Published in final edited form as:

Mol Biochem Parasitol. 2012 December ; 186(2): 139–142. doi:10.1016/j.molbiopara.2012.09.003.

Chromatin immunoprecipitation and microarray analysis reveal that TFIIB occupies the SL RNA gene promoter region in *Trypanosoma brucei* chromosomes

Wenzhe Liu^{a,d}, Anish Das^a, Rachel Morales^a, Mahrukh Banday^a, Virginie Aris^c, David M. Lukac^a, Patricia Soteropoulos^{a,c}, David A. Wah^b, Jennifer Palenchar^e, and Vivian Bellofatto^{a,*}

^aDepartment of Microbiology and Molecular Genetics, University of Medicine and Dentistry-New Jersey Medical School, Newark, NJ, USA 07103

^bDepartment of Biochemistry, University of Medicine and Dentistry-New Jersey Medical School, Newark, NJ, USA 07103

^cPublic Health Research Institute, University of Medicine and Dentistry-New Jersey Medical School, Newark, NJ, USA 07103

^dVaxInnate Corporation, Cranbury, NJ, USA 08512

^eDepartment of Chemistry, Villanova University, Villanova, Pa 19087

Abstract

RNA polymerase II (RNAP-II) synthesizes the m⁷G-capped Spliced Leader (SL) RNA and most protein-coding mRNAs in trypanosomes. RNAP-II recruitment to DNA usually requires a set of transcription factors that make sequence-specific contacts near transcriptional start sites within chromosomes. In trypanosomes, the transcription factor TFIIB is necessary for RNAP-II-dependent SL RNA transcription. However, the trypanosomal TFIIB (tTFIIB) lacks the highly basic DNA binding region normally found in the C-terminal region of TFIIB proteins. To assess the precise pattern of tTFIIB binding within the SL RNA gene locus, as well as within several other loci, we performed chromatin immunoprecipitation/microarray analysis using a tiled gene array with a probe spacing of 10 nucleotides. We found that tTFIIB binds non-randomly within the SL RNA gene locus mainly within a 220-nt long region that straddles the transcription start site. tTFIIB does not bind within the small subunit (SSU) rRNA locus, indicating that trypanosomal TFIIB is not a component of an RNAP-I transcriptional complex. Interestingly, discrete binding sites were observed within the putative promoter regions of two loci on different chromosomes. These data suggest that although trypanosomal TFIIB lacks a highly basic DNA binding region, it nevertheless localizes to discrete regions of chromatin that include the SL RNA gene promoter.

© 2012 Elsevier B.V. All rights reserved.

*CORRESPONDING AUTHOR: Vivian Bellofatto, bellofat@umdnj.edu, phone: (973) 972-4483, ext. 2-4406, fax: (973) 972-3644. Mailing address: Dept. of Microbiology and Molecular Genetics, UMDNJ-NJ Medical School, International Center for Public Health, 225 Warren St., Newark, NJ 07103.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Trypanosoma brucei; trypanosomes; RNA synthesis; transcription; SL RNA genes; chromatin immunoprecipitation; TFIIB

Trypanosomatids, which are vector-transmitted parasites, require RNA polymerase II (tRNAP-II) to transcribe two independent RNAs for the production of a translatable mRNA (reviewed in [1, 2]). The smaller RNA, which is the ~140-nt Spliced Leader (SL) RNA, contributes its m⁷G-capped 5'-terminal 39 nt (the SL) to the longer protein-coding pre-mRNA, which contains multiple open reading frames (ORFs) arranged head-to-tail. The protein-coding pre-mRNA is processed into functional mRNAs by a coupled reaction that adds a *trans*-spliced SL upstream and a polyadenylated tail downstream from each ORF. The central role of the SL RNA in mRNA production is reflected in the genomic arrangement of SL RNA genes in trypanosomes. For example, in the case of *Trypanosoma brucei* strain 927, the SL RNA genes are present as ~30, 1.4 kb tandemly reiterated units within chromosome (Chr) 9 [3]. The influence that this genomic structure has on SL RNA gene expression is unknown.

SL RNA gene expression relies on the trypanosome version of the general transcription factor TFIIB, which is an essential component of RNAP-II-dependent transcription in the well-studied mammalian and yeast systems [4–6]. The well-studied TFIIB proteins function in concert with other basal transcription factors, integrated into a preinitiation complex (PIC), to direct RNAP-II to promoters. Within the PIC, TFIIB facilitates TATA-binding protein (TBP)-DNA interactions, determines RNAP-II direction along the DNA, and influences start site selection [7, 8]. In some cases, TFIIB binds to a TFIIB-recognition element (BRE) in the promoter [9].

The human TFIIB structure consists of an amino-terminal zinc ribbon and B finger that contact RNAP-II, and a C-terminal region containing tandem cyclin repeats that contact TBP and the promoter. Trypanosomal TFIIB (tTFIIB) shares its amino-terminal zinc ribbon, B finger region and its first cyclin repeat with the human TFIIB protein. However, tTFIIB lacks the highly basic DNA binding surface that is present in human TFIIB and other well-studied TFIIB proteins [4, 5, 10]. Given this unusual carboxyl terminal structure of trypanosomal TFIIB, and the repetitive configuration of the TFIIB-dependent SL RNA genes, we sought to determine the pattern of the TFIIB-SL RNA gene interactions *in vivo*.

We also explored the possibility that tTFIIB functions in the initiation of polycistronic pre-mRNA gene expression. Genomic structure analysis predicts that transcription start sites may localize within divergent strand switch regions (divergent SSRs), called as such because transcription emanates bidirectionally from them [11–16]. In support of this prediction are recent data that include specific histone modification-based marking of polycistronic transcription units and single nucleotide resolution mapping of RNAP-II transcription start sites within divergent SSRs [11, 16, 17]. Given the role of TFIIB in transcription initiation of protein-coding genes in other eukaryotes, it is possible that tTFIIB binding sites exist in divergent SSRs. However, as tTFIIB contains a neutral surface instead of the highly basic DNA binding surface found in other characterized TFIIB homologs, and as there are a lack of discernable BREs within putative RNAP-II-dependent promoter regions [16–19], the presence of tTFIIB in divergent SSRs is not a foregone conclusion.

We produced a high-resolution microarray by tiling DNA sequences from selected genomic regions totaling 48.6 kb within four megabase chromosomes (Fig. S1 contains all queried genomic coordinates). We tiled genomic regions from an entire 1.4 kb SL RNA gene repeat

unit, which contains the 140-nt-transcribed SL RNA and its intergenic regions, and a 3.6 kb alpha-beta tubulin gene repeat unit, which contains the polycistronic alpha/beta tubulin pre-mRNA within the *T. brucei* genome (Fig. 1). We also tiled DNA sequences from a 12.0 kb SSU rRNA region that included the RNAP-I promoter for the 18S, 5.8S and 28S rRNA, as a non-RNAP-II-dependent gene control. In addition, we tiled two representative divergent SSRs (including several flanking ORFs). They are the ~2.5 kb SSR from chromosome 3, which brackets protein-coding ORFs Tb972.3.5489 and Tb927.3.4900, and the ~7.0 kb SSR from chromosome 7, which brackets protein-coding ORFs Tb972.7.920 and Tb927.7.930. These SSRs were chosen because they represent the apparent size variation in SSRs (we consider 2.5 kb a short SSR, and 7.0 kb a long SSR), and are free of tRNA-coding genes, which bind RNAP-III-related transcription complexes and thus may complicate our analysis.

The *T. brucei* genomic tiled array was accomplished using 9694 sixty-mer DNA oligonucleotide probes representing one DNA strand at a resolution of 10 bp with a 55-base overlap (Agilent Technologies). Non-*T. brucei* probes were added to determine background tTFIIB binding levels. In our experiments, six individually grown *T. brucei* procyclic cultures (biological replicates) were analyzed. Parasites were treated with formaldehyde, cells were lysed, chromatin was sheared, and tTFIIB-bound chromatin was precipitated using a rabbit polyclonal antibody that specifically recognized tTFIIB. A limited number of *T. brucei* probes on the array showed median signal intensity up to a 100 fold above the background non-*T. brucei* control probes, suggesting that tTFIIB bound to discrete regions of the genome.

We determined the pattern of tTFIIB-SL RNA gene interaction using our ChIP-chip approach. A complete 1.4-kb gene copy of an SL RNA gene, comprising the promoter, coding region, and intergenic region from Chr 9 was interrogated. tTFIIB occupancy on the 1.4-kb SL RNA gene was exclusively at a ~220 bp region (limit of detection, due to fragmentation sizes and tiled array hybridization parameters) (Fig. 1A, peak 1). This region is centered at the SL RNA gene promoter, defined as the -100/-1-bp region upstream from the $_{(+1)}\text{AACGAACU}_{(+8)}$ transcription start site [2, 20]. There was no tTFIIB occupancy outside of the SL RNA gene promoter, within the 3.6-kb tubulin locus, within the 12.0-kb rRNA locus (Fig. 1B and C), or within the non-*T. brucei* control DNA. Subsequent ChIP-sequence data was consistent with limited tTFIIB occupancy within a ~200-nt sequence centered on the -100/-1-nt SL RNA gene promoter (unpublished). These data, which represent the first *in vivo* fine-structure mapping of tTFIIB binding, show that tTFIIB is indeed localized specifically and exclusively at the promoter that drives SL RNA transcription. In light of the unusual C-terminal domain of trypanosomal TFIIB, and the lack of a discernable BRE within the SL RNA gene promoter, we predict that the selected TFIIB-binding we observe occurs through other general transcription factors. In support of this suggestion, previous experiments show that tTFIIB does associate with the general transcription factors tSNAPc, tTBP, and tTFIIH, which comprise a multi-protein preinitiation complex, assembled at the SL RNA gene promoter [4–6, 21]. Given the neutral surface and the absence of the second cyclin repeat in tTFIIB, direct interaction between tTFIIB and the SL RNA gene promoter is an interesting question that awaits experimental evaluation.

Next, we tested whether tTFIIB localized to genomic regions that included two representative, divergent SSRs. We queried the 10.8 kb region of Chr 3 that contains a short, 2.5 kb SSR region, located between the ORFs of genes Tb972.3.4890 and Tb972.3.4900, as well as the ORFs for three protein-coding genes on each side of the divergent SSR. We also queried a larger area, a 20.8-kb region on Chr 7, that contains a long, 7.0 kb divergent SSR, located between the ORFs of genes Tb927.7.920 and Tb927.7.930, as well as the two ORF coding regions of these genes. Three narrow (~0.08 kb) regions of tTFIIB occupancy were

observed within the 10.8-kb Chr3 loci and two narrow regions of tTFIIB occupancy were observed within the 20.8-kb Chr 7 locus (Fig. 1D and E). Intriguingly, four of the five peaks observed in the queried ~30-kb region were within putative transcription initiation regions of the chromosomes (Fig. 1D and E, peaks a, b, d and e) [11, 14, 16, 17]. The fifth peak (Fig. 1D, peak c), was between two ORFs. Interestingly, transcription start sites are posited to be between ORFs within gene clusters, as well as upstream of them [17]

Comparing our data to the relevant literature, we find some discrepancies and many analogies. For example, tTFIIB did not occupy any discrete regions within the SSU rRNA locus. This is in contrast to recent tTBP and tSNAP₅₀ ChIP-chip analysis in *Leishmania major* that suggests that some basal transcription factors bind within rRNA coding regions [16]. Recent ChIP-chip data in *L. major* demonstrated tTBP and tSNAP50 binding to 2–3-kb regions within many divergent SSRs [16]. In our experiments, tTFIIB was present in the two SSRs examined. However, tTFIIB did not show a robust presence within the tested divergent SSR regions. This may reflect low levels of transcription from the SSRs examined in this work; thus, a relatively low abundance of tTFIIB at these loci. It also could reflect a limited role of tTFIIB in these strand switch regions. Indeed, one role of TFIIB in other characterized systems is to orient the transcriptional machinery. If productive transcription can proceed in either direction from an SSR, perhaps this role of tTFIIB is eschewed in SSR-related transcription. The finding that divergent SSRs are replete with histone acetylation in both in *L. major* and *T. brucei* [11, 16] may indicate that SSR transcription initiation is directed at an epigenetic level and/or may involve yet unidentified factors. Finally, open chromatin regions of the genome, possibly modulated by modified histones, may help transcription factors access their cognate binding sites. In support of this, we note that the peaks a and b that represent TFIIB occupancy on Chr 3 in our *T. brucei* study coincide with a histone H4K10ac peak observed in the Siegel, et al *T. brucei* study [11]. Moreover, we note that peaks a and peak d that represent TFIIB occupancy on Chr 3 and Chr 7, respectively, in our *T. brucei* study correspond to tSNAP and TBP binding sites observed in the Thomas et al *L. major* study [16] (and P. Myler, personal communication.)

Our findings are reflective of tTFIIB-SL RNA gene interactions observed in immunoprecipitation assays, in *in vitro* DNA pull-down assays, and transcription experiments [4–6]. Employing an overlapped-probe tiled array, we were able to increase the resolution of the TFIIB- SL RNA gene interaction and demonstrate that tTFIIB, albeit unusual in its carboxyl terminal domain, nevertheless exhibits binding specificity and thus may contribute to transcription start site selection within the SL RNA gene promoter. As the first four nucleotides of the spliced leader become the hypermethylated cap 4, trypanosomatid mature mRNAs, correct initiation of SL RNA gene transcription may be a key job for tTFIIB.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Stacey Garcia for critical reading of the manuscript. This work was supported by NIH-NIAD grants AI-29478 and AI-53835, a grant from the Foundation of UMDNJ to VB, and American Heart grants to AD and DL.

References

1. Matthews KR. Controlling and coordinating development in vector-transmitted parasites. *Science*. 2011; 331(6021):1149–53. [PubMed: 21385707]
2. Palenchar JB, Bellofatto V. Gene transcription in trypanosomes. *Mol Biochem Parasitol*. 2006

3. Aslett M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae 10.1093/nar/gkp851. *Nucl Acids Res.* 2010; 38(suppl_1):D457–462. [PubMed: 19843604]
4. Palenchar JB, et al. A Divergent Transcription Factor TFIIB in Trypanosomes Is Required for RNA Polymerase II-Dependent Spliced Leader RNA Transcription and Cell Viability. *Eukaryot Cell.* 2006; 5(2):293–300. [PubMed: 16467470]
5. Schimanski B, et al. A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Res.* 2006; 34(6):1676–84. [PubMed: 16554554]
6. Lee JH, et al. A TFIIB-associated mediator head is a basal factor of small nuclear spliced leader RNA gene transcription in early-diverged trypanosomes. *Molecular and cellular biology.* 2010; 30(23):5502–13. [PubMed: 20876299]
7. Tsai FT, Sigler PB. Structural basis of preinitiation complex assembly on human pol II promoters. *Embo J.* 2000; 19(1):25–36. [PubMed: 10619841]
8. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol.* 2004; 11(5):394–403. [PubMed: 15114340]
9. Deng W, Roberts SG. Core promoter elements recognized by transcription factor IIB. *Biochemical Society transactions.* 2006; 34(Pt 6):1051–3. [PubMed: 17073748]
10. Ibrahim BS, et al. Structure of the C-terminal domain of transcription factor IIB from *Trypanosoma brucei*. *Proc Natl Acad Sci U S A.* 2009; 106(32):13242–7. [PubMed: 19666603]
11. Siegel TN, et al. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.* 2009
12. Berriman M, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science.* 2005; 309(5733):416–22. [PubMed: 16020726]
13. El-Sayed NM, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science.* 2005; 309(5733):404–9. [PubMed: 16020724]
14. Martinez-Calvillo S, et al. Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryotic Cell.* 2004; 3(2):506–17. [PubMed: 15075279]
15. Respuela P, et al. Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *trypanosoma cruzi*. *J Biol Chem.* 2008
16. Thomas S, et al. Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics.* 2009; 10(1):152. [PubMed: 19356248]
17. Kolev NG, et al. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS pathogens.* 2010; 6(9):e1001090. [PubMed: 20838601]
18. Siegel TN, et al. Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. *Trends in parasitology.* 2011; 27(10):434–41. [PubMed: 21737348]
19. Siegel TN, et al. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic acids research.* 2010; 38(15):4946–57. [PubMed: 20385579]
20. Campbell DA, Sturm NR, Yu MC. Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol Today.* 2000; 16(2):78–82. [PubMed: 10652494]
21. Ruan JP, et al. Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. *Mol Cell Biol.* 2004; 24(21):9610–8. [PubMed: 15485927]
22. Harlow, E.; Lane, D., editors. *Antibodies: A Laboratory manual*. Cold Spring Harbor Laboratory; 1988.
23. Wirtz E, et al. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol.* 1999; 99(1):89–101. [PubMed: 10215027]
24. Lowell JE, Cross GA. A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. *J Cell Sci.* 2004; 117(Pt 24):5937–47. [PubMed: 15522895]
25. Loayza D, De Lange T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature.* 2003; 423(6943):1013–8. [PubMed: 12768206]

Highlights

Localization of *T. brucei* TFIIB at a subset of chromatin sites.

Chromatin Immunoprecipitation and microarray analysis in *T. brucei*.

Discrete TFIIB binding sites within the SL RNA gene array.

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1. Analysis of tTFIIB occupancy at five genomic loci in *T. brucei* by chromatin immunoprecipitation/microarray (ChIP-chip)

tTFIIB-DNA complexes were identified by chromatin immunoprecipitation (ChIP) using anti-tTFIIB antibody. Polyclonal tTFIIB-specific antibody [4] was immunopurified from rabbit sera by GST affinity chromatography followed by recombinant tTFIIB affinity chromatography [22]. Purified antibody detected a single polypeptide in a total *T. brucei* extract by Western blot analysis. Protein-DNA complexes were obtained from 4×10^8 *T. brucei* procyclic (tsetse midgut form) wild-type Lister strain 427, cultured as described previously [23], and mixed with formaldehyde as described in [24]. Chromatin was sheared using the Bioruptor™ 200 (Diagenode) to 200–400 bp fragments (assessed by ethidium bromide staining of DNA fragments separated by electrophoresis on 1.1% agarose gels). For ChIP, anti-tTFIIB or preimmune serum was used. Immunoprecipitates were extensively washed as described [25]. Preliminary experiments that compared the ability of preimmune and immune serum to capture specific regions of chromatin showed that no sequence enrichment occurred in the preimmune sample. Thus, in the six pairs of biological replicates performed with different parasite cell cultures, and two independently purified anti-tTFIIB antibody preparations, we used anti-tTFIIB antibody and input DNA to generate our six data sets. After reverse cross-linking DNA from protein (65° C in carbonate buffer), DNA was amplified using the WGA2 materials and protocol (Sigma) and labeled with Cy5 or Cy 3 using the Genomic Labeling Kit PLUS (Agilent). In alternate hybridizations, the two labeling dyes were reversed. Custom-made glass slide 4x44000 format (Agilent) were blocked and then hybridized (3 mg of labeled DNA, 24 hrs, 60° C), and washed as described in the Mammalian ChIP-chip protocol, version 10.0 (Agilent). The array contained triplicate printings of 60-mer oligonucleotides (genomic *T. brucei* from GeneDB and TriTrypDB [3]), permuted every 5 nts, and thus overlapped for 55 nt. The coordinates represented on the

array are indicated. (note: The SL RNA gene array sequences were sequencing retrieved from coordinates 2,266,146 to 2,267,562 on Chr 9. They correspond to 2,265,476 to 2,266,892, which represents a different permutation of the 1.4 kb repeat. The terminal sequences of the tiled coordinates generated an artifactual, short gap the hybridization pattern). A single strand of DNA, equivalent to the strand presented in GeneDB and TriTrypDB, was represented within the oligonucleotides. Slides were scanned on a GenePix 4000B scanner (Molecular Devices) and the images processed using GenePix 5.1. Raw data from microarray hybridizations was normalized according to the Agilent definition incorporated into their feature extraction software. Specifically, hybridization to 1227 sixty-mer Agilent control probes provided a nonspecific hybridization background level of signal. The background signal was subtracted from *T. brucei* probes using an algorithm built into the Agilent analysis package. We analyzed consecutive, triplicate spots in each of the six data sets and removed those that had a p-value great than 0.05. Spots that failed this test in 50 % of the data sets were not considered. The $\log_{10}$ ratio was taken of the immunoprecipitated versus input DNA. To identify regions of enrichment, we filtered probe intensity to include only points with at least two out of three triplicates with a p-value below 0.05 and 50% of the samples with a p-value below 0.05. The averaged probe values in each data set were then fit to a normal distribution, and the average of standard deviations from the mean were calculated for each probe. Probes with greatest deviations from the mean corresponded to those that were most significantly enriched relative to each data set. The peaks in panels B-E are the loci where there were 2 data points that generated the peak, and the standard deviation from the mean was 1.2. We used more stringent criteria to establish the peaks in panel A; since the SL RNA-Chr 9 genes are highly reiterated sequences, the peaks in panel A are the loci where 2 data points generated the peak, and the standard deviation from the mean was 1.8. Further analysis of the region designated with an asterisk was not confirmed, although promoter binding was reinforced (unpublished ChIP-sequence data). The diagrams below each graph include the regions on the microarray (in brackets) as well as the surrounding regions of the chromosomes. Y-axis shows values 1.5 standard deviations from the mean level of tTFIIB ChIP-chip quantization. Two-part arrows designate established promoter regions and one-part arrows designate transcriptional direction. The graduated grey area is the SL RNA gene promoter. (The SL RNA transcription initiates as coordinate #266,142). The black area is the transcribed SL RNA. (The wider box designates the 39-nt SL component of 140-nt SL RNA.) Solid grey boxes show chromosome regions that produce stable RNAs. Genomic coordinates that define the queried region are marked.