

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

The epigenome of *Trypanosoma brucei*: A regulatory interface to an unconventional transcriptional machine



Johannes P. Maree, Hugh -G. Patterton*

Advanced Biomolecular Research Cluster, University of the Free State, PO Box 339, Bloemfontein 9332, South Africa

ARTICLE INFO

Article history:

Received 21 February 2014

Received in revised form 9 May 2014

Accepted 28 May 2014

Available online 3 June 2014

Keyword:

Epigenetics

Chromatin

Transcription

Gene regulation

African sleeping sickness

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 kilobases 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 modulation of gene activity, particularly that of pol I transcribed genes, 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 mapped to essential genetic functions. Some epigenetic mechanisms have also 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 highlight the utility of epigenetic targets in the possible development of new therapies for human African trypanosomiasis.

© 2014 Elsevier B.V. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Trypanosoma brucei, the causative agent of African sleeping sickness, is an extracellular, flagellated parasite that is transferred to the human host during a blood meal by a *Glossina spp.* fly. 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 form [2], 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 genes [3]. Many of these genes appear to be regulated by epigenetic mechanisms, implicating chromatin in *T. brucei* gene regulation.

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 and associated with histone H1. 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 a molecular message board to genetic processes, and, equally, as a dynamic binding surface for proteins involved in gene regulation and transcriptional control.

Core histones are globular proteins with a characteristic histone fold domain and N-terminal extensions or “tails” emanating from the central fold. An extensive range of post-translational modifications (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 [4]. Extensive studies have shown that histone PTMs can function either singularly or in combination with other PTMs, referred to as histone “cross-talk” [5]. An insight into the organization of nucleosomes in a genome, as well as the distribution of histone variants and the presence of PTMs, is essential to understand the regulatory role of chromatin in genome function.

* Corresponding author. Tel.: +27 51 401 3052.

E-mail address: patterrh@ufs.ac.za (H.-G. Patterton).

The aim of this review is to integrate recent data gathered from the fields of genomics, transcriptomics and proteomics to understand the epigenetic mechanisms that are employed by *T. brucei* to control its gene expression programs.

2. Genome organization

The haploid genome of *T. brucei* is 26–35 Mb in size, depending on the strain [3,6], and is composed of 11 megabase chromosomes (MBC) [7], 1–5 intermediate chromosomes (IC) (300–900 kb), and approximately 100 minichromosomes (MC) (50–150 kb) [8]. MC account for approximately 10% of the nuclear genome, and about half of each MC is composed of 177 bp repeats, as well as silent variable surface glycoprotein (VSG) genes and pseudogenes [9].

The housekeeping portion of the genome, encoded by genes on the MBC, exists as long, non-overlapping, polycistronic transcription units (PTUs). Adjacent PTUs may be located on the same DNA strand, arranged in a head to tail fashion, or on different strands, separated by convergent or divergent strand switching regions (SSRs). The latter refers to the direction of transcription of the adjacent PTUs (Fig. 1). The MBCs contain nearly 8800 non-redundant protein-coding genes, including about 500 pseudogenes [10], organized into unidirectional gene clusters that are interrupted by tRNA, snRNA, siRNA and rRNA genes. It is unusual for protein coding genes to be organized in directional PTUs on a genome wide scale [11] as is observed in *Trypanosoma*. Unlike generic prokaryotic PTUs, which contain genes encoding proteins that act in combination to perform a specific function, genes in trypanosomal polycistrons are usually functionally unrelated. 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, as well as a number of positions internal to PTUs [12,13]. Although all genes in a PTU are constitutively expressed, the location of a gene within the PTU was shown to impact on the transcript level of that gene [14].

3. Transcription in *T. brucei*

The initiation of transcription represents a key 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 leishmanian or trypanosomal protozoa. 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 [15]. Although the *Leishmania major*, *T. cruzi* and *T. brucei* genomes encode all five subunits common to the three classes of RNA polymerases [16], *T. brucei* employs conventional polymerases for alternative functions.

3.1. DNA-dependent RNA 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) [17]. The VSG gene, encoding the BF stage cell surface antigen, is also transcribed by pol I as part of the polycistronic bloodstream expression site, discussed below. Pol I probably allows expression of high levels of a single transcript from a monoallelic transcription unit.

Pol II transcribes the majority of the PTUs, initiating from regions enriched for specific epigenetic features (see Fig. 1). Unlike in other eukaryotes, the *T. brucei* pol II promoter is weakly defined, and lacks a canonical TATA box and initiator sequence [18], although a TBP-like protein, TbTrf4, was identified [19]. Siegel and colleagues reported that oligo[dG]-runs, located between divergent SSRs, may act as an initiator element providing directionality to transcription [20]. 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 [21]). The SL RNA genes contain the only well-defined promoter of a trypanosomal pol II transcribed gene. The promoter structure consists of a bipartite upstream sequence element and an initiator on which a class II pre-initiation complex nucleates, which contain orthologues of known transcription factors necessary for pol II transcription in other model eukaryotes [22–24]. In *L. tarentolae* the SL RNA genes and promoters were shown to be depleted of nucleosomes [25].

McAndrew et al. [26] demonstrated α -amanitin sensitive transcription from a T3 polymerase promoter in *T. brucei* in the presence of T3 polymerase. This was interpreted as the possible opening of the chromatin structure by T3 polymerase, allowing the entry of pol II and subsequent transcription initiation [26]. This, taken together with the lack of identifiable promoter elements and the enrichment of specific histone PTMs and histone variants at SSRs, suggests that epigenetic control mechanisms play a central role in the modulation of pol II transcription initiation and termination in *T. brucei*.

Trypanosomal tRNAs, transcribed with other non-coding RNAs by pol III, are interspersed between pol II PTUs. Since the tRNA gene itself may be in the process of transcription by pol III, or associated with regulatory proteins [27], the presence of a tRNA gene presents a kinetic block to a transcribing pol II, and contributes to the termination of transcription, as seen in other eukaryotes [28]. It is therefore possible that the presence of a tRNA gene may contribute to the termination of pol II transcription at the end of a trypanosomal PTU.

3.2. Long non-coding RNA

In one transcriptomic study, 103 transcripts, ranging in size from 154–2229 bp, were identified that did not possess recognizable coding potential [13]. Long non-coding RNAs (lncRNAs) have been shown to play a critical role in gene regulation in model eukaryotes [29,30]. lncRNAs can act locally or globally as epigenetic regulators, like the

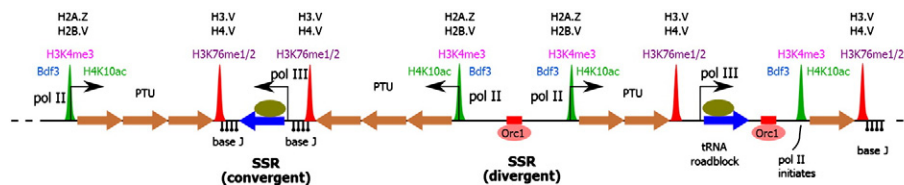


Fig. 1. The epigenetic signals that demarcate transcription units and regulate the expression of genes in *T. brucei*. Pol II transcription initiates from weakly defined promoters in divergent SSRs as well as at some internal positions in PTUs. These initiation loci are enriched for TbBDF3, H4K10ac, H3K4me3 and the H2A.Z and H2B.V 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 H3.V and H4.V histone variants. TTSs often contain an active pol III transcribed tRNA gene. Replication origins, nucleated by TbORC1, occur at the boundaries of PTUs and have an effect on the expression level of adjacent genes.

Xist and HOTAIR lncRNAs, respectively [30,31], affecting DNA–protein interactions, chromatin condensation and gene activity. One can speculate 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.

3.3. 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, from a selection of approximately 1500 VSG genes. The active VSG gene is co-transcribed with a set of expression site associated genes (ESAGs) 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. Telomeres are generally associated with transcriptionally repressive heterochromatic structures, where the repressive effect diminishes with distance from the telomeric ends. The ES promoter is preceded by an array of 50 bp repeat sequences stretching for ~10–50 kb [32]. A total of 14 distinct ESs were identified in the Lister 427 *T. brucei* strain [33], of which the single, active ES is located in a sub-nuclear structure, the expression site body [34]. The canonical structure and associated proteins of the ES and telomeric region are shown in Fig. 2. 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 by pol II, the ESs, as well as the procyclin loci, are subject to tight transcriptional regulation. Research has clearly demonstrated the involvement of epigenetic control mechanisms at 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 [35]. TbTDP1, an HMG box protein, was enriched at active ESs [36]. HMG Box proteins are capable of facilitating chromatin decondensation, thus making chromatin more accessible to regulatory factors, and facilitating the recruitment of transcription activators [37, 38]. TbTDP1 was also enriched at the 50 bp repeats adjacent to ESs and immediately downstream of the rDNA promoter, binding to the entire rDNA 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 24 h. 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 [39,40], 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* [41].

TbISWI also contributed to the down regulation of PF-specific procyclin genes, non-transcribed VSG arrays and minichromosomes. The involvement of chromatin remodellers in gene repression was previously shown in the temporal regulation of the *pS2* gene [42]. SWI/SNF and NuRD were suggested to be required for resetting the local nucleosomal structures to allow transcriptional shutdown of *pS2* in the absence of transcriptional activators [42]. The histone deacetylase, TbHDAC1, antagonized basal telomeric repression in BF cells, and TbHDAC3 was required for VSG ES promoter silencing in both PF and BF cells [43].

In the mammalian telomere complex, TRF2 is bound to duplex telomeric DNA, and serves as a recruitment anchor for another telomeric protein, RAP1 [44,45]. *T. brucei* possesses functional orthologues of both these telomeric proteins, termed TbTRF2 and TbRAP1 [39]. TbRAP1 is found at telomeres, and is essential for growth and critical for ES silencing. Knockdown of TbRAP1 led to a graduated derepression of silent ESs [46]. TbRAP1-mediated silencing increased within the terminal 10 kb of the telomeres, supporting the suggestion that telomere structure were essential for the regulation of VSG expression [39]. It was also shown that telomeric repression decreased with distance from the telomere [39]. In *Saccharomyces cerevisiae*, the telomere repeat binding protein Rap1, together with the Sir proteins, was shown to be required for telomere proximal silencing as well as for position effect variegation [47]. TbSIR2RP1, a Sir2 related protein in *T. brucei*, colocalized with telomeric sequences, and appeared to be involved in the establishment of a silencing gradient at the telomeres in the BF parasite [40]. 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*. This is perhaps not surprising, since Sir3 only appeared in the *S. cerevisiae* genome by gene duplication of Orc1 after evolutionary divergence of the trypanosomes [48]. 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 [49], and switch VSGs more frequently [50,51]. 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 [52].

3.4. 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 kinetoplasts as well as in *Diploleonea* and *Euglena* [53]. 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

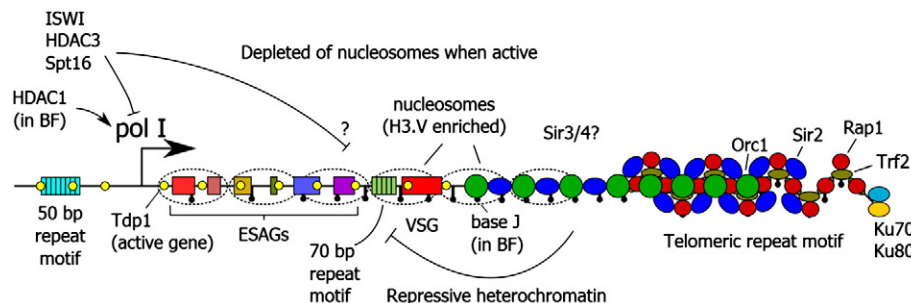


Fig. 2. The epigenetic marks that define the transcriptional state of an ES. A repressive chromatin structure is formed by TbTRF2 and TbRAP1 (which may recruit Sir2) as well as TbORC1, propagating to sub-telomeric regions. It is not known whether other proteins fulfil the roles of yeast Sir3 and Sir4, 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 H3.V, 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 are required for efficient ES silencing, and TbHDAC1 is required for activated expression.

sites (TTS) [54]. Base J was particularly enriched at silent VSG expression sites, forming an increasing gradient towards the telomere termini (Figs. 1 and 2).

Base J is developmentally regulated, and is only found in the BF stage of the *T. brucei* life cycle [55]. 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* [56], *T. brucei* strains in which both enzymes had been knocked-out exhibited no serious growth defects [57].

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 [58]. 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.

3.5. Replication origin complex and gene silencing

DNA replication, similar to transcription, 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 [59]. Genome-wide analysis of TbORC1/CDC6 (subsequently referred to as TbORC1) binding sites in *T. brucei* revealed an overlap between replication origins and the boundaries of PTUs. All functional replication origins occurred in chromosome core regions, associated with transcription initiation and termination [60]. 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 [48]. Surprisingly, TbORC1 does not contain an identifiable BAH domain, but was shown to be required for efficient subtelomeric repression and ES silencing [60–62]. It is not currently known whether all TbORC1-binding sites in *T. brucei* represent active origins, or whether a subset, specifically those located in subtelomeric regions or at silent VSG arrays, function exclusively in gene silencing, similar to *S. cerevisiae* and other yeasts [62]. Indeed, it was shown that the genes that were present adjacent to the boundaries of PTUs, proximal to sites of TbORC1 localization, displayed higher transcript levels following RNAi mediated TbORC1 knock-down [60].

4. Nucleosomal organization

Genome-wide maps of the nucleosome organization in model organisms show a common arrangement of nucleosomes at specific functional locations. A nucleosome free region (NFR), exposing part of the proximal pol II promoter, is seen in yeast [63,64], *Caenorhabditis elegans* [65], *Drosophila* [66], and in humans [67]. The NFR is bordered by two well positioned nucleosomes: –1 on the upstream and +1 on the downstream side of the NFR, followed by a nucleosomal array extending over the gene. High levels of histone variants and post-translational modifications are observed for nucleosomes flanking the NFR [68].

A genome-wide map of nucleosome positions in *T. brucei* has not yet been published. However, looking at the genome-wide map of another protist, *Plasmodium falciparum*, insight into the possible nucleosomal organization of *T. brucei* may be derived. In *P. falciparum* a nucleosomal organization different from other model eukaryotes was seen [69]. Nucleosomes were found to be associated with coding regions and generally absent from intergenic and promoter regions. The high AT-content of the intergenic regions of *Plasmodium* may selectively exclude nucleosomes, allowing easy access to polymerases and associated factors [70]. Indeed, it has been suggested that AT-richness in the *Plasmodium* genome may serve as intrinsic signals for nucleosome positioning [71]. The effect of the many AT-rich repeats, including

those associated with centromeres [72], on nucleosome positioning in the *T. brucei* genome is currently unknown, and may contribute to the epigenetic regulation employed by this kinetoplastid.

5. Histone epigenetic patterns

5.1. 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 arrays and inactive VSG ESs. TbH1 is not only required to down-regulate silent VSG ESs, but may also suppress VSG switching [74].

5.2. 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] (Fig. 3A). 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* H2AK118 and H2AK122 aligns with *Drosophila* H2AK118 and human H2AK119, sites of ubiquitination associated with transcriptional repression [78,79]. Human H2AK125, corresponding to trypanosomal H2AK128, is a possible target for either acetylation or crotonylation [80,81]. However, neither H2AK118 nor K125 appears to be modified in *T. brucei* [76].

5.3. H2B

H2B is the least conserved of the four core histones [82] 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] (Fig. 3B).

Evidence of life stage-dependent modifications is also seen *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] (Fig. 3B).

5.4. 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, identification of PTMs has been complicated by the acetylated N-terminal serine, which blocks Edman degradation, and 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] (Fig. 3C). 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

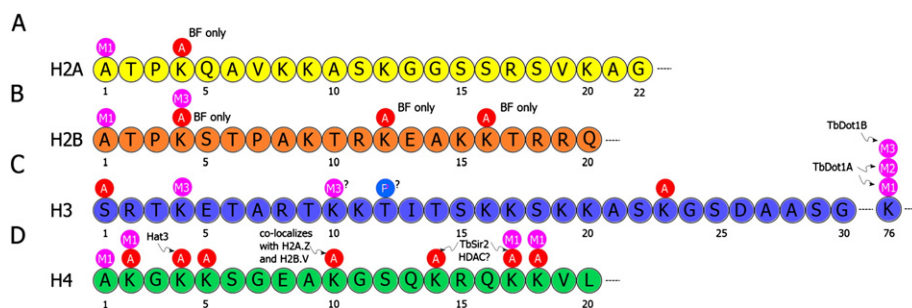


Fig. 3. Epigenetic modifications of the *T. brucei* core histone N-terminal tails. Modifications mapped to specific residues and enzymes involved in the modulation of some modifications are shown. Life stage specific modifications of the parasite are also identified. It is not currently known whether 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.

been functionally described in other organisms [4,83], the functional roles of these PTMs in *T. brucei*, with the exception of K76, is not known.

TbDOT1A is responsible for mono- and di-methylation of H3K76. The RNAi knock-down of TbDOT1A resulted in severe cell cycle defects [84,85]. A clear correlation exists between H3K76 mono- and di-methylation and transcription termination sites (TSSs), suggesting a role in transcription termination [85]. Tri-methylation of H3K76 was mediated by TbDOT1B, which was not essential for viability [85]. Mono-, di-, and tri-methylation of K76 was implicated in several processes, including replication control, antigenic variation, and developmental differentiation [82,85]. K76 di-methylation is only detectable during mitosis [84].

Mandava and co-workers [86] reported that the H2B variant, H2B.V, was present in mononucleosomes enriched for tri-methylated H3K4 and K76, and suggested that H2B.V 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 [87]. 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.

5.5. 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% (Fig. 3D). 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 [88].

ChIP-seq studies showed twin peaks of acetylated H4K10 at divergent SSRs [20]. 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 within a PTU, for example, 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 [20]. Distribution profiles of this modification were remarkably similar between parasite

life stages, with only two life stage-specific peaks being observed (on chromosomes 7 and 11).

The bromodomain factor 3 protein (TbBDF3) was shown to bind to acetylated lysines [89]. It co-localizes 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 [89,90]. TbBDF3 was essential for cellular viability, and RNAi mediated knockdown caused an immediate growth defect where most cells died within 48 h [20].

6. Histone variants

T. brucei encodes four histone variants: H2A.Z, H2B.V, H3.V and H4.V. Nucleosomes that contain H2A.Z are less stable than nucleosomes containing canonical H2A [20,91], and data also suggested that H2A.Z containing chromatin is less condensed, and thus primed for transcription. H2A.Z was shown to be associated exclusively with H2B.V in *T. brucei* [92,93], exhibiting virtually identical ChIP profiles and similar genomic distributions during the cell cycle. Both histone variants were shown to be essential for cell viability [92].

ChIP-seq of H2A.Z and H2B.V revealed a genomic distribution almost indistinguishable from that of acetylated H4K10. Distinct matched peaks were observed at divergent SSR as well as single H2B.V peaks at non-SSR, coincident with that of H4K10ac. H2B.V was also shown to be present in nucleosomes that were enriched for trimethylated H3K4 and H3K76, PTMs typically associated with transcriptionally active chromatin [4].

T. brucei encodes two H3 histones, H3 and the variant H3.V, which shares ~60% identity [94,95]. *T. brucei* appears to lack a centromere-specific CenH3 orthologue [95,96]. H3.V was found to be highly enriched at telomeric repeats and subtelomeric regions (see Fig. 2), but not at the 177 bp minichromosome repeat or 5S rDNA loci [94]. Single peaks of H3.V 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 9–15 guanine residues at SSRs. Generation of H3.V null strains indicated that H3.V is not essential for viability, mini-chromosome segregation, telomere maintenance or transcriptional silencing of ESs and does not contribute to the open chromatin structure observed at the active ES [94,97].

The distribution of H4.V was found to be similar to that of H3.V throughout the genome [20]. H4.V was, however, less enriched compared to H3.V at sub-telomeric and telomeric sites [20]. Both H3.V and H4.V were found to be significantly enriched immediately downstream of the last coding sequence of a PTU (see Fig. 1). This suggested that H3.V–H4.V 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.

7. 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 of the genes encoding the major cell surface proteins, is closely tied to chromatin. In other areas of the genome, promoters as well as transcription stop sites may be determined, not by DNA sequence, but by chromatin structure, which is established by chromatin remodellers, readers, writers, and erasers. 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. 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. This is particularly relevant in the light of increasing resistance to widely used drug combinations reported for *T. brucei* field strains.

8. Where do we go from here?

The sequencing of the *T. brucei* genome has revealed that many epigenetic modifiers, readers, erasers and chromatin remodelling enzymes are encoded by this kinetoplastid [7]. The localization of specific histone variants, modified histones [21] and glycosylated thymidines [49] at sites of pol II transcription initiation and termination is striking. Also, ORC1, implicated in the nucleation of a repressive heterochromatin at the silent mating type loci in *S. cerevisiae*, was shown to be recruited to PTU flanks as well as the silent VSG arrays in *T. brucei*. Thus, there are many bits of enticing data that suggest an important role for chromatin structure and epigenetics in the modulation of gene expression in *T. brucei*. A number of questions that beg for investigation include the role of histone variants and modified histones in defining regions of pol II initiation. By what mechanism are these proteins and modifications enriched at these loci? Although the pol II promoter remains undefined, is there a sequence recognized by a DNA-binding protein that subsequently recruits additional factors and modifiers?

The silent bloodstream ESs show strong evidence of telomeric repression. The role of TbRAP1 and TbORC1 in this process was clearly shown [34,54]. The repressive effect extends beyond the TTAGGG telomeric repeats. This suggests the propagation of a repressive chromatin structure from the telomeric ends, similar to the process seen in yeast and in *Drosophila* [59,60,62]. In yeast the Orc1 paralogue, Sir3, together with Sir4 and the histone deacetylase, Sir2, are required for the propagation of the telomeric repression effect. In *Drosophila* the HP1 protein is involved in heterochromatic structures. What structural proteins contribute to extended telomeric repression in *T. brucei*?

The pursuit of these questions will deepen our understanding of the epigenome of *T. brucei* and the mechanisms by which genome function in this kinetoplastid is modulated. It is likely that expansion of our knowledge in this area would include the identification of novel, epigenetic targets that could form part of new therapies directed at the control of African trypanosomiasis.

Acknowledgements

The authors thank Megan Povelones for critical and useful comments on the manuscript. This study was supported by Award Number 1 U01 HG007465-01 (to HGP) administered by the National Human

Genome Research Institute as part of the NIH Common Fund H3Africa Initiative.

References

- [1] G. Hide, A. Tait, Molecular epidemiology of African sleeping sickness, *Parasitology* 136 (2009) 1491–1500, <http://dx.doi.org/10.1017/S0031182009990333>.
- [2] J.D. Barry, S.V. Graham, M. Fotheringham, V.S. Graham, K. Kobryn, et al., VSG gene control and infectivity strategy of metacyclic stage *Trypanosoma brucei*, *Mol. Biochem. Parasitol.* 91 (1998) 93–105.
- [3] J.-P. Daniels, K. Gull, B. Wickstead, Cell biology of the trypanosome genome, *Microbiol. Mol. Biol. Rev.* 74 (2010) 552–569, <http://dx.doi.org/10.1128/MMBR.00024-10>.
- [4] T. Kouzarides, Chromatin modifications and their function, *Cell* 128 (2007) 693–705, <http://dx.doi.org/10.1016/j.cell.2007.02.005>.
- [5] W. Fischle, Y. Wang, C.D. Allis, Histone and chromatin cross-talk, *Curr. Opin. Cell Biol.* 15 (2003) 172–183, [http://dx.doi.org/10.1016/S0955-0674\(03\)00013-9](http://dx.doi.org/10.1016/S0955-0674(03)00013-9).
- [6] M. Berriman, E. Ghedin, C. Hertz-fowler, G. Blandin, D.C. Bartholomeu, et al., The genome of the African trypanosome *Trypanosoma brucei*, *Science* 309 (2005) 416–422.
- [7] S. Melville, V. Leech, C. Gerrard, The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers, *Mol. Biochem. Parasitol.* 94 (1998) 155–173.
- [8] K. Ersfeld, Nuclear architecture, genome and chromatin organisation in *Trypanosoma brucei*, *Res. Microbiol.* 162 (2011) 626–636, <http://dx.doi.org/10.1016/j.resmic.2011.01.014>.
- [9] B. Wickstead, K. Ersfeld, K. Gull, The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes, *Genome Res.* 14 (2004) 1014–1024, <http://dx.doi.org/10.1101/gr.2227704>.
- [10] M. Aslett, C. Aurrecochea, M. Berriman, J. Brestelli, B.P. Brunk, et al., TriTrypDB: a functional genomic resource for the *Trypanosomatidae*, *Nucleic Acids Res.* 38 (2010) D457–D462, <http://dx.doi.org/10.1093/nar/gkp851>.
- [11] T. Blumenthal, D. Evans, C.D. Link, A. Guffanti, D. Lawson, et al., A global analysis of *Caenorhabditis elegans* operons, *Nature* 417 (2002) 851–854, <http://dx.doi.org/10.1038/nature00831>.
- [12] S. Martínez-Calvillo, S. Yan, D. Nguyen, M. Fox, Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region, *Mol. Cell* 11 (2003) 1291–1299.
- [13] N.G. Kolev, J.B. Franklin, S. Carmi, H. Shi, S. Michaeli, et al., The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution, *PLoS Pathog.* 6 (2010) e1001090, <http://dx.doi.org/10.1371/journal.ppat.1001090>.
- [14] S. Kelly, S. Kramer, A. Schwede, P.K. Maini, K. Gull, et al., Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes, *Open Biol.* 2 (2012) 120033, <http://dx.doi.org/10.1098/rsob.120033>.
- [15] J.B. Palenchar, V. Bellofatto, Gene transcription in trypanosomes, *Mol. Biochem. Parasitol.* 146 (2006) 135–141, <http://dx.doi.org/10.1016/j.molbiopara.2005.12.008>.
- [16] S. Kelly, B. Wickstead, K. Gull, An *in silico* analysis of trypanosomatid RNA polymerases: insights into their unusual transcription, *Biochem. Soc. Trans.* 33 (2005) 1435–1437, <http://dx.doi.org/10.1042/BST20051435>.
- [17] A. Günzl, T. Bruderer, G. Laufer, L. Tu, H. Chung, et al., RNA polymerase I transcribes procyclin genes and Variant Surface Glycoprotein gene expression sites in *Trypanosoma brucei*, *Eukaryot. Cell* 2 (2003) 542–551, <http://dx.doi.org/10.1128/EC.2.3.542>.
- [18] H. Luo, G. Glinger, D. Mukherjee, V. Bellofatto, 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, *J. Biol. Chem.* 274 (1999) 31947–31954, <http://dx.doi.org/10.1074/jbc.274.45.31947>.
- [19] J. Ruan, G.K. Arhin, E. Ullu, C. Tschudi, Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes, *Mol. Cell Biol.* 24 (2004) 9610–9618, <http://dx.doi.org/10.1128/MCB.24.21.9610-9618.2004>.
- [20] T.N. Siegel, D.R. Hekstra, L.E. Kemp, L.M. Figueiredo, J.E. Lowell, et al., Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*, *Genes Dev.* 23 (2009) 1063–1076, <http://dx.doi.org/10.1101/gad.1790409>.
- [21] X. Liang, A. Haritan, S. Uljel, S. Michaeli, Trans and cis splicing in trypanosomatids: mechanism, factors, and regulation, *Eukaryot. Cell* 2 (2003) 830–840, <http://dx.doi.org/10.1128/EC.2.5.830-840.2003>.
- [22] J.H. Lee, H.S. Jung, A. Günzl, Transcriptionally active TFIIB of the early-diverged eukaryote *Trypanosoma brucei* harbors two novel core subunits but not a cyclin-activating kinase complex, *Nucleic Acids Res.* 37 (2009) 3811–3820, <http://dx.doi.org/10.1093/nar/gkp236>.
- [23] J.H. Lee, T.N. Nguyen, B. Schimanski, A. Günzl, Spliced leader RNA gene transcription in *Trypanosoma brucei* requires transcription factor TFIIB, *Eukaryot. Cell* 6 (2007) 641–649, <http://dx.doi.org/10.1128/EC.00411-06>.
- [24] B. Schimanski, J. Brandenburg, T.N. Nguyen, M.J. Caimano, A. Günzl, A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*, *Nucleic Acids Res.* 34 (2006) 1676–1684, <http://dx.doi.org/10.1093/nar/gkl090>.
- [25] R. Hitchcock, S. Thomas, D. Campbell, N.R. Sturm, The promoter and transcribed regions of the *Leishmania tarentolae* spliced leader RNA gene array are devoid of nucleosomes, *BMC Microbiol.* 7 (2007) 44, <http://dx.doi.org/10.1186/1471-2180-7-44>.
- [26] M. McAndrew, S. Graham, C. Hartmann, C. Clayton, Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription, *Exp. Parasitol.* 76 (1998) 65–76.

- [84] C.J. Janzen, S.B. Hake, J.E. Lowell, G.A.M. Cross, Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in *Trypanosoma brucei*, *Mol. Cell* 23 (2006) 497–507, <http://dx.doi.org/10.1016/j.molcel.2006.06.027>.
- [85] A. Gassen, D. Brechtefeld, N. Schandry, J.M. Arteaga-Salas, L. Israel, et al., DOT1A-dependent H3K76 methylation is required for replication regulation in *Trypanosoma brucei*, *Nucleic Acids Res.* (2012) 1–10, <http://dx.doi.org/10.1093/nar/gks801>.
- [86] V. Mandava, C.J. Janzen, G.A.M. Cross, Trypanosome H2Bv replaces H2B in nucleosomes enriched for H3 K4 and K76 trimethylation, *Biochem. Biophys. Res. Commun.* 368 (2008) 846–851, <http://dx.doi.org/10.1016/j.bbrc.2008.01.144>.
- [87] S. Rea, F. Eisenhaber, D. O'Carroll, B.D. Strahl, Z.W. Sun, et al., Regulation of chromatin structure by site-specific histone H3 methyltransferases, *Nature* 406 (2000) 593–599, <http://dx.doi.org/10.1038/35020506>.
- [88] T.N. Siegel, T. Kawahara, J.A. Degrasse, C.J. Janzen, D. Horn, et al., Acetylation of histone H4K4 is cell cycle regulated and mediated by HAT3 in *Trypanosoma brucei*, *Mol. Microbiol.* 67 (2008) 762–771, <http://dx.doi.org/10.1111/j.1365-2958.2007.06079.x>.
- [89] F. Winston, C.D. Allis, The bromodomain: a chromatin-targeting module? *Nat. Struct. Mol. Biol.* 6 (1999) 601–604.
- [90] A.H. Hassan, P. Prochasson, K.E. Neely, S.C. Galasinski, M. Chandy, et al., Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes, *Cell* 111 (2002) 369–379.
- [91] C. Jin, G. Felsenfeld, Nucleosome stability mediated by histone variants H3.3 and H2A.Z, *Genes Dev.* 21 (2007) 1519–1529, <http://dx.doi.org/10.1101/gad.1547707>.
- [92] J.E. Lowell, F. Kaiser, C.J. Janzen, G.A.M. Cross, Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*, *J. Cell Sci.* 118 (2005) 5721–5730, <http://dx.doi.org/10.1242/jcs.02688>.
- [93] J.Y. Fan, F. Gordon, K. Luger, J.C. Hansen, D.J. Tremethick, The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states, *Nat. Struct. Mol. Biol.* 9 (2002) 172–176, <http://dx.doi.org/10.1038/nsb767>.
- [94] J.E. Lowell, G.A.M. Cross, A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*, *J. Cell Sci.* 117 (2004) 5937–5947, <http://dx.doi.org/10.1242/jcs.01515>.
- [95] P.B. Talbert, S. Henikoff, Chromatin-based transcriptional punctuation, *Genes Dev.* 23 (2009) 1037–1041, <http://dx.doi.org/10.1101/gad.1806409>.
- [96] B. Akiyoshi, K. Gull, Discovery of unconventional kinetochores in kinetoplastids, *Cell* 156 (2014) 1247–1258, <http://dx.doi.org/10.1016/j.cell.2014.01.049>.
- [97] L.M. Figueiredo, G.M. Cross, Nucleosomes are depleted at the VSG expression site transcribed by RNA polymerase I in African trypanosomes, *Eukaryot. Cell* 9 (2010) 148–154, <http://dx.doi.org/10.1128/EC.00282-09>.