The Genome-wide nucleosome positions in *Trypanosoma brucei* procyclic and Bloodstream forms

by

Johannes P. Maree

Submitted in accordance with the requirements for the degree

Magister Scientiae

In the

Faculty of Natural and Agricultural Sciences,

Department of Microbial, Biochemical and Food Biotechnology

University of the Free State

Bloemfontein

South Africa

Supervisor: Prof Hugh G. Patterton

January 2014

Acknowledgements

I wish to thank the following:

God, for giving me strength and wisdom.

Prof H. G. Patterton, for his help, patience, guidance, advice and encouragement.

Drs. Megan Povelones (PSU) and Kathrin Witmer (ICL), for assistance, mentoring and discussions.

Dr. Gloria Rudenko (ICL), for accommodating me in her laboratory, guidance and discussions.

Dr. David Clark (NIH), for advice on core particle preparation and sequencing.

Friends and family, for support and encouragement.

National Research Foundation, **RSA**, for financial support.

National Institutes of Health, **USA**, for financial support.

Table of Contents:

Chater 1: Literature review

1.1) Abstract **9**

1.2) Introduction **10**

1.3) Materials and Methods **11**

1.4) Genome organisation **14**

1.5) Transcription in T. brucei **17**

1.5.1) RNA dependent polymerases **17**

1.5.2) Long non-coding RNA **21**

1.6) Telomeric silencing and bloodstream expression sites **21**

1.7) Base J **25**

1.8) Replication origin complex and gene silencing **26**

1.9) Nucleosomal organization **27**

1.10) Histone epigenetic patterns **28**

1.10.1) H1 **28**

1.10.2) H2A **28**

1.10.3) H2B **29**

1.10.4) H3 **29**

1.10.5) H4 **30**

1.11) Histone variants **34**

1.12) Conclusions **36**

Chapter 2: Nucleosomal positioning in procyclic and bloodstream form *Trypanosoma brucei*.

2.1) Abstract **38**

2.2) Introduction **39**

2.3) Materials and Methods

In vitro

2.3.1) Trypanosome strains and culture **48**

2.3.2) Core particle preparation. **50**

2.3.3) Mononucleosomal DNA isolation **51**

2.3.4) Histone H1 analysis **51**

2.3.5) Nucleosome repeat length determination **52**

2.3.6) Paired-end sequencing **54**

In silico

2.3.7) Alignment using Bowtie **57**

2.3.8) Visual inspection of data **58**

2.3.9) Data normalization and Bin analysis **59**

2.3.10) Average dyad distribution analysis **61**

2.3.11) Sequence motif analysis **62**

Base frequency and AT skew analysis **62**

Dinucleotide distribution **63**

Fast Fourier transform **64**

2.4) Results

In vitro

2.4.1) Core particle preparation **65**

2.4.2) Nucleosome repeat lengths **67**

2.4.3) Histone H1 knock-down **68**

In silico

2.4.4) Sequence alignment **69**

2.4.5) Visual inspection **72**

2.4.6) Data normalization and Bin analysis **75**

Bin analysis **76**

Bin analysis of Tandem repeats **78**

Pol II transcribed genes **80**

Pol I transcribed genes **82**

Pol III transcribed genes **83**

Subtelomeric regions **84**

2.4.7) Nucleosomal architecture **87**

Pol II **87**

Pol I **95**

Pol III **97**

Effect of H1 knock-down **99**

2.4.8) Intrinsic sequence motifs of pol II transcribed regions **105**

Base frequency upstream of PTU **105**

Sequence motifs **105**

AT skew **107**

2.4.9) Dinucleotide distribution **116**

2.5) Discussion and Conclusions **121**

2.6) References **127**

2.7) Supplementary material **138**

Summary **151**

Keywords **152**

Opsomming **153**

Sleutelwoorde **154**

List of abbreviations used

BES	Bloodstream Expression Site
BF	Bloodstream Form
CDS	Coding Sequence
ES	Expression Site
ESAG	Expression Site Associated Gene
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation sequencing
FFT	Fast Fourier Transform
IC	Intermediate Chromosome
KD	Knock-down
LncRNA	long non-coding RNA
LRRP	Leucine-rich Repeat Protein
MBC	Megabase Chromosome
MC	Minichromosome
MNase	Micrococcal Nuclease
MVSG	Metacyclic Variable Surface Glycoprotein
MS	Mass Spectrometry
NDR	Nucleosome Depleted region
NRL	Nucleosome Repeat Length
ORF	Open Reading Frame
PAS	Poly-Adenylation Site
PF	Procyclic Form
PIC	Pre-initiation Complex
PTM	Post Translational Modification
PTU	Polycistronic Transcription Unit
RHS	Retrotransposon Hotspot
RNAi	RNA interference
SAS	Splice Acceptor Site
SSR	Strand Switching Regions
TBP	TATA-binding Protein
TF	Transcription Factor
TR	Tandem Repeat
TSS	Transcription Start Site
TTS	Transcription Termination Site
VSG	Variable Surface Glycoprotein
UTR	Untranslated Region

Chapter 1

Literature Review

1.1) Abstract

The epigenome represents a major regulatory interface to the eukaryotic genome. Nucleosome positions, histone variants, histone modifications and chromatin associated proteins all play a role in the epigenetic regulation of DNA function.

Trypanosomes, an ancient branch of the eukaryotic evolutionary lineage, exhibit some highly unusual transcriptional features, including the arrangement of functionally unrelated genes in large, polymerase II transcribed polycistronic transcription units, often exceeding hundreds of kb in size. It is generally believed that transcription initiation plays a minor role in regulating the transcript level of genes in trypanosomes, which are mainly regulated post-transcriptionally.

Recent advances have revealed that epigenetic mechanisms play an essential role in the transcriptional regulation of *Trypanosoma brucei*. This suggested that the regulation of gene activity is, indeed, an important control mechanism, and that the epigenome is critical in regulating gene expression programs that allow the successful migration of this parasite between hosts, as well as the continuous evasion of the immune system in mammalian hosts. A wide range of epigenetic signals, readers, writers and erasers have been identified in trypanosomes, some of which have been observed to be unique to trypanosomes. We review recent advances in our understanding of epigenetic control mechanisms in *T. brucei*, the causative agent of African sleeping sickness, and discuss the possible role that these mechanisms may play in the life cycle of the parasite.

1.2) Introduction

Trypanosoma brucei, the causative agent of African sleeping sickness, is an extracellular, flagellated parasite that is transferred into the human host during a blood meal by a *Glossina spp.* fly (Figure 1). In its initial haemolymphatic phase, bloodstream form (BF) *T. brucei* invades the bloodstream, interstitial spaces, and lymph system where it divides asexually [1]. With prolonged infection, the parasite crosses the blood brain barrier and enters an encephalitic stage, where the patient exhibits the typical clinical signs of the disease from which the name is derived. Without treatment, African sleeping sickness is lethal.

If the *Glossina spp.* fly feeds on an infected host, trypanosomes may be taken up, and will transform to a procyclic form (PF) trypomastigote in the insect midgut, where the parasites again multiply by asexual cell division. From the midgut, the parasite moves to the salivary gland of the fly, transforming to a metacyclic epimastigote, capable of infecting a new mammalian host. The migration of the parasite from a mammalian to an insect host is accompanied by the activation and shutdown of several transcription programs [2]. Many of these programs appear to be regulated by epigenetic mechanisms, implicating chromatin in *T. brucei* gene regulation.

The early divergence of Kinetoplasts from the main eukaryotic lineage contributes to some of the unusual features seen in the *T. brucei* genome organisation and replicative and transcriptional processes. Kinetoplastids, one of the earliest mitochondrial containing eukaryotes, harbour the mitochondrial genome in a body called the kinetoplast, separating mitochondrial and nuclear DNA [3]. Mitochondrial DNA amounts to ~ 30% of the total DNA compliment [4]. A striking feature of the parasite's nuclear genome is the size variability of

sister chromatids and the nuclear DNA complement, as well as gene organisation and expression. *T.brucei* seems to have a more relaxed approach when it comes to transcription and replication. This could possibly be because of "generalized" demands of a parasitic life style.

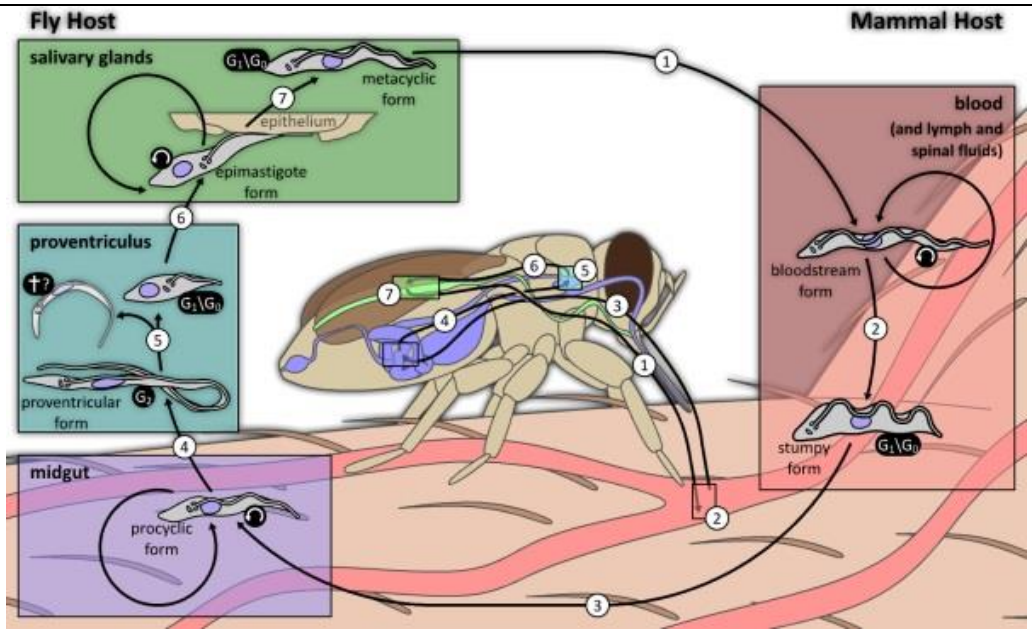


Figure 1: Life cycle of *T. brucei* as it progresses from fly to man and back again. (1) Metacyclic trypanosomes are transferred from the fly to a mammalian host during a bloodmeal. Parasites transform from metacyclic to bloodstream form cells in the interstitial spaces, bloodstream, and lymph system proliferates through binary fission. (2) A proportion of BF trypanosomes transforms to the non-dividing stumpy form, (3) ready for uptake by the insect vector. In the insect midgut, the parasite transforms to the procyclic form cells. Procyclics then migrate (4 -5) to the salivary gland of the fly, transforming to the (6) epimastigote stage where it attaches to the epithelium (7) and transform to the infective metacyclic form. Reprinted with permission (http://www.richardwheeler.net)

Both DNA transcription and replication appears to be lacking certain factors which are fundamental to these processes in the eukaryotic domain. These processes in *T. brucei* also show similarities with organisms from another ancient domain – the Archaea [5].

The nuclear genome is delineated by a multitude of epigenetic marks, demarcating transcriptional boundaries and adding to the mechanisms this organism utilises to control a genome with seemingly relaxed transcription. Some of these epigenetic entities, like histone variants and post-translational modifications (PTM), are part of a larger structure, the nucleosome, which packs the DNA into chromatin and provides a regulatory interface to the genome.

Chromatin is composed of repetitive arrays of nucleosomes, which are formed by 168 bp of DNA wrapped in two negative supercoils onto a histone octamer. Although nucleosomes represent the basic structural unit of chromatin, facilitating the compaction of poly-anionic DNA molecules to a level where it can fit into a cell nucleus, nucleosomes also serve as dynamic binding surfaces for proteins involved in gene regulation and transcriptional control.

Core histones are globular proteins with a characteristic histone fold domain and N-terminal "tails" extending from the central fold. An extensive range of PTMs occur on the tails that influence many biological processes, including chromatin condensation and the recruitment of DNA-binding proteins such as chromatin readers, writers and erasers [6]. Extensive studies have shown that histone PTMs can function either singularly or in combination with other PTMs, referred to as histone "cross-talk" [7]. Insight into the organisation of nucleosomes in a genome, as well as the distribution of histone variants and the presence of post-translational modifications, is essential to understand the regulatory role of chromatin. Here we aimed to integrate recent data gathered from the fields of genomics, transcriptomics and proteomics in order to understand the epigenetic mechanisms that are employed by *T. brucei* to control its gene expression programs.

1.3) Materials and Methods

Data presented here was gathered in a systematic search for research and review articles, in accordance with the PRISMA workflow (see supplementary material S 1.1) [4]. Original research articles was the primary focus, recent and relevant reviews was considered and mentioned if applicable or of importance for further reading/background. Areas of focus included fields included or related to genomics, epigenomics, transcriptomics and proteomics. Studies uncovering life-cycle dependent data was also targeted. Primary search results were not restricted to a specific time period, as to not exclude older, yet relevant studies. More value was given to research articles published from 2010 to 2013.

Primary and secondary search strategies as well as search engines and databases accessed are summarized in figure 2.

Articles identified from primary searches were assessed by considering the type, focus and date of each study. More refined and specific secondary searches were performed using keywords identified from the primary search. From these searches 174 papers were identified that conformed to one or more of the search criteria of which 97 was selected for this review and read in full. Data from highly cited older studies were also consolidated and related to more recent studies made possible by contemporary technological advances.

Primary search strategy

Trypanosoma, T. brucei

OR

T. cruzi, Leishmania (kinetoplasts), *Plasmodium, Saccharomyces, Caenorhabditis, Drosophila, Mus*, etc.

+

Genome, epigenetic, histone, nucleosome, chromosome, chromatin DNA modification, Base J, life cycle, polymerase I, II, III,

Secondary search strategy

Specific key words were identified from initial search strategy to allow investigation of individual topics (eg. Specific DNA binding proteins like TbRap1).

Search engines used for primary and secondary searches include: PUBMED, Google Scholar, TriTrypDB, MRS (searching SwissProt), psi-Blast.

Figure 2: Summary of strategies employed and databases accessed for literature retrieval.

1.4) Genome organisation

The haploid genome of *T. brucei* is 26 - 35 Mb in size, depending on the strain [8,9], and is composed of 11 megabase chromosomes (MBC), 1-5 intermediate chromosomes (IC) (300 – 900 kb), and approximately 100 minichromosomes (MC) (50 – 150 kb) [10]. MCs account for approximately 10% of the nuclear genome, and about half of each MC is composed of 177 bp AT repeats, as well as silent Variable Surface Glycoprotein (*VSG*) genes and pseudogenes [11].

The housekeeping portion of the genome, encoded by genes on the MBC, exists as long, non-overlapping, polycistronic transcription units (PTUs). Adjacent PTUs are separated by convergent or divergent strand switching regions (SSRs), referring to the directions of transcription of the bordering PTUs (Figure 3). The MBCs contain nearly 8800 non-redundant protein-coding genes, including about 500 pseudogenes [12], organised into unidirectional gene clusters that are interrupted by *tRNA*, *snRNA*, *siRNA* and *rRNA* genes. It is unusual for protein coding genes to be organised in directional PTUs on a genome wide scale [13] as is observed in *Trypanosoma*. Unlike generic prokaryotic PTUs, genes in trypanosomal polycistrons are not functionally related. Analysis of the *T. brucei* transcriptome revealed that RNA polymerase II (pol II) transcription initiates bidirectionally from putative pol II transcription start sites (TSS) at divergent SSRs, and could also occur at positions internal to PTUs [14,15].

Analysis on the control of gene expression and mRNA stability in response to heat shock showed a reduction in Pol II transcription initiation and mRNA half-life, as well as selective stabilization and translation of heat-shock protein mRNAs by specific RNA-protein interactions[16]. These studies also revealed cell cycle dependence of mRNA abundance corresponding to the position of a gene relative to the TSS within a PTU. Although early studies reported constitutive expression of all genes within a PTU, the regulation of transcript levels of single coding regions within a PTU has been observed [17].

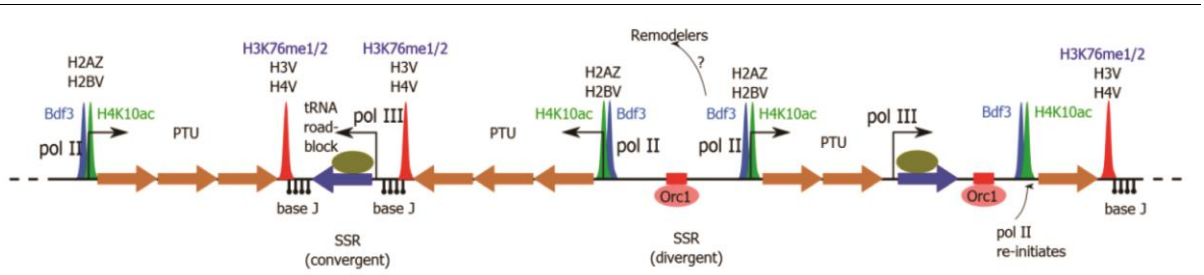


Figure 3: The epigenetic signals that demarcate transcription units and regulate the expression of genes in T. brucei. Pol II transcription initiates from weakly defined promoters in divergent SSRs with loci enriched for TbBDF3, H4K10ac and the H2AZ and H2BV histone variants. Transcription proceeds through polycistronic units that may span hundreds of kilobases that contain functionally unrelated genes. Transcription terminates in a region enriched for the modified thymidine base J, H3K76me1/2, and the H3V and H4V histone variants. TTSs often contain an active pol III transcribed tRNA gene. Where a tRNA gene interrupts a PTU, a region enriched for TbBDF3 and H4K10ac immediately downstream of the tRNA gene probably facilitates pol II re-initiation. Replication origins, nucleated by TbORC1, occur at the boundaries or upstream or re-initiation sites in PTUs.

1.5) Transcription in *T. brucei*

The initiation of transcription represents a key regulatory point for controlling the levels of gene products in most eukaryotes. A series of events involving *cis*- and *trans*-acting factors binding to specific DNA sequences, collectively functioning to recruit a specific polymerase complex and ultimately initiating mRNA synthesis, is the standard mechanism for regulation of eukaryotic gene expression. However, this paradigm does not seem to apply to *T. brucei*. The lack of classic pol II promoters, activators and co-activators, as well as basal transcription factors, coupled with constitutive polycistronic transcription, suggested that transcription initiation was not a fundamental regulatory event in mRNA synthesis [18]. Although the *T. brucei* genome encodes all five subunits common to the three classes of RNA polymerases [19], the kinetoplastids employ conventional polymerases for alternative functions.

1.5.1) RNA dependent polymerases

In *T. brucei*, pol I, apart from transcribing the *rRNA* genes, also transcribes two essential, life cycle specific genes that encode cell surface proteins. Procyclin, the major cell surface protein expressed in PF *T. brucei*, is transcribed from two polycistronic gene loci (GPEET and EP1) [20]. The *VSG* gene, encoding the BF stage cell surface antigen, is also transcribed by pol I. Pol I probably allows expression of high levels of a single transcript from a monoallelic transcription unit. This high level of expression allows the generation of an exceptionally dense VSG coat, which effectively shields the invariable cell surface antigens from the host immune system [21], allowing the parasite to escape immune clearance.

Pol II transcribes the majority of the PTUs, initiating mostly from divergent SSRs. Unlike in other eukaryotes, the *T. brucei* pol II promoter is weakly defined, and lacks a canonical TATA box and initiator sequence [22], although a TBP-like protein, TbTrf4, was identified [23]. Siegel and colleagues reported that oligo-G, locted between divergent SSRs, may act as an initiator element providing directionality to transcription [24]. The long, resulting, polycistronic RNA is spliced into individual, stable, translatable mRNA molecules by the co-transcriptional *trans*-splicing of a capped 39 bp spliced leader (SL) RNA, coupled with polyadenylation (reviewed in reference [25]).

Interestingly, these processes occur independent of the class of the transcribing polymerase, allowing production of mature mRNAs by polymerases other than pol II, crucial for mRNA synthesis by pol I. Because the polycistronic pre-mRNA contains numerous coding sequences and the active *VSG* is transcribed at extremely high rates, SL RNA must be produced at elevated levels to avoid mRNA production being a rate limiting factor in protein synthesis. Arrays of monocistronic, 1.4 kb tandem repeats of SL RNA are located on chromosome 9. These genes are transcribed at high rates by pol II, and contain the only defined and described pol II promoter in T. brucei [26,27].

An interesting feature of the *T. brucei* largest pol II sub-unit is the absence of the heptapeptide sequence repeat in the C-terminus of the protein. These heptad repeats, present in the pol II C-terminal domains of higher eukaryotes, contain serines which are phosphorylated, leading to promoter escape and progressive transcription [28,29]. *T. brucei* does, however, possess di-serines in the C-terminal of pol II and phosphorylation of pol II has been reported, but

it is unclear if this phosphorylation occurs in the C-terminal of the protein [20,29].

McAndrew *et al*. [30] suggested that an open chromatin structure was sufficient to initiate transcription in a trypanosome. This, taken together with the lack of identifiable promoter elements and the enrichment of specific histone PTMs and histone variants at SSRs, suggested that epigenetic control mechanisms played a central role in the modulation of pol II transcription initiation and termination in *T. brucei.*

Trypanosomal *tRNAs*, transcribed with other non-coding RNAs (snRNAs, SRP-RNAs) by pol III, are interspersed in and between PTUs. Since the *tRNA* gene itself may be in the process of transcription by pol III, or may be associated with regulatory proteins [31], the presence of a *tRNA* gene in a PTU presents a kinetic block to a transcribing pol II [32]. It is therefore likely that pol II would need to re-initiate downstream of a *tRNA* gene present internally in a PTU [9,24]. Figure 4 summarizes the components of RNA polymerase I, II and III present in budding yeast, human and *T. brucei*. Further details about specific components, sub-units and associated transcription factors indicated in table 1 as well as information regarding the search strategies and engines can be seen in the supplementary material (S 1.2 – 1.9).

RNA Polymerase I subunits

	RPA190	RPA135	RPA43	**RPC40**	**RPB5**	**RPB6**	**RPB8**	RPA14	RPA12
Yeast	X	X	X	X	X	X	X	X	X
Human	X	X	X	X	X	X	X	X	X
T. brucei	X	X	O	X	XX	XX	X	X*	X
	RPB10	**RPA19**	**RPB12**	RPA49	RPA34.5	RPA31			
Yeast	X	X	X	X	X	O			
Human	X	X	X	X	X	O			
T. brucei	X	X	X	O*	O	X			

RNA Polymerase II subunits

	RPB1	RPB2	RPB7	RPB3	**RPB5**	**RPB6**	**RPB8**	RPB4	RPB9
Yeast	X	X	X	X	X	X	X	X	X
Human	X	X	X	X	X	X	X	X	X
T. brucei	XX	X	X	X	XX	XX	X	X	X
	RPB10	RPB11	**RPB12**						
Yeast	X	X	X						
Human	X	XXX	X						
T. brucei	X	X	X						

RNA Polymerase III subunits

	RPB160	RPC128	RPC25	**RPC40**	**RPB5**	**RPB6**	**RPB8**	RPC17	RPC11
Yeast	X	X	X	X	X	X	X	X	X
Human	X	X	X	X	X	X	X	X	X
T. brucei	X	X	X	X	XX	XX	X	O	X
	RPB10	**RPC19**	**RPB12**	RPC82	RPC52	RPC37	RPC34	RPC31	
Yeast	X	X	X	X	X	X	X	X	
Human	X	X	X	X	X	X	X	X	
T. brucei	X	X	X	X	O	O	X	O	

Figure 4: RNA polymerase components of *S. cerevisiae*, *H. sapiens* and *T. brucei*. X indicates the presence of the sub-unit in the particular polymerase; two or more Xs indicate that the sub-unit contains multiple peptides. O indicates the absence of the sub-unit. **Bold** subunits are shared among polymerases. The asterisks indicate that the sub-units RPA49 and RPA14 might be replaced with RPB7 and RPB4 paralogues in *T. brucei* (Kelly *et al.,* 2005).

1.5.2) Long non-coding RNA

Abortive transcription can lead to a distribution of related RNA molecules. In one transcriptomic study, 103 transcripts, ranging in size from 154 – 2229 bp, were identified that did not possess recognizable coding potential [15]. Long non-coding RNAs (lncRNAs, defined as RNA > 100 nt) have been shown to play a critical role in gene regulation [33,34]. LncRNAs can act locally or globally as epigenetic regulators, like the Xist and HOTAIR lncRNAs, respectively [34,35], affecting DNA-protein interactions, chromatin condensation and gene activity. It is likely that some of the putative *T. brucei* lncRNAs similarly act at an epigenetic level, adding another layer of control to the regulation of trypanosome gene expression.

1.6) Telomeric silencing and bloodstream expression sites

The BF stage of *T. brucei* evades clearance by the human host immune system by periodically switching the monoallelically expressed *VSG* gene, and thus the VSG coat protein, from a selection of approximately 1500 *VSG* genes. The active *VSG* gene is co-transcribed with a set of expression site associated genes (*ESAG*s) from a single subtelomeric polycistronic unit known as the bloodstream expression site (ES). ESs are transcribed from a single RNA pol I promoter located 30-60 kb upstream of the telomeric repeats. The promoter is preceded by an array of 50 bp repeat sequences stretching for ~10-50 kb [36]. A total of 14 distinct ESs were identified in the Lister 427 *T. brucei* strain [37], of which the single, active ES is located in a sub-nuclear compartment, the expression site body [38]. The canonical structure and associated proteins of the ES and telomeric region are shown in figure 5. Due to its high relevance to

immune clearance in humans and the development of possible therapies, the ES has been the subject of intense research. Unlike the PTUs, which are constitutively transcribed, the ESs, as well as the procyclin loci, are subject to transcriptional regulation. Research has clearly demonstrated the involvement of epigenetic control mechanisms for these genomic loci.

The active ES in BF *T. brucei* was shown to be depleted of nucleosomes compared to silent ESs, a phenomenon probably related to transcriptional activity [39]. TbTDP1, an HMG box protein, was enriched at active ESs [40]. HMG Box proteins are capable of facilitating chromatin decondensation, thus making chromatin more accessible to regulatory factors, and facilitating the recruitment of transcription activators [41,42]. TbTDP1 was also enriched at the 50 bp repeats adjacent to ESs and immediately downstream of the *rRNA* promoter, binding to the entire *rRNA* locus. Diminishing TbTDP1 synthesis by RNAi resulted in an increase in histone abundance on pol I transcription units and a concomitant reduction in pol I transcriptional activity, leading to a growth arrest within 24h. TbTDP1 was essential for active pol I transcription, and was enriched at highly transcribed regions which were generally depleted of nucleosomes, including the active ES. This, along with recent results [43,44], strongly suggested the involvement of chromatin remodelling in the regulation of the transcriptional state of an ES. Indeed, the chromatin remodeller TbISWI was shown to play a role in repression of pol I transcribed ESs in both BF and PF stages of *T. brucei* [45]. TbISWI also contributed to the down regulation of PF-specific procyclin genes, non-transcribed *VSG* arrays and minichromosomes. The involvement of chromatin modellers in gene repression was previously shown in higher eukaryotes to be involved in the temporal regulation of the *pS2* gene, and SWI/SNF and NuRD were suggested to be required for resetting

the local nucleosomal structures to allow transcriptional shutdown in the absence of transcriptional activators [46]. The histone deacetylase, TbHDAC1, antagonised basal telomeric repression in BF cells, and TbHDAC3 was required for *VSG* ES promoter silencing in both PF and BF cells [47].

In the mammalian telomere complex, TRF2 is bound to duplex telomere DNA, and serves as a recruitment anchor for another telomeric protein, RAP1 [48,49]. *T. brucei* possesses functional orthologues of both these telomeric proteins, termed TbTRF2 and TbRAP1 [43]. TbRAP1 is found at telomeres, and is essential for growth and critical for ES silencing. Knock-down of TbRAP1 led to a graduated derepression of silent ESs. TbRAP1-mediated silencing increased within the terminal 10 kb of the telomeres, supporting the suggestion that telomeres were essential for *VSG* expression regulation [43]. In *Saccharomyces cerevisiae*, the telomere repeat binding protein Rap1, together with the Sir proteins, were shown to be required for telomere proximal silencing as well as for position effect variegation [50]. TbSIR2RP1, a Sir2 related protein, co-localized with telomeric sequences, and appeared to be involved in the establishment of a silencing gradient at the telomeres in the BF parasite [44]. Interestingly, orthologues to the yeast Sir3 and 4 proteins, which are recruited by Rap1 and Sir2 to form propagative, repressive chromatin structures at the telomeres and silent mating type loci in *S. cerevisiae*, appear absent in *T. brucei* (JPM and HGP unpublished data). This is perhaps not surprizing, since Sir3 only appeared in the *S. cerevisiae* genome by gene duplication of Orc1 after evolutionary divergence of the trypanosomes [51]. It is not clear what proteins, if any, may function with TbTRF2, TbRAP1 and TbSIR2 to establish a telomere-proximal repressive domain in *T. brucei*.

In contrast to laboratory strains, *T. brucei* field strains possess shorter telomeres [52], and switch *VSG*s more frequently [53,54]. A recent study revealed that telomere length is correlated with *VSG* switching frequency, and demonstrated that the shorter the telomere structure at an active ES, the more frequently *VSG* switching occurred [55].

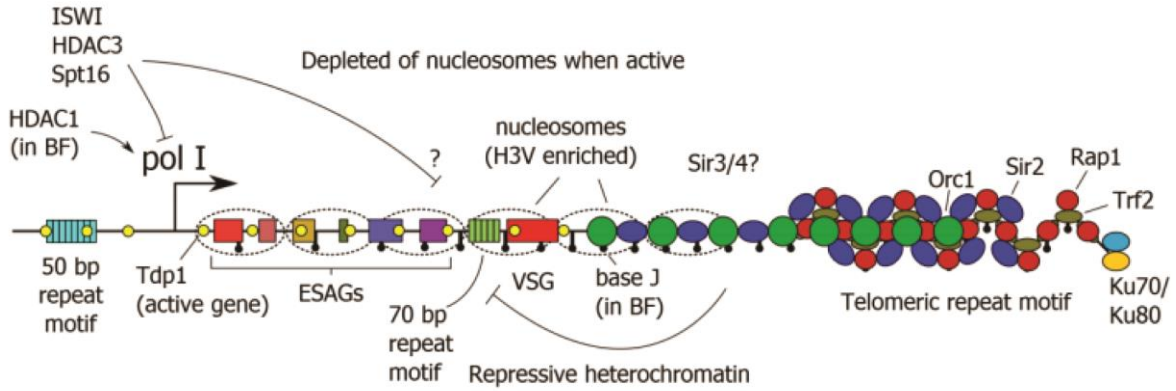


Figure 5: The epigenetic marks that define the transcriptional state of an ES. A repressive chromatin structure is formed by TbTRF2 and TbRAP1 (which may recruit Sir2) as well as TbORC1, propagating to sub-telomeric regions. It is not known whether other proteins fulfil the roles of yeast Sir3 and Sir4, for which orthologues are absent in T. brucei. Base J is present at an increasing density towards the telomere termini, and is required for ES silencing. Nucleosomes present on a silent ES are enriched for the transcriptional terminating variant H3V, and are depleted on an active ES. The HMG box protein TbTDP1 is present on the active ES, and is associated with chromatin decondensation. The histone deacetylase TbHDAC3 and the chromatin remodeller TbISWI is required for efficient ES silencing, and TbHDAC1 is required for activated expression.

1.7) Base J

One of many unusual epigenetic features found in *T. brucei* is the modified thymidine residue β-D-glucosyl-hydroxymethyluracil, designated base J, which is found in all kinetoplastids as well as in *Dipolonema* and *Euglena* [56]. In *T. brucei*, J was primarily associated with repetitive DNA elements such as the telomeric, 50, 70, and 177 bp repeats, and was also shown to localize at PTU flanks and at transcription termination sites (TTS) [57]. Base J was particularly enriched at silent *VSG* expression sites, forming an increasing gradient towards the telomeric termini (Figure 3 and 5).

Base J is developmentally regulated, and is only found in the BF stage of the *T. brucei* life cycle [58]. Two thymidine hydroxylases that are involved in the synthesis of J have been identified: TbJBP1 bound to J DNA and stimulated conversion of adjacent thymine residues to base J, whereas TbJBP2 was capable of *de novo* J synthesis. Deletion of these enzymes eliminated the first step of J biosynthesis. Although a JBP1 knock-out was lethal in *Leishmania* [59], *T. brucei* strains in which both enzymes had been knocked-out exhibited no serious growth defects [60]. The two-step J synthesis is depicted in figure 6.

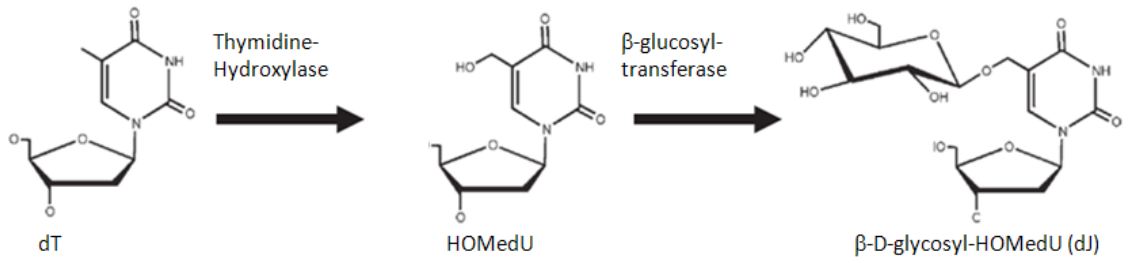


Figure 6: Biosynthesis of base J in a two-step pathway. Step one involves oxidation of thymidine (dT) forming HOMedU. Step two involves glucosylation of HOMedU to base j (dJ). Adapted from Cliffe *et al.* 2009.

In *Leishmania* the efficient termination of pol II transcription did not occur in the absence of J, unless pol II was terminated by a transcribing pol III [61]. Although the function of J in *T. brucei* remains unclear, it appears highly likely to interfere with pol II elongation, acting as a transcriptional terminator and epigenetic repressor.

1.8) Replication origin complex and gene silencing

Similar to transcription, DNA replication initiates with the assembly of a pre-replication complex at an origin of replication sequence. This complex is composed of the Origin Recognition Complex (ORC), Cdc6, Cdt1 and MCM [62]. Genome-wide analysis of TbORC1/CDC6 (referred to as TbORC1 henceforth) binding sites in *T. brucei* revealed an overlap between replication origins and the boundaries between PTUs. All replicated origins occurred in chromosomal core regions associated with transcription initiation and termination [63]. In *S. cerevisiae*, silent genomic regions such as the silent mating type loci are bordered by A-boxes, sequences recognized and bound by ORC1. ORC1, which contains a nucleosome binding BAH domain, nucleates a complex that is essential in facilitating transcriptional silencing in the adjacent genome [51]. Surprisingly, TbORC1 does not contain an identifiable BAH domain, but was shown to be required for efficient sub-telomeric repression and ES silencing [63,64]. It is not currently known whether all TbORC1-binding sites in *T. brucei* represent active origins, or whether a subset, specifically those located in sub-telomeric regions or bound to silent *VSG* arrays, functions in gene silencing, similar to *S. cerevisiae* and other yeasts.

1.9) Nucleosomal organization

Genome-wide maps of the nucleosome organization in model organisms show a common arrangement of nucleosomes at specific genomic features. A nucleosome depleted region (NDR), exposing part of the proximal pol II promoter, is seen in yeast [65,66], *Caenorhabditis elegans* [67], *Drosophila* [68], and in humans [69]. The NDR is bordered by two well positioned nucleosomes: -1 on the upstream and +1 on the downstream side of the NDR, followed by a nucleosomal array extending over the gene. High levels of histone variants and post-translational modifications are observed for nucleosomes flanking the NDR [70].

In contrast to the above model, where generally only the promoter region is exposed, genome-wide mapping of nucleosomes in *Plasmodium falciparum* revealed a different picture [71]. Nucleosomes were found to be associated with coding regions and generally absent from intergenic and promoter regions. In addition, the nucleosomal organization of several TSSs did not correlate with either nucleosome-free or intergenic regions. The high AT-content of the intergenic regions of *Plasmodium* may selectively exclude nucleosomes, allowing easy access to polymerases and associated factors [72].

1.10) Histone epigenetic patterns

1.10.1) H1

Trypanosomal histone H1 differs noticeably from that of other eukaryotes. *T. brucei* H1 is comprised of a single domain corresponding to the lysine rich C-terminal domain of higher eukaryotic histone H1. This arrangement is similar to *Tetrahymena* H1 [73], which lacks the central winged helix domain. A recent study demonstrated the involvement of TbH1 in maintaining a condensed state of chromatin at non-transcribed regions, including the silent *VSG* basic copy arrays and inactive *VSG* ESs. TbH1 is not only required to down-regulate silent *VSG* ESs, but may also suppress *VSG* switching [74].

1.10.2) H2A

Several studies of *T. brucei* H2A PTMs revealed the absence of modifications that were well conserved in other eukaryotes [75,76]. Additionally, a number of trypanosome-specific PTMs were also identified [75,76].

Analysis of the first 22 amino acid residues of histone H2A revealed 60% monomethylation of A1 and an ~1% acetylation of K4 [76] (Figure 7 A). H2A displayed a complex pattern of multiple PTMs of the C-terminus, including 6 acetylated lysines (K115, K119, K120, K122, K125, and K128) of which three (K120, K122, and K128) corresponded to conserved lysine residues with defined epigenetic marks in other species.

It is possible that *T. brucei* H2AK122 could be ubiquitinated, since it is the only lysine in the H2A C-terminus adjacent to a potential phosphorylation target, S123. It was suggested that phosphorylation influenced the ubiquitination of neighbouring lysines [77]. In addition, *T. brucei* H2AK122 aligns with human

H2AK119, a site of ubiquitination associated with transcriptional repression [78].

1.10.3) H2B

H2B is the least conserved of the four core histones [79] in *T. brucei*, and analysis has revealed only 4 PTMs. The same degree of methylation of A1 and acetylation of K4 was observed as for H2A. Tandem MS analysis showed minor acetylation of K12 and K16 [76] (Figure 7 B).

Evidence of life stage-dependent modifications is also seen in *T. brucei*. Acetylated lysine residues are observed at K4 and K122 in H2A and at K4, K12 and K16 of H2B in BF trypanosomes, but not in the procyclic form [76] (Figure 7 B).

1.10.4) H3

Histone H3 as well as its N-terminal tail is highly conserved from human to yeast, where the tail is subjected to an extensive range of PTMs. In *T. brucei* H3, however, only a few PTMs have been mapped to specific residues. MS analyses revealed that S1 and K23 were acetylated, K4 and K32 were tri-methylated, and K76 could be mono-, di- or tri methylated [76] (Figure 7 C). Internal sequences of the H3 tail diverge sharply from that of canonical H3, but sequence alignment suggests that *T. brucei* K19, K23, K32 and K76 could be equivalent to K23, K27, K36, and K79 of other eukaryotes, respectively. Although many of above PTMs have been functionally described in other organisms [6,80], the functional roles of these PTMs in *T. brucei,* with the exception of K76, is not known.

TbDOT1A was responsible for mono- and di-methylation of H3K76. The RNAi knock-down of TbDOT1A resulted in severe cell cycle defects [81,82]. A clear correlation exists between H3K76 mono- and di-methylation and transcription termination sites, suggesting a role in transcription termination [82]. Tri-methylation of H3K76 was mediated by TbDOT1B, which was not essential for viability [82]. Mono-, di-, and tri-methylation of K76 was implicated in several processes, including replication control, antigenic variation, and developmental differentiation [79,82]. K76 di-methylation is only detectable during mitosis [81].

Mandava and co-workers [83] reported that the H2B variant, H2BV, was present in mononucleosomes enriched for tri-methylated H3K4 and K76, and suggested that H2BV can replace canonical H2B, permitting H3K4 and K76 methylation. A puzzling feature among kinetoplastids is the absence of the almost universally conserved H3K9, implicated in gene repression in its tri-methylated state [84]. It is not yet clear whether K10 is the equivalent residue, although the sequence context of *T. brucei* H3K10 is markedly different from that of K9 in other eukaryotes.

1.10.5) H4

Of all trypanosomal histones, H4 is the most conserved. As in H2A and H2B, H4A1 is also monomethylated to a level of approximately 60% (Figure 7 D). K4, K5, K10, and K14 were observed to be acetylated, and K2, K17, and K18 were acetylated or methylated to various extents [76]. Sequence alignment showed the presence of lysine residues at both position 4 and 5 in trypanosomes. In other eukaryotes glycine is the conserved residue at position 4 in H4. H4K4 is the most commonly acetylated histone tail residue in *T. brucei* [75], suggesting that *T. brucei* K4 was the functional equivalent of K5 present in other

eukaryotes. The histone acetyltransferase TbHAT3 was responsible for H4K4 acetylation in both PF and BF life stages. This non-essential, MYST-type acetyltransferase seemed to acetylate H4 upon import into the nucleus for packaging of newly-replicated DNA [85].

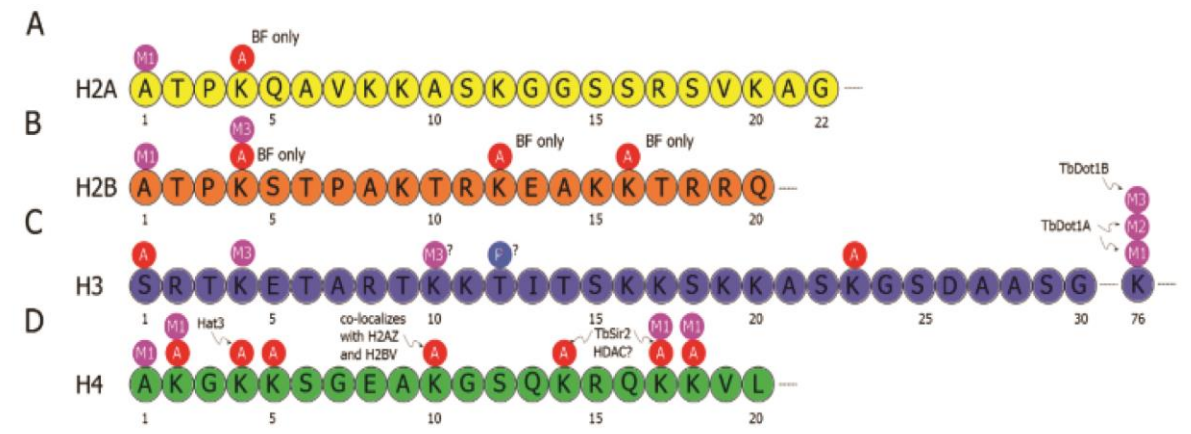


Figure 7: Epigenetic modifications of the *T. brucei* core histone N-terminal tails. Modifications mapped to specific residues and enzymes involved in the modulation of some modifications are shown. Life stage specific modifications of the parasite are also identified. It is not currently known whether H3K10 or H3K11 is the equivalent of the highly conserved H3K9 present in other eukaryotes, and whether H3T12 is the equivalent of H3S11, a known phosphorylation target. A = Acetylation, Me = Methylation, P = Phosphorylation.

ChIP-seq studies showed twin peaks of acetylated H4K10 at divergent SSRs [24]. A number of single acetylated K10ac peaks were found at non-SSRs, many of which occurred downstream of *tRNA* genes. Most *tRNA* genes are located at convergent SSRs, and of those located at non-SSR, all but 3 of 38 were located upstream of a single acetylated K10ac peak. If a *tRNA* gene represents a roadblock to pol II transcription, for example, within a PTU, pol II would need to re-initiate downstream of the *tRNA* gene, within the region enriched for H4K10ac. This, together with the observation that pol II

transcription initiated at divergent SSRs, suggested a link between transcription initiation and acetylated H4K10. Higher levels of this modification were observed upstream of the first and downstream of the last PTU of each chromosome [24]. Distribution profiles of this modification were remarkably similar between parasite life stages, with only two life stage-specific peaks being observed (on chromosome 7 and 11).

The Bromodomain Factor 3 (TbBDF3) was shown to bind to acetylated lysines [86]. It co-localises with acetylated H4K10 and is concentrated towards the upstream end of H4K10ac peaks. It was suggested that TbBDF3 is involved in targeting chromatin remodelling complexes to TSSs [86,87]. TbBDF3 was essential for cellular viability, and RNAi mediated knock-down caused an immediate growth defect where most cells died within 48h [24].

Chaperone proteins assist with the assembly or disassembly of macromolecular structures. Three *T. brucei* chaperone proteins, TbASF1A, TbCAF-1b, and the FACT complex were shown to be important for the maintenance and inheritance of epigenetically determined states of silent ESs [88,89].

The FACT (Facilitates Chromatin Transcription) complex is capable of deposition of core histones as well as binding and displacing an H2A/H2B dimer from a nucleosome [90,91]. The *T. brucei* FACT complex was hypothesised to play roles in progressive pol I transcription of the active *VSG* ES, as well as the establishment and maintenance of heterochromatin at centromeric sequences. The FACT subunit TbSPT16 was shown to bind to ESs, and was enriched at silent ES promoters. RNAi of TbSPT16 resulted in derepression of *VSG* ESs in both life cycles, disruption of minichromosome segregation, and also in cell cycle arrest [89].

TbASF1A is involved in the recycling and assembly of histone H3-H4 dimers during DNA replication and transcription. TbCAF-1b, in contrast, is predominantly a replication dependant histone chaperone, and also involved in H3-H4 dimer recycling and assembly. TbASF1A RNAi mediated depletion resulted in *VSG* ES derepression at all cell cycle stages, suggesting a replication independent role, whereas TbCAF-1b knock-down resulted in *VSG* ES derepression primarily in S and G2/M cell cycle stages [88].

Nucleoplasmins are small proteins that function as histone chaperones [92]. TbNLP is a nucleoplasmin-like protein containing an AT-hook motif [93]. Although not strictly a nucleoplasmin, homology suggested that TbNLP could interact with histones. TbNLP bound to both active and silent ESs, and could function in facilitating and inhibiting transcription, depending on the epigenetic context of its molecular environment. Consistent with this, RNAi mediated depletion of TbNLP caused derepression of silent ESs as well as a reduction in progressive transcription of the active *VSG* ES, leading to a growth arrest within 24h. TbNLP appears to be a general transcription regulator in both life-cycle stages, since it binds other transcriptionally inactive genomic regions, including the 50 and 177 bp repeats, *VSG* basic copy arrays and the procyclin loci [93].

1.11) Histone variants

T. brucei encodes four histone variants: H2AZ, H2BV, H3V and H4V. Nucleosomes that contain H2AZ are less stable than nucleosomes containing canonical H2A [24,94], and data also suggested that H2AZ containing chromatin is less condensed, and thus primed for transcription. H2AZ was shown to be associated exclusively with H2BV in *T. brucei* [95,96], exhibiting virtually identical ChIP profiles and similar genomic distributions during the cell cycle. Both histone variants were shown to be essential for cell viability [95].

ChIP-seq of H2BV revealed a genomic distribution almost indistinguishable from that of acetylated H4K10. Distinct matched peaks were observed at divergent SSR as well as single H2BV peaks at non-SSR, coincident with that of H4K10ac. Since H2AZ exclusively dimerized with H2BV, it can be presumed that H2AZ would show a similar genomic distribution. H2BV was also shown to be present in nucleosomes that were enriched for trimethylated H3K4 and H3K76, PTMs typically associated with transcriptionally active chromatin [6].

The variant H3V was found to be highly enriched at telomeric repeats and subtelomeric regions (see Figure 3), but not at the 177 bp minichromosome repeat or 5S *rRNA* loci [97]. Single peaks of H3V nucleosomes were located at convergent SSR and upstream of all H4K10ac-rich regions not associated with a SSR. Sequence analysis of regions rich in H4K10ac revealed G-rich stretches of 9 to 15 guanine residues at SSRs.

The distribution of H4V was found to be similar to that of H3V throughout the genome [24]. H4V was, however, less enriched compared to H3V at sub-telomeric and telomeric sites [24]. Both H3V and H4V were found to be significantly enriched immediately downstream of the last coding sequence of a PTU (see Figure 3 and 5). This suggested that H3V-H4V containing nucleosomes were enriched at presumed pol II TTS, and thus serves as epigenetic markers for the end of transcription units.

Collectively, these findings suggest that putative RNA polymerase transcription start and termination sites are demarcated by specific histone variants and PTMs, likely conferring defined structural states to local chromatin regions, and recruiting functionally important chromatin associated proteins to such regions.

1.12) Conclusions

The many studies cited in this review have provided ample evidence that in *T. brucei*, rather than being a constitutive, unregulated process, where transcript levels are only controlled post-transcriptionally, gene expression, particularly as regards the genes encoding the major cell surface proteins, is closely tied to chromatin. Therefore, although there is little regulatory control at the level of transcription at the PTUs, chromatin plays a key role in delineating *T. brucei* transcription units, and in controlling the initiation of transcription as well as DNA replication. These boundaries are demarcated by an assortment of other epigenetic signals, like histone PTMs and histone variants. Histone "cross-talk" has also been observed in *T. brucei* nucleosomes. These marks are deposited by specific readers and writers. A multitude of chromatin remodellers (TbHDACs, TbSIR2, TbISWI, Chaperone proteins, TbNLP, TbBDF3, TbTDP1, TbTRF2, TbRAP1) were found co-localising with putative TSS, TSS as well ORC sites. This marvellous coordination of epigenetic factors functions altogether to regulate life cycles in the absence of the normal eukaryotic control systems.

Chromatin is a dynamic structure, providing an interface to the regulation of DNA function. This interface includes the composition and positioning of nucleosomes, determined by intrinsic DNA sequence preferences, transcription factors, chromatin remodellers, and by active transcription. Furthermore, specific histone variants and histone modification states synergise to provide a rich regulatory interface to control gene expression. This regulation of gene transcription by the epigenome provides the exciting possibility that epigenetic components may represent novel drug targets, and that epigenetic therapies may be developed to treat this lethal disease in future.

Chapter 2

Nucleosomal architecture in insect and bloodstream form *Trypanosoma brucei.*

2.1) Abstract

Trypanosoma brucei is an extracellular parasite of the mammalian bloodstream that causes African sleeping sickness in humans. This protist displays highly unusual genomic characteristics, like the transcription of protein coding genes by pol I, the absence of canonical pol II promoters, and polycistronic gene organisation on a genome-wide scale. These sites have also been found to be enriched with a myriad of epigenetic markers, like histone PTMs and variants. Work over the past decade has revealed that epigenetics plays a key part in genome regulation and antigenic variation in *T. brucei*. These epigenetic marks are associated with nucleosomes, which packs DNA into chromatin and has been revealed to have a major impact on gene expression and silencing. We produced a map of all nucleosomes covering the megabase chromosomes in both procyclic and bloodstream form *T. brucei* by MNase digestion of chromatin and paired-end sequencing of mononucleosomal DNA fragments. The fragments were realigned and nucleosome positions determined. This revealed nucleosomal architectures surrounding pol II transcribed PTUs to be comparable to those of other model eukaryotes. Remarkably, there appears to be no significant difference in nucleosomal architecture between the two life cycles. The effect of histone H1 knock-down by RNAi revealed a change in nucleosomal patterns in the BF, with little effect observed in the PF. Sequence analysis also revealed the use of intrinsic sequence positioning signals, which directly oppose DNA-octamer binding, to position nucleosomes relative to specific genetic marks. It also appears that the preferential distributions of A/T and G/C dinucleotides are employed to impart a rotational position on well positioned nucleosomes. This could indicate the need for well positioned nucleosomes at putative transcription start sites which can possibly assist in nucleosome-mediated transcription start site selection.

2.2) Introduction

The kinetoplast *Trypanosoma brucei* is a unicellular flagellated protist that causes Human African Trypanosomiasis, also known as African sleeping sickness. The parasite is transmitted by a *Glossina spp.* fly to the mammalian host during a blood meal, invades the interstitial spaces, lymph and bloodstream, and multiplies asexually.

During its two life stages, the Procyclic form (PF) and Bloodstream form (BF) occurring in the insect and mammal hosts, respectively, different coat proteins cover the cell surface [98]. Procyclin is expressed only in the PF stage of the parasite when residing in the insect mid-gut. Procyclic trypanosomes travel to the salivary glands of the fly and exchange the procyclin coat protein for a Metacyclic Variable Surface Glycoprotein (MVSG) [99], ready to infect a mammalian host. Once transmitted to a mammal, the parasite expresses a single Variable Surface Glycoprotein (VSG), covering the cell surface in a dense (10^7) coat which effectively shields non-variable cell surface antigens from being recognised by the host immune system [100]. A single *VSG* is transcribed mono-allelically by polymerase I from one of about 15 sub-telomeric bloodstream Expression Sites (ES) [8,37]. To avoid clearance by the host adaptive immune-response, the parasites periodically switches the *VSG* with one of ~ 2000 alternate *VSG* genes in a process called antigenic variation [101]. This results in oscillating waves of infection and ensuing partial immunological clearance. With prolonged infection, in the absence of treatment, the parasite eventually crosses the blood-brain barrier and invades the central nervous system, presenting the typical symptoms of the disease from which the name is derived.

The nuclear genome of *T. brucei* consists of 11 megabase chromosomes (MBC, >1 Mb), 1 – 5 intermediate chromosomes (IC, 300 – 900 kb) and ~ 100 minichromosomes (MC, 50 – 150 kb), totalling 26 – 35 Mb per haploid genome, depending on the strain [8,9]. This variation in genome size is caused by the variation in population size of the MC, which totals ~10% of the total nuclear genome and carries a vast repertoire of silent *VSG* genes [11]. The MC and IC contain no housekeeping genes and are composed of silent *VSG* genes and simple AT-rich repeats. These nuclear entities effectively enlarge the silent *VSG* repertoire which is carried in the sub-telomeric regions of the MBCs.

The housekeeping portion of the genome resides on the 11 MBCs (Fig 2.1 A) and is encoded as long, unidirectional polycistronic transcription units (PTUs) separated by strands switching regions (SSR) [9]. These SSRs can be either convergent or divergent, depending on the direction of transcription (Figure 2.1 B). Polycistronic transcription is not unique to kinetoplasts and was observed in other eukaryotes, like *Caenorhabditis elegans* [13]. However, PTUs in *T. brucei* can cover tens of kilobases and unlike bacterial operons, does not contain functionally related genes.

The sub-telomeric loci of the MBCs typically contain silent *VSG* arrays or BESs. Figure 2.1 C shows the typical architecture of a BES; a polycistronic unit containing a pol I promoter and a set of expression site associated genes (ESAGs) which are co-transcribed with the active *VSG* [37]. The precise function of these ESAGs are not yet elucidated, but seem to contribute to parasite virulence or enables the organism to parasitize different mammalian hosts [37].

MBC in *T. brucei* are richly interspersed with simple AT-rich tandem repeats (TRs), some of which have been mapped as centromeric repeats in chromosome 1 – 8 [102], while others (specifically those in sub-telomeric regions – the 50 and 70 bp repeats, fig 2.1 C) facilitate antigenic variation by homologous recombination [11]. Another interesting feature of nuclear *T. brucei* DNA is the presence of a hypermodified thymidine, β-D-glucosyl-hydroxymethyluracil, or Base J [103]. Base J primarily associates with repetitive DNA sequences in BF trypanosomes and were shown to be non-essential in *T. brucei,* although knock-out of key enzymes in Base J synthesis was lethal in a related Leishmanian kinetoplast [57,59].

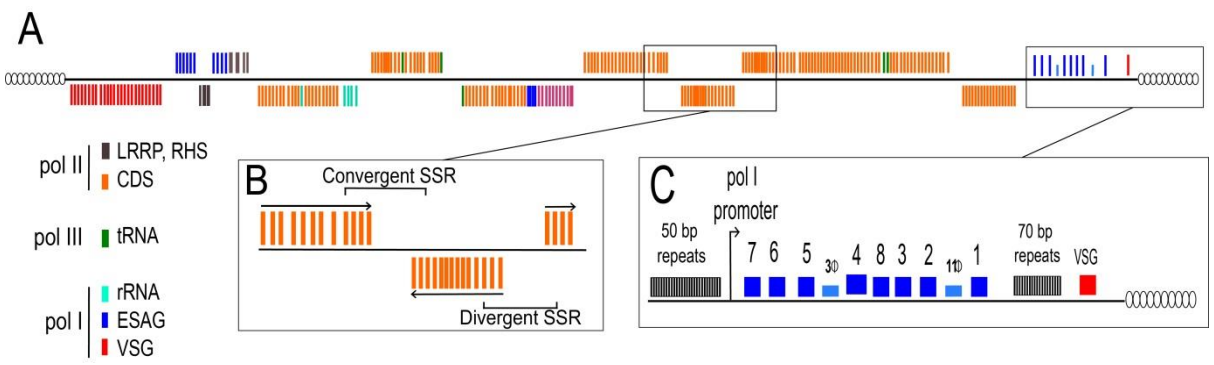


Figure 2.1: Schematic representation of a generic *T. brucei* megabase chromosome (panel A), showing gene organisation by their class and transcribing polymerase. Coloured bars indicate gene positions and lengths of different genetic elements. Panel B illustrates the strand switching regions, which are either convergent or divergent, depending on the direction of transcription (black arrows). Panel C shows a generic bloodstream expression site, containing a single terminal *VSG* which is co-transcribed along with multiple ESAGs as a single PTU by pol I.

Also interspersed through the MBCs are genes transcribed by polymerases I (*rRNAs*, excluding 5S *rRNA*) and III (*tRNAs*), with some occurring internal to pol II PTUs (Fig 2.1 A). As these genes might themselves be in the process of transcription or might be stably associated with transcription factors (TFs), this could present a roadblock to a transcribing pol II. The polymerase might pause and then continue transcription after the template has been cleared of molecular obstructions, or it could terminate transcription and re-initiate downstream of the pol I or III roadblock.

Another striking feature of *T. brucei* is the absence of readily identifiable pol II promoter sequences [26,27]. Rather than being defined by consensus sequences, pol II transcription start and termination sites (TSS and TTS, respectively) are demarcated by co-localized epigenetic marks, like histone PTMs and variants [24]. The lack of canonical pol II regulatory sequences and localization of specific epigenetic signals seems to imply that chromatin plays a role in transcriptional regulation.

To solve the spatial and steric hindrances DNA is packaged into chromatin, visible as sister chromatids during mitosis, in order to fit into the confines of the eukaryotic nucleus. Chromatin is composed of ~168 bp of DNA wound around a histone octamer composed of two copies each of histone H2A, H2B, H3 and H4, and associated histone H1, functionally described as a nucleosome. To further compact and stabilize the DNA in higher order chromatin structures, the linker histone H1 interacts with both the nucleosome and the linker DNA [104].

Histone H1 is usually associated with repressive heterochromatin structures and may function as a general transcription repressor in eukaryotes, yet the precise function of H1 has not yet been elucidated [105,106]. Knock-out of H1

affects a small number of genes and has been shown to be dispensable in several unicellular eukaryotes. *T. brucei* H1 does not contain the central globular domain conserved in other eukaryotes, thought necessary for interaction with the nucleosome, and instead consists of a single domain corresponding to the C-terminal of H1 of higher eukaryotes. It has been shown that RNAi-mediated TbH1 knock-down resulted in significant changes in chromatin structure and increased sensitivity to endonucleases in BF *T. brucei* but not in the PF. Histone H1 has also been implicated in silencing VSG BES promoters and supressing VSG switching [74].

Core histones (H2A, H2B, H3 and H4) are highly conserved throughout the eukaryotic domain. However, *T. brucei* again display significant sequence divergence from the canonical histone sequences as well as what seems to be trypanosome specific PTMs [76].

It has been demonstrated that nucleosome positioning can have a major effect on gene expression in eukaryotes. Nucleosomes can hamper transcription and high nucleosomal occupancies on DNA can have a transcriptionally repressive effect [107]. It has been demonstrated that upon gene activation extensive nucleosomal remodelling and eviction takes place [108]. Recent advances in sequencing technology allowed the generation of massively paralleled sequencing and determination of precise locations of individual nucleosomes on a genome-wide scale. A common theme emerged as genome-wide nucleosomal maps were generated and provided insights into the nucleosomal architecture around protein coding genes. Figure 2.2 provides a generic architecture of a eukaryotic pol II transcribed gene [109].

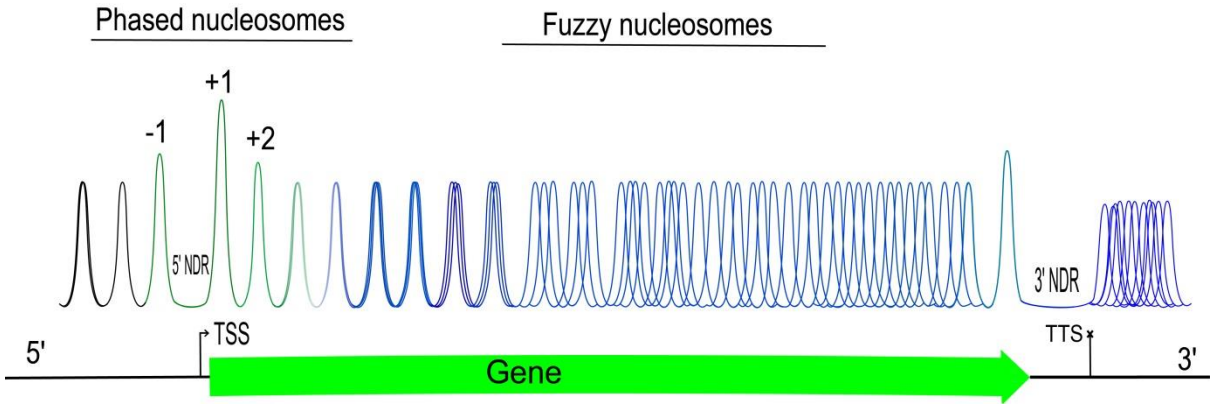


Figure 2.2: Generic nucleosomal architecture surrounding pol II transcribed genes in yeast. Upstream of the TSS, a 5' nucleosome depleted region flanked by two well positioned nucleosomes, the -1 and +1 nucleosomes are present. These nucleosomes are enriched with histone variants and PTMs (indicated by green shading). Following the +1 nucleosome is an array packing the gene which diminishes in phasing and histone modifications. At the 3' end of the gene is a well-positioned nucleosome preceding the 3' NDR and the TTS.

Regulatory elements, such as upstream promoter elements and core promoter sequences, reside in a 5' nucleosome depleted region (NDR), directly upstream of the gene. Flanking the 5' NDR are two well positioned nucleosomes, the -1 and +1 nucleosomes at 5' and 3' positions relative to the NDR, respectively. These nucleosomes are well positioned and highly phased, assuming very precise locations relative to the TSS. They also possess high levels of histone variants and PTMs, which may assist the pre-initiation complex (PIC) assembly (indicated by green shading). Following the +1 nucleosome is the +2 nucleosome which appears to be less well positioned and contain less histone PTMs or variants. This decreasing trend continues in downstream nucleosomes which packages the gene, although the most extreme 3' nucleosomes appears to be slightly more phased than nucleosomes in the gene interior. At the end of the gene there appears a 3' NDR where pol II transcription terminates. Nucleosomes tend to be uniformly spaced with a

fixed distance from each other, known as the nucleosome repeat length (NRL). The budding yeast *Saccharomyces cerevisiae* has a NRL of 165 bp, *C. elegans* 175 bp and humans a 185 bp NRL. This plot is a simplification of the nucleosomal organisation, which generally show significant heterogeneity within a population of cells [31].

Specific nucleosome positions are maintained by a combination of chromatin remodellers and intrinsic DNA sequence preferences. A 10 periodicity in the distribution of AT and GC dinucleotides have been shown, with a 5 bp offset between them. The 10 bp periodicity of these dinucleotides probably provides a rotational setting for nucleosome-bound DNA as AT nucleotides tens to expand the major groove of DNA, while GC nucleotides contracts the major groove[109,110], resulting in an anisotropically flexible length of DNA.

It is known that poly:(dA/dT) runs are resistant to bending, as was seen from a crystal structure of an oligo-A run where bifurcated hydrogen bonds was visible between A and T nucleotides at positions *n* and *n+1* on opposite strands. The bifurcated hydrogen bonds were proposed to increase the stiffness of the oligo-A run, which was preferentially excluded from the internal regions of nucleosomes, where the DNA must bend through 360° over 80 bp of duplex, thus creating NDRs [110].

The NDR usually contain regulatory elements to which TFs are recruited and facilitate pre-initiation complex (PIC) formation. However, in the absence of these elements, it is possible that positioned nucleosomes may assist in PIC formation. TATA binding protein (TBP) is a universal eukaryotic basal transcription factor, that has a critical role in pol II transcription complex formation [111]. Although most genes contain regulatory elements to which PIC proteins, such as TFIID containing TBP, are recruited, there are cases of

transcription in the absence of a TATA box containing promoter. Here, the TBP is recruited by protein-protein interactions and then moves to the promoter by sliding along the DNA or looping over DNA [112]. How then, in the absence of core promoter elements, does the PIC establish formation at the transcription start site in *T. brucei*?

As some PIC components contain nucleosome binding subunits (in yeast the bromodomain containing factor 1, BDF1, of TFIID), it could be that positioned nucleosomes may define TSS by positioning the PIC. TSSs are often associated with PTMs like acetylated H3K4, H3K9, and H3K14 which can be recognised by chromatin binding proteins. These proteins can bind acetylated nucleosomes and recruit PIC formation, thereby positioning the transcription machinery at the promoter [109].

T. brucei does not have canonical pol II promoters containing a TATA box, but does however, have a TBP related protein, TRF4, which is thought to bind a TTTT box and was found associated with pol I and III transcribed genes as well as the SL RNA promoters [26,27]. This, taken together with TSS being demarcated by specific histone PTMs, might indicate that such a mechanism can function in *T. brucei*.

T. brucei employs a multitude of epigenetic readers, writers and erasers to deposit and regulate epigenetic markers. The marks are frequently in the form of histone PTMs or variants, thereby providing a dynamic interface to the genome via the nucleosomes. This suggests that nucleosome positioning and perturbation is required for regulating genetic function in this trypanosome. The absence of canonical pol II regulatory elements defining TSS and putative pol II TSS being demarcated by epigenetic markers implicates nucleosome positioning in the placement of pol II at putative TSS. Genome wide

nucleosomal mapping in *S. cerevisiae* [113], *Schizosaccharomyces pombe* [114], *Drosophila* [115], *C. elegans* [67], *Plasmodium falciparum* [71] and human T-cells [69] have provided valuable insight into the nucleosomal organization around regulatory elements and genes transcribed by different polymerases. The effects of gene activation on nucleosomal organization, and in contrast, gene silencing by nucleosomal restructuring, has proven valuable in elucidating life-cycle dependent differential gene expression patterns. Insight into the nucleosomal organization of the *T. brucei* genome will therefore provide valuable information regarding the role of chromatin and epigenetics in the expression and regulation of DNA in this kinetoplast.

In this study, we aimed to generate genome wide nucleosomal maps of both PF and BF *T. brucei* Lister 427 trypanosomes to understand the epigenetic regulation that takes place in these two life cycles. This was achieved by isolating mononucleosomal core particles following MNase digestion, which was then sequenced using the Illumina platform, allowing massively paralleled sequencing of isolated mononucleosomal DNA. The sequence reads were mapped back to the *T. brucei* Lister 427 strain using Bowtie 2 [116]. Nucleosomal maps of pol I, II and III transcribed genes were generated, revealing specific nucleosomal positioning around these loci. We also explored the DNA context of pol II PTUs regarding DNA base composition and showed a preferential AT skew upstream of PTUs which might indicate sequence specific nucleosomal positioning.

We investigated the effect of histone H1 depletion on nucleosomal positioning and overall chromatin architecture by RNAi mediated knock-down of H1 in both PF and BF trypanosomes.

2.3) Materials and Methods

All reagents used were of Molecular Biology grade purity, unless otherwise stipulated. This section is functionally separated into *in vitro* and *in silico* sections, dividing wet-work and computational analysis.

In vitro

2.3.1) Trypanosome strains and culture

Bloodstream form (BF) Lister 427 (HNI-VO2) *T. brucei* was cultured in HMI-9 medium (1L Isocove Modified Dubecco's medium (IMDM) powder (Sigma), 0.136g hypoxanthine, 0.1817g L-cysteine, 0.1100g sodium pyruvate, 0.0387g thymidine, 0.0282g bathocuproine disulphonic acid, 3.024g sodium bicarbonate, pH 7.4) supplemented with 15% v/v foetal calf serum (FCS) and 0.1mM β-mercaptoethanol and appropriate drugs (2 μg/ml neomycin). Cells were cultured at 37^{o}C with 5% CO_2 as previously described [117].

For H1 RNAi studies, BF *T. brucei* strains were grown as described above. *T. brucei* RYT3 (stabilite no. mp 100) was used as parental strain, which was then transfected to produce a tetracycline inducible RNAi-mediated histone H1 knock-down strain, termed RYT3_H1C3 (stabilite no. mp 228). For information regarding the construction of this strain, see *Povelones et al.,* 2012 [74]. H1 RNAi knock-down trypanosomes were cultured in HMI-9 medium as discussed above with appropriate drugs (2 μg/ml neomycin, 2.5 μg/ml phleomycin, 5 μg/ml hygromycin, 5 μg/ml blasticidin). Histone H1 RNAi was induced by addition of 1μg/ml tetracycline and cultured for 48h before harvesting.

Procyclic form (PF) *T. brucei* (Amsterdam wild-type) was cultured in SDM-79 medium (SDM-79, 2 g/l sodium hydrogen carbonate pH 7.3), supplemented with 10% FCS and 5 μg/ml hemin at 27^{o}C.

For H1 RNAi studies, PF *T. brucei* strains were grown as described above. *T. brucei* DsRed (stabilite no. mp 113) was used as parental strain, which was then transfected to produce a tetracycline inducible RNAi-mediated histone H1 knock-down strain, termed DsRed_H1_1.1c (stabilite no. mp 237). For information regarding the construction of this strain, see *Povelones et al.,* 2012 [74]. PF parental and RNAi knock-down strains were grown as described with appropriate drugs (20 μg/ml neomycin, 2.5 μg/ml phleomycin, 25 μg/ml hygromycin, 20 μg/ml blasticidin).

Figure 3.1 summarises the strains used in this study. BF and PF "wild type" trypanosomes (HNI_VO2 and Amsterdam WT, respectively) were cultured as technical replicates, as well as RYT3 and DsRed (BF and PF, respectively) to serve as biological replicates.

Sample	Strain	Form	RNAi
1	HNI_VO2	BF	
2	HNI_VO2	BF	
3	Amsterdam WT	PF	
4	Amsterdam WT	PF	
5	RYT3 H1	BF	
6	RYT3 H1	BF	
7	RYT3 H1c3	BF	-
8	RYT3 H1c3	BF	+
9	DsRed H1	PF	
10	DsRed H1	PF	
11	DsRed H1_1.1c3	PF	-
12	DsRed H1_1.1c3	PF	+

Figure 3.1: Trypanosome strains used in this study. HNI_VO2 and Amt WT served as "wild type" trypanosomes and were cultured in duplicate. We also included RYT3 and DsRed (BF and PF, respectively) to control for possible strain effects. Also indicated are strains in which H1 RNAi KD were uninduced and induced.

2.3.2) Core particle preparation

Micrococcal nuclease (MNase) digestion of chromatin was performed as described [39]. Per digestion, 5 x 10^7 cells were harvested, washed with phosphate buffered saline (PSG for BFs (PBS, 1% w/v glucose), pH 7.4) and pelleted at 1200 x*g* for 5 min. Cells were resuspended to a density of 2 x 10^8 cells/ml in permeabilization buffer (100 mM KCl, 10 mM TRIS.Cl pH 8.0, 25 mM EDTA, 1 mM DTT) and permeabilized with 400 µM digitonin for 5 min at room temperature with agitation. Cells were harvested at 1200 x*g* for 5 min at 4°C. Hereafter, cells were washed with 1 ml ice cold isotonic buffer (100 mM KCl, 10 mM TRIS.Cl pH 8.0, 10 mM $CaCl_2$, 5% v/v glycerol, 1 mM DTT, 1 mM PMSF, 1µg/ml pepstatin A, 1µg/ml aprotinin, 1µg/ml leupepsin) and pelleted at 1200 x*g* for 5 min at 4°C. The pellets were resuspended in MNase digestion buffer (10 mM TRIS.Cl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 3 mM $CaCl_2$, 0.25 M sucrose and 100 µM PMSF) to a density of 5 x 10^7 cells per 100 µl. Chromatin was digested with an increasing concentration of MNase (Worthington units; ranging from 0.5 to 32U) for 5 min at 37°C. MNase digestion was quenched by addition of EDTA and EGTA to a final concentration of 10 mM each. Cells were harvested at 4°C at 10, 000 x*g* for 10 min and resuspended in 200 µl resuspension buffer (10 mM TRIS.Cl pH 7.4, 200 mM NaCl, 3 mM $MgCl_2$, 3 mM $CaCl_2$, 5 mM EDTA, 5 mM EGTA, 0.05% NP-40). Samples were centrifuged for 10 min at 10, 000 x*g* at 4°C and the supernatant (containing solubilised chromatin) saved. SDS and proteinase K were added to the supernatant to a final concentration of 0.5% v/v and 0.2 µg/µl, respectively, and incubated at 56°C for 60 min. After protein digestion, RNase A was added to a final concentration of 0.2 µg/µl and incubated at 37°C for 60 min. Following RNA digestion, DNA was extracted using phenol:chloroform:isoamylalcohol (25:24:1 v/v). DNA was then ethanol precipitated as follows: The volume of

aforementioned eluent was measured, and, in this order, 1/10 vol. 3M NaAc (pH 5.2), 2 µl of 20 mg/ml glycogen (as nucleation particle), and 3x vol. ice-cold 98% (v/v) EtOH added sequentially, vortexed, and the DNA precipitated at -20°C overnight. The DNA was pelleted in a bench top micro centrifuge at 4°C for 10 min at maximum velocity, washed with 70% (v/v) room temperature ethanol, and air dried. The DNA pellet was resuspended in 15 µl TE buffer (10 mM TRIS.Cl pH 8.0, 0.1 mM EDTA).

2.3.3) Mononucleosomal DNA isolation

A 2.0% (w/v) agarose gel, containing ethidium bromide (0.1µg/ml), was used to separate mononucleosomal DNA fragments of 147 bp from total precipitated DNA (containing higher order nucleosomal bands corresponding to di-, tri- and tetranucleosomes) as seen in figure 4.1. The bands corresponding to the 147 bp core particle were excised and DNA purified using the freeze/squeeze method (BioRad) according to manufacturer's instructions. DNA was thereafter ethanol precipitated as previously described. DNA was resuspended in 30 µl TE buffer (10 mM TRIS.Cl pH 8.0, 0.1 mM EDTA) and stored at – 80°C.

2.3.4) Histone H1 analysis

To analyse the depletion of Histone H1, 5 x 10^6 trypanosomes were harvested at 1200 x*g* for 7 min and washed twice with PSG buffer. Cells were pelleted at 3000 x*g* for 5 min and the supernatant removed. Pre-heated Laemmli sample buffer was added to each sample to yield a final concentration of 1 x 10^5 cell equivalents per microliter, i.e. 50 µl sample buffer per sample. Samples were boiled for 5 min at 100°C and vortexed at high speed for 10 sec. For Western blot analysis, 10 µl of this whole cell lysate (1 x 10^6 cell equivalents) was loaded onto a 16% Laemmli SDS-PAGE gel and separated electrophoretically at 100 V

for 90 min. Proteins were transferred from the SDS-PAGE gel to a nitrocellulose membrane at 100 V for 75 min.

Knock-down of Histone H1 was analysed by probing with an anti-histone H1 antibody (a kind gift from G. Rudenko), and developed using anti-rabbit IgG (Invitrogen) conjugated to horseradish peroxidase using standard protocols (Figure 4.3 top panel). Westerns were stripped and re-probed with anti-BiP, as previously described [118], used as loading control (Figure 4.3 bottom panel).

2.3.5) Nucleosome repeat length determination

Nucleosome repeat lengths (NRLs) were calculated by firstly estimating the molecular weight (bp) for each nucleosome number (mono-, di-, tri- and tetranucleosomes) using Quantity One software (Bio-Rad) from agarose gel photographs. Mono-, di-, tri- and tetranucleosomes were selected and fragment size estimated by defining the band size for the molecular markers used (Figure 3.2 A for PF, and B for BF).

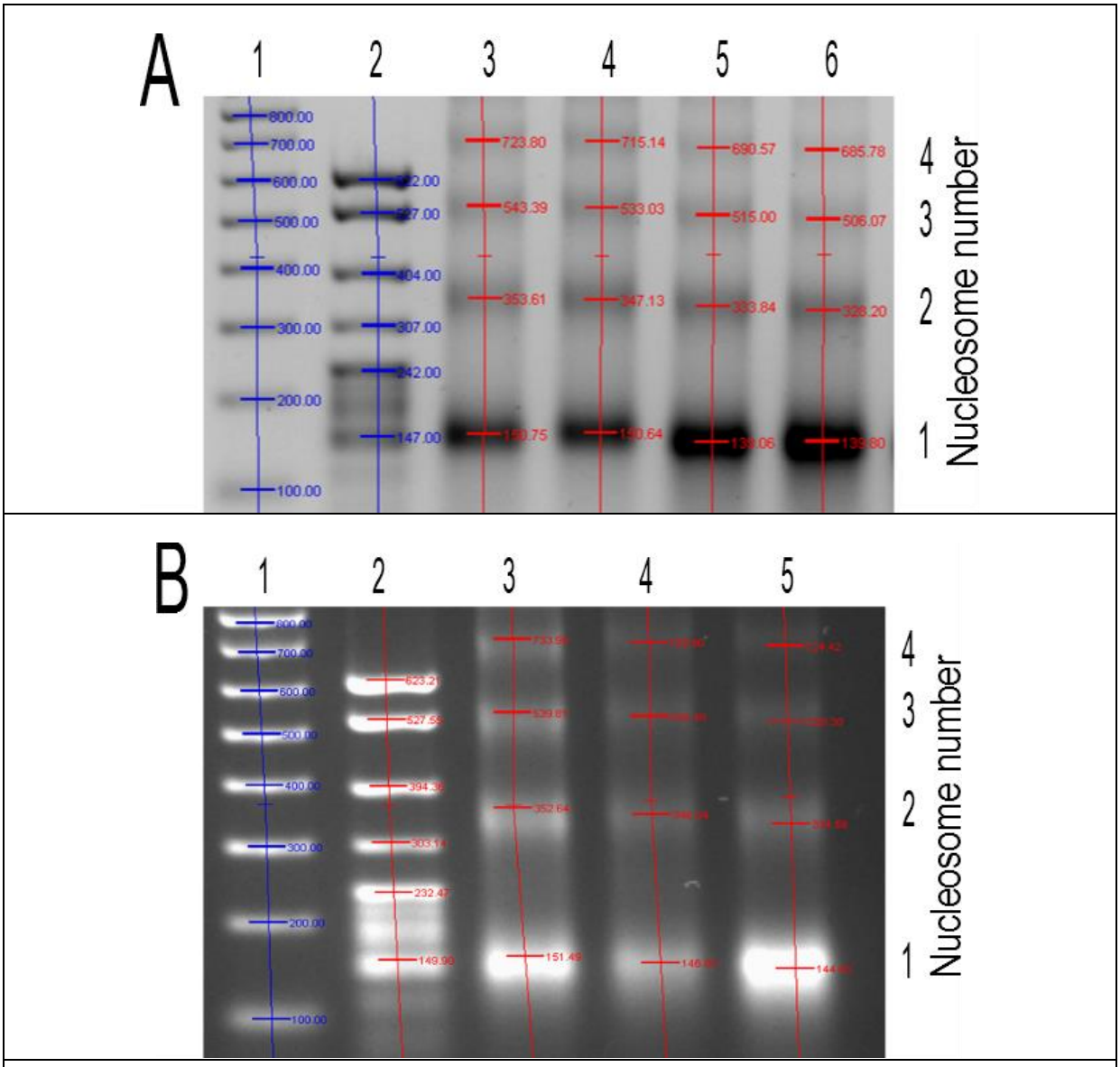


Figure 3.2: Nucleosomal fragment size estimation using Quantity One. Fragments size was estimated by firstly defining marker fragment sizes and applying the standard curve obtained to the nucleosomal fragments. Panel A = PF, B = BF, Lane 1 = 100 bp PCR marker, 2 = pBR322 MspI (in both panel A and B). Panel A lanes 3 – 6 = 1, 2, 4, and 8U MNase. Panel B lanes 3 – 5 = 4, 8, and 16U MNase. Blue shading represents molecular weight standards (in bp) and red shading the molecular weight of the bands as calculated by the software.

The size of each nucleosomal band was plotted, and a line fitted by least squares regression. The slope of this line corresponded to the apparent NRL for that particular sample (lane on the gel, corresponding to a specific [MNase]) (Figure 3.3 A for PF, B for BF).

For PF NRL determination, the chromatin was digested with 1, 2, 4, and 8U MNase. Linear regression produced slopes of 190.9, 188.1, 183.57 and 181.84, respectively. For the BF sample, chromatin was digested with 4, 8 and 16U MNase, representing slopes of 193.3, 193.34 and 191.38, respectively.

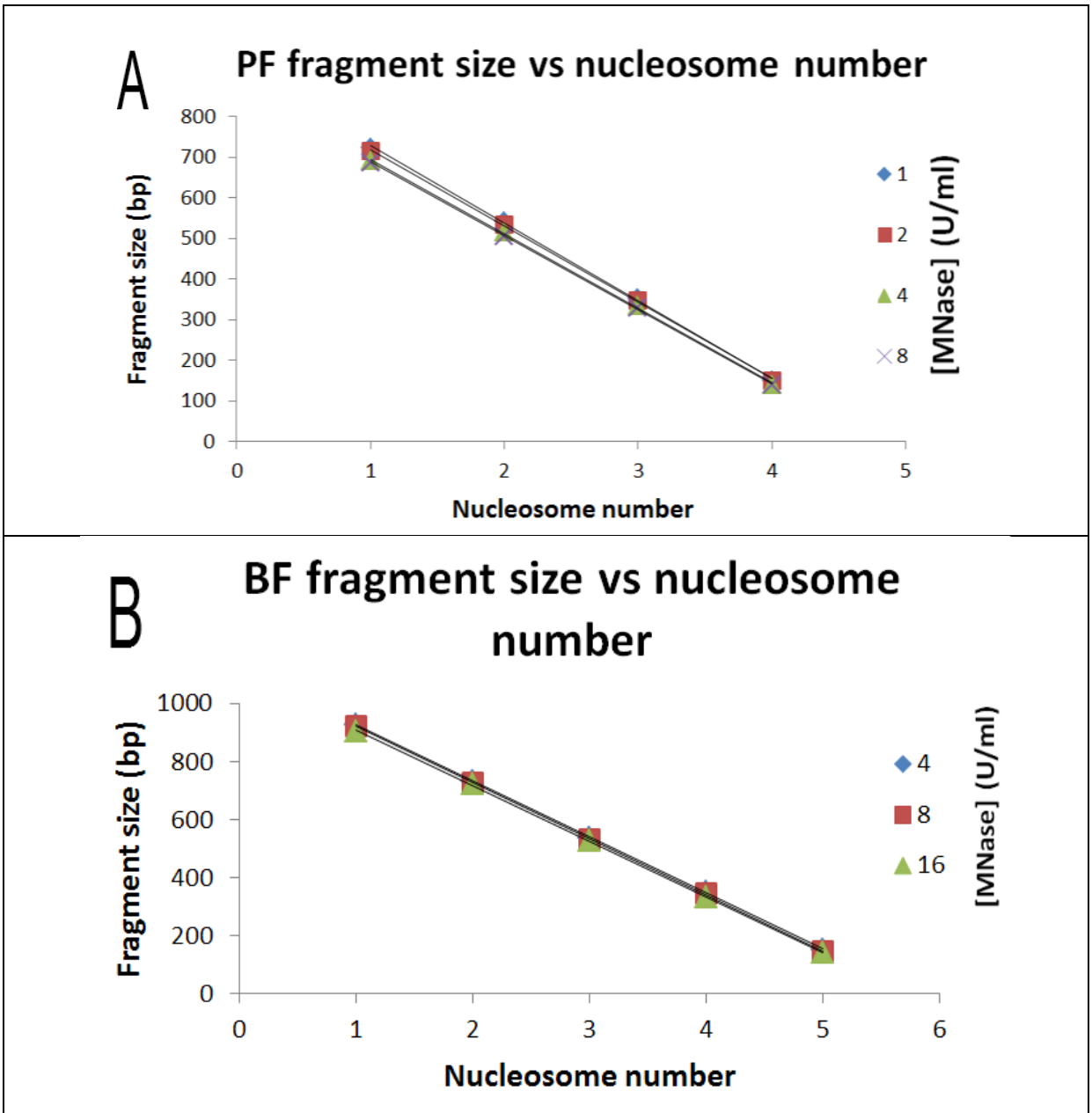


Figure 3.3: Plot of the nucleosomal fragment size (bp) for each MNase concentration used for digestion. Standard curve fitted by least squares regression. Gradients obtained from curve fitting represents the NRL for each digestion.

Since increasing MNase concentrations and more extensive digestion of chromatin can result in the sliding of nucleosomes on the DNA, apparent NRLs measured at high concentrations of MNase can be artefactually shortened due to the shifting of nucleosomes closer together. To correct for this possible discrepancy, the slopes obtained for the different MNase concentrations was plotted against the [MNase], and a line fitted by linear regression. This line was then extrapolated to 0U MNase to correct for nucleosome sliding at high MNase concentrations (Figure 3.4 A and B), and represent the real NRL.

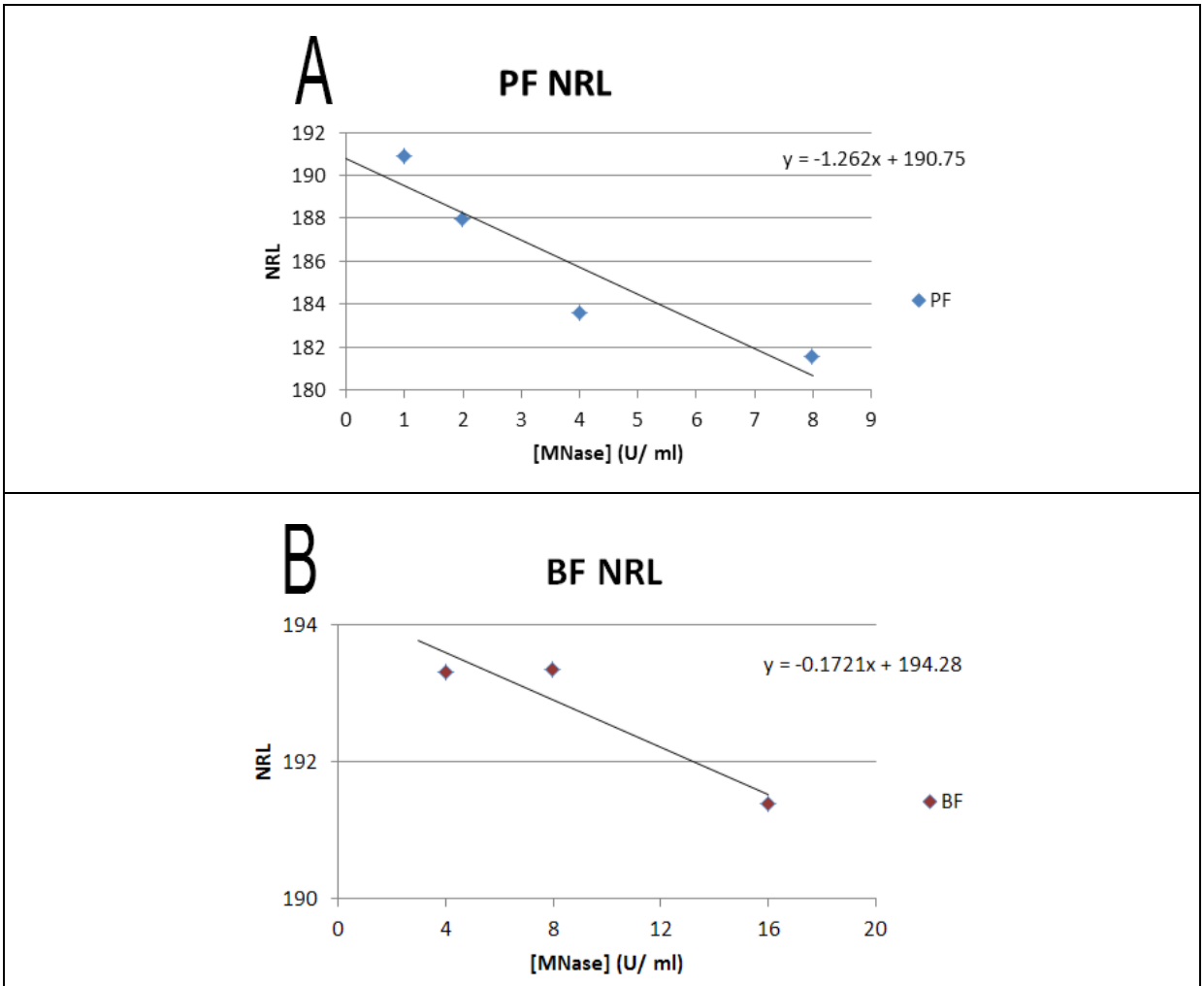


Figure 3.4: NRL estimation of PF (A) and BF (B) trypanosomes extrapolated to 0U MNase. Y-intercepts indicate the NRLs of 190.75 bp (PF) and 194.28 (BF).

2.3.6) Paired-end sequencing

Following DNA isolation, samples were sequenced using the Illumina platform. Samples were repaired using a NEB PreCR Repair Mix (M0309) and analysed using a Bioanalyzer 2100. All samples were then prepared for paired-end sequencing by ligation of mononucleosomal DNA fragments to Illumina paired-end adapter as well as index primers, as provided by Illumina (Figure 3.5). A 30 μl mixture, containing 0.42 ng/μl of each sample was used in one flow cell for sequencing purposes. All reactions were carried out to manufacturer's instructions.

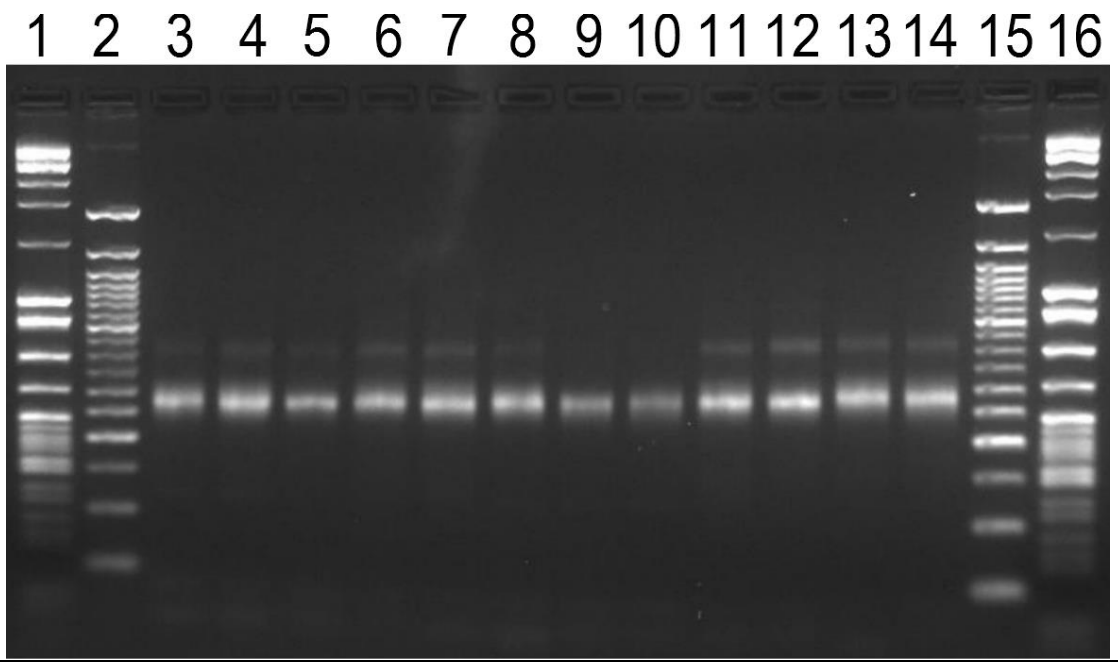


Figure 3.5: Adapter ligation to mononucleosomal fragments yields a band at ~270 bp, indicative of successful primer-fragment ligation. Weak bands at ~ 420 bp could indicate ligation of two core particles, which was not sequenced. Lanes 1 and 16 = 1 kb marker (NEB) + pBR 322 MspI (NEB), lanes 2 and 15 = 50 bp ladder, lanes 3 – 14 = samples 1 – 12.

In silico

All programs developed in the laboratory for this study were coded in standard C++ (ISO/IEC 14882:2011) using the 64 bit Gnu compiler collection (GCC) version 4.6 (http://gcc.gnu.org/), and Eclipse code editor version 4.3.0.I20130605-2000 (http://www.eclipse.org/). Code were compiled on a 64 bit Windows version 8.1 platform (http://www.microsoft.com/) using a Mingw toolchain (http://mingw-w64.sourceforge.net/).

2.3.7) Alignment using Bowtie

Illumina sequencing files were recovered in FASTQ format. A *T. brucei* 427 Lister reference genome was constructed from the version 4.2 genome downloaded from TriTrypDB (tritrypdb.org). This genome was used to construct a genome searchable with *bowtie2* version 1 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), using the *bowtie-build* program, which is part of the *bowtie2* program suite [116]. Bowtie 2 was used to map the short sequence fragments in the Illumina FASTQ file to the *T. brucei* 427 Lister genome. Commands used to perform alignment are as follows:

```
bowtie2 --no-hd -I 145 -X 155 -x /share/Genomes/T_brucei/427 -1
sequence_file1.fastq -2 sequence_file2.fastq -S bowtie_sam_output.txt
```

Where:

--no-hd: no headers

-I: minimum concordant fragment size of 145 bp

-X: maximum concordant fragment size of 155 bp

-x: indexed genome data (built with bowtie-build.exe)

-1: primer 1 sequence file (files in FASTQ format)

-2: primer 2 sequence file (files in FASTQ format)

-S: SAM output file (saves SAM fields with each aligned sequence).

This command was repeated for each set of files for each experimental condition.

After bowtie sequence alignment, concordant primer pairs were extracted from the alignment SAM files data using *Nucfrag,* a C++ program developed in this laboratory. Commands used for concordant sequence extraction are as follows:

```
-d -f -s -w 7 -i 145 -x 155 -s bowtie_sam_input.txt -o output_file.txt
```

Where:

-d: generates a dyad file

-f: generates a digital footprint file

-s: generates a smoothed digital footprint file

-w: running window size of 7 bp.

-i: minimum concordant fragment size of 145 bp

-x: maximum concordant fragment size of 155 bp

-o: Text output file

In this study, only data files representing the number of dyads at each nucleotide position was used.

2.3.8) Visual inspection of data

Data file outputs from *Nucfrag* were converted to wiggle format and uploaded and visualized as custom tracks using the TriTrypdb genome browser (tritrypdb.org). Samples were inspected in batches of 4 tracks at a time which included two BF wild type samples and two PF wild type samples. Each chromosome of each sample was visually inspected and regions of interest noted.

2.3.9) Data normalization and Bin analysis

To allow comparison of nucleosome positioning on different chromosomes within a sample of a single experimental condition, as well as between experimental conditions, it is necessary to normalise the dyad data files. We assumed that the average nucleosome density on all chromosomes, as well as between experimental conditions, would be identical. This seems a fair assumption, since the NRL did not appear to change much for different experimental conditions, and, similarly, nucleosomal density was not observed to decrease with H1 knock-down or deletion in *S. cerevisiae* [119]. Therefore, the value of the number of dyads present at each location was multiplied by a factor representing the ratio between the largest chromosome (chromosome 11, ~ 5 Mb) and the chromosome in question. This adjustment was followed by normalising the total number of dyads present on all chromosomes of an experimental condition to be identical to that of all other experimental conditions.

We observed that the tandem repeat sequences in the genome appeared to be associated with very high levels of nucleosomes. Although this can reflect a real biological phenomenon, we argue that this high apparent occupancy is due to the presence of the identical tandem repeat sequences on the intermediate and minichromosomes, which do not constitute part of the available *T. brucei* strain 427 genome sequence. Thus, following Illumina sequencing of the nucleosome core particle fragments, sequences were generated from the intermediate and minichromosomes. These sequences were subsequently aligned by bowtie to the megabase chromosome genome sequence. Thus, the alignment of sequences to megabase genome tandem repeats would over-represent these alignments, resulting in an artefactually

high nucleosomal occupation. For this reason we decided to exclude the tandem repeats from the analysis. The position of identified tandem repeat sequences were downloaded from the TriTrypDB (tritrypdb.org) as a GFF format text file. The program *strip_tandem_repeats* was used to insert a "-1" at every position in the dyad data file corresponding to a nucleotide in any tandem repeat. The dyad file from which the tandem repeat data had been stripped were written to disk, and used in all subsequent analyses, except where noted. The "-1" entry was recognised by subsequent analysis programs as a tandem repeat nucleosome position, and appropriately processed where averages were calculated.

The binning analysis was performed with the program *dyads_on_genomic_features*. This program reads a text file of dyad data file paths, and reads in the dyad file data. It reads a text file of genomic features listed in the format "Tb427_1_v4: 60182 - 60254 (+)", referring the specific chromosomes and genomic regions as well as the Watson or Crick strand. The program produces a text file output of the number of dyad axes in bins ranging from 0 to that of the highest number of dyad axes identified on one or more nucleotides in the analysed data. A listing of the source code of *dyads_on_genomic_features* is provided in the supplementary material (S 2.1). Gnuplot version 4.6 (http://www.gnuplot.info/) was used to construct plots from the bin data files.

2.3.10) Avarage dyad distribution analysis

The cumulative dyad distributions of aligned genomic features such as the region centred on the translation start codon of all ORFs were generated with the program *align_dyads*. This program reads the dyad data files from disk and a list of genomic positions, in the format "chromosome number, genomic position, strand". The dyads from every nucleotide within a defined window centred on the genomic position of every entry in the file of genomic positions was calculated, and written to disk. Plots of the data files were constructed with gnuplot. A listing of the source code of *align_dyads* is provided in the supplementary material (S 2.2). Analyses were typically performed for window sizes of 500 or 1000 bp upstream and downstream of the genomic positions. The genomic positions of the translation start of the first ORF as well as all ORFs in every PTU were retrieved from TriTrypDB. The positions of the STOP codon of every ORF, as well as *tRNA* and *rRNA* gene positions were similarly retrieved. The positions of splice acceptor sites (SAS, henceforth referred to as the 5'-untranslated region (UTR)) and polyadenylation sites (PAS) were recovered from the data of George Cross [120]. The most upstream 5' UTR identified for divergent SSRs and the most downstream PAS detected for convergent SSRs were manually acquired from TriTrypDB (tritrypdb.org). Only curated entries with high alignment scores and no ambiguity was used for average dyad distribution analysis.

2.3.11) Sequence motif analysis

Base frequency and AT skew analysis

The frequency of each nucleotide on each chromosome was determined for version 4.2 of the *T. brucei* 427 Lister genome using the program, *AT_enrichment*. The results are shown in figure 3.6. The A+T, A and T richness of either the Watson or Crick strand at specific genomic loci were determined with the program *AT_enrichment*, using appropriate settings. The program reads in a list of FastA format sequences to be analysed. We retrieved sequences of interest from TriTrypDB using the "Sequence Retrieval" tool. The *AT_enrichment* program calculates the ratio of A (or another selected nucleotide) to all nucleotides in a sliding 10 nt window and expresses that as the $\log_2$ ratio to the average genomic frequency of that nucleotide, indicated in figure 3.6. This gives an indication of the fold enrichment or depletion of the selected nucleotide in the window relative to the genomic average. The data was plotted using gnuplot. A listing of the program *AT_enrichment* is provided in the supplementary material (S 2.3).

Chromosome	A	T	G	C	Total
1	264770	267455	230265	218944	981434
2	312177	315948	249860	261998	1139983
3	434193	422572	375190	377554	1609509
4	414923	404190	346261	377616	1542990
5	388645	418397	355307	348804	1511153
6	404149	419783	346004	352221	1522157
7	570901	573553	500815	524464	2169733
8	633685	630473	582256	571098	2417512
9	787976	771613	652733	665405	2877727
10	1046364	1044977	967338	961101	4019780
11	1250235	1241506	1140995	1130712	4763448
Total	**6508018**	**6510467**	**5747024**	**5789917**	**24555426**
%	**26,50338**	**26,51335**	**23,40429**	**23,57897**	**100**

Figure 3.6. The occurrence of each nucleotide in the MBC of *T. brucei*.

Poly (A/T) runs

The presence of continuous runs of A or T in the genome sequence was analysed with the program *polyA_enrich*. This program reads a file of aligned sequences in FastA format, retrieved from the TriTrypDB site with the sequence retrieval tool, and adds up the number of occurrences of A (or T) runs of defined length initiating at each position in the aligned sequences. The data is written to disk as a list of values representing the fractional occurrence at each aligned nucleotide position. A listing of *polyA_enrich* is provided in the supplementary material (S 2.4).

Sequence Motif detection

The presence of sequence motifs were analysed using Meme version 4.9.1 (http://meme.nbcr.net/meme/cgi-bin/meme.cgi). Sequences for analysis were retrieved from TriTrypDB in FastA format, and uploaded to the online version of Meme. Default settings of number of motifs=3, minimum width=6 bp, maximum width=50 bp, minimum sites=2 and maximum sites=100 were used. To gain insight into the possible statistical significance of identified motifs, a search was repeated with the same settings, but with the randomly shuffled sequences.

Dinucleotide distribution

The dinucleotide distribution was calculated with the program *dinucleotide_distribution*. The program reads a text file of sequences in FastA format. The occurrence program then determines the number of occurrences of each of the sixteen possible dinucleotides at each position in the aligned sequences. This occurrence is smoothed with a 10 bp running average, and written to a data file. Plots of the dinucleotide distribution were generated

with gnuplot. A listing of the program *dinucleotide_distribution* is provided in the supplementary material (S 2.5).

Fast Fourier transform

The periodicity of select dinucleotide pairs in aligned nucleosomal sequences of identical length was studied by Fourier transformation. Fast Fourier transforms were performed by an online FFT tool (http://www.sooeet.com/math/fast-fourier-calculator.php). The data of the dinucleotide distribution of the central 128 bp of the data file generated by *dinucleotide_distribution* was copied to the input window of the FFT tool and the Fourier transform calculated at a sampling rate of 1 bp. The calculated transform was used in a Fourier synthesis, and confirmed that the original dinucleotide distribution was re-generated. The square of the imaginary component of the Fourier transform (known as the Fourier amplitude) was saved to disk, and plotted as impulses using gnuplot.

2.4) Results

In vitro

2.4.1) Core particle preparation

Micrococcal nuclease is a non-specific endo- exonuclease capable of digesting double stranded DNA. It also exhibits significant single strand exonuclease activity. The wrapping of DNA around the histone octamer effectively shields nucleosome-associated DNA from digestion by MNase. Therefore, MNase will digest the linker DNA, leaving the nucleosome-bound DNA intact [104].

Chromatin was digested with increasing concentrations of MNase to allow estimation of the optimal MNase concentration for digestion of chromatin to mononucleosomal "core particles". These core particles consist of 147 bp of DNA wrapped around a histone octamer. Treatment of solubilized core particles with Proteinase K and RNase A digests the histone proteins and any residual RNA molecules, which were then removed by phenol: chloroform extraction. The resulting supernatant contained DNA digested to various degrees, i.e. mono-, di-, tri-, tetranucleosomes, etc. These fragments were precipitated and separated on 2% w/v agarose gel (Figure 4.1). Characteristic nucleosomal ladders were observed, representing under-digested higher order chromatin structures. Digestion by 4, 8 and 16U MNase yielded the highest amount of mononucleosomal DNA (Figure 4.1, lanes 5 – 7). Also visible in Figure 4.1 is the presence of a DNA band migrating together with the pentanucleosome band (800 – 900 bp). This band is thought to originate from the kinetoplast (mitochondrial) maxi circle DNA (G. Rudenko, personal communication).

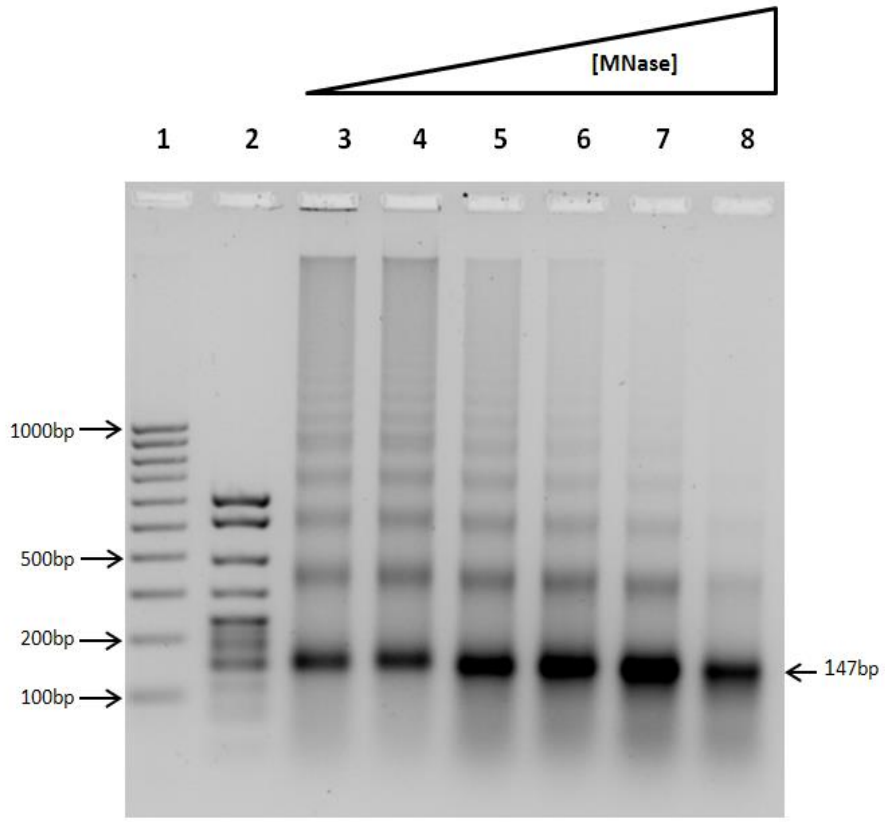


Figure 4.1: MNase digested chromatin from Amsterdam wt PF trypanosomes, visualized in a 2% (w/v) ethidium bromide stained agarose gel. Chromatin was digested with 1 – 32U MNase, with a two-fold increase per lane. The mononucleosomal DNA fragment at ~147 bp is indicated. Optimal digestion of higher order chromatin was achieved at 4, 8 and 16U MNase. Markers are (1) 100 bp PCR marker; (2) pBR322 *Msp*I which yields a band at 147 bp (NEB); (3-8) MNase titration [1U, 2U, 4U, 8U, 16U, 32U].

For sequencing purposes, chromatin of the PF H1 knock-down was digested with a MNase concentration of 0.5 – 2.0 U/ml MNase (supplementary material S 2.6), as it was previously observed that depletion of H1 in *T. brucei* increases the MNase sensitivity of chromatin [74]. Prior to excising the core particle length DNA bands, agarose gels were photographed on a glass backplate to minimise exposure of the DNA to UV (supplementary material S 2.6).

2.4.2) Nucleosome repeat lengths

Nucleosomes tend to be uniformly spaced from one another, located on a stretch of 147 bp of DNA separated from adjacent nucleosomes by linker DNA [109]. This length of DNA (147 bp + linker DNA) is known as the nucleosome repeat length (NRL), and exhibits species and tissue specific variation in length [106]. The NRL for PF and BF were calculated from the MNase digestion profiles as described. Figure 4.2 shows the graphs representing the NRL at different MNase concentrations. A standard curve, fitted by least squares regression, was extrapolated to determine the NRL at 0 U MNase. This corrects for possible sliding of the nucleosomes at higher MNase concentrations, which would result in an artifactual short NRL. Figure 4.2 A represents BF, and 4.2 B PF NRLs.

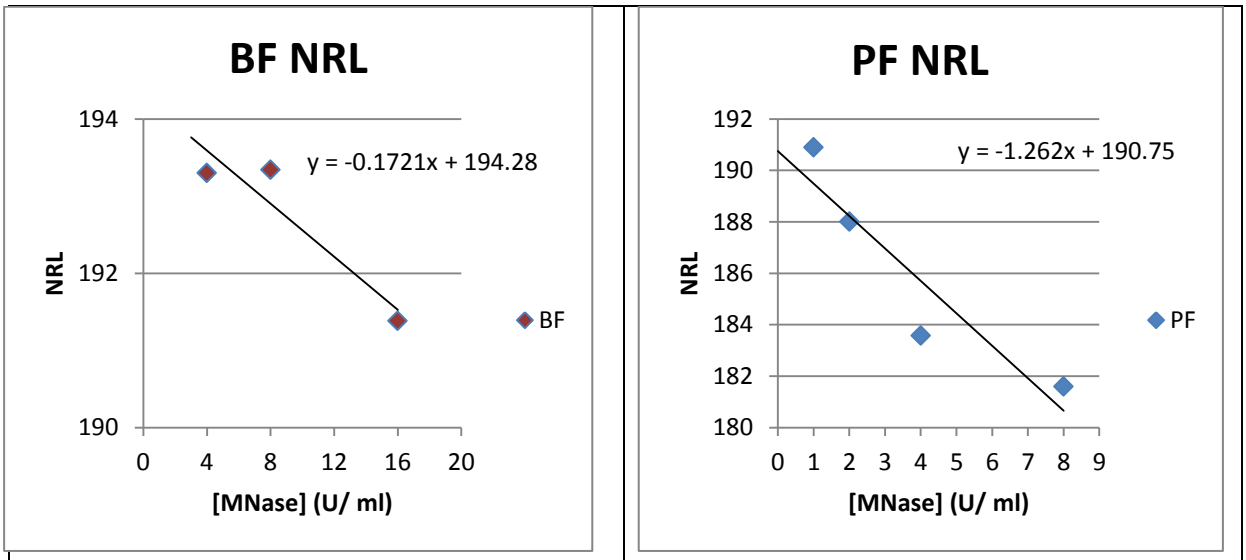


Figure 4.2: Nucleosome repeat length was determination for BF (panel A) and PF (panel B) trypanosomes. Extrapolation of a standard curve fitted by least squares regression indicates a NRL of 194.28 (±0.9 bp) and 190.75 (±4.2 bp) bp DNA for the different life cycles, respectively.

It was determined that the NRL for BF trypanosomes was 194.28 bp (±0.9 bp) and for PF 190.75 bp (±4.2 bp). There was no statistical significant difference in NRL between the different life forms (Student T-test,$p < 0.05$). Therefore, with a NRL of 194, the linker DNA is ~23 nucleotides in length.

2.4.3) Histone H1 knock-down

To study the function of the linker histone in nucleosomal organisation in *T. brucei*, we used tetracycline inducible RNAi-mediated H1 knock-down strains to deplete chromatin of H1 before cell harvesting. Following tetracycline induction, maximal levels of H1 depletion were reached at 48h and 96h for BF and PF trypanosomes, respectively. This was confirmed by western blot analysis using rabbit anti-H1 antibody (Figure 4.3).

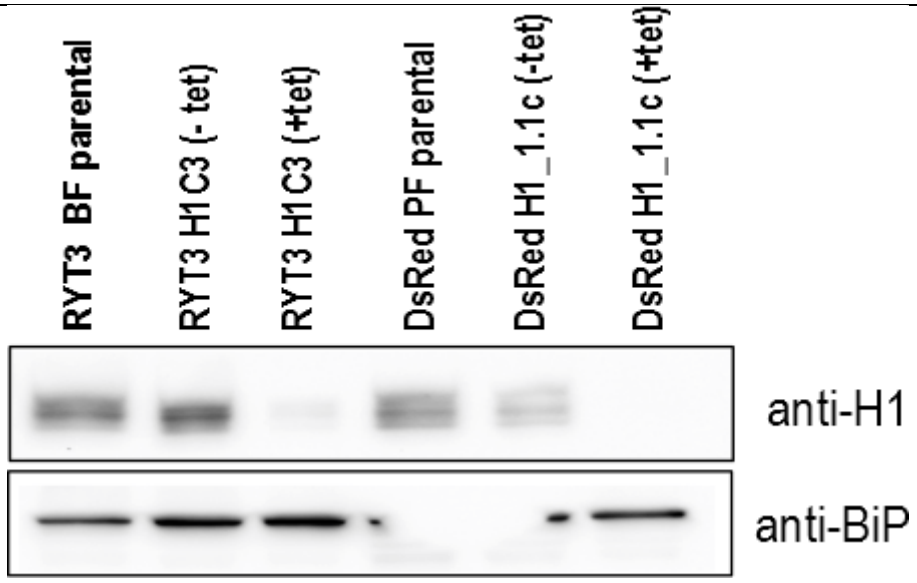


Figure 4.3: Western blotting with anti-H1 shows the depletion of H1 after incubation with tetracycline for 48h for BF and 96h for PF trypanosomes (top panel). Severe depletion H1 was observed in the tetracycline induced PF sample. The bottom panel shows anti-BiP probing as load control. Uninduced RNAi = -tet, induced = +tet.

H1 levels were greatly depleted in both life cycles, with severe depletion observable in the PF. Therefore, MNase digestion of chromatin was done in almost complete absence of H1 in these samples. Uninduced samples also displayed a degree of H1 depletion, which was likely caused by leaky expression of the RNAi construct. However, a meaningful amount of H1 was still present in the chromatin, as seen from figure 4.3.

Also shown in figure 4.3 is the BiP-probing which served as loading control. BiP is a HSP70 molecular chaperone located in the lumen of the endoplasmatic reticulum, and is present in equal amounts in all cells and life forms in *T. brucei* [118]. BiP protein bands of DsRed PF parental and DsRedH1_1.1c uninduced did not completely transfer to the nitrocellulose membrane. However, it was assumed that the intensity of the sides of these bands that did transfer was representative of the whole band, and therefore comparable to the intensities of other BiP bands.

In silico

2.4.4) Sequence alignment

Sequencing of these short 147 bp DNA fragments by Illumina paired-end sequencing yielded short sequences of 50 bp that was mapped back to the *T. brucei* 427 Lister genome. Bowtie 2 was designed to map short sequence fragments to large reference genomes, and also supports gapped paired-end alignment modes [116].

After sequencing bowtie 2 was used to align the short reads to the genome by applying basic nucleosome constraints, i.e. the aligned sequence pairs must be

between 145 – 155 bp apart from end to end. This resulted in an alignment rate of ~70% for all samples (Figure 4.4.). This alignment level was expected since the mitochondrial DNA as well as the intermediate and minichromosome sequences, constituting approximately 30% of the total DNA compliment [4], did not form part of the reference genome

Sample nr:	Total reads	Concordant alignments (1 match)	Concordant alignments (>1 match)	Discordant alignments	No match	Alignment rate (%)
1	11829835	5423632	2163826	98774	4242377	68
2	13828097	6459546	2573654	91225	4794897	69
3	11962093	5606947	2394925	37317	3960221	70
4	12519958	6074342	2587847	43107	3857769	72
5	12878856	6469307	2581145	51115	3828404	73
6	13843890	6786489	2668326	43988	4389075	71
7	13323741	6570607	2458943	73560	4294191	71
8	11422590	5378641	2109908	59914	3934041	69
9	13473485	6464713	2809759	24169	4199013	71
10	12744135	5796994	2532050	118453	4415091	70
11	14172570	6659217	2915285	24302	4598068	70
12	12480529	5840765	2489286	25762	4150478	69
Average	12873315	6127600	2523746	57640	4221968	70.25

Figure 4.4: Summary of sequence alignments by Bowtie 2. Sequencing produced an average of 12.8 million reads per sample, of which ~ 70% was realigned to the *T. brucei* 427 Lister genome.

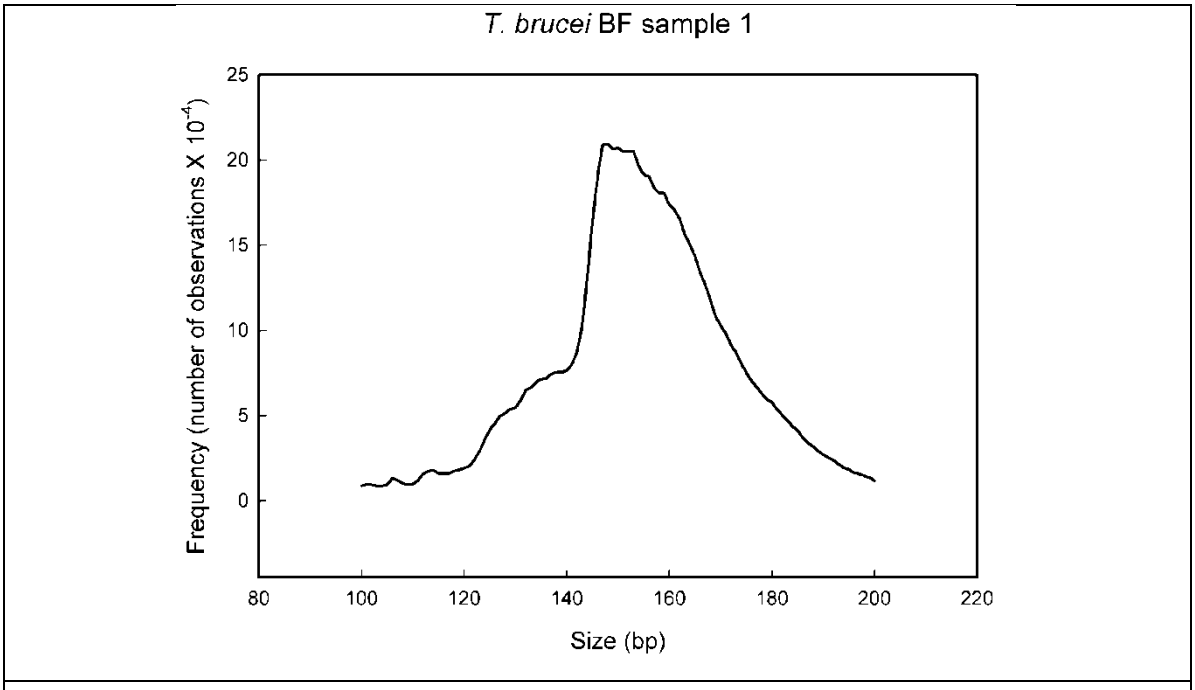


Figure 4.5: The size distribution of concordant, single match alignments from which alignments of 145 – 155 bp were extracted by Nucfrag. This graph is representative of all samples.

The first anticipated artefact resulting from the bowtie 2 alignment was the alignment of non-concordant sequences. Therefore, concordant aligned sequence pairs of correct length were extracted using an in-house developed program, NucFrag, again applying mononucleosomal constraints of 145 – 155 bp (Figure 4.5). This program generated files that contained the number of dyads (centre of the identified nucleosomal fragments), for every position in each chromosome. Although MNase is routinely used to generate core particles, the enzymatic activity does possess some sequence bias, leading to imprecise trimming of core nucleosomal DNA resulting in over or under trimmed or internally nicked DNA strands. Data filtering using nucleosome constraints and concordant alignment extraction would, in part, remove this bias. The Nucfrag output files, containing the concordantly aligned nucleosomal positions, were then visually inspected.

2.4.5) Visual inspection

Despite generating the nucleosomal positions using strict constraints, alignment artefacts could still be present and included in the nucleosomal occupancy maps. Each chromosome of each sample was therefore visually inspected using the TriTrypDB genome browser and regions of interest were noted.

Centromeres and Tandem repeats

The centromeric repeat arrays for chromosomes 1 to 8 of *T. brucei* have recently been identified [102]. These 177 bp AT-rich arrays were found to vary from 20 to 100 kb, with the size of each array roughly proportional to the chromosome size. This is in contrast with the centromere size in *S. pombe* which shows an inverse proportionality [121].

The *T. brucei* genome contains a multitude of simple sequence AT repeats. Almost half of the IC and MC are composed of 177 bp AT-repeats [11] and the MBCs are richly interspersed with these repeats (Figure 4.6). An interesting feature of these repetitive sequences is the positions of the repeats relative to functional regions on the MBC. Figure 4.6 shows the overlap of tandem repeats between divergent SSRs (~0.52 Mb and ~0.95 Mb). Repeats were also visible at genes transcribed by pol III (*tRNAs* at ~0.33 Mb).

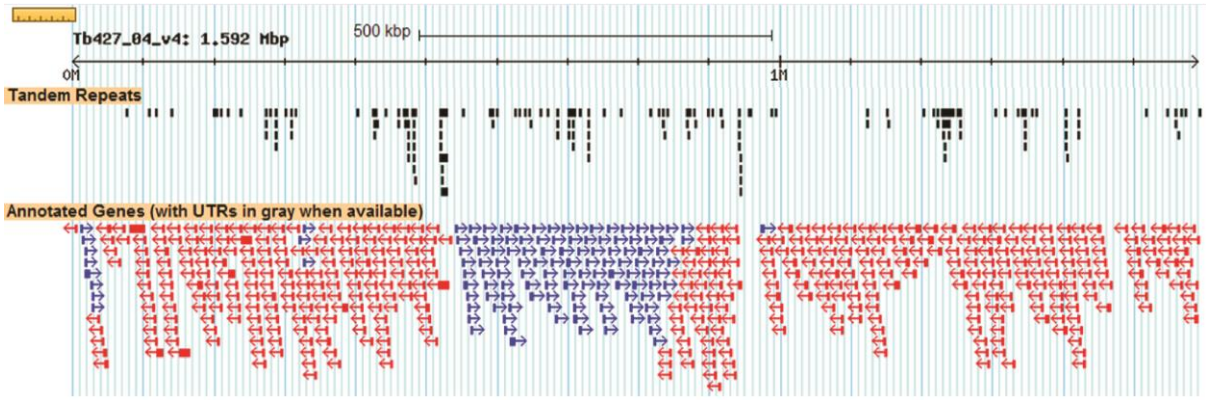


Figure 4.6: Tandem repeats interspersed throughout the MBCs. Shown is the tandem repeats (black boxes) as well as annotated genes (blue = forward strand, red = reverse strand) present in MBC 4. This image is representative of all chromosomes.

These repeats posed a unique challenge in the generation of nucleosome maps. When bowtie 2 finds that a sequence can be perfectly aligned at more than one site, it will randomly assign the alignment to one of the possibilities. Thus, if there were tandem repeat sequences on the intermediate and minichromosomes, these sequences would be assigned randomly to matching tandem repeats on the MBC, thus artefactually inflating the number of nucleosome dyads generates close to tandem repeats. This phenomenon is shown in Figure 4.7. Some of these regions of well positioned nucleosomes correlated with the centromeric positions identified on chromosomes 1 to 8 by Echeverry and colleagues [102]. However, most of these alignments were artefactual due to the repetitive nature of these sequences and their presence in the *T. brucei* genome.

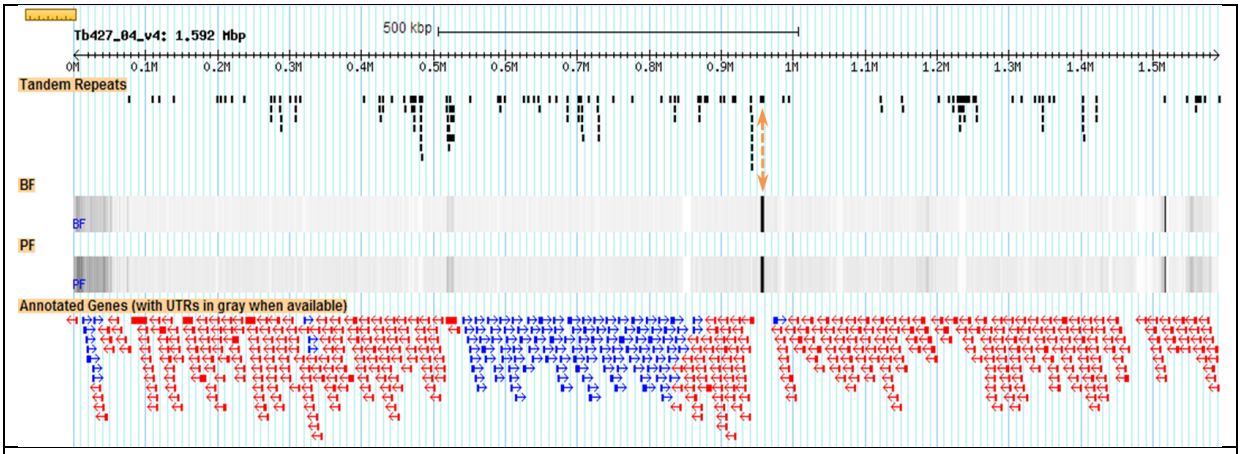


Figure 4.7: BF and PF nucleosomal density map for *T. brucei* MBC 4. The intensity of horizontal bands (in the BF and PF tracks) indicates the occupancy and phasing of nucleosomes at these regions. The orange double arrow shows the centromeric repeats and the high levels of well positioned nucleosomes at this repeat.

Figure 4.7 also shows nucleosomal densities fluctuating throughout chromosome 4. The dense nucleosomal occupancy at 0 – 0.7 Mb includes DNA that contains numerous retrotransposon (RHS), leucine-rich repeat (LRRP), and other sequences associated with "subtelomeric" regions (discussed below). The dense nucleosomal occupancy at ~1.50 – 1.59 Mb correlates with silent "*VSG* arrays" (discussed below). Similar patterns were observed on other chromosomes, indicating that functionally different regions were occupied by different nucleosomal densities.

Another interesting feature seen upon visual inspection was the overall distribution of nucleosomes across pol II PTUs. As seen in figure 2.2, nucleosomes are arranged in a particular fashion around genes in eukaryotes, where the upstream nucleosomes are highly phased and well positioned. However, the phasing decreases and becomes fuzzy towards the 3' end of the gene.

If one would imagine a pol II PTU of *T brucei* to be functionally similar to a single gene in *S. cerevisiae,* one would imagine the same setting as seen in figure 2.2, with well positioned 5' nucleosomes that exhibit a phasing that diminishes towards the 3' end of the gene. However, this was not the case. Well positioned nucleosomes were observed throughout the PTU, with no weakening of phasing detected. Indeed, nucleosomes seemed to be deposited at preferential positions relative to genes and genomic regions, not only between biological replicates, but also, more strikingly, between life cycles.

2.4.6) Data normalization and Bin analysis

As the protocol used for core particle preparation did not yield the same total amount of DNA per sample, the amount of total reads observed during sequencing also fluctuated, as seen from figure 4.4. This complicated possible comparison between samples of the nucleosomal positioning data.

Thus, the number of mapped reads per base pair varied from sample to sample. To correct for this "total reads discrepancy" and to allow comparison of differential nucleosomal densities between samples, the total reads per sample were normalized relative to one another, yielding data files where the ratio of dyads to nucleotide was identical for all chromosomes and between samples. An implicit assumption of this normalisation approach is that there is no striking difference in nucleosome densities between chromosomes or between samples. We could not find any study that contradicted this assumption.

Bin analysis

The *T. brucei* genome can be demarcated into functional regions depending on the contents or "features" of each region. These features are grouped based on regional (eg. divergent or convergent SSR), transcriptional (eg. pol I, II, or III transcribed genes) or sequence (genic, intergenic, repeats) characteristics.

Because of the observed differences of nucleosomal occupancy across the MBCs, bin analysis was used to explore the compaction and occupation of these genomic regions by nucleosomes. All nucleotides of the MBCs associated with a single dyad were summed, and assigned to the bin of value 1. Similarly, all nucleotides that were associated with two dyads were summed, and placed in the bin of value two. This process was repeated up to the maximum dyad value observed in the region under analysis. However, when performing a bin analysis of a region, it is important to consider the inherent length, in bp, of a region or features. This will have an impact on both intra- (eg. PTUs ranging from 10 – 100 kbp) and inter- (eg. *tRNAs, rRNAs*, tandem repeats) feature comparison. Therefore, the binned data was further normalized to the number of nucleotides in the analysed region to allow comparison of binned data between regions or features of different length. This was done by expressing the bin values as a fraction of the total bp length of each of these features (total bp of ORFs bin value x bin value / total bp of feature). As the genic (pol II PTUs) features are by far the largest in size (bp), all binned values were normalized to this feature. Therefore, for genes like *tRNAs*, which are considerably shorter than pol II genic features, the binned values were adjusted according to the ratio (bp *tRNA*: bp genic). Independent analysis of PF vs BF samples showed no significant difference in binning analysis and results presented includes both PF and BF life cycles.

Figure 4.8 demonstrates the effect of normalization on the plot data of *tRNA* genes. The number of dyads grouped in bin sizes of 1 to 10 is shown for unnormalized (solid circles) and normalized data (unfilled circles). Using the normalization rationale explained above, a normalization factor of 2741 was used for *tRNA* genes.

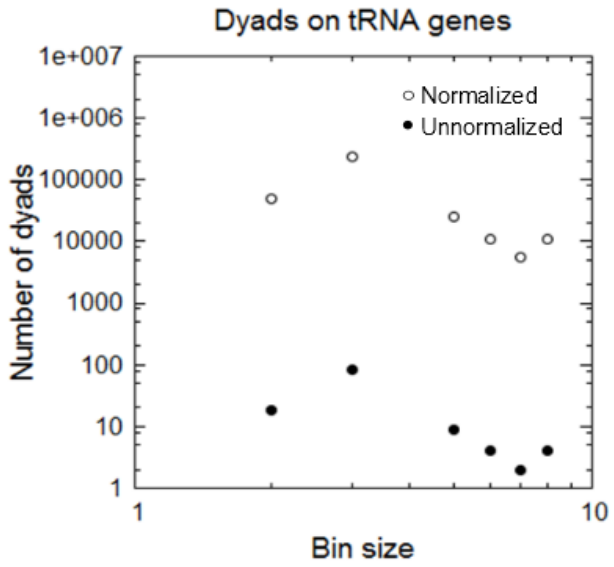


Figure 4.8: Effect of data normalization on *tRNA* genes. Normalisation relative to the ORF data was carried out by correcting for the difference in the number of nucleotides that are present in ORFs compared to the number of nucleotides present in *tRNA* genes, resulting in a normalization factor of 2741 for *tRNAs*. Unnormalized data = solid circles, normalized data = unfilled circles.

Bin analysis of Tandem repeats

As we observed exceptionally high densities of nucleosomes mapped to TRs, we first sought to analyse the distribution of dyads on tandem repeats relative to that of genic pol II transcribed ORFs.

Bin analysis of nucleosomes mapped to tandem repeats confirmed the presence of exceptionally high amounts of well positioned nucleosomes. Figure 4.9 shows nucleosomal distribution on tandem repeats and ORFs as a reference. These repeats showed a much wider distribution compared to ORFs, with a less negative overall slope, and clearly showed large amount of nucleosomes associated with these sequences. The positive shift observed on the Y-axis for the TRs data was due to the normalization for total length, compared to the ORFs.

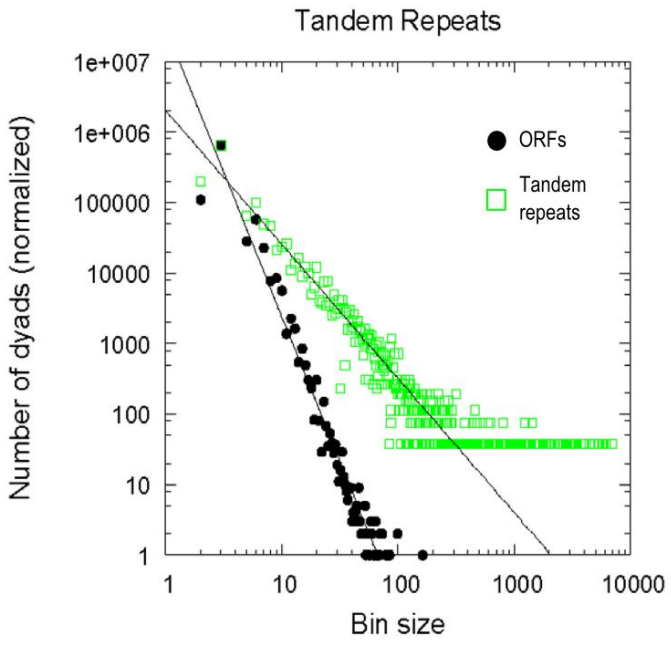


Figure 4.9: Bin analysis of nucleosomal distribution on Tandem Repeats. Bin analysis confirmed that tandem repeats displayed high level of well positioned nucleosomes. ORF bin analysis included as reference. ORFs = solid circles, green squares = Tandem Repeats

Based on this data it was decided to omit data obtained from these tandem repeats, as it was clearly an artefact arising from the repetitive sequences in the *T. brucei* genome which has been realignment by bowtie 2. These repeats may, however, function in genome regulation between life forms or as chromosomal centromeric repeats. These AT rich repeats might also serve to selectively exclude nucleosomes from specific regions and allow access to re-initiating polymerases, transcription factors or other DNA binding proteins.

We anticipated that the exclusion of the nucleosomal dyads mapped to TRs would influence the overall distribution of both genic and intergenic dyads. Therefore, to explore the effect on dyad distribution over these regions in the presence and absence of TR-mapped dyads, we examined the binning analysis of these regions under these defined circumstances. Figure 4.10 shows the effect of TR removal on the binning analysis of genic ORFs (4.10 A) and intergenic regions (4.10 B). A modest difference was observed over both regions after TR-dyad subtraction, with the effect being a bit more pronounced over intergenic regions, where most TRs are located. Note the presence of an extensive number of high value bins for intergenic regions when the TRs data is part of the analysis set. This could indicate that TRs are more enriched on intergenic regions.

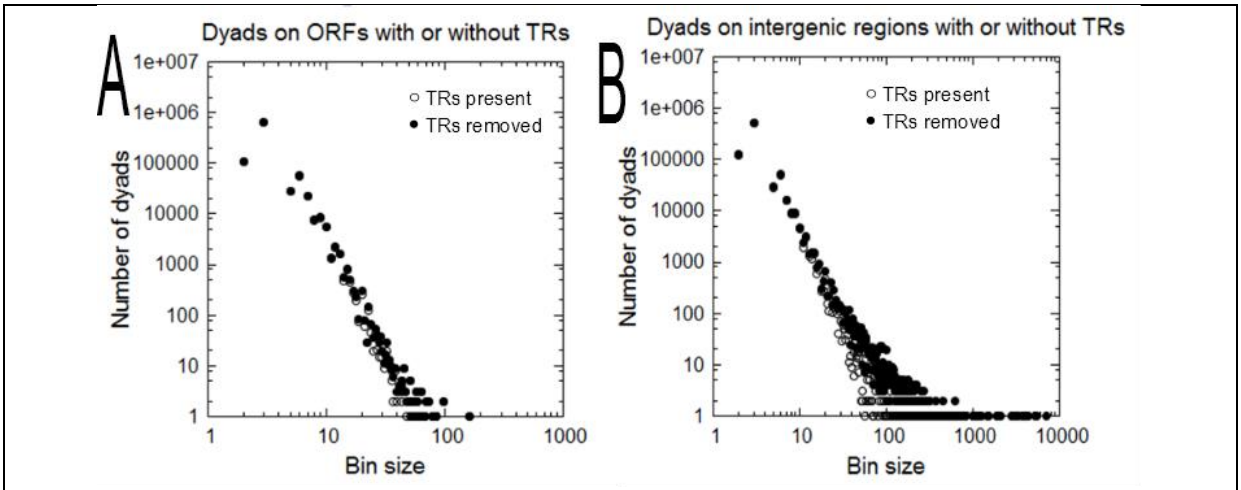


Figure 4.10: Effect of tandem repeats on binning analysis of genic (panel A) and intergenic (panel B) regions. Distribution of the number of dyads on features was performed on datasets with (unfilled circles) or without (solid circles) TR data.

Pol II transcribed genes

Next, we sought to analyse the nucleosomal packaging of all constitutively transcribed pol II ORFs (Figure 4.11 A). The slope obtained from the dyad densities of pol II PTUs was used as standard for comparison to other genetic features.

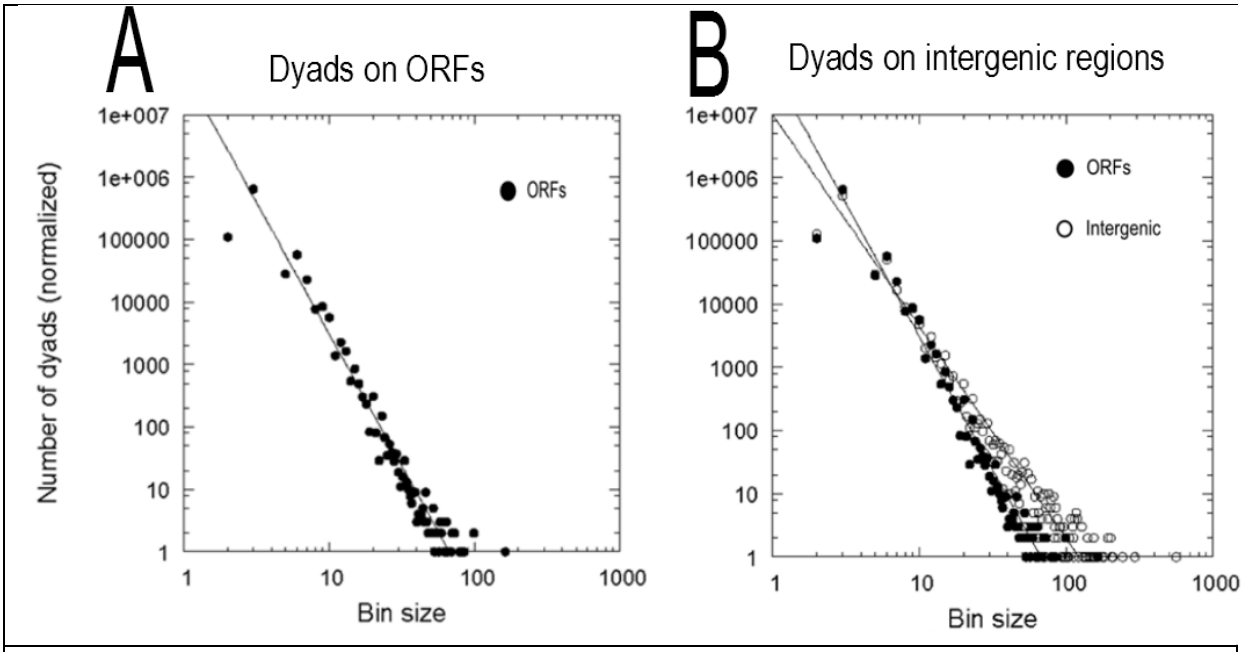


Figure 4.11: Bin analysis of dyad axis distribution on ORFs of all MBCs. A range of dyad densities are visible on pol II ORFs of the MBCs of *T. brucei* panel A), represented using a logarithmic scale, indicating an exponential relationship between these values. Panel B shows a higher compaction of nucleosomes on intergenic regions compared to constitutively transcribed pol II ORFs.

A total of 164 values were used for bin analysis, containing a total of 3116739 dyads spread over 12.98 Mb, which produced an avaraged density of 0.248 dyads/ bp (±0.098 bp).

We next compared the dyad distributuion of ORFs to intergenic regions; i.e., non-coding regions like SSRs (Figure. 4.11 B).

Intergenic regions displayed a higher dyad density than pol II genic regions. This suggested that intergenic regions of the genome were more compacted by nucleosomes than coding regions. Nucleosomes were shown to be evicted by a transcribing polymerase in budding yeast, in part explaining the lower occupancy of nucleosomes on actively transcribed gene regions, as demonstrated by Cole and co-workers [122]. This is consistent with the fact that pol II PTUs are constitutively transcribed in *T. brucei*.

Pol I transcribed genes

In *T.* brucei pol I transcribes not only the *rRNA* genes but also procyclin in the PF, and VSG in BF stages. However, only the *rRNA* genes are actively transcribed in both life forms. Analysis of dyad distribution of *rRNA* genes revealed an averaged density of 0.250281 (± 0.17). This was comparable to the dyad densities of ORFs. However, the SD for *rRNAs* was x 1.7 that of ORFs, indicating a broader distribution of nucleosomal occupancy on *rRNAs* compared to ORFs. This can also be seen when represented graphically (Figure 4.12). This could be indicative of pol I transcribed genes being depleted of nucleosomes [20].

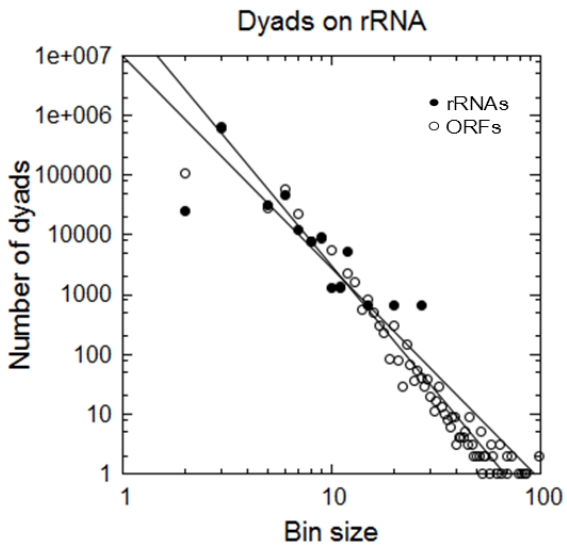


Figure 4.12: Dyad distribution on *rRNA* genes. Bin analysis performed the *rRNA* genes (solid circles) compared to ORFs (unfilled circles). The dyad densities observed on the *rRNA* genes are comparable to that of the average ORF.

Pol III transcribed genes

Next, we analysed the dyad distribution around pol III transcribed *tRNAs* (Figure 4.13). *tRNAs* appeared significantly depleted of canonical nucleosomes, displaying very low bin values compared to ORFs. It has been found that pol III transcription factors stably associate with *tRNA* sequences in *S. cerevisiae* [31], thereby pushing nucleosomes to adjacent sites or evicting nucleosomes altogether. It is possible that the low nucleosomal density observed over *tRNAs* can be an effect of transcription factors binding to these genes and pushing nucleosomes off these genes.

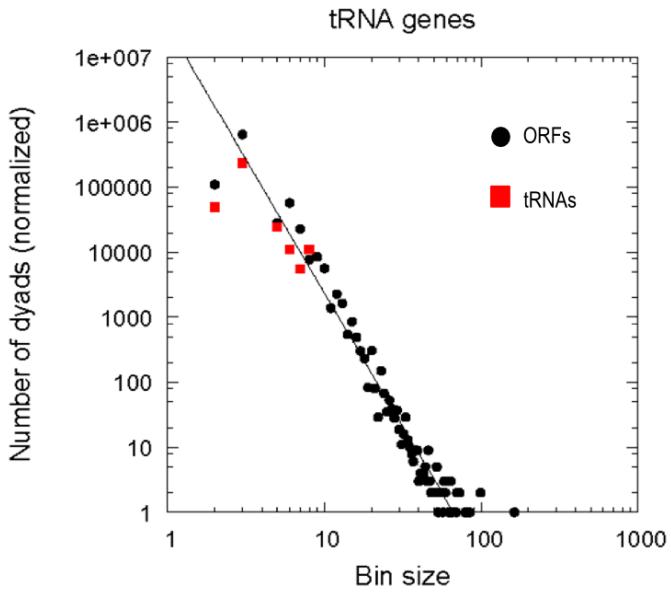


Figure 4.13: Bin analysis of dyad axis distribution on *tRNAs*. Compared to ORFs, *tRNAs* appear depleted of canonical nucleosomes, a possible result of pol III transcription factor binding to *tRNA* regulatory elements.

Subtelomeric regions

It has been demonstrated that the *T. brucei* MBCs can be demarcated into three different functional regions to analyse feature distributions globally, as was performed by Tiengwe and colleagues [63]. These regions, illustrated in figure 4.14, are as follows: 1) CDS, which contain pol II transcribed housekeeping genes, 2) subtelomeric, containing a combination of functional and non-functional genetic elements typical of subtelomeric regions ((pseudo)ESAGs, LRRPs, RHS, INGI and RIME elements), and 3) *VSG* arrays, containing silent *VSG*s. Subtelomeric regions were further divided into telomere-proximal (≤ $^{1}/_{10}$ of the chromosome length away from telomere) and chromosome internal (> $^{1}/_{10}$ of the total chromosome length away from telomere) subtelomeric regions.

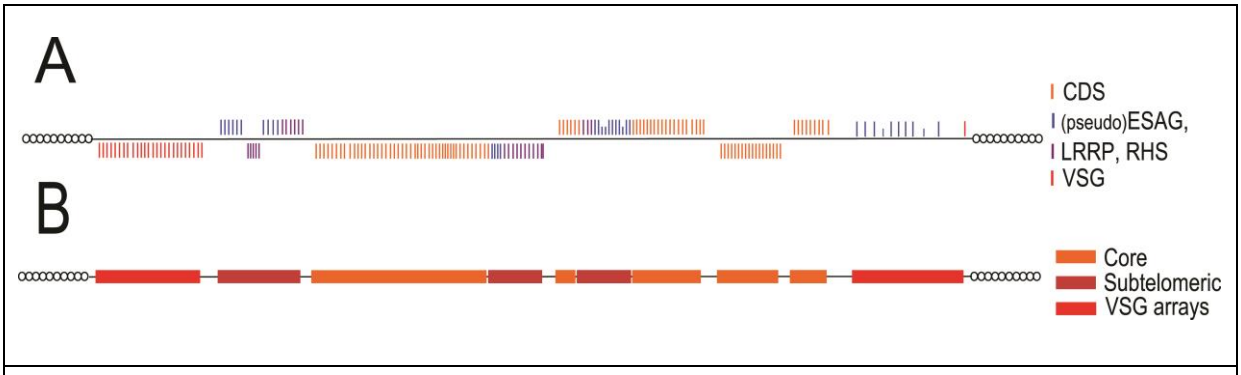


Figure 4.14: Sub-division of megabase chromosomes into three functional regions: Core, Subtelomeric and VSG array (panel B), based on the genetic content of these regions (panel A).

Bin analysis of subtelomeric regions revealed a difference in the compaction of telomere-proximal vs. chromosome internal subtelomeric regions (Figure 4.15 A and B, respectively).

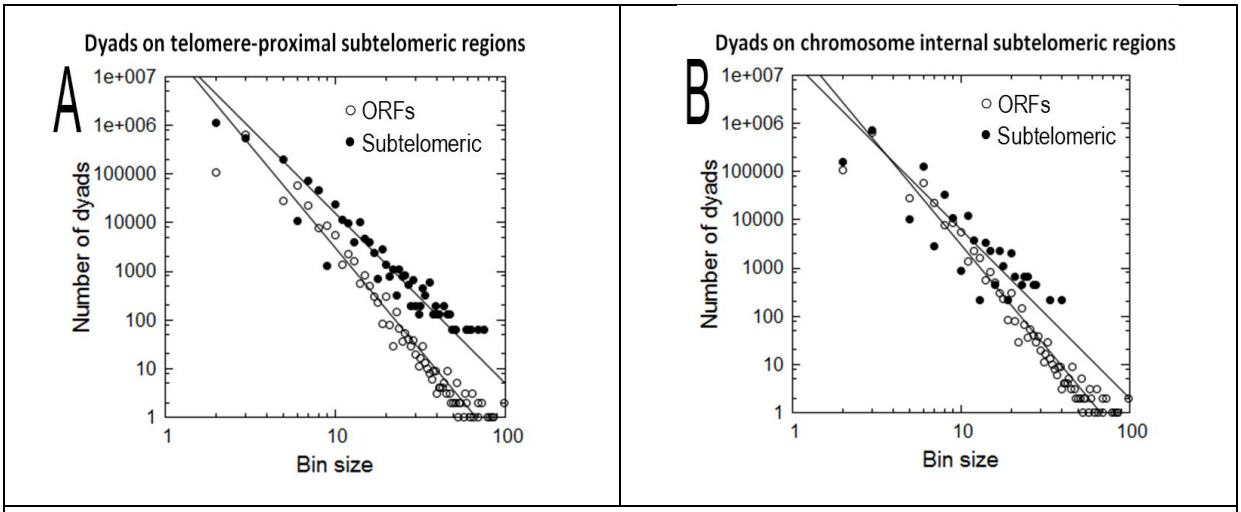


Figure 4.15: Bin analysis on telomere-proximal (panel A) and chromosome internal (panel B) subtelomeric regions. For comparison, dyad bin distributions of pol II ORFs (unfilled circles) are shown.

From figure 4.15 A it was clear that there is a higher dyad density on telomere-proximal regions compared to ORFs. The dyad densities of chromosome internal regions appeared comparable to ORFs (Figure 4.15 B).

Indeed, analysis of 114 values for telomere-proximal subtelomeric genes produced an average dyad density of 0.478 dyads/ bp (± 0.34). This was twice the density of pol II ORFs, indicating a higher degree of compaction by nucleosomes, as is expected in regions near the telomeres. Bin analysis of 41 values for chromosome internal subtelomeric genes exhibited a 0.272 (±0.18) average dyad density, which was close to the average ORF density. The telomere-proximal subtelomeric genes exhibited a higher degree of compaction than chromosome internal subtelomeric genes, indicating that the difference in dyad densities were an effect of positioning rather than genetic content of these regions.

Investigation of dyad densities on 63 values representing *VSG* gene arrays revealed an average dyad density of 0.34 dyads/ bp (± 0.33). From figure 4.16, it is clearly visible that these *VSG* arrays are occupied by a higher dyad density compared to the ORFs. This is consistent with the necessity for transcriptional silencing of these *VSG*s and could indicate the usage of nucleosomes that tightly pack DNA as transcriptional repressive structures.

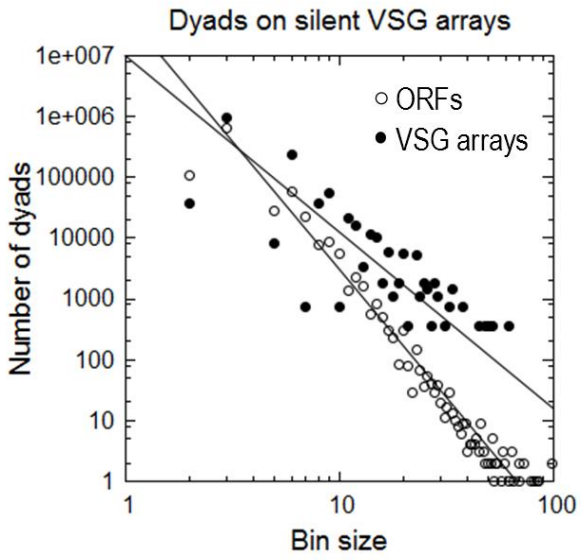


Figure 4.16: Dyad density on silent *VSG* arrays. Bin analysis reveals a higher nucleosomal occupation on *VSG* gene arrays (solid circles) compared to ORFs (unfilled circles).

Taken together, these results indicate that different features of the *T. brucei* MBCs are occupied by different levels of nucleosomes, which shows uniqueness regarding transcriptional activity and class, as well as abundance and length of these features. Pol II transcribed ORF seemed to have the lowest nucleosomal compaction of all features and regions analysed, as evident from plots above and the low average dyad density of this feature. This could be reflective of the constitutive transcription of PTUs by pol II. Silenced *VSG* arrays and telomere-proximal subtelomeric regions displayed the highest

compaction by nucleosomes. The nucleosomal density of these two regions were very similar, which is expected as they both reside within 100 kb of the telomeres. This is consistent with subtelomeric regions typically harbouring genes required to be in a transcriptionally silenced state, which can be facilitated by the close proximity of the telomeres.

2.4.7) Nucleosomal architecture

Nucleosomes tend to be arranged around genes in a manner typical to the gene class encoded by the DNA associated with nucleosomes. This is a result of a combination of factors, such as gene regulatory elements (and associated transcription factors), chromatin remodelling and intrinsic sequence preferences. We therefore sought to explore the nucleosomal architecture surrounding genes transcribed by the three polymerases of *T. brucei*.

Pol II

As seen in figure 2.2 (introduction), nucleosomes tend to have a specific distribution profile around pol II transcribed genes in model organisms. These genes are transcribed monocistronically in eukaryotes, with each gene possessing regulatory elements to which regulatory proteins can bind and facilitate PIC formation and nucleosomal distribution. Pol II transcribed genes in *Trypanosoma* exists as large PTUs, containing multiple ORFs, which are trans-spliced and polyadenylated to form mature mRNAs from an mRNA precursor [25]. These pol II PTUs does however not contain readily identifiable canonical regulatory elements, which may imply that nucleosomal organisation may facilitate pol II positioning or transcription start site selection/ demarcation. Therefore we next sought to elucidate the nucleosomal distribution around the beginning and end of pol II PTUs.

Pol II 5' PTU regions

Nucleosomal maps around the beginning of genes are normally generated by aligning mapped nucleosomes to the pol II TSS in eukaryotes. This was not possible in *T. brucei* as transcription start sequences have not been identified. Nucleosomes were therefore aligned relative the first ORF translation start codon (ATG) as this provided a consistent site across all PTUs (Figure 4.17 A and C). However, this site serves as a signal for the translational process, and has no transcriptional function and may be many bp away from the true TSS. To facilitate alignment of nucleosomes to a site closer to the true TSS, we also aligned nucleosomes to the 5' untranslated region (5'-UTR) of the first ORF (Figure 4.17 B and D), which have been mapped for each transcript in *T. brucei* [120].

The sites used for these alignments were the 5' UTR upstream of the first ORF of divergent PTUs and was obtained from TriTripDB. This site resides within divergent SSR and is partnered with another 5' UTR of another gene on the opposite strand.

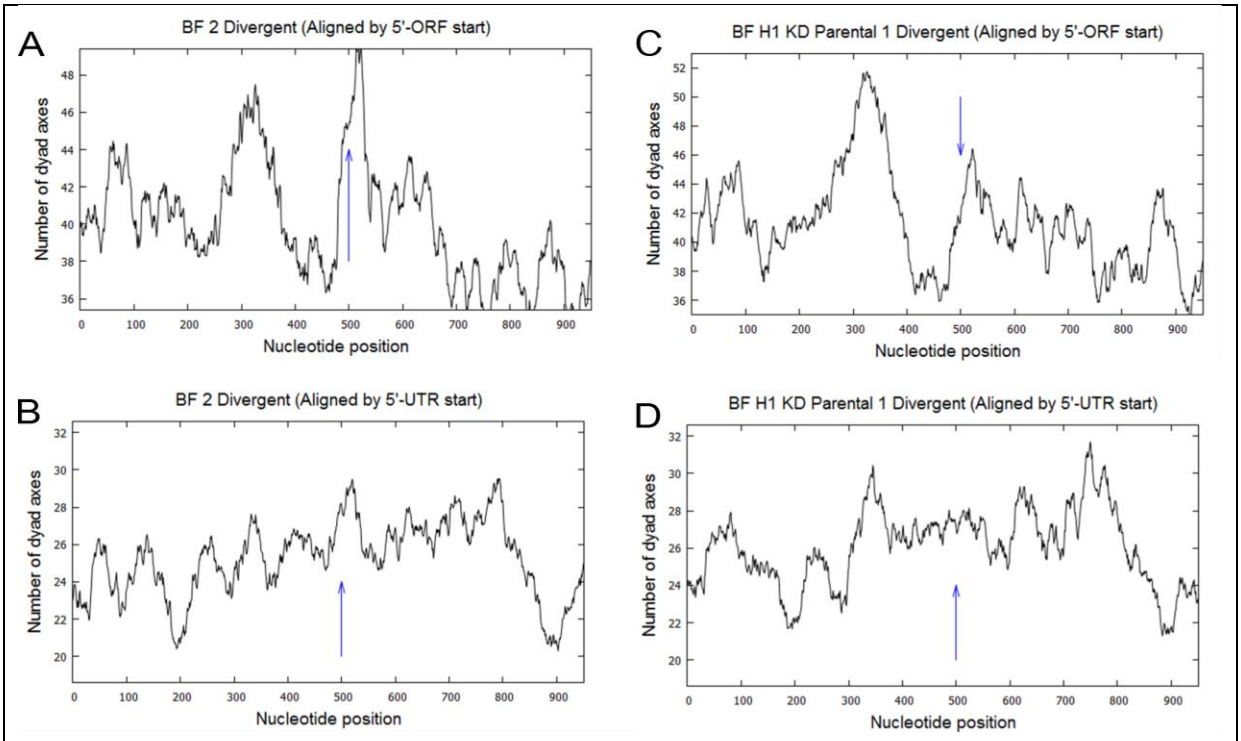


Figure 4.17: Dyad distribution aligned relative to first ORF translation start codon (top panels A and C) and the most upstream 5'-UTR, mapped by Siegel and colleagues (bottom panels B and D). The two BF strains depicted are HNI_VO2 (as BF 2, panel A and B) and RYT3 (as BF KD parental 1). ORF start and 5' UTR are indicated with blue arrows in respective panels.

Dyads aligned relative to the ORF start in BF trypanosomes (Figure 4.17 A and C) showed patterns comparable to that seen in other eukaryotes; a NDR flanked by two positioned nucleosomes on either side. The NDR occurred immediately upstream of the first ORF. The translation start site of the ORF (blue arrow) seems to be enclosed in the upstream region of +1 nucleosome. This could possibly contribute to pol II pausing just before transcribing the first ORF of a PTU. Upstream of the NDR, centred ~180 bp upstream of the start site, a -1 nucleosome was visible. Although the -1 and +1 nucleosomes seem well positioned, there appeared little evidence for phased nucleosomes in the surrounding 1000 bp depicted.

When aligned to the most upstream 5'-UTR start site, no evidence for either positioned or phased nucleosomes were observed (Figure 4.17 C, D). The peaks present in these plots were too closely spaced to present phased nucleosomes. However, there appeared to be a NDR ~ 400 bp after the 5' UTR (at nucleotide position 900), seen in all BF technical and biological replicates (supplementary material S 2.7). This is striking as the 5' UTR lengths differed from 2 – 1860 nucleotides in length.

In summary, nucleosomes surrounding the 5'-UTR start seemed poorly defined, possibly indicating the need for nucleosome arrays packing these regions rather than well positioned nucleosomes. The low definition could also possibly be caused by nucleosomes being evicted by a transcribing pol II.

The nucleosomal distribution at the beginning of PTUs in PF trypanosomes showed a pattern similar to that observed in the BF. An NDR, immediately upstream of the ORF start, flanked by a -1 and +1 nucleosome, with the ORF start again enclosed with the 5' border of the +1 nucleosome (Figure 4.18 A, C). The -1 nucleosomal dyad was again centred at ~180 bp upstream of the translation start site. The PF +1 nucleosome seemed less well positioned, with lower occupancy compared to that of the BF. As in the case of the BF trypanosomes, there seemed little evidence for well positioned nucleosomes in the 1000 bp region surrounding the ORF, with the exception of -1 and +1 nucleosomes. Interesting to note was that the dyad position of the -1 nucleosome, in both PF and BF trypanosomes, was predominantly positioned at ~ 180 bp upstream of the ATG start site.

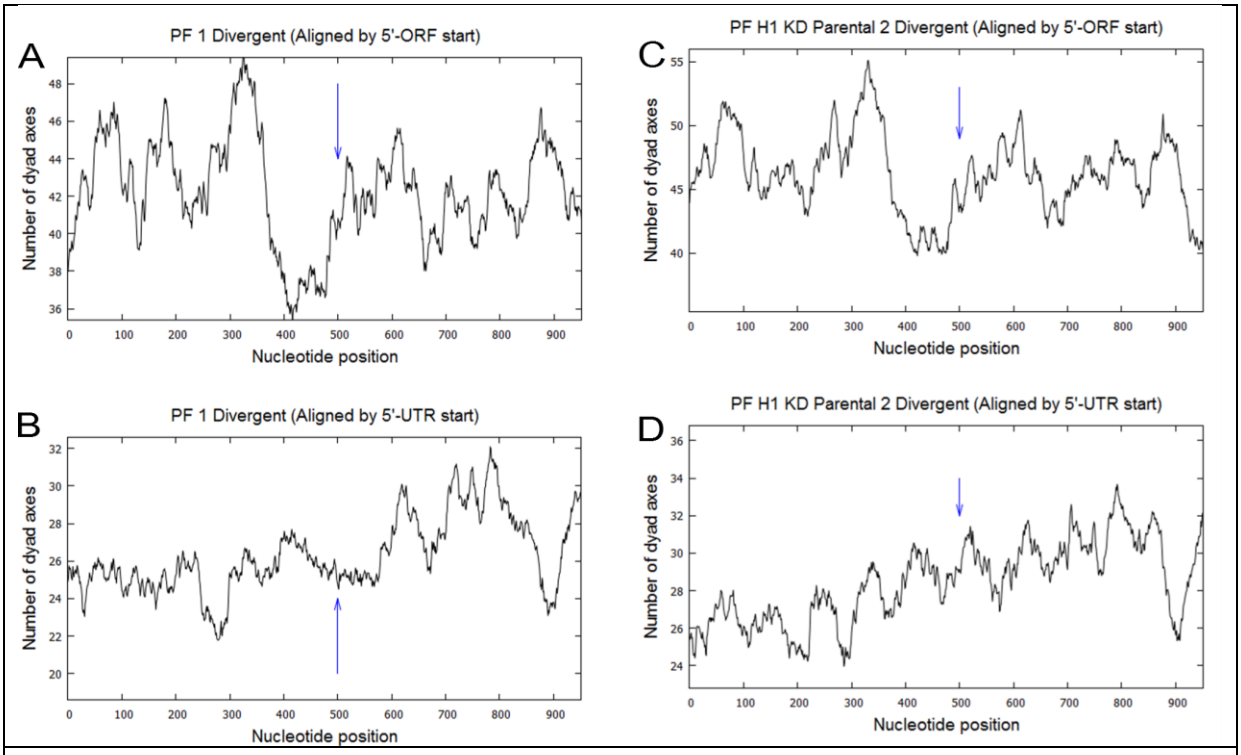


Figure 4.18: Dyad distribution in PF cells aligned relative to first ORF translation start codon (top panels A and C) and the most upstream 5'-UTR, mapped by Siegel and colleagues (bottom panels B and D). The two PF strains depicted are Amsterdam WT (as PF 1, panel A and B) and DsRed (as PF KD parental 2). ORF start and 5' UTR are indicated with blue arrows in respective panels.

When aligned relative to the 5' UTR, there was an absence of any observable nucleosomal pattern of well positioned nucleosomes (panels B, D). The nucleosome peaks observed in the 1000 bp region around the 5' UTR seemed fuzzy and could represent a nucleosomal array that packs this DNA region which was disrupted by a transcribing polymerase. Again, a NDR appeared at position 900 nt across all replicates (supplementary material S 2.8). These trends were observed across all "wild type" replicates and strains aligned relative to the ORF start and 5' UTR (supplementary material S 2.8).

Pol II 3' PTU regions

Next, we looked at the nucleosomal distribution at the 3' end of pol II PTUs. This is usually achieved by aligning nucleosomes to the transcription stop site, which has also not yet been elucidated in *Trypanosoma*.

Nucleosomes were therefore aligned relative to the translation stop codon (Figure 4.19, panel A and C) as well as the 3' PAS (Figure 4.19, panel B and D), which have also been mapped in *T. brucei* [15,120]. The genes selected for the alignment were selected if they possessed a mapped 3' UTR forming a convergent SSR with another gene on the opposite strand.

Again, specific dyad distribution profiles appeared relative to these landmarks.

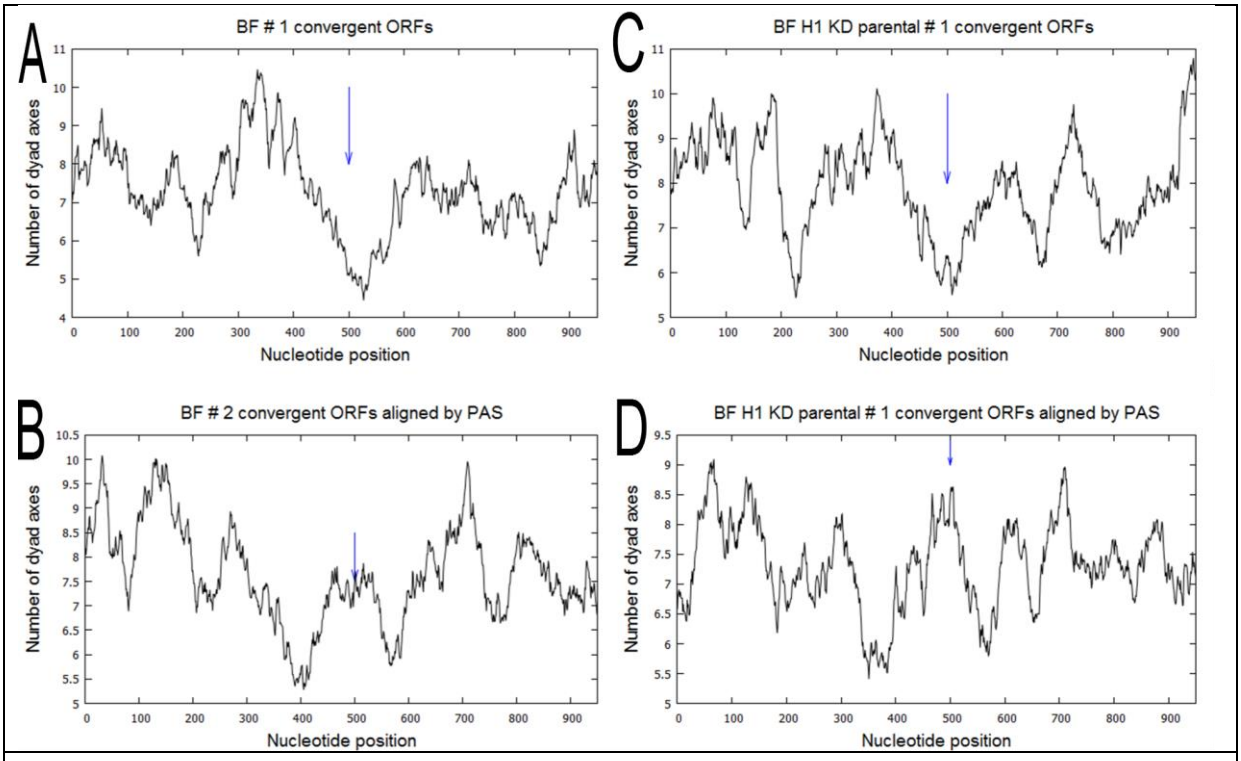


Figure 4.19: Nucleosomal organisation around the ends of pol II PTUs in BF trypanosomes. Dyads are aligned relative to the stop codon of the last ORF (top panels A and C) or to the most 3' polyadenylation site (bottom panels B and D). ORF start and 5' UTR are indicated with blue arrows in respective panels.

When aligned relative to the stop codon of the last ORF in each PTU, a NDR was seen over the 3' end of the ORF, with an enrichment of dyads flanking the NDR on each side. However, the nucleosome peaks in the surrounding 1000 bp window were positioned too close to each other to represent well phased nucleosomes although a degree of positioning was evident.

When aligned to the 3' PAS, which may present a better indication of transcription termination, a nucleosome was observed covering the PAS, although not strongly positioned. This 3' nucleosome occupying the PAS was flanked on either side by a NDR. Also observable was the presence of 3 positioned nucleosomes upstream of the PAS and a less organised region following the PAS. As it is anticipated that a transcribing pol II will disrupt nucleosomes upstream of the PAS, this was somewhat unexpected. This region enriched for positioned nucleosomes at the PAS may function in retarding the transcribing pol II near the TSS. However, the absence of a well phased nucleosomal array may indicate that nucleosomes do not act as transcription terminators, and explains the occurrence/need of *tRNA* genes and/or histone PTMs and variants at TSS to act as transcriptional terminators.

Looking at dyad distribution in PF trypanosomes, the same patterns described above was observed (Figure 4.20). The ORF stop codon was also contained within a NDR flanked by a nucleosome on either side with little evidence of nucleosomal phasing in the 1000 bp window surrounding the stop codon. When aligned to the 3' PAS, a region of intermediate nucleosomal occupancy was seen, flanked on both sides with a NDR. Three positioned nucleosomes upstream of the PAS were observed, but this positioning disappeared downstream of the PAS.

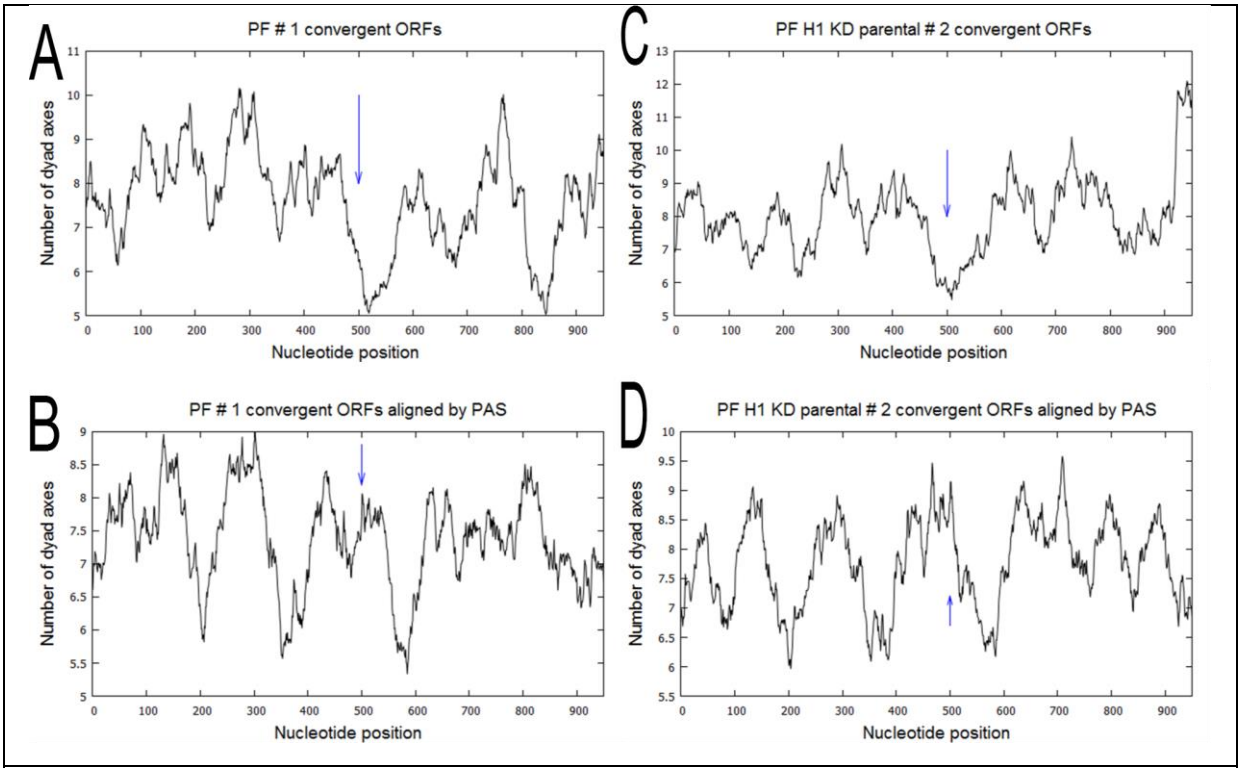


Figure 4.20: Nucleosomal organisation around the ends of pol II PTUs in PF trypanosomes. Dyads are aligned relative to the stop codon of the last ORF (top panels A and C) or to the 3' polyadenylation site (bottom panels B and D). ORF start and 5' UTR are indicated with blue arrows in respective panels.

The same patterns of dyad distribution observed in BFs were seen in PF trypanosomes when aligned to the stop codon (Figure 20 A and C). A NDR covering the translation stop site, bordered by a -1 and +1 nucleosome, were observed in the immediate vicinity of the ORF stop.

When aligned to the PAS, a single nucleosome covering the PAS was observed with a NDR flanking each side of the nucleosome (Figure 4.20 B, D). Overall, alignment to the stop site and PAS of PF were comparable to that of BF trypanosomes with little to no differences observed between the two life cycles. These patterns were observed across all biological "wild-type" replicates and strains (see supplementary material S 2.9, S 2.10).

Taken together, the nucleosomal organisation observed around the beginning and end of pol II transcribed PTUs were comparable to those seen in other eukaryotes. However, this was striking because of the (1) lack of regulatory elements in both the 5' and 3' regions of PTUs and formation of patterns around transcriptional signals, (2) *T. brucei* PTUs containing numerous genes and (3) the polycistronic nature of transcription in *T. brucei*.

Pol I

In *T. brucei*, the multifunctional polymerase I transcribes not only the rRNA genes but also procyclin and VSG genes [18].

The nucleosomal organisation around the *40S* and *60S rRNAs* (Figure 4.21 A and B) appeared to lack well positioned nucleosomes. The visible nucleosomal peaks could represent poorly positioned nucleosomes which were present as arrays packaging the DNA, which may be caused by the constitutive transcription of the genes by pol I. The *40S* and *60S* genes do, however, occur internal to pol II PTUs and might represent a roadblock to a transcribing pol II.

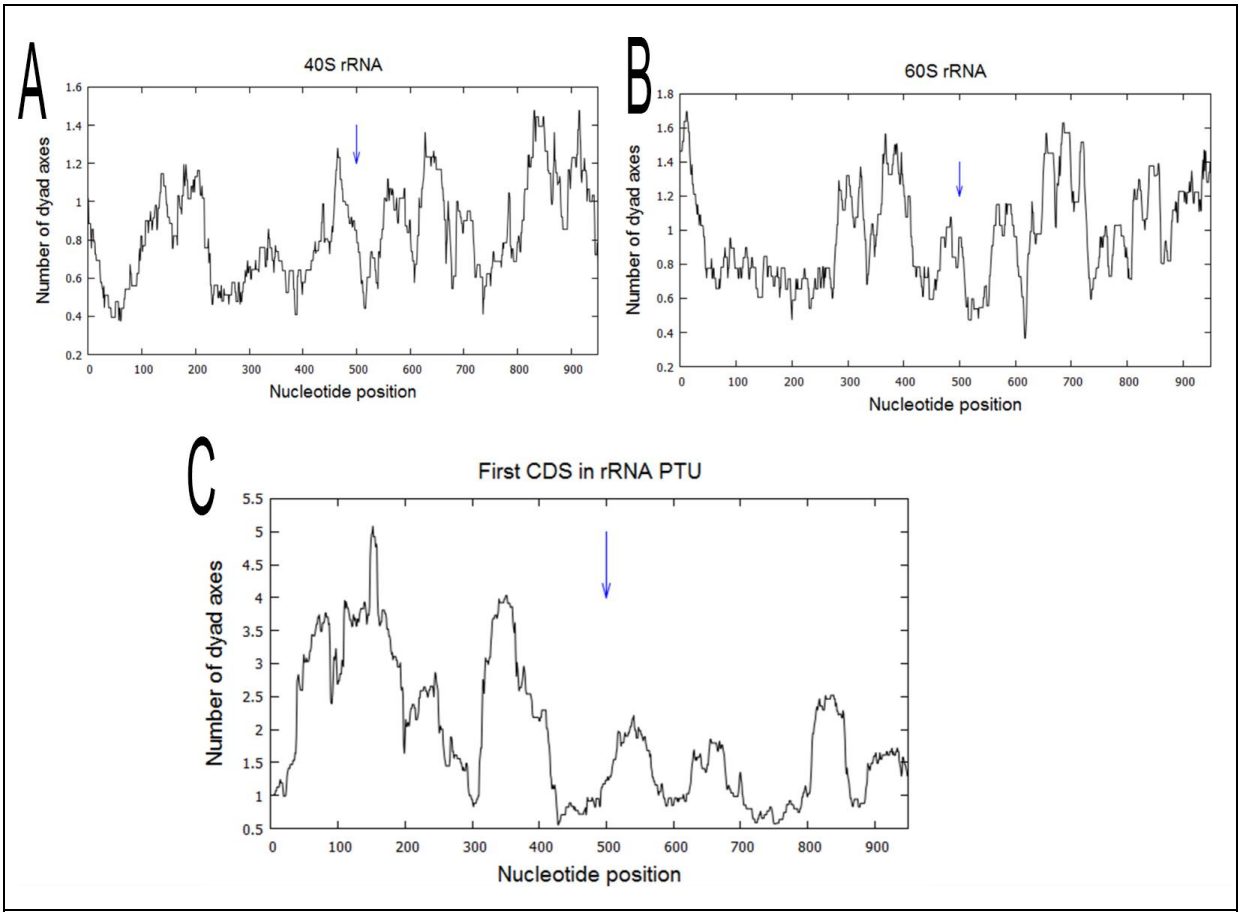


Figure 4.21: Nucleosomal organisation around pol I transcribed rRNA genes. No distinct patterns are visible around the 40S and 60S rRNA genes (panel A and B, respectively). However, pol I transcribed rRNA PTUs (panel C) does display some degree of organisation preceding the rRNA CDS start which disappears when proceeding from the CDS start.

Looking at the other pol I transcribed *rRNAs* (Figure 4.21 C), there was visible evidence of nucleosomal positioning around these genes. These genes occur as pol I transcribed PTUs (3 – 5 genes per PTU), and show a -1 nucleosome and an NDR immediately upstream of the first *rRNA* CDS start. Proceeding from the CDS start (blue arrow), there seemed to be a reduction in nucleosomal occupancy and positioning as the spacing between the concentrations of dyad peaks present were too small to represent individual nucleosomes.

Pol III

Figure 4.22 presents an aggregate plot of dyad axis of all *tRNAs*. Bin analysis indicated that *tRNAs* were depleted of canonical nucleosomes compared to surrounding regions. No strong evidence of a strong +1 nucleosome was observed for *tRNAs*, however, nucleosomes appeared moderately depleted directly upstream of the *tRNA* start and extended over the coding sequence. Strongly positioned nucleosomes do appear upstream of the NDR and downstream of the coding sequence.

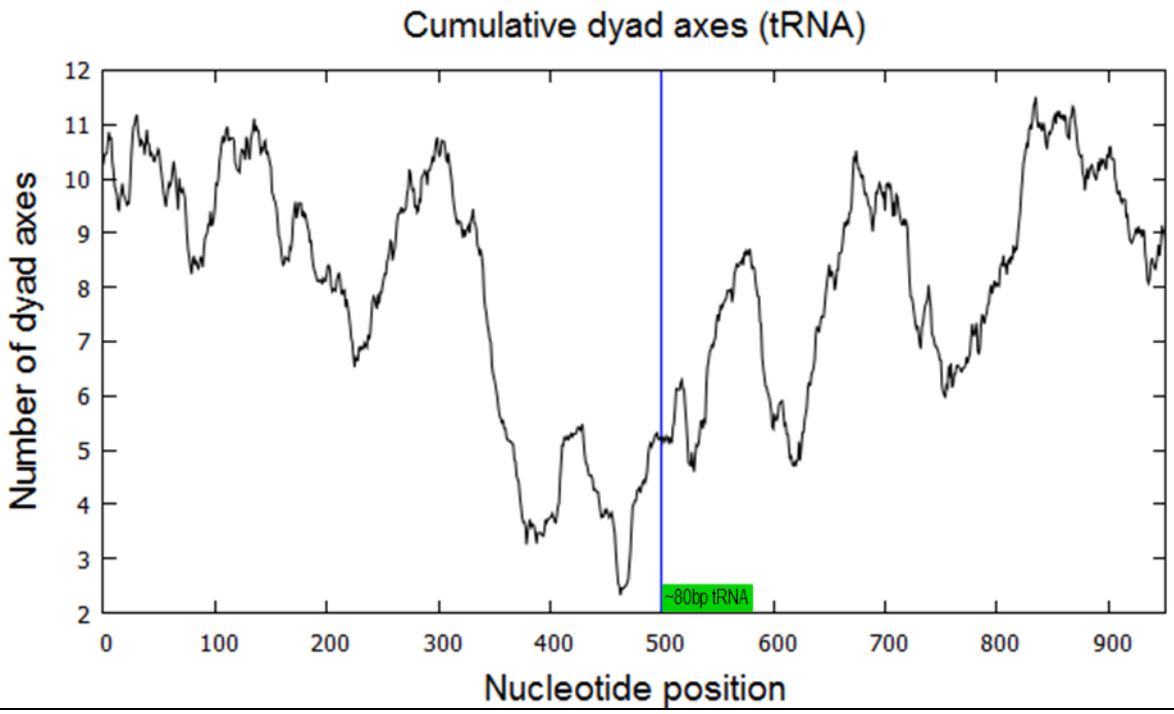


Figure 4.22: Dyad axis distribution around *tRNA* genes (green box). Nucleosomes were aligned relative to the first codon of the *tRNA* gene, indicated by the blue line.

It has been shown that in *S. cerevisiae*, *tRNA* A and B boxes are stably associated with the TFIIIB-TFIIIC complex which protects ~150 bp of DNA from MNase digestion, mimicking a nucleosome peak [31]. *T. brucei* does not possess a detectable TFIIIC, but possesses a single protein (TbBRF) subunit of the TFIIIB complex, and functional *tRNA* A and B boxes [27]. However, it is not clear whether the slight nucleosomal signals contained in the NDR are weakly positioned nucleosomes or pol III transcription factors associated with the *tRNA* promoter elements.

tRNA transcription is an example of a class II, TATA-less pol III transcription, although TbTRF4 is needed for *tRNA* expression. It was proposed that TFs may be recruited to *tRNAs* via protein-protein interactions between TF sub-units (containing nucleosome binding domains) and histones containing PTMs. Could it be that *tRNA* transcription in *T. brucei* is facilitated by positioned nucleosomes? This might be the case as *tRNAs* are flanked by nucleosomes containing histone PTMs (H4K10ac) and variants (H3V, H4V).

Effect of H1 knock-down on nucleosomal architecture of pol II PTUs

The linker histone H1, which interacts with both the nucleosome and linker DNA, is associated with repressive heterochromatin structures in eukaryotes and may act as a transcriptional repressor [123]. Knock-out of H1 in lower, single cell eukaryotes did not show major effects on transcription although yeast cells lacking H1 demonstrated genomic instability [124]. It has been shown that RNAi-mediated TbH1 knock-down resulted in significant changes in chromatin structure and increased sensitivity to endonucleases in BF *T. brucei* but not in the PF. Histone H1 has also been implicated in silencing *VSG* BES promoters and suppressing *VSG* switching.

5' PTU region

Looking at the nucleosomal distribution around the 5' ORF start of BF trypanosomes, the NDR observed in Figure 4.23 (A and B) just upstream of the first ORF was almost completely abolished after H1 KD with the appearance of a NDR following the +1 nucleosome (Figure 4.23 C).

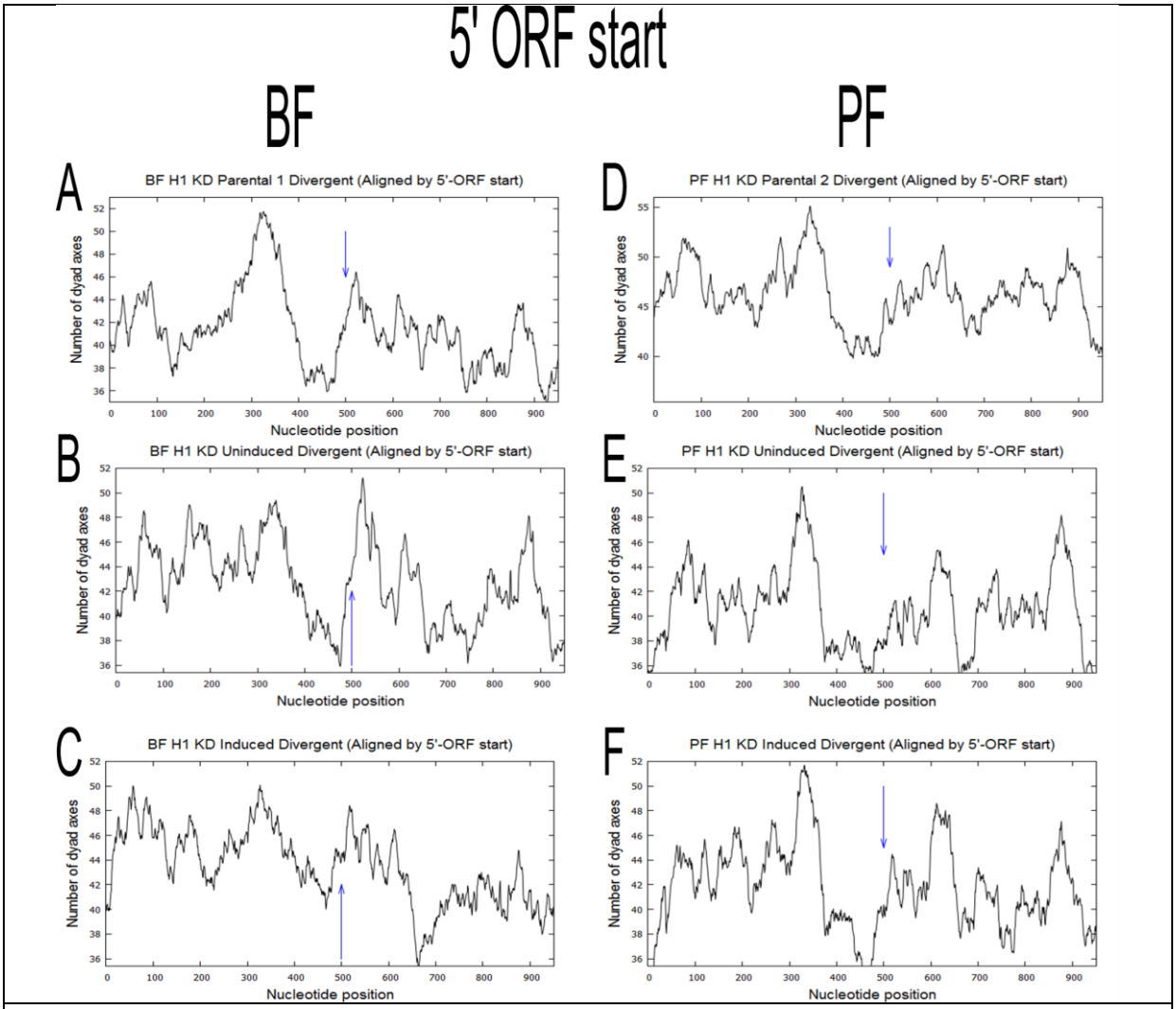


Figure 4.23: The effect of Histone H1 depletion on the nucleosomal organisation at the beginning of pol II PTUs. Dyad axes were aligned relative to the 5' ORF start site in BF and PF trypanosomes. Depicted are, as reference, the parental strains (panel A for BF, D for PF) as well as uninduced (panel B for BF, E for PF) and induced samples (panel C for BF, F for PF). The 5'-ORFstart sites are indicated by the blue arrow.

PF trypanosomes showed a pattern reminiscent of wild type PFs (Figure 4.23 C, D), with a NDR flanked by a -1 and +1 nucleosome. Following RNAi-mediated knock-down of H1 (Figure 4.23 F), no striking change in the dyad distribution surrounding the ORF start site was visible, contrary to BF where H1 is knocked down.

Nucleosomal distribution around the 5’-UTR showed no significant changes following H1 knock-down compared to either the parental or uninduced samples (Figure 4.24 C, A, B respectively). There also appeared to be little difference between BF and PF trypanosomes in this region. The close spacing of nucleosomes in these regions indicated weakly positioned nucleosomes, as observed for wild type cells.

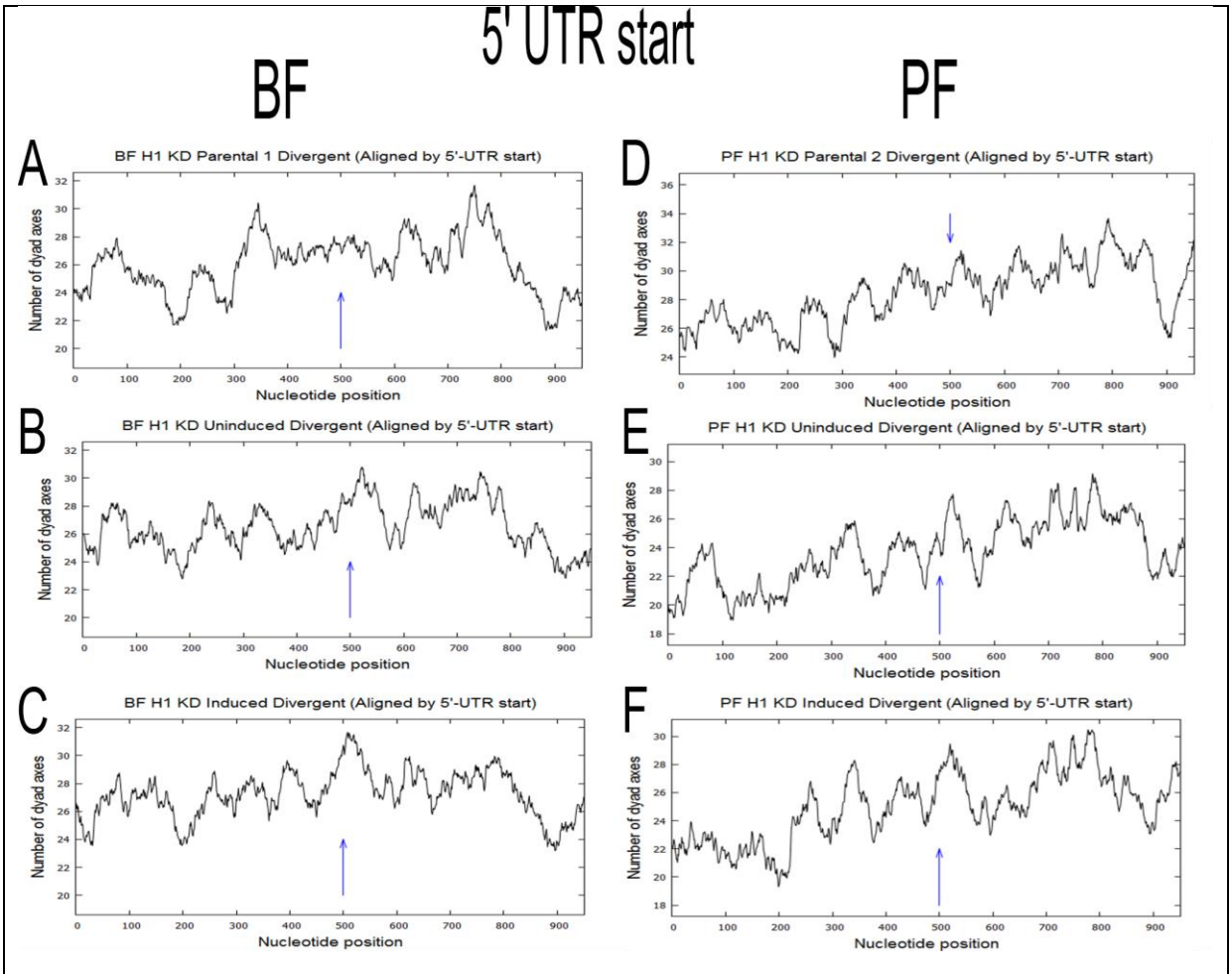


Figure 4.24: The effect of Histone H1 depletion on the nucleosomal organisation at the beginning of pol II PTUs. Dyad axes were aligned relative to the 5’-UTR start site in BF and PF trypanosomes. Depicted are the parental strains (panel A for BF, D for PF) as well as uninduced (panel B for BF, E for PF) and induced samples (panel C for BF, F for PF). The 5’-ORFstart site is indicated by the blue arrow.

3' PTU region

When aligned relative to the stop codon of PTUs, both BF and PF trypanosomes showed a NDR with nucleosomal dyad enrichment flanking both sides (Figure 4.25 A, B, E, F). Upon H1 depletion this organisation seemed to disappear, particularly in the BF sample, where nucleosomes in the 1000 bp surrounding the ORF stop became fuzzier with no evidence of positioned nucleosomes (Figure 4.25 C, F).

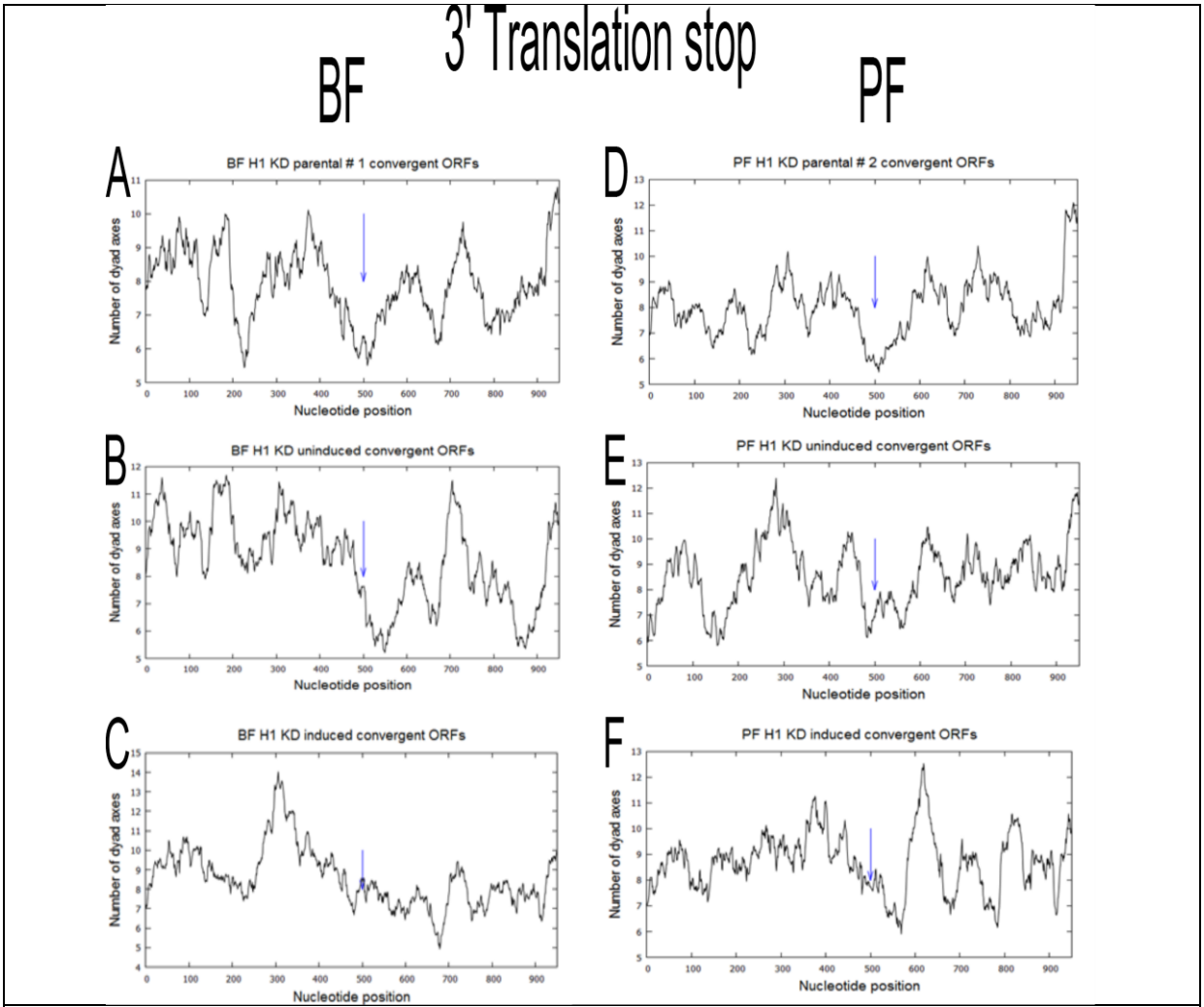


Figure 4.25: The effect of H1 depletion on nucleosomal organisation surrounding the stop site of the last ORF. Dyads were aligned relative to the translation stop codon of the last ORF at convergent ends of PTUs. Shown are the parental strains (panel A for BF, C for PF), the uninduced sample (panel B for BF, E for PF) and the induced samples (C for BF, F for PF). Translation stop indicated by the blue arrow.

When aligned relative to the PAS, both BF and PF wild type trypanosomes showed nucleosomal enrichment over the PAS with regions of relative depleted levels of nucleosomes bordering this PAS (Figure 4.26 A, B, D, E). Following H1 knock-down, this ordering seemed to disappear in both life forms, resulting in a fuzzy nucleosomal array occupying this region (Figure 4.26 C, F).

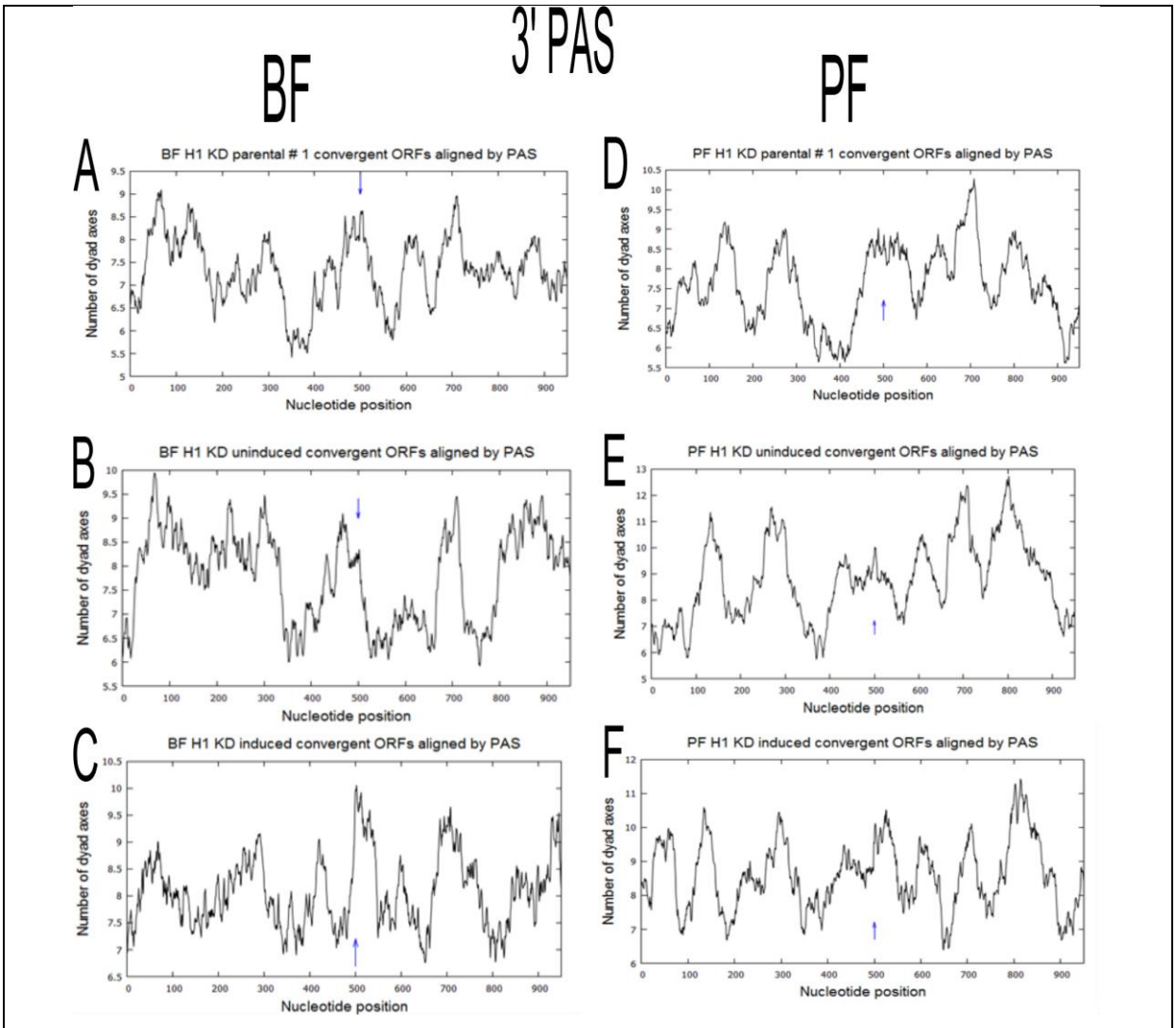


Figure 4.26: The effect of H1 depletion on regions surrounding the 3' PAS. Shown are the parental strains (panel A for BF, C for PF), the uninduced sample (panel B for BF, E for PF) and the induced samples (C for BF, F for PF). PAS indicated by the arrow.

Overall, the effect of H1 knock-down could be seen in both the BF and PF trypanosomes. The result of H1 KD around the 5' ORF region of pol II PTUs was more severe in the BF than in the PF, although no obvious differences in nucleosomal distribution was observed surrounding the 5'-UTR. H1 KD affected the nucleosomal organisation around the 3' PTU regions (translation stop and PAS) in both life forms, disrupting the patterns observed in wild type trypanosomes. The effect of H1 depletion was also more obvious in the BF cells compared to PF cells, although western blotting suggests that PFs were more depleted of H1 than BFs (Figure 4.3). This is in agreement with the observation by Povelones, where the effect of depletion of H1 on chromatin structure was more pronounced in BF trypanosomes [125].

2.4.8) Intrinsic sequence motifs of pol II transcribed regions

Base frequency upstream of PTU

Analysis of the 300 bp upstream region of the first ORFs of pol II PTUs using WebLogo revealed general thymidine enrichment in the region from 20 – 180 bp upstream of the ATG site (Figure 4.27). These thymidines may be present as oligo-T runs which possibly functions as nucleosomal positioning signals.

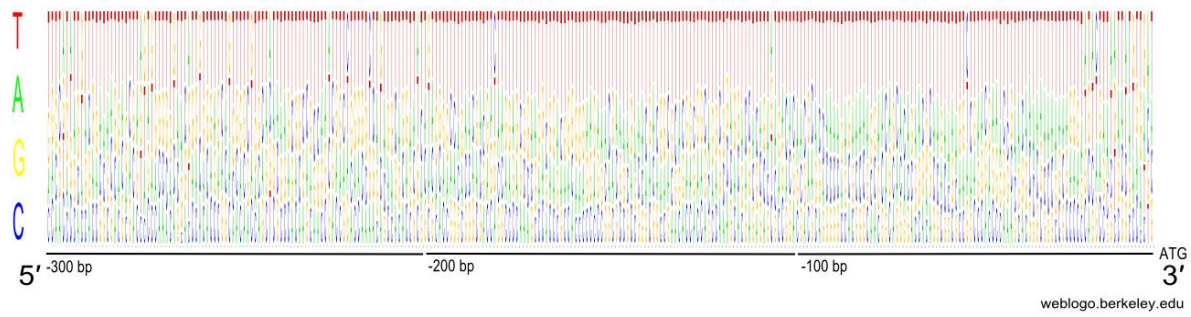


Figure 4.27: Base frequency analysis of the 300 bp upstream of the first ORF of pol II transcribed PTUs. A general enrichment of thymidine residues are seen in the region 20 – 180 bp upstream of the ATG start site.

The observation of T enrichment led to the exploration of the upstream regions for sequence motifs which might act as nucleosomal positioning signals.

Sequence motifs

Typically, the upstream regions of pol II transcribed genes in eukaryotes contain multiple regulatory elements, such as upstream promoter elements (eg. CCAAT box) and core promoters (includes TATA box, initiator element and translation start codon). Putative pol II TSS in *T. brucei* are devoid of these elements, with only an oligo-G run identified between divergent SSR [24]. As

putative pol II TSS occurs at divergent SSRs, these oligo-G runs have been suggested to provide directionality to transcription of these regions.

Noting the general enrichment of AT nucleotides upstream of the first ORF, we explored the 100 bp upstream regions of pol II PTUs for sequence motifs which might function in the transcription process or aid in nucleosome positioning, using MEME (available at http://meme.ncbr.net). Figure 4.28 shows the results of this analysis.

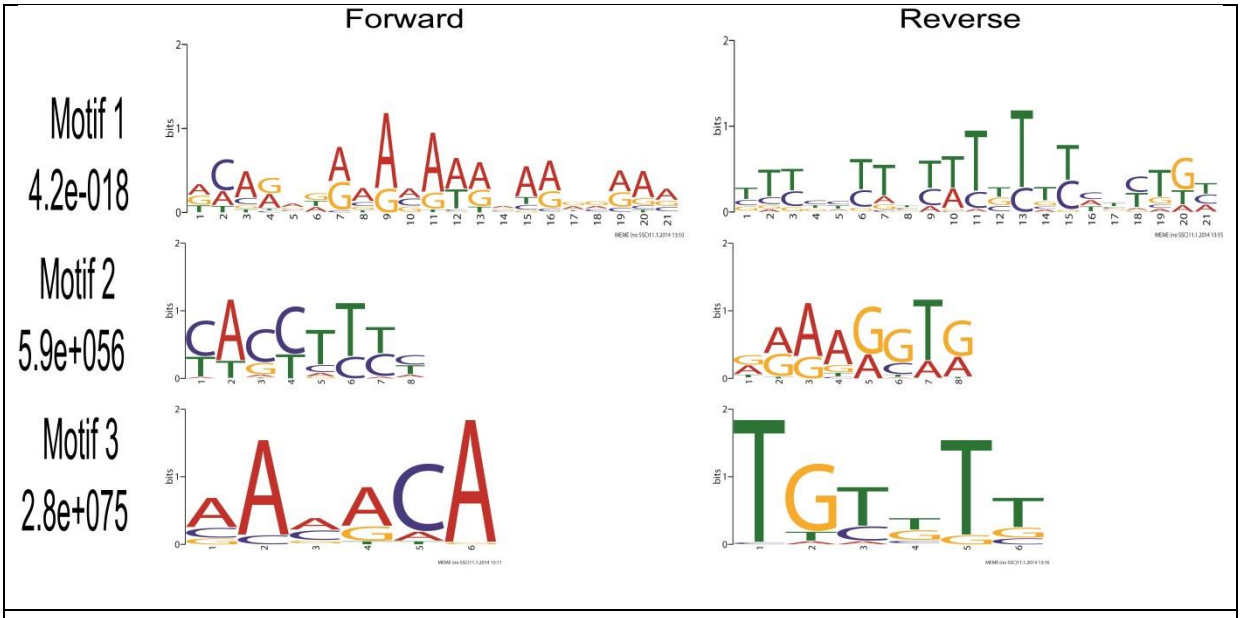


Figure 4.28: Motif analysis by MEME analyser detected three possible motifs in the 100 bp region upstream of pol II ORF starts. Only motif 1 seems to be statistically significant.

Three motifs were detected in the 100 bp upstream region of pol II PTUs by MEME. Motif 1, being statistically significant with an E-value of 4.2×10^{-18}, was investigated further to examine any possible AT bias/ skew in the 1000 bp surrounding the 5' ORF region.

AT skew

In the absence of selective pressure and under random distribution of nucleotides, the bases A, T, G and C would be evenly distributed in DNA. However, in living organisms there appears an asymmetric distribution of these bases when aligned relative to certain genomic landmarks. This is referred to as the AT or GC skew, represented as: AT skew= (A – T)/ (A + T).

Pol II 5' PTU regions

Analysis of the base frequency indicated an enrichment of T nucleotides upstream of the pol II ORF of *T. brucei,* implying a possible AT skewness in the region surrounding the pol II ORFs. AT skew analysis by "AT enrichment" confirmed that there was indeed an AT skewness in this region (Figure 4.30).

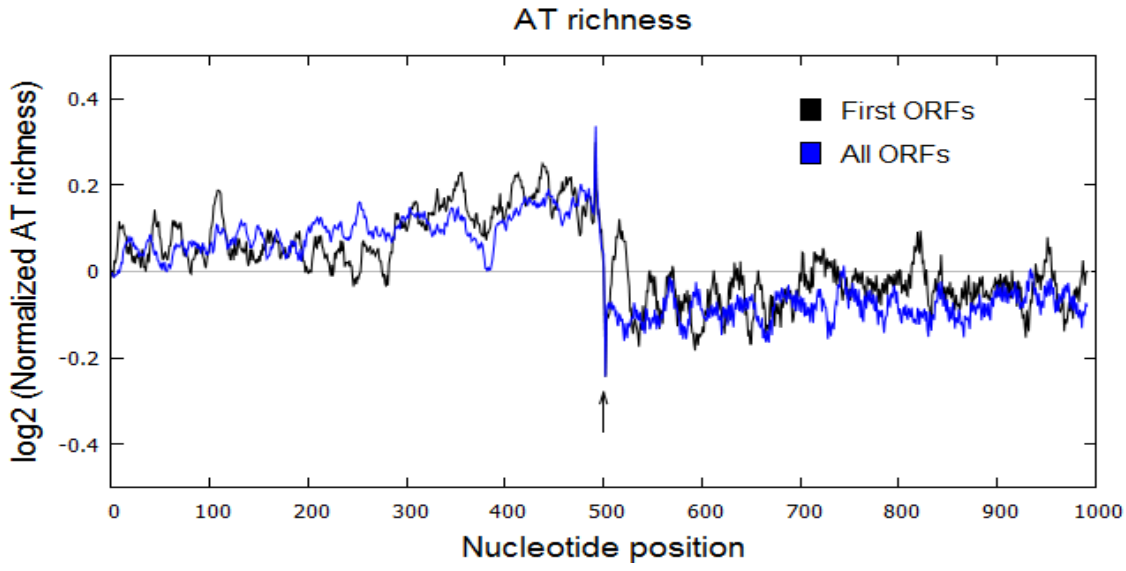


Figure 4.30: AT enrichment at the beginning of pol II PTUs. A clear AT enrichment can be observed in the 200 bp preceding the first ORF (black line), comparable to AT enrichment preceding all ORFs (blue line).

Figure 4.30 shows a significant drop in AT richness, normalized to the average of the genome, when proceeding from the translation start site into the coding

region. The 200 bp immediately upstream of the translation start site appeared AT enriched compared to the genome average, with little difference in AT richness in this region between the first ORF (black) and all ORFs (blue).

When looking at A enrichment on the non-coding strand (Figure 4.31 A), the 100 bp region upstream of all ORFs seemed more A rich compared to both the genome average and the first ORF, but this trend disappeared in the preceding 400 bp region. The A richness of the first ORF was comparable with the genome average, although the 500 bp preceding the first ORF seemed less A rich than the genome average. The first ORF also seems depleted of A in the region directly following the ATG start. This showed a decided A rich skewness between the coding and non-coding strands.

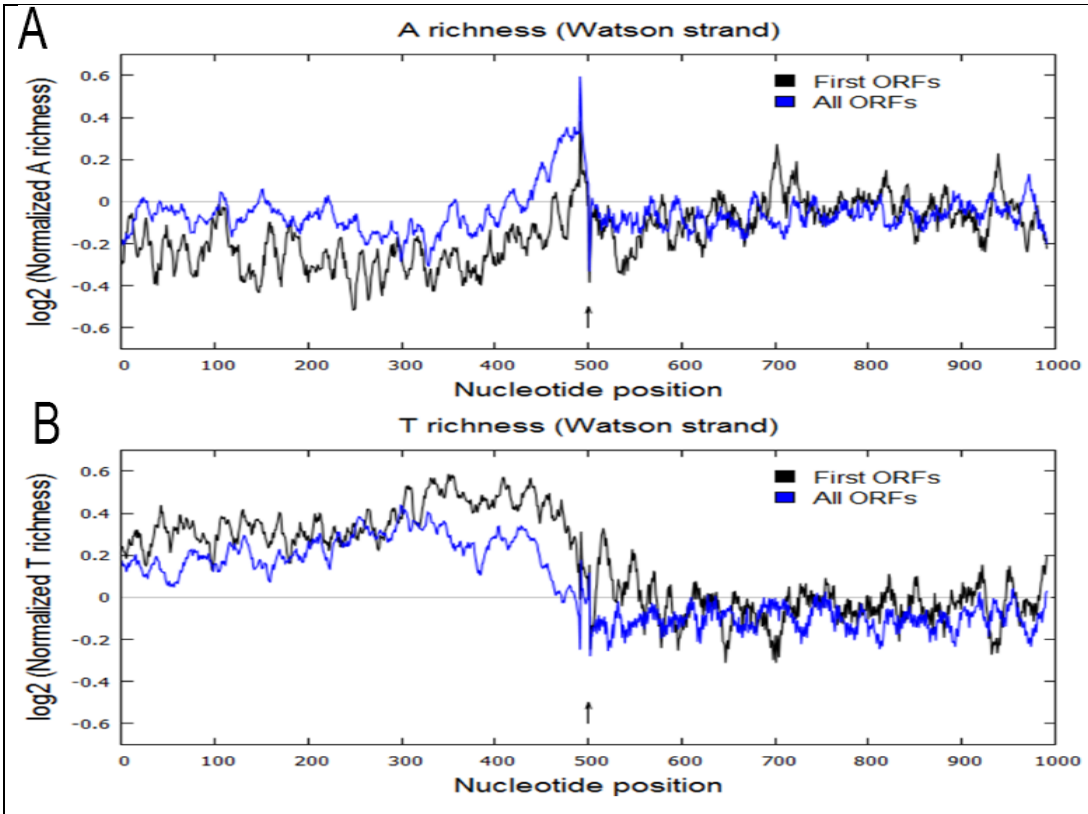


Figure 4.31: A and T richness in the 1000 bp surrounding the ATG start codons (panel A and B, respectively). These graphs indicate a general enrichment of Ts preceding both the first (black line) and all ORFS (blue line).

Looking at the T richness of the non-coding strand (Figure 4.31), an inverse relationship to A richness was seen, as expected. The entire 500 bp region upstream of the ATG site of both the first and all ORFs appeared enriched for T relative to the genome average. The 200 bp preceding the ATG site of the first ORF seems more T enriched than that of all ORFs, which was in agreement with the observed T enrichment detected by the base frequency analysis (Figure 4.27).

Beyond 200 bp upstream of the ATG site the T richness of the first ORF and of all ORFs appeared similar. Interestingly, the first 50 bp of the CDS of the first ORF appeared more T rich than the average ORF, but beyond this region, the T richness was slightly below the genome average.

It is known that poly(dA/dT) tracts are resistant to bending and can selectively exclude nucleosomes. Is it possible that this AT skewness at the beginning of ORFs can account for the well positioned nucleosomes observed in these regions and are there extended runs of T's on the Watson strand that serves to exclude nucleosomes from specific localities?

Plotted in figure 4.32 A is 100X number of T runs normalized to the number of sequences analysed, and smoothed with a 50 bp running average window. Run lengths of T5 to T14 were identified. The same was done for figure 4.32 B to explore A enrichment.

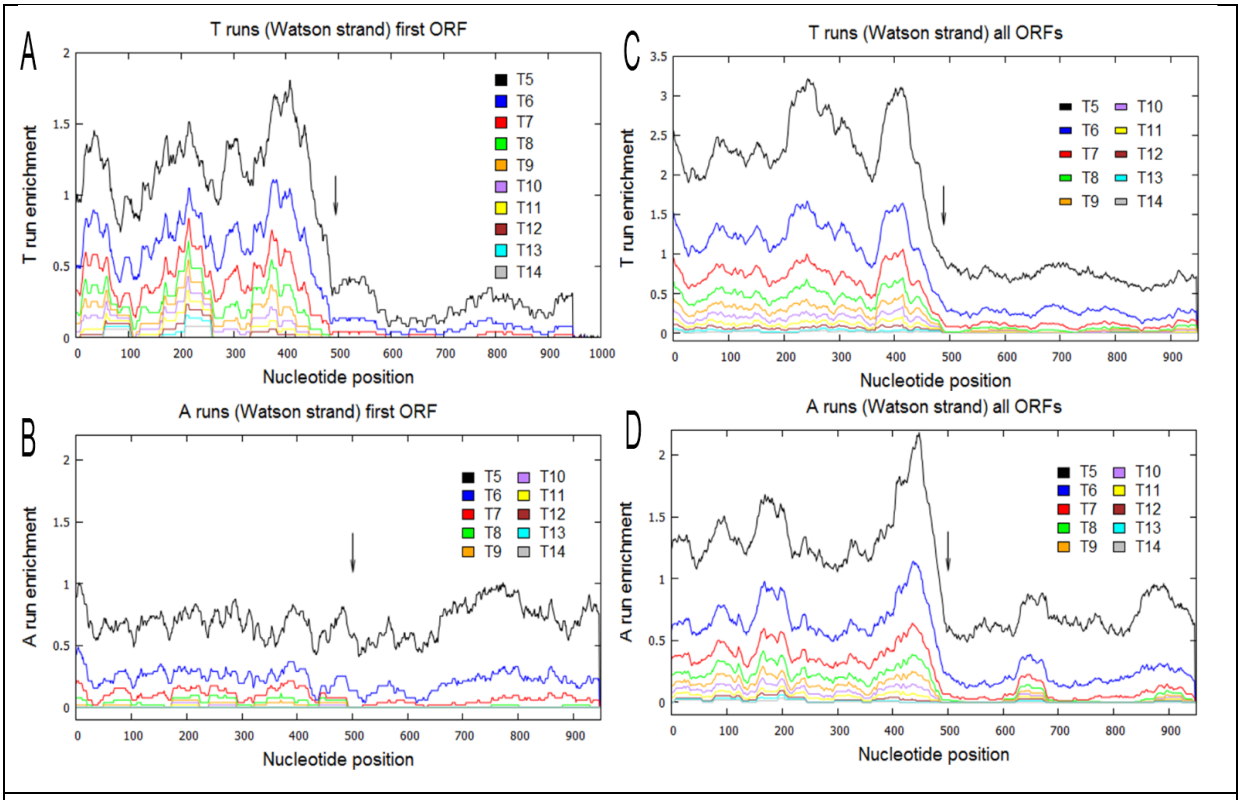


Figure 4.32: Analysis of T and A runs of 5 – 14 nucleotides in length in the 1000 bp surrounding the first ORFs (panels A and B, respectively) and all ORFs (Panel C and D, respectively).

The difference between T runs before and after the ATG site was as expected from the T richness plot above. Figure 4.32 A shows a number of T runs occurring upstream of the ATG site of the first ORF. There was no striking enrichment of A runs over the entire region analysed for the first ORF of the PTU (Figure 4.32 B). The slight oscillation visible in the region upstream of the ATG site was not of high magnitude and likely due to noise. When analysing the regions surrounding all ORFs, a number of T runs again appeared upstream of all ORFs (Figure 4.32 C). It appeared that the largest of the T runs were distributed in a number of discrete peaks in the case of the first ORF as compared to the distribution T runs of all the ORFs. In agreement with the observed A enrichment preceding the ATG site of all ORFs, a large number of short A repeats were observed in the 500 bp region preceding the ATG site

(Figure 4.32 D). Looking only at the distribution of T10 runs in the upstream region, it was striking that a clear periodic distribution is seen in the case of the first ORF, compared to the average of all the ORFs, where no periodic modulation was visible (Figure 4.33 A). Also clearly visible in the T10 distribution around the first ORF were areas of T depletion, centred approximately 200 and 380 bp upstream of the ATG site.

As previously stated, poly(dA/dT) runs plays a role in nucleosomal positioning by selectively excluding nucleosomes. It is thus likely that the distribution of T runs seen in the upstream region of the first ORF could influence the placement of nucleosomes in this region.

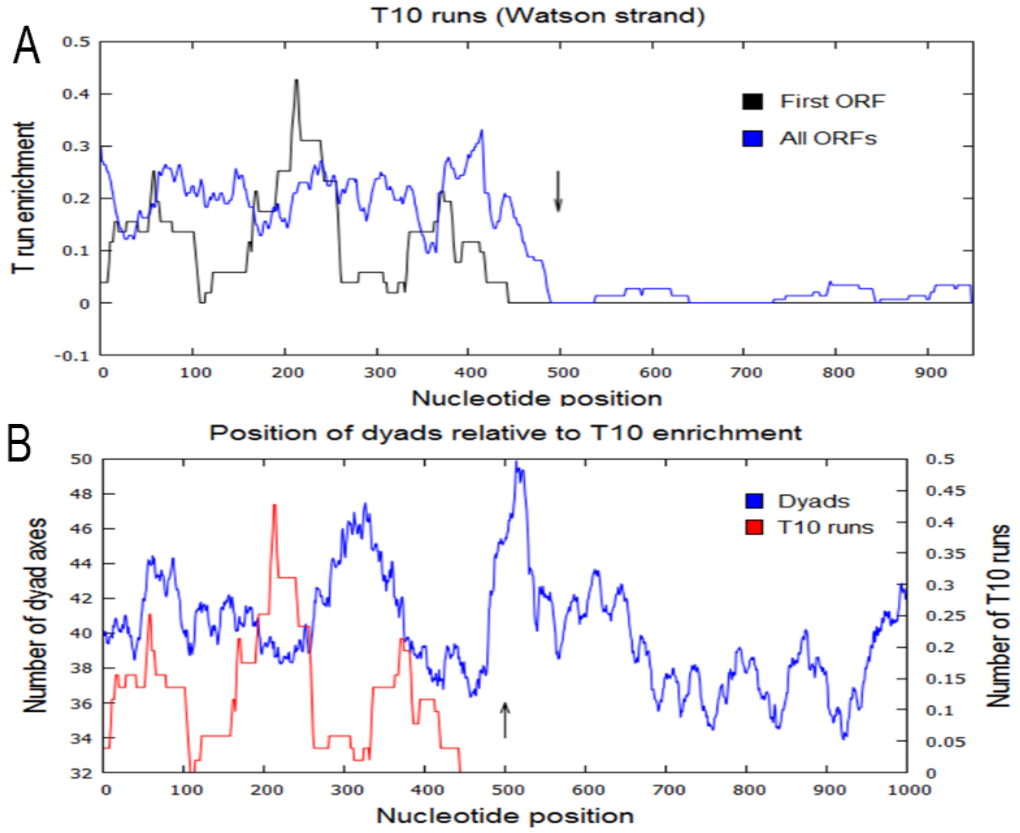


Figure 4.33: Distribution of T10 runs preceding the first and all ORFs (panel A) in the 500 bp preceding the ATG site. Panel B illustrates the superimposition of the T10 runs to the average nucleosomal distribution around the first ORF of pol II PTUs.

When the positions of these T10 runs were superimposed to the average dyad distribution of the first ORF, an inverse correlation between the T10 runs and positioned nucleosomes appeared (Figure 4.33 B).

It appears that *T. brucei* has evolved an asymmetric distribution of T runs to indicate directionality at the promoter, and are using oligo T runs to precisely oppose the upstream nucleosome to leave the putative transcriptional start site free for binding by the transcriptional machinery at the beginning of the pol II PTU.

Pol II 3' PTU regions

From analysis of the AT skew surrounding the ends of PTUs, aligned relative to the translation stop codon on the non-coding strand, there was an observable increase in AT richness when proceeding from the coding sequence to the 3' UTR (Figure 4.34 A). This probably reflects the GC richness of the CDS. There also appeared a drop in AT content ~30 bp upstream of the translation stop of the last ORF. Except for this region there appeared little difference between the AT content of the last ORF compared to all ORFs. Analysis of A richness showed no striking enrichment or depletion for As surrounding the stop codon for either the last of all ORFs (Figure 4.34 B). However, there was an increase in T richness when proceeding from the ORF stop downstream towards the 3' UTR (Figure 4.34 C). This accounts for the increase in AT richness following the stop codon. No significant difference was present between the T richness of the last and all ORF stops.

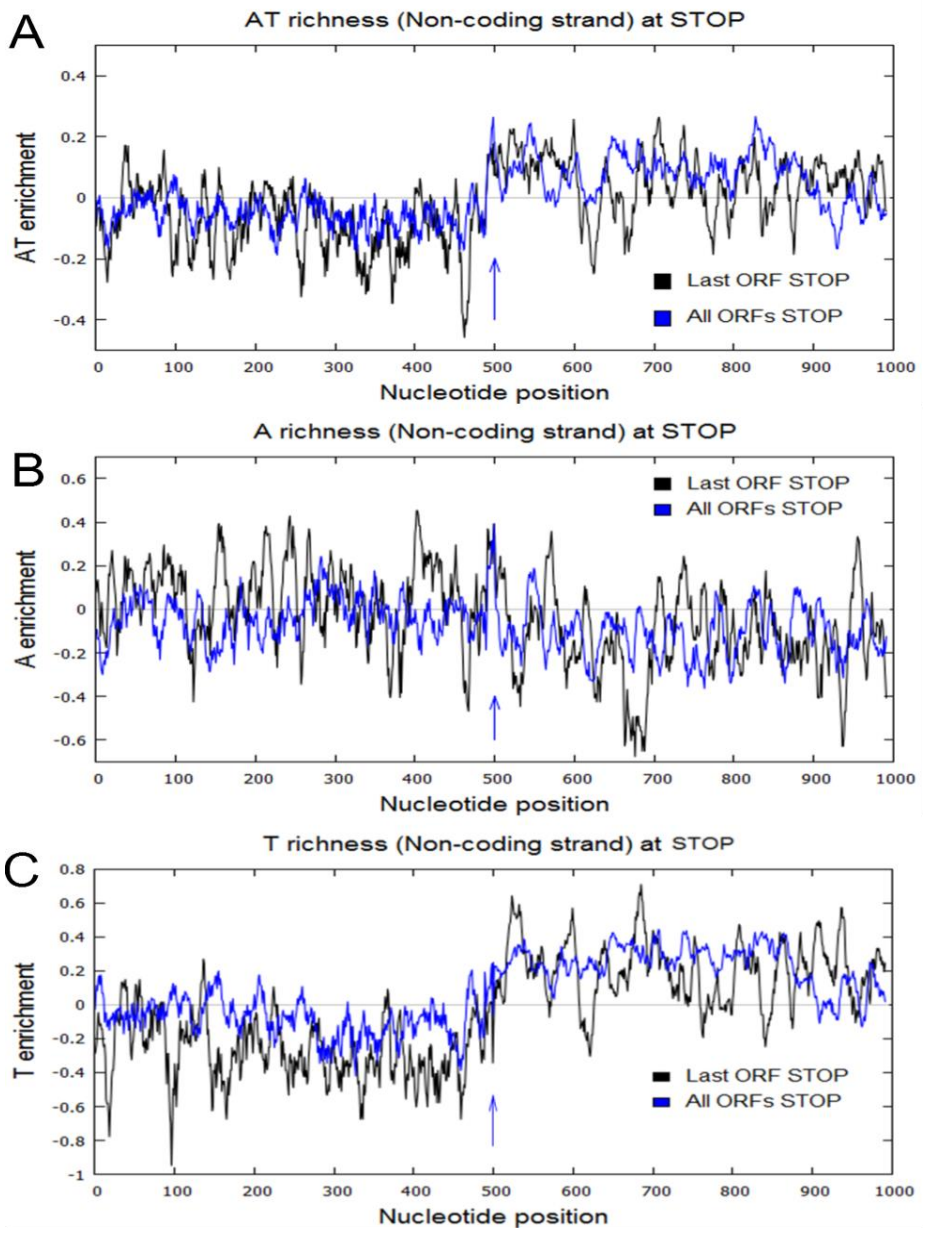


Figure 4.34: Analysis of the AT skew surrounding pol II PTU ends. Panel A shows the combined AT skew for the 1000 bp region surrounding the stop site (Blue arrow), panels B and C represent A and T enrichment over this region, respectively.

Aligned relative to the 3' PAS, a narrow region of AT depletion was observed immediately following the PAS (Figure 4.35 A). Although the distribution of A richness in the 1000 bp surrounding the PAS appeared reasonably evenly distributed relative to the genomic average (Figure 4.35 B), there appeared a slight decrease in T richness on the non-coding strand at the PAS (Figure 4.35 C)

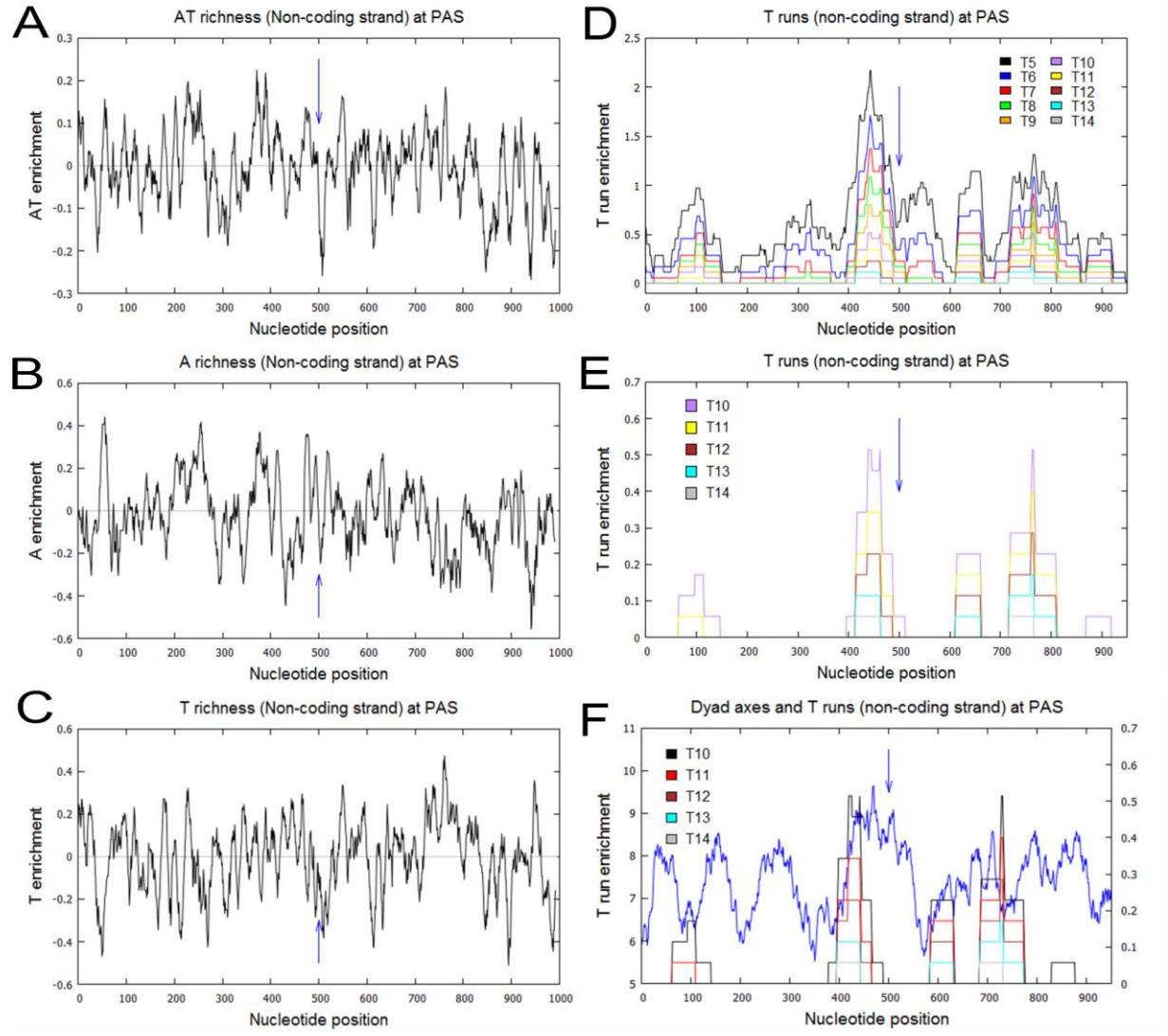


Figure 4.35: AT skew analysis of the 1000 bp region around the PAS of pol II PTUs (panel A). Panels B and C shows the separate A and T enrichment of this region, respectively. T-runs of 5 – 14 residues in length around the PAS is shown in panel D, with long T runs (≥ 10 bp) represented in panel E. Panel F shows the superimposition of average nucleosomes relative to the long T runs.

Looking at the distribution of T5 – T14 runs, a region of enrichment of long T runs immediately upstream of the PAS was seen (Figure 4.35 D). Is there any sequence involved in pol II pausing or processing of the PAS and are these T runs also involved in nucleosomal organization of the PAS? Figure 4.35 E represents T runs of T 10 – T14 in length. The region immediately upstream of the stop codon appeared particularly enriched for long oligo-T runs.

We have previously shown that there appeared 3 well positioned nucleosomes upstream of the PAS, and a single nucleosome aligned with the PAS (Figure 4.20). Superimposing the long T runs to the dyad axis distribution around PAS the nucleosomes upstream of the stop codon appeared herded into positions by the T runs at positions 100 and ~420 (Figure 4.35 F). The T runs directly flanking the PAS (at ~420 and 600) may act to shepherd the nucleosome at the PAS into position.

There was a subtle indication that the region left of the range of positioned nucleosomes (~420) depleted of nucleosomes, was centred at position 350. It is not clear whether the T runs at position ~420 is responsible for this, although the peaks do not quite align. The region downstream of the PAS appeared more random. This may be because of a combination of *tRNA* genes in this region of some of the genes, which adds noise to the combined signal.

It seems that the T runs directly upstream of the PAS may be involved in pushing nucleosomes onto the PAS. It is unclear whether this contributes to pol II pausing or termination of transcription at or directly beyond the PAS.

A certain percentage of thymidine residues are converted to Base J in BF *T. brucei*, and was found to be enriched both up- and downstream of pol II PTUs and particularly enriched at repeat sequences. It is possible that the oligo-T

runs detected up- and downstream of pol II PTUs may contain J residues. However, the effect of J on DNA rigidity and nucleosomal positioning is not known and would need further elucidation.

Taken together, it seems that *T. brucei* may use intrinsic DNA sequences to position nucleosomes at the beginning and end of individual genes and whole polycistrons by directly opposing nucleosomal binding at specific sites.

2.4.9) Dinucleotide distribution

It is clear that DNA sequences containing oligo-T runs can selectively exclude nucleosomes from binding and thus influence global nucleosomal positioning in eukaryotes [126]. This strategy seems to also be employed by *T. brucei* to position nucleosomes relative to genomic landmarks. Well positioned nucleosomes have been observed to be associated with DNA in which 10 bp periodical dinucleotides, like AA and GG, are present [110]. AA dinucleotides can more easily be accommodated in the narrow minor groove, and is this located at positions in the nucleosomal DNA where the minor groove points towards the octamer surface and the DNA minor groove is compressed. We next asked if *T. brucei* also employs the use of dinucleotides to provide a rotational setting of the DNA on the octamer surface.

We performed a dinucleotide frequency analysis on 139 values, representing the first ORFs of PTUs at divergent strand switching regions. Sequences of 147 bp in length, centred at position 180 nt upstream of the ATG start codon, was retrieved from TriTrypDB. These sequences represent the approximate boundaries of the well positioned -1 nucleosome observed in the nucleosomal architecture around pol II transcribed 5' PTU region.

The distribution of all 16 possible di-nucleotides in the aligned sequences was determined using the program "dinucleotide distribution". The two di-nucleotides that showed the strongest periodic distribution was AA/TT and GG/CC.

The plot of the number of AA dinucleotides at each position in the aligned 147 bp sequences is shown in figure 4.36. The grey, vertical bars are at 10 bp intervals, centred at 74 nt, the location of the dyad axis in a 147 bp nucleosome sequence (red arrow).

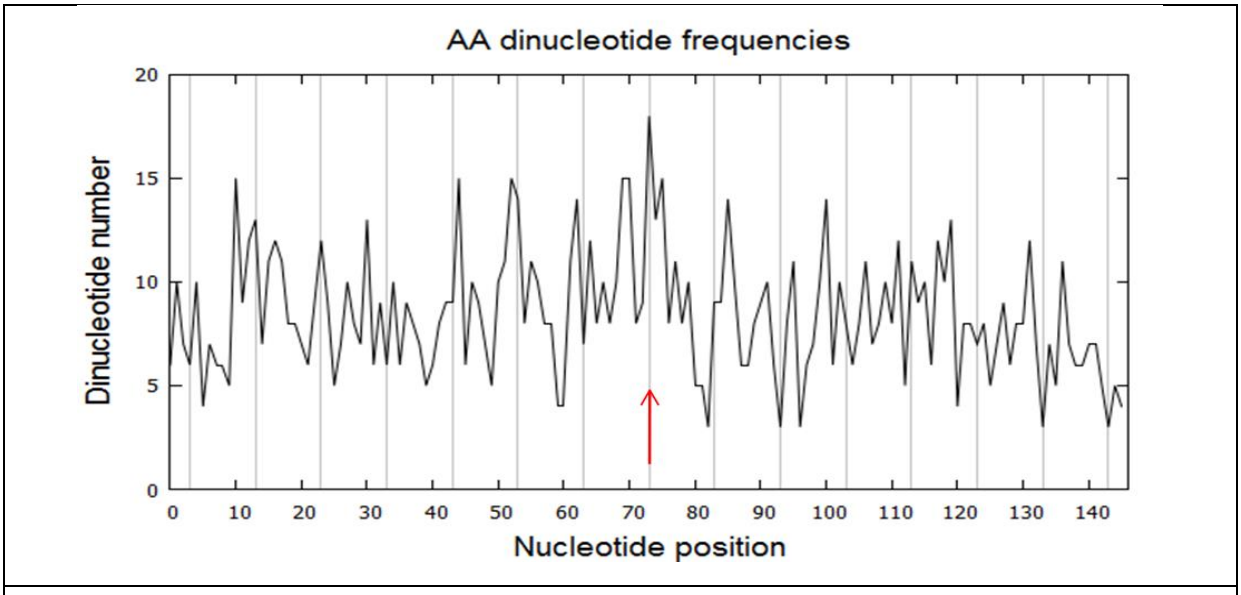


Figure 4.36: AA dinucleotide distribution of the 147 bp sequences of the -1 nucleosome of ORFs. Grey vertical bars are at 10bp intervals, centred at 74 nt, the dyad axis (red arrow). A series of major AA peaks can be observed with a periodicity of ~10 nt.

Prominent AA peaks are visible at positions 10, 23, 43 52, 62, 74, 85, 99, 110, 128 and 130, with the strongest signal on the dyad axis. Although there appears a number of minor peaks between, the major peaks appear to follow a periodical frequency of approximately 10 bp. To investigate the possible periodicity of dinucleotides, we performed a fast Fourier transform (FFT) on the central 128 bp of the dinucleotide distribution data (Figure 4.37).

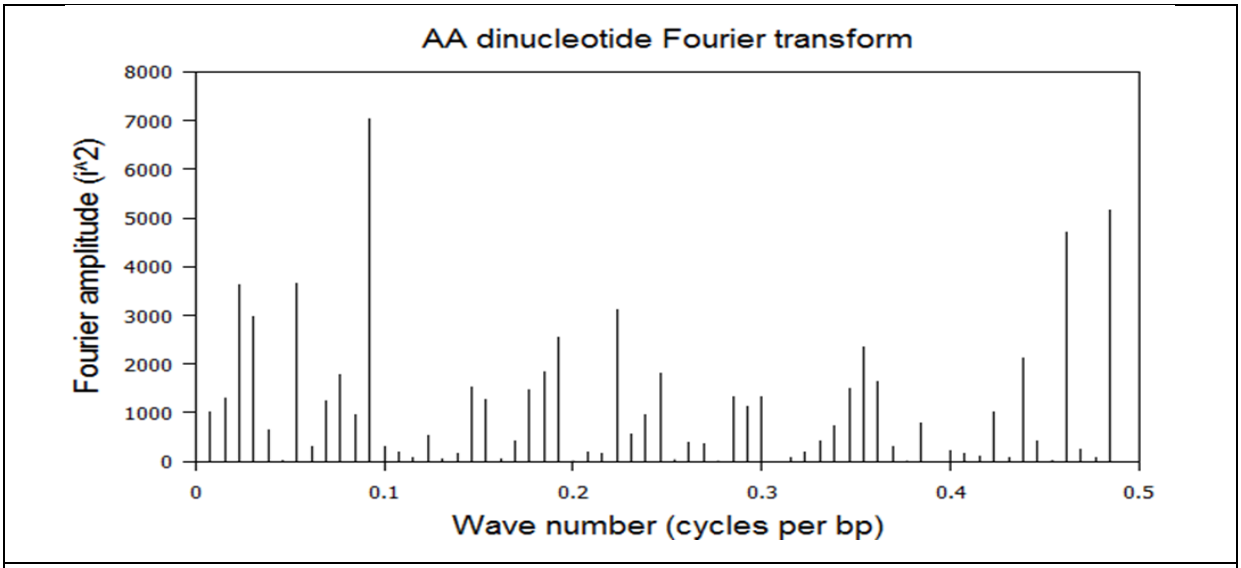


Figure 4.37: Fourier transform analysis reveals a major component of dinucleotide distribution at ~0.1 cycles per bp, corresponding to a dinucleotide periodicity of ~10bp.

Plotting the Fourier amplitude, a prominent component is seen at approximately 0.1 cycles per sampling interval of 1 bp. Therefore, the major component has a period of $(^{1}/_{0.1})$ = 10 bp per cycle. This clearly demonstrates that AA dinucleotides are distributed at an average period of 10 bp in the aligned nucleosomal sequences. This period closely matches that of the nucleosomal DNA, with a base pair periodicity of 10.2 bp per helical turn (corrected for the writhe component of the DNA on the octamer surface) and means that the AA dinucleotide will typically appear on the same side of the helix throughout the span of the nucleosomal DNA. This distribution is in agreement with the distribution observed by Drew and colleagues in chicken nucleosomal DNA [110].

Looking at the distribution of GG dinucleotides (Figure 4.38 A), there are less GG present on the dyad axis (at 74 bp, red arrow), but major peaks are observed at 5, 14, 35, 48, 58, 68, 77, 89,98, 108, 118, 128, 134. Again, the major peak series appear to follow a periodical frequency of approximately 10 bp. To calculate the frequency of the major peaks we performed an FFT of the data.

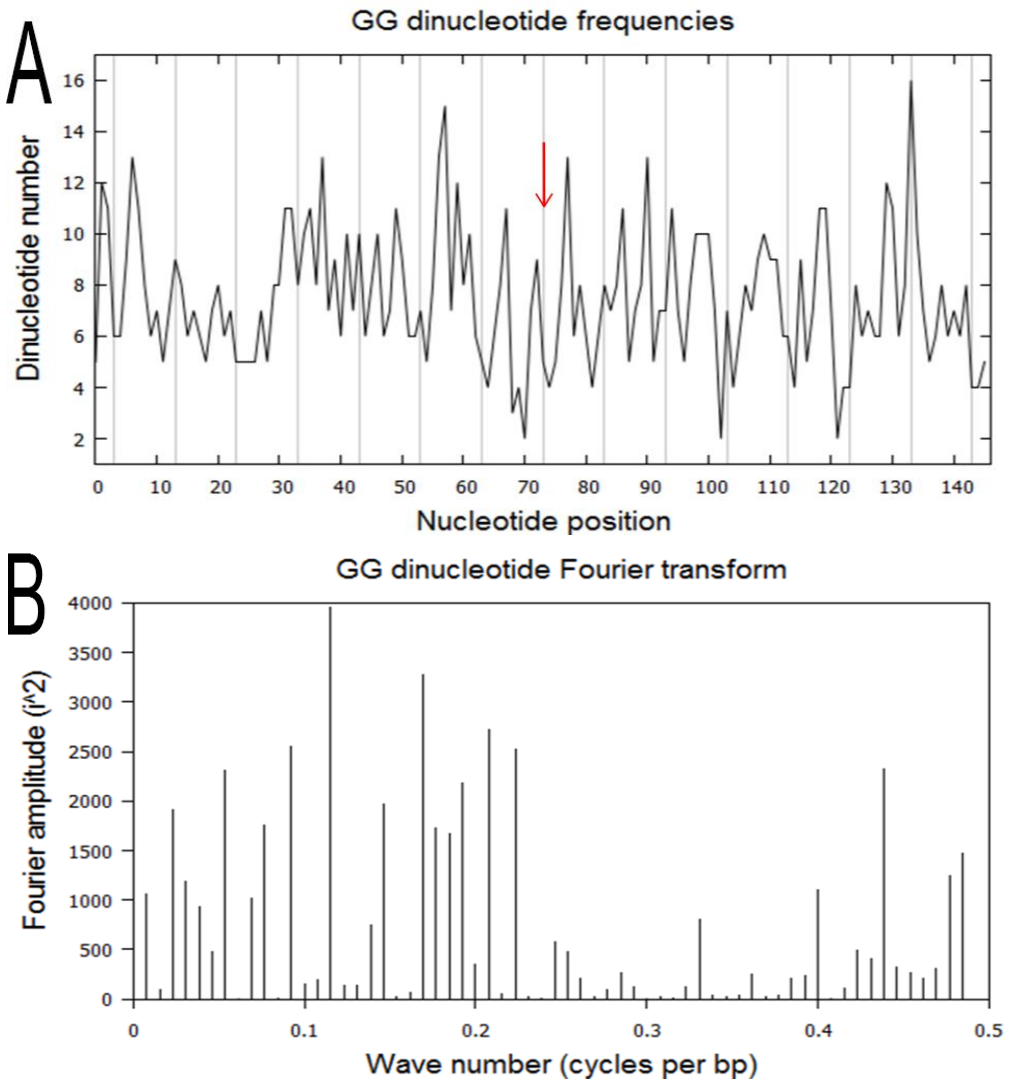


Figure 4.38: GG dinucleotide distribution of the 147 bp sequences of the -1 nucleosome of ORFs (panel A). Grey vertical bars are at 10bp intervals, centred at 74 nt, the dyad axis (red arrow). A series of major GG peaks can be observed with a periodicity of ~10 nt. Fourier transform analysis (panel B) reveals a major component of dinucleotide distribution at ~0.1 cycles per bp, corresponding to a dinucleotide periodicity of ~10bp

A clear peak in the Fourier amplitude is observable at ~ 0.1 cycles per bp, corresponding to a period of 10 bp (Figure 4.38 B). This showed that GG dinucleotides, like AA, are distributed at a periodicity of approximately 10 bp in the nucleosomal DNA sequences. Interestingly, the periodicities of GG and AA dinucleotides appear out of phase by 180°, (corresponding to 5 nucleotides). Therefore, GG dinucleotides appear in regions where AAs are depleted. This is due to the fact that GG cannot be accommodated in a compressed minor groove, but can more easily be accommodated in an expanded minor groove. The GG di-nucleotides are therefore located at positions where the minor groove points out and away from the octamer surface, where the minor groove is expanded, and is exactly out of phase with the distribution of AA. This was also previously observed by Satchwell *et al.* in chicken nucleosomal DNA [110].

Taken together, this data indicates that the strongly positioned -1 nucleosome is located on DNA that may impart a strong rotational frame. Such a strongly, rotationally positioned nucleosome may be related to the exposure of specific cis-elements in the NDR, or may facilitate the interaction of complexes that form on either side of the strongly positioned nucleosome.

2.5) Discussion and Conclusions

The condensation of DNA into chromatin by nucleosomes not only serves in packaging eukaryotic DNA, but also influences gene expression and regulation [127]. The positioning of nucleosomes on DNA is far from random and exhibits species and tissue specific distribution [70]. Indeed, nucleosomes show preferential positioning relative to specific genomic landmarks. However, the positions of nucleosomes on DNA are not fixed and display transient variations within a cell population [104]. This may be caused by several factors such as nucleosome breathing, nucleosome turnover, eviction by a transcribing polymerase, and modification by chromatin remodellers.

It has become increasingly clear that chromatin must play an important role in transcription regulation in *T. brucei* [44]. The co-localization of chromatin binding proteins, histone PTMs and variants with putative TSS and TTS indicates that the epigenome of *T. brucei* may function in genome regulation. A high resolution map of individual nucleosomes was generated by paired-end sequencing of mononucleosomal DNA, obtained from MNase digested chromatin. These short reads (~50 bp) was then mapped to the *T. brucei* 427 Lister genome using bowtie 2 [116]. An overall alignment rate of ~70 % was achieved for all samples.

Bin analysis of the dyad densities on the demarcated features in the genome revealed a lower nucleosomal occupation on pol II transcribed genic regions when compared to intergenic regions (Figure. 4.11). This could be reflective of nucleosome eviction by a transcribing pol II which constitutively transcribes these PTUs. This is in agreement with the decrease in nucleosome occupancy observed upon gene activation in yeast [122]. Dyad distribution on constitutively pol I transcribed *rRNAs* appears comparable with that of genic

regions (Figure 4.12). However, *tRNA* genes appeaed significantly depleted of canonical nucleosomes (Figure 4.13), possibly resulting from pol III transcription factors binding to regulatory *tRNA* sequences and driving nucleosomes into neighbouring areas. Bin analysis also revealed the substantial enrichment of nucleosomes on silent *VSGs* telomere-proximal subtelomeric regions (Figures 4.15, 4.16). Association of these regions with high amounts of nucleosomes may serve as transcriptional repressive structures, consistent with the necessity for transcriptional silence of these genes. This shows that the different features of the MBCs of *T. brucei* are occupied by different levels of nucleosomes, depending on the transcriptional activity and class, as well as abundance and length of these features.

The nucleosomal architecture surrounding pol II PTUs revealed correlations with nucleosomal patterns surrounding similar regions in other eukaryotes. The typical organisation around eukaryotic TSSs, with a NDR flanked by well positioned -1 and +1 nucleosomes, are also seen in *T. brucei*. However, there are three important differences that should be noted: 1) pol II transcribed genes occured as single entities in the majority of eukaryotes, but as long PTUs in *T. brucei*, 2) conserved pol II TSS sequence motifs are absent in *T. brucei* and 3) this organisation appears around the translation start codon which has no transcriptional significance. TSS are demarcated by histone PTMs (H4K10ac) and variants (H2AZ, H2BV) [24]. The co-localization of these epigenetic marks with the bromodomain factor BDF3 could imply a mechanism of pol II recruitment and positioning by nucleosome mediated TSS selection. Figure 5.1 provides an overview of the general nucleosomal architecture around a pol II PTU.

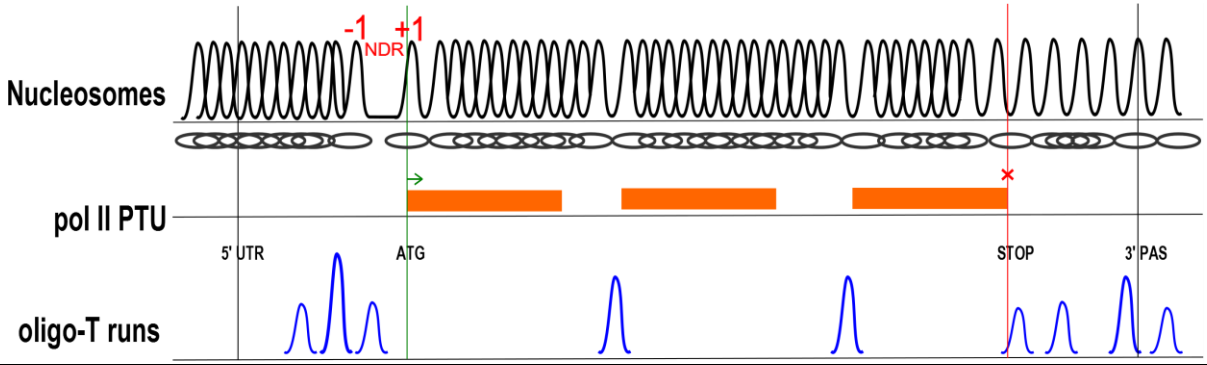


Figure 5.1: Visual representation of the average nucleosomal positions relative to genetic land marks. Also shown is the enrichment of oligo-T runs.

The ATG site occurs just inside of the +1 nucleosome border, which may allow the nucleosome to contribute pol II pausing or abortive transcription. The occurrence of the ATG site within the 5' border of a well-positioned nucleosome is also seen in yeast [113] and *C. elegans* [128], suggesting an evolutionary conservancy of the ATG site being contained within a nucleosome. These patterns appear strikingly similar in the PF and BF trypanosomes.

Covering the translation stop codon of the last ORF in PTUs is a NDR flanked by a -1 and +1 nucleosome. However, there appears an enrichment of nucleosomal dyads covering the 3' PAS, which might be nearer to the true TTS. This positioned nucleosome might exert a transcriptional hindrance resulting in transcription termination by pol II. Pol II TTSs are demarcated by the histone variants H3V and H4V [24]. However, their role in transcriptional termination is unknown.

There appears no noticeable effect of histone H1 on the average nucleosomal distributions around the 5' UTR in either PF or BF trypanosomes (Figure 4.24). However, the NDR located immediately upstream of the ATG codon seemed to shift ~ 150 bp downstream after H1 knock-down in the BF (Figure 4.23), but no

striking difference is observed in PF after H1 knock-down. Similarly, the NDR covering the ORF stop sites again seemed to shift ~150 bp downstream in BFs (Figure 4.25). It seemed that H1 depletion causes the loss / downstream shift of the NDRs covering pol II PTU translation start and stop sites only in BF but not PF trypanosomes. This is in agreement with H1 depletion having a lower effect on global chromatin structure [125]. No remarkable change was observed in nucleosomes surrounding the 3' PAS, although nucleosomes seemed to be less well positioned following H1 knock-down (Figure 4.26)

Nucleosomes surrounding *rRNAs* displayed preferential positioning only when the rRNA genes occurred in pol I transcribed PTUs (of ~3 – 5 genes / PTU). When these *rRNAs* occurred internal to pol II PTUs, no obvious patterning was visible (Figure 4.21). *tRNA* genes appeared devoid of canonical nucleosomes (Figure 4.22). It has been demonstrated that pol III transcription factors associate with the A and B boxes of *tRNAs* in yeast, thereby preventing nucleosome formation on the *tRNA* gene [31]. It is possible that the NDR appearing upstream and extending over the *tRNA* could be caused by pol II transcription factor binding. In yeast, association of a TFIIIB-TFIIIC complex with the A and B boxes of *tRNAs* protects ~ 150 bp of DNA from MNase digestion, mimicking a nucleosome but exhibiting an indicative "bootprint" profile. *T. brucei* does however not possess any detectable TFIIIB homologues, and it is thus unclear whether the slight signals observed preceding *tRNAs* are caused by associated pol III transcription factors or weakly positioned nucleosomes.

We also explored the DNA sequence around the beginning of pol II PTUs for motifs possibly associated with nucleosomal positioning. There is a general enrichment of thymidine residues in the 300 bp preceding the ORF of PTUs

(Figure 4.27) and motif detection revealed the presence of a poly-A/T runs (Figure 4.28). However, this motif, although statistically significant, does not localize to a specific location relative to the ORFs (data not shown). It is an exciting prospect that this motif could reflect putative pol II starting sites, but the correlation between the 5' UTR positions and the locality of this motif requires further investigation.

We observed a negative AT skew preceding the first and all ORFs (Figure 4.31). Analysis of T runs revealed the possibility that *T. brucei* has evolved an asymmetric T-distribution to indicate directionality at promoters and uses oligo-T runs to precisely opposition the -1 nucleosome to leave the putative transcriptional start site free for binding by the transcriptional machinery (Figure 4.33).

Looking at the nucleotide distribution of the DNA associated with the well positioned -1 nucleosome of pol II PTUs, there appears a periodical distribution of A/T and G/C dinucleotides peaking every ~10 bp (Figure 4.36 and 4.38, respectively). Fast Fourier transform revealed a wave periodicity of ~0.1 cycles/bp, corresponding to a periodicity of ~10 bp (Figure 4.37 and 4.38). However, the periodicity of these A/T and G/C dinucleotides are out of phase by 180°, corresponding to 5 bp. This observation is in accordance with the observation by Satchwell and co-workers in chicken nucleosomal DNA patterns [110]. This indicated that *T. brucei* could employ periodical dinucleotides to rotationally position the -1 nucleosome. It also appears that oligo-T runs are used to position the nucleosome covering the PAS (Figure 4.35). It is possible that the ordered nucleosomal pattern observed preceding the PAS are a result of intrinsic DNA sequence preferences and statistical positioning.

Although *T. brucei* displays some uncharacteristic genomic features, the nucleosomal organisation observed around the beginning and end of genes seemed to correlate to those observed in other model eukaryotes. It seems that intrinsic DNA sequences are used to position nucleosomes by directly opposing nucleosome formation at specific sites, probably exposing functional sequence elements on which the PIC can be formed. The precise positioning of the -1 nucleosome by dinucleotide distribution may indicate the need for a well-positioned nucleosome, preceding a NDR and the TSS, which could facilitate nucleosome/epigenetic mediated TSS selection.

Although we excluded data in which dyads have been mapped to tandem repeats, it seems intuitive that these AT rich regions may influence nucleosomal positioning. Therefore, elucidation of the genomic function of tandem repeats calls for future analysis. The influence of the modified thymidine residue, base J, on histone-DNA interaction and nucleosomal positioning is as yet unknown, and requires further investigation.

This study will hopefully provide insight into the unique epigenome of *T. brucei* and further existing knowledge on how this organism employs epigenetic marks to control a genome devoid of canonical transcriptional control.

2.6) References

Doi included where available.

1. Hide G, Tait A (2009) Molecular epidemiology of African sleeping sickness. Parasitology 136: 1491–1500. doi:10.1017/S0031182009990333.

2. Kramer S (2012) Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. Molecular and biochemical parasitology 181: 61–72. doi:10.1016/j.molbiopara.2011.10.002.

3. Carpenter LR, Englund PT (1995) Kinetoplast maxicircle DNA replication in Crithidia fasciculata and Trypanosoma brucei. Molecular and cellular biology 15: 6794–6803.

4. Vargas-Parada L (2010) Kinetoplastids and Their Networks of Interlocked DNA. Nature Education 3: 63.

5. Tiengwe C, Marcello L, Farr H, Gadelha C, Burchmore R, et al. (2012) Identification of ORC1/CDC6-interacting factors in Trypanosoma brucei reveals critical features of origin recognition complex architecture. PloS one 7: e32674. doi:10.1371/journal.pone.0032674.

6. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705. doi:10.1016/j.cell.2007.02.005.

7. Fischle W, Wang Y, Allis CD (2003) Histone and chromatin cross-talk. Current Opinion in Cell Biology 15: 172–183. doi:10.1016/S0955-0674(03)00013-9.

8. Berriman M, Ghedin E, Hertz-fowler C, Blandin G, Bartholomeu DC, et al. (2005) The genome of the African trypanosome Trypanosoma brucei. Science 309: 416–422.

9. Daniels J-P, Gull K, Wickstead B (2010) Cell biology of the trypanosome genome. Microbiology and molecular biology reviews : MMBR 74: 552–569. doi:10.1128/MMBR.00024-10.

10. Ersfeld K (2011) Nuclear architecture , genome and chromatin organisation in Trypanosoma brucei. Research in Microbiologoy 162: 626–636. doi:10.1016/j.resmic.2011.01.014.

11. Wickstead B, Ersfeld K, Gull K (2004) The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome research 14: 1014–1024. doi:10.1101/gr.2227704.

12. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, et al. (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic acids research 38: D457–62. doi:10.1093/nar/gkp851.

13. Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, et al. (2002) A global analysis of Caenorhabditis elegans operons. Nature 417: 851–854. doi:10.1038/nature00831.

14. Martínez-Calvillo S, Yan S, Nguyen D, Fox M (2003) Transcription of Leishmania major Friedlin Chromosome 1 Initiates in Both Directions within a Single Region. Molecular cell 11: 1291–1299.

15. Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, et al. (2010) The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS pathogens 6: e1001090. doi:10.1371/journal.ppat.1001090.

16. Droll D, Minia I, Fadda A, Singh A, Stewart M, et al. (2013) Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. PLoS pathogens 9: e1003286. doi:10.1371/journal.ppat.1003286.

17. Kelly S, Kramer S, Schwede A, Maini PK, Gull K, et al. (2012) Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. Open biology 2: 120033. doi:10.1098/rsob.120033.

18. Palenchar JB, Bellofatto V (2006) Gene transcription in trypanosomes. Molecular and biochemical parasitology 146: 135–141. doi:10.1016/j.molbiopara.2005.12.008.

19. Kelly S, Wickstead B, Gull K (2005) An in silico analysis of trypanosomatid RNA polymerases: insights into their unusual transcription. Biochemical Society transactions 33: 1435–1437. doi:10.1042/BST20051435.

20. Günzl A, Bruderer T, Laufer G, Tu L, Chung H, et al. (2003) RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei. Eukaryotic cell 2: 542–551. doi:10.1128/EC.2.3.542.

21. Cross GAM (1975) Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology 71: 393–417. doi:http://dx.doi.org/10.1017/S003118200004717X.

22. Luo H, Gilinger G, Mukherjee D, Bellofatto V (1999) Transcription Initiation at the TATA-less Spliced Leader RNA Gene Promoter Requires at Least Two DNA-binding Proteins and a Tripartite Architecture That Includes an Initiator Element. Journal of Biological Chemistry 274: 31947–31954. doi:10.1074/jbc.274.45.31947.

23. Ruan J, Arhin GK, Ullu E, Tschudi C (2004) Functional Characterization of a Trypanosoma brucei TATA-Binding Protein-Related Factor Points to a Universal Regulator of Transcription in Trypanosomes Functional Characterization of a Trypanosoma brucei TATA-Binding Protein-Related Factor Points to a Un. Molecular and cellular biology 24: 9610–9618. doi:10.1128/MCB.24.21.9610-9618.2004.

24. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, et al. (2009) Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes & development 23: 1063–1076. doi:10.1101/gad.1790409.

25. Liang X, Haritan A, Uliel S, Michaeli S (2003) trans and cis Splicing in Trypanosomatids : Mechanism , Factors , and Regulation. Eukaryotic cell 2: 830–840. doi:10.1128/EC.2.5.830–840.2003.

26. Das A, Zhang Q, Palenchar JB, Cross GAM, Bellofatto V, et al. (2005) Trypanosomal TBP Functions with the Multisubunit Transcription Factor tSNAP To Direct Spliced-Leader RNA Gene Expression Trypanosomal TBP Functions with the Multisubunit Transcription Factor tSNAP To Direct Spliced-Leader RNA Gene Expression. Molecular and cellular biology 25: 7314–7322. doi::10.1128/MCB.25.16.7314–7322.2005.

27. Schimanski B, Nguyen T, Günzl A (2005) Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Molecular and cellular … 25: 7303–7313. doi:10.1128/MCB.25.16.7303-7313.2005.

28. Talbert PB, Henikoff S (2009) Chromatin-based transcriptional punctuation. Genes & development 23: 1037–1041. doi:10.1101/gad.1806409.

29. Gilinger G, Bellofatto V (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic acids research 29: 1556–1564.

30. McAndrew M, Graham S, Hartmann C, 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. Experimental parasitology 76: 65–76.

31. Nagarajavel V, Iben JR, Howard BH, Maraia RJ, Clark DJ (2013) Global “bootprinting” reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex in vivo. Nucleic acids research 41: 8135–8143. doi:10.1093/nar/gkt611.

32. Hull MW, Erickson J, Johnston M, Engelke DR (1994) tRNA Genes as Transcriptional Repressor Elements. Molecular and cellular biology 14: 1266–1277. doi:10.1128/MCB.14.2.1266.

33. Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136: 629–641. doi:10.1016/j.cell.2009.02.006.

34. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annual review of biochemistry 81: 145–166. doi:10.1146/annurev-biochem-051410-092902.

35. Lee JT (2012) Epigenetic regulation by long noncoding RNAs. Science (New York, NY) 338: 1435–1439. doi:10.1126/science.1231776.

36. Sheader K, Berberof M, Isobe T, Borst P, Rudenko G (2003) Delineation of the regulated Variant Surface Glycoprotein gene expression site domain of Trypanosoma brucei. Molecular and Biochemical Parasitology 128: 147–156. doi:10.1016/S0166-6851(03)00056-2.

37. Hertz-Fowler C, Figueiredo LM, Quail M a, Becker M, Jackson A, et al. (2008) Telomeric expression sites are highly conserved in Trypanosoma brucei. PloS one 3: e3527. doi:10.1371/journal.pone.0003527.

38. Navarro M, Gull K (2001) A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. Nature 414: 759–763. doi:10.1038/414759a.

39. Stanne TM, Rudenko G (2010) Active VSG expression sites in Trypanosoma brucei are depleted of nucleosomes. Eukaryotic cell 9: 136–147. doi:10.1128/EC.00281-09.

40. Narayanan MS, Rudenko G (2013) TDP1 is an HMG chromatin protein facilitating RNA polymerase I transcription in African trypanosomes. Nucleic acids research 41: 2981–2992. doi:10.1093/nar/gks1469.

41. Stros M (2010) HMGB proteins: interactions with DNA and chromatin. Biochimica et biophysica acta 1799: 101–113. doi:10.1016/j.bbagrm.2009.09.008.

42. Reeves R (2010) Nuclear functions of the HMG proteins. Biochimica et biophysica acta 1799: 3–14. doi:10.1016/j.bbagrm.2009.09.001.

43. Yang X, Figueiredo LML, Espinal A, Okubo E, Li B (2009) RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in Trypanosoma brucei. Cell 137: 99–109. doi:10.1016/j.cell.2009.01.037.

44. Rudenko G (2010) Epigenetics and transcriptional control in African trypanosomes. Essays in biochemistry 48: 201–219. doi:10.1042/bse0480201.

45. Stanne TM, Kushwaha M, Wand M, Taylor JE, Rudenko G (2011) TbISWI regulates multiple polymerase I (Pol I)-transcribed loci and is present at Pol II transcription boundaries in Trypanosoma brucei. Eukaryotic cell 10: 964–976. doi:10.1128/EC.05048-11.

46. Métivier R, Penot G, Hübner MR, Reid G, Brand H, et al. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell 115: 751–763.

47. Wang Q, Kawahara T, Horn D (2010) Histone deacetylases play distinct roles in telomeric VSG expression site silencing in African trypanosomes. Molecular microbiology 77: 1237–1245. doi:10.1111/j.1365-2958.2010.07284.x.

48. De Lange T (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. Genes & development 19: 2100–2110. doi:10.1101/gad.1346005.

49. Li B, Oestreich S, de Lange T (2000) Identification of Human Rap1. Cell 101: 471–483. doi:10.1016/S0092-8674(00)80858-2.

50. Gotta M, Laroche T, Formenton A, Maillet L, Scherthan H, et al. (1996) The Clustering of Telomeres and Colocalization with Rapl, Sir3, and Sir4 Proteins in Wild-Type Saccharomyces cerevisiae. Journal Of Cell Biology 134: 1349–136.

51. Hickman M a, Rusche LN (2010) Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proceedings of the National Academy of Sciences of the United States of America 107: 19384–19389. doi:10.1073/pnas.1006436107.

52. Dreesen O, Cross GAM (2008) Telomere length in Trypanosoma brucei. Experimental Parasitology 118: 103–110.

53. Lamont GS, Tucker RS, Cross GAM (1986) Analysis of antigen switching rates in Trypanosoma brucei. Parasitology 92: 355–367. doi:http://dx.doi.org/10.1017/S003118200006412X.

54. Robinson NP, Burman N, Melville SE, Barry JD (1999) Predominance of Duplicative VSG Gene Conversion in Antigenic Variation in African Trypanosomes Predominance of Duplicative VSG Gene Conversion in Antigenic Variation in African Trypanosomes. Molecular and cellular biology 19: 5839–5846.

55. Hovel-Miner G a, Boothroyd CE, Mugnier M, Dreesen O, Cross GAM, et al. (2012) Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. PLoS pathogens 8: e1002900. doi:10.1371/journal.ppat.1002900.

56. Leeuwen F van, Kieft R, Cross M, Borst P (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by β-d-glucosyl-hydroxymethyluracil in Trypanosoma brucei. Molecular and biochemical … 109: 133–145.

57. Cliffe LJ, Siegel TN, Marshall M, Cross GAM, Sabatini R (2010) Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of Trypanosoma brucei. Nucleic acids research 38: 3923–3935. doi:10.1093/nar/gkq146.

58. Van Leeuwen F, Wijsman ER, Kieft R, van der Marel GA, van Boom JH, et al. (1997) Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. Genes & Development 11: 3232–3241. doi:10.1101/gad.11.23.3232.

59. Genest P-A, ter Riet B, Dumas C, Papadopoulou B, van Luenen HGAM, et al. (2005) Formation of linear inverted repeat amplicons following targeting of an essential gene in Leishmania. Nucleic acids research 33: 1699–1709. doi:10.1093/nar/gki304.

60. Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, et al. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic acids research 37: 1452–1462. doi:10.1093/nar/gkn1067.

61. Van Luenen HGAM, Farris C, Jan S, Genest P, Tripathi P, et al. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell 150: 909–921. doi:10.1016/j.cell.2012.07.030.

62. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M (2010) Eukaryotic chromosome DNA replication: where, when, and how? Annual review of biochemistry 79: 89–130. doi:10.1146/annurev.biochem.052308.103205.

63. Tiengwe C, Marcello L, Farr H, Dickens N, Kelly S, et al. (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei. Cell reports 2: 185–197. doi:10.1016/j.celrep.2012.06.007.

64. Benmerzouga I, Concepción-Acevedo J, Kim H-S, Vandoros A V, Cross G a M, et al. (2013) Trypanosoma brucei Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. Molecular microbiology 87: 196–210. doi:10.1111/mmi.12093.

65. Lee W, Tillo D, Bray N, Morse RH, Davis RW, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nature genetics 39: 1235–1244. doi:10.1038/ng2117.

66. Brogaard K, Xi L, Wang J-P, Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. Nature 486: 496–501. doi:10.1038/nature11142.

67. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, et al. (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome research 18: 1051–1063. doi:10.1101/gr.076463.108.

68. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, et al. (2008) Nucleosome organization in the Drosophila genome. Nature 453: 358–362. doi:10.1038/nature06929.

69. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, et al. (2011) Determinants of nucleosome organization in primary human cells. Nature 474: 516–520. doi:10.1038/nature10002.

70. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nature reviews Genetics 10: 161–172. doi:10.1038/nrg2522.

71. Westenberger SJ, Cui L, Dharia N, Winzeler E, Cui L (2009) Genome-wide nucleosome mapping of Plasmodium falciparum reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. BMC genomics 10: 610. doi:10.1186/1471-2164-10-610.

72. Segal E, Widom J (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. Current opinion in structural biology 19: 65–71. doi:10.1016/j.sbi.2009.01.004.

73. Shen X, Yu L, Weir JW, Gorovsky MA (1995) Linker Histones Are Not Essential and Affect Chromatin Condensation In Vivo. Cell 82: 47–56.

74. Povelones ML, Gluenz E, Dembek M, Gull K, Rudenko G (2012) Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in Trypanosoma brucei. PLoS pathogens 8: e1003010. doi:10.1371/journal.ppat.1003010.

75. Janzen CJ, Fernandez JP, Deng H, Diaz R, Hake SB, et al. (2006) Unusual histone modifications in Trypanosoma brucei. FEBS Letters: 2306–2310.

76. Mandava V, Fernandez JP, Deng H, Janzen CJ, Hake SB, et al. (2007) Histone modifications in Trypanosoma brucei. Molecular and biochemical parasitology 156: 41–50. doi:10.1016/j.molbiopara.2007.07.005.

77. Fischle W, Wang Y, Allis CD (2003) Binary switches and modification cassettes in histone biology and beyond. Nature 425: 475–479. doi:10.1038/nature02017.

78. Wang H, Zhai L, Xu J, Joo H-Y, Jackson S, et al. (2006) Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. Molecular cell 22: 383–394. doi:10.1016/j.molcel.2006.03.035.

79. Figueiredo LM, Cross GAM, Janzen CJ (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. Nature reviews Microbiology 7: 504–513. doi:10.1038/nrmicro2149.

80. Garcia BA, Hake SB, Diaz RL, Kauer M, Morris SA, et al. (2007) Organismal differences in post-translational modifications in histones H3 and H4. The Journal of biological chemistry 282: 7641–7655. doi:10.1074/jbc.M607900200.

81. Janzen CJ, Hake SB, Lowell JE, Cross GAM (2006) Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in Trypanosoma brucei. Molecular cell 23: 497–507. doi:10.1016/j.molcel.2006.06.027.

82. Gassen A, Brechtefeld D, Schandry N, Arteaga-Salas JM, Israel L, et al. (2012) DOT1A-dependent H3K76 methylation is required for replication regulation in Trypanosoma brucei. Nucleic acids research: 1–10. doi:10.1093/nar/gks801.

83. Mandava V, Janzen CJ, Cross GAM (2008) Trypanosome H2Bv replaces H2B in nucleosomes enriched for H3 K4 and K76 trimethylation. Biochemical and biophysical research communications 368: 846–851. doi:10.1016/j.bbrc.2008.01.144.

84. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, et al. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406: 593–599. doi:10.1038/35020506.

85. Siegel TN, Kawahara T, Degrasse JA, Janzen CJ, Horn D, et al. (2008) Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in Trypanosoma brucei. Molecular microbiology 67: 762–771. doi:10.1111/j.1365-2958.2007.06079.x.

86. Winston F, Allis CD (1999) The bromodomain : a chromatin-targeting module ? Nature structural biology 6: 601–604.

87. Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, et al. (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell 111: 369–379.

88. Alsford S, Horn D (2012) Cell-cycle-regulated control of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in Trypanosoma brucei. Nucleic acids research 40: 10150–10160. doi:10.1093/nar/gks813.

89. Denninger V, Fullbrook A, Bessat M, Ersfeld K, Rudenko G (2010) The FACT subunit TbSpt16 is involved in cell cycle specific control of VSG expression sites in Trypanosoma brucei. Molecular Microbiology 78: 459–474. doi:10.1111/j.1365-2958.2010.07350.x.

90. Belotserkovskaya R, Saunders A, Lis JT, Reinberg D (2004) Transcription through chromatin: understanding a complex FACT. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1677: 87–99.

91. Hondele M, Stuwe T, Hassler M, Halbach F, Bowman A, et al. (2013) Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature 499: 111–114. doi:10.1038/nature12242.

92. Lawrence E (2008) Henderson's Dictionary of Biology 14th edt.

93. Narayanan MS, Kushwaha M, Ersfeld K, Fullbrook A, Stanne TM, et al. (2011) NLP is a novel transcription regulator involved in VSG expression site control in Trypanosoma brucei. Nucleic acids research 39: 2018–2031. doi:10.1093/nar/gkq950.

94. Jin C, Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development 21: 1519–1529. doi:10.1101/gad.1547707.

95. Lowell JE, Kaiser F, Janzen CJ, Cross GAM (2005) Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in Trypanosoma brucei. Journal of cell science 118: 5721–5730. doi:10.1242/jcs.02688.

96. Fan JY, Gordon F, Luger K, Hansen JC, Tremethick DJ (2002) The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nature structural biology 9: 172–176. doi:10.1038/nsb767.

97. Lowell JE, Cross GAM (2004) A variant histone H3 is enriched at telomeres in Trypanosoma brucei. Journal of cell science 117: 5937–5947. doi:10.1242/jcs.01515.

98. Vickerman K (1985) Developmental cycles and biology of pathogenic trypanosomes. British medical bulletin 41: 105–114.

99. Ginger ML, Blundell PA, Lewis AM, Browitt A, Günzl A, et al. (2002) Ex Vivo and In Vitro Identification of a Consensus Promoter for VSG Genes Expressed by Metacyclic-Stage Trypanosomes in the Tsetse Fly Ex Vivo and In Vitro Identification of a Consensus Promoter for VSG Genes Expressed by Metacyclic-Stage Trypanosomes in. Eukaryotic cell 1: 1000–1009. doi:10.1128/EC.1.6.1000.

100. Barry JD, McCulloch R (2001) Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite. Advances in parasitology 49: 1–70. doi:10.1016/S0065-308X(01)49037-3.

101. Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? Trends in genetics : TIG 22: 614–620. doi:10.1016/j.tig.2006.08.003.

102. Echeverry MC, Bot C, Obado SO, Taylor MC, Kelly JM (2012) Centromere-associated repeat arrays on Trypanosoma brucei chromosomes are much more extensive than predicted. BMC genomics 13: 29. doi:10.1186/1471-2164-13-29.

103. Simpson L (1998) A base called J. Proceedings of the National Academy of Sciences of the United States of America 95: 2037–2038.

104. Cole HA, Howard BH, Clark DJ (2012) Genome-wide mapping of nucleosomes in yeast using paired-end sequencing. Methods in enzymology 513: 145–168. doi:10.1016/B978-0-12-391938-0.00006-9.

105. Woodcock CL, Skoultchi AI, Fan Y (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology 14: 17–25. doi:10.1007/s10577-005-1024-3.

106. Routh A, Sandin S, Rhodes D (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proceedings of the National Academy of Sciences of the United States of America 105: 8872–8877. doi:10.1073/pnas.0802336105.

107. Bai L, Morozov A V (2010) Gene regulation by nucleosome positioning. Trends in genetics : TIG 26: 476–483. doi:10.1016/j.tig.2010.08.003.

108. Lieb JD, Clarke ND (2005) Control of transcription through intragenic patterns of nucleosome composition. Cell 123: 1187–1190. doi:10.1016/j.cell.2005.12.010.

109. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nature reviews Genetics 10: 161–172. doi:10.1038/nrg2522.

110. Satchwell SC, Drew HR, Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. Journal of molecular biology 191: 659–675.

111. Green MR (2000) TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. Trends in biochemical sciences 25: 59–63.

112. Latchman DS (2008) Eukaryotic Transcription Factors.

113. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, et al. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science (New York, NY) 309: 626–630. doi:10.1126/science.1112178.

114. Lantermann AB, Straub T, Strålfors A, Yuan G-C, Ekwall K, et al. (2010) Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nature structural & molecular biology 17: 251–257. doi:10.1038/nsmb.1741.

115. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, et al. (2008) Nucleosome organization in the Drosophila genome. Nature 453: 358–362. doi:10.1038/nature06929.

116. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357–359. doi:10.1038/nmeth.1923.

117. Hirumi H, Hirumi K (1989) Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. The Journal of parasitology 75: 985–989.

118. Bangs JD, Uyetake L, Brickman MJ, Balber AE, Boothroyd JC (1993) Molecular cloning and cellular localization of a BiP homologue in Trypanosoma brucei. Divergent ER retention signals in a lower eukaryote. Journal of cell science 105 (Pt 4: 1101–1113.

119. Patterton HG, Landel CC, Landsman D, Peterson C, Simpson R (1998) The Biochemical and Phenotypic Characterization of Hho1p, the Putative Linker Histone H1 of Saccharomyces cerevisiae. Journal of Biological Chemistry 273: 7268–7276. doi:10.1074/jbc.273.13.7268.

120. Siegel TNT, Hekstra DRD, Wang X, Dewell S, Cross GAM (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. Nucleic acids … 38: 4946–4957. doi:10.1093/nar/gkq237.

121. Wood V, Gwilliam R, Rajandream M-A, Lyne M, Lyne R, et al. (2002) The genome sequence of Schizosaccharomyces pombe. Nature 415: 871–880. doi:10.1038/nature724.

122. Cole H a, Nagarajavel V, Clark DJ (2012) Perfect and imperfect nucleosome positioning in yeast. Biochimica et biophysica acta 1819: 639–643. doi:10.1016/j.bbagrm.2012.01.008.

123. Juan L-J, Utley RT, Vignali M, Bohm L, Workmann JL (1997) H1-mediated Repression of Transcription Factor Binding to a Stably Positioned Nucleosome. Journal of Biological Chemistry 272: 3635–3640. doi:10.1074/jbc.272.6.3635.

124. Downs JA, Kosmidou E, Morgan A, Jackson SP (2003) Suppression of Homologous Recombination by the Saccharomyces cerevisiae Linker Histone. Molecular Cell 11: 1685–1692. doi:10.1016/S1097-2765(03)00197-7.

125. Povelones ML, Gluenz E, Dembek M, Gull K, Rudenko G (2012) Histone H1 plays a role in heterochromatin formation and VSG expression site silencing in Trypanosoma brucei. PLoS pathogens 8: e1003010. doi:10.1371/journal.ppat.1003010.

126. Kaplan N, Moore I, Fondufe-Mittendorf Y (2008) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458: 362–366. doi:10.1038/nature07667.

127. Bai L, Morozov A V (2010) Gene regulation by nucleosome positioning. Trends in genetics : TIG 26: 476–483. doi:10.1016/j.tig.2010.08.003.

128. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, et al. (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome research 18: 1051–1063. doi:10.1101/gr.076463.108.

2.6) Supplementary material

Chapter 1

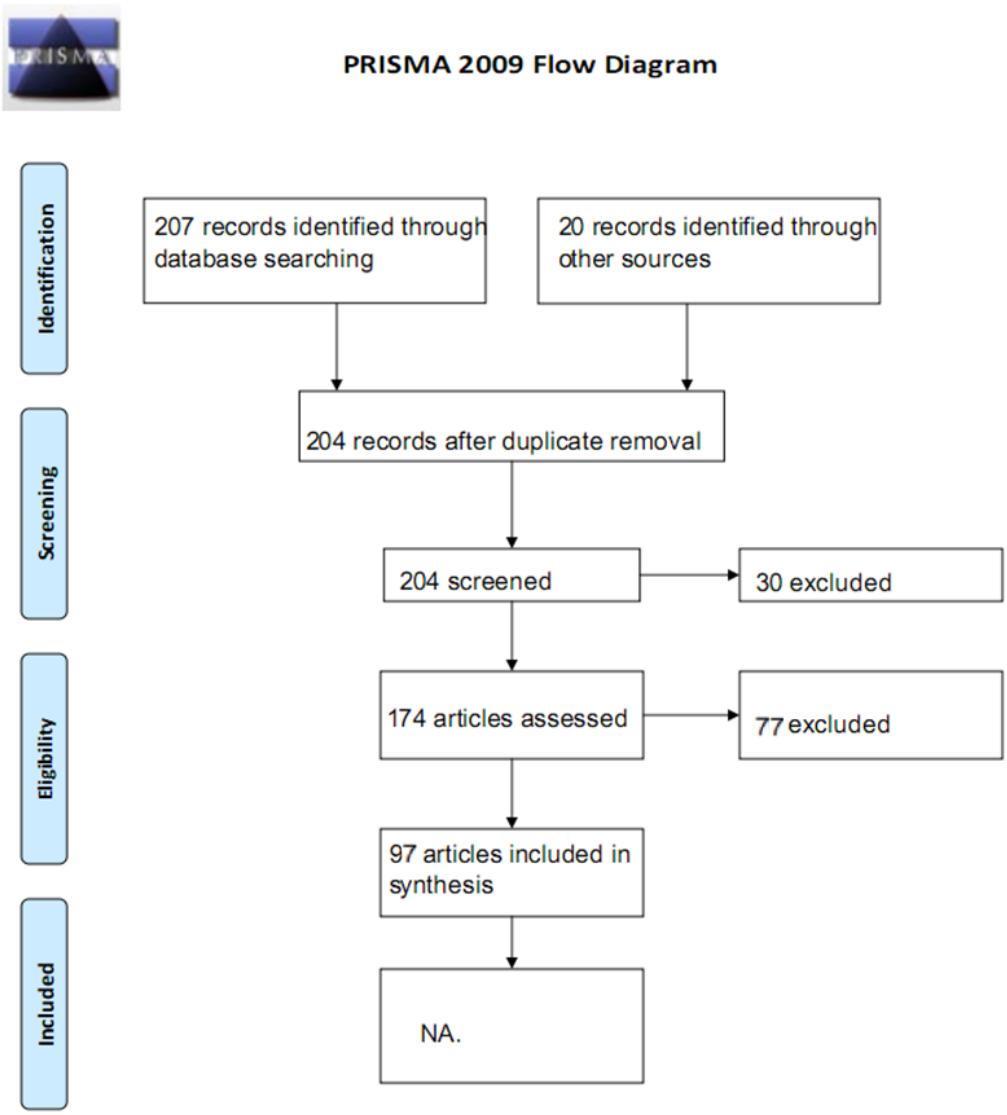


Figure S 1.1: PRISMA work flow diagram indicating article retrieval, filtering and selection.

Polymerase subunits and PIC components

Pol I

Subunits	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity	E – value
RPA190	RPA1 (P10964)	RPA1 (O95602)	Q57UI1	31	1e-167
RPA135	RPA2 (P22138)	RPA2 (Q9H9Y6)	Q387I8	32	3e-131
A49	RPA49 (Q01080)	RPA49 (Q9GZS1)			
A43	RPA43 (P46669)	RPA43 (Q3B726)		Absent (Kelly 2005)	
RPC40	**RPAC (P07703)**	**RPAC1 (O15160**)	**Q580S5**	36	2e-24
A34.5	RPA34 (P47006)	RPA34 (O15446)			
RPB5	**RPAB1 (P20434)**	**RPAB1 (P19388)**	**Q388R5** **Q388R6**	36 // 30	1e-31 // 8e-14
RPB6	**RPAB2 (P20435)**	RPAB2 (P61218)	**Q585J9** **Q387N1**	55 35	5E-31 4E-11
AC19	**RPAC2 (P28000)**	**RPAC2 (Q9Y2S0)**	**Q384D9**	36	9E-16
RPB8	**RPAB3 (P20436)**	**RPAB3 (P52434)**	**Q384M6**	29	4e-10
A14	RPA14 (p50106)	RFA3 (Q549U6)	Q387T8	Absent (Kelly 2005)	
A12.2	RPA12 (P32529)	RPA12 (Q9P1U0)	Q383N7	30	1e-08
RPB12	**RPAB4 (P40422)**	**RPAB4 (P53803)**	**Q1XEE3**	Present (Kelly 2005)	
RPB10	**RPAB5 (P22139)**	**RPAB5 (P62875)**	**Q57XV8**	55	3e-12
RPA31			Q38BP5	Unique to T brucei	

Figure S 1.2: Pol I subunits of *S. cerevisiae, H. sapiens* and subunits we found in *T. brucei* using PSI-BLAST. A14 + A49 sub-units might be substituted by RPB4 and RPB7 paralogues (Kelly 2005). Subunits in bold are shared between polymerases. Indicated in brackets are accession numbers

Protein	Subunits	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity	E –
TBP		**TBP (P13393)**	**TBP (P20226)**	**TRF4 (Q387R9)**	(Ruan 2004)	
Yeast Core Factors	Rrn6	RRN6 (P32786)				
	Rrn7	RRN7 (P40992)				
	Rrn11	RRN11 (Q04712)				
Upstream Activating Factor	Rrn5	RRN5 (Q02983)	KD3A(Q9Y4C1)			
	Rrn9	RRN9 (P53437)		Q57UB6	22	0.02
	Rrn10	RRN10 (P38204)				
	UAF30	UAF30 (Q08747)				
	RRN3	RRN3 (P36070)	RRN3(Q9NYV6)	Q584B3	50	0.42
TIF-IA			TIF-1A(Q9NYV6)			
TIF-IB			TAF1B(Q53T94)			
TIF-IC			TAF1C(Q15572)			
TIF-1D			TAF1D(Q9H5J8)			
UBF /UAF			UBF1(P17480)	Q57XJ9	26	2e-05

Figure S 1.3: Transcription factors associated with pol I PIC formation

Pol II

T. brucei pol II transcribes: *mRNA*, *SLRNA*, *snoRNA*.

S. cerevisiae	*H. sapiens*	*T. brucei*
YSPTSPS	YSPTSP(V/S/T/K)	No obvious pattern of >4 repeats

Figure S 1.4: Pol II hepta-peptide C-terminal repeats.

Sub-units	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity
RPB1	RPB1 9(P04050)	RPB1 (P24928)	Q584K3 / Q57YU2	38 /38
RPB2	RPB2 (P08518)	RPB2 (P30876)	Q583H5	48
RPB3	RPB3 (P16370)	RPB3 (P19387)	Q580S5	29
RPB4	RPB4 (P20433)	RPB4 (O15514)	Q580U8	43
RPB5	**RPAB1 (P20434)**	**RPAB1 (P19388)**	**Q388R5** **Q388R6**	36 // 30
RPB6	**RPAB2 (P20435)**	**RPAB2 (P61218)**	**Q585J9** **Q387N1**	55 35
RPB7	RPB7 (P34087)	RPB7 (P62487)	Q381Y8	26
RPB8	**RPAB3 (P20436)**	**RPAB3 (P52434)**	**Q384M6**	30
RPB9	RPB9 (P27999)	RPA9 (P36954)	Q384T7	32
RPB10	**RPAB5 (P22139)**	**RPAB5 (P62875)**	**Q57XV8**	56
RPB11 *RPB11-a* *RPB11-b* *RPB11-c*	RPB11 (P38902)	POLR2J RPB11 (P52435) RPB1B (Q9GZM3) RPB1C (Q9H1A7)	Q386T9	38 38 38
RPB12	**RPAB4 (P40422)**	**RPAB4 (P53803)**	**Q1XEE3**	(Kelly 2005)

Figure S 1.5: RNA polymerase II sub-units of *S. cerevisiae, H. sapiens* and *T. brucei* psi-BLAST hits. Pol IIA (TbRPB1) + pol IIB (TbRPB2) are non-allelic genes.

POL II complex in *T. brucei* (Ruan 2004; Schimanski 2005, 2006; Lee 2007, 2009)		
Sub-units		Accession
TRF4		Q387R9
TFIIA large		Q599N4
TFIIA small		Q599N3
TFIIB	(70kDa)orthologue	Q38EY4
TbSNAP50	(55kDa)	Q70GM9
TbSNAP3	26kDa Orthologue to human and insect SNAP43	Q599N5
TbSNAP2	46kDa orthologue	Q599N6
TFIIH	10 SU – no-cyclin activating kinase complex	10 SU complex

Figure S 1.6: Components of the pol II PIC of *Trypanosoma brucei* as identified by Ruan *et al.*2004; Schimanski *et al.* 2005, 2006; Lee *et al.* 2007, 2009.

Protein	Subunits	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity	E – value
TFIIA	Large		TF2A (P52655)	Q387R9		
TFIIA	Small		T2AG (P52657)	Q599N4		
TFIIB		TF2B (P29055)	TF2B (Q00403)	Q38EY4	Present (Schimanski 2006)	
TFIID	**TBP**	**TBP (P13393)**	**TBP (P20226)**	**TRF4 (Q387R9)**	Present (Ruan 2004)	
TFIIE	Alpha	T2EA (P36100)	T2EA (P29083)			
TFIIE	Beta	T2EB (P36145)	T2EB (P29084)			
TFIIF	tfg1	T2FA (P41895)	T2FA (P35269)			
	tgf2	T2FB (P41896)	T2FB (P13984)	Q385X9	27 (with yeast)	0.64
	Tgf3	TAF14 (P35189)		Q57ZF8	31	4e-06
TFIIH				Contains a 10 subunit TFIIH - no-cyclin activating kinase complex.		
Human 7 subunits	SU1		TF2H1 (P32780)			
	SU2		TF2H2 (Q13888)	Q57TV9	26	7e-20
	SU3		TF2H3 (Q13889)			
	SU4		TF2H4 (Q92759)	Q38B82	26	2e-16
	SU5		TF2H5 (Q6ZYL4)			
Yeast 7 subunits	RAD3	RAD3 (P06839)		Q57YJ9	41	0.0
	SSL1	SSL1 (Q04673)		Q57TV9	24	8e-31
	SSL2	RAD25 (Q00578)		Q381F9	34	3e-107
	TFB1	TFB1 (P32776)				
	TFB2	TFB2 (Q02939)		Q38B82	24	6e-14
	TFB4	TFB4 (Q12004)				
	TFB5	TFB5 (Q3E7C1)				

Figure S 1.7: Transcription factors associated with Pol II PIC formation

Pol II

T. brucei pol III transcribes *tRNAs*, 5S *RNA*, signal recognition particle RNA (*snRNA*);

Sub-units	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity	E – value
RPC160	RPC1 (P04051)	RPC1 (O14802)	Q38BW7	41	0.0
RPC128	RPC2 (P22276)	RPC2 (Q9NW08)	Q4GZE8	50	0.0
RPC82	RPC3 (P32349)	RPC3 (Q9BUI4)	Q587G4	24	7e-08
RPC52	RPC4 (P25441)	RPC4 (P05423)			
RPC40	**RPAC (P07703)**	**RPAC1 (O15160)**	**Q580S5**	28	2e-23
RPC37	RPC5 (P36121)	RPC5 (Q9NVU0)			
RPC34	RPC6 (P32910)	RPC6 (Q9H1D9)	Q57WC7	22	4e-09
RPC31	RPC7 (P17890)	RPC7 (O15318)			
RPB5	**RPAB1 (P20434)**	**RPAB1 (P19388)**	**Q388R5** **Q388R6**	36 // 30	1e-31 // 8e-14
RPC25	RPC8 (P35718)	RPC8 (Q9Y535)	Q382L2	58	6e-13
RPB6	**RPAB2 (P20435)**	**RPAB2 (P61218)**	**Q585J9** **Q387N1**	67	6E-31 4E-09
RPC 19	**RPAC2 (P28000)**	**RPAC2 (Q9Y2S0)**	**Q384D9**	36	6E-17
RPC17	RPC9 (P47076)	RPC9 (O75575)			
RPB8	**RPAB3 (P20436)**	**RPAB3 (P52434)**	**Q384M6**	30	7e-20
RPC11	RPC10 (Q04307)	RPC10 (Q9Y2Y1)	Q387Z6	33	1e-13
RPB12	**RPAB4 (P40422)**	**RPAB4 (P53803)**	**Q1XEE3**	Present (Kelly 2005)	
RPB10	**RPAB5 (P22139)**	RPAB5 (P62875)	**Q57XV8**	56	3e-22

Figure S 1.8: RNA polymerase III sub-units of *S. cerevisiae, H. sapiens* and *T. brucei* psi-BLAST hits. RPC17 might be substituted by a RPB4 paralogue (Kelly *et al.* 2005)

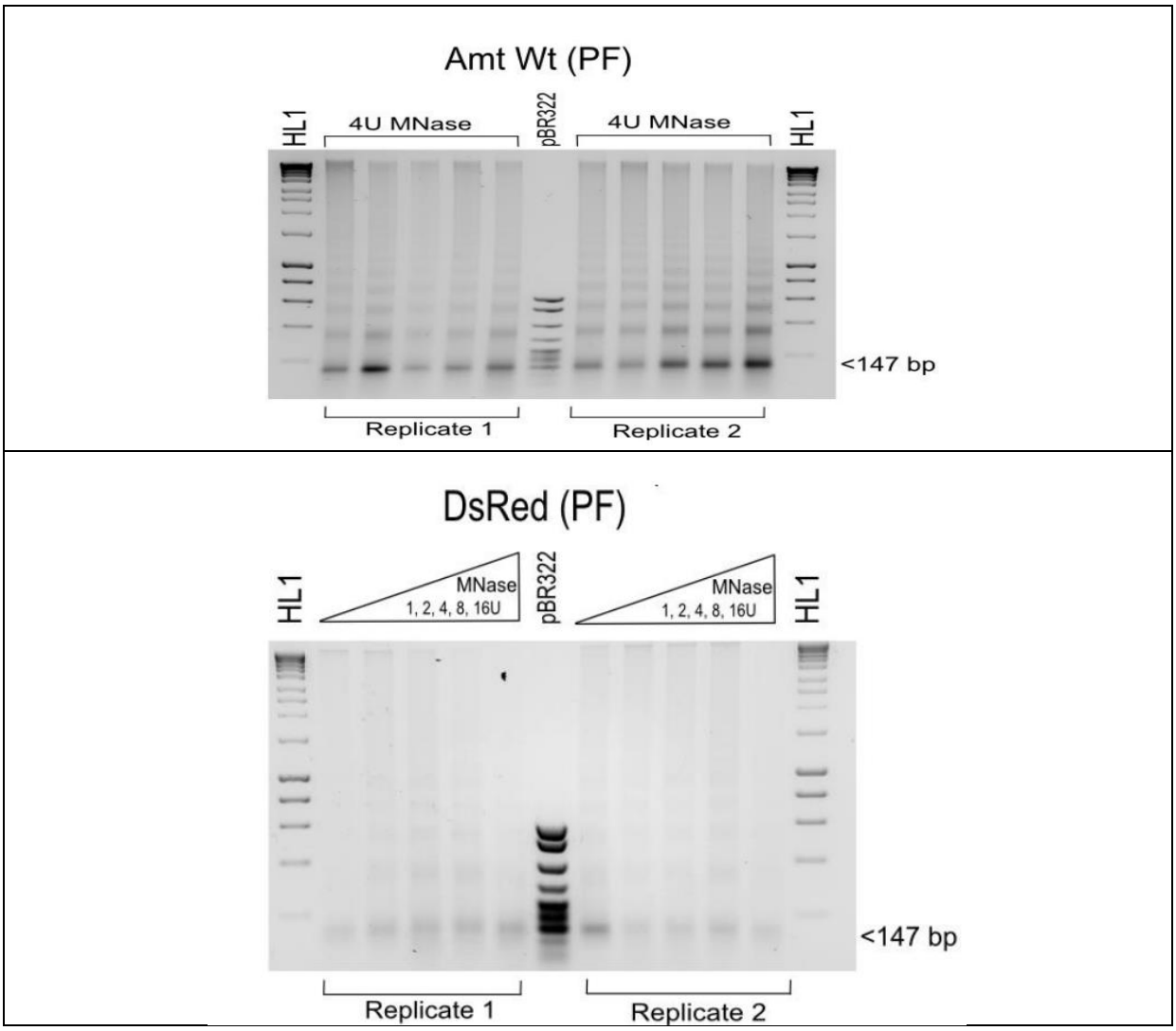
	Subunits	*S. cerevisiae*	*H. sapiens*	*T. brucei*	% Identity	E – value
TFIIIA		TF3A (P39933)	TF3A (Q92664)			
TFIIIB	TBP			Q387K4	TBP of POL III	
	90kDa		TF3B (Q92994)	Q387K4	30	2e-36
	BRF	TF3B (P29056)		Q387K4	30	3e-35
	Beta	TFC5 (P46678)				
	50kDa		BRF2 (Q9HAW0)			
TFIIIC						
Yeast	TFC1	TFC1 (P32367)				
	TFC3	TFC3 (P34111)				
	TFC4	TFC4 (P33339)		Q4GYJ1	26	0.001
	TFC6	TFC6 (Q06339)				
	TFC7	TFC7 (Q12415)				
	TFC8	TFC8 (Q12308)				
Human	SU1		TF3C1 (Q12789)			
	SU2		TF3C2 (Q8WUA4)			
	SU3		TF3C3 (Q9Y5Q9)	Q4GYJ1	28	2e-11
	SU4		TF3C4 (Q9UKN8)			
	SU5		TF3C5 (Q9Y5Q8)			
	SU6		TF3C6 (Q969F1)			

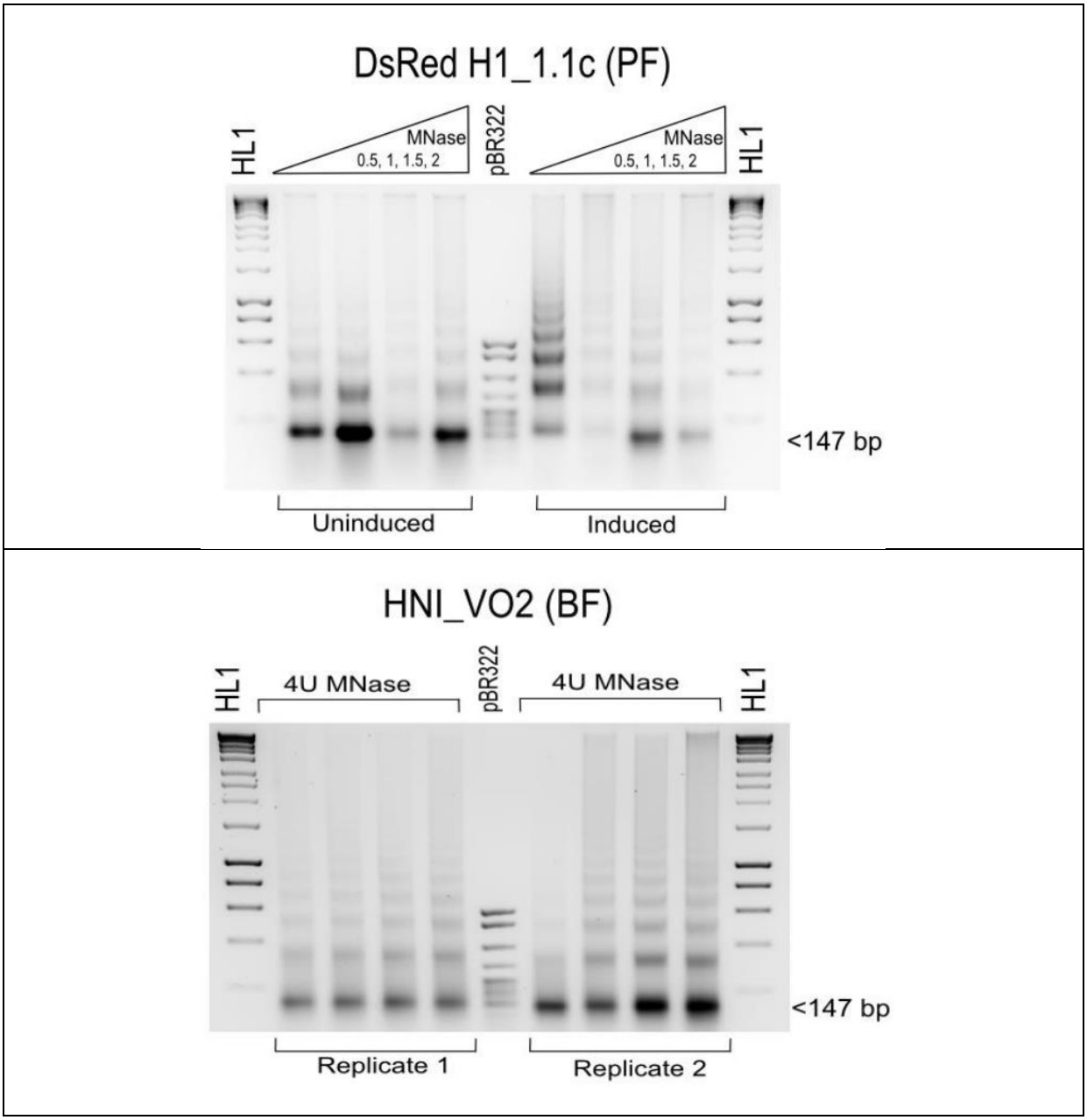
Figure S 1.9: Transcription factors associated with Pol II PIC formation

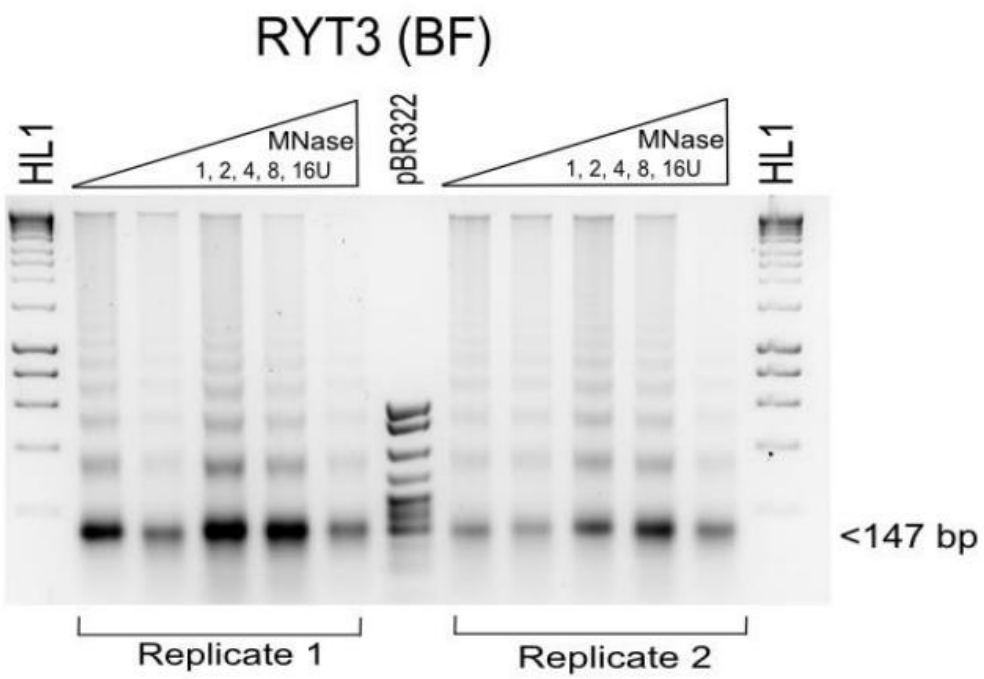
Sequences from all the subunits of each polymerase as well as transcription associated factors identified from literature were obtained using the MRS tool [1] (version 6; http://77.235.253.122:18090/) searching SwissprotDB. These sequences were then used to perform a Position Specific Iterated (PSI) BLAST, using a BLOSUM62 matrix, against the *Trypanosoma brucei* TRUE927 and Lister 427 genomes.

Chapter 2

A listing of the source codes of *dyads_on_genomic_features* (S 2.1), *align_dyads* (S 2.2), *AT_enrichment* (S 2.3), *polyA_enrich* (S 2.4) and *dinucleotide_distribution* (S 2.5) are provided in text (.txt) format on the enclosed CD.




DsRed H1_1.1c (PF)
HL1
MNase
0.5, 1, 1.5, 2
pBR322
MNase
0.5, 1, 1.5, 2
HL1
<147 bp
Uninduced
Induced
HNI_VO2 (BF)
HL1
4U MNase
pBR322
4U MNase
HL1
<147 bp
Replicate 1
Replicate 2



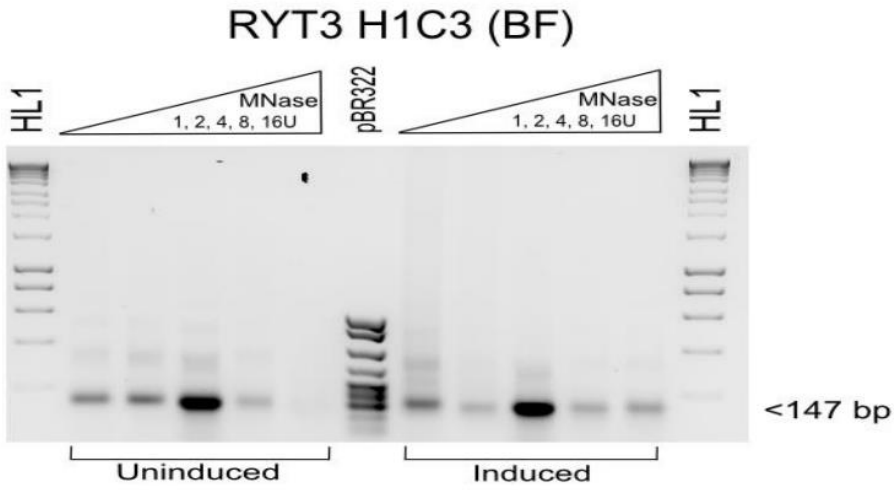


Figure S 2.6: MNase digested chromatin, visualised on a 2% (w/v) agarose. Each image shows DNA markers HL1 = 1 kb Hyperladder (NEB) and pBR322 MspI (NEB), as well as the MNase concentration or titration used. Indicated below the image is either a technical replicate number or RNAi induction. Also indicated are mononucleosomal fragments are indicated (~147 bp). Agarose gels were photographed on glass backplates to minimize UV exposure. DNA bands seem quantitatively equivalent in all samples, except samples which contained the RNAi construct (DsRed H1_1.1c and RYT3 H1C3).

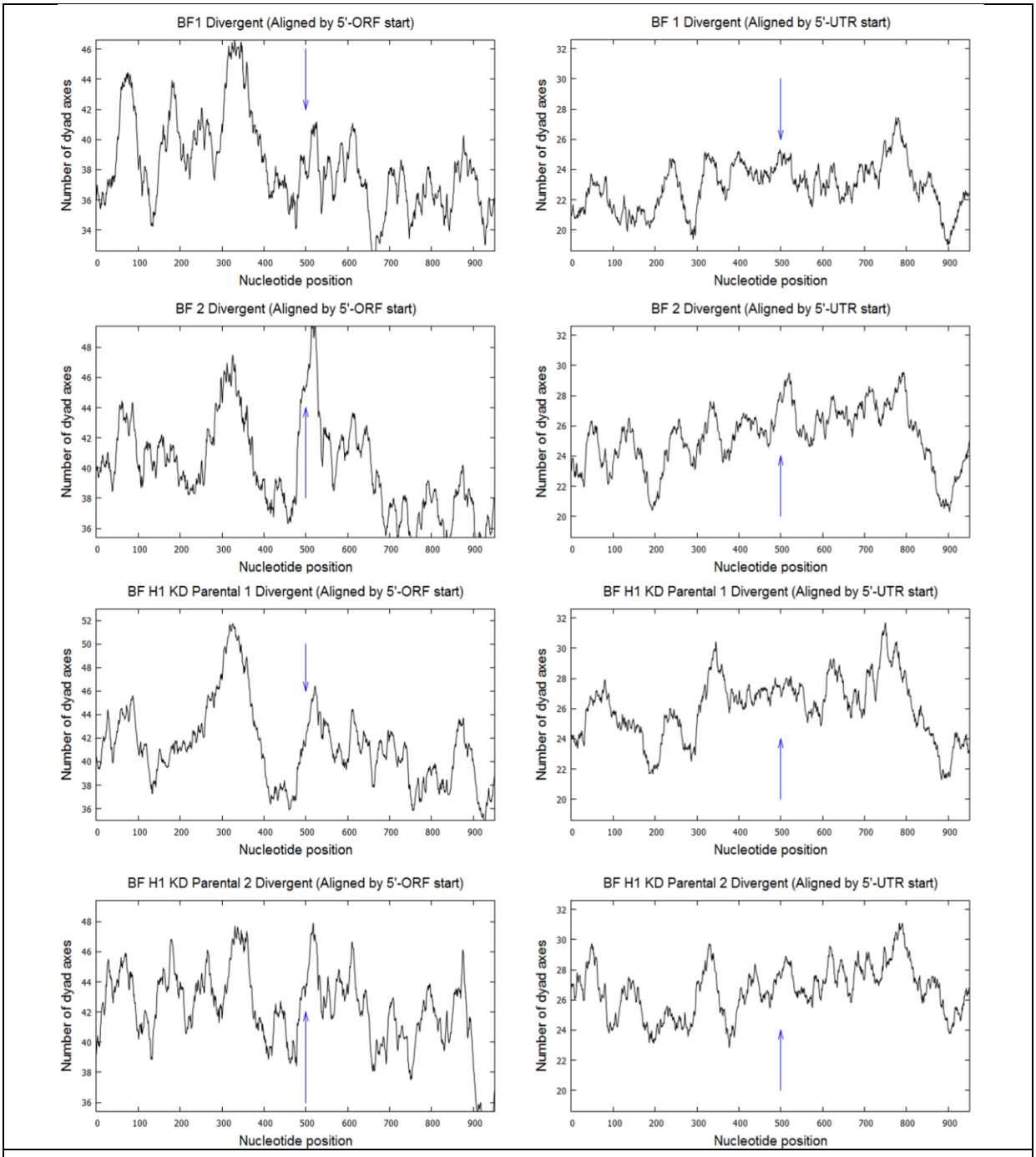


Figure S 2.7: Average dyad distributions in BF trypanosomes, aligned relative to the ORF start (left) or the 5'-UTR start (right) of pol II PTUs. A NDR preceding the ORF start were flanked by areas of dyad enrichment. It seemed as if the -1 nucleosome was better positioned than the +1 nucleosome. Aligned relative to the 5'-UTR, no clear nucleosomal phasing was apparent, except for an area of relative nucleosomal depletion ~ 400 bp downstream of the 5'-UTR.

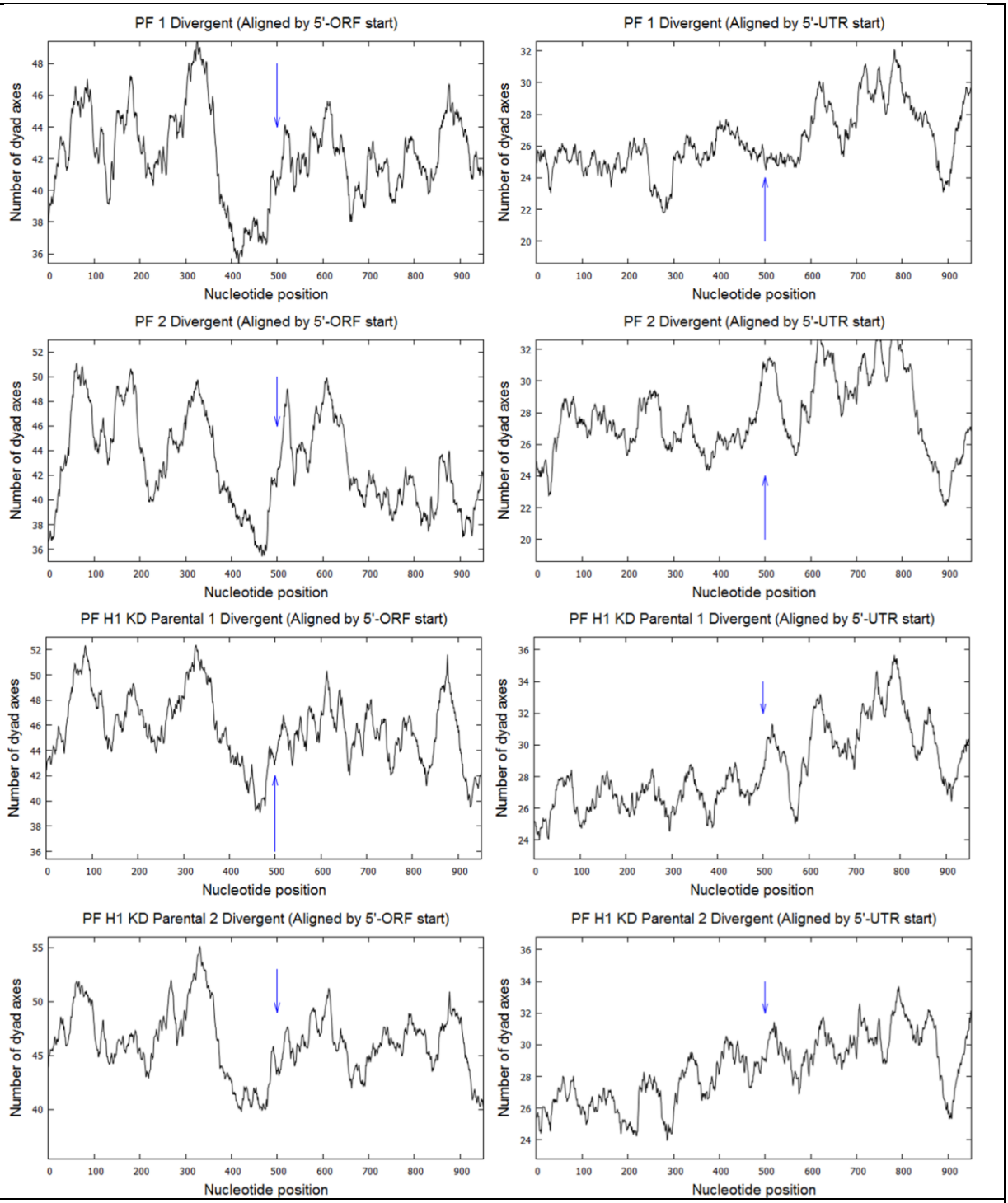


Figure S 2.8: Average dyad distributions of PF trypanosomes, aligned relative to the ORF start (left) or the 5'-UTR start (right). Average nucleosomal distributions appeared comparable to that seen in the BF. Preceding the ORF start was a NDR, bordered by a -1 and +1 nucleosome. The 5'UTR nucleosomes showed no obvious pattern, except for a NDR ~400 bp downstream of the 5'-UTR site.

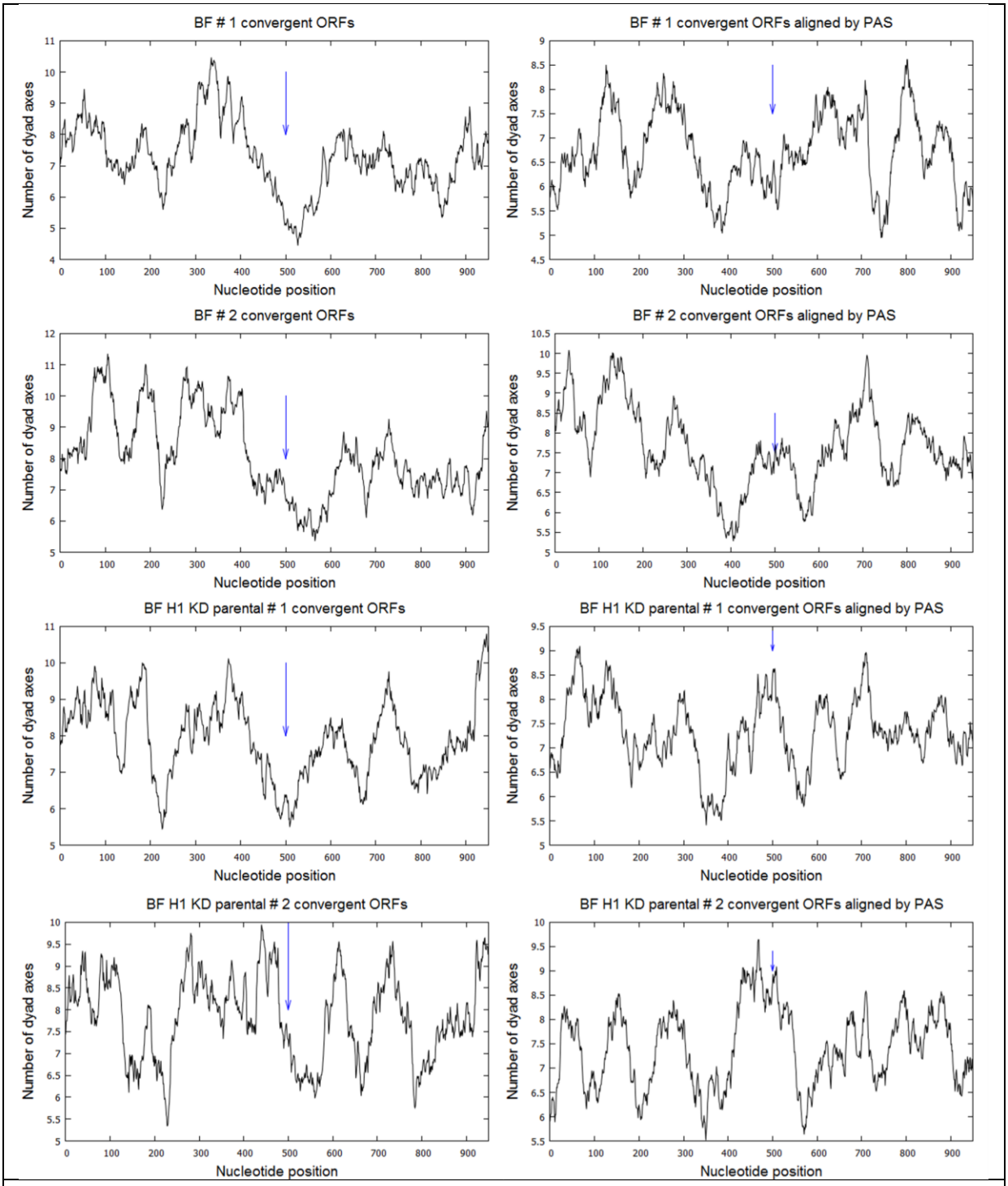


Figure S 2.9: Nucleosomal distribution around the ends of pol II PTUs of BF cells. The ORF stop site appeared devoid of nucleosomes. However, the PAS seemed to be covered by a nucleosome flanked by regions of low occupancy.

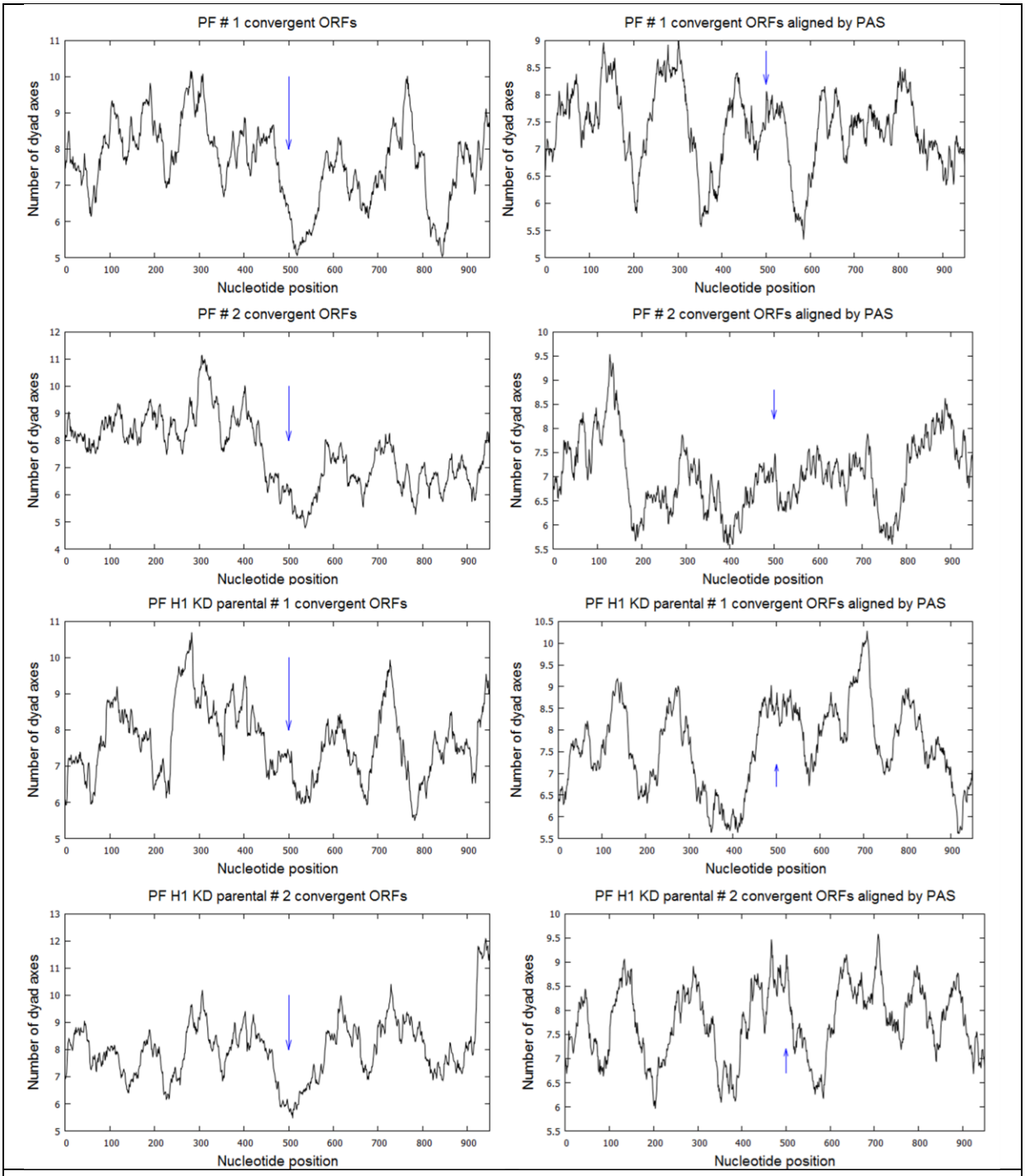


Figure S 2.10: Nucleosomal distribution around the ends of pol II PTUs of PF cells. The ORF stop site appeared devoid of nucleosomes, similar to patterns observed in BF cells. Again, the PAS seemed to be covered by a nucleosome flanked by NDRs.

Summary

Trypanosoma brucei is an extracellular parasite of the mammalian bloodstream that causes African sleeping sickness in humans. Trypanosomes are an ancient branch of the eukaryotic evolutionary lineage, and exhibit some highly unusual transcriptional features. These include the arrangement of functionally unrelated genes in large, pol II transcribed polycistronic transcription units, often exceeding hundreds of kb in size, as well as the transcription of protein coding genes by pol I. Furthermore, no canonical pol II promoters have yet been identified.

Transcription start and termination sites have been shown to be enriched with many epigenetic markers, like histone PTMs and variants. Recent advances have revealed a wide range of epigenetic signals, readers, writers and erasers in trypanosomes, some of which have been observed to be unique to trypanosomes. The epigenome represents a major regulatory interface to the eukaryotic genome. Nucleosome positions, histone variants, histone modifications and chromatin associated proteins all play a role in the epigenetic regulation of DNA function.

Epigenetic marks are often associated with nucleosomes, the basic structural unit of chromatin. Wrapping of 168 bp of DNA in two negative supercoils around a histone octamer surface facilitates the compaction of the poly-anionic DNA molecules to a level where it can fit into a cell nucleus; nucleosomes also serve as dynamic binding surfaces for proteins involved in gene regulation and transcriptional control. Genome-wide maps of the nucleosome organization in model organisms have showed a common arrangement of nucleosomes at specific genomic features, and have provided valuable information on nucleosome mediated gene regulation.

In this study, we have produced a map of nucleosomes covering the megabase chromosomes in both procyclic and bloodstream form *T. brucei*. This was achieved by the MNase digestion of chromatin, followed by paired-end sequencing of mononucleosomal DNA fragments. The fragments were realigned to the reference genome, and nucleosome positions determined. This revealed that the nucleosomal architectures of pol II transcribed PTUs were comparable to that of other model eukaryotes. Remarkably, there appears to be no significant difference in nucleosomal architecture between the two life cycles of *T. brucei*. The histone H1 knock-down by RNAi revealed a change in nucleosomal patterns in the BF, but little effect was observed in the PF. Sequence analysis also revealed the use of intrinsic sequence positioning signals to position nucleosomes relative to specific genetic marks. It also appeared that the preferential distributions of A/T and G/C dinucleotides were employed to impart a rotational position on well positioned nucleosomes. This could indicate the need for well positioned nucleosomes at putative transcription start sites, possibly assisting in nucleosome-mediated transcription start site selection.

Keywords

Trypanosome brucei, African sleeping sickness, epigenetics, chromatin, nucleosome, histone, micrococcal nuclease, paired-end sequencing, transcription.

Opsomming

Trypanosoma brucei is 'n ekstrasellulere parasiet van die soogdier bloedstroom wat Afrika slaapsiekte veroorsaak in die mens. Trypanosome is 'n antieke tak van die eukariotiese evolusionêre boom en toon hoogs ongewone transkripsionele einskappe. Dit sluit in die organiseering van die funksioneel onverwante gene in groot, pol II getranskripteerde polysistroniese transkripsie eenhede (PTE), dikwels meer as honderde kb in grootte, sowel as die transkripsie van proteïen koderende gene deur pol I. Daar is tot dusver geen tipiese pol II promotors geïdentifiseer nie.

Transkripsie begin en eind posisies is getoon om verryk te wees met ruim epigenetiese merkers, soos histoon na-translationele modifikasies (NTM) en variante. Onlangse ontwikkelings het 'n wye verskeidenheid van epigenetiese seine, lesers, skrywers en uitveërs in trypanosome aan die lig gebring, waarvan sommige uniek is. Die epigenoom verteenwoordig 'n groot regulerende koppelvlak tot die eukariotiese genoom. Nucleosoom posisies, histoon variante, histoon veranderinge en chromatien geassosieer proteïene speel almal 'n rol in die epigenetiese regulering van DNA funksie.

Epigenetiese merke word dikwels geassosieer met nukleosome, die basiese strukturele eenheid van chromatien. Die opdraai van 168 bp DNA in twee negatiewe superspoelle rondom 'n histoon oktameer oppervlak fasiliteer die kompaktering van die poli-anioniese DNA molekules tot 'n vlak waar dit kan inpas in die sel kern. Nukleosome dien ook as 'n dinamiese binding oppervlak vir proteïene betrokke in geenregulering en transkripsionele beheer.

Genoom-wye kaarte van die nucleosoom organisasie in model organismes toon 'n gemeenskaplike organiseering van nukleosome rondom spesifieke genomiese elemente, en het waardevolle inligting oor die nucleosoom bemiddelde geenregulering verskaf.

In hierdie studie het ons 'n kaart van nukleosome van die mega-chromosome in beide procycliese en bloedstroom vorm *T. brucei* saamgestel. Dit is bereik deur die MNase vertering van chromatien, gevolg deur gepaarde-end volgordebepaling van mononucleosomale DNA-fragmente. Die fragmente is herbelyn na die verwysing genoom, en nucleosoom posisies bepaal. Dit blyk dat die nucleosomale argitektuur van pol II getranskripteerde PTEs vergelykbaar was met dié van ander model eukariote. Dit blyk of daar geen beduidende verskille in die nucleosomale argitektuur tussen die twee lewensiklusse van *T. brucei* is nie. Die uit-klop van histoon H1 deur RNAi het 'n verandering in nucleosomale patrone in die BV getoon, maar wienige invloed is in die PV waargeneem. Volgorde analise het die gebruik van intrinsieke volgorde posisioneerings seine onthul om nukleosome relatief tot spesifieke genetiese punte te posisioneer. Dit het ook geblyk dat die voorkeur verspreiding van A / T en G / C dinucleotides gebruik word om 'n radiale posisie op goed geposisioneerde nukleosome oor te dra. Dit kan dui op die behoefte aan goed geposisioneerde nukleosome op waarskynlike transkripsie begin posisies wat moontlik bydra in nucleosoom-gemedieerde transkripsie begin posisie seleksie.

Sleutelwoorde

Trypanosoma brucei, Afrika slaapiekte, epigenetika, chromatien, nucleosome, histone, MNase, gepaarde-end volgordebepaling, transkripsie.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.