Molecular & Biochemical Parasitology 172 (2010) 141–144

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

Short communication

Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*

Jessica R. Wright, T. Nicolai Siegel[1], George A.M. Cross*

Laboratory of Molecular Parasitology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

ARTICLE INFO

Article history:
Received 2 March 2010
Accepted 19 March 2010
Available online 27 March 2010

Keywords:
Trypanosoma brucei
Histone modification
Transcription regulation
Chromatin immunoprecipitation
High-throughput DNA sequencing

ABSTRACT

Recent studies have identified histone modifications and suggested a role for epigenetic gene regulation in *Trypanosoma brucei*. The histone modification H4K10ac and histone variants H2AZ and H2BV localize to probable sites of transcription initiation. Although all *T. brucei* histones have very evolutionarily divergent N-terminal tails, histone H3 shows conservation with other eukaryotic organisms in 6 of 8 amino acids encompassing lysine 4. Tri-methylation of H3K4 is generally associated with transcription. We therefore generated a specific antibody to *T. brucei* H3K4me3 and performed chromosome immunoprecipitation and high-throughput sequencing. We show that H3K4me3 is enriched at the start of polycistronic transcription units at divergent strand-switch regions and at other sites of RNA polymerase II transcription reinitiation. H3K4me3 largely co-localizes with H4K10ac, but with a skew towards the upstream side of the H4K10ac peak, suggesting that it is a component of specific nucleosomes that play a role in Pol II transcription initiation.

Trypanosoma brucei branched early in evolution and shows some conservation to higher eukaryotes, but also striking divergences, making it an interesting model for a "simpler" eukaryotic organism. One of the unique differences relates to the initiation of gene transcription. *T. brucei* genes do not have traditional promoter elements. Instead, RNA polymerase II transcribes polycistronic units that become mature monocistronic mRNAs after polyadenylation and *trans*-RNA splicing [1]. Polycistronic transcription units can be convergent or divergent, and divergent strand-switch regions (SSR) have been hypothesized to be transcription start sites (TSS) [2].

In addition to the lack of promoter elements, there appear to be no traditional primary sequences directing RNA Pol II to TSS [2,3]. This suggests that epigenetic phenomena may be key players in *T. brucei* gene regulation. Indeed, several histone modifications have recently been identified and have been implicated in various cellular processes, including life-cycle regulation and antigenic variation [4–11]. Direct evidence that epigenetics plays a role in gene transcription was recently provided by ChIP-sequencing (ChIP-seq), which showed that H4K10 is likely 100% acetylated in all nucleosomes at TSS of *T. brucei* [3]. H4K10ac represents only 10% of the total H4 and its associated acetyltransferase (HAT) is essential, supporting its role as a key player in a regulated cellular process [8]. Histone variants have also been shown to be important players in gene regulation and are often associated with sites of active transcription and more open chromatin conformation [12,13]. *T. brucei* has four histone variants, two of which, H2AZ and H2BV, form a complex *in vivo* [14,15] and co-localize with H4K10ac at TSS [3].

As with gene transcription, the *T. brucei* histone code is highly divergent. An exception is the region surrounding lysine 4 on the H3 N-terminal tail, which shows conservation to higher eukaryotes (Fig. 1, inset (A)). Although several modifications were identified in the H3 N-terminal tail by mass spectrometry, they could not be assigned to a specific amino acid: a blocked N-terminus prevented Edman sequencing and the abundance of lysines and lack of other useful cleavage sites confounded mass spectrometry assignment [4]. Nevertheless, because of sequence conservation and the key role of H3K4 tri-methylation in other organisms, antibodies were made against *T. brucei* H3 N-terminal nonapeptides that were either un-methylated or tri-methylated at K4 [16]. H3K4me3 antibody bound to *T. brucei* H3 and co-immunoprecipitated with H2AZ and H2BV [4,16]. Tri-methylated H3K4 is traditionally associated with sites of active transcription [17], suggesting that this modification could form part of the active nucleosome complex at SSR. Indeed, there is modest enrichment of H3K4me3 at divergent SSR in *Trypanosoma cruzi*, as demonstrated by ChIP-CHIP [18].

To investigate the distribution of H3K4me3 in *T. brucei*, ChIP-grade antibody was generated by affinity purification with positive

Abbreviations: BDF, bromodomain-containing proteins; ChIP, chromatin immunoprecipitation; ChIP-seq, ChIP-sequencing; SSR, strand-switch regions; TSS, transcription start sites.

* Corresponding author. Tel.: +1 212 327 7571; fax: +1 212 327 7572.
E-mail address: george.cross@rockefeller.edu (G.A.M. Cross).
[1] Present address: Unité de Biologie des Interactions Hôte-Parasite, CNRS URA2581, Institut Pasteur, 25 Rue du Dr. Roux, 75724 Paris, France.

0166-6851/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2010.03.013

selection for an H3K4me3 peptide and negative selection against an unmodified H3K4 peptide. As assayed by peptide competition, the resulting antibody showed strong specificity for the H3K4me3 peptide and no reactivity to unmodified H3K4 (data not shown). However, the antibody cross-reacted with tri-methylated H3K76, a known trypanosome modification [4]. When purified on an H3K76me3 column, the resulting antibody concentration was too low for ChIP analysis and no further selectivity was achieved. H3K76 is tri-methylated by the Dot1B enzyme [6] and this modification is not essential. We thus pursued all further experiments using a Dot1B null strain. The purified antibody showed no significant cross reactivity to either H3K76me2 or to tri-methylated H3K32 (Fig. 1, inset (B)).

Using the purified H3K4me3 antibody, we performed ChIP-seq using previously described methodology [3]. Briefly, chromatin was extracted from 1×10^8 Dot1B null cells, cross-linked with formaldehyde, sonicated (Bioruptor, Wolf Laboratories Ltd.: 15 cycles of 30 s on/30 s off, on the highest setting) to generate DNA fragments < 300 bp, then immunoprecipitated with 10 μg of purified H3K4me3 antibody [19]. The resulting DNA was prepared for sequencing by amplification with standard adapters and oligonucleotides (Illumina), and sequenced with a Solexa sequencer (Illumina). DNA sequences were aligned to the trypanosome genome using the BLAT algorithm [20] and enrichment was determined by graphing the number of hits per 100 bp windows for each chromosome (using MATLAB, Mathworks). As the BLAT analysis assigns repetitive regions to the first chromosome hit, these sequences can result in artifactual enrichment peaks. To eliminate this, we excluded repetitive regions from the analysis.

ChIP-seq with the H3K4me3 antibody resulted in a series of peaks of strong enrichment across the genome (Fig. 1). When compared with peaks generated from an H4K10ac antibody [3], there was extensive overlap of each modification, as illustrated in a higher resolution rendering of a portion of chromosome 10 (Fig. 2). The overlap included all divergent SSR, as well as other peaks of H4K10ac enrichment, such as the regions flanked by tRNA genes, which are transcribed by Pol III (Fig. 2). Closer inspection reveals that H3K4me3 has a narrower peak of enrichment than H4K10ac, which extends slightly in the direction of transcription (Fig. 2).

These results strongly suggest that the H3K4me3 histone associates with H4K10ac, H2AZ, and H2BV. The presence of another histone modification at divergent SSR, especially one that is known to correlate with active transcription in other organisms [17], provides further support for the role of several histone modifications in transcriptional activation in *T. brucei*. Each of these modifications has been linked to a more open chromatin structure. As there

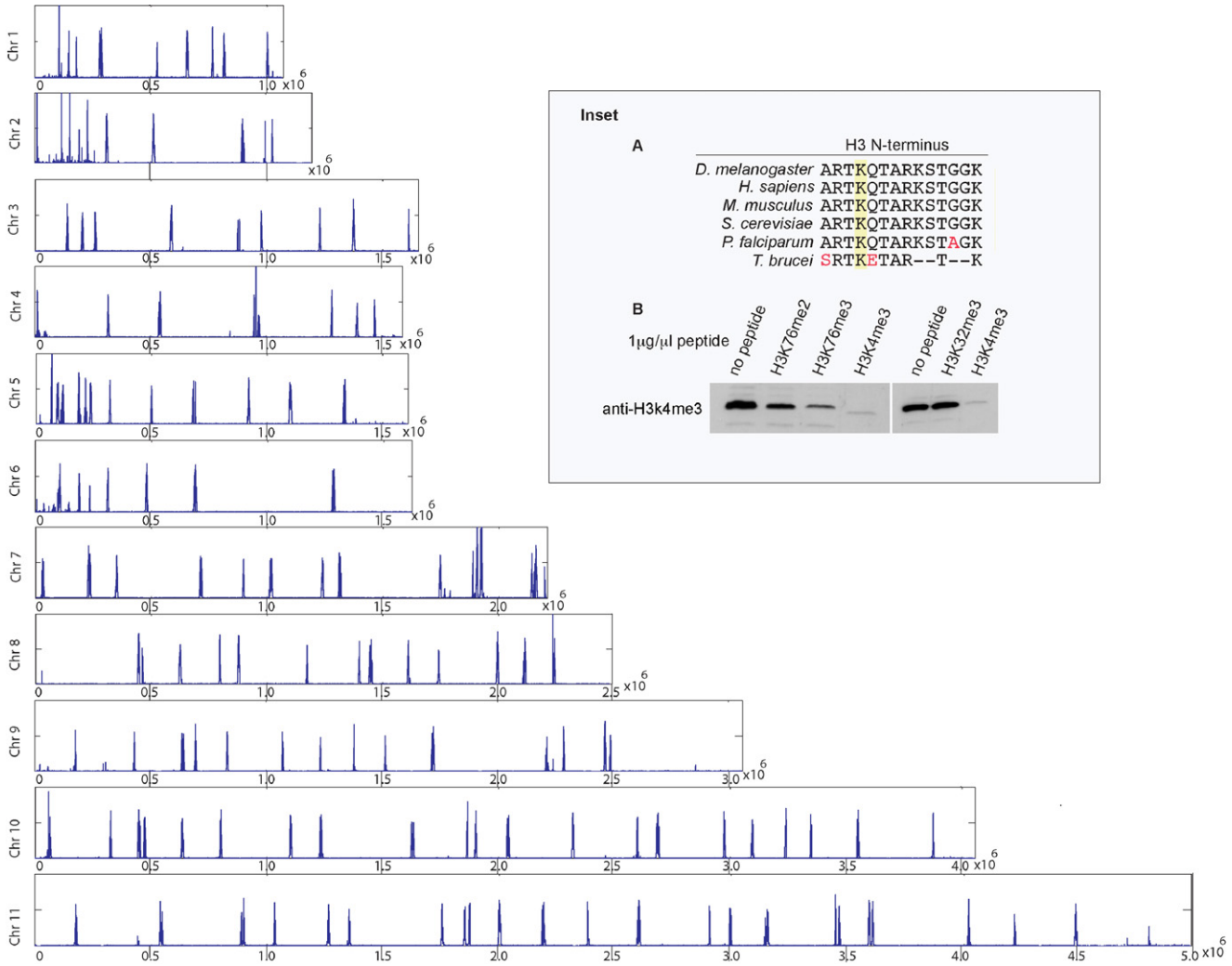

Fig. 1. Genome-wide distribution of H3K4me3. Nucleotide hits from 36-bp sequences were averaged over a 100-bp window. Inset: (A) The N-terminal sequence of *T. brucei* H3 is evolutionarily conserved. Divergent residues are shown in red and the conserved K4 lysine is highlighted. (B) Specificity of H3K4me3 ChIP-grade antibody. The total extract from 2×10^6 Dot1B null cells was electrophoresed and blotted with purified H3K4me3 antibody that was pre-incubated for 1 h with 1 μg/μl of the indicated peptides. Peptide sequences are as follows: H3K4me3 RTKETARTC; H3K73me3 VSGAQKEGLRFC; H3K76me2 VSGAQKEGLRFC (provided by Christian Janzen); H3K32me3 ASGVKTAQRC.