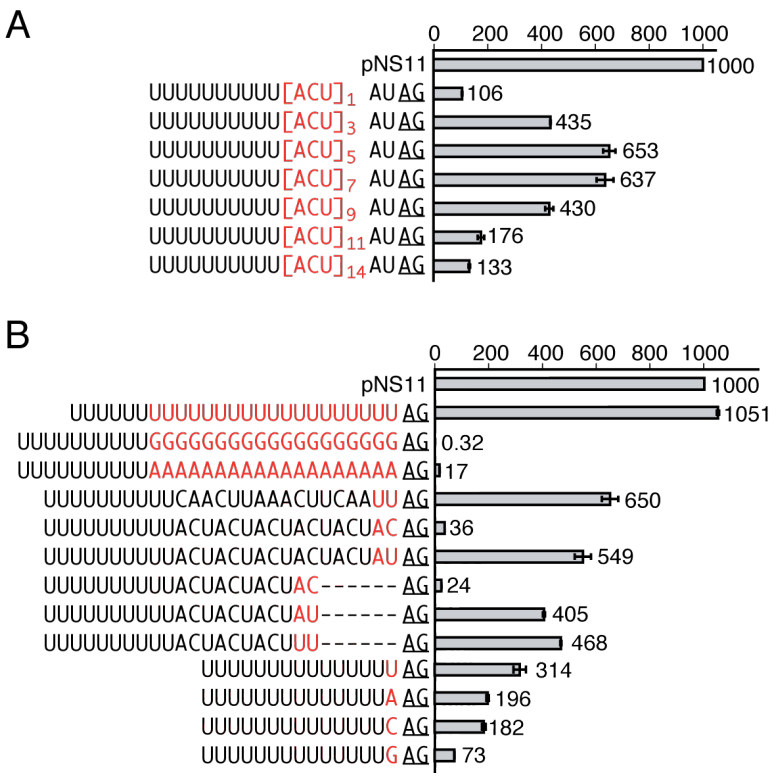


Figure 6.4. Effects of spacer length and composition on *trans*-splicing efficiency

Constructs containing a 10-bp poly(Y) tract were tested for the effect of variations in the length **(A)** and composition **(B)** of the spacer region (red). Changes of particular interest at positions -3 and -4 are marked in red in panel B.

composition on *trans*-splicing efficiency (Figure 6.4B). With homopolymeric spacer tracts, it was clear that a 3´SS that was preceded solely by uridines was significantly better than the heterogeneous spacer sequences included in this data set and at least as good as the natural upstream sequence derived from the highly expressed procyclin locus that was used as the positive control for all experiments in this study. The deleterious effects of a poly(A) spacer can probably be attributed to

the formation of an A-U duplex region that prevents the poly(U) tract from being recognized, and the extreme inhibition of *trans* splicing in the poly(G) spacer construct can probably be explained by the tendency of consecutive guanosine residues to form secondary structures that could block any scanning machinery from reaching the 3´SS or prevent splicing factors from binding to the UR.

The most interesting and surprising result from this set of constructs was the 20-fold drop in luciferase activity when the 3´SS AG dinucleotide was preceded by AC rather than AU (Figure 6.4B). This effect was independent of spacer length. No consensus sequence apart from the AG dinucleotide at the 3´SS has been documented in *T. brucei*, whereas humans and yeast appear to require a pyrimidine at the -3 position (YAG) (Burge et al., 1999). I therefore looked at the effect of the nucleotide at the -3 position in a poly(U) background [poly(U) tract and poly(U) spacer] on *trans*-splicing efficiency. I observed significant decreases in *trans*-splicing efficiency as I changed the nucleotide at the -3 position as follows: U to G, a fourfold decrease; U to A or C, a twofold decrease (Figure 6.4B). I will return to this conundrum in the last section.

Native poly(Y) tracts and spacer regions

My results identified several trends that affected splicing efficiency when synthetic UR sequences were used. Although a few studies had

compared the *trans*-splicing efficiencies of similar genes with different URs, I wanted to confirm that my reporter system would be influenced by differences in naturally occurring URs. I therefore tested the *trans*-splicing efficiency obtained with the native URs of several *T. brucei* genes. The *T. brucei* genome encodes three isoenzymes of phosphoglycerate kinase (PGK) that are coexpressed as consecutive sequences on a polycistronic pre-mRNA. This pre-mRNA gives rise to unequal amounts of PGK A, B, and C mRNAs. Very low levels of PGK A mRNA can be detected in procyclic *T. brucei*, compared to intermediate levels of PGK C mRNA and high levels of PGK B mRNA (Le Blancq et al., 1988; Gibson et al., 1988). Placing the corresponding PGK UR sequences in my reporter system, I observed very high levels of luciferase activity with the UR of PGK B, intermediate levels with the UR of PGK C, and levels barely above the background with PGK A (Figure 6.5), reflecting the natural mRNA levels. In a previous study, when the UR from each PGK gene was placed upstream of the luciferase ORF and the resultant constructs transiently transfected into *T. brucei* procyclic cells, luciferase activity levels also indicated different levels of splicing and correlated with mRNA levels (Kapotas and Bellofatto, 1993). Enzyme activity was low when the UR from the A gene was present but indistinguishable when the B and C URs were compared. A review of these results, however, suggests that limiting reaction conditions had probably prevented any difference in the B and C sequences from being observed. Surprisingly, the UR of the gene

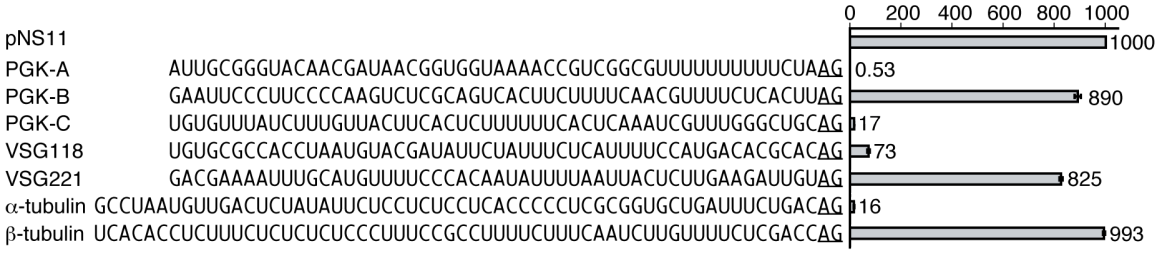


Figure 6.5. *Trans*-splicing efficiencies of selected native URs
The sequences upstream of the proven 3´SS for the three tandem PGK genes, two VSGs, and α- and β-tubulin are shown.

encoding variant surface glycoprotein 118 (VSG118) was rather poorly spliced (the VSG represents 10% of the total cellular protein of bloodstream *T. brucei*) compared to efficient synthetic URs or to the UR of VSG221. Interestingly, the UR of VSG118 has an AC dinucleotide preceding the AG dinucleotide, which I had observed to result in very low levels of *trans* splicing.

The observation that the URs of α- and β-tubulin led to drastically different levels of *trans* splicing (Figure 6.5) was even more surprising, as tubulin exists as a heterodimer containing one α and one β subunit, thus requiring equal amounts of the two subunits. The α-tubulin UR also contains an AC dinucleotide preceding the AG at the 3´SS. One possible explanation for my observation was that the previous 3´SS assignment was incorrect (Patzelt et al., 1989). However, I discarded this hypothesis after sequencing of α-tubulin cDNA clones confirmed the original assignment, which was also verified by primer extension (see below). A

second explanation could be that efficient *trans* splicing of α-tubulin genes depends upon exonic splice-enhancing elements within the α-tubulin 5´ UTR. Such splice-enhancing elements have been described in higher eukaryotes and are capable of activating weak upstream splice sites (Schaal and Maniatis, 1999a; Schaal and Maniatis, 1999b). Splicing enhancer elements consist of short RNA sequences, which are recognized by a number of splicing factors with characteristic serine/arginine (SR)-rich domains. Two such SR proteins have been described in *T. brucei*, but their exact function is unknown (Manger and Boothroyd, 1998; Ismaili et al., 1999; Ismaili et al., 2000). All of my constructs contained the 5´ UTR of the procyclin gene, which might be well suited for most genes, but some URs with weak poly(Y) tracts or a suboptimal 3´SS might require additional splice-enhancing sequences in their 5´ UTR. Previously, López-Estraño and colleagues demonstrated that block substitution mutations in the α-tubulin 5´ UTR negatively affect *trans* splicing at the upstream 3´SS AG (Lopez-Estrano et al., 1998). I decided to extend the study and pinpoint the responsible UTR element.

Role of the α-tubulin 5´ UTR in *trans* splicing

All of the experiments described so far used constructs containing a procyclin 5´ UTR in which I replaced four cryptic AG sites with AA or AC. To test the hypothesis that the 5´ UTR contains sequence elements that, in some contexts, are important for efficient *trans* splicing, I first

changed the 5´ UTR in my constructs in small blocks to match the 5´ UTR from α-tubulin. I observed a 60-fold increase in luciferase expression after changing nucleotides 9 to 18 nt downstream of the 3´SS to match the sequence present in endogenous α-tubulin 5´ UTRs (Figure 6.6). These 10 replacement nucleotides contained an AG dinucleotide, and a single-nucleotide change showed that high luciferase expression depended upon this AG (compare pNS10/74 and 10/75). Several other single-nucleotide substitutions in the vicinity of the AG site had no effect on *trans*-splicing efficiency (data not shown). I therefore needed to exclude the possibility that the second AG site was being used as the 3 ´SS. Because of the low luciferase mRNA levels in transiently transfected trypanosomes and my inability to distinguish the two potential alternative products by quantitative reverse transcription-PCR, I performed primer extension analyses on clones that had been stably transfected with a version of pNS10/74 (pNS20/74) in which *lacZ* had been replaced with a gene encoding phleomycin resistance. With two primer pairs, I demonstrated that the first AG site functioned exclusively as a splice site; no product corresponding to the second AG was observed (Figures 6.7A and C). Primer extension experiments with the endogenous α-tubulin and β-tubulin mRNAs in these transfected clones confirmed the previously assigned splice sites (Figures 6.7B and C). Based on these findings, I concluded that the presence of a second AG site in the 5´ UTR

enhances *trans* splicing in constructs that contain a suboptimum sequence (AC) immediately upstream of the 3´SS.

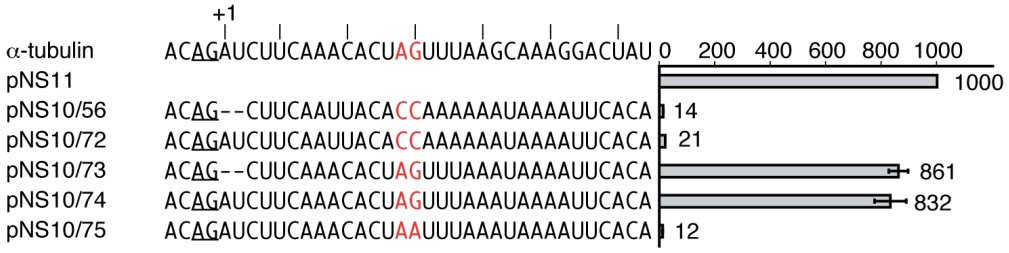


Figure 6.6. One nucleotide in the 5´ UTR of α-tubulin is critical for splicing
The upper sequence shows the endogenous α-tubulin 5´ UTR. Sequences are shown from the -4 position relative to the 3´SS, and all constructs contain the α-tubulin UR. Only the 5´ UTR sequences differ. pNS10/56 contains the procyclin 5´ UTR lacking cryptic AG sites (red). In pNS10/72, 8 nt downstream of the 3´SS have been changed to match the sequence found in the endogenous α-tubulin 5´ UTR. pNS10/73 contains a 5´ UTR sequence identical to the α-tubulin 5´ UTR for 20 nt downstream of the 3´SS, except for the omission of nucleotides at positions +1 and +2. pNS10/74 is identical to pNS10/73 but with the +1 and +2 nucleotides intact. In pNS10/75, the AG within the 5´ UTR of pNS10/74 was changed to AA (red).

Discussion

Role of poly(Y) tract

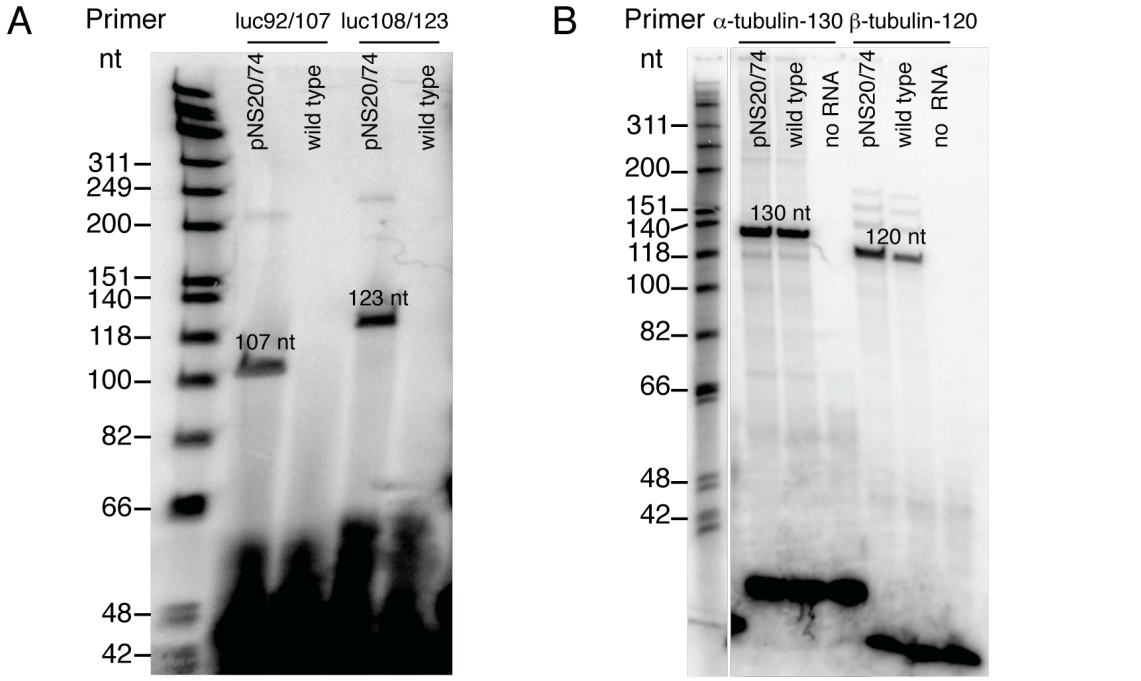
In this study, I demonstrated that *trans* splicing in *T. brucei* depends upon the length, composition, and position of a poly(Y) tract. Furthermore, my data suggest that changes in length between the branch point and splice site are acceptable as long as the poly(Y) tract and spacer region fulfill certain requirements. My observations are

Figure 6.7. Primer extension analysis of the 3´SS used in pNS20/74 and by endogenous α- and β-tubulin

(A) Twenty micrograms of cellular RNA and primers specific for pNS20/74. The entire reaction product was loaded, and the gel was exposed to a phosphorimager screen for 60 hours.

(B) Ten micrograms of RNA and primers specific for α-tubulin and β-tubulin. Half the reaction product was loaded, and the gel was exposed to a phosphorimager screen for 15 hours. One major product was seen in each analysis.

(C) Schematic depiction of possible primer extension products. The expected lengths include the 39 nt of the miniexon. Cryptic splice sites are marked in red. AUG (blue) marks the beginning of the ORF.



C

pNS20/74

acAGAUCUUCAAACACUAGUUUAA[N]$_{20}$AUUCCUUUGUGUUACAUUCUUGAUCGCUCGCACUGACAUUACCAUU[N]$_{17}$ AUG...

107 nt ← 92 nt ← GGAAACACAATGTAAGAACTAGCGA primer: luc92/107

123 nt ← 108 nt ← GAACTAGCGAGCGTGACTGTAATGGT primer: luc108/123

α-tubulin

acAGAUCUUCAAACACUAGUUUAAG[N]$_{26}$AGCAACAGUAGGUACUAGCACCACUAACAACAACAACAAAGCAC[N]$_{15}$ AUG...

130 nt ← TCGTGGTGATTGTTGTTGTTTCGTG

β-tubulin

ccAGUCCUACUAGAGGAGAAAGAAUAG[N]$_{15}$AGAAAAUAGUUCAAACGAAUUAUGCGCGAAAUCGUCUGCGUUC...

120 nt ← ATACGCGCTTTAGCAGACGCAAG

consistent with recently published data indicating that *T. brucei* genes do not necessarily contain a single branch point and that different nucleotides may serve as branch points (Lucke et al., 2005). One of the early events in spliceosome assembly is the recruitment of U2 snRNP to the branch point sequence. In mammalian cells, U2 recruitment has been confirmed to depend upon the two subunits of U2AF. The 65-kDa subunit has been shown to bind to the poly(Y) tract, while the 35-kDa subunit binds to the 3´SS AG (Wu et al., 1999; Zamore et al., 1992). Further evidence that poly(Y) tract variations affect the first of the two catalytic steps stems from research with HeLa cells, where no poly(Y) tract is necessary for the second step (Anderson and Moore, 1997; Chen et al., 2000).

No trypanosome homolog to the human U2AF65 subunit has been found, although database mining has revealed proteins with domains that could potentially bind to pyrimidine-rich RNA tracts. In contrast, the *T. brucei* genome contains a relatively conserved U2AF35 homolog. A similar homolog has been studied more closely in *T. cruzi* and, interestingly, like the fission yeast homolog, it is missing the C-terminal SR domain. In addition, the *T. cruzi* homolog is lacking Trp 134, which is highly conserved in other organisms and has been shown to allow the human U2AF35 subunit to participate in a tongue-in-groove interaction with its larger partner U2AF65 (Kielkopf et al., 2001; Vazquez et al., 2003; Wentz-Hunter and Potashkin, 1996). The absence of these

conserved residues appears plausible in the absence of U2AF65.

Poly(Y) tract lengths shorter than 10 pyrimidines may interfere with the binding of such poly(Y)-binding proteins, thereby affecting initial assembly of the spliceosome machinery. This hypothesis is in agreement with my data showing that shorter poly(Y) tracts are only well tolerated as long as they are not positioned too close to the 3´SS. My results suggest the absence of a threshold poly(Y) tract length for optimum binding. Incremental increases in poly(Y) tract length led to a continuous increase in *trans*-splicing efficiency, which ultimately reaches a plateau.

Identification of the 3´SS

In humans, there is a very strong preference for a pyrimidine preceding the AG at the 3´SS and a strong preference against G (Burge et al., 1999). In *T. brucei*, there is a small but statistically significant underrepresentation of G in the -3 position, but the other nucleotides are found with equal frequency (Shuba Gopal, personal communication).

However, with the procyclin 5´ UTR, I observed very low splicing efficiency when the -3 and -4 positions were occupied by A and C, respectively. These results suggest, as did earlier work (Lopez-Estrano et al., 1998), that certain URs might require distinct splice enhancer motifs in the 5´ UTR for efficient *trans* splicing. In HeLa cells, a very heterogeneous population of splice enhancer sequences has been described (Graveley et al., 2001; Schaal and Maniatis, 1999b). In

contrast, I was able to show that a single downstream AG site was required for efficient *trans* splicing at the upstream AG site. This result is consistent with earlier observations that show that an AG dinucleotide, newly introduced upstream of the original 3´SS AG dinucleotide, can function as a new 3´SS as long as the downstream AG is present (Zhuang and Weiner, 1990). These findings can be explained by the requirement of U2AF35 binding to the 3´SS splice during early stages of spliceosome assembly.

My data can be explained assuming, first, that the first AG site functions as a splice site and is recognized by a scanning mechanism that identifies the first AG downstream of the branch point and poly(Y) tract; second, that the first AG site does not function as a U2AF35 binding site, possibly due to unfavorable residues at positions -3 and -4; and third, that U2AF35 does not have to bind at the 3´SS AG itself as long as it can bind to an AG close by, allowing interactions with a putative poly(Y) tract binding protein, helping the latter to bind to the poly(Y) tract.

The strong effect of the 5´ UTR on *trans*-splicing efficiency observed in my study, compared to a more moderate effect observed by others (Lopez-Estrano et al., 1998), probably stems from the fact that I previously deleted all other cryptic AG sites, whereas (Lopez-Estrano et al., 1998) only deleted one AG site per block substitution mutation.

Data from studies with HeLa cells suggest a scanning model for the

recognition of the 3´SS AG prior to the second catalytic step (Smith et al., 1993; Smith et al., 1993). This model is supported by data showing that sequences that can form stable secondary structures lead to a dramatic decrease in splicing efficiency, potentially by blocking factors scanning along the RNA from the branch point to the splice site (Chen et al., 2000). I observed little splicing of constructs containing poly(A) or poly(G) spacers, which would probably form secondary structures, with the upstream poly(U) tract or otherwise.

The data obtained in this study should aid in the development of algorithms to identify splice sites computationally in *T. brucei* and in other members of the order *Kinetoplastida*. There may be differences within this broad family, however, as the poly(Y) tracts of *L. major* are predominantly C, whereas those of *T. brucei* are predominantly U (Shuba Gopal, personal communication).

Predicted *trans*-splicing efficiency is in agreement with levels of mRNA

In this study, I exploited the advantage of enzymatic assays (luciferase and β-galactosidase) to study *trans*-splicing efficiency. However, many factors besides *trans*-splicing efficiency might influence enzyme levels in a cellular context (RNA stability or protein degradation, for example). I am confident, however, that my assays accurately reflect the amount of *trans*-spliced RNA, since only exonic sequence elements

should influence RNA stability, and no changes in the 5´ UTR or coding region were introduced when analyzing the UR. In my analysis of the 5´ UTR, I did change parts of the mRNA sequence, but it seems unlikely that those changes, as little as a single nucleotide change generating an additional AG site, affected RNA stability, since the deletion of four cryptic AG sites from the procyclin 5´ UTR had no effect on luciferase activity.

Previous work showed that the PGK A, B, and C mRNA levels differ, although they are transcribed as a polycistronic unit (Gibson et al., 1988). These results led the authors to suggest that the differences in mRNA concentration must be controlled post-transcriptionally, by differences in *trans*-splicing efficiency. I was able to confirm that the different URs for PGK A, B, and C indeed resulted in very different levels of *trans*-splicing efficiency. These results strongly suggest that *trans* splicing can play an important role in posttranscriptional gene control.

Chapter 7: Cell-cycle assignment by quantitative DAPI imaging

Summary

T. brucei has two DNA compartments – the nucleus and the kinetoplast. DNA replication of these two compartments only partially coincides. Woodward and Gull (Woodward and Gull, 1990) comprehensively studied the relative timing of the replication and segregation of nuclear DNA (nDNA) and kinetoplast DNA (kDNA). Others have since assumed the consistency of morphological indicators of cell-cycle stage among strains and conditions. I used quantitative DAPI imaging to determine the cell-cycle stage of individual procyclic cells. Using this approach, I found that kinetoplast elongation occurs mainly during nuclear S phase and not during G2, as previously assumed. I confirmed this finding by sorting cells by DNA content, followed by fluorescence microscopy. In addition, simultaneous quantitative imaging at two wavelengths can be used to determine the abundance of cell-cycle-regulated proteins during the cell cycle. I demonstrate this technique by co-staining for the non-acetylated state of histone H4K4, which is enriched during nuclear S phase.

Introduction

Members of the order Kinetoplastidae are characterized by the presence of the kinetoplast, the uniquely structured mitochondrial DNA that consists of an interlocked network of several thousand minicircles and more than 20 maxicircles. Understanding the mechanism of kDNA replication has been of great interest (Barry et al., 2007; Liu et al., 2005). Initiation of kDNA and nDNA replication probably coincides, but kDNA segregation is completed before the onset of mitosis (Woodward and Gull, 1990). Any cell can hence be assigned to a specific cell-cycle stage by comparing nuclear and kinetoplast morphology. For example, cells with an elongated kinetoplast would be in the G2. However, my findings described in Chapter 3 suggest that elongated kinetoplasts exist in nuclear S phase as well.

The goal of this study was to determine the cell-cycle stage of individual cells solely by their nDNA content. Using fluorescence microscopy and deconvolution algorithms, I quantified nDNA by DAPI-staining. In contrast to Woodward and Gull (Woodward and Gull, 1990), I found that many cells with elongated kinetoplasts are still in nuclear S phase. This approach allows cell-cycle stage determination without relying on kDNA morphology, and avoids the more complicated approach of antibody staining for known cell-cycle markers. It also can be used to quantify the abundance of cell-cycle-regulated proteins or protein modifications, during the cell cycle. To validate my cell-cycle

assignments, I performed fluorescence microscopy on sorted live cells whose DNA was stained with DyeCycle Orange (DCO, Invitrogen). Advances in FACS technology and the development of new dyes make cell sorting a good approach for analysis of cells that cannot be easily synchronized.

Results and Discussion

Quantification of images requires images of high quality. Spherical aberration, non-uniform illumination, and differential response of sensor pixels need to be avoided or corrected for. Spherical aberration occurs when light waves passing through the periphery of an objective are focused to a different position on the *Z*-axis than those passing through the center of the objective. This creates a significant asymmetry of the point-spread function (the observed three-dimensional distribution of light intensity from an ideal point source). In practice, this asymmetry appears as a series of rings on either side of the focal plane (depending on imaging conditions). Spherical aberration can best be identified when rotating the Z-stack of images and looking along the *Z*-axis (Hiraoka et al., 1990; North, 2006). The image should be symmetric, out-of-focus light appearing as much below as above the focal plane. Most modern objectives are designed to minimize spherical aberration artifacts under specific conditions such as coverslip thickness (generally 170 μm), short distance between specimen and coverslip (less than a few microns),

specified temperature (generally 25°C), and immersion-oil refractive index (RFI), usually 1.516–1.518.

I used a DeltaVision image restoration system to acquire Z-stacks of images. In order to image a large number of cells, I chose an objective lens with 60-fold magnification and number 1.5 glass coverslips (Fisher Scientific). Fisher and other vendors supply coverslips within a thickness range, which is supposedly 160–190 μm for number 1.5 coverslips. Using a micrometer (Mitutoyo Digimatic), I measured several of my coverslip batches and found them to be between 180 μm and 190 μm. Differences in what the objective 'expects' and what is provided can be compensated for by changing the RFI of the immersion oil accordingly. Ten μm of extra coverslip thickness can be compensated for by reducing the RFI of the immersion oil by 0.007. Similarly, an increase of temperature should be accompanied by a drop of RFI (0.001/2.5°C). To reduce the distance between coverslip and specimen, cells were not settled onto the slide but were settled directly onto aminopropyltriethoxysilane-coated coverslips (see Chapter 2). In addition, the attached cells were centrifuged at 800 × *g* for 2 min at room temperature, to achieve flat adhesion, as spherical aberration increases with specimen thickness. Coverslips were mounted in antifade solution containing DAPI (Vectashield, Vecta Laboratories). Different mounting media did not seem to affect spherical aberration, which I minimized by using a graded set of immersion oils (Applied Precision).

Statistically valid cell-cycle analysis requires comparison of several hundred cells across many images. Even after careful alignment of the light bulb, I noticed uneven illumination of the field of view. Uneven illumination can be best seen after strongly increasing the contrast of the image. To compensate for uneven illumination and differential response of pixels, I used a calibration slide (Applied Precision), focused within the slide, to avoid variations on the slide surface, and generated a calibration standard according to the manufacturer's instructions. Once the calibration standard has been recorded, it can be applied automatically to all images during the imaging process. Calibrations need to be performed for each wavelength and image frame resolution. Since image quality was visibly better towards the center of the maximum image frame of 1024 × 1024 pixels, I decided to use an image frame of 768 × 768 pixels.

Three-dimensional deconvolution microscopy (Agard et al., 1988) requires acquisition of a stack of images at narrowly spaced parallel focal planes. Microscope objectives also collect light from out-of-focus objects, which normally blurs each image. For a stack of images, deconvolution algorithms iteratively approximate the true distribution of objects by reassigning out-of-focus light, using a point spread function previously measured for each objective. This sharpens the image and preserves the total light intensity. I use a proprietary deconvolution algorithm included in the softWoRx 3.51 software package (Applied Precision). SoftWoRx

contains different deconvolution algorithms and, after comparison of the default 'ratio' and the milder 'enhanced additive' methods, I chose the latter, as this seemed to better handle residual spherical aberration apparent at the top or bottom of the specimen. Residual mismatch between the observed image and the expected image based on the inferred true distribution of objects reached its best value after about seven deconvolution iterations, which I confirmed visually.

To test the suitability of my image-based quantification, I analyzed two sets of 22 Z-stacks. Z-stacks containing 100 images were acquired at 100 nm *z*-intervals. To achieve consistency among images, each set of Z-stacks was recorded in one session from one coverslip, at a constant temperature and constant exposures. Temperature was kept at 25°C using a Weather Station (Precision Control). To avoid pixel saturation, optimal exposure was determined prior to image acquisition. On average, a Z-stack yielded ~12 usable cells in which the kinetoplast did not overlap with the nucleus in the Z-projection (Figure 7.1A). Cells for which the nDNA content could be measured were grouped manually, based on the presence of one short kinetoplast, one elongated kinetoplast, or two short kinetoplasts. For the purpose of this study, late mitotic cells and cells with completely separated nuclei were grouped with G1/G0 cells. I classified all kinetoplasts that were clearly elongated, dumb-bell shaped, or bi-lobed, as elongated. Surprisingly, the majority of cells with an

elongated kinetoplast had an nDNA content corresponding to nuclear S phase (Figure 7.1B).

nDNA fluorescence-based cell-cycle assignment could be especially useful for quantification of cell-cycle regulation of proteins or protein modifications. To test this application, I imaged cells fluorescently stained with an antibody to the unmodified histone H4K4 site. As described in Chapter 3, this histone mark is strongly enriched during nuclear S phase. If fluorescence emission is measured at different wavelengths, it is important to correct for chromatic aberration. Both magnification and focal length vary with wavelength, so objects may appear not to co-localize even if they actually do. This is especially important in three-dimensional microscopy, as most microscopes are corrected only for lateral aberration (North, 2006). In addition, even most modern microscopes are only properly corrected for chromatic aberration between green and red wavelengths (North, 2006). To correct for chromatic aberration, I took advantage of the long emission tail of the DAPI signal. I excited DAPI at 360 nm and measured its light emission not only at its peak (457: filter bandwidth 50 nm) but also in the green (528: filter bandwidth 38 nm) area of the spectrum. In the absence of chromatic aberration, the images recorded at 457 nm and 528 nm should co-localize. For the objective used, the blue and green images were displaced by 0.6 μm in the Z plane. I corrected for this shift using softWoRx, then scored nuclei as either positive or negative for antibody

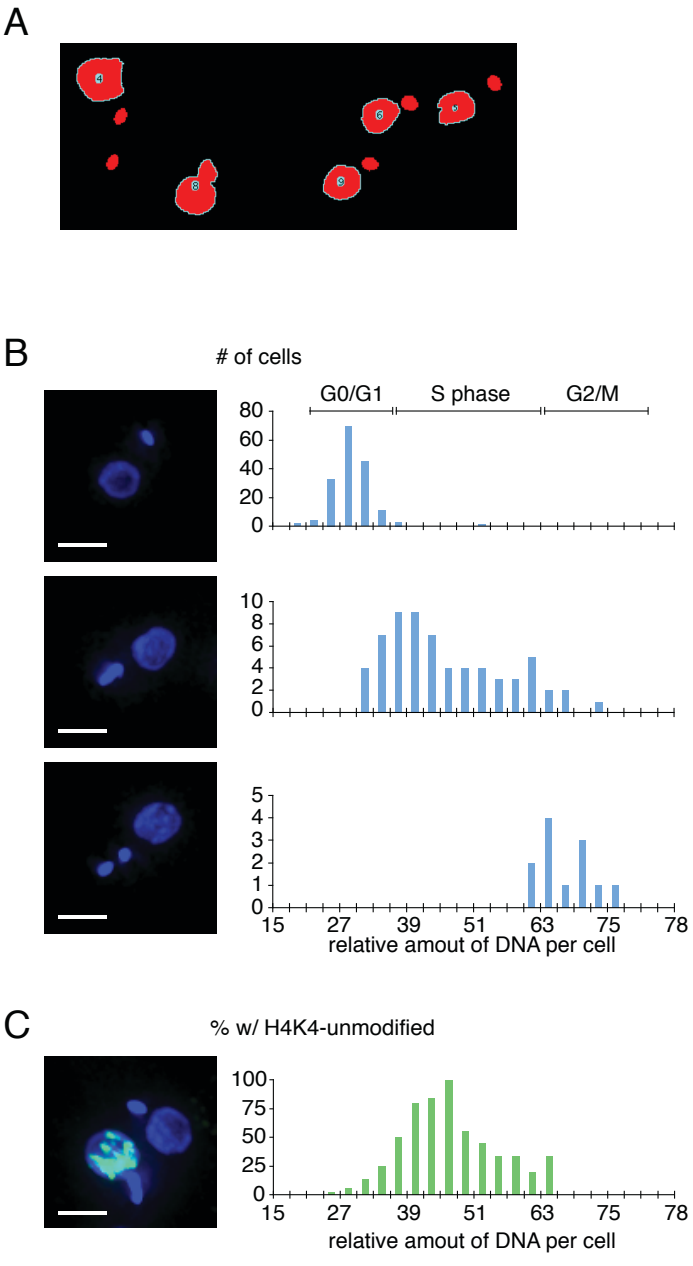
Figure 7.1. Fluorescence microscopy and signal quantification

Exponentially growing (~1.0 × 10^7 cells/ml) PF cells of *T. brucei* strain Lister 427 were used for all experiments. Images were acquired on a DeltaVision image restoration system (IX70 Olympus microscope) with a Plan Apo N 60x/1.42 NA objective (Olympus) and Photometrics CoolSnap QE camera. Deconvolution and pseudo-coloring were performed using softWoRx™ v3.5.1.

(A) To measure DAPI fluorescence from individual cells, background fluorescence was subtracted and fluorescent objects were identified in the individual image layers based on a signal intensity threshold. Next, the layer intensities were summed, signal from the kinetoplast was excluded based on a size threshold, and all remaining objects were numbered. To automate this process, I wrote a script for the freely available imaging program, imageJ (http://rsb.info.nih.gov/ij/). The figure shows the nuclear boundary and numbering determined by the automated script.

(B) Representative images are shown. For 244 cells, signal intensity was measured using imageJ 1.38v, as illustrated in panel (A).

(C) Images were collected and quantified as in (B) except that unmodified H4K4 was stained with a specific rabbit antibody and a secondary antibody conjugated to Alexa Fluor 488, green (Invitrogen). 230 cells were analyzed. Scale bar represents 3 μm.


A
B
# of cells
G0/G1
S phase
G2/M
80
60
40
20
0
10
8
6
4
2
0
5
4
3
2
1
0
15
27
39
51
63
75
78
relative amout of DNA per cell
C
% w/ H4K4-unmodified
100
75
50
25
0
15
27
39
51
63
75
78
relative amout of DNA per cell

staining, based on the summed intensity in the green channel over DAPI-positive pixels. To remove background signal, I only counted nuclei as positive for H4K4-unmodified staining if the summed intensity was above 50,000 relative light units (18.6% of the maximum intensity observed). Results indicating that unmodified H4K4 is enriched in S-phase cells are shown in Figure 7.1C.

Lastly, I validated my findings using flow cytometry. DNA of fixed or live cells was stained with propidium iodide or with DCO, respectively. Cells were sorted according to their total DNA content (this reflects nDNA content because kDNA represents only ~5% of total DNA in *T. brucei* (Borst et al., 1982)), then analyzed by fluorescence microscopy (see Chapter 2). Tubes to collect the sorted cells were coated with bovine serum albumin, to prevent cells from sticking to the wall of the tubes. The sorted cells were fixed in suspension with formaldehyde and analyzed by fluorescence microscopy. In agreement with published literature (Woodward and Gull, 1990), G1 cells contained one nucleus and one kinetoplast and G2/M cells contained one nucleus and two kinetoplasts (Figure 7.2A, top and bottom panels). The majority of nuclear S-phase cells contained an elongated kinetoplast (Figure 7.2A, middle panel), confirming my observations from nDNA fluorescence-based cell-cycle determination and agreeing with the findings described in Chapter 3. Results from fixed and live cells were comparable, but FACS sorting of live cells generated significantly cleaner sorts, maybe

because live cells are less sticky than fixed cells. Post-sort analysis of live cells confirmed very homogenous populations and no further cell-cycle progression was observed (Figure 7.2B).

Previously, the non-coincident division of the nucleus and kinetoplast has served as a convenient marker for cell-cycle determination. In their seminal studies, Woodward and Gull (Woodward and Gull, 1990) outlined the relationship between nuclear DNA replication and kinetoplast replication and division, and observed kinetoplast division (elongated kinetoplast) during G2. By my approach of image-based cell-cycle determination, I surprisingly found that the majority of cells with an elongated kinetoplast were in nuclear S phase. FACS sorting of cells based on relative DNA content followed by fluorescence microscopy confirmed this finding. The observed difference may be due to differences in cell-doubling time. The cells used in the original studies had a doubling time of 8.65 hours compared to 12.5 hours in my experiments. It has been observed that inhibition of nuclear S phase but not kinetoplast S phase leads to an increase in cells containing one nucleus and two kinetoplast, suggesting that kinetoplast division can proceed despite nuclear S-phase arrest (Ono Ploubidou et al., 1999; Ono and Nakabayashi, 1980). If nuclear and kinetoplast cell-cycle progression are uncoupled, a longer nuclear G0/G1 or S phase may lead to a relatively earlier onset of kinetoplast division, which would explain the observed difference.

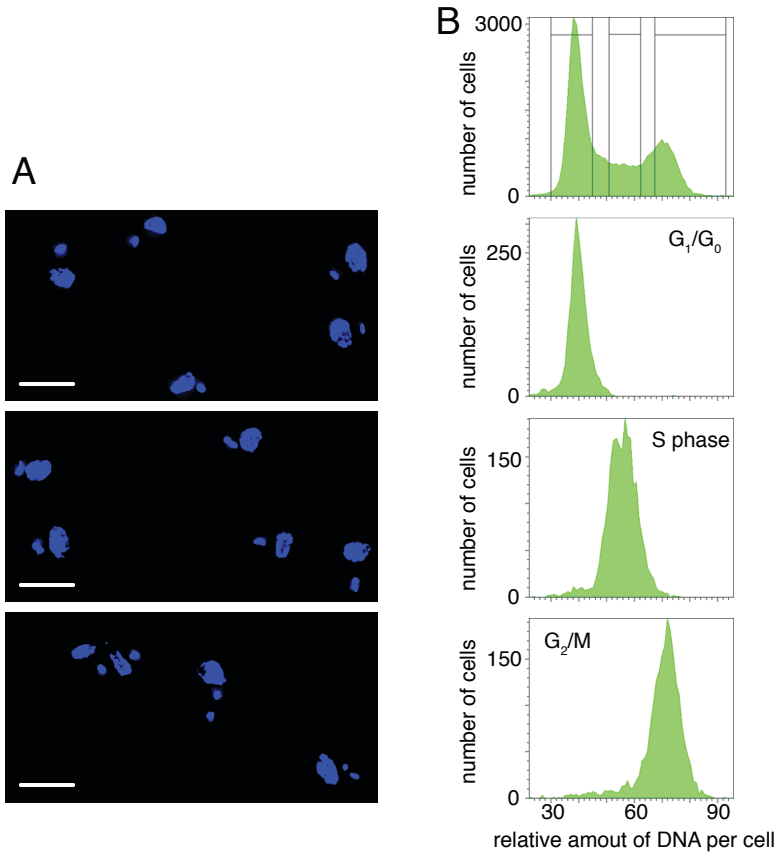


Figure 7.2. FACS sorting followed by IF microscopy

Unfixed PF (~1.3 × 10^7 cells/ml) cells were stained in 5 mM DyeCyclin Orange (Invitrogen) in SDM-79 (Brun and Schonenberger, 1979) for 30 min at 27°C. Cells were sorted based on total relative DNA content, using a FACSAria (BD Biosciences).

(A) The sorted cells were fixed with formaldehyde and imaged as described in Figure 7.1. Panels from top to bottom represent cells in nuclear G0/G1, S phase and G2/mitosis, respectively. Scale bar represents 5 µm.

(B) FACS profiles for unsorted (top panel) and sorted cell populations (lower panels).

Conclusion

In this study, I have established a method for image-based cell-cycle determination. This approach can be used to assign a cell-cycle-regulated protein to a specific cell-cycle stage. Since the cell-cycle determination was based solely on DNA fluorescent staining and did not rely on any other markers, no additional emission channels were occupied. This leaves both the green and red channel to stain proteins of interest. By measuring the fluorescence from DAPI and H4K4-unmodified, I was able to prove the feasibility of the approach and to confirm that H4 unmodified at lysine 4 is strongly enriched in S-phase cells, where visual inspection of the kinetoplast would misleadingly have assigned these cells to G2.

The advantages of using live cells include faster set-up, cleaner sorts, and easy post-sort analysis by SDS-PAGE. FACS sorting of live cells followed by SDS-PAGE analysis appears to be a good alternative to synchronizing cells, which has been a great challenge in *T. brucei*.

Chapter 8: General Discussion and Future Directions

What is the function of H4K4ac?

I generated specific antibodies to acetylated and unmodified histone H4K4 and, using IF microscopy and western blotting analyses, I observed a strong enrichment of unmodified H4K4 in S phase. Furthermore, my studies suggested a G1/G0-specific masking of H4K4 due to a non-covalently binding factor. Finally, I was able to demonstrate acetylation of H4K4 by the non-essential HAT3.

Eighty percent of H4K4 sites are acetylated, making it one of the more, if not the most, highly acetylated site in *T. brucei*. Yet, my data indicate that this acetyl mark is not essential and therefore raise the question, why trypanosomes would acetylate H4K4 on almost all histones if they grow well without it?

Redundancy among histone H3 and histone H4 acetylation sites is well established in other eukaryotes. However, because in *T. brucei* no other lysine on histone H4 is acetylated more than 10% of the time, redundancy does not answer the question why 80% of H4K4 sites are acetylated if a second minor acetyl mark could substitute for the loss of the major H4K4 acetylation. Even upon deletion of HAT3 and loss of H4K4 acetylation, no newly acetylated sites could be detected, and acetylation levels of all previously identified H4 acetylation sites

remained unchanged. Because I quantified the levels of individual acetyl marks only on histone H4, one possibility is that, upon deletion of HAT3, an acetyl mark on another histone is upregulated and compensates for the loss of H4K4ac. This explanation appears unlikely because only five KATs have been identified in *T. brucei*, two of which (Elp3a and Elp3b) show highly localized distribution and are thus unlikely to be able to compensate for the loss of a widespread acetyl mark like H4K4. Instead, it may be that acetylation of H4K4 is not essential under the conditions tested, i.e. standard culture conditions or growth in mice and rats. The following two scenarios appear more plausible:

First, H4K4ac confers a subtle, yet undetected, growth advantage throughout the parasite life cycle. For example, such an advantage could arise through a role of H4K4ac in DNA repair.

Second, H4K4ac is only essential during a distinct stage of the life cycle that has not been reproduced under laboratory conditions. For example, H4K4ac may be required for survival in the tsetse midgut or in the human bloodstream. Alternatively, H4K4ac could play a role during differentiation from one life-cycle stage to the next.

To address the question whether H4K4ac confers a growth advantage and to detect even minimal differences in growth rate, one could mix a $HAT3^{+/+}$ and a $HAT3^{-/-}$ cell line and determine, after weeks or months of continuous subculturing, whether one cell type outgrows the other. Such a comparison could be performed by analysis of genomic

DNA or, if the cell lines are tagged with fluorescent markers, by FACS analysis. To avoid effects specific to the respective fluorescent markers, a control should contain the reverse combination of markers.

Of course, even if loss of H4K4ac led to a growth defect, further work would be required to identify the function of the mark. In other eukaryotes, acetylation of H4K5, possibly the homologous site to *tb*H4K4, has been shown to play a role in histone deposition (Sobel et al., 1994; Sobel et al., 1995), cell cycle progression (Megee et al., 1995), transcription activation (Schiltz et al., 1999), and DNA damage repair (Bird et al., 2002).

The role of H4K4ac in DNA repair has been investigated by members of the David Horn Laboratory (David Horn, personal communication). Initially they tested the role of H4K4ac in DNA damage repair by exposing $HAT3^{+/+}$ and $HAT3^{-/-}$ cell lines to the DNA alkylating agent methyl methanosulfonate, but no reproducible effect on cell growth could be detected. However, in subsequent studies it was observed that RAD51, a factor of the double stranded break (DSB) repair pathway that formed distinct foci at DSBs in wild type cells (Glover et al., 2008), failed to localize to DSBs in $HAT3^{-/-}$ cells. This finding suggests that HAT3 may play a role in RAD51-mediated DSB repair. Incidentally, just like HAT3, RAD51 is not essential for viability in *T. brucei* (McCulloch and Barry, 1999). Given that most H4K4 sites are acetylated, it may also be interesting to investigate the role of HDACs in DNA repair. Maybe in the

vicinity of DSBs H4K4 is deacetylated and the effect of HAT3 deletion on Rad51 localization in only indirect.

Are H4K10ac, H2AZ and H2BV epigenetic marks?

I established a ChIP-seq protocol for *T. brucei* and used custom-made antibodies and tagged cell lines to determine the genome-wide distribution of H4K10ac, H2AZ, H2BV, H3V, H4V and BDF3.

My data reveal a strong enrichment of H4K10ac, H2AZ, H2BV and BDF3 at probable RNA pol II TSSs. In contrast, the variants of histone H3 and of histone H4 are enriched at convergent-SSR and other sites corresponding to putative ends of RNA pol II PTUs. Taken together, my ChIP-seq data suggest that *T. brucei* has a dedicated set of histone variants enriched at RNA pol II TSSs and TTSs. Finally, computational analysis of RNA pol II TSSs revealed a peculiar motif containing long stretches of guanines that I speculate could lead to directional RNA pol II transcription.

The apparent lack of RNA pol II promoters for protein-encoding genes and my findings that H4K10ac and four histone variants mark the probable boundaries of RNA pol II transcription units raise a number of interesting questions, including the following: Is RNA pol II transcription regulated differently than RNA pol I or RNA pol III transcription? What other factors do H4K10ac and the four histone variants interact with?

How are they deposited at TSSs and TTSs? Are they inherited in a DNA sequence-independent manner, i.e. are they true 'epigenetic marks'?

While all these questions are worth pursuing, in the next paragraphs I will explain why I find the question especially interesting whether H4K10ac and H2AZ/H2BV could be considered 'epigenetic marks'.

If one defines 'epigenetics' as heritable changes in gene expression caused by mechanisms other than alterations of DNA sequence, an 'epigenetic mark' would have to be stably inherited in a DNA sequence-independent manner. Thus far, most histone modifications or histone variants that are considered epigenetic marks are repressive marks like DNA methylation (Bostick et al., 2007; Sharif et al., 2007; Woo et al., 2007), H3K9m3 (Bannister et al., 2001; Lachner et al., 2001), H3K27me3 (Hansen et al., 2008) or CENP-A (Henikoff et al., 2001) and they are maintained by replication-coupled or replication-independent mechanisms (reviewed in Probst et al., 2009). In contrast, in *T. brucei* H4K10ac, H2AZ, and H2BV are marks correlating with transcription initiation.

A popular model of replication-coupled inheritance of histone modifications proposes that DNA replication is followed by random distribution of parental H3/H4 dimers between the old and new DNA strands. Such a random distribution would initially lead to a 'dilution' of histone modifications that were present on old H3/H4 dimers but not on

newly synthesized H3/H4 dimers. However, reader proteins could subsequently bind to the existing histone modifications and recruit histone-modifying enzymes, which could modify neighboring, newly synthesized histones until pre-replication modification levels are reached (Probst et al., 2009; Morris and Moazed, 2007). Inheritance of histone variants could follow a similar pattern. One obvious problem with such a model is the finite chance of all old H3/H4 dimers assembling onto one strand, in which case the histone mark would be lost on the other strand. The shorter the region containing a particular histone mark, the higher the chance that it will be lost during random distribution. For example, a histone mark spanning three nucleosomes would be lost during every eighth (2^3) mitotic division.

A well-studied example of replication-independent propagation of an epigenetic mark is the maintenance of CENP-A (Shelby et al., 2000). CENP-A is a variant form of histone H3 that marks centromeres and is maintained independently of DNA sequence (see Chapter 1 and Henikoff et al., 2001). Following replication, CENP-A is distributed over both new and old strands, but pre-replication CENP-A levels are only reestablished after replenishment of CENP-A during G1 phase (Jansen et al., 2007).

Much less is known about inheritance of transcription-activating histone modifications or histone variants. Only one group reported the successful maintenance of a transcriptionally active chromatin state through multiple rounds of mitotic cell divisions. In that study, the

authors performed nuclear transfer experiments in *Xenopus* and observed that a particular gene, which was active in the donor cell, remained active in its new cell environment, a process that appeared to be dependent on the histone variant H3.3 (Ng and Gurdon, 2005; Ng and Gurdon, 2008).

Could H4K10ac, H2AZ and H2BV represent epigenetic marks that are responsible for transcriptional activation in *T. brucei*? While my findings do not answer the question whether H4K10ac, H2AZ and H2BV can be maintained in a DNA sequence-independent manner, the following three findings would be 'expected' from an epigenetic mark: no cell cycle or life cycle regulation, localization over large regions, and no correlation between specific DNA sequence motif and incorporation of marks.

IF microscopy indicated that H4K10ac, H2AZ and H2BV co-localized and that all three marks could be detected in all cells. While IF analysis could not exclude the possibility that the marks are absent from some sites for some time, this possibility appears highly unlikely given the remarkably constant levels of DNA immunoprecipitated from the different RNA pol II TSSs. If H4K10ac, H2AZ or H2BV were absent from a RNA pol II TSS for a short period during the cell cycle, the time of absence would have to be the same for all RNA pol II TSSs. Furthermore, Edman data demonstrated that 10% of H4K10 sites are acetylated. Given that there are 207 H4K10ac/H2AZ/H2BV-rich regions, about 8.3 kb in

width, and a genome size of ~35 Mb, all 207 sites should be acetylated throughout the cell cycle.

My data indicate that in *T. brucei* H4K10ac/H2AZ/H2BV-containing nucleosomes mark mostly divergent-SSRs and can be found over large regions with an average width of 8.3 kb (> 40 nucleosomes). Given the number and width of sites marked by H4K10ac/H2AZ/H2BV and assuming random distribution of these marks during replication, a H4K10ac/H2AZ/H2BV-rich region should be lost less than once in every ~6 x 10^{10} cell divisions. A loss of H4K10ac/H2AZ/H2BV from a divergent-SSR would probably be lethal and thus undetectable, but in principle new sites of H4K10ac/H2AZ/H2BV enrichment should form sporadically, as has been described in other organisms for neocentromere formation (Nasuda et al., 2005; Maggert and Karpen, 2001; Amor et al., 2004).

I have performed ChIP-seq for H4K10ac in BF and PF cells and observed one unique peak in each cell line. In order to determine whether H4K10ac should be considered an epigenetic mark, it will be crucial to determine if these unique H4K10ac peaks serve a life cycle-specific function or if they have arisen spontaneously. To do so, one could perform ChIP-seq for H4K10ac in BF, differentiate the cells to PF and repeat the H4K10ac ChIP in the newly differentiated cells. If the previously observed differences in H4K10ac-peak patterns between the two life cycle stages can be reproduced, they probably serve a life cycle-

specific function. If not, they could represent spontaneously arisen peaks.

To determine the general level of conservation among H4K10ac/ H2AZ/H2BV peaks outside of divergent-SSR, one could perform ChIP-seq experiments using different wild-type isolates of *T. brucei* that have been cultured in isolated laboratory conditions for several decades. A high degree of variation would support the hypothesis that these chromatin marks can be maintained in a DNA-sequence-independent manner. Alternatively, one could manipulate the DNA sequence in H4K10ac/H2AZ/H2BV-rich regions outside of divergent-SSR and look for changes in the chromatin structure. Maintenance of the original chromatin structure would suggest that H4K10ac, H2AZ and H2BV are epigenetic marks.

Overexpression of CENP-A in human cells has led to CENP-A incorporation outside of the original centromeric region (Van Hooser et al., 2001; Lam et al., 2006). Similarly, one could overexpress H2AZ, H2BV and HAT2, the enzyme responsible for H4K10 acetylation, and test by ChIP-seq if this leads to new sites of H4K10ac/H2AZ/H2BV enrichment. If new sites of enrichment can be detected, one would have to compare several clones and determine if these new sites formed at random sites or at similar locations. Only a random occurrence of H4K10ac/H2AZ/H2BV-rich sites would suggest a DNA sequence-independent incorporation, while a non-random distribution would

suggest that a certain DNA sequence composition is necessary but not sufficient for H4K10ac/H2AZ/H2BV incorporation.

Even after extensive sequence analysis, which I performed with help from Doeke Hekstra, I was unable to identify a specific RNA pol II promoter motif that could serve as putative transcription factor binding site. Such binding sites would be necessary if acetylation of H4K10 or the incorporation of H2AZ and H2BV were DNA sequence-dependent.

To exclude the possibility that some highly degenerate DNA sequence motif exists, one could duplicate the DNA sequence found at divergent-SSR, clone it into a site within a PTU and perform ChIP-seq experiments with the newly generated cell line. Alternatively, one could clone the sequence into an artificial linear mini chromosome or episome; both techniques have been successfully employed in PF (Patnaik et al., 1993; Patnaik et al., 1996). Should H4K10ac, H2AZ and H2BV incorporation be detected from the newly generated site, one would have clear evidence that DNA sequence plays a role in incorporation of these marks. Thus, they would not be epigenetic marks.

In Chapter 5, I described the presence of poly-guanine tracts at the 5´ ends of probable RNA pol II TSSs. In my opinion, the correlation between runs of guanine and H4K10ac/H2AZ/H2BV-rich regions does not exclude the possibility that H4K10ac, H2AZ and H2BV are epigenetic marks. If one assumes that incorporation of these marks generates an environment permissive to transcription initiation, such transcription

would probably not be unidirectional. Bidirectional transcription initiation outside of a SSR should generate coding RNA transcript as well as non-coding RNA. The latter should lead to RNAi-mediated downregulation of gene expression and would probably be lethal. In Chapter 5, I proposed that the presence of poly-guanine tracts could hinder RNA pol II from proceeding and thereby promote unidirectional transcription. Thus, H4K10ac/H2AZ/H2BV-rich regions may occur spontaneously at random sites, but cells only survive if these sites are flanked by a poly-guanine tract that prevents RNA-mediated downregulation of essential genes. A first step in testing this hypothesis would be to determine if poly-guanine tracts have the ability to hinder RNA pol II transcription elongation by cloning a luciferase reporter gene into a PTU with and without an upstream poly-guanine tract.

Post-transcriptional regulation of gene expression

Since the discovery of polycistronic transcription in *T. brucei*, it has been assumed that regulation of transcription initiation could only contribute minimally to regulation of gene expression. My data reveal a remarkable degree of conservation of chromatin structure among different probable RNA pol II TSSs, suggesting that there may be little difference in regulation even among different PTUs. Given the apparent lack of regulation at the level of transcription initiation, it will be important to determine to what extent processes like *trans* splicing,

polyadenylation, RNA degradation, RNA export, translation efficiency (codon usage), and protein stability contribute to post-transcriptional regulation of gene expression in *T. brucei*.

Thus far, most research has been directed at understanding mechanisms responsible for *trans* splicing and RNA degradation. The efficiency of both processes can vary greatly and appears to be dependent on regulatory DNA motifs located in intergenic regions and UTRs, as my analysis indicated for the case of *trans* splicing (see Chapter 6).

Attempts to identify regulatory DNA motifs that influence RNA stability have focused on relatively small groups (<10 genes) of co-regulated genes because genome-wide information on RNA transcript levels is not available; nor is the length of 5´ UTR and 3´ UTR for most genes. Availability of both genome-wide transcript levels and UTR length information would allow researchers to group genes based on relative RNA transcript levels or based on life cycle-specific regulation and perform a large-scale search for regulatory DNA motifs in the different groups.

By performing high throughput sequencing of cDNA from BF and PF cells, I was able to generate information on transcript levels for more than 7000 genes (out of 8774 total). Using the generated sequence tags, George Cross was able to determine the 5´ UTR lengths for 6,200 genes and the 3´ UTR lengths for 1,500 genes. Whereas about 4,000 genes had

a single splice site, polyadenylation could occur at several positions in a short window. In future studies, I hope to use these data to identify regulatory DNA motifs correlating with relative gene expression levels and with life cycle-specific regulation of genes.

References

Agard, D. A., Hiraoka, Y., and Sedat, J. W. (1988). Three-dimensional light microscopy of diploid *Drosophila* chromosomes. Cell Motil. Cytoskeleton *10*, 18-27.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell *9*, 1191-1200.

Alibu, V. P., Storm, L., Haile, S., Clayton, C., and Horn, D. (2005). A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *139*, 75-82.

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in regulation of RNA synthesis. Proc. Natl. Acad. Sci. USA *51*, 786-794.

Allfrey, V. G., Littau, V. C., and Mirsky, A. E. (1963). On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. Proc. Natl. Acad. Sci. USA *49*, 414-421.

Allfrey, V. G., and Mirsky, A. E. (1962). Evidence for the complete DNA-dependence of RNA synthesis in isolated thymus nuclei. Proc. Natl. Acad. Sci. USA *48*, 1590-1596.

Allis, C. D., Glover, C. V., Bowen, J. K., and Gorovsky, M. A. (1980). Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, *Tetrahymena thermophila*. Cell *20*, 609-617.

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.

Amor, D. J., Bentley, K., Ryan, J., Perry, J., Wong, L., Slater, H., and Choo, K. H. (2004). Human centromere repositioning "in progress". Proc. Natl. Acad. Sci. USA *101*, 6542-6547.

Anderson, K., and Moore, M. J. (1997). Bimolecular exon ligation by the human spliceosome. Science *276*, 1712-1716.

Bangs, J. D., Crain, P. F., Hashizume, T., McCloskey, J. A., and Boothroyd, J. C. (1992). Mass spectrometry of mRNA cap 4 from

trypanosomatids reveals two novel nucleosides. J. Biol. Chem. *267*, 9805-9815.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Barry, J. D., and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. In Advances in Parasitology, eds. (London: Academic Press Ltd), pp. 1-70.

Barry, J. D., McCulloch, R., Mottram, J. C., and Acosta-Serrano, A. (2007). Trypanosomes: After the Genome (Norfolk: Horizon Bioscience).

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and K., Z. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bastin, P., Bagherzadeh, A., Matthews, K. R., and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *77*, 235-239.

Ben Amar, M. F., Jefferies, D., Pays, A., Bakalara, N., Kendall, G., and E, P. (1991). The actin gene promoter of *Trypanosoma brucei*. Nucleic Acids Res. *19*, 5857-5862.

Benne, R., van den Burg, J., Brakenhoff, J. P. J., Sloof, P., vanBoom, J. H., and Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819-826.

Bernstein, E., and Hake, S. B. (2006). The nucleosome: a little variation goes a long way. Biochem. Cell. Biol. *84*, 505-517.

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

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

Bird, A. W., Yu, D. Y., Pray-Grant, M. G., Qiu, Q., Harmon, K. E., Megee, P. C., Grant, P. A., Smith, M. M., and Christman, M. F. (2002). Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. Nature *419*, 411-415.

Blower, M. D., and Karpen, G. H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat. Cell Biol. *3*, 730-739.

Borst, P., van der Ploeg, M., van Hoek, J. F., Tas, J., and James, J. (1982). On the DNA content and ploidy of trypanosomes. Mol. Biochem. Parasitol. *6*, 13-23.

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Boucher, N., Wu, Y., Dumas, C., Dube, M., Sereno, D., Breton, M., and Papadopoulou, B. (2002). A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element. J. Biol. Chem. *277*, 19511-19520.

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311-322.

Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D., and Broach, J. R. (1993). Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes & Dev. *7*, 592-604.

Brogna, S., and Ashburner, M. (1997). The Adh-related gene of *Drosophila melanogaster* is expressed as a functional dicistronic

messenger RNA: multigenic transcription in higher organisms. EMBO J. *16*, 2023-2031.

Brownell, J. E., and Allis, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr. Opin. Genet. Dev. *6*, 176-184.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). *Tetrahymena* histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Bruce, D. (1895). Tsetse Fly Disease Or Nagana (London: Harrison and Sons).

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

Burge, C. B., Tuschl, T., and Sharp, J. A. (1999). Splicing of precursors to mRNA by the splicosome. In The RNA world : the nature of modern RNA suggests a prebiotic RNA, Gesteland, R. F., T. Cech, and J. F. Atkins, eds. (Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press), pp. 525-559.

Cairns, B. R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. Nat. Struct. Mol. Biol. *14*, 989-996.

Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984). Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature *311*, 350-355.

Chen, S., Anderson, K., and Moore, M. J. (2000). Evidence for a linear search in bimolecular 3' splice site AG selection. Proc. Natl. Acad. Sci. USA *97*, 593-598.

Chiara, M. D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. EMBO J. *16*, 4746-4759.

Chicoine, L. G., Schulman, I. G., Richman, R., Cook, R. G., and Allis, C. D. (1986). Nonrandom utilization of acetylation sites in histones isolated from *Tetrahymena*. Evidence for functionally distinct H4 acetylation sites. J. Biol. Chem. *261*, 1071-1076.

Chowdhury, A. R., Zhao, Z., and Englund, P. T. (2008). Effect of hydroxyurea on procyclic *Trypanosoma brucei:* an unconventional

mechanism for achieving synchronous growth. Eukaryot. Cell *7*, 425-428.

Chua, K., and Reed, R. (2001). An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol. Cell. Biol. *21*, 1509-1514.

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

Cross, G. A. M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology *71*, 393-417.

Cross, G. A. M. (1996). Antigenic variation in trypanosomes: secrets surface slowly. Bioessays *18*, 283-291.

Csank, C., Taylor, F. M., and Martindale, D. W. (1990). Nuclear pre-mRNA introns: analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. Nucleic Acids Res. *18*, 5133-5141.

da Cunha, J. P., Nakayasu, E. S., de Almeida, I. C., and Schenkman, S. (2006). Post-translational modifications of *Trypanosoma cruzi* histone H4. Mol. Biochem. Parasitol. *150*, 268-277.

Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A. M., and Bellofatto, V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol. Cell. Biol. *25*, 7314-7322.

Daury, L., Chailleux, C., Bonvallet, J., and Trouche, D. (2006). Histone H3.3 deposition at E2F-regulated genes is linked to transcription. EMBO Rep. *7*, 66-71.

Dawe, R. K., and Henikoff, S. (2006). Centromeres put epigenetics in the driver's seat. Trends. Biochem. Sci. *31*, 662-669.

de Lange, T., and Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature *299*, 451-453.

De Lange, T., Michels, P. A., Veerman, H. J., Cornelissen, A. W., and Borst, P. (1984). Many trypanosome messenger RNAs share a common 5' terminal sequence. Nucleic Acids Res *12*, 3777-3790.

Deshler, J. O., and Rossi, J. J. (1991). Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. Genes & Dev. *5*, 1252-1263.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Di Noia, J. M., D'Orso, I., Sanchez, D. O., and Frasch, A. C. (2000). AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. J. Biol. Chem. *275*, 10218-10227.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873-885.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005a). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science *309*, 409-415.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram,

J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., and Hall, N. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. Science *309*, 404-449.

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

Ersfeld, K., Docherty, R., Alsford, S., and Gull, K. (1996). A fluorescence in situ hybridisation study of the regulation of histone mRNA levels during the cell cycle of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *81*, 201-209.

Etzioni, S., Yafe, A., Khateb, S., Weisman-Shomer, P., Bengal, E., and Fry, M. (2005). Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. J. Biol. Chem. *280*, 26805-26812.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Ferguson, M. A., Homans, S. W., Dwek, R. A., and Rademacher, T. W. (1988). Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Science *239*, 753-759.

Fernandes, A. P., Nelson, K., and Beverley, S. M. (1993). Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. Proc. Natl. Acad. Sci. USA *90*, 11608-11612.

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

Forsythe, G. R., McCulloch, R., and Hammarton, T. C. (2009). Hydroxyurea-induced synchronisation of bloodstream stage *Trypanosoma brucei*. Mol. Biochem. Parasitol. *164*, 131-136.

Freistadt, M. S., Cross, G. A. M., Branch, A. D., and Robertson, H. D. (1987). Direct analysis of the mini-exon donor RNA of *Trypanosoma brucei:* detection of a novel cap structure also present in messenger RNA. Nucleic Acids Res. *15*, 9861-9879.

Freistadt, M. S., Cross, G. A. M., and Robertson, H. D. (1988). Discontinuously synthesized mRNA from *Trypanosoma brucei*

contains the highly methylated 5' cap structure, m7GpppA*A*C(2'-O)mU*A. J. Biol. Chem. *263*, 15071-15015.

Furger, A., Schurch, N., Kurath, U., and Roditi, I. (1997). Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. Mol. Cell. Biol. *17*, 4372-4380.

Gao, L., Cueto, M. A., Asselbergs, F., and Atadja, P. (2002). Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. J. Biol. Chem. *277*, 25748-25755.

Garcia-Salcedo, J. A., Gijon, P., Nolan, D. P., Tebabi, P., and Pays, E. (2003). A chromosomal SIR2 homologue with both histone NAD-dependent ADP-ribosyltransferase and deacetylase activities is involved in DNA repair in *Trypanosoma brucei*. EMBO J. *22*, 5851-5862.

Garcia-Salcedo, J. A., Gijon, P., and Pays, E. (1999). Regulated transcription of the histone H2B genes of *Trypanosoma brucei*. Eur. J. Biochem. *264*, 717-723.

Georgakopoulos, T., and Thireos, G. (1992). Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO J. *11*, 4145-4152.

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

Glover, L., McCulloch, R., and Horn, D. (2008). Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic Acids Res. *36*, 2608-2618.

Goldknopf, I. L., French, M. F., Musso, R., and Busch, H. (1977). Presence of protein A24 in rat liver nucleosomes. Proc. Natl. Acad. Sci. USA *74*, 5492-5495.

Gopal, S., Cross, G. A. M., and Gaasterland, T. (2003). An organism-specific method to rank predicted coding regions in *Trypanosoma brucei*. Nucleic Acids Res. *31*, 5877-5885.

Gorovsky, M. A. (1973). Macro- and micronuclei of *Tetrahymena pyriformis*: a model system for studying the structure and function of eukaryotic nuclei. J. Protozool. *20*, 19-25.

Gottschling, D. E., Aparicio, O. M., Billington, B. L., and Zakian, V. A. (1990). Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. Cell *63*, 751-762.

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. *313*, 903-919.

Graveley, B. R., Hertel, K. J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA *7*, 806-818.

Guillemette, B., Bataille, A. R., Gevry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005). Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol. *3*, e384.

Guldner, H. H., Lakomek, H. J., and Bautz, F. A. (1984). Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. Clin. Exp. Immunol. *58*, 13-20.

Haanstra, J. R., Stewart, M., Luu, V. D., van Tuijl, A., Westerhoff, H. V., Clayton, C., and Bakker, B. M. (2008). Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. J. Biol. Chem. *283*, 2495-2507.

Hammarton, T. C. (2007). Cell cycle regulation in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *153*, 1-8.

Hanas, J., Linden, G., and Stuart, K. (1975). Mitochondrial and cytoplasmic ribosomes and their activity in blood and culture form *Trypanosoma brucei*. J. Cell. Biol. *65*, 103-111.

Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291-1300.

Harlow, E., and Lane, D. (1999). Using Antibodies : A Laboratory Manual (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Hassan, A. H., Awad, S., Al-Natour, Z., Othman, S., Mustafa, F., and Rizvi, T. A. (2007). Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. Biochem. J. *402*, 125-133.

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., and Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hausler, T., and Clayton, C. (1996). Post-transcriptional control of hsp70 mRNA in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *76*, 57-71.

Hebbes, T. R., Clayton, A. L., Thorne, A. W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J. *13*, 1823-1830.

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

Heitz, E. (1928). Das Heterochromatin der Moose. Jahrb. Wiss. Botanik *69*, 762-818.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat. Rev. Genet. *9*, 15-26.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. Annu. Rev. Cell. Dev. Biol. *21*, 133-153.

Henikoff, S., Ahmad, K., and Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science *293*, 1098-1102.

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Taylor, J. E., Walker, D., White, B., Young, R., Cross, G. A. M., Rudenko, G., Barry, J. D., Louis, E. J., and Berriman, M. (2008). Telomeric expression sites are highly conserved in *Trypanosoma brucei*. PLoS ONE *3*, e3527.

Hiraoka, Y., Sedat, J. W., and Agard, D. A. (1990). Determination of three-dimensional imaging properties of a light microscope system. Partial confocal behavior in epifluorescence microscopy. Biophys. J. *57*, 325-333.

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.

Honda, B. M., Dixon, G. H., and Candido, E. P. (1975). Sites of in vivo histone methylation in developing trout testis. J Biol Chem *250*, 8681-8685.

Horn, D. (2008). Histone deacetylases. Adv. Exp. Med. Biol. *625*, 81-86.

Horn, D., and Cross, G. A. M. (1995). A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei*. Cell *83*, 555-561.

Horn, D., and Cross, G. A. M. (1997). Position-dependent and promoter-specific regulation of gene expression in *Trypanosoma brucei*. EMBO J. *16*, 7422-7431.

Horn, P. J., and Peterson, C. L. (2002). Chromatin Higher Order Folding: Wrapping up Transcription. Science *297*, 1824-1827.

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

Hotz, H. R., Lorenz, P., Fischer, R., Krieger, S., and Clayton, C. (1995). Role of 3'-untranslated regions in the regulation of hexose transporter mRNAs in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *75*, 1-14.

Huang, J., and Van der Ploeg, L. H. (1991). Requirement of a polypyrimidine tract for *trans*-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. EMBO J. *10*, 3877-3885.

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.

Hull, M. W., Erickson, J., Johnston, M., and Engelke, D. R. (1994). tRNA genes as transcriptional repressor elements. Mol. Cell. Biol. *14*, 1266-1277.

Huppert, J. L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. *35*, 406-413.

Illumina (2008). Instruction Manual: preparation of gDNA for high throughput sequencing.

Imai, S., Armstrong, C. M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature *403*, 795-800.

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

Ismaili, N., Perez-Morga, D., Walsh, P., Cadogan, M., Pays, A., Tebabi, P., and Pays, E. (2000). Characterization of a *Trypanosoma brucei* SR domain-containing protein bearing homology to *cis*-spliceosomal U1 70 kDa proteins. Mol. Biochem. Parasitol. *106*, 109-120.

Ismaili, N., Perez-Morga, D., Walsh, P., Mayeda, A., Pays, A., Tebabi, P., Krainer, A. R., and Pays, E. (1999). Characterization of a SR protein from *Trypanosoma brucei* with homology to RNA-binding *cis*-splicing proteins. Mol. Biochem. Parasitol. *102*, 103-115.

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

Jackson, V., Shires, A., Tanphaichitr, N., and Chalkley, R. (1976). Modifications to histones immediately after synthesis. J. Mol. Biol. *104*, 471-483.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jansen, L. E., Black, B. E., Foltz, D. R., and Cleveland, D. W. (2007). Propagation of centromeric chromatin requires exit from mitosis. J. Cell Biol. *176*, 795-805.

Janzen, C. J., Fernandez, J. P., Deng, H., Diaz, R., Hake, S. B., and Cross, G. A. M. (2006a). Unusual histone modifications in *Trypanosoma brucei*. FEBS Lett. *580*, 2306-2310.

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

Jin, C., and Felsenfeld, G. (2006). Distribution of histone H3.3 in hematopoietic cell lineages. Proc. Natl. Acad. Sci. USA *103*, 574-579.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Dev. *21*, 1519-1529.

Johnson, C. N., Adkins, N. L., and Georgel, P. (2005). Chromatin remodeling complexes: ATP-dependent machines in action. Biochem. Cell. Biol. *83*, 405-417.

Johnson, P. J., Kooter, J. M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273-281.

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., Leproust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., and Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kapotas, N., and Bellofatto, V. (1993). Differential response to RNA *trans*-splicing signals within the phosphoglycerate kinase gene cluster in *Trypanosoma brucei*. Nucleic Acids Res. *21*, 4067-4072.

Kaufman, P. D., Kobayashi, R., Kessler, N., and Stillman, B. (1995). The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. Cell *81*, 1105-1114.

Kawahara, T., Siegel, T. N., Ingram, A. K., Alsford, S., Cross, G. A. M., and Horn, D. (2008). Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Mol. Microbiol. *69*, 1054-1068.

Kayne, P. S., Kim, U. J., Han, M., Mullen, J. R., Yoshizaki, F., and Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell *55*, 27-39.

Kendall, A., Hull, M. W., Bertrand, E., Good, P. D., Singer, R. H., and Engelke, D. R. (2000). A CBF5 mutation that disrupts nucleolar localization of early tRNA biosynthesis in yeast also suppresses tRNA gene-mediated transcriptional silencing. Proc. Natl. Acad. Sci. USA *97*, 13108-13113.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. Genome Res. *12*, 656-664.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595-605.

Korber, P., and Horz, W. (2004). SWRred not shaken; mixing the histones. Cell *117*, 5-7.

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science *184*, 868-871.

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

Kruhlak, M. J., Hendzel, M. J., Fischle, W., Bertos, N. R., Hameed, S., Yang, X. J., Verdin, E., and Bazett-Jones, D. P. (2001). Regulation of global acetylation in mitosis through loss of histone acetyltransferases and deacetylases from chromatin. J. Biol. Chem. *276*, 38307-38319.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lam, A. L., Boivin, C. D., Bonney, C. F., Rudd, M. K., and Sullivan, B. A. (2006). Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. Proc. Natl. Acad. Sci. USA *103*, 4186-4191.

Le Blancq, S. M., Swinkels, B. W., Gibson, W. C., and Borst, P. (1988). Evidence for gene conversion between the phosphoglycerate kinase genes of *Trypanosoma brucei*. J. Mol. Biol. *200*, 439-447.

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

Lee, M. G. (1996). An RNA polymerase II promoter in the hsp70 locus of *Trypanosoma brucei*. Mol. Cell. Biol. *16*, 1220-1230.

Lee, S. J. (1991). Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. Proc. Natl. Acad. Sci. USA *88*, 4250-4254.

Lee, T. I., Johnstone, S. E., and Young, R. A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protocols *1*, 729-748.

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

Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E., and Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. Trends. Parasitol. *21*, 363-369.

Lopez-Estrano, C., Tschudi, C., and Ullu, E. (1998). Exonic sequences in the 5' untranslated region of alpha-tubulin mRNA modulate *trans* splicing in *Trypanosoma brucei*. Mol. Cell. Biol. *18*, 4620-4628.

Lowell, J. E., and Cross, G. A. M. (2004). A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. J. Cell Sci. *117*, 5937-5947.

Lowell, J. E., Kaiser, F., Janzen, C. J., and Cross, G. A. M. (2005). Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*. J. Cell Sci. *118*, 5721-5730.

Lucke, S., Jurchott, K., Hung, L. H., and Bindereif, A. (2005). mRNA splicing in *Trypanosoma brucei*: branch-point mapping reveals

differences from the canonical U2 snRNA-mediated recognition. Mol. Biochem. Parasitol. *142*, 248-251.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Ma, X. J., Wu, J., Altheim, B. A., Schultz, M. C., and Grunstein, M. (1998). Deposition-related sites K5/K12 in histone H4 are not required for nucleosome deposition in yeast. Proc. Natl. Acad. Sci. USA *95*, 6693-6698.

Maggert, K. A., and Karpen, G. H. (2001). The activation of a neocentromere in *Drosophila* requires proximity to an endogenous centromere. Genetics *158*, 1615-1628.

Mandava, V. Histone Modifications in *Trypanosoma brucei*. New York: The Rockefeller University; 2007. 200 p. Dissertation.

Mandava, V., Janzen, C. J., and Cross, G. A. M. (2008). Trypanosome H2Bv replaces H2B in nucleosomes enriched for H3 K4 and K76 trimethylation. Biochem. Biophys. Res. Commun. *368*, 846-851.

Mandava, V., Fernandez, J. P., Deng, H., Janzen, C. J., Hake, S. B., and Cross, G. A. M. (2007). Histone modifications in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *156*, 41-50.

Manger, I. D., and Boothroyd, J. C. (1998). Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from *cis*-splicing systems. Mol. Biochem. Parasitol. *97*, 1-11.

Mannironi, C., Bonner, W. M., and Hatch, C. L. (1989). H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. Nucleic Acids Res. *17*, 9113-9126.

Marchetti, M. A., Tschudi, C., Silva, E., and Ullu, E. (1998). Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersion of RNA polymerase II- and III-transcribed genes. Nucleic Acids Res. *26*, 3591-3598.

Margottin, F., Dujardin, G., Gerard, M., Egly, J. M., Huet, J., and Sentenac, A. (1991). Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science *251*, 424-426.

Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P. J. (2003). Transcription of *Leishmania* major Friedlin chromosome 1 initiates in both directions within a single region. Mol. Cell. *11*, 1291-1299.

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

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

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

McAndrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. Exp. Parasitol. *90*, 65-76.

McCulloch, R., and Barry, J. D. (1999). A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. Genes & Dev. *13*, 2875-2888.

McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., and Felsenfeld, G. (1981). A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion. Cell *27*, 45-55.

Megee, P. C., Morgan, B. A., Mittman, B. A., and Smith, M. M. (1990). Genetic analysis of histone H4: essential role of lysines subject to reversible acetylation. Science *247*, 841-845.

Megee, P. C., Morgan, B. A., and Smith, M. M. (1995). Histone H4 and the maintenance of genome integrity. Genes & Dev. *9*, 1716-1727.

Meneghini, M. D., Wu, M., and Madhani, H. D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725-736.

Metzenberg, S., and Agabian, N. (1996). Human and fungal 3' splice sites are used by *Trypanosoma brucei* for *trans* splicing. Mol. Biochem. Parasitol. *83*, 11-23.

Milhausen, M., Nelson, R. G., Sather, S., Selkirk, M., and Agabian, N. (1984). Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? Cell *38*, 721-729.

Mito, Y., Henikoff, J. G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090-1097.

Morris, C. A., and Moazed, D. (2007). Centromere assembly and propagation. Cell *128*, 647-650.

Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., and Dimitrov, S. (1998). Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. Mol. Cell. Biol. *18*, 6293-6304.

Nakaar, V., Dare, A. O., Hong, D., Ullu, E., and Tschudi, C. (1994). Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. Mol. Cell. Biol. *14*, 6736-6742.

Nakaar, V., Günzl, A., Ullu, E., and Tschudi, C. (1997). Structure of the *Trypanosoma brucei* U6 snRNA gene promoter. Mol. Biochem. Parasitol. *88*, 13-23.

Nasuda, S., Hudakova, S., Schubert, I., Houben, A., and Endo, T. R. (2005). Stable barley chromosomes without centromeric repeats. Proc. Natl. Acad. Sci. USA *102*, 9842-9847.

Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., and Senda, T. (2007). Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. Nature

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. EMBO J. *18*, 2265-2272.

Ng, R. K., and Gurdon, J. B. (2005). Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. Proc. Natl. Acad. Sci. USA *102*, 1957-1962.

Ng, R. K., and Gurdon, J. B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat. Cell Biol. *10*, 102-109.

Noma, K., Cam, H. P., Maraia, R. J., and Grewal, S. I. (2006). A role for TFIIIC transcription factor complex in genome organization. Cell *125*, 859-872.

North, A. J. (2006). Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. J. Cell Biol. *172*, 9-18.

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.

Ogbadoyi, E., Ersfeld, K., Robinson, D., Sherwin, T., and Gull, K. (2000). Architecture of the *Trypanosoma brucei* nucleus during interphase and mitosis. Chromosoma *108*, 501-513.

Olins, A. L., and Olins, D. E. (1974). Spheroid Chromatin Units (ν Bodies). Science *183*, 330-332.

Ono, T., and Nakabayashi, T. (1980). Studies on the effect of bleomycin on *Trypanosoma gambiense*. Biken J. *23*, 143-155.

Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H., and Margolis, R. L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. USA *88*, 3734-3738.

Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S., and Margolis, R. L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J. Cell Biol. *104*, 805-815.

Paro, R., and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. Proc. Natl. Acad. Sci. USA *88*, 263-267.

Parsons, M., Nelson, R. G., Watkins, K. P., and Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. Cell *38*, 309-316.

Parthun, M. R., Widom, J., and Gottschling, D. E. (1996). The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell *87*, 85-94.

Patnaik, P. K., Axelrod, N., Van der Ploeg, L. H., and Cross, G. A. M. (1996). Artificial linear mini-chromosomes for *Trypanosoma brucei*. Nucleic Acids Res. *24*, 668-675.

Patnaik, P. K., Kulkarni, S. K., and Cross, G. A. M. (1993). Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. EMBO J. *12*, 2529-2538.

Patterson, B., and Guthrie, C. (1991). A U-rich tract enhances usage of an alternative 3' splice site in yeast. Cell *64*, 181-187.

Patzelt, E., Perry, K. L., and Agabian, N. (1989). Mapping of branch sites in *trans*-spliced pre-mRNAs of *Trypanosoma brucei*. Mol. Cell. Biol. *9*, 4291-4297.

Pehrson, J. R., and Fried, V. A. (1992). MacroH2A, a core histone containing a large nonhistone region. Science *257*, 1398-1400.

Perry, K. L., Watkins, K. P., and Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. USA *84*, 8190-8194.

Pillus, L. (2008). MYSTs mark chromatin for chromosomal functions. Curr. Opin. Cell. Biol. *20*, 326-333.

Ploubidou, A., Robinson, D. R., Docherty, R. C., Ogbadoyi, E. O., and Gull, K. (1999). Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. J. Cell Sci. *112*, 4641-4650.

Priest, J. W., and Hajduk, S. L. (1994). Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. J. Bioenerg. Biomembr. *26*, 179-191.

Probst, A. V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. *10*, 192-206.

Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., Rando, O. J., and Madhani, H. D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233-248.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D. J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J. *22*, 1599-1607.

Redmond, S., Vadivelu, J., and Field, M. C. (2003). RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *128*, 115-118.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Curr. Opin. Genet. Dev. *12*, 162-169.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr. Opin. Cell Biol. *12*, 340-345.

Respuela, P., Ferella, M., Rada-Iglesias, A., and Aslund, L. (2008). Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi*. J. Biol. Chem. *283*, 15884-15892.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858-5868.

Ross, R., and Thomson, D. (1910). A Case of Sleeping Sickness Studied by Precise Enumerative Methods: Regular Periodical Increase of the Parasites Disclosed. Proceedings of the Royal Society of London. Series B *82*, 411-415.

Rout, M. P., and Field, M. C. (2001). Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. identification of a major repetitive nuclear lamina component. J. Biol. Chem. *276*, 38261-38271.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. *132*, 365-386.

Ruskin, B., Krainer, A. R., Maniatis, T., and Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell *38*, 317-331.

Ruthenburg, A. J., Allis, C. D., and Wysocka, J. (2007a). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell *25*, 15-30.

Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007b). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983-994.

Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., Weaver, M., Shafer, A., Lee, K., Neri, F., Humbert, R., Singer, M. A., Richmond, T. A., Dorschner, M. O., McArthur, M., Hawrylycz, M., Green, R. D., Navas, P. A., Noble, W. S., and Stamatoyannopoulos,

J. A. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat. Methods *3*, 511-518.

Satchwell, S. C., Drew, H. R., and Travers, A. A. (1986). Sequence periodicities in chicken nucleosome core DNA. J. Mol. Biol. *191*, 659-675.

Scahill, M. D., Pastar, I., and Cross, G. A. M. (2008). CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *157*, 73-82.

Schaal, T. D., and Maniatis, T. (1999a). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol. Cell. Biol. *19*, 261-273.

Schaal, T. D., and Maniatis, T. (1999b). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol. Cell. Biol. *19*, 1705-1719.

Schiltz, R. L., Mizzen, C. A., Vassilev, A., Cook, R. G., Allis, C. D., and Nakatani, Y. (1999). Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. J. Biol. Chem. *274*, 1189-1192.

Schimanski, B., Nguyen, T. N., and Günzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. Mol. Cell. Biol. *25*, 7303-7313.

Schwartz, Y. B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. Curr. Opin. Cell. Biol. *20*, 266-273.

Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., and Clayton, C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Res. *36*, 3374-3388.

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Shalgi, R., Lapidot, M., Shamir, R., and Pilpel, Y. (2005). A catalog of stability-associated sequence elements in 3' UTRs of yeast mRNAs. Genome Biol. *6*, R86.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.,

Tajima, S., Mitsuya, K., Okano, M., and Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shelby, R. D., Monier, K., and Sullivan, K. F. (2000). Chromatin assembly at kinetochores is uncoupled from DNA replication. J. Cell Biol. *151*, 1113-1118.

Shi, B., Guo, X., Wu, T., Sheng, S., Wang, J., Skogerbo, G., Zhu, X., and Chen, R. (2009). Genome-scale identification of Caenorhabditis elegans regulatory elements by tiling-array mapping of DNase I hypersensitive sites. BMC Genomics *10*, 92.

Shia, W. J., Li, B., and Workman, J. L. (2006). SAS-mediated acetylation of histone H4 Lys 16 is required for H2A.Z incorporation at subtelomeric regions in *Saccharomyces cerevisiae*. Genes & Dev. *20*, 2507-2512.

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr. Opin. Cell. Biol. *20*, 341-348.

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Siegel, T. N., Kawahara, T., Degrasse, J. A., Janzen, C. J., Horn, D., and Cross, G. A. M. (2008). Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in *Trypanosoma brucei*. Mol. Microbiol. *67*, 762-771.

Smith, C. W., Chu, T. T., and Nadal-Ginard, B. (1993). Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol. Cell. Biol. *13*, 4939-4952.

Sobel, R. E., Cook, R. G., and Allis, C. D. (1994). Non-random acetylation of histone H4 by a cytoplasmic histone acetyltransferase as determined by novel methodology. J. Biol. Chem. *269*, 18576-18582.

Sobel, R. E., Cook, R. G., Perry, C. A., Annunziato, A. T., and Allis, C. D. (1995). Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. Proc. Natl. Acad. Sci. USA *92*, 1237-1241.

Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993). Operons in *C. elegans:* polycistronic mRNA precursors are processed by *trans*-splicing of SL2 to downstream coding regions. Cell *73*, 521-532.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sun, Z. W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Taddei, A., Roche, D., Sibarita, J. B., Turner, B. M., and Almouzni, G. (1999). Duplication and maintenance of heterochromatin domains. J. Cell. Biol. The Journal of cell biology *147*, 1153-1166.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992). brahma: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell *68*, 561-572.

Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Tornaletti, S., Park-Snyder, S., and Hanawalt, P. C. (2008). G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. J. Biol. Chem. *283*, 12756-12762.

Tschudi, C., and Ullu, E. (1988). Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. EMBO J. *7*, 455-463.

Tse, C., Sera, T., Wolffe, A. P., and Hansen, J. C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol. Cell. Biol. *18*, 4629-4638.

Tyler, J. K., Adams, C. R., Chen, S. R., Kobayashi, R., Kamakaka, R. T., and Kadonaga, J. T. (1999). The RCAF complex mediates chromatin assembly during DNA replication and repair. Nature *402*, 555-560.

Van der Ploeg, L. H., Valerio, D., De Lange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982). An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant

surface glycoproteins are clustered in the genome. Nucleic Acids Res. *10*, 5905-5923.

Van Hooser, A. A., Ouspenski, I. I., Gregson, H. C., Starr, D. A., Yen, T. J., Goldberg, M. L., Yokomori, K., Earnshaw, W. C., Sullivan, K. F., and Brinkley, B. R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J. Cell Sci. *114*, 3529-3542.

Vazquez, M., Atorrasagasti, C., Bercovich, N., Volcovich, R., and Levin, M. J. (2003). Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor. Mol. Biochem. Parasitol. *128*, 77-81.

Verreault, A., Kaufman, P. D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell *87*, 95-104.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. J. Cell Sci. *5*, 163-193.

Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Waterborg, J. H. (2002). Dynamics of histone acetylation in vivo. A function for acetylation turnover? Biochem. Cell Biol. *80*, 363-378.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848-856.

Wentz-Hunter, K., and Potashkin, J. (1996). The small subunit of the splicing factor U2AF is conserved in fission yeast. Nucleic Acids Res. *24*, 1849-1854.

WHO (2004). The World Health Report 2004. Changing History.

Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. Q. Rev. Biophys. *34*, 269-324.

Wilusz, C. J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet. *20*, 491-497.

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

Winston, F., and Allis, C. D. (1999). The bromodomain: a chromatin-targeting module? Nat. Struct. Biol. *6*, 601-604.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G. A. M. (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.

Wirtz, E., Hoek, M., and Cross, G. A. M. (1998). Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. Nucleic Acids Res. *26*, 4626-4634.

Woo, H. R., Pontes, O., Pikaard, C. S., and Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes & Dev. *21*, 267-277.

Woodward, R., and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. J. Cell Sci. *95*, 49-57.

Wu, J. I., Lessard, J., and Crabtree, G. R. (2009). Understanding the words of chromatin regulation. Cell *136*, 200-206.

Wu, R. S., and Bonner, W. M. (1981). Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. Cell *27*, 321-330.

Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E. S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature *434*, 338-345.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J. E. J. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res. *13*, 1863-1872.

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.

Yao, M. C., and Gorovsky, M. A. (1974). Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. Chromosoma *48*, 1-18.

Zamore, P. D., and Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc. Natl. Acad. Sci. USA *86*, 9243-9247.

Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhang, H., Roberts, D. N., and Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/ basal promoters for activation through histone loss. Cell *123*, 219-231.

Zhuang, Y., and Weiner, A. M. (1990). The conserved dinucleotide AG of the 3' splice site may be recognized twice during *in vitro* splicing of mammalian mRNA precursors. Gene *90*, 263-269.

Zubiaga, A. M., Belasco, J. G., and Greenberg, M. E. (1995). The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. Mol. Cell. Biol. *15*, 2219-2230.