EPIGENETIC REGULATION OF TRANSCRIPTION AND VIRULENCE IN *TRYPANOSOMA CRUZI* BY O-LINKED THYMIDINE GLUCOSYLATION OF DNA

by

DILRUKSHI KUMARI EKANAYAKE

(Under the Direction of Robert Sabatini)

ABSTRACT

In trypanosomes, unlike most eukaryotes, genes transcribed by RNA polymerase II (Pol II) are arranged in polycistronic transcription units (PTUs). Based on this organization, it has been held that trypanosomes rely solely on post-transcriptional processes to regulate gene expression. However, very little is known about how transcription is regulated in trypanosomes. Here, we have characterized the function of a novel glucosylated thymidine DNA residue (base J) in *Trypanosoma cruzi.* We found that base J localizes to promoter regions of PTUs throughout the *T. cruzi* genome. Loss of base J following genomic deletion of the thymidine hydroxylase involved in J synthesis (JBP1) led to a global increase in RNA polymerase II transcription and gene expression. We characterized changes in chromatin structure following base J depletion, and observed that loss of base J coincides with decreased nucleosome occupancy, increased histone acetylation, and increased RNA polymerase II occupancy at promoter regions. Changes in gene expression due to loss of base J may increase the ability of *T. cruzi* to invade mammalian cells and decrease its capacity to egress from host cells *in vitro*. These studies present the first detailed analysis of the role of base J in chromatin structure and regulation of Pol II-dependent transcription in kinetoplastids.

INDEX WORDS: Trypanosomes, Base J, Gene expression, Chromatin, RNA polymerase II

EPIGENETIC REGULATION OF TRANSCRIPTION AND VIRULENCE IN *TRYPANOSOMA CRUZI* BY O-LINKED THYMIDINE GLUCOSYLATION OF DNA

by

DILRUKSHI KUMARI EKANAYAKE

Doctor of Veterinary Medicine, University of Peradeniya, Sri Lanka, 2003

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

EPIGENETIC REGULATION OF TRANSCRIPTION AND VIRULENCE IN *TRYPANOSOMA CRUZI* BY O-LINKED THYMIDINE GLUCOSYLATION OF DNA

by

DILRUKSHI KUMARI EKANAYAKE

Major Professor:	Robert Sabatini
Committee:	Stephen Hajduk Rick Tarleton Michael McEachern

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011

DEDICATION

I dedicate this work to my parents.

ACKNOWLEDGEMENTS

I thank my advisor, Dr. Robert Sabatini, for his guidance, support, advice and flexibility. I thank Dr. Steve Hajduk for his encouragement and support. I thank the other members of my committee, Dr. Rick Tarleton and Dr. Michael McEachern, for their thoughtful discussions and constructive criticisms. I thank past and current members of the Sabatini, Tarleton, Docampo Moreno and Hajduk labs for their help. I thank my husband, Benjamin Alper, for editing my English, his love and encouragement. I thank my friends and family for their love and their support.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW 1

Chagas disease 1

The *T. cruzi* life cycle 2

Host–parasite interactions and proteins important for parasite survival 3

Base J biosynthesis and possible function in trypanosomatids 5

General mechanism for control of gene expression 8

Epigenetic mechanisms for control of gene expression 9

Control of gene expression in trypanosomatids 11

Chromatin structure and regulation of eukaryotic transcription 14

Chromatin structure and regulation of transcription in trypanosomatids 15

References 18

2 TELOMERIC CO-LOCALIZATION OF THE MODIFIED BASE J AND CONTINGENCY GENES IN THE PROTOZOAN PARASITE *TRYPANOSOMA CRUZI* 28

Abstract 29

Introduction 30

Materials and Methods……...32

Results 36

Discussion 43

Acknowledgements……49

References 50

3 EPIGENETIC REGULATION OF TRANSCRIPTION AND VIRULENCE IN *TRYPANOSOMA CRUZI* BY O-LINKED THYMIDINE GLUCOSYLATION OF DNA 67

Abstract 68

Introduction 69

Materials and Methods……...71

Results 78

Discussion 87

Acknowledgements……91

References 93

4 EPIGENETIC REGULATION OF POL II TRANSCRIPTION INITIATION IN *TRYPANOSOMA CRUZI*: MODULATION OF NUCLEOSOME ABUNDANCE, HISTONE MODIFICATION AND POLYMERASE OCCUPANCY BY O-LINKED THYMIDINE DNA GLUCOSYLATION 116

Abstract 117

Introduction 118

Materials and Methods……122

Results 124

Discussion 130

Acknowledgements……..134

References 135

5 CONCLUSIONS, DISCUSSION, AND FUTURE DIRECTIONS 146

Conclusions 146

Mechanism of base J in regulating gene expression..... 147

Biological significance of J regulation of gene expression…….....148

Other possible functions of base J 150

Summary……..156

References……157

LIST OF TABLES

Page

Table 2.1: Distribution of base J in tandem repeats of *T. cruzi* 55

Table 2.2: Telomeric versus internal repeat scaffold 56

Table 3.1: Increased transcription rate in JBP1 KO by run-on analysis 97

Table 3.2: Spliced leader trapping analysis of epimastigote cells 97

Table S3.1: Number of surface proteins upregulated for different life stages of JBPKOs 107

Table S3.2: PTU distribution of regulated genes of JBP2dKO and JBP1dKO 108

LIST OF FIGURES

Page

Figure 1.1: *T. cruzi* life cycle 26

Figure 1.2: Regulation of J biosynthesis 27

Figure 2.1: Developmental regulation of J biosynthesis in *T. cruzi* 57

Figure 2.2: Quantitating the fraction of base J present outside the telomeric repeat regions of *T. cruzi* 58

Figure 2.3: Telomeric-localized repetitive regions contain base J 60

Figure 2.4: Localization of J in the subtelomeric region 61

Figure 2.5: J localization within three additional subtelomeres 63

Figure 2S.1: PCR analysis of subtelomeric versus internal repeat sequences 64

Figure 3.1: JBP knockout (JBPKO) cells have reduced levels of base J 98

Figure 3.2: JBP1 and JBP2 regulate base J synthesis in polycistronic unit (PTU)-flanking regions of *T. cruzi* 99

Figure 3.3: Micro array analysis of WT and JBPdKO cell lines 100

Figure 3.4: Base J regulates Pol II transcription rate 102

Figure 3.5: Transcription initiation and termination from sense (s) and antisense (as) strands within divergent and convergent SSRs 103

Figure 3.6: Whole-genome comparison of gene expression levels in WT cells versus JBP1KOs 104

Figure 3.7: Base J regulates *T. cruzi* virulence 106

Figure 3S.1: JBP1 and JBP2 regulate J synthesis in telomeric DNA,

related to Figure 3.1..... 110

Figure 3S.2: Micro array analysis of the JBP1dKO and JBP2dKO,

related to Figure 3.3..... 111

Figure 3S.3: RNA stability of selected genes assayed by qRT-PCR,

related to Figure 3.4..... 112

Figure 3S.4: Effect of base J on *in vivo* transcription rate, related

related to Figure 3.4..... 113

Figure 3S.5: SLT and run-on analysis of SSRs,

related to Figure 3.5..... 114

Figure 3S.6: Virulence phenotype analysis of JBP KO cell lines,

related to Figure 3.7..... 115

Figure 4.1: FAIRE and histone IP analyses show that divergent SSRs are depleted of

nucleosomes and undergo increased histone acetylation upon the loss of base J.....141

Figure 4.2: The spliced leader transcription unit becomes more organized in an open

chromatin structure upon the loss of base J 144

Figure 4.3: The loss of base J at Pol II promoters leads to increased Pol II occupancy……145

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Chagas disease

Trypanosoma cruzi is a major cause of cardiac disease in Central and Southern America, where it infects approximately 18 million people [1]. The parasite causes Chagas cardiomyopathy, an unusually complex chronic illness with poorly understood pathogenesis. 5% to 10% of patients produce typical clinical conditions of chagasic megaesophagus and megacolon [1]. Acute infection occurs following exposure to parasitic infection, with detectable parasitemia in the peripheral blood, pronounced inflammation and myocardial damage. The acute phase can be asymptomatic or symptomatic with fever, conjunctivitis, myocarditis and meningeoencephalitis [2,3]. The majority of individuals surviving acute infection enter a clinical quiescence period of 10 to 15 years. During the clinical course, striated muscles including cardiomyocytes and skeletal muscles, and autonomous ganglia of the nervous system are severely affected. Chronic disease is characterized by the presence of severe cardiomyopathies [2,3]. The pathogenesis of clinical cardiomyopathy may be due either to the persistence of the parasites in the myocardium, or to autoimmunity induced by the disease process [4].

To date, there are no effective drugs for treatment of chronic *T. cruzi* infections, and available drugs for treatment of acute infections are highly toxic. The main control strategy for Chagas disease is the prevention of parasite transmission by eliminating the triatomine insect vectors [5]. Approximately 130 species of triatomine bugs have been reported in transmission of

T. cruzi. However, *Triatome infestan*, *Triatome brasiliensis*, *Triatome Dimidiata,* and *Rhodnius prolixus* are the main vectors of Central and Southern America [6]. The risk of acquiring Chagas disease outside of Latin America is rare. However, transmission of *T. cruzi* by blood transfusion or by organ donation outside of endemic areas has been reported [7]. Disease incidence within the United States is also indicated by the recent United States Food and Drug Administration (US FDA) approval of a test for *T. cruzi* in donated human blood [5].

The *T. cruzi* life cycle

T. cruzi is a flagellate protozoan of kinetoplastid order *Trypanosomatidae*. The parasite adopts a complex life cycle, with two host and four developmental stages (Figure 1.1.) [8]. The epimastigote life stage is the replicative form that proliferates within the insect. It has a flagella pocket in the anterior position in the nucleus. *T. cruzi* epimastigotes develop in the hindgut of the triatomine insect vector. Triatomine insects acquire their infections by feeding on infected mammals, and it is within the insect host that *T. cruzi* epimastigotes differentiate to form the metacyclic typomastigotes. The kinetoplast of metacyclic trypomastigotes is located in the extremity posterior to the nucleus [9]. Infective metacyclic *T. cruzi* accumulate in the distal region of the insect intestine, and may be deposited on mammalian hosts along with insect feces[8]. The parasite can then gain access to the vertebral blood stream via skin abrasions or by entering through mucosal membranes. In turn, blood-borne *T. cruzi* invade a wide variety of host cells and can differentiate to form amastigotes. Amastigotes are a replicative, rounded form of *T. cruzi* with short, inconspicuous flagella. Following several binary divisions, amastigotes differentiate to form trypomastigotes, and upon cell lysis, tissue-derived trypomastigotes escape into the blood stream. The transmission cycle closes when a triatome vector feeds on a *T. cruzi*

infected mammal[8]. During sustained infection, life stages of *T. cruzi* continually adapt as they encounter different host environments. Therefore, host-parasite interactions is a key step determining the disease process.

Host-parasite interactions and proteins important for parasite survival

Pathogens encounter multiple immune challenges during host infection and invasion. These challenges may be overcome by activating genes to change surface phenotypes, or by generating a broad array of these phenotypes through a process termed antigenic diversity [10,11]. The cell surface presents an important interface between a parasite and its environment. In a number of eukaryotic pathogens, including *Trypanosoma brucei, Plasmodium falciparum, Pneumocyctis carinii* and *Candida glabrata*, surface protein genes undergo rapid evolution and are subject to highly regulated expression to allow fast adaptation to the host immune system [12,13].

Extensive variation in surface protein gene structure is thought to represent an important aspect of *T. cruzi* pathogenesis. Proliferation is ensured by the parasite's ability to regulate expression of surface antigens that prevent adequate immune response [14,15,16,17] and allow invasion of host cells [16,18,19]. Many surface protein genes in *T. cruzi* are highly mutable, having higher than normal mutation rates compared with the broader genome, thereby leading to enhanced antigenic diversity [10]. Intriguingly, the trans-sialidase and mucin surface protein gene families are among the most highly mutable gene families in *T. cruzi* [10].

The *T. cruzi* genome contains approximately 1000 trans-sialidase (TS) genes. The organism does not posses the ability to produce sialic acid, and is therefore obligated to express TS to cleave sialic acid from host glycoconjugates [20,21,22]. Thus obtained, sialic acid is incorporated into various surface molecules, including *O*-linked oligosaccharides that are

important for immune system evasion (*i.e.,* mucins). Enzymatically active TSs produced by epimastigotes are not released into the extracellular medium and are thought to play a role in protecting the parasite from enzymes in the gut of the insect vector. TS activity is also important for the metacyclic and trypomastigote stages of the *T. cruzi* life cycle, as the transfer of host sialic acid onto parasite surface molecules enhances the host-parasite interaction and facilitates invasion [21]. In metacyclics, sialic acid is used to form the Ssp3 epitope on the parasite cell surface, and its sialation enhances the initial recognition host cells targeted for invasion [21,23,24]. One mucin can harbor several host sialic acids and may prevent the interaction of the trypomastigote surface with host cell receptors of phagocytic cell membranes, thereby protecting the parasite from complement mediated killing [21,25].

Mucin-like glycoproteins (TcMUCs) are the major surface glycoproteins of *T. cruzi*. These glycoproteins are encoded by a very large family of genes, with >500 copies present in the genome [26]. TcMUC sequences are rich in serine and threonine residues and act as the acceptor sites for the addition of sialic acids [27]. In the insect-borne *T. cruzi* life stages (epimastigotes and metacyclics), expressed mucins are relatively short and are comprised of comparatively simple oligosaccharides [21]. They are highly resistant to protease activity, and protect the parasite from digestive enzymes of the insect gut. Metacyclic TcMUCs can trigger Ca^{2+} mobilization from host tissues and facilitate the invasion process. By contrast, trypomastigotes express mucins that are relatively large and contain complex type *O*-oligosaccharides. Trypomastigote TcMUCs share the sialic acid-containing epitope Ssp3, which is an important component of mammalian cell attachment and invasion [24,28]. During differentiation *T. cruzi* exchanges a homogenous mucin coat for a highly heterogeneous surface coat. This mosaic mucin coat may help the parasite to evade the host immune response by antagonizing the effector T cell

response, and by facilitating attachment and invasion of a broad spectrum of mammalian cell types [28].

The functional diversity of *T. cruzi* TSs and TcMUCs, and the nature of their interactions with various host cell types and the vertebrate immune system are subjects of great interest. However, little is known about how *T. cruzi* TS and TcMUC genes are organized, expressed, or rearranged to promote the genetic diversity that leads to sustained parasitemia and pathogenesis. In this work, we have identified an epigenetic DNA modification in *T. cruzi* that colocalizes with members of the TS and TcMUC gene families, and sought to clarify its role in regulation of gene expression.

Base J biosynthesis and possible functions in trypanosomatids

β-D-glycosylhydroxymethyluracil, or base J, is a novel modification of DNA in kinetoplastids and related organisms (*e.g., Diplonema* and *Euglena*) [29,30]. Base J was first discovered in the blood stream form of *T. brucei,* but was not detected in the insect form of this organism. In *T. brucei,* base J is lost during procyclic differentiation within the insect gut, owing not to its active removal, but rather to its gradual dilution during replication. However, in *T. cruzi* and *L. major,* base J is present in both insect and mammalian forms [29]. Base J was initially discovered within the telomeric and subtelomeric repeat regions of *T. brucei*.

Base J is synthesized through a two-step mechanism. First, a thymidine hydroxylase converts a thymidine residue in DNA to hydroxymethyldeoxyuridine (HOMedU), and second, a glucosyltransferase converts HOMedU into base J by addition of glucose (Figure 1.2). Two thymidine hydroxylases (TH) have been identified in trypanosomes [30]. The 90 kDa J-binding protein 1 (JBP1) was first identified in nuclear extract of *Crithidia fasciculate,* and later in other

kinetoplastids [31]. Subsequently, a JBP1 sequence homology search conducted in *T. brucei* identified the homologous 120 kDa JBP2 protein [32]. The N-terminal halves of JBP1 and JBP2 show 34% sequence similarity and are required for TH activity of the respective enzymes. Both JBP1 and JBP2 catalyze a thymidine oxidation reaction using oxygen as the reactive agent, ferrous iron as cofactor, and 2-oxyglutarate as the co-substrate [33]. These two enzymes belong to the AlkB super family, and are thought to share a β-strand fold within a highly conserved motif of their N-termini. This N-terminal motif consists of four amino acids, including alanine and serine residues that are critical for function [32,34]. JBP1 and JBP2 double null mutants completely lack base J, indicating the presence of only two TH enzymes in the *T. brucei* genome [35]. Genome-wide sequencing analysis has since identified JBP1 and JBP2 homologues in *T.cruzi*, as well [36].

JBP1 binds to base J containing duplex DNA with high affinity, as evidenced by gel shift experiments performed using recombinant JBP1 protein and J modified DNA oligonucleotides. In order to bind to J modified DNA, JBP1 exhibits a strict requirement for J in one helical turn of duplex DNA, suggesting that DNA structure is essential for JBP1 binding [34]. JBP1 contains a 20 kDa J modified DNA binding domain within the center of the protein. Mutation of residues within this domain inhibits the J modified DNA binding by JBP1 *in vitro*. However mutations within the N-terminal region of JBP1, which contains the TH domain, did not affect J modified DNA binding [30]. JBP1 can also bind to unmodified DNA, albeit with 100-fold lower affinity than J modified DNA. JBP1 does not require any sequence-specific contacts in order to bind to J modified DNA, as demonstrated through methylation interference and DNA footprinting analysis [34]. However, JBP1 has a high binding affinity for telomeric repeat sequences when base J is present, yet does not appear to have the ability to initiate the base J synthesis at those

sites [30,33]. JBP2 in WT cells specifically provides the basal J modified DNA that is important for directing JBP1-stimulated synthesis. This was confirmed through telomeric fragmentation analyses using blood stream JBP2 null *T. brucei*. In these cells, telomeric fragmentation resulted in growth of telomeric repeats that lacked base J [37]. JBP1 knockouts in *T .brucei* reduced J modified DNA levels 20-fold, with all DNA sequences containing base J appearing to be similarly affected [31]. This observation suggested a role for JBP1 in maintaining base J in DNA. When JBP1 was added back to JBP1 and JBP2 null background, JBP1 could stimulate *de novo* J synthesis, but the mark was rapidly lost during subsequent cell proliferation. Therefore, maintenance of J at JBP1 specific regions appears to be dependent on JBP2 SWI/SNF and TH activity [38].

Deletion of JBP2 reduced J levels less than 5-fold in *T. brucei,* suggesting that JBP1 is the key regulator of base J maintenance. However, JBP2 can initiate *de novo* synthesis of base J without the presence of J modified DNA. Ectopic expression of JBP2 in the insect form of *T. brucei* resulted in *de novo* site-specific J synthesis. J modified DNA levels were further amplified by co-expressing JBP1 [32]. When JBP2 was added to a JBP1 and JBP2 null background in *T. brucei*, JBP2 stimulated base J synthesis within telomeric and subtelomeric repeat regions [32]. However, JBP2 null trypanosomes were unable to incorporate base J into newly generated telomeric DNA [37]. Collectively, these observations suggest that JBP2 is the key regulator of *de novo* J biosynthesis.

In addition to subtelomeric repeat regions, base J was also found to localize to silent variable surface glycoprotein (VSG) expression sites in *T. brucei*. The trypanosome genome contains ~1000 VSGs, yet only expresses one VSG on the cell surface at a given time. Trypanosomes regularly change their surface coat, a process known as antigenic variation, to

evade the host immune challenge. Intriguingly, the expressed VSG is present in only one telomeric expression site, and base J is absent from this region. This finding led to the hypothesis that base J may function in regulation of gene expression and antigenic variation in *T. brucei* [29].

General mechanisms for control of gene expression

Transcriptional regulation is a key step in controlling gene expression. In most organisms, control of transcription is mediated at the molecular level by numerous *cis*-and *trans*-acting factors, including canonical promoter sequences that serve as anchors to both general and specific transcription factors. Under certain conditions, these transcription factors may act to form a pre-initiation complex, which can in turn be recognized by RNA polymerase II to initiate the synthesis of mRNA [39]. Once the full length mRNA is transcribed, the elongating polymerase complex terminates transcription. Newly transcribed mRNA is then spliced, edited, modified and transported to the cytoplasm before being translated by ribosomes to form proteins and other peptides. The entire transcription process is tightly controlled by dynamic regulation within the cell [40,41].

The eukaryotic genome consists of a large number of genes, and activation of gene expression often requires that cells alleviate repression of genes that are transcriptionally silent. For example, transcriptional activity often spikes in response to environmental cues and differentiation. A transcription cascade including activating transcription factors (TFs) can initiate such expression changes. Both in eukaryotes and prokaryotes, transcription is controlled by TFs that bind to specific regulatory sequences and modulate the activity of RNA polymerases.

In most eukaryotes, the formation of a pre-initiation complex at promoter sites involves general transcription factors TFII A, B, D, E, F and H. These TFs bind to *cis*-acting regulatory sequences such as the AT-rich TATA boxes and Inr sequences within the promoters of genes transcribed by RNA polymerase II (Pol II) [39]. *Trans*-acting regulatory elements can function as either activators or repressors of eukaryotic protein-coding genes. For example, the binding of a repressor protein near the transcription start site can block the interaction of Pol II or general transcription factors with the promoter, while transcriptional activators, like Sp1, may bind to regulatory DNA sequences and stimulate transcription [42]. Similar mechanisms exist to control transcription regulation in many prokaryotes [43].

A high percentage of the prokaryotic genome consists of DNA operons in which a single promoter controls the expression of a series of genes physically linked to one another in functionally related units. Transcription initiation results in the production of long polycistronic mRNAs that contain multiple gene transcripts in tandem. This operon architecture allows multiple genes to be expressed and regulated sequentially as a single transcriptional unit. However, functionally related genes are not physically clustered on the genomes of eukaryotes, and promoters from each gene regulate transcription individually. Epigenetic mechanisms that produce changes in gene expression without altering the underlying gene structure contribute to control of transcription in both prokaryotic and eukaryotic organisms.

Epigenetic mechanisms for control of gene expression

Epigenetic mechanisms such as DNA methylation and the controlled incorporation of diverse histone modifications and histone variants present additional means for controlling gene expression that can be inherited or subject to developmental cues.

DNA methylation represents a central mechanism of epigenetic control. This covalent modification of DNA is associated with transcriptional gene silencing and the stabilization of skipper retrotransposons. Cytosine residues in vertebrate DNA can be modified by the addition of methyl groups at the number 5 position within the pyrimidine ring (m5C). m5C methylation is correlated with reduced transcriptional activity of genes that contain high frequencies of CpG dinucleotides in the vicinity of their promoters. The formation of m5C is catalyzed by methyl-5-cytosine methyltransferases that use S-adenosylmethionine as a cofactor (examples include DNMTs in mammals and the Dam methyltransferases in prokaryotes).

DNMTs and Dam methyltransferases both establish and maintain CpG methylation patterns on genomic DNA [44]. The presence of CpG methylation in promoters and other regulatory sequences can affect gene expression by influencing the binding of RNA polymerases or other transcriptional regulators [45]. Gene silencing via m5C methylation is established by recruitment of methylated CpG-binding proteins to methylated cytosine nucleotides.

Interestingly, one of the better characterized CpG-binding proteins, human MeCP2, functions in complex with a histone deacetylase activity, thereby linking DNA methylation to histone modifications that are also important for transcription silencing [46]. Many prokaryotic promoter regions are regulated by Dam methylation. *E. coli* and *S. typhimurium* Dam deletion mutants exhibit increased expression of genes that are thought to be controlled by promoter methylation in the native state. Intriguingly, Dam methylation regulates the invasion genes and pathogenicity of *S. typhimurium,* and Dam mutant strains of *S. typhimurium* display numerous virulence related defects [47].

Although most eukaryotes are thought to control gene expression primarily at the level of transcription initiation, regulation of gene expression can also be achieved by coordinated

control of transcription elongation, mRNA processing, stability, and even translation [48]. Many studies have been directed towards understanding the post-transcriptional mechanisms controlling specific RNA–protein interactions that target mRNA degradation or prevent access of the ribosome to the translation initiator (start codon). In order to protect the mRNA chain from degradation, the 5' mRNA terminus is capped by 7-methylguanosine, and the 3' end may be trailed by a poly (A) tail of up to 200 adenosine residues. mRNA stability is further determined by internal sequence elements found at the 5' untranslated region (UTR), coding regions within the body of the ORF, and in the 3' UTR. For example, sequences within the mammalian c-Fos coding region and 3' UTR contain AU-rich elements (AREs) that have been shown to be important for mRNA stability [49]. AREs are also found in the 3' untranslated region of transcripts encoding several oncoproteins and cytokines. The ARE binding protein HuR is commonly important for their mRNA stability [49]. Unlike eukaryotes, trypanosomal gene expression is thought to regulate primarily by posttranscriptional mechanisms.

Control of gene expression in trypanosomatids

Trypanosomatid chromosomes contain large polycistronic gene clusters, which can be divergently or convergently transcribed. Mature mRNAs are generated from constitutively transcribed primary transcripts by *trans*-splicing and polyadenylation of their 5' and 3' ends, respectively. A 39 nucleotide miniexon or spliced leader is added to the 5' terminus of an mRNA during *trans*-splicing, and after nuclear processing and transport to the cytoplasm, a tightly controlled turnover mechanism determines the final gene product [48]. DNA regions that separate two polycistronic units are called strand switch regions (SSRs). It is thought that transcription initiates in the SSR and terminates at a downstream SSR or at a specific site of the

same polycistron. Nuclear run-on analysis conducted in *L. major* showed that Pol II dependent transcription of the entire chromosome initiates within the SSR, proceeds bidirectionally in divergent SSRs, and terminates at tRNA-containing convergent SSRs [50]. Notably, sequences specifying transcription initiation and termination have not been identified, with the exception of the Pol II promoter sequence of the SL (spliced leader) gene in trypanosomatids. Using a combination of biochemistry and bioinformatic approaches directed towards study of the SL RNA gene promoter, trypanosomal homologs of TFIIB, TATA-binding protein (TBP), and TFIIH were uncovered. RNAi gene knockdown experiments subsequently revealed the role of these factors in regulation of transcription initiation. Importantly, TFIIH knockdowns decreased the efficiency of SL RNA transcription. Transcription of genes encoding histones and tubulin transcribed by RNA Pol II, and transcription of 18S ribosomal subunits transcribed by RNA Pol I were reduced concomitantly [51].

There are comparatively small numbers of transcription factors and other factors thought to be involved in transcriptional regulation in trypanosomes. For this reason, and because of the presence of a polycistronic transcription mechanism without clear operon arrangements, trypanosomal gene expression is thought to be controlled almost exclusively through post-transcriptional mechanisms [52]. Specific nucleotide sequences within transcribed but untranslated regions 3' of many *T. cruzi* ORFs have been found to control the stability of mRNA in this organism [52]. Importantly, differential expression of stage-specific *T. cruzi* surface proteins can also be controlled via post-transcriptional regulation of gene expression [53], and several recent studies point towards mechanisms by which post-transcriptional events regulate gene expression in a developmentally coordinated manner [54]. This may be achieved through binding of *cis*-elements within the 3' UTR regions of a given mRNA by stage-specific RNA

binding proteins (*trans*-acting factors). For example, one 44 nucleotide AU–rich element (ARE) present in the 3' region of TcMUC mRNA has been shown to be destabilized specifically in the metacyclic *T. cruzi* life stage, but not in epimastgotes [55]. Conversely, the *cis*-acting element stabilizing TcMUC mRNA in epimastigotes was determined to be a 27 nucleotide G–rich element (GRE). *Trans*-acting factors binding to either of these *cis*-elements (AREs or GREs, respectively) were found to differ during the parasite life cycle [56].

DNA sequencing analyses of *L. major*, *T. brucei* and *T. cruzi* revealed that their genomes also encode putative DNA methyltransferases [57]. Consistent with this observation, *T. brucei* and *T. cruzi* have been shown to incorporate 5-methylcytosine in their genomes [58,59]. In *T. brucei,* m5C was identified in large percentage of retrotransposons, intergenic regions, and VSGs that are subject to gene silencing [58]. Hypermethylation of *T. cruzi* DNA sequences induces active cell proliferation, suggesting that DNA methylation may be important for cell division and pathogenesis [59]. However, further studies will be necessary to confirm the significance of DNA methylation in this organism.

Using genome-wide ChIP-seq technology, β-D-glycosylhydroxymethyluracil or base J, was found to colocalize to transcription initiation and/or termination sites of *T. brucei* [38]. This finding suggests that analogous to the role of DNA methylation in repression of gene expression in many other eukaryotes, base J may play an important role in controlling gene expression in trypanosomes. However, these distinct covalent DNA modifications may also have different underlying mechanisms for regulating gene expression. For example, base J does not appear to act as a binding platform for *trans*-acting silencing complexes as does DNA methylation. To date, no single protein binding to J modified DNA has been identified other than JBP1 [34].

We hypothesize that base J reduces the accessibility of DNA to various binding factors due to physical hindrance. Addition of a glucosyl group to DNA makes thymidine residues bulky and modified nucleotides project prominently from the major grove of DNA. Conceivably, base J might sterically hinder the binding of chromatin remolding factors and transcription factors, thereby inhibiting the formation of pre-initiation complexes and limiting the recruitment of RNA polymerase. As demonstrated in the current studies, similar base J localization patterns to those found in *T. brucei* are also observed within *T. cruzi,* indicating that base J may play an important role in control of chromatin structure and Pol II dependent gene expression.

Chromatin structure and regulation of eukaryotic transcription

In eukaryotic cells, nuclear DNA is organized into a compact DNA/protein structure called chromatin. Nucleosomes are the primary units of chromatin, with each nucleosome comprising approximately 147 bp of DNA wrapped around a basic core of histone octamers [46]. Modulation of chromatin structure is essential for regulation of gene expression as it determines the accessibility and recruitment of regulatory proteins to DNA. Chromatin structure can be altered through changes in its physical constitution (*e.g.,* via the incorporation of multiple histone variants) or by covalent modification of its components (*e.g.,* via DNA modifications, post-translational modification of histones) [60].

Post-translational modifications of histone H3 and H4 N-termini dictate the affinity of their interaction with DNA and chromatin associated proteins, and thereby serve to regulate transcription [61]. Common post-translational modifications include methylation of arginine; methylation, acetylation and ubiqutinylation of lysine; and phosphorylation of serine and threonine residues. Post-translational modifications generally associated with active

transcription include hyperacetylation of histone H3 and histone H4, and di- or tri-methylation of H3 at position lysine 4 (H3K4) [46]. Acetylated histones are presumed to be easier to displace from DNA due to changes in net charge, and facilitate active transcription [62]. Conversely, modifications associated with heterochromatin include H3K9 and H3K27 methylation [46]. Repetitive DNA sequences such as those found at the satellite repeats of eukaryotic centromeres and repetitive sequence elements located near the ends of telomeres are generally packaged into constitutive heterochromatin [60,63]. More transient facultative heterochromatin is sometimes found at genomic regions that are subject to developmentally regulated silencing [60].

Chromatin structure and regulation of transcription in trypanosomatids

As in other eukaryotes, developmental regulation of gene expression controlled by changes in chromatin structure has been reported in trypanosomes. In *T. cruzi* epimastigotes, electron-dense chromatin is present in small areas of the nuclear periphery. However, during differentiation into trypomastigotes, electron-dense chromatin extends throughout the nucleus and correlates with a decrease in the rate of transcription [64]. Trypanosomes have a standard complement of core histones (H2A, H2B, H3, and H4), but these histones are more loosely packed than mammalian chromosomes. They also contain histone H2 and H3 variants, such as H2Az, H2BV, and H3V [65]. *T. brucei* and *T. cruzi* also have histone H4 homologs, which share 85% and 96% identity with canonical H4, respectively. The trypanosomatid H2A is homologous to H2A.Z of eukaryotes, whereas trypanosomatid H2B is a trypanosome-specific variant. Both H2A and H2B are associated with repetitive DNA in trypanosomes. The trypanosomatid H3 variant has clear homology to the eukaryotic H3 variant CENP-A [66]. However, trypanosomatid H3 is not incorporated into centromeric nucleosomes, but is only enriched at telomeres [63]. As in most

eukaryotes, the inclusion of distinct histone variants and the distribution of post-transcriptional modifications of histone tails are closely related to functional status and control of gene expression in trypanosomes.

A number of covalent histone modifications, including phosphorylation, acetylation, and methylation, are present in trypanosomes, though these modifications are thought to be less abundant in trypanosomes than in higher eukaryotes [66]. For example, *T. cruzi* histone H1 is has been shown to be phosphorylated, histones H2A and H4 are acetylated and methylated, and H2B and H3 are often methylated, as well. In histone H4, lysine residues K4, K10, K14 and K54 can be acetylated, whereas residues K18 and R53 can also be methylated [67].

Multiple histone modifications have been implicated in the regulation of transcription initiation and termination in trypanosomatids. For example, H4K10 acetylation and H3K4 methylation are concentrated at divergent SSRs from which transcription initiates, and are absent from convergent SSRs at which transcription frequently terminates in *T. cruzi* [68]. Moreover, genome-wide chromosome immunoprecipitation (ChIP-seq) studies of *T. brucei* and *L. major* have shown similar enrichment of acetylated and methylated histones and resultant recruitment of transcription factors to SSRs in each of these organisms [69,70]. The presence of histone H3 acetylation on residues K9 and K14 in transcription initiation sites of *L. major* also appears to coincide with the presence of the transcription factors TRF4 and SNAP50 [69].

Histone modifications are broadly held to create a "histone code" which controls gene expression by creating a chromatin environment that is either permissive or refractory to transcription. In most eukaryotes, various effector complexes are thought to bind to different histone modifications to initiate chromatin remolding and transcription. *T. cruzi* bromodomain factor 2 (BDF2) is the only factor that has been proven to bind to acetylated histones in this

organism [71]. However, in *T. brucei,* BDF3 localization overlapped with genomic regions of enriched H4K10 acetylation, and was enriched across the 5' ends of transcription units. It has therefore been hypothesized that H4K10 acetylation may serve to create a binding site for BDF3. BDF3 association with acetylated histone H4 may serve to recruit transcription factors and chromatin remodeling complexes that incorporate H2AZ and H2BV nucleosomes. Localized H2AZ and H2BV enrichment could in turn facilitate polymerase binding, thereby creating a chromatin environment that is permissive to transcription [70].

More generally, epigenetic regulation of chromatin structure is thought to play a significant but poorly understood role in transcriptional control within trypanosomatids. In the work that follows, we present a novel epigenetic mechanism for regulation of gene expression in the trypanosome *T. cruzi.* We demonstrate that the modified thymidine derivative base J is an important factor controlling gene expression in this organism. Loss of base J coincides with decreased nucleosome abundance, increased levels of histone H3 and H4 acetylation, and increased Pol II occupancy at several promoter regions within the *T. cruzi* genome. Moreover, these changes lead to genome-wide increases in Pol II transcription of PTUs, and global changes in gene expression. We predict that the presence of base J at promoter regions alters chromatin structure and changes the accessibility of DNA binding proteins including RNA Pol II, and we provide considerable evidence in support of this hypothesis. We therefore speculate that base J plays a critical role in regulating gene expression in *T. cruzi*.

REFERENCES

1. Rassi AJ, Rassi A, Marin-Neto JA (2010) Chagas disease. Lancet 375: 1388-1402.

2. Manoel-Caetano Fda S, Silva AE (2007) Implications of genetic variability of *Trypanosoma cruzi* for the pathogenesis of Chagas disease. Cad Saude Publica 23: 2263-2274.

3. Marin-Neto JA, Cunha-Neto E, Maciel BC, Simões MV (2007) Pathogenesis of chronic Chagas' heart disease. Circulation 115: 1109-1123.

4. Bonney KM, Engman DM (2008) Chagas heart disease pathogenesis: one mechanism or many? Curr Mol Med 8: 510-518.

5. Biolo A, Ribeiro AL, Clausell N (2010) Chagas cardiomyopathy-where we stand after a hundred years? Prog Cardiovasc Dis 52: 300-316.

6. Schofield CJ, Jannin J, Salvatella R (2006) The future of Chagas disease control. Trends Parasitol 12: 583-588.

7. Kirchhoff LV, Paredes P, Lomelí-Guerrero A, Paredes-Espinoza M, Ron-Guerrero CS, et al. (2006) Transfusion-associated Chagas disease (American trypanosomiasis) in Mexico: implications for transfusion medicine in the United States. Transfusion 2: 298-304.

8. Tyler KM, Engman DM (2001) The life cycle of *Trypanosoma cruzi* revisited. Int J Parasitol 31: 472-481.

9. De Souza W (2002) Basic cell biology of *Trypanosoma cruzi*. Curr Pharm Des 8.

10. Myler PJ (1993) Molecular variation in trypanosomes. Acta Trop 53: 205-225.

11. Zambrano-Villa S, Rosales-Borjas D, Carrero JC, Ortiz-Ortiz L (2002) How protozoan parasites evade the immune response. Trends Parasitol 18: 272-278.

12. Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33: 29-45.

13. Figueiredo L, Scherf A (2005) Plasmodium telomeres and telomerase: the unusual actors in an unusual scenario. Chromosome res 13: 517-524.

14. Cross GA, Takle GB (1993) The surface trans-sialidase family of *Trypanosoma cruzi*. Annu Rev Microbiol 47: 385-411.

15. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of Trypanosoma trans-sialidase. Annu Rev Microbiol 48: 499-523.

16. Chuenkova M, Pereira ME (1995) *Trypanosoma cruzi* trans-sialidase: enhancement of virulence in a murine model of Chagas' disease. J Exp Med 181: 1693-1703.

17. Leguizamon MS, Campetella OE, Gonzalez Cappa SM, Frasch AC (1994) Mice infected with *Trypanosoma cruzi* produce antibodies against the enzymatic domain of trans-sialidase that inhibit its activity. Infect Immun 62: 3441-3446.

18. Abuin G, Colli W, de Souza W, Alves MJ (1989) A surface antigen of *Trypanosoma cruzi* involved in cell invasion (Tc-85) is heterogeneous in expression and molecular constitution. Mol Biochem Parasitol 35: 229-237.

19. Lima MF, Villalta F (1989) *Trypanosoma cruzi* trypomastigote clones differentially express a parasite cell adhesion molecule. Mol Biochem Parasitol 33: 159-170.

20. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of Trypanosoma trans-sialidase. Annu Rev Microbiol 48: 499-523.

21. Frasch AC (2000) Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. Parasitol Today 16: 282-286.

22. Cross GA, Takle GB (1993) The surface trans-sialidase family of *Trypanosoma cruzi*. 47: 385-411.

23. Schenkman S, Eichinger D, Pereira ME, Nussenzweig V (1994) Structural and functional properties of Trypanosoma trans-sialidase. Annu Rev Microbiol 48: 499-523.

24. Epting CL, Coates BM, Engman DM (2010) Molecular mechanisms of host cell invasion by *Trypanosoma cruzi*. Exp Parasitol 126: 283-291.

25. Chuenkova M, Pereira ME (1995) *Trypanosoma cruzi* trans-sialidase: enhancement of virulence in a murine model of Chagas' disease. J Exp Med 181: 1693-1703.

26. Urban I, Boiani Santurio L, Chidichimo A, Yu H, Chen X, et al. (2011) Molecular diversity of the *Trypanosoma cruzi* TcSMUG family of mucin genes and proteins. Biochem J.

27. Acosta-Serrano A, Almeida IC, Freitas-Junior LH, Yoshida N, Schenkman S (2001) The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. Mol Biochem Parasitol 114: 143-150.

28. Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. 4: 229-236.

29. Borst P, van Leeuwen F (1997) beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. Mol Biochem Parasitol 90: 1-8.

30. Sabatini R, Cliffe L, Vainio S, Borst P (2009) Enzymatic Formation of the Hypermodified DNA Base J (β-D-Glucopyranosyloxymethyluracil) DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.

31. Cross MR, Kieft R, Sabatini M, Wilm M, de Kort GA, et al. (1999) The Modified Base J is the Target for a Novel DNA-Binding Protein in Kinetoplastid Protozoans. EMBO J 18: 6573-6581.

32. DiPaolo C, Kieft R, Cross M, Sabatin R (2005) Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Mol Cell 17: 441-451.

33. Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62: 235-251.

34. Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, et al. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res 35: 2107-2115.

35. 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 Res 37: 1452-1462.

36. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309.

37. Kieft R, Brand V, Ekanayake DK, Sweeney K, DiPaolo C, et al. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form *Trypanosoma brucei*. Mol Biochem Parasitol 156: 24-31.

38. Cliffe LJ, Siegel TN, Marshall M, Cross GA, 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 Res 38: 3923-3935.

39. Morse RH (2007) Transcription factor access to promoter elements. J Cell Biochem 2007 Oct 15;102(3):560-70 102: 560-570.

40. Hager GL, ., McNally JG, Misteli T (2009) Transcription dynamics. Mol Cell 35: 741-753.

41. Yosef N, Regev A (2011) Impulse control: temporal dynamics in gene transcription. Cell 144: 886-896.

42. Roeder RG (2005) Transcriptional regulation and the role of diverse coactivator in animal cells. . FEBS Letters 579: 909-915.

43. Perez JC, Groisman EA (2009) Transcription factor function and promoter architecture govern the evolution of bacterial regulons. Proc Natl Acad Sci 106: 4319-4324.

44. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31: 89-97.

45. Bergman Y, Mostoslavsky R (1998) DNA demethylation: turning genes on. Biol Chem 379: 401-407.

46. Quina AS, Buschbeck M, Di Croce L (2006) Chromatin structure and epigenetics. Biochem Pharmacol 72: 1563-1569.

47. Heusipp G, Falker S, Schmidt MA (2007) DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol 297: 1-7.

48. Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010: 525241.

49. Peng SSY, Chen CA, Xu N, Shyu AB (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. The EMBO J 17: 3461 - 3470.

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

51. Lecordier L, Devaux S, Uzureau P, Dierick JF, Walgraffe D, et al. (2007) Characterization of a TFIIH homologue from *Trypanosoma brucei*. Mol Microbiol 5.

52. Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156: 93-101.

53. Abuin G, Freitas-Junior LH, Colli W, Alves MJ, Schenkman S (1999) Expression of trans-sialidase and 85-kDa glycoprotein genes in *Trypanosoma cruzi* is differentially regulated at the post-transcriptional level by labile protein factors. J Biol Chem 274: 13041-13047.

54. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics 10: 495.

55. Di Noia JM, D'Orso I, Sánchez DO, Frasch AC (2000) AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. J Biol Chem 275: 10218-10220.

56. D'Orso I, Frasch AC (2001) Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. J Biol Chem 276: 15783-15793.

57. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005 Jul 15;309(5733):409-15 309: 409-415.

58. Militello KT, Wang P, Jayakar SK, Pietrasik RL, Dupont CD, et al. (2008) African Trypanosomes Contain 5-Methylcytosine in Nuclear DNA. Eukarotic Cell 7: 2012-2016.

59. Rojas MV GN (1991) Relationship between DNA methylation and cell proliferation in *Trypanosoma cruzi*. FEBS Lett 295: 31-31.

60. Mellor J (2006) Dynamic nucleosomes and gene transcription. Trends Genet 22: 320-329.

61. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707-719.

62. Mizzen CA, Allis CD (1998) Linking histone acetylation to transcriptional regulation. Cell Mol Life Sci 54: 6-20.

63. Lowell JE, Cross GA (2004) A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. J Cell Sci: 5937-5947.

64. Elias MC, Faria M, Mortara RA, Motta MC, de Souza W, et al. (2002) Chromosome localization changes in the *Trypanosoma cruzi* nucleus. Eukaryot Cell 1: 944-953.

65. Daniels JP, Gull K, Wickstead B (2010) Cell biology of the trypanosome genome. Microbiol Mol Biol Rev 74: 552-569.

66. Horn D (2007) Introducing histone modification in trypanosomes. Trends Parasitol 6: 239-242.

67. da Cunha JP, Nakayasu ES, de Almeida IC, Schenkman S (2006) Post-translational modifications of *Trypanosoma cruzi* histone H4. Mol Biochem Parasitol 150: 268-277.

68. Respuela P, Ferella M, Rada-Iglesias A, Aslund L (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in T*rypanosoma cruzi*. J Biol Chem 283: 665-673.

69. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. BMC Genomics 8: 152-158.

70. 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 Dev 23: 1063-1076.

71. Villanova GV, Nardelli SC, Cribb P, Magdaleno A, Silber AM, et al. (2009) *Trypanosoma cruzi* bromodomain factor 2 (BDF2) binds to acetylated histones and is accumulated after UV irradiation. Int J Parasitol 39: 665-673.

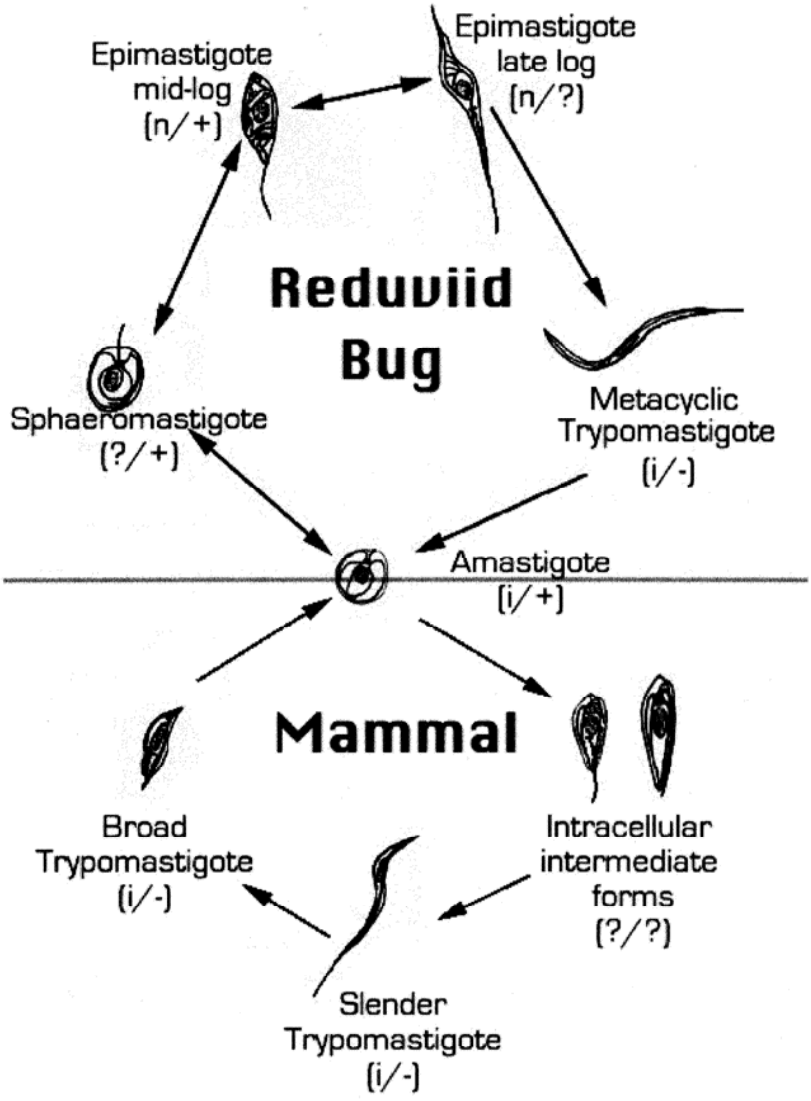


Figure1.1 The life cycle of *Trypanosoma cruzi* adapted from Tyler et al., 2001; (i) is an infective form; (n) is non-infective; (+) represents a proliferative form; and (−) is non-proliferative.

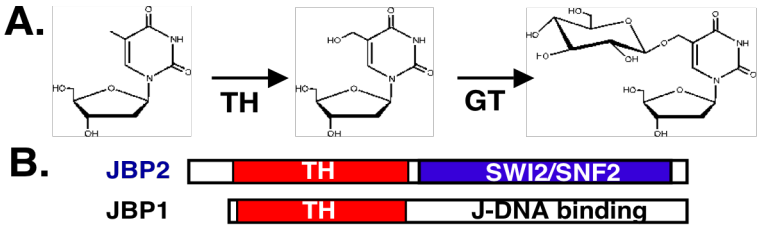


Figure 1.2 A. Regulation of J Biosynthesis. The proposed pathway of J bio synthesis indicating the two thymidine hydroxylases (JBP1 and JBP2) involved in first step of J bio synthesis. TH, Thymidine Hydroxylase, GT, Glucosyl transferase. **B**. Alignment of JBP1 and JBP2 showing the TH domain and J binding domains

CHAPTER 2

TELOMERIC CO-LOCALIZATION OF THE MODIFIED BASE J AND CONTINGENCY GENES IN THE PROTOZOAN PARASITE *TRYPANOSOMA CRUZI*

---

[1]**Ekanayake DK**, Cipriano MJ, Sabatini R. **Nucleic Acids Res**. 2007;35 (19):6367-77.

ABSTRACT

Base J or α-D-glucosylhydroxymethyluracil is a modification of thymine residues within the genome of kinetoplastid parasites. In organisms known to contain the modified base, J is located mainly within the telomeric repeats. However, in *Trypanosoma brucei,* a small fraction of J is also located within the silent subtelomeric Variant Surface Glycoprotein (VSG) gene expression sites, but not in the active expression site, suggesting a role for J in regulating telomeric genes involved in pathogenesis. With the identification of surface glycoprotein genes adjacent to telomeres in the South American Trypanosome, *Trypanosoma cruzi*, we became interested in the telomeric distribution of base J. Analysis of J and telomeric repeat sequences by J immunoblots and Southern blots following DNA digestion, reveals ~25% of J outside the telomeric repeat sequences. Moreover, the analysis of DNA sequences immunoprecipitated with J antiserum, localized J within subtelomeric regions rich in lifestage-specific surface glycoprotein genes involved in pathogenesis. Interestingly, the pattern of J within these regions is developmentally regulated. These studies provide a framework to characterize the role of base J in the regulation of telomeric gene expression/diversity in *T. cruzi*.

INTRODUCTION

Trypanosoma cruzi is a parasitic protozoan causing Chagas' disease, a debilitating and incurable disease affecting millions of people in Latin America [1]. The lifecycle of *T. cruzi* shows multiple developmental stages both in the insect vector (triatome) and the mammalian host. Within the insect vector there are two stages, the replicative epimastigotes in the mid gut and metacyclic trypomastigotes in the distal intestine. Infective metacyclic forms, in insect faeces, invade a wide variety of mammalian cells through skin abrasion or mucous membranes of mammalian hosts. They transform into amastigotes and after several intracellular binary divisions, tissue derived trypomastigotes escape into the blood stream [2]. Their survival within the insect gut and mammalian host depends on the parasites ability to regulate the expression of a large number of stage specific surface antigens that prevent an adequate immune response and allow optimal interactions with, and invasion of, the host cell. Among these antigens are the large multi-gene family described as mucin and trans-sialidase (TS) superfamilies, which encode lifestage-specific surface glycoproteins. Proteins of the TS superfamily are likely to be the most abundant proteins on the surface of the infective forms of *T. cruzi*, and genes encoding these proteins may represent about 5% of the parasite genome [3]. While there is great interest in the functional diversity and involvement of these proteins in interactions with cells and immune system of the hosts, little progress has been made concerning how these genes are organized in the genome and expressed.

For many pathogenic organisms, telomeric regions of the genome are involved in homologous and ectopic recombination events and acquisition of heterologous gene sequences to diversify, as well as regulate the expression of, genes involved in virulence [4]. For example, the telomere and subtelomeric regions of the blood protozoan parasites *Trypanosoma brucei* and

Plasmodium falciparum are specialized in the expression of specific variant-surface genes that allow the parasite to evade the host immune system [5]. In the case of *T. brucei*, which reside primarily within the mammalian bloodstream, evasion of the host immune response is achieved by regularly changing their variant surface glycoprotein (VSG) coat in a process termed antigenic variation. While each trypanosome contains ~1000 VSG genes, the cell only expresses one at a time. The active VSG is found within one of about 20 telomeric expression sites (ESs). In order to exclusively express a particular VSG, only one ES is active at a time, while the others are silenced. The preferred subtelomeric region in the genome of these organisms is believed to facilitate gene switching and expression, and the generation of new variants.

Analysis of the silent telomeric expression sites in *T. brucei* led to the discovery of the novel modified DNA base called α-D-glucosylhydroxymethyluracil, or base J [6] . This hypermodified base, consisting of a glucose moiety attached to thymine residues, is present in telomeric DNA of all major representatives of kinetoplastidea as well as in the DNA of two distantly related organisms; *Diplonema* and *Euglena* [7,8]. Base J is most likely synthesized in two steps: first a thymine hydroxylase converts a thymidine residue in DNA to hydroxymethyl deoxyuridine (HOMedU), then a glucosyltransferase converts HOMedU into J by the addition of a glucose moiety [9]. In *T. brucei*, J is also found in other DNA repeat sequences such as the subtelomeric 70bp, 50bp and 177bp repeats, and the SL RNA and 5S rRNA genes [10]. The presence of J within the ~19 inactive telomeric VSG ESs but not within the single active ES has suggested that base J is involved in the repression of telomeric gene expression and/or DNA recombination and thus, to the regulation of antigenic variation [11].

In the case of *T. cruzi*, it is becoming increasingly clear that representative members of the TS superfamily, and other surface glycoprotein genes involved in pathogenesis, are localized

to the subtelomeric regions of the chromosome [12,13,14]. It is speculated that telomeric localization in *T. cruzi* may play a role in the evolution of the genetic diversity of these surface glycoproteins and thus increasing their capability to endure the host defense (5,14). While studies of base J in *T. brucei* have suggested its role in telomeric gene regulation, the organization and function of the modified base is poorly studied in other kinetoplastids including *T. cruzi*. To begin to explore the role of J in telomeric gene expression/diversity in *T. cruzi,* we have characterized the distribution of the modified base in more detail. We find a significant fraction of the total J (~25%) is present outside the telomeric repeats. Analysis of the subtelomeric regions has localized J within sequences up to 50 kb from the telomeric repeats. Interestingly, we find the developmental regulation of J biosynthesis within several telomeric regions that contain members of the TS superfamily and other surface glycoprotein genes involved in pathogenesis. The significance of these results in terms of the biological function of J in *T. cruzi*, as well as the global function of base J in kinetoplastids, will be discussed.

MATRIALS AND METHODS

Maintenance of parasite cultures

T.cruzi epimastigotes of the Y strain was grown in LIT media containing 10% fetal calf serum at 28^0C as described earlier [15] . In vitro metacyclogenesis was performed according to the previously described methods. Briefly, epimastigotes growing in LIT were harvested by centrifugation and resuspended in Triatome Urine media (TAU) to a concentration of $3x10^5$ /ml and incubated for 2h at 28^0C. Parasites are then transferred to 150cm^2 culture flasks containing

TAU3AAG media (supplemented with 10mM Proline, 50mM L-Glutamate and 50mM L-Aspartate and 10mM Glucose). Parasites are then incubated for 2-7 days at 28^0C [16].

Metacyclics were added to Vero cells at 1:10 cells to parasite ratio and grown in Eagle's Minimum Essential Medium (MEM) supplemented with 10% fetal calf serum. Unattached metacyclics are removed by washing after 24h of growth and released trypomastigotes were harvested every day.

Anti J Immunoblotting of terminal restriction digests and dot blots

Anti J immunoblots were used to quantify the genomic level of J as previously described [7]. Briefly, serially diluted epimastigotes and trypomastigotes DNA was blotted onto nitrocellulose and incubated with anti -J antisera and detected with HRP conjugated goat-anti-rabbit antibodies and visualized with ECL. The level of DNA loading was determined by hybridization using a ^{32}P random-prime labeled tubulin probe.

In order to quantitate the amount of J outside telomeres in *T. cruzi*, J-containing fragments were detected on Southern blots as described in [10]. Briefly, 200ng of epimastigotes and trypomastigotes DNA was digested using frequently cutting restriction enzymes and size fractionated in a 0.5% agarose gel. Fragments were then blotted on to a nylon membrane (Hybond N+, Amersham Bioscience, Upsala, Sweden) and blocked for 2h in TBST with 5% milk. Immuno detection was performed using HRP conjugated goat-anti-rabbit antibodies in combination with enhanced chemi-luminescence (Amersham Bioscience) after incubation with anti-J anti sera. The membrane was then hybridized with a ^{32}P end-labeled telomeric oligonucleotide $(TTAGGG)_5$ probe in hybridization buffer at 55^0C and washed 3 times with 6X SSC. The resulting signals from the telomeric Southern blot and the J-immunoblot were then

compared. The regions of the blot reacting with J antisera but not with the telomeric probe were quantitated by phosphoimager analysis. The optical density of the non-telomeric localized J was divided by the optical density of the whole lane following correction for background.

Analysis of telomeric regions of the *T.cruzi* genome

T. cruzi CL Brener strain scaffold (CH473309-CH473946) and contig (AAHK01000001-AAHK01032746) information was downloaded from Genbank genome project database (AAHK00000000). A total of 32,746 singleton contigs and 29,495 scaffolds were imported into a local database. Telomeric repeats were annotated by searching the *T. cruzi* genome using known hexameric repeat sequence (5'-CCCTAA-3'). It searches for the known telomere repeat sequence, though requiring at least 3 copies of the sequence in a row in order to annotate that area. Original gene calls were imported from the Genbank entries. No additional gene calls were automatically made. Contig entries were reverse complemented from their original Genbank entries in order for scaffolds to show correct contig alignment. We refer to these regions as S-00 to denote the representative super-contig of the *T.cruzi* sequencing database. Any questionable overlapping regions during the assemblies were confirmed by PCR. Gene calls were annotated by blastp searching against nr, swissprot, mitop, and HMM searching against Pfam_ls [17]. All contigs between 100bp and 10kb were submitted to nr and swissprot via blastx algorithm. Intergenic regions were obtained and searched with blastx against nr and swissprot. TMHMM and SignalP algorithms were applied to all gene calls [18,19]. Interpro searches were also performed on all gene calls [20]. The whole assembly was run through nucmer searching for intra-genomic matches of size 40 or greater and annotated within the genome.

J-Immunoprecipitation

J-DNA containing fragments were immunoprecipitated and quantitated by dot-blot analysis as described in [7]. Briefly, genomic DNA was sonicated to 0.5-3kb fragments and incubated with J-antisera in TBSTE buffer containing 10mM Tris pH 8.0, 150 mM NaCl, 0.2% Tween-20, 2 mM EDTA, 0.1 mg tRNA/ml and 1 mg BSA/ml. The DNA was immunoprecipitated by the addition of protein G agarose, washed 4 times and spotted onto nitrocellulose along with a fraction (10%) of the input DNA. The membrane was hybridized with random-primed labeled PCR probes corresponding to the sequences of interest. The exact sequences of the DNA oligonucleotides used to make the probes are available upon request. The hybridization signals were quantitated by phosphoimager analysis.

In order to localize J within subtelomeric regions of *T. cruzi*, genomic DNA was digested with various restriction enzymes prior to J-immunoprecipitation. Restriction enzymes were selected based on the contig assembly to isolate specific fragments for Southern hybridization and PCR analysis. The immunoprecipitated DNA was either fractionated on a 0.5% agarose gel and analyzed by Southern hybridization (as described above) or used in a PCR reaction. Specific primers used in the PCR analysis were designed using the *T. cruzi* sequence database for selected regions. Sequences of all primers used in this analysis are available upon request.

RESULTS

Base J is developmentally regulated and present outside the telomeric repeat sequences in *Trypanosoma cruzi*

In all kinetoplatids analyzed thus far, base J is abundantly found within telomeric repeat sequences [7] . In the African trypanosome, J biosynthesis is restricted to the mammalian bloodstream life-stage where 50% of the total amount of J is localized within the telomeric repeats. Previous analysis of J in *T. cruzi* identified the modified base in the telomeric repeats and suggested its synthesis is developmentally regulated [7]. However, it was unclear from these studies whether the DNA samples representing the different life-stages were from the same strain of *T. cruzi*. It has since been shown that telomere length varies significantly between different strains of *T. cruzi* [21]. Thus, the level of J between different strains may be difficult to compare if the amount of J is proportional to the length of the telomere. In order to alleviate this problem, we isolated DNA from Y strain cultures of *T. cruzi* representing epimastigote and trypomastigote life-stages and analyzed the levels of base J by dot blot hybridization using anti-J antibodies. DNA loading was checked by southern hybridization with a *T.cruzi* specific trans-sialidase gene probe. As shown in Figure 2.2A, there is about a 2-fold increase in the amount of base J in the mammalian life stage. Similar results were found with the CL-Brener strain (data not shown).

In order to examine the regulated J-synthesis in telomeric sequences, genomic DNA fragments immunoprecipitated with the J-specific antisera were analyzed by dot-blot hybridization (Figure 2.1B). In agreement with the previous study, we found that there is a similar ~2-fold up regulation of the level of J in the telomeric repeats of trypomastigotes compared to epimastigotes. These results confirm the regulation of base J synthesis in *T.cruzi*

life cycle and suggest that telomeric localized J accounts for a significant fraction of the regulated J synthesis.

In order to quantitate the fraction of J localized outside the telomeric repeats, we performed terminal restriction digestions of genomic DNA followed by J immunoblot and Southern hybridization. To do this, we used a series of four base cutters (RsaI, AluI and MspI restriction enzymes) to separate out telomeric repeat regions from the rest of the chromosome. Telomeric repeats are not digested by these enzymes and result in a telomeric fragment of ~500bp, corresponding to the mean telomeric length in the Y strain [21]. The digested DNA was size fractionated by electrophoresis in an agarose gel, blotted and incubated with the J-antisera followed by hybridization with a telomeric probe. As shown in Figure 2.2A, most of the telomeric fragments co-migrated with the J DNA suggesting that the telomeric repeats contain high quantities of the modified base. However, a fraction of the digest did not hybridized with the telomeric probe but did react with anti J antibody. To examine the sensitivity of our assays, we increased the amount of digested genomic DNA in two fold dilution series (Figure 2.2B). By this approach, we found that even with increased amounts of DNA we did not detect smaller telomeric fragments within 100-250bp range whereas, the smaller J-DNA fragments were readily detectable by the J immunoblot. Furthermore, we see equivalent levels of detection of telomeric fragments of *T.cruzi* (~500bp) and *T.brucei* (~10kb). Thus, the lack of detectable shorter telomeric DNA fragments in *T. cruzi* is not due to any differential ability to detect 500bp verses 250bp of telomeric DNA on the Southern blot.

Quantification of the J-blot in Figure 2.2A indicates ~23% and ~26% of the total J was found outside the telomeric repeats in the epimastigote and trypomastigote life-stage, respectively.

These results demonstrate that there is a similar and significant level of base J present outside the telomeric repeat regions of the insect and mammalian life stages of *T. cruzi*.

Localization of base J in tandem repeat arrays

In *T.brucei* DNA base J was found in tandem repeat sequences that are primarily subtelomeric and some repeat arrays that are chromosomal internal. This includes the subtelomeric 50bp and 70bp repeats, the 177bp repeats (satellite repeats) in the minichromosomes, and the chromosome internal long arrays of 5S RNA and SL RNA gene repeats [10]. Many of these J containing sequences in *T. brucei* are lacking in *T. cruzi* [22]. Therefore, we tested several of the known repetitive sequences of *T. cruzi*, such as the satellite sequences (195 bp repeats), DIRE, L1Tc, VIPER and DGF-1. We also looked for the presence of J in repetitive genes like the SL RNA, the 18S and 24S rRNA as well as the □-tubulin gene cluster. The results of these experiments are presented in Table 2.1. Using sonicated DNA, we find significant immunoprecipitation of all the repetitive sequences tested with the exception of the 195 bp satellite repeats. The 0.003% immunoprecipitation of tubulin fragments is consistent with background immunoprcipitation observed with DNA without J [7,10,23].

All trypanosomatid ribosomal gene sequences and SL RNA genes exist in tandem arrays [24]. In order to determine whether the ribosomal genes are modified, we used J-immunoprecipitation of sonicated DNA followed by dot-blot hybridization. As shown in Table 2.1, the small and large subunit rRNA genes are significantly immuno-precipitated. The trans-spliced 39mer region of SL RNA gene is highly conserved between species while the 3' non-spliced portion varies in size and sequences. At least 10% of SL RNA genes in *T. cruzi* are interrupted by a site-specific retrotransposon element, called cruzi associated retrotransposon

(CZAR), between nucleotides 11 and 12 of the SL 39mer [25]. To determined whether SL RNA gene units are modified, we used sonicated DNA and immuno-precipitated with anti J antibodies, dot blotted onto a nylon membrane and hybridized using probes representing the SL RNA transcribed region and non-transcribed region. As indicated in Table 2.1**,** the SL RNA transcribed region showed a slight immuno-precipitation whereas the non-transcribed region showed a significant immuno-precipitation. These results were confirmed by Southern blot analysis (data not shown) and are comparable with the analysis of J within the SL RNA coding and non transcribed region in *T. brucei* [10]. However, the CZAR site-specific retro-element of *T.cruzi* did not contain base J (data not shown).

Subtelomeric localization of J

As shown in Table 2.1, we obtained significant immunoprecipitation from repeat arrays, including DGF-1, DIRE, L1Tc and VIPER. These repeat sequences are known to be adjacent to telomeric repeats in some chromosomes and may have been pulled down in sonicated DNA because of their linkage to the J containing telomeric repeats. To more closely examine the presence of J within these sequences, we needed a restriction map of representative subtelomeric arrays and re-examine the ability to pull-down specific repeat sequences that have been cleaved from the telomeric repeat arrays. Furthermore, sequence assemblies representing telomeric end of several chromosomes would allow the detailed localization of base J throughout the subtelomeric region. In order to assemble large telomeric scaffolds (20-400 kb in size), we pulled out 70 *T. cruzi* telomeric contigs from *T. cruzi* DB using the known hexameric repeat sequence (5'-CCCTAA-3'). Contigs of these subtelomeres were further assembled into scaffolds to comprehensively expand the telomeric details. These assembled telomeres were annotated

using genbank annotations and GMOD and GBrowse was used to visualize the genes, pseudogenes and repeat regions. All of the subtelomeric assemblies have the varying length of telomeric repeats followed by telomere associated 189 bp telomeric junction sequences specific to *T.cruzi* [22]. At least 25 of them consisted of trans-sialidase genes or pseudogenes and many contain retro elements like VIPER, DIRE and L1Tc. VIPER sequences, DGF-1 and retro transposon hot spot genes are particularly abundant in these subtelomeres. Ten percent of the subtelomeres we assembled have mucin genes, pseudogenes and strand switch regions within 40 kb of the telomeric repeats. Since we performed our initial assemblies, *T. cruzi* DB has updated their database to include similar scaffold assemblies including telomeric details. Our assemblies are consistent with the scaffold details now available from *T.cruzi* DB.

Utilizing this telomeric contig database, we were able to examine the localization of J within internal versus telomeric localized repeat arrays. To do this, we designed specific primers to distinguish internal versus subtelomeric members of the same repeat sequences (see supplementary Figure 2.1A and B for more detail). Furthermore, we cloned and sequenced all the PCR products to confirm the specific identity of these regions. Internal repeat sequences were chosen based on their location in the genome scaffolds and the presence of adjacent internal gene arrays (i.e. ribosomal genes). See Table 2.2 for a summary of the assemblies analyzed representing telomeric versus internal repeat regions. Digested DNA was immunoprecipitated and analyzed by PCR using primers against the internal or subtelomeric localized sequences. As demonstrated in Figure 2.3, internal members of DGF-1, DIRE and L1Tc did not contain DNA base J but their subtelomeric counterparts were readily immunoprecipitated with anti J antibodies. This result was consistent between epimastigotes and trypomastigote life stages. All PCR products were cloned and sequenced to confirm identities. VIPER sequences are generally

found in subtelomeres, and thus, we were able to get a positive reaction for subtelomeric VIPER sequences (Figure 2.5 and data not shown). However, due to the fact that we were unable to distinguish internal versus subtelomeric members of VIPER, we are unable to conclude that only subtelomeric VIPER sequences are modified by base J. These results demonstrate that DNA base J is present in the subtelomeric localized repeat arrays of *T.cruzi* chromosomes throughout the life-cycle.

To determine the distribution of base J in more detail, we generated restriction maps for four subtelomeric regions (see Figures 2.4 and 2.5). Genomic DNA digested with the indicated restriction enzymes was immunoprecipitated with J antibodies and used in a PCR reaction with specific primers designed to amplify defined regions of the subtelomere using a similar approach described for the analysis of the telomeric repeat arrays (Figure 2.3 and Supplementary Figure 2.1). As shown in Figure 2.4B, all regions analyzed, up to 30 kb from the telomeric repeats, in contig 197 are immunoprecipitated using DNA from the trypomastigote life-stage. In contrast, while J is present up to 30 kb from the telomeric repeats, two regions of the subtelomere in epimastigote DNA lacks J. As a negative control, tubulin gene was not immunoprecipitated with J antibodies. Specificity of the PCR assay was confirmed by direct sequencing of the PCR products (data not shown) and by Southern analysis (Figure 2.4C-E). In the Southern blot analysis, immunoprecipitated DNA was size fractionated and hybridized using the PCR fragments as probes. In many cases, complete digestion and the developmental regulation of J in particular regions is clearly verified by the Southern blot analysis (for example see Figure 2.4E). The subtelomeric regions analyzed are rich in sequences representing members of large gene families (i.e. TS and RHS). Therefore, while the PCR primers are specific, the amplified products may cross-hybridize to various fragments on the Southern blot. Interestingly, only a

select fraction of the cross-hybridized fragments are immunoprecipitated. Whether these cross-hybridized fragments localize to telomeric regions of other chromosomes remains to be tested.

In order to further examine the developmental regulation and subtelomeric localization of J, we analyzed three additional telomeric contigs. The results from the PCR analysis of epimastigote and trypomastigote J-DNA immunoprecipitation in these regions, is presented in Figure 2.5. All results were confirmed by southern hybridization (data not shown). The lack of J in regions immediately adjacent to telomeric repeats provides additional support that the digests were complete and false positive IP due to the linkage to telomeric repeats is unlikely. Therefore, our analysis of four telomeric contigs indicates that subtelomeric regions of *T. cruzi* extending to at least 50 kb from the telomeric repeats, including trans-sialidase genes and pseudogenes and repeat regions such as VIPER and DGF-1, are modified by base J. Interestingly, some regions containing trans-sialidase genes and pseudogenes (S-197, S-673, S-651) showed a developmental regulation where base J was absent in epimastigote stage. This suggests that there is not only localization pattern but also developmentally regulated distribution pattern of J biosynthesis in this region the *T.cruzi* genome.

DISCUSSION

β-D-glucosyl-hydroxymethyluracil or base J is a unique DNA modification found in kinetoplastids. It was originally discovered in bloodstream *T. brucei* as a DNA modification localized to the silent telomeric VSG expression sites. Quantitative analysis of *T. brucei* genomic DNA has indicated that base J is primarily present in simple repeat sequences, including the subtelomeric repeats flanking and within the VSG expression sites, with approximately 50% located specifically within the telomeric repeats [11,26]. However, it is the correlation of the developmental regulated localization of J within the subtelomeric region and the silencing of telomeric expression sites in *T. brucei* that has suggested the biological role of the modified base in regulating antigenic variation. J is also conserved in the telomeric repeat sequences of other kinetoplastids including *T.cruzi*, *Leishmania* and *Crithidia* except in *Euglena* where total J is found exclusively outside the telomeres [7,8]. A recent study on *Leishmania* and *Crithidia* showed that ~98% of the modified base is localized in its telomeric repeat sequences [27]. This would suggest that J cannot be involved in telomeric gene silencing in *Leishmania* and *Crithidia*, since J is mostly a modification of telomeric sequences. There is little knowledge on the localization and distribution pattern of base J in the South American trypanosome, *T. cruzi*.

It is becoming increasingly clear that members of the stage-specific surface glycoprotein gene families of *T. cruzi* that have been implicated in pathogenesis (i.e. TS and mucins) are localized to telomeric regions of the chromosome. For many pathogenic organisms, the telomeric regions are involved in homologous and ectopic recombination events and acquisition of heterologous gene sequences to diversify contingency gene families involved in virulence [4]. In the case of *T. cruzi*, it has been speculated that telomeric localization of these surface

glycoproteins may play a role in the evolution of their genetic diversity and thus increasing the parasites ability to endure host defenses [4,14,28,29]. In order to begin to explore the role of J in telomeric gene expression/diversity in *T. cruzi*, we have characterized the distribution of the modified base in more detail.

We find that base J synthesis is developmentally regulated in *T.cruzi*, with 2-fold up-regulation during the infective mammalian life stage, and ~25% of the total J localized outside of the telomeric repeats. Similar to the distribution of J in *T. brucei*, we find J primarily within simple repeat sequences in *T. cruzi*. These include the SL RNA and the 24S and 5S RNA gene repeats. While the amount of J may be low in these internal repetitive gene arrays, as indicated by the % IP, telomeric localized repeat sequences (i.e. VIPER, DIRE) contain significant levels of J. However, when these same repeat sequences are examined when they are present within the chromosome, they do not contain the modified base. The lack of J within these sequences when they are localized within the chromosome, as well as the lack of J within the 195 bp satellite repeats, suggests that it is not the sequence or repetitive nature per se that determines whether a particular DNA region contains J, only that it is localized near the telomeric end of the chromosome. The ability to analyze the same DNA sequence in different locations in the *T. cruzi* genome has allowed us, for the first time, to directly examine the effect of genome context on the regulation of J biosynthesis. Presumably, some aspect of the telomeric context is able to optimally recruit JBP2; the key thymine hydroxylase involved in stimulating de-novo J synthesis [30,31].

Using the telomeric contig assemblies, we have localized the modified base within the subtelomeric regions of *T. cruzi* that are rich in stage-specific contingency genes involved in pathogenesis. In order to examine the localization of J within four distinct telomeric assemblies,

we used different restriction enzymes to separate out the telomeric repeat sequences from the regions of interest. We are aware that partial restriction enzyme digestions and the sensitivity of the PCR technique may give rise to false positive results. The efficiency of the restriction digestion was evaluated by repeating the PCR analysis using different genomic DNA preps and digestions at least three times. In two subtelomeric regions examined, large areas adjacent to the telomeric repeats sequences were found to be negative for base J. This further supports the efficiency of restriction digestion as otherwise those regions would have been amplified along with the incomplete digested fragments containing telomeric repeat sequences. We also performed Southern hybridization on the IP fractions for all the regions tested with PCR to confirm the results. *T.cruzi* genome is highly redundant and contain large amount of pseudogenes due to the constant duplication and recombination events [29]. This complicated the Southern hybridization reactions as most of the probes cross hybridize to similar sequences throughout the genome and thus different restriction fragments. We attempted to minimize this by searching for sequences within subtelomeres which failed to indicate significant similarities to other regions of the genome via blast analysis of the *T. cruz*i database. While the cross-hybridization problem complicated the southern blot analysis, we were able to generate similar overall J distribution profiles as determined via the PCR procedure when analyzing a given subtelomeric region. The 5' flanking regions of the targeted telomeric gene sequences was used to ensure the specificity of the PCR reaction. The lack of J in regions immediately adjacent to telomeric repeats, including the 189 bp junction region (data not shown), and the reproducible differential localization of J throughout the subtelomeric region of epimastigote versus trypomastigote DNA, further verifies the specificity of the analysis. These results also suggest

that the J present in the subtelomeric region is specific and not due to any simple diffusion of J synthesis in from the telomeric repeats.

In the analysis of the distinct telomeric assemblies, we observed a differential localization pattern for the distribution of base J within the subtelomeric regions corresponding to epimastigote versus trypomastigote life-stages. Interestingly, this pattern was observed primarily within regions rich in trans-sialidase genes and pseudogenes. The biological significance of this apparent developmentally regulated J localization pattern is currently unclear. The TS gene family has a stage specific expression pattern where a different set of TS genes are expressed in the insect versus mammalian life stage [32,33]. However, it is not clear to what extent the telomeric localized members of any of the surface glycoprotein gene families are expressed. Whether base J is involved in the regulated expression of these stage specific genes is under investigation.

The results presented here suggest that like in *T. brucei*, base J is mainly a modification of telomeric localized repetitive DNA sequences in *T. cruzi*. Base J is thought to be involved in generalized repression of transcription as it was initially found in inactive expression sites of *T.brucei* suggesting a role in antigenic variation. The significant fraction of J within DNA repeats in *T. brucei* also suggested that base J is involved in repression of recombination between repetitive sequences or retro elements and thus, stabilizing the genome. In fact, the majority of VSG switching events are due to DNA rearrangements within the subtelomeric repetitive DNA sequences. Furthermore, manipulation of the levels of J in the genome of *T. brucei* has supported its role in regulating telomeric DNA rearrangements [23]. It would appear that *T.cruzi* adopted a different mechanism to survive in the host environment. Rather than the mono-allelic expression of a single variant surface glycoprotein, *T. cruzi* cells simultaneously expresses large

number of surface glycoproteins including trans-sialidases and mucins [13,34]. Although there is no evidence for telomeric expression sites in *T. cruzi*, we frequently find base J in these subtelomeric contingency gene families along with repetitive sequences/retro elements. Therefore, it is tempting to propose a role for base J in regulating recombination in the subtelomeric region of *T.cruzi.*

Along with surface glycoprotein genes, *T.cruzi* subtelomeres are enriched in retro elements such as VIPER, DIRE and L1Tc. We find base J in the subtelomeric retroelements of *T.cruzi* while their internal counterparts lack this modification. Retro elements are the mobile genetic elements found in the genome of many organisms and represent a potent force or driver of genomic evolution [35,36]. These regions are the sites for genetic exchange between reciprocal chromosomes and also from ectopic loci. Studies in yeast have shown that intra-chromosomal cross over due to the homologous recombination events between retro elements cause deletion, duplication and inversions of flanking regions whereas inter-chromosomal crossover resulted from ectopic recombination leads to translocation [37]. Retro elements constitute approximately 45% of the human genome and they are modified by methylcystosine [38]. DNA methylation of these regions suppresses the activation of the retro elements and prevents illegitimate recombination between elements located in the different sites. It has been shown that hypomethylation of LINE element (non- LTR-retro element) in various human cancer cells facilitate illegitimate recombination contributing chromosomal instability [39]. Whether J plays a similar role in DNA recombination and genome stability in kinetoplastids, remains to be seen.

Leishmania species, the pylogenetically distant kinetoplastid of trypanosomatids, do not undergo antigenic variation and base J is restricted to the telomeric repeat sequences [27]. They

exhibit only a limited genetic variability in comparison with other kinetoplastids and do not contain active non LTR retro elements in their genome [40], [41]. Therefore, telomeric recombination may not be a vital event for its survival. This may explain the apparent lack of J outside the telomeric repeats in *Leishmania*. In contrast, base J is abundant in telomeric localized repeat arrays and essential gene ORFs in *T. brucei* and, as we demonstrate here, in *T. cruzi*. We have recently generated a *T. brucei* bloodstream form cell line that is unable to synthesize base J. Consistent with the proposed role of J in regulating DNA recombination, preliminary analysis of the J-null trypanosome indicates a significant increase in telomeric exchange based VSG switching events and no transcriptional de-repression of the 19 silent telomeric expression sites (Sabatini et. al., unpublished results). Future studies will focus on the consequence of a lack of J on the subtelomeric localized contingency genes and repeat arrays of *T. cruzi*.

ACKNOWLEDGEMENT

We are grateful to Laura Cliffe, Rudo Kieft, Justin Widener, Kate Sweeney, Shanda Birkeland, Henri van Luenen, Paul-Andre Genest and Saara Vainio for critical reading of the manuscript. We would also like to thank Piet Borst, Steve Hajduk, Heather Goldstone and members of the Borst and Hajduk lab for their helpful advice and support. Anti-J antisera was a gift from Piet Borst. This work was funded by a grant from the NIH (grant A1063523) to R.S.

REFERENCES

1. WHO (2002) Control of Chagas disease. WHO Technical Report Series 905: 82-83.

2. Tyler K, Engman, DM . (2001) The life cycle of *Trypanosoma cruzi* revisited. Int J Parasitol 31: 3916-3923.

3. El-Sayed N, Myler P, Bartholomeu D, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309: 409-415.

4. Barry J, Ginger, ML., Burton, P., McCulloch, R. (2003) Why are parasite contingency genes often associated with telomeres? Int J of Parasitol 33: 29-45.

5. Voss T, Healer, J., Marty, AJ., Duffy, MF., Thompson, JK., Beeson, JG., Reeder, JC., Crabb, BS., Cowman, AF. (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. Nature 439: 1004-1008.

6. Borst P, Leeuwen Fv (1997) β--glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida Mol Biochem Parasitol 80: 1-8.

7. Leeuwen Fv, Taylor MC, Mondragon A, Moreau H, Gibson W, et al. (1998) b-D-Glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres Proc Natl Acad Sci U S A 95: 2366-2371.

8. Dooijes D CI, Kieft R, Dirks-Mulder A, Martin W, Borst P. (2000) Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. Nucleic Acids Res 28: 3017-3021.

9. Leeuwen Fv, Kieft R, Cross M, Borst P (1998) Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. Mol Cell Biol 18 5643-5651.

10. Leeuwen Fv, Kieft R, Cross M, Borst P (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. Mol Biochem Parasitol 109: 133-145.

11. Leeuwen Fv, Wijsman ER, Kieft R, Marel GAvd, Boom JHv, et al. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei* Gen Develop 11: 3232-3241.

12. Pereira-Chioccola V, Schenkman S (1999) Biological role of *Trypanosoma cruzi* trans-sialidase. Biochem Soc Trans 27: 516-518.

13. Buscaglia C, Campo V, Frasch A, Di Noia J (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. Nature Rev Microbiol 4: 229-236.

14. Kim D, Chiurillo, MA., El-Sayed, N,. Jones, K,. Santos, MR., Porcile, PE., Andersson, B., Myler, P., da Silveira, JF., Ramirez, JL. (2005) Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruz*i telomeres. Gene 346: 153-161.

15. Contreras V, Morel C, Goldenberg S (1985) Stage specific gene expression precedes morphological changes during *Trypanosoma cruzi* metacyclogenesis. Mol Biochem Parasitol 14: 83-96.

16. Contreras V, Salles J, Thomas N, Morel C, Goldenberg S (1985) In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions. Mol Biochem Parasitol 16: 315-327.

17. Finn R, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34: D247-D251.

18. Bendetsen T, Nielsen H, Heijne G, Brunak S (2004) Improved prediction of signal peptides: SingnalP 3.0. JMol Biol 340: 783-795.

19. Krogh A, Larsson B, Heijne Gv, Sonnhammer E (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genome. J Mol Biol 305: 567-580.

20. Quevilon E, Silventoinen V, Pillai S, Hark N, Mulder N, et al. (2005) Interproscan: Protein domains identifier. Nucleic Acids Res 33: W116-w120.

21. Freitas-Junior L, Porto R, Pirrit L, Schenkman S, Scherf A (1999) Identification of the telomere in *Trypanosoma cruzi* reveals highly heterogeneous telomere lengths in different parasite strains. Nucleic Acids Res 27: 2451-2456.

22. Wickstead B, Ersfeld, K., Gull K. (2003) Repetitive elements in genomes of parasitic protozoa. Micro Mole Biol Rev 45: 360-375.

23. Leeuwen Fv, Kieft R, Cross M, Borst P (1998) Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. Mol Cel Biol 18 5643-5651.

24. Aksoy S, Lalor TM, Martin J, Van der Ploeg LH, FF R (1987) Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosomes, *Trypanosoma gambience*. EMBO J 6: 3819-3826.

25. Villanueva MS, Williams SP, Beard CB, Richards FF, Aksoy S (1991) A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. Mol Cell Biol 11: 6139-6148.

26. Leeuwen Fv, Wijsman ER, Kuyl-Yeheskiely E, Marel GAvd, Boom JHv, et al. (1996) The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. Nuc Aci Res 24 2476-2482.

27. Genest P, Riet, B., Cijsouw, T., Luenen, HV., Borst P. (2007) Localization of the modified DNA base J in the genome of the protozoan parasite *Leishmania* Nucleic Acids Res Inpress.

28. Ruef B, Dawson, BD., Tewari, D., Fouts, DL., Manning, JE. (1994) Expression and evolution of members of the *Trypanosoma cruzi* trypomastigote surface antigen multigene family. Mol Biochem Parasitol 63: 109-120.

29. Bogliolo A, Lauria-Pires L, Gibson W (1996) Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. *Acta Tropica* 61: 31-40.

30. C D, R K, M C, R S (2005) Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. Mol Cell 17: 441-451.

31. Yu Z, Genest P-A, Riet Bt, Sweeney K, DiPaolo C, et al. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase Nucl Aci Res.

32. Briones M, Egima C, . SS (1995) *Trypanosoma cruzi* trans-sialidase gene lacking C-terminal repeats and expressed in epimastigote forms. Mol Biochem Parasitol 70: 9-17.

33. Chaves LB, Briones MRS, Schenkman S (1993) Trans-sialidase from *Trypanosoma cruzi* epimastigotes is expressed at the stationary phase and is different from the enzyme expressed in trypomastigotes. Mol Biochem Parasitol 61: 97-106.

34. Frasch ACC (2000) Functional Diversity in the Trans-sialidase and Mucin Families in *Trypanosoma cruzi*

. Parasitol Today 16: 282-286.

35. Bhattacharya S, Bakre, A., Bhattacharya, A. (2002) Mobile genetic elements in protozoan parasites. J Genetics 81: 73-86.

36. Kazazian HJ (2004) Mobile elements: drivers of genome evolution. Science 303: 1626-1632.

37. Mieczkowski P, Lemoine, FJ., Petes, TD. (2006) Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. DNA Repair 5: 1010-1020.

38. Lander E, Linton, LM., Birren, B., Nusbaum, C., Zody, MC., et al, (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

39. Schulz W (2006) L1 retrotransposons in human cancers. JBiomed Biotechnol 2006: 1-12.

40. Bringaud F, Ghedin, E., Blandin, G., Bartholomeu, DC., Caler, E., Levin, MJ., Baltz, T., El-Sayed, NM. (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. MolBiochem Parasitol 145: 158-170.

41. Kissinger J (2006) A tale of three genomes: the kinetoplastids have arrived. TrendsParasitol 22: 240-243.

Table 2.1: Distribution of base J in tandem repeats of *T.cruzi*

Sequence	Unit size	Copy numbers/array	% IP	± SE	n	Density[a]	Ref [b]
VIPER	2.3kb	70	10.49	0.12	3	0.065	[3-22]
DGF-1	10kb	~100	12.47	0.91	3	0.012	[3-22]
L1Tc	5kb	300-500	4.25	0.46	3	0.002	[3-22]
DIRE	260bp	150	10.56	0.99	3	0.270	[3-22]
Tubulin	3.6kb	10	0.003	0.004	3	NA	[3]
LS-rRNA	18kb	80	9.39	0.55	3	0.006	[3]
SS- rRNA	740bp	100	8.58	0.80	3	0.115	[3]
SL RNA coding	120bp	1~200	0.92	0.32	3	0.076	[24-25]
SL RNA nts	200bp-5kb*	ND	5.55	0.41	3	ND	[24-25]
Sat repeats	195bp	20,000	0.02	0.03	3	NA	[22]

Density = %IP / (unit size x copy number)

Reference for repeat organization

Based on southern hybridization in ref and analysis performed here

NA- Not applicable since %IP indicates the absence of base J in these regions

ND- Not determined/not

Table 2.2: Telomeric versus internal repeat scaffold

Repeat region	Telomeric		Internal	
	Scaffold	Distance[1]	Scaffold	Length[2]
DIRE	CH473486	10kb	CH473327	100kb
DGF-1	CH473498	10kb	CH473404	177kb
L1Tc	CH473485	50kb	CH473516	90kb

1. Represents the distance of the end of the repeat sequence to the telomeric repeat.
2. 2. Represents the distance of the end of the repeat sequence to the closest end of the scaffold. Thus
3. indicating the minimal distance of the repeat to the telomeric repeat.
4. 3. Based on http://www.tcruzidb.org

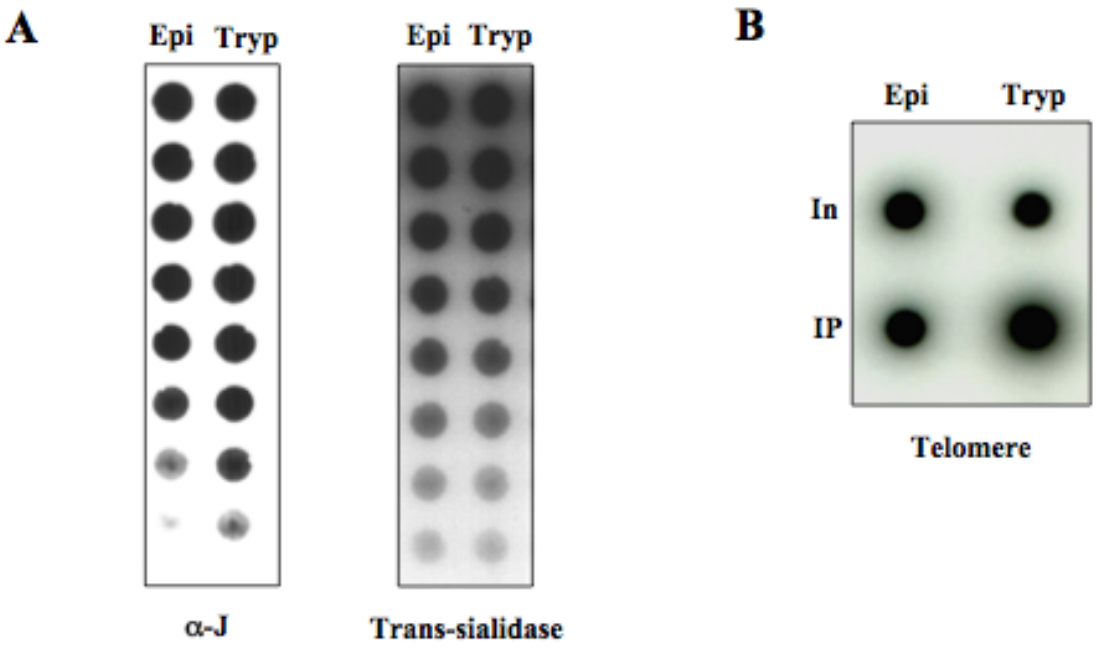
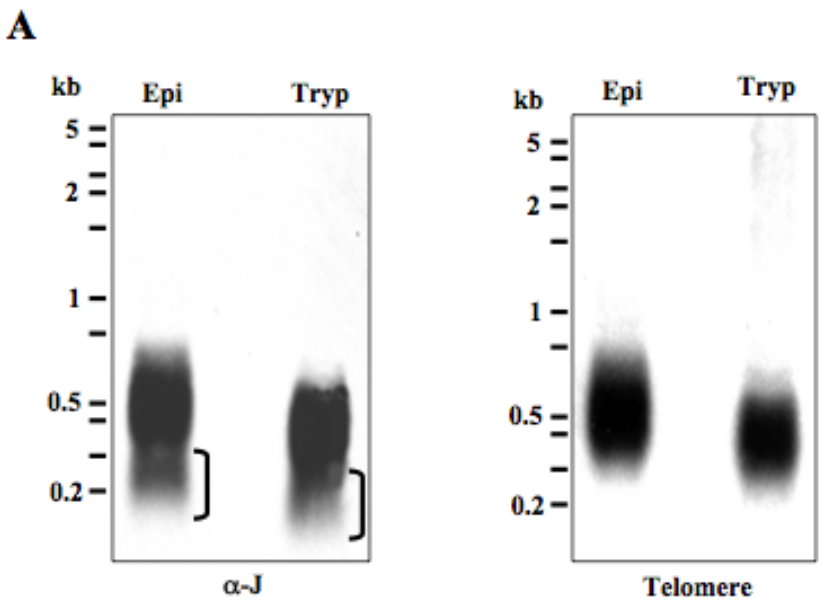


Figure 2.1: Developmental regulation of J-biosynthesis in *T. cruzi*. (**A**) A total of 200 ng of DNA from each life stage was serially diluted, spotted onto nitrocellulose and incubated with α-J antibodies. Signals were detected using enhanced chemiluminescence with a secondary antibody linked to HRP. Left lane represents the dilution series for the epimastigotes (Epi) and right lane represents the trypomastigotes (Tryp) life stage. The same blot was stripped and hybridized with a *T. cruzi*-specific trans-sialidase (TS) gene probe to check the DNA loading. (**B**) Epimastigotes and trypomastigotes DNA was sonicated and immunoprecipitated with α-J antibodies. Immunoprecipitated DNA (IP) along with 10% input DNA (In) was used in dot-blot assay and hybridized with an oligonucleotide probe of telomeric repeats.



Epi 23.8 ± 0.76

Tryp 26.12 ± 0.45

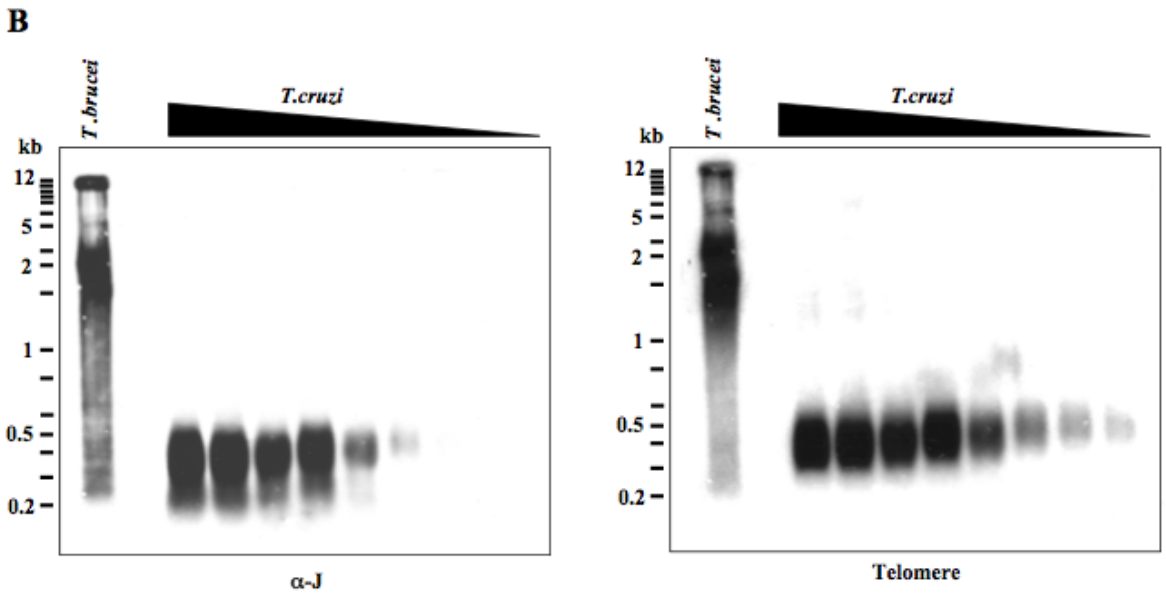


Figure 2.2: (**A**) **Quantitating the fraction of base J present outside the telomeric repeat regions of *T. cruzi***. Southern blot and J immunoblot of genomic DNA from Y strain *T. cruzi* epimastigotes and trypomastigotes digested with frequently cutting restriction enzymes. DNA was digested with a mix of terminal restriction enzymes (MspI, AluI and RsaI) to generate the 500 bp telomeric repeat fragment. This digested DNA was size fractionated by gel electrophoresis and blotted onto a nylon membrane. The blot was incubated with the α-J antibodies and signals were detected by ECL (J Blot). The stripped blot was then hybridized with radioactively labeled telomeric repeat oligonucleotide probe (Telomere). Brackets indicate DNA fragments that contain base J but do not correspond to telomeric sequence. Quantitation of the percent J outside the telomeric repeats from three separate experiments is provided below. (**B**) Determination of the sensitivity of the Southern blot. Two-fold serially diluted genomic DNA from T. cruzi digested with a mix of AluI, MspI and RsaI restriction enzymes (starting with 250 ng digested DNA) and 4 μg of DNA of T. brucei digested with mix of AluI, AvaII, HinfI, RsaI and SspI restriction enzymes was size fractionated in 0.5% agarose gel and blotted onto nylon membrane. The blot was incubated with the α-J antibodies and signals were detected by ECL. The stripped blot was then hybridized with radioactively labeled telomeric repeat probe.

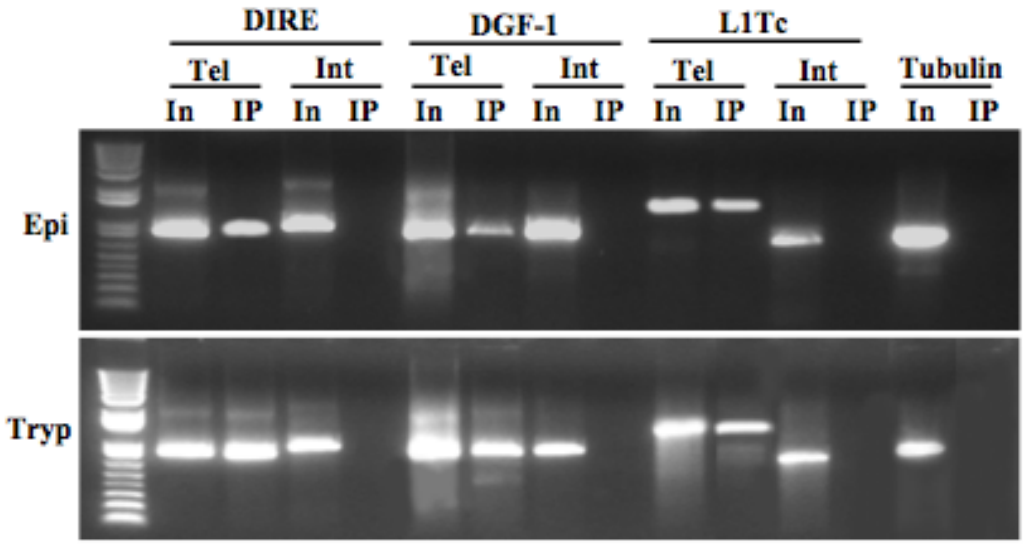


Figure 2.3: Telomeric-localized repetitive regions contain base J. IP and PCR analysis of the internal versus subtelomeric repeat arrays of T. cruzi. Genomic DNA from both life stages was digested using different restriction enzymes and immunoprecipitated using α-J antibodies. PCR was performed for immunoprecipitated DNA (IP) along with 10% of input DNA (In). Specific primers were designed to distinguish internal versus subtelomeric copy of the same repeat based on the T. cruzi database. The same digested DNA sample was used in the analysis of the internal and subtelomeric copy of a given repeat. The tubulin gene was used as a control.

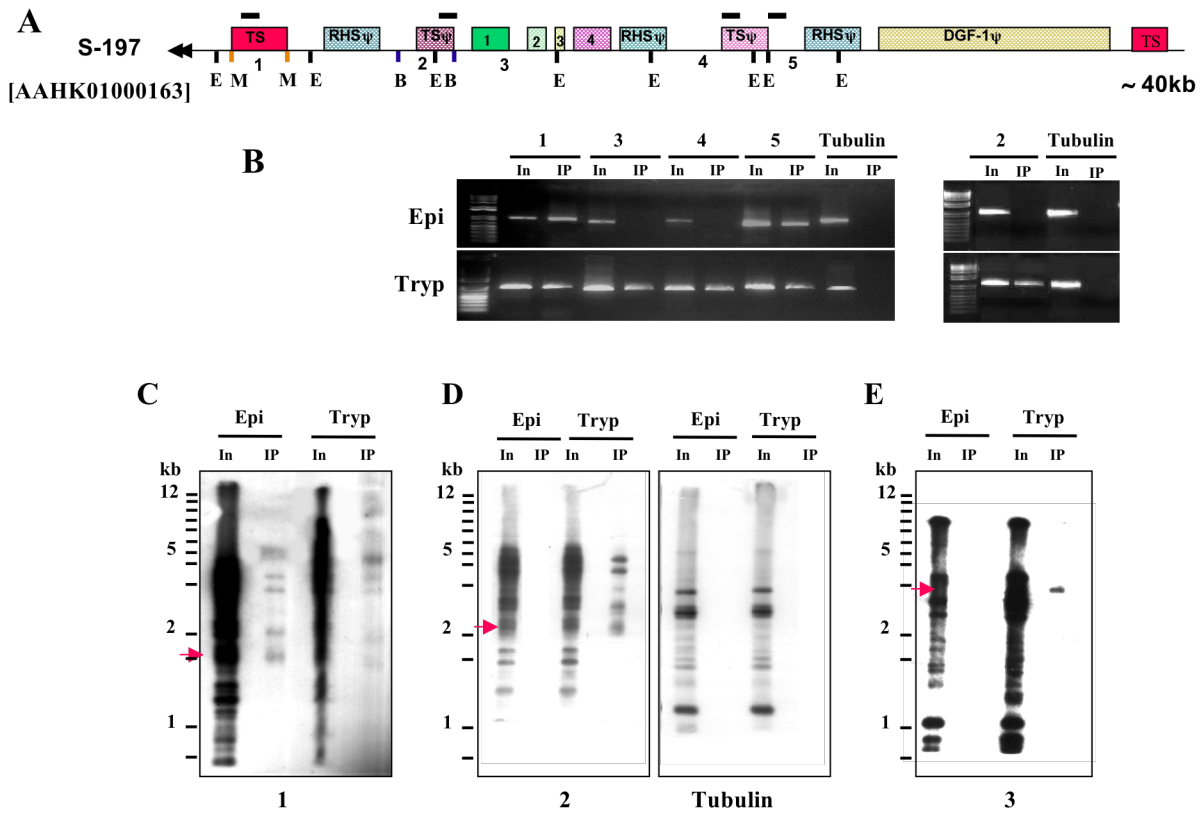


Figure 2.4: Localization of J in the subtelomeric region. (**A**) Schematic diagram and restriction map of the selected subtelomere using contig assembly of T. cruzi database. Vertical lines of the diagram of the subtelomere indicate the restriction site for the enzyme used for PCR and Southern hybridization; EcoRI (E), MscI (M) and BsrGI (B). The nomenclature used here for the subtelomeres and genes is according to the T. cruzi database and Genbank annotations. The Genbank scaffold/contig accession number is indicated in parenthesis. Arrowheads indicate the repetitive regions and horizontal lines represent the PCR amplified regions (also used as probes for Southern hybridization). Numbers represent the regions of interest. (**B**) PCR analysis. EcoRI-digested DNA was immunoprecipitated with α-J antibodies. Ten percent input and

immunoprecipitated DNA was used in PCR reactions along with specific primers to amplify the indicated regions. Tubulin gene primers were used as a control. Upper panel represents the IP/PCR reactions for DNA of epimastigotes and lower panel for trypomastigotes. IP/PCR reaction for fragment 2 was from BsrGI digestion and amplified using the same oligos used for the analysis of fragment 3 following EcoRI digestion. It is shown separately with the corresponding Tubulin control for both epimastigotes and trypomastigotes. (**C–E**). Southern blot analysis. DNA was digested using different restriction enzymes (C, MscI; D, BsrGI and E, EcoRI) to isolate fragments of the regions analyzed above. The digested DNA was immunoprecipitated, size fractionated in an agarose gel and blotted onto a nylon membrane along with 10% input DNA. PCR products, representing the indicated regions, were used as the probe for Southern hybridization. Left lanes demonstrate the hybridization signals for epimastigotes and right lanes for trypomastigotes. Small arrows are used to point out the expected restriction digested fragments. The tubulin control hybridization is shown for the BsrGI digest/IP reaction.

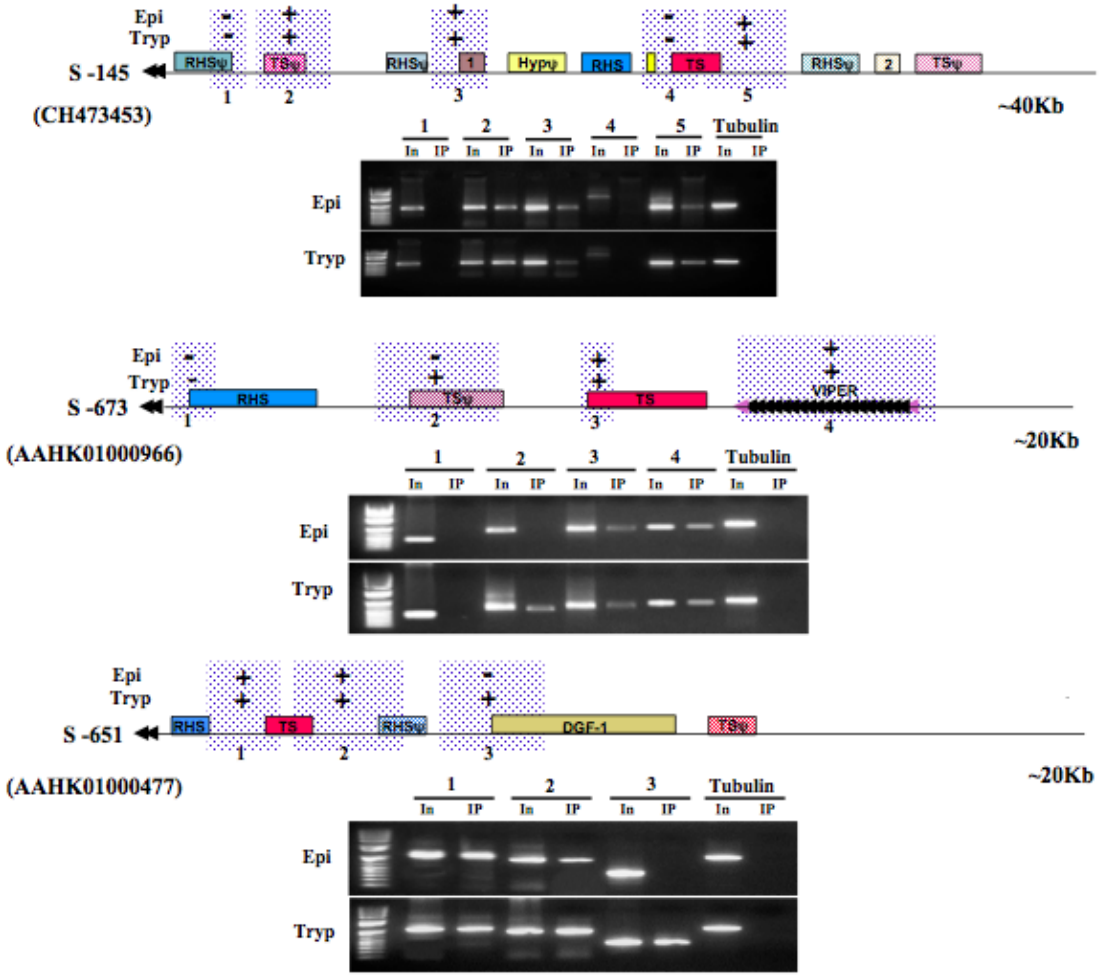
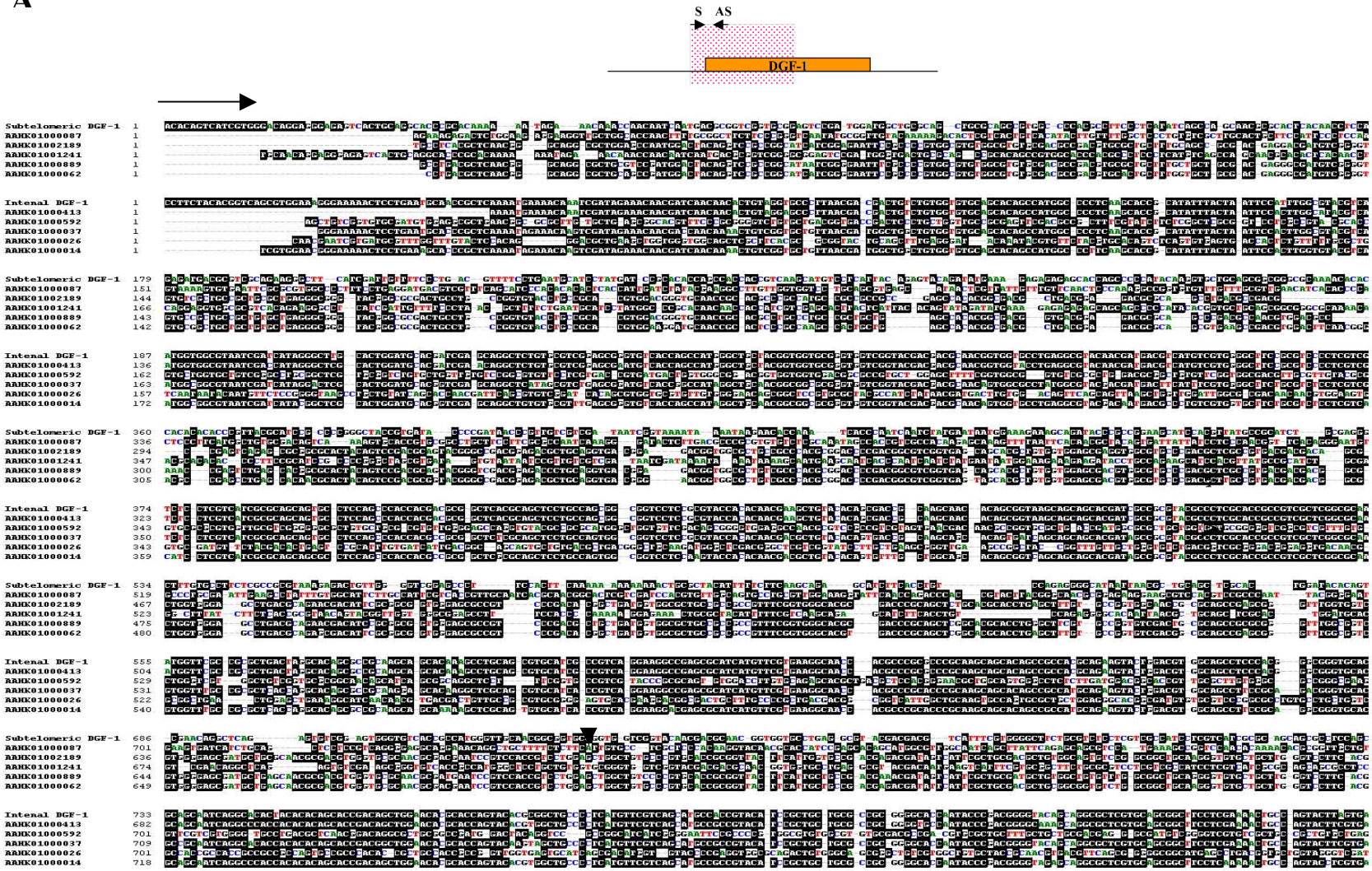


Figure 2.5: J localization within three additional subtelomeres. PCR data for epimastigotes and trypomastigotes is presented along with the schematic diagram of the representative subtelomeres. The nomenclature used here is according to the T. cruzi database and Genbank annotations and as described in Figure 2.4. Patch boxes show the restriction digested fragment analyzed and plus/minus signs indicate the presence and absence of base J for respective life stage.

A


S
AS
DGF-1
Subtelomeric DGF-1
Intenal DGF-1
RAHK01000087
RAHK01002189
RAHK01001241
RAHK01000889
RAHK01000062
RAHK01000413
RAHK01000592
RAHK01000037
RAHK01000026
RAHK01000014

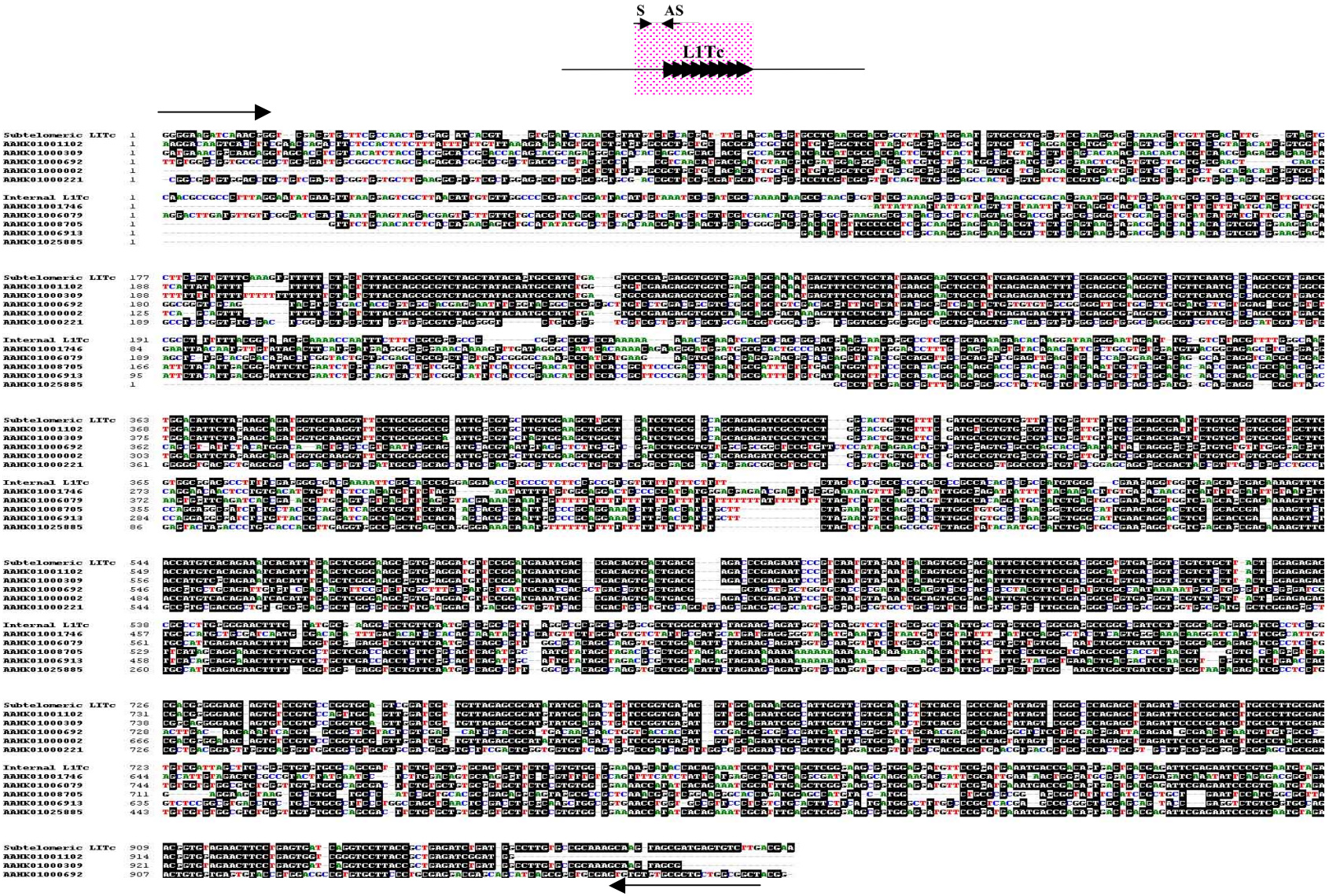

S
AS
LITc
Subtelomeric LITc
AAHK01001102
AAHK01000309
AAHK01000692
AAHK01000002
AAHK01000221
Internal LITc
AAHK01001746
AAHK01006079
AAHK01008705
AAHK01006913
AAHK01025885

Figure S2.1: Subtelomeric verses internal repeat sequences.

Schematic diagram of the subtelomeric copy of the DGF-1 (A) and L1Tc (B) along with the nucleotide alignments to show the specificity of the primers used for the differentiation of subtelomeric verses internal sequences. The patch boxs indicate the restriction fragments used for the analysis and the arrows show the location of the sense and anti-sense primers. 5' end is from the flanking region of the gene/repeat and 3' end is from the open reading frame/repeat region. Predicted PCR amplifying sequences from subtelomeric and internal regions were blasted against the *T.cruzi* genome. The fragments producing high similarity scores were selected and aligned using ClustalW multiple alignment tool. First five sequences are from fragments with high homology with subtelomeric DGF-1/L1Tc and the second five sequences are the fragments producing the highest homology with internal DGF-1/L1Tc sequences. Black shaded indicates the conserved sequences of all aligned fragments. Forty percent threshold score was selected for shading. Right arrows above the alignment represent the 5'oligo and left arrows for the 3' oligo for both subtelomeric and internal DGF-1/L1Tc. The arrowheads pointed to the start of open reading frames/repeat region.

CHAPTER 3

EPIGENETIC REGULATION OF TRANSCRIPTION AND VIRULENCE IN *TRYPANOSOMA CRUZI* BY O-LINKED THYMINE GLUCOSYLATION OF DNA

[1]**Ekanayake DK**, Minning T, Weatherly B, Gunasekera K, Nilsson D, Tarleton R, Ochsenreiter T, Sabatini R. Mol Cell Biol. 2011; 31(8): 1690-1700.

Reprinted here with permission of publisher

ABSTRACT

Unlike other eukaryotes, the protein coding genes of *Trypanosoma cruzi* are arranged into large polycistronic gene clusters transcribed by polymerase II (Pol II). Thus, it is thought that trypanosomes rely solely on post-transcriptional processes to regulate gene expression. Here, we show that the glucosylated thymine DNA base (Beta-D-glucosyl hydroxymethyluracil or base J) is present within sequences flanking the polycistronic units (PTU) in *T. cruzi.* The loss of base J at sites of transcription initiation, via deletion of the two enzymes that regulate J synthesis (JBP1 and JBP2), correlates with an increased rate of Pol II transcription and subsequent genome-wide increase in gene expression. Affected genes include virulence genes and the resulting parasites are defective in host cell invasion and egress. These studies indicate that base J represents an epigenetic factor regulating Pol II transcription initiation in kinetoplastids, and provides the first biological role of the only hyper-modified DNA base in eukaryotes.

INTRODUCTION

Trypanosoma cruzi, the protozoan parasite that causes Chagas' disease, is the major cause of cardiac disease in South and Central America (1). The parasite has a complex life cycle with two hosts and four developmental stages. Epimastigotes develop in the hindgut of the triatomine insect vector and differentiate into metacyclics. Infective metacyclics enter the vertebrate host, invade the host cell and differentiate to form amastigotes. Trypomastigotes released from the infected cell are able to reinvade a wide variety of host cells. Success of the parasite throughout the lifecycle is ensured by the regulated expression of surface proteins such as mucin and trans-sialidase, which allow differential adherence and evasion of the host immune responses (2). Members of the surface glycoprotein gene family co-localize with a novel hyper-modified DNA base, base J, suggesting an epigenetic mechanism of regulating *T. cruzi* pathogenesis (3).

Base J consists of O-linked glucosylation of the thymine base in telomeric DNA of all kinetoplastid flagellates and some closely related unicellular flagellates, but is not present in the genome of other protozoa or metazoa [1,2]. Base J was initially discovered based on its distinct presence within the 19 silent telomeric variant surface glycoprotein (VSG) expression sites (ES) of *T.brucei* but absence from the single transcribed ES, suggesting its role in the regulation of telomeric VSG gene expression [1,3] Recent genome-wide analysis revealed that base J is also present throughout the *T. brucei* genome, enriched at regions flanking polymerase II polycistronic transcription units (PTU) [4] . PTU are large gene clusters that are co-transcribed by Pol II to yield polycistronic pre-mRNAs that are then processed into mature mRNAs by trans-splicing and polyadenylation [5]. The localization of base J at PTU flanking regions suggests a

role for the modified base in regulating Pol II transcription initiation and termination and thus, trypanosome gene expression. However, the regulation of trypanosome gene expression is thought to occur primarily via differential mRNA decay or other post-transcriptional mechanisms (for reviews see [5,6]. To date, there is no evidence for regulation of gene expression at the level of Pol II transcription in kinetoplastids. Currently, very little is understood about the DNA sequences and proteins involved in transcription initiation and termination. With the exception of the spliced-leader RNA promoter, Pol II promoters and associated factors have not been identified. Thus, how Pol II transcription is initiated/regulated remains an unanswered question. While the localization of base J in the genome of *T. brucei* has suggested its involvement in regulating antigenic variation and Pol II transcription, no direct evidence has been provided.

Base J is synthesized in a two step pathway, where a thymidine hydroxylase enzyme converts a thymidine residue in DNA to hydroxymethyldeoxyuridine (HOMedU), and then a glucosyltransferase converts HOMedU into base J by addition of glucose [1,7]. Previous work has identified two distinct thymidine hydroxylases involved in J biosynthesis, JBP1 and JBP2 [8,9,10]. Deletion of JBP1 and JBP2 in *T. brucei* resulted in 20- and 8-fold reduction in J levels, respectively [8,11,12]. Deletion of both enzymes resulted in a trypanosome cell line (J null) unable to synthesize base J [8]. Analysis of JBP function in *T. brucei* has indicated that JBP1 and JBP2 are each able to stimulate de-novo J synthesis but have different chromatin/substrate specificities [4]. JBP1 preferentially stimulates de-novo J synthesis at PTU flanking regions, while JBP2

prefers telomeric substrates. However, optimal maintenance of J at the PTU flanks by JBP1 requires JBP2 function.

To elucidate the function of base J in *T. cruzi*, we manipulated J levels through deletion of JBP1 or JBP2. We show that similar to *T. brucei*, base J is localized at sequences involved in Pol II transcription initiation and termination in *T. cruzi* and its synthesis is regulated by JBP1 and JBP2. The analysis of the JBP mutant cell lines demonstrates that base J regulates the rate of Pol II transcription, gene expression and virulence of *T. cruzi*. These results provide a clear biological function of the only hyper-modified DNA base identified in eukaryotes, and demonstrate its role in regulating transcription in an important pathogenic organism where this process is not well understood. We also describe, for the first time, epigenetic regulation of virulence through transcriptional control in *T. cruzi*.

MATERIALS AND METHODS

Growth of cells. Y strain *T. cruzi* epimastigotes were grown in liver infusion Tryptose media as previously described [13]. Cultures were harvested at mid log phase by centrifugation. Metacyclics were obtained from epimastigotes by axenic induction as described previously [14]. Briefly, epimastigotes were transferred to Grace's insect media supplemented with 10% fetal bovine serum and grown for 10-14 days at 37°C. Differentiated metacyclic stages were purified using TAAUAG media and was confirmed by their morphology. These metacyclics were then used to infect irradiated vero and myoblast cells as previously described [15]. Trypomastigotes were then grown in monolayers of Vero cells/myoblasts in Modified Eagle's Media containing 20% fetal

bovine serum. Emergent trypomastigotes were harvested daily and examined by light microscopy. Preparations containing >95% of trypomastigotes were used in the subsequent assays.

Targeted deletion of JBP1 and JBP2. In order to generate *T. cruzi* JBP1 and JBP2 double KO cell lines, we adopted the same technique used to generate the *T. brucei* JBP KO cell lines [8,16]. Briefly, 5′and 3′UTR sequences of JBP1 and JBP2 were inserted flanking the drug markers of the deletion plasmid constructs PTREX-Neo, PTREX-hygro and pTub-Puro and pTub-Blast [17]. PCR was used to generate 5' and 3' non-coding regions (250-300bp) of JBP1 and JBP2. 20-50 μg of KpnI and XbaI digested plasmid DNA was used for each transfection and cells were selected under drug pressure for 5-10 weeks [18]. We used 100 ug/ml neomycin (G418), 100 ug/ml of hygromycin, 5 ug/ml of puromycin and 10 ug/ml of blastocidin for selection. This allows the targeted deletion of the JBP1 and JBP2 ORFs and insertion of the drug markers via homologous recombination. Drug resistant clones were then selected after limited dilution. Clones were screened by PCR for the absence of ORFs and used in subsequent analysis.

Microarray analysis. Total RNA isolation, cDNA synthesis, labeling and array analysis was performed as previously described [19]. Briefly, total RNA was isolated using TRIZOL and cleaned up with RNAeasy kit (QIAGEN, Valencia, CA) per the manufacture's instruction. cDNA was synthesized using superscript II reverse transcriptase (Life technologies, Grand Island, NY) and Cy3 or Cy5-dUTP (Amersham Phamacia Biotech) included in the reaction mix. Unincorporated fluorescence nucleotide

was removed using GFX columns per manufacture's instructions. Microarrays were obtained from the Pathogen Functional Genomic resource center, and contained an array of 12,288 70mer oligonuclotides, each with complementarity to an individual *T. cruzi* open reading frame (representing a total of 7,865 genes on the array). Microarray was prehybridized immediately before hybridization. Three hybridizations were performed for each life stages. Arrays were scanned on a ScanArray 4000 (GSI Lumonics, Wilmington, MA) to obtain tagged Image File format (TIFF) images. Microarray scans were quantified using The Institute of Geonomic Research (TIGR) SpotFinder module of TM4 http://www.tm4.org. Quantification was carried out as previously described [19]. Further, files generated from SpotFinder were imported to TIGR Microarray Data analysis system (MIDAS) module of TM4 http://www.tm.org and further analyzed using TIGR MultiExperimental Viewer (MeV) software as previously described. We analyzed the results files using Multiclass SAM for triplicates and hierarchical cluster analysis. Since the JBPKO cell lines were derived from Y strain *T. cruzi*, we utilized the Esmeraldo allele of the annotated CL Brenner genome to identify the localization of affected genes. Data was submitted to NCBI and can be viewed at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=xpctnuqceqwwgje&acc=GSE23855 .

Spliced Leader Trapping. RNA extraction, library construction and bioinformatic analysis was performed essentially as previously described **[20]**. Briefly, first strand cDNA was synthesized from poly(A) RNA using random hexamers and Superscript II reverse transcriptase. In order to process multiple samples in one sequencing channel a

four-nucleotide barcode was added to the 3'end of the cDNAs. After amplification and size fractionation to 120-160 base pairs the cDNA library was sequenced on a Genome Analyzer (GA-II, Illumina) using the Chrysalis 36 cycles v3.0 sequencing kit and 76 cycles. Base calling was performed using the Genome Analyzer Pipeline and linker sequences were removed while separating reads according to identified barcodes. Only sequence reads of inserts with a length of at least 24 nucleotides were retained. The reads were mapped to the genome sequence of *T. cruzi* CL Brenner strain using maq (http://maq.sourceforge.net) with n=3 and an effective first read length of 24. Single mapping reads were separated from multi mapping reads by an alignment quality threshold of 30. Tag counts were normalized to the library size (number of reads of length 24 or more) and scaled linearly to reflect counts of tags per million (TPM). Mapped tags were assigned to the annotated protein coding gene 3' of the tag. Data was exported in tabular and GFF format (http://www.sanger.ac.uk/Software/formats/GFF/) and then visualized using Gbrowse. Differences in expression levels of a gene between WT and JBP1KO was tested for significance according to Audic and Claverie with a threshold of $P<10^{-5}$. Antisense splice sites were detected using bioperl. R and ad hoc perl scripts were used for statistical analysis and plotting. All bioinformatic tools, programs, pipelines used in this study will be provided upon request.

Quantitative RT-PCR. Total RNA was obtained from three biological replicates of each life stage as described above. First strand cDNA was synthesized from 1µg of total RNA using ISCRIPT cDNA synthesis kit (Bio-Rad Laboratories, Hercules, CA) per manufacture's instruction. Heat inactivated cDNA reactions were finally treated with

RNaseH at 37°C for 45min. Quantification of selected up and down regulated genes were performed on an iCycler with an iQ5 Multicolor Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA). Primer sequences used in the analysis is available upon request. Reaction mix contains 5 pmole forward and reverse primer, 2x iQ SYBR Green Super Mix (Bio-Rad Laboratories, Hercules, CA) and 2μl of template cDNA. Standard curves were prepared for each gene using ten folds dilutions of known quantity (15ng/μl) of WT DNA. Quantifications were calculated using iQ5 optical detection system software. Each sample was normalized to the average of three unaffected genes including 24S rRNA. The average ratios were calculated for each gene using 3 stage and 3 reference sample comparisons.

Transcription assays. *In vivo* transcription assay was performed as previously described [21]. Briefly, epimastigotes and fresly released trypomastigotes were pre incubated for 1h with media (LIT or MEM) containing 10%FBS previously dialysed against 0.25N NaCl. $5x10^6$ of parasites were incubated with 4μCi of (5,6-^{3}H) labeled uridine (37ci/mmol, Perkin Elmer, Waltham, MA) containing media for 20hr. At least 5 samples were taken during different time points. Following labeling, the cells were isolated by centrifugation and lysed in 0.01% sodium dodecyl sulfate. The lysate was vigorously mixed with 4.5 ml of 10% ice cold TCA solution and left on ice for 30-60 mm. The acid-precipitable material was recovered on Whatman GF/C glass fiber filters and washed 3 times with *5 %* TCA, followed by one wash with 95 % ethanol. The dried filter pads were counted in a toluene-based scintillation fluid in a Beckman liquid scintillation counter.

In vitro transcription was performed using lysolesithin permeabilized cells as previously described [22]. Briefly, $2x10^8$ log phase epimastigotes were washed twice with transcription buffer (20mM potasium glutamate, 3mM $MgCl_2$, 150mM sucrose, 10μg /ml leupeptin and 1mM dithiothreitol), and chilled on ice for 5min. palmitoyl-L-α-lysophospotiduylcholine was added at 500 μ/ml final concentration to permeabelized the cells. After 1min, parasites were centrifuged and resuspended in 200μl of transcription buffer containing 2mM dATP, 1mM dCTP, 1mM dGTP, 6mg/ml creatine kinase, 25mM creatine phosphate and 100μCi of (α^{32}P)-UTP (300Ci/mmol). After 30min at 30^0C parasites were lysed and RNA was extracted using TRIZOL. Labeled RNA was then hybridized to slot blot containing 5-10 μg of plasmid DNA in hybridization buffer (5X saline/sodium phosphate/EDTA, 5X Denhardts, 0.1mg/ml tRNA and 0.1%SDS). Slot blots were prepared, as previously described [22] using denatured plasmid DNA containing probes for multigene family members/ multi-copy genes and other single-copy genes and regions. Primer sequence used to generate these probes can be provided upon request. Dot blots to detect transcription of sense and antisense strands within a SSR were prepared by spotting 2ug of 50 mers as previously described [23]. Oligonucleotides used in this run-on are available upon request. Hybridization was carried out at 65ºC for 48hr and washed as a regular northern blot. Data was quantified using phosphoimager.

Mammalian cell invasion assay. Mammalian invasion assays were performed using previously described methods [24]. Vero cells were plated on 13-mm round coverslips at a density of 3×10^4 cells in MEM medium with 10% FCS and cultivated in 24-well plates for 24 h at 37°C in a 5% CO_2 atmosphere. Coverslips with attached cells were then

washed three times with PBS to remove the unattached cells and debris. Freshly released trypomastigotes from cultures of infected Vero cells and purified metacyclics were used for invasion assays. Parasites were centrifuged to remove the cell debris and seeded onto the Vero cells in equal number (5×10^6 parasites/well). After 3h at 37°C, the interaction was stopped by removing parasites and washing cells three times with PBS. Monolayers were fixed and stained with amastigote-specific antibodies and/or Giemsa. Invasion was quantified by counting the number of intracellular parasites in a total of 100 cells per coverslip. Values represent means ± SD of at least three independent experiments.

Anti J dot-blot and immuno-precipitation. Anti J immunoblots were used to determine the global genomic levels of base J as previously described [25]. Serially diluted DNA from WT and JBPdKOs were blotted onto nitrocellulose. Membranes were then incubated with anti J antibodies, detected with HRP conjugated goat anti-rabbit antibodies and visualized with ECL. DNA loading was determined by hybridization using a ^{32}P-labeled tubulin probe. J-DNA containing fragments were immunoprecipitated and quantitated by dot-blot analysis as previously described [25].The hybridization signals were quantitated by phosphorimager analysis.

mRNA degradation experiments. Total RNA was obtained from WT and JBP1dKO samples of $5x10^6$ cells taken at different times after the addition of 100μg/m1 of Actinomycin D. Total RNA was extracted from each sample and qRT-PCR was performed using the method described above. mRNA half lives were estimated after normalizing to 24S rRNA values and plotted on semi-logarithmic scales. Half-life of the

mRNA is calculated as the amount of time required for a transcript to decrease 50% of initial amount. Half-life values were compared using Student's t-test and a P value <0.001 was considered statistically significant.

RESULTS

JBP1 and JBP2 knockout cell lines of *T. cruzi* contain reduced levels of base J.

In order to characterize the role of base J in *T. cruzi*, we generated thymidine hydroxylase mutant cell lines by targeted removal of JBP1 and JBP2 from the genome (Figure 3.1). After transfection with the appropriate JBP KO construct (Figure 3.1A and B), drug resistant parasites were screened by PCR and southern analysis to confirm proper integration (data not shown) and the loss of JBP1 and JBP2 gene sequence (data not shown and Figure 3.1C and D). The corresponding loss of JBP mRNA was confirmed by northern analysis (data not shown) and microarray analysis (see below). Similar to the effect of JBP loss of function in *T. brucei*, JBP2 and JBP1 dKO *T. cruzi* cell lines have 8- and 20-fold reduction in J levels, respectively (Figure 3.1E).

Earlier studies have shown base J to be distributed in telomeric and sub-telomeric regions of *T. cruzi* [25]. This includes the telomeric repeat arrays and sub-telomeric localized repeat elements (VIPER), retrotransposons (L1Tc) and genes encoding surface proteins. Analysis of the JBP mutant cell lines by anti-J IP and dot-blot hybridization/PCR indicates that the amount of base J is either reduced or absent in these telomeric and sub-telomeric regions (Figure S3.1A-C). This is particularly apparent in JBP1dKO cell lines, where the levels of base J are reduced by 40% in telomeric repeats, by 80% within sub-telomeric localized VIPER and LITc elements and to undetectable

levels in other sub-telomeric regions. Overall, while there is an apparent global reduction in base J levels in the mutant *T. cruzi* cell lines, the reduction in specific regions differs due to the loss of JBP1 or JBP2. These results are consistent with previous analysis of JBP function in *T. brucei* [4].

JBP1 and JBP2 regulate J synthesis in Polymerase II polycistronic transcription unit flanking regions. We have recently shown that base J is localized within regions flanking polymerase II polycistronic transcription units (PTU) of *T. brucei* and *Leishmania* [4]. These regions include convergent and divergent strand-switch regions (SSR), which demarcate sequences involved in Pol II transcription initiation and termination [23,26,27]. In order to investigate the internal localization of base J in *T. cruzi*, we analyzed 13 randomly chosen SSRs by anti-J ChIP/PCR. Base J was found in all convergent and divergent SSRs analyzed in WT cells (Figure 3.2). Deletion of JBP1 resulted in the loss of J from all of the SSRs, while several of these regions retain significant levels of J upon the deletion of JBP2 (Figure 3.2). This result is consistent with our previous observations in *T. brucei* that JBP1 represents the key regulator of J synthesis at PTU flanking regions [4]. In *T. cruzi* the requirement for JBP2 is particularly evident at divergent SSRs. Overall, the results indicate that similar to other trypanosomatids, base J is localized at Pol II transcription start sites (TSS) and termination sites (TTS) in the genome of *T. cruzi*. Furthermore, the synthesis of base J within these regions is regulated by JBP1 and JBP2.

Microarray analysis of *T. cruzi* JBP1 and JBP2 knockouts. To explore the role of base J in the regulation of transcription, microarray analysis of WT and JBP KO cell lines corresponding to epimastigote, metacyclic and trypomastigote life-stages was performed. Duplicate analysis of JBP1dKO epimastigotes showed an average of 160 genes significantly (>2-fold) up regulated while 110 genes showed significant down regulation compared to WT (Figure 3.3A). We then extended the study to include triplicate analysis of trypomastigotes and metacyclic life-stages upon differentiation of the JBP1 and JBP2dKO. Significant analysis of microarray (SAM) of the three biological replicates from trypomastigotes determined that 658 transcripts (representing 8% of the genes on the array) exhibited statistically significant up or down regulation in the JBP1dKO with a median false discovery rate (FDR) of 0.7% while 417 transcripts (5.8% of the genes on the array) were up or down regulated in JBP2 dKO to statistically significant levels (FDR of 0.3%). 482 and 341 of these genes were up regulated in JBP1dKO and JBP2dKO, respectively (Figure 3.3A). The changes and consistency between biological replicates are apparent in Figure S3.2A. The genome-wide distribution of regulated genes is apparent in Table S3.1. SAM analysis of the metacyclic data showed a high false discovery rate due to the significant variation of number of affected genes between biological replicates (Figure S3.2B). This may be due to defects in the efficiency of metacyclogenesis for the JBP KO cell lines (data not shown).

Changes in gene expression of several up and down regulated genes indicated by the array analysis were confirmed by qRT-PCR (Figure 3.3B and Figure S3.2C). For 10 out of 10 genes showing significant up or down regulation by array analysis in trypomastigotes, qRT-PCR data agreed in the direction of regulation (Figure 3.3B).

Similar agreement was seen in the case of epimastigotes (9 out of 10 genes) (Figure S3.2C). While the extent of up or down regulation estimated by qRT-PCR and microarray analysis may differ slightly, the direction of regulation is consistent. These differences were likely a result of the differing dynamic range of the two platforms [19] as well as the fact that the RNA samples used for the array and qRT-PCR were not identical. Analysis of the same samples by both techniques gave more closely matching results (Figure S3.2D). In conclusion, the decrease (JBP2 KO) or loss (JBP1 KO) of base J at Pol II TSS and TTS in *T. cruzi,* led to significant changes in Pol II gene expression of at least 200-400 genes in all three life-stages analyzed.

Previous studies have demonstrated the ability of a nucleoside analog of thymidine, BrdU, to reduce the levels of base J in kinetoplastids by an unknown mechanism (van Leeuwen 1998 and Vanio 2008). We show that growth in 100ug/ml BrdU results in a ~16-fold reduction in J levels in *T. cruzi* (Figure S3.1D). The levels of J in WT cells fed BrdU approximates the levels in the JBP1 KO and results in similar changes in gene expression (Figure S3.2C). Altered regulation of gene expression in cells with reduced J in a WT JBP background supports the conclusion that the changes in gene expression observed in the JBP KO cell lines are due solely to defects in J synthesis.

RNA polymerase activity increases in the JBPdKO cell lines. Since base J represents a covalent modification of DNA, we proposed that any changes in gene expression in the JBP KO cell lines would be due to effects of the modified base on transcription rather than post-transcription. To address this, we analyzed the mRNA turnover rate of genes that were up and down regulated in the JBP1dKO epimastigotes and trypomastigotes. We

treated cells with 100mg/ml actinomycin D to inhibit transcription and calculated the half-life of specific transcripts. The mRNA degradation kinetics of all four up regulated genes analyzed was not significantly affected between the WT and JBP1dKO (Figure S3.3A-B and E-F). Therefore, the increased abundance of the transcripts following the reduction in base J levels in the genome is not due to any increase in mRNA stability. In contrast, all four of the down regulated transcripts in the JBP1KO epimastigotes and trypomastigotes showed an increased mRNA turnover rate compared to WT (Figure S3.3C-D and G-H).

In order to directly address the role of base J in regulating transcription, we compared (^{3}H)-uridine incorporation by WT and JBP1dKO cell lines. As shown in Figure S3.4 (A and B), ~3-fold more (^{3}H)-uridine was incorporated into RNA by epimastigote and trypomastigote JBP1dKO cell lines when compared to WT cells. As this difference in labeling efficiency could be due to other possible metabolic defects, we chased the (^{3}H)-uridine incorporation by adding actinomycin D. As shown in Figure S3.4C, RNAs decay similarly in JBP1dKO and WT cell lines, showing that the difference in (^{3}H)-uridine incorporation was in fact due to the higher transcription rate in the JBP1dKOs. In fact, the slightly increased decay rate of RNA in the JBP1 dKO would suggest that these experiments under-represent the overall increased rate of transcription.

Nuclear run-on assay quantifies the elongations of nascent mRNA chains already initiated *in vivo* and thus provides a direct measure of the transcription rate of a particular gene or PTU. To do this, we extracted RNA labeled by *in vitro* incorporation of ^{32}P-UTP in lysolecithin permeable epimastigotes of both WT and JBP1dKO, and used them to probe membranes containing representative genes that are transcribed by Pol I, II and III.

As shown in Figure 3.4, JBP1dKO epimastigotes show a significantly higher labeling for a majority of the genes analyzed when compare to WT. However, a few genes (i.e. 24S) did not change and act as a loading control. These results were consistently obtained in at least four independent experiments. mRNA encoding Pol II transcribed multi-gene family/copy genes (TS, Gp85, amastin, tubulin and SL) were transcribed on average ~3.6 fold more in JBP1dKO than the WT (Figure 3.4A and Table 3.1). We also see increased transcription of single-copy genes, including enolase and three genes in the divergent PTU discussed below. The hypothetical gene (Hypo) represents a Pol II transcribed single-copy gene whose steady-state mRNA levels decreased in the JBP1 KO. Moreover, we saw 5-fold higher labeling for Pol III transcribed 5S genes and 3-fold higher labeling for Pol I transcribed SSU-rRNA gene copies that are, in many cases, localized within a SSR. α-amanitin titration confirmed that these signals are coming from the appropriate polymerase (data not shown). To evaluate the role of base J in regulating Pol II transcription at the PTU level, we investigated the rate of transcription within a divergent PTU (SSR 7739) using probes against the first three single-copy genes in the unit (Figure 3.4A and B). While the steady-state transcript level of one gene increased and another decreased, the transcription rate increased for all three genes in the PTU following the loss of base J.

To verify the strand bias of transcription at TSS and TTS, nuclear run-on analysis was performed using strand-specific oligomers spanning SSRs at divergent and convergent PTUs (Figure 3.5A and B). In all three divergent regions analyzed, transcription of both strands initiated bidirectionally at similar sites within the SSR in WT and JBP1 KO (Figure 3.5A). Additionally, the increased rate of strand-specific Pol II

transcription in the JBP1 KO is evident in all regions examined, especially at regions closest to the 'promoter' (Figure 3.5A and S3.5). Similar analyses of three convergent SSR indicate that the loss of J has no affect on Pol II transcription termination (Figure 3.5B). Interestingly, the increased transcription of sense strand is evident throughout the entire nascent transcript (Figure 3.5B and S3.5). In conclusion, the increased rate of Pol II transcription throughout the PTU correlates with the loss of an epigenetic mark at TSS and the loss of J at TTS does not lead to detectable defects in transcription termination.

Genome-wide transcriptome analysis. To further examine for defects in transcription termination and confirm changes in gene expression, we performed a genome wide transcriptome analysis using spliced leader trapping (SLT; [20]). This analysis produced more than 3 million sequence tags in wild type and JBP1 KO epimastigote cell lines. The vast majority (≥93%) of these tags mapped to the sense strand in both wild type and JBP1 KO cells respectively (Table 3.2). Overall, this analysis indicates a genome-wide increase in Pol II transcript abundance in the epimastigote JBP1 KO (Figure 3.6). If the loss of J at TTS results in defects in termination, we would expect a decrease in the transcript abundance of the genes proximal to the convergent strand switch region due to an increase in anti-sense RNA and/or colliding Pol II. However, the number of antisense tags ranged from 0.7% to 0.5% in the wild type and JBP1 KO, respectively. Since tags in the zones between individual transcription units (including SSRs) are often times difficult to assign to the + or – strand, we determined the number of those separately. In the wild type 6.2% of tags located in between PTU while 5.6% where assigned to these regions in the JBP1 KO cell line. The localization of tags in SSR is consistent with the TTSs

mapped via run-on analysis (Figure 3.5 and S3.5 and data not shown). When we analyzed the transcript abundance of the 5 genes proximal to the TSS or proximal to the TTS for all transcription units with 10 or more genes, we could show a significant difference in how the knockout was affecting the transcript levels in these different groups (p=0.02, Fisher's exact test). 807 genes at the start of PTUs showed increased transcript abundance in the JBP1KO/WT while 481 decreased, whereas 602 of the genes proximal to the TTS were increased and 439 decreased in transcript abundance. Similarly, when we analyzed the transcript abundance of genes proximal to 157 converging strand switch regions, we found no significant change in transcript abundance/ratios in the JBP1KO/WT (mean of JBP1KO/WT ratio 1.3, p=0.18; Table S3.2 and Figure S3.5). Thus, there is statistically significant increased transcript abundance at the start of TUs and no clustering of down regulated genes localizing to the ends of convergent PTUs or increased antisense RNA produced in the JBP1 KO. In conclusion, genome wide analysis of transcript abundance supports the overall finding that the loss of base J within SSRs leads to increased Pol II transcription of PTUs and no defect in transcription termination.

Base J regulates the virulence of *T. cruzi*. Transcriptome analysis revealed a significant change in the expression of genes associated with virulence and cell homeostasis in the JBPdKO cell lines. For example, several genes encoding surface proteins are up regulated in the infective life-stages of the JBPdKO cell lines compared to WT (Table S3.3). Immunoflorescence analysis of the JBP1dKO confirm that trypomastigote mutant cells co-express trypomastigote and amastigote specific surface proteins (Figure S3.6A and B). One would predict that alterations in the expression of surface proteins would result in

significant changes in phenotype (i.e. the ability of the parasite to interact and invade the host cell). To address this, an *in vitro* invasion assay using both infective life-stages of the parasite and mammalian cell cultures was performed. A significant (3-5 fold) increase in the efficiency of invasion of metacyclic and trypomastigote parasites was detected corresponding to both JBP1 and JBP2 KO cell lines (Figure 3.7A). In fact, there was a concomitant increase in invasion efficiency upon the deletion of each JBP allele and subsequent decrease in J levels (Figure S3.6C and Figure 3.1E). Moreover, JBP mutant cells showed a significant delay in egress. WT *T. cruzi* parasites emerge from the mammalian cell within 4-7 days post infection (Figure 3.6B). However, the JBP1dKO and JBP2dKO parasites remain intracellular for 4-5 weeks. Intracellular amastigote growth rates were similar between WT and JBP1KO cell lines during initial 7 days of culture (Figure S3.6D). Presumably, a delay in the transition of amastigote to trypomatigote of the JBP mutants within the mammalian cell leads to a delay in egress. In conclusion, our data strongly supports the hypothesis that epigenetic regulation of gene expression by base J regulates *T. cruzi* virulence.

DISCUSSION

The recent identification of the novel glucosylated thymine DNA base at Pol II TSS and TTS in *T. brucei* and *Leishmania* suggested a role of base J in transcription initiation and termination [4]. Here, we demonstrate that base J is also present at TSS and TTS in the genome of *T. cruzi,* and reduced or loss of base J at these sites results in an increased rate of Pol II transcription and genome-wide changes in gene expression. These results identify a role of J in regulating initiation and/or elongation of Pol II transcription. However, the specific localization of base J at TSS, verses within the PTU, indicates that increased Pol II activity in the JBP KO cell lines is primarily due to increased initiation events. These data provide the first example of epigenetic regulated Pol II transcription in organisms where gene expression is thought to mainly occur at the level of RNA stability, translation and protein stability. Furthermore, these studies provide a biological function for the only hyper-modified DNA base identified in eukaryotes.

Epigenetic regulation of transcription has been well studied in mammalian and bacterial systems where methylation of DNA is necessary for heritable gene silencing and regulated gene expression [28,29]. In these organisms, the presence of the methylated DNA base at promoter regions or regulatory sequences hinders the binding of Pol II and other transcriptional regulators [30]. Alternatively, methylated DNA recruits methylated CpG binding domain (MBD) proteins that interact with histone deacetylase thus altering chromatin structure leading to gene silencing [31,32]. We believe that base J regulates transcription in a similar fashion, through regulated binding of Pol II and other regulatory proteins via physical hindrance of the glucose moiety present in the major groove of J-DNA and/or modulating chromatin structure. As most transcriptional factors bind the

major groove of DNA, the idea that J can induce steric hindrance is highly plausable. The ability of base J to recruit specific proteins leading to repressed chromatin is unlikely since the only J-DNA binding protein identified in trypanosome nuclear extracts is the thymidine hydroxylase involved in base J synthesis, JBP1 [33]. Gel-shift analysis of extracts from the JBP1 KO cell line failed to indicate any other J-DNA binding activity. Therefore, either the presence of base J modifies DNA structure by another mechanism or regulation is due to the glucose moiety interfering with formation of the Pol II initiation complex. Further studies are required to address the mechanism of base J regulation of transcription in *T. cruzi.*

Regulation of trypanosome mRNA abundance is usually determined by trans-acting factors interacting with cis-elements in each mRNA [34]. Transcriptome analysis showed significant up regulation of several RNA binding proteins in the JBP KO, including a family of proteins (i.e PUF 8) that are known to affect mRNA stability [35]. Therefore, increased abundance of certain trans-acting factors following the reduction in base J would lead to the destabilization of specific mRNA transcripts. This would explain the decreased steady-state abundance (and decreased half-life) of certain transcripts in the JBPKO cell lines following an overall increase in PTU transcription rate.

The presence of base J at convergent SSRs suggests that glucosylated DNA may also play a role in regulating transcription termination. However, the data presented here do not support a significant role for base J in Pol II termination in *T. cruzi*. Defects in termination would likely, as in yeast, lead to decreased initiation of downstream PTU and restricted elongation of convergent PTU due to polymerase collision [36]. Furthermore, we would see increased antisense RNA and subsequent decrease in transcript levels for

genes located at the 3' end of PTU. However, upon the loss of base J at TTSs of epimastigote and metacyclic cells, more genes are up regulated than down regulated and down regulated genes do not cluster at 3' ends of convergent PTU. In fact, many convergent regions have ORFs that are up regulated in the JBP KO in all life-stages analyzed. Transcriptome analysis of WT and JBP1 KO failed to detect any significant difference in the levels of antisense mRNA following the loss of J at TTS. Furthermore, direct analysis of nascent sense and antisense mRNA synthesis within three convergent SSRs failed to detect any changes in Pol II termination upon the loss of base J.

The analysis of the JBP1 and JBP2 mutant cell lines, with their differing levels and localization of base J and corresponding differences in number of genes affected, indicate that modulating base J can have specific gene expression/phenotypic consequences. This correlation is also apparent upon the deletion of each individual JBP allele. For example, upon the generation of the JBP1sKO and JBP1dKO we see a stepwise decrease in J levels in the genome (Figure 3.1E) as well as parasite metacyclogenesis (data not shown), and the corresponding increased ability of the parasite to invade the mammalian cell (Figure S3.7C). We believe that the JBP specific localization and levels of J at PTU flanking regions leads to differing levels and/or localization of transcriptional regulation within the genome. Presumably, in the case of the *in vitro* cell culture assays utilized here, the increase in gene expression changes upon deletion of both alleles of JBP is extensive (i.e. significant number of genes representing the various surface protein families are up regulated) such that the differences in the total number of genes affected between the JBP1KO and JBP2 KO is not reflected by a significant change in virulence.

The distribution of base J throughout the *T. cruzi* genome is unknown. Genome-wide analysis of base J in *T. brucei* indicated that ~70% of the PTU contain detectable levels of J within flanking regions [4]. Therefore, many PTU in the *T. brucei* genome lack significant levels of base J at sites of transcription initiation. Genome-wide analysis of J distribution combined with transcriptome analysis of additional JBP mutants (i.e. J null cell line) and life-stages will provide a more detailed analysis of J function in gene expression in *T. cruzi*. Interestingly, we have thus far been unable to generate a *T. cruzi* cell line lacking both JBP1 and JBP2. As previously demonstrated in Leishmania (vainio 2008), the essential nature of J in *T. cruzi* is supported by the dramatic concentration dependent hypersensitivity to BrdU exhibited by the JBP1KO and JBP2 KO cell lines (Ekanayake and Sabatini, unpublished results).

Reduction in base J levels in *T.cruzi* generated parasites with increased invasion efficiency and delayed egress. Whether such changes occur during infection of the mammalian host and lead to changes in pathogenesis are future topics to investigate. The ability of base J to regulate the transcription rate of virulence genes may have an evolutionary advantage for the long-term survival of the parasite in the mammalian host. Therefore, regulatable presence of J at the Pol II transcription initiation sites may provide an important control point for gene expression and parasite survival during its lifecycle.

ACKNOWLEDGEMENTS

The authors would like to thank Sergio Schenkman for assistance with nuclear run-on experiments, Melina Galizzi for help irradiating our mammalian cells and Norma Andrews for providing the amastin and trypomastigote specific antisera. We also thank Laura Cliffe for reading the manuscript. We would also like to thank Gwen Hirsch and Laura Cliffe for assistance with qRT-PCR analysis. Funding support was from National Institutes of Health Grant AI063523 (R.S), American Heart Foundation Grant 09PRE2290187 (D.K.E), AI044979 (R.T.) and from the Swiss National Science Foundation 31003A_125194/1 (T.O) and Novartis Foundation (T.O)

REFERENCES

1. Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62: 235-251.
2. Borst P, van Leeuwen F (1997) beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. Mol Biochem Parasitol 90: 1-8.
3. van Leeuwen F, Kieft R, Cross M, Borst P (1998) Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in Trypanosoma brucei. Mol Cell Biol 18: 5643-5651.
4. Cliffe LJ, Siegel TN, Marshall M, Cross GA, 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 Res 38: 3923-3935.
5. Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156: 93-101.
6. Ouellette M, Papadopoulou B (2009) Coordinated gene expression by post-transcriptional regulons in African trypanosomes. J Biol 8: 100.
7. Sabatini R, Cliffe L, Vainio S, Borst P (2009) Enzymatic Formation of the Hypermodified DNA Base J (β-D-Glucopyranosyloxymethyluracil) DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.
8. 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 Res 37: 1452-1462.

9. Vainio S, Genest PA, ter Riet B, van Luenen H, Borst P (2008) Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. Mol Biochem Parasitol 164: 157-161.

10. Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, et al. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res 35: 2107-2115.

11. Kieft R, Brand V, Ekanayake DK, Sweeney K, DiPaolo C, et al. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form Trypanosoma brucei. Mol Biochem Parasitol 156: 24-31.

12. Cross M, R. Kieft, Sabatini, R., A. Dirks-Mulder, I. Chaves and P. Borst (2002) J-Binding Protein Increases the Level and Retention of the Unusual Base J in Trypanosome DNA. Mol Micro 46: 37-47.

13. Camargo EP (1964) Growth and Differentiation in Trypanosoma Cruzi. I. Origin of Metacyclic Trypanosomes in Liquid Media. Rev Inst Med Trop Sao Paulo 6: 93-100.

14. Contreras VT, Araque W, Delgado VS (1994) Trypanosoma cruzi: metacyclogenesis in vitro--I. Changes in the properties of metacyclic trypomastigotes maintained in the laboratory by different methods. Mem Inst Oswaldo Cruz 89: 253-259.

15. da Silva CV, Kawashita SY, Probst CM, Dallagiovanna B, Cruz MC, et al. (2009) Characterization of a 21kDa protein from Trypanosoma cruzi associated with mammalian cell invasion. Microbes Infect 11: 563-570.

16. Vainio S GP, ter Riet B, van Luenen H, Borst P (2008) Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. Mol Biochem Parasitol 164: 157-161.

17. Kelly JM, Ward HM, Miles MA, Kendall G (1992) A shuttle vector which facilitates the expression of transfected genes in Trypanosoma cruzi and Leishmania. Nucleic Acids Res 20: 3963-3969.

18. Kelly JM, Taylor MC, Rudenko G, Blundell PA (1995) Transfection of the African and American trypanosomes. Methods Mol Biol 47: 349-359.

19. Minning TA, Weatherly DB, Atwood J, 3rd, Orlando R, Tarleton RL (2009) The steady-state transcriptome of the four major life-cycle stages of Trypanosoma cruzi. BMC Genomics 10: 370.

20. Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, et al. (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. PLoS Pathog 6.

21. Elias MC, Marques-Porto R, Freymuller E, Schenkman S (2001) Transcription rate modulation through the Trypanosoma cruzi life cycle occurs in parallel with changes in nuclear organisation. Mol Biochem Parasitol 112: 79-90.

22. Abuin G, Freitas-Junior LH, Colli W, Alves MJ, Schenkman S (1999) Expression of trans-sialidase and 85-kDa glycoprotein genes in Trypanosoma cruzi is differentially regulated at the post-transcriptional level by labile protein factors. J Biol Chem 274: 13041-13047.

23. Respuela P FM, Rada-Iglesias A, Aslund L. (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in Trypanosoma cruzi. J Biol Chem 283: 665-673.

24. Caler EV, Vaena de Avalos S, Haynes PA, Andrews NW, Burleigh BA (1998) Oligopeptidase B-dependent signaling mediates host cell invasion by Trypanosoma cruzi. EMBO J 17: 4975-4986.

25. Ekanayake DK, Cipriano MJ, Sabatini R (2007) Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite Trypanosoma cruzi. Nucleic Acids Res 35: 6367-6377.

26. Frasch AC (2000) Functional diversity in the trans-sialidase and mucin families in Trypanosoma cruzi. Parasitol Today 16: 282-286.

27. 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 Dev 23: 1063-1076.

28. Bergman Y, Mostoslavsky R (1998) DNA demethylation: turning genes on. Biol Chem 379: 401-407.

29. Heusipp G, Falker S, Schmidt MA (2007) DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol 297: 1-7.

30. Marinus MG, Casadesus J (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev 33: 488-503.

31. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H (2003) The role of DNA methylation in setting up chromatin structure during development. Nat Genet 34: 187-192.

32. Fuks F (2005) DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev 15: 490-495.

33. Cross MR, Kieft R, Sabatini M, Wilm M, de Kort GA, et al. (1999) The Modified Base J is the Target for a Novel DNA-Binding Protein in Kinetoplastid Protozoans. EMBO J 18: 6573-6581.

34. Queiroz R, Benz C, Fellenberg K, Hoheisel JD, Clayton C (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. BMC Genomics 10: 495.

35. Droll D, Archer S, Fenn K, Delhi P, Matthews K, et al. (2010) The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. FEBS Lett 584: 156-162.

36. Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci 99: 8796-8801.

Table 3.1. Increased transcription rate in JBP1 KO by run-on analysis

	Fold upregulation*	±SD
24s	1.00	
SSU-RNA	2.71	0.80
5S	3.65	2.43
SL-RNA	6.58	1.50
Gp85	3.63	1.51
Amastin	6.71	3.65
MASP	0.92	0.15
TS-Enzyme	2.97	1.39
Enolase	3.62	1.88
Tubulin	8.22	2.86
ABC1	1.22	0.24
Hypo	2.00	0.69
P1	3.12	0.86
P2	3.31	0.58
P3	4.96	2.12

*** WT vs JBP1KO**

Table 3.2. Spliced leader trapping analysis of epimastigote cells

	WT	JBP1 KO
Number of reads	3381708	3247634
Sense tags	93%	94%
Antisense tags	0.7%	0.5%
SSR tags	6.2%	5.6%

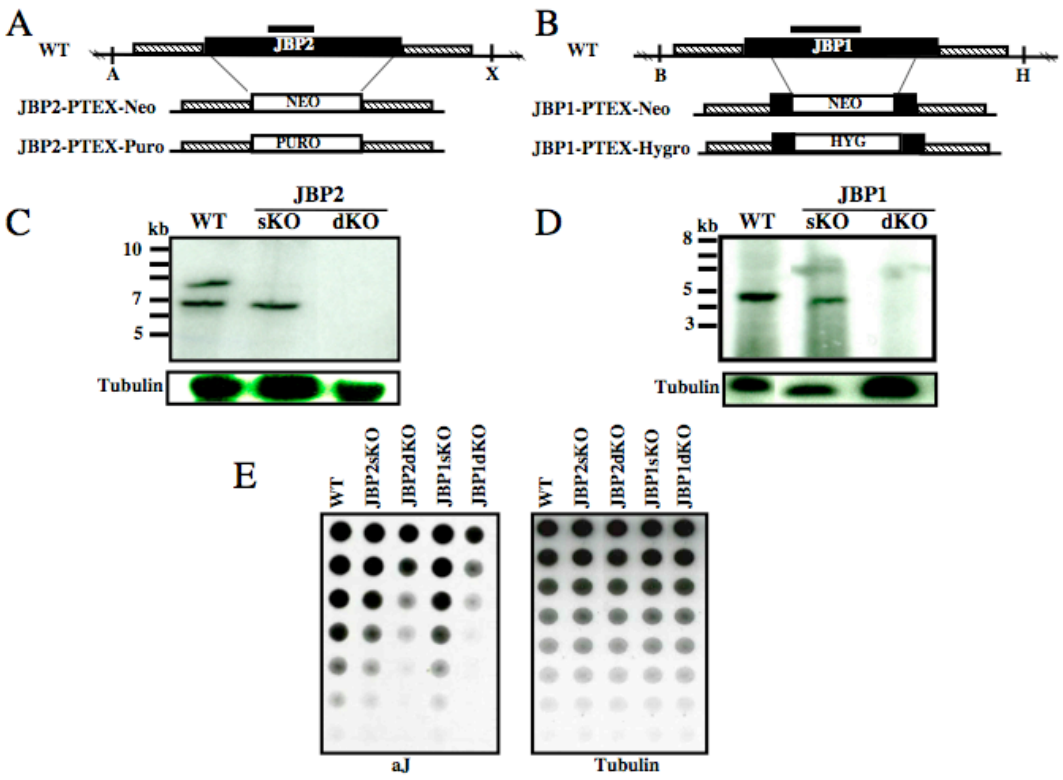


Figure 3.1. JBP knockout (JBPKO) cells have reduced levels of base J. (**A** to **D**) Generation of the JBP2 double knockout (JBP2dKO) and JBP1dKO cell lines. (**A** and **B**) Schematic diagrams of targeted deletion of JBP2 and JBP1 ORFs using neomycin (Neo), hygromycin (Hyg), and puromycin (Puro) deletion constructs. Homologous regions are indicated by the hatched bars. The black bars above the JBP ORF indicate the positions of the probes used for Southern blotting. A, AlwNI; X, Xcm1; B, BamHI; H, HindIII. (**C** and **D**) Southern blot analysis to confirm the loss of JBP2 and JBP1 ORFs. AlwNI-and-XcmI- or BamH-and-HindIII-digested genomic DNAs from the indicated cell lines (WT, single knockout [sKO], and double knockout [dKO]) were fractionated by gel electrophoresis, blotted onto nylon membranes, and hybridized using radioactive probes

for the corresponding JBP ORFs. The stripped blots were hybridized with radioactively labeled tubulin probe as a control for DNA loading. Southern blot analysis for the JBP2 ORF results in two bands due to a restriction site polymorphism existing in the two alleles. (**E**) Reduction of global levels of base J in JBP1 and JBP2 knockout cells. DNA was isolated from the indicated cell lines, blotted onto nitrocellulose membrane in a 2-fold dilution series, and incubated with anti-base J antisera (αJ). The membrane was then hybridized with a radioactive tubulin probe to control for DNA loading.

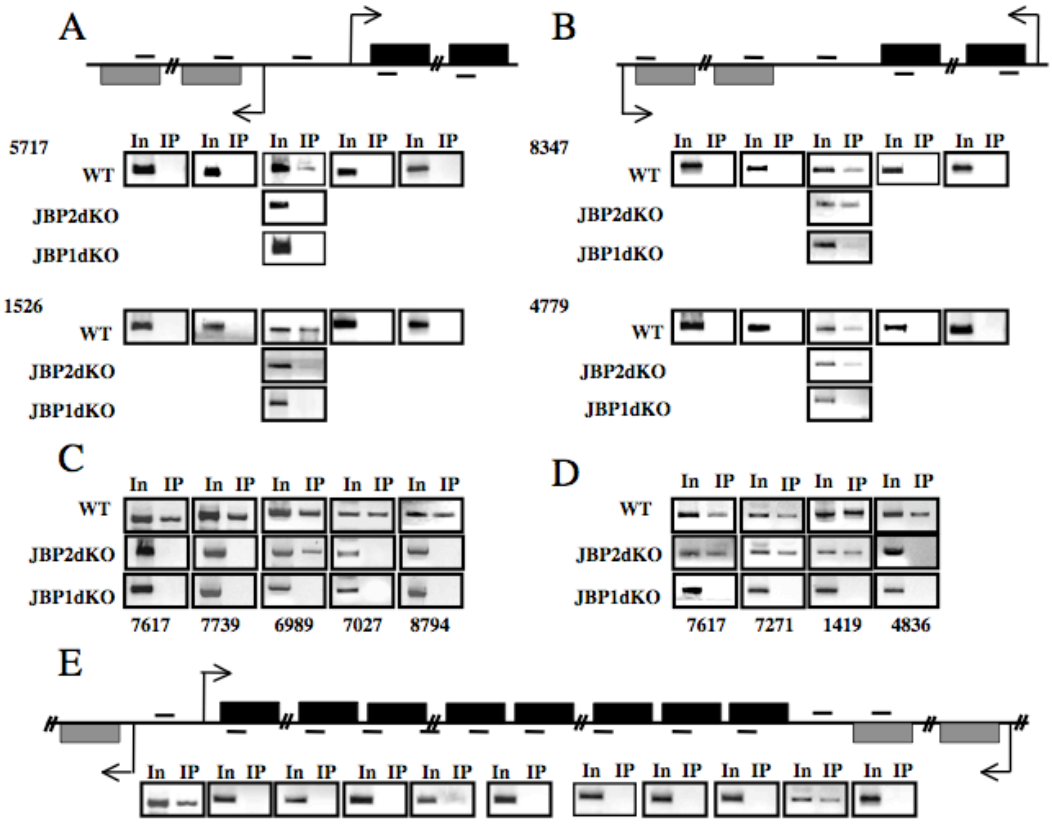


Figure 3.2. JBP1 and JBP2 regulate base J synthesis in polycistronic unit (PTU)-flanking regions of *T. cruzi*. (**A** and **B**) Schematic diagram of a divergent strand switch region (SSR) (**A**) and a convergent SSR (**B**) and the distribution of base J analyzed by anti-base J immunoprecipitation (IP) and PCR of representative regions of the T. cruzi genome (below). The boxes represent ORFs. The black lines indicate the PCR-amplified regions tested in the IP reactions shown below. The identities of the SSRs are indicated as previously described (30). (**C** and **D**) Anti-base J IP and/or PCR analysis of additional diverging SSRs (**C**) and converging SSRs (**D**). The identity of the SSR is indicated below. (E) Distribution of base J within an entire ~200-kb-long PTU (AAHK01000001) and the flanking divergent and convergent SSRs. PCR-amplified regions are spaced approximately 20 kb apart. The map is not drawn to scale.

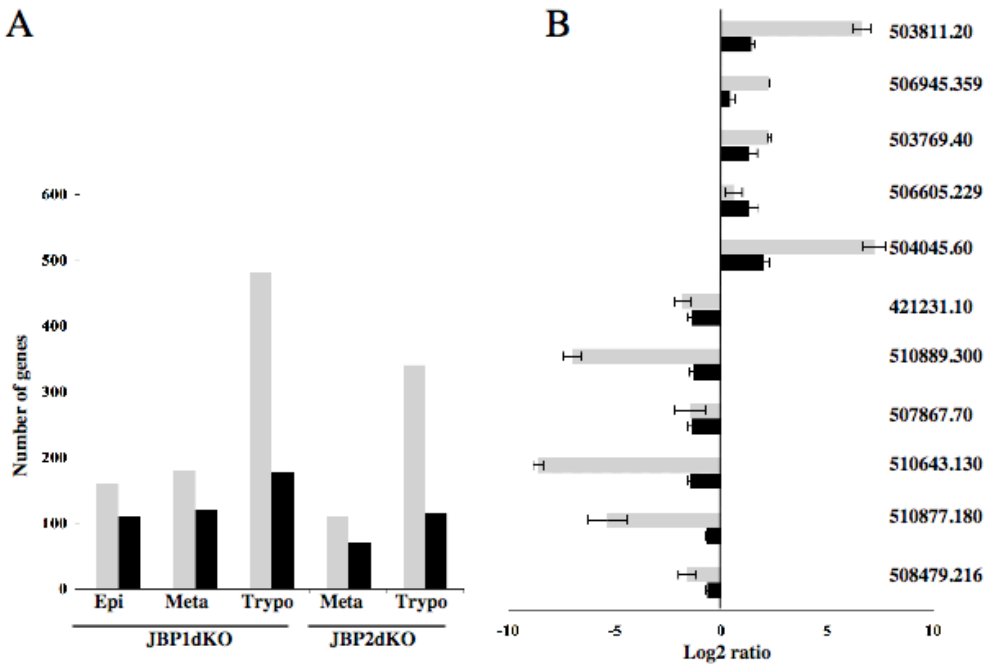


Figure 3.3. Microarray analysis of WT and JBPdKO cell lines. (**A**) Number of affected genes in JBP1dKO and JBP2dKO cells of epimastigote (Epi), metacyclic (Meta), and trypomastigote (Trypo) life stages as indicated by microarray analysis. The black bars indicate downregulated genes, and the gray bars indicate upregulated genes. (**B**) Quantitative RT-PCR analysis. Mean log2 ratio (stage/reference) for selected significantly up- or downregulated genes determined by microarray of JBP1dKO trypomastigotes (black bars). The gray bars indicate the genes significantly up- or downregulated genes by qRT- PCR analysis. The values are means ± standard deviations (SDs) (error bars) for three biological replicates. The ORFs are labeled using their GeneIDs (Tc00.1047053 is truncated).

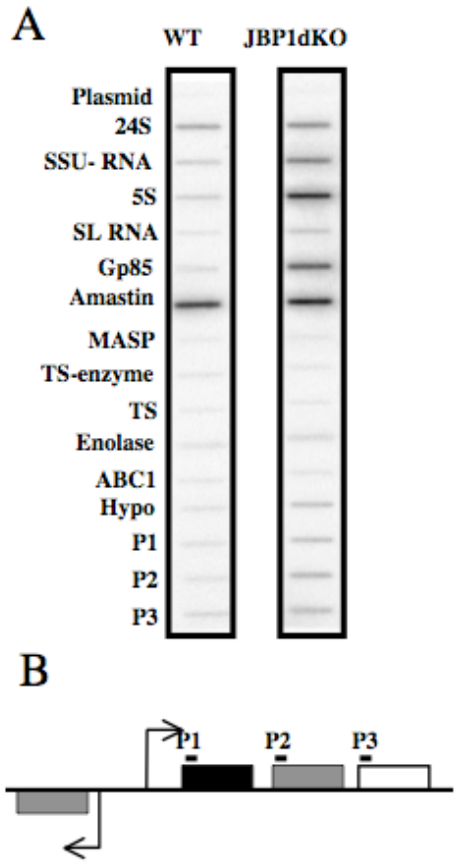


Figure 3.4. Base J regulates Pol II transcription rate. (**A**) Lysoleci- thin-permeabilized parasites from WT and JBP1dKO epimastigotes were labeled with [α-32P]uridine for 30 min. The total RNA was extracted and used to probe nitrocellulose filters containing probes for the indicated genes. The blot shown is representative of the results of five independent experiments. The RNAs shown include plasmid RNA, 24S RNA, small-subunit rRNA (SSU- rRNA), spliced-leader (SL) RNA, 85-kDa surface protein gene family (Gp85), mucin-associated surface proteins (MASP), trans-sialidases (TS-enzyme and TS), hypothetical Tc00.1047053508773.9 (Hypo), and ABC1 (Tc00.1047053508221.610). P1, P2, and P3 indicate the sequences from 3 regions within

a PTU from divergent SSR 7739 (chromosome 11, 120,001 bp to 160,000 bp). (**B**) Schematic repre- sentation of divergent SSR 7739 to indicate the locations of P1, P2, and P3 probes (black lines). ORFs not represented on the array (gray boxes), ORFs with increased steady-state transcript levels (black box), and ORFs with no change in transcript levels (white box) are indicated.

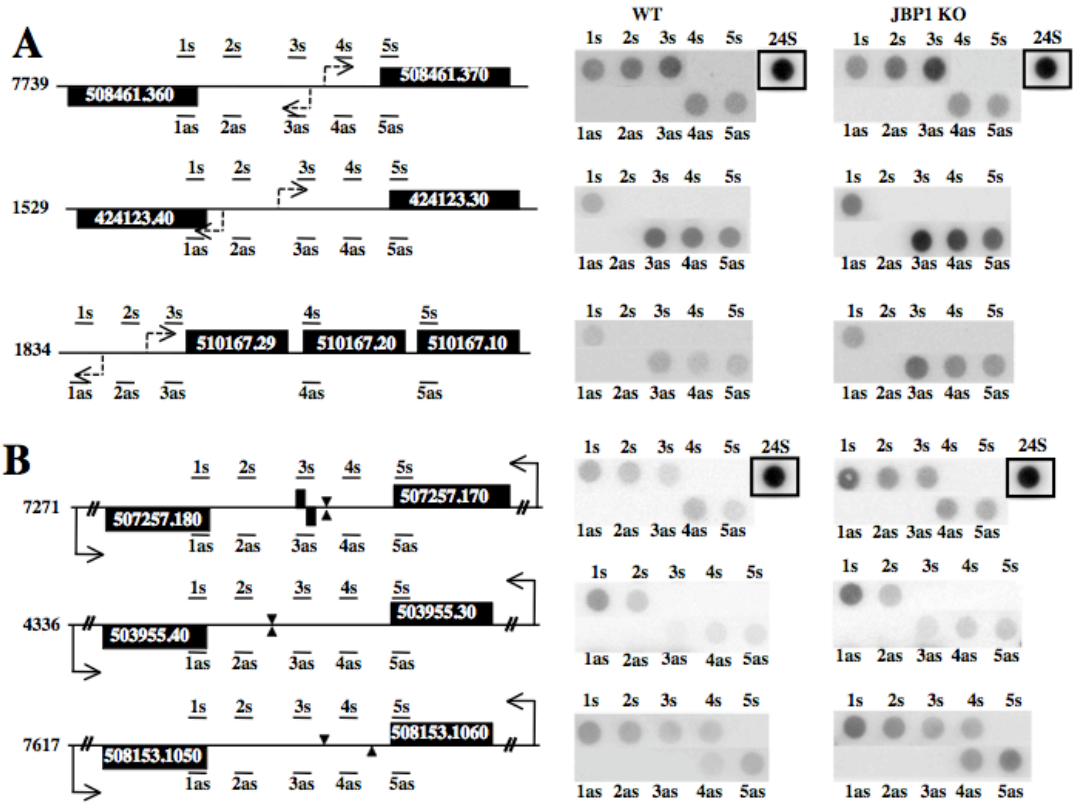


Figure 3.5. Transcription initiation and termination from sense(s) and antisense(as) strands within divergent and convergent SSRs. Nuclearrun-on assay was used to map the transcription start sites (TSSs) and transcription termination sites (TTSs) and measure the strand bias of transcription in divergent and convergent SSRs. Labeled nascent RNA was hybridized to filters spotted with 2 μg of 50-mers corresponding to both the sense (1s

to 5s) and antisense strands (1as to 5as) along the SSR. Cartoons represent the divergent SSR (**A**) and convergent SSR (**B**) and depict the locations of oligomers used in nuclear run-on assays. The broken-line arrows in panel A represent the approximate locations of the TSSs. The black triangles in panel B indicate the approximate region of transcription termination. The small black boxes represent tRNA. Numbers refer to the SSR analyzed as previously described (30). (Right) Results of run-on analysis of WT and JBP1 dKO epimastigotes. The blots shown are representative of the blots in 2 to 4 experiments. The 24S probe, whose signal does not change in WT and JBP1KO cells (Figure 3.4 and Table 3.1), is included as a loading control.

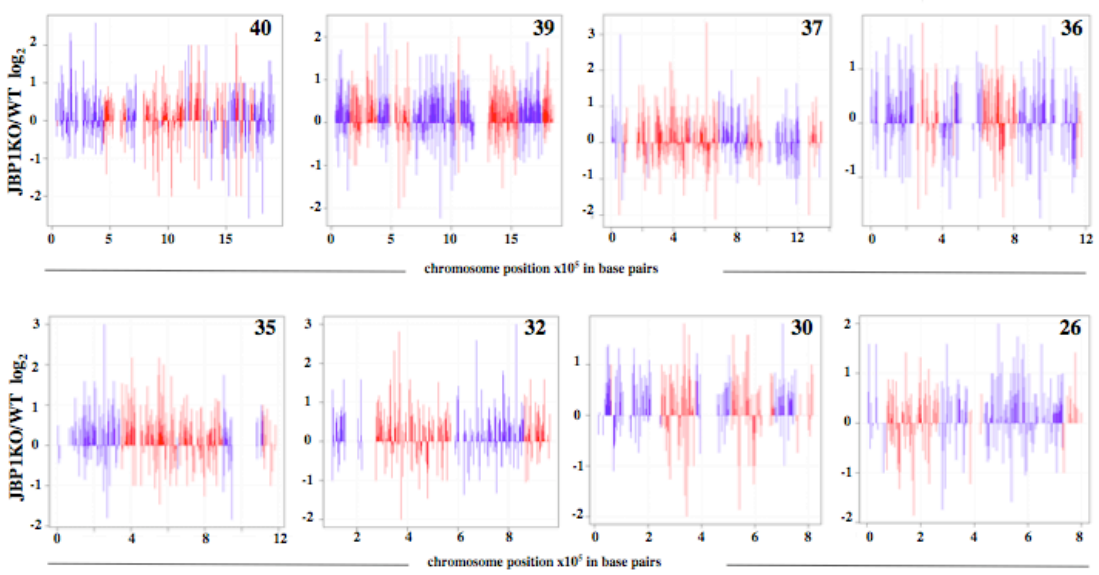


Figure 3.6. Whole-genome comparison of gene expression levels in WT cells versus JBP1KO cells. The gene expression ratios (log2 of the ratio) along the whole chromosome of the expression of genes in JBP1KO epimastigotes compared to that in WT epimastigotes using RNA sequencing are shown. Each line represents the gene expression ratio of one gene with the color indicating the strandedness (red for the sense strand and blue for the antisense strand). The numbers on the x[r] axis indicate the position on the respective chromosome in base pairs (~10^5). The boldface numbers in the top right corner of each graph indicate the chromosome number.

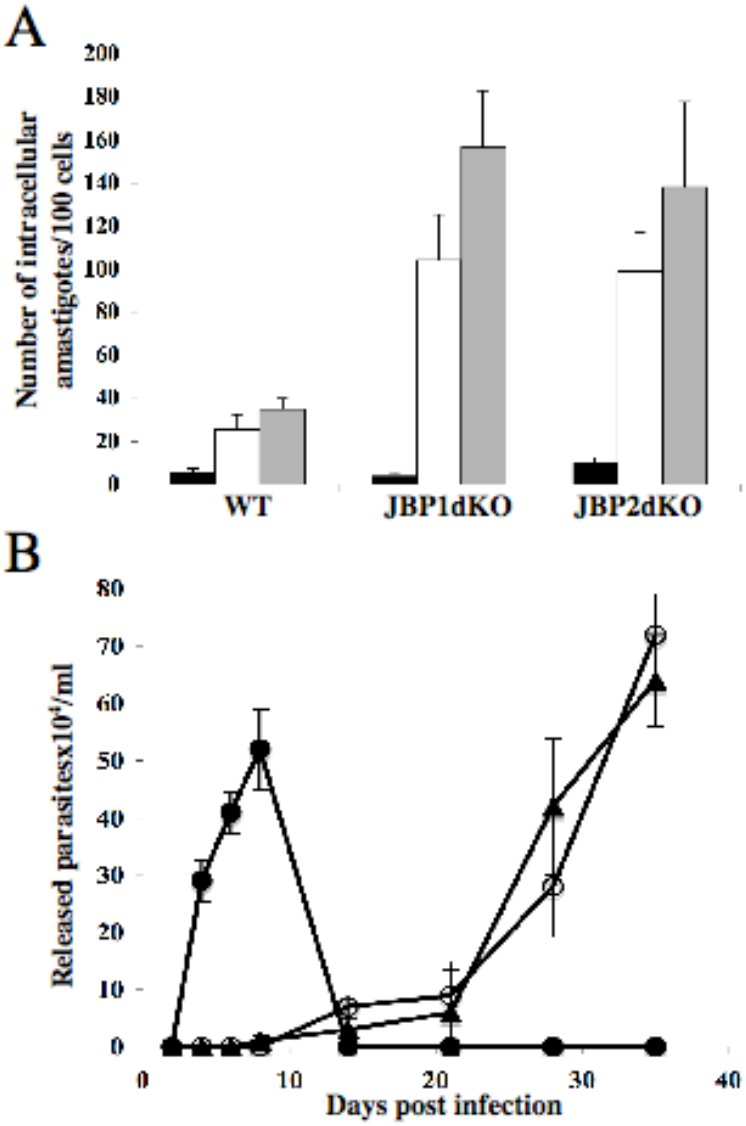


Figure 3.7. Base J regulates T. cruzi virulence. (**A**) Mammalian invasion. WT and JBPKO parasites were allowed to invade Vero cells/ myoblasts for 3 h. The number of parasites/100 cells was determined as described in Materials and Methods. Values are means ± SDs (error bars) of reactions done in triplicate. The stage of the parasite is indicated as follows: black bars, epimastigote; light gray bars, metacyclic form; dark gray bars, trypomastigote. (**B**) Parasite egress. WT and JBPdKO amastigote parasites were

grown in irradiated Vero cells, and released trypomastigotes were counted on the indicated days post-infection. Symbols: black circles, WT cells; white circles, JBP1dKO cells; black triangles, JBP2dKO cells. The error bars represent the standard deviations of three independent experiments.

Table 3S.1: Number of surface proteins upregulated for different life stages of JBPKOs

Multi gene family genes	Micro array				SLT	
	JBP1dKO		JBP2dKO		JBP1dKO	
	Meta	Tryp	Meta	Tryp	Meta	Tryp
MASP	0	1	1	1	10	61
Amastin	0	1	0	1	0	0
Gp63	0	1	1	1	11	94
Trans-sialidase	4	12	11	15	19	243
DGF-1	0	2	0	2	32	72
TuMuc	4	3	3	10	4	72
TcMucI	N/A	N/A	N/A	N/A	2	6
TcMucII	N/A	N/A	N/A	N/A	1	37
TcMucIII	N/A	N/A	N/A	N/A	0	1
TcSMUG	N/A	N/A	N/A	N/A	0	15
TcSMUGL	N/A	N/A	N/A	N/A	0	8
TcSMUGS	N/A	N/A	N/A	N/A	0	7

*Different members of the gene family are affected in each life
N/A, not on the micro array

Table S2. PTU distribution of regulated genes of JBP2dKO and JBP1dKO

Chr	Number of Genes	Number of genes on the array	Number of PTUs	JBP1KO		JBP2KO	
				Upregulated genes (*location)	Downregulated genes (*location)	Upregulated genes (*location)	Downregulated genes (*location)
41	1500	425	39	8(2,5,7,11,12,19,24,26,27)	1(17)	9(6,13,19,22,26,33,39)	2(11,19
40	1185	527	22	15(1,2,4,14,15,19,21,22)	0	6(1,2,4,55,14,21)	1(21)
39	1132	764	10	14(2,3,4,5,7,8,9,10)	0	5(2,4,5,7,9)	1(7)
38	751	244	19	9(2,4,19)	1(2)	1(10)	2(2,10)
37	1001	559	11	14(2,4,5,6,7,8,10)	3(4,5)	3(4,8,9)	2(4,10)
36	896	430	7	15(1,2,3,4,5,6)	1(6)	8(1,2,3,6)	2(4,6)
35	845	432	6	11(2,4,5,6)	1(2)	2(1,5)	2(2)
34	706	379	5	11(1,2,3,4)	2(1,2)	4(1,2)	1(1)
33	660	264	7	6(1,3,6)	6(6,7)	3(2,3,7)	3(6,7)
32	807	422	4	12(1,2,3,4)	4 (1,2,3,)	5(1,2,3)	6(2,3,4)
31	545	251	10	12(1,3,4,8,10)	1(8	1(3,10)	0
30	619	10	5	17(1,3,4,5,6,7,8,9)	2(7,9)	1(7)	0
29	524	208	14	2(4,14)	1(5)	8(8,9,13,14)	0
28	429	122	10	2(6,8)	1(2)	1(10)	0
27	605	316	3	4(1,2,3)	3(2,3	4(3,4)	3(1,3)
26	572	276	7	3(6,4)	3(4,6)	0	0
25	427	171	4	3(1,2	2(3,4)	1(1	0
24	411	157	4	3(1,2,4)	0	1(1)	1(3)
23	441	193	7	4(6,7)	1(6	3(1,6)	2(1,5)
22	467	195	6	6(2,6)	2(1,5	6(2,6)	1
21	420	191	5	6(3,4,6)	1(5,6)	2(3)	0
20	533	250	6	7(1,2,3,4)	0	4(2,3,4)	0
19	338	163	4	3(1,2,4)	2 (2,3)	1(1)	1(3)
18	438	139	18	2(10,13)	0	1(16)	0

17	396	170	4	7(1,2,4)	5(1,2)	4	0
16	475	209	7	9 (1,2,3,4,5)	1(2)	1(2)	1(2)
15	322	135	9	6 (3,4)	1 (4)	7 (2,3,4,5)	0
14	388	188	7	2(1,4)	0	1(1)	0
13	368	185	5	4(1,2)	2 (1,4)	4 (1,3)	1(4)
12	307	146	4	5 (2,3)	1(3)	0	1(3)
11	336	160	4	7(1,2)	3(2,3,4)	1(4)	1(2)
10	374	185	5	3 (2,3,4,5)	1	5(2,3,4)	0
09	331	159	5	7 (1,2,3)	1(2)	2(1,2)	0
08	265	135	6	3(1,2)	1(1)	3(1,3)	0
07	252	117	3	3 (1)	1(1)	1(1)	1(1)
06	274	141	5	3(1,2,3)	2 (2,3)	1(2)	2(2,3)
05	149	81	3	2(1)	1(1)	0	0
04	131	66	3	3(1)	1(1)	2(2)	1(b)
03	130	60	3	0	2(1)	0	2(1,2)
02	119	48	2	3(1)	1(1)	0	1(2)
01	44	20	1	1(1)	1	0	0 (1)

* PTU number starting from left to right representing each gene

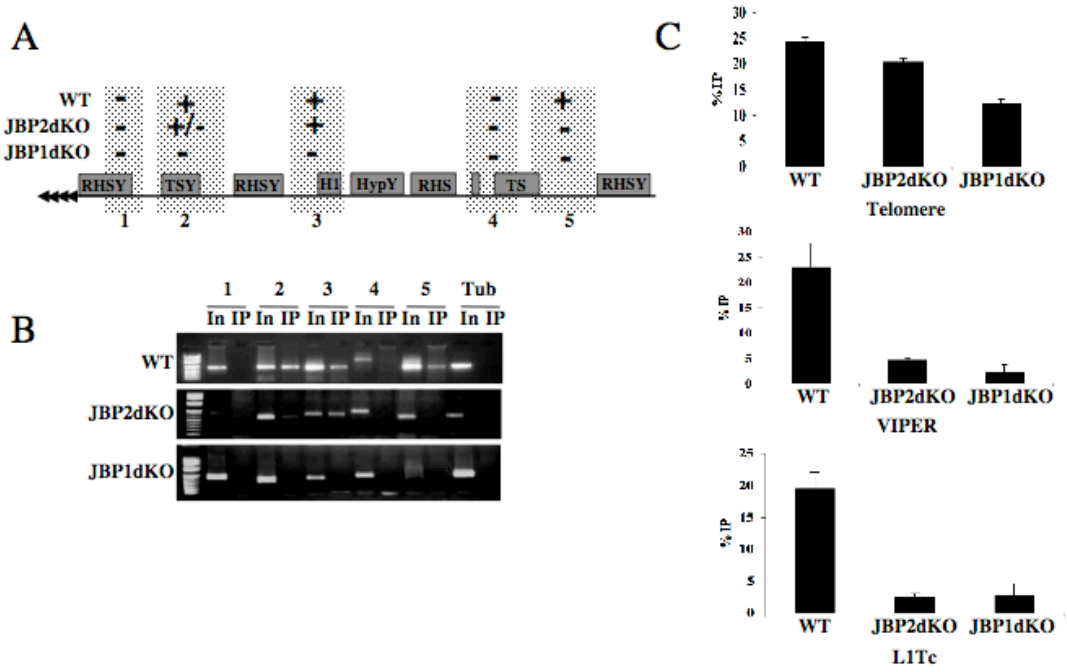


Figure 3S.1: JBP1 and JBP2 regulate J synthesis in telomeric DNA, related to Figure 3.1.

(**A-C**) Telomeric and subtelomeric localization of base J. (**A**) Map of assembled telomeric contig CH473453 (Ekanayake et al., 2007). Patch boxes indicate the regions assayed by anti-J ChIP assay where EcoRI restricted digested fragments were assayed by immunoprecipitation with anti-J as described in (Ekanayake et al., 2007). RHSΨ, retrotransposon hot spot peudo gene; TSΨ, trans-sialidase pseudo gene; HypoΨ, hypothetical pseudo gene; H1 and H2, hypothetical genes. (**B**) PCR analysis of the IP reactions performed using primers specific for the indicated regions (1-5) for WT, JBP2dKO and JBP1dKO. Input DNA (IN) and Immunoprecipitated DNA (IP) from the indicated cell lines were used for PCR reactions. Tubulin primers are used as a negative control. (**C**) Phosphoimage quantification of imunoprecipitated fraction of base J within the indicated repeat regions of WT and knockout cell lines. (**D**) Reduction of global levels of base J by BrdU. WT cells were grown in the presence of 100 ug/ml BrdU for 5 days. DNA was

isolated from WT cells grown in the absence (WT) and presence (BrdU) of BrdU and J levels determined as in Figure 3.1E.

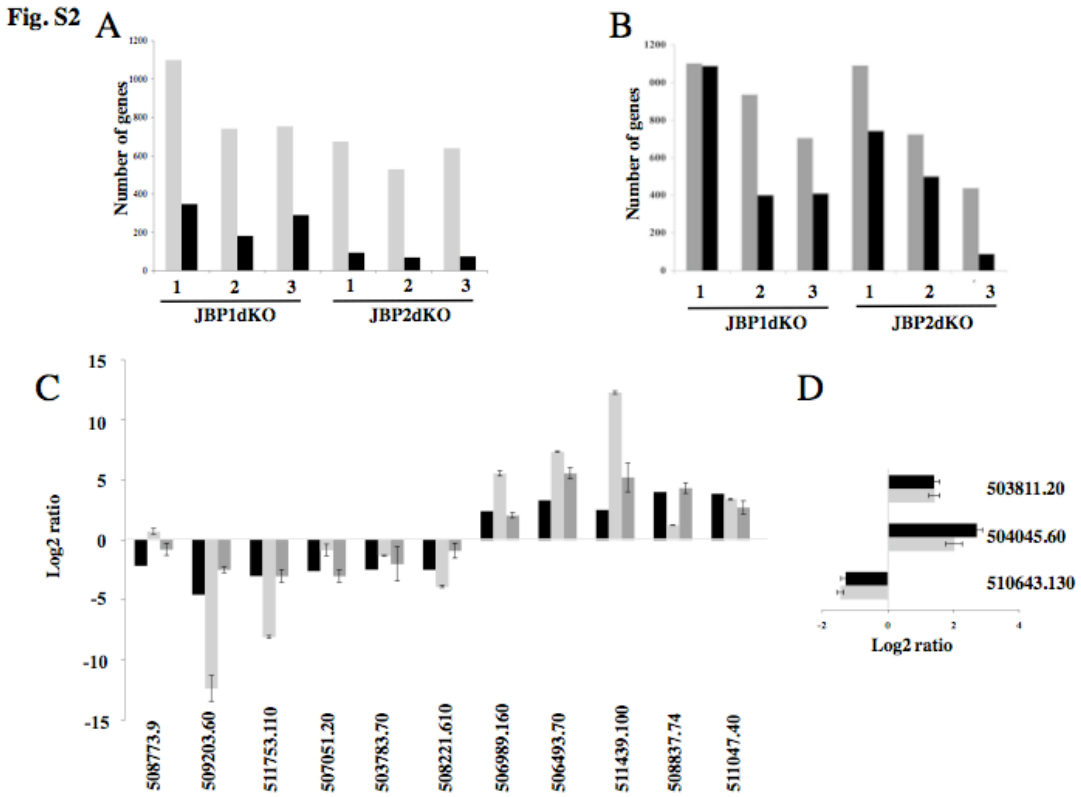


Figure 3S.2: **Microarray analysis of the JBP1dKO and JBP2dKO, related to Figure 3.3.** (**A-B**) Number of deregulated genes in three biological replicates of JBP1dKO and JBP2dKO trypomastigotes (**A**) and metacyclics (**B**). Grey bars represent the up regulated genes and black bars represent the down regulated genes. (**C**). Quantitative RT-PCR analysis. Mean log2 ratio (stage/reference) for selected significantly up and down regulated genes determined by microarray of JBP1dKO epimastigotes. Black bars represent microarray analysis and light grey bars represent qRT-PCR analysis of changes in expression in the JBP1 KO. Dark grey bars

represent qRT-PCR analysis of changes in expression of WT epimastigotes fed BrdU as described in Figure S3.1D. Error bars are standard deviations for three biological replicates. (**D**) RT-PCR analysis of identical trypomastigote RNA samples used in the array analysis as in Figure 3.3B.

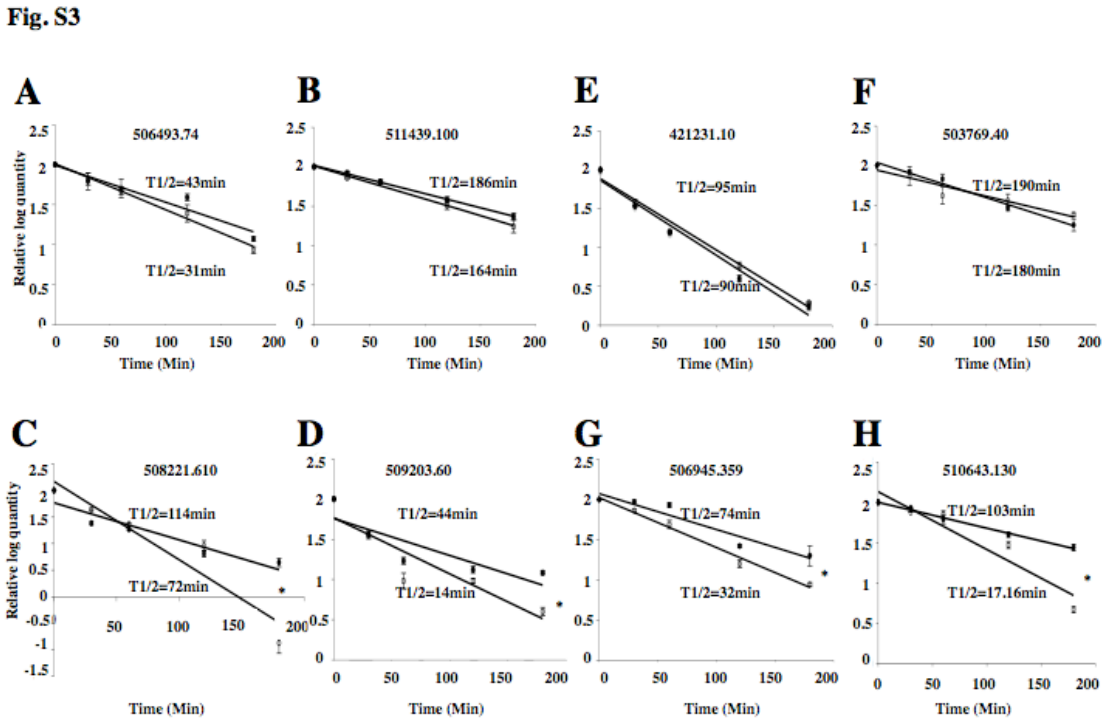


Figure 3S.3: RNA stability of selected genes assayed by qRT-PCR, related to Figure 3.4. Total RNA was obtained from WT and JBP1dKO epimastigote (**A-D**) and metacyclic (**E- H**) samples at the indicated times after the addition of 100µg/m1 of Actinomycin D. mRNA levels were determined by qRT-PCR and mRNA half lives were estimated as described in Experimental Procedures (WT, solid circles; JBP1dKO, open circles). Identity of up regulated transcripts (**A-B**, **E-F**) and down regulated transcripts (**C-D**, **G-H**) is indicated on the graph

(geneID as in Figure 2). Half-lives are indicated on the representative graphs. Half-life values between WT and JBP1dKO were compared using Student's t-test and *P* value <0.001 (*) was considered statistically significant.

Fig. S4

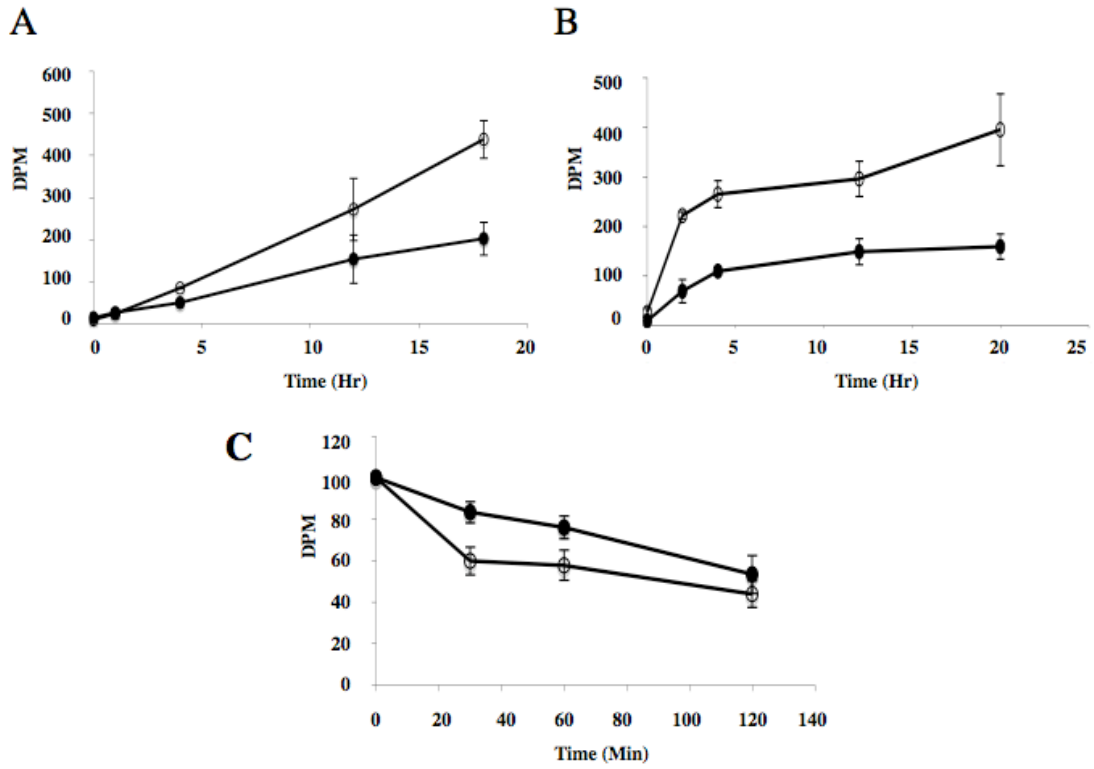


Figure 3S.4: Effect of base J on *in vivo* transcription rate, related to Figure 3. 4.

Epimastigotes (**A**) and trypomastigotes (**B**) forms of WT and JBP1dKO were incubated with [3H] uridine and at the indicated periods of time total RNA was extracted and the labeled RNA was counted by liquid scintillation. (**C**) WT and JBP1dKO epimastigotes were incubated with [3H] uridine for 4hrs and Actinomycin D was added at 100µg/ml concentration. Total RNA was extracted from the sample taken at the indicated time period and labeled RNA was counted by liquid scintillation. (WT, solid circles; JBP1dKO, open circles).

Fig. S5

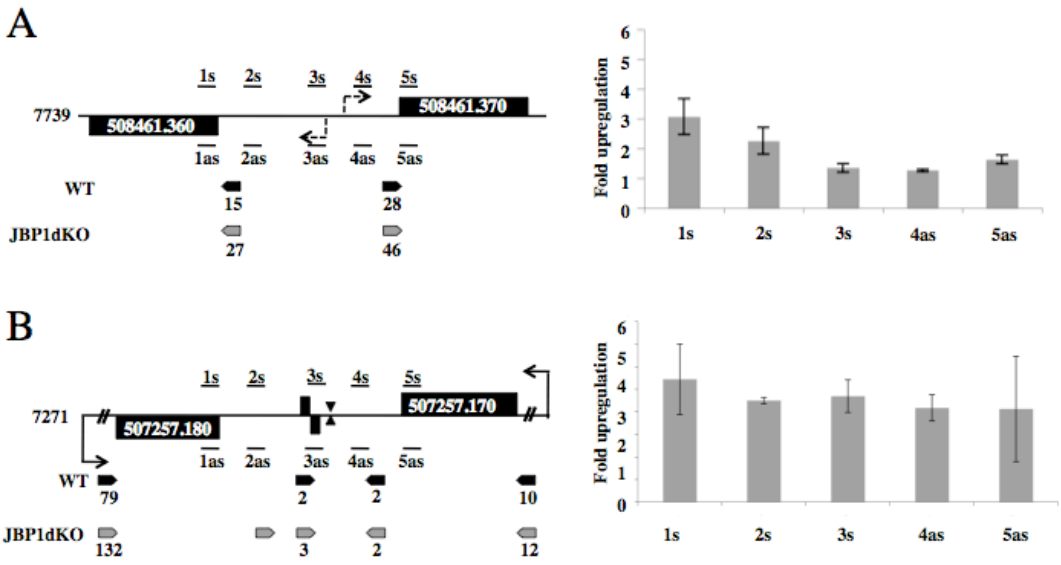


Figure 3S.5: SLT and run-on analysis of SSRs, related to Figure 3.5. (**A**) Divergent and convergent (**B**) transcription unit with sequence tags from spliced leader sequencing (SLT tags) in tags/million, representing the steady state abundance of the corresponding mRNAs in WT and JBP1 KO cell lines. Arrows indicate the site of transcription initiation (**A**) and termination (**B**) determined in Figure 3.5. On the right is quantification of the nuclear run-on analysis for the region (see Figure 3.5). The average increase in transcription rate for all regions analyzed within the three diverging and three converging SSRs is 2.5 and 2.4 fold, respectively.

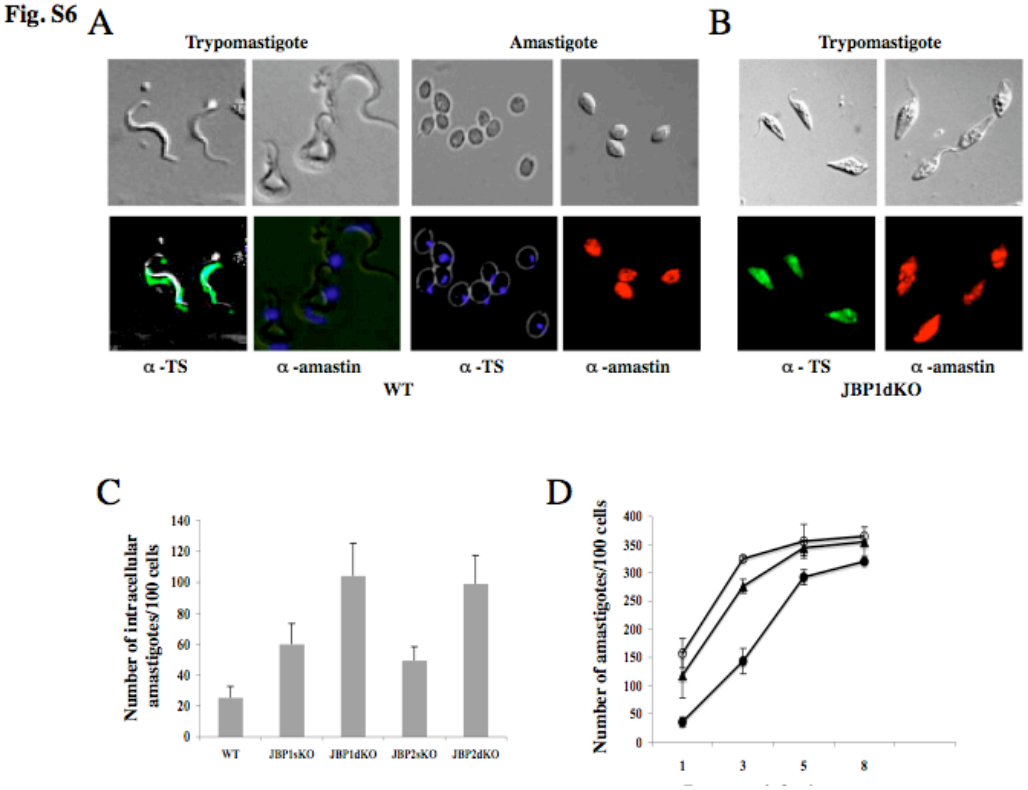


Figure 3S.6. Virulence phenotype analysis of JBP KO cell lines, related to Figure 3. 7. (**A** and **B**) Expression of stage specific amastigotes and trypomastigotes specific surface proteins determined by indirect immuno-florescence. Insets of FITC (α-trypo specific TS) and rhodamine (α-amastin) and DAPI staining WT trypomastigotes and amastigotes (**A**) and JBP1dKO trypomastigotes (**B**). (**C**) Correlation of virulence and the amount of base J. Vero/ myoblast cells were infected with tissue culture derived trypomastigotes of the indicated cell lines and the number of parasites/100 cells was determined as described in Experimental Procedures. Values represent the mean ± SD of 3 independent experiments. (**D**) Intracellular growth of JBPKO parasites. Number of intracellular immunoflorescence positive amastigotes was counted during first 5days in WT and JBPdKO cell lines to determine the intracellular growth rates. Solid circles, WT; open circles, JBP1dKO; solid triangles, JBP2dKO. Error bars represent the standard deviation of three independent experiments.

CHAPTER 4

EPIGENETIC REGULATION OF POL II TRANSCRIPTION INITIATION IN *TRYPANOSOMA CRUZI*: MODULATION OF NUCLEOSOME ABUNDANCE, HISTONE MODIFICATION AND POLYMERASE OCCUPANCY BY O-LINKED THYMINE DNA GLUCOSYLATION

[1] Dilrukshi K Ekanayake and Robert Sabatini. *To be submitted*

ABSTRACT

Very little is understood regarding how transcription is initiated/regulated in the early-diverging eukaryote, *Trypanosoma cruzi*. Unusually for a eukaryote**,** genes transcribed by RNA polymerase (Pol) II in trypanosomes including *T. cruzi* are arranged in polycistronic transcription units (PTUs). Based on this gene organization it was previously thought that trypanosomes rely solely on posttranscriptional processes to regulate gene expression. We recently localized a novel glucosylated thymine DNA base, called base J, to potential promoter regions of PTUs throughout the trypanosome genome. Loss of base J, following the deletion of JBP1, a thymidine hydroxylase involved with synthesis, led to a global increase in Pol II transcription rate and gene expression. In order to determine the mechanism by which base J regulates transcription, we have characterized changes in chromatin structure and Pol II recruitment to promoter regions following the loss of base J. The loss of base J coincides with a decrease in nucleosome abundance, increased histone H3/H4 acetylation and increased Pol II occupancy at promoter regions, including the well characterized SL RNA gene promoter. These studies present the first direct evidence for epigenetic regulation of Pol II transcription initiation via DNA modification and chromatin structure in kinetoplastids as well as providing a mechanism for regulation of trypanosome gene expression via the novel hyper-modified base J.

INTRODUCTION

Trypanosoma cruzi is a major human parasitic pathogen with complex life cycle, alternating between an insect vector and mammalian hosts [1]. In both, *T. cruzi* undergoes differential morphological and functional changes that require rapid and selective changes in gene expression profile. Unlike other eukaryotes, the genes in *T. cruzi* are arranged in large polycistronic bi-directional gene clusters [2]. These polycistronic transcription units (PTUs) are transcribed by RNA polymerase (Pol) II to yield polycistronic pre-mRNAs that are then processed into mature mRNAs by trans-splicing and polyadenylation. Given this genome arrangement, and the apparent lack of functional relationship of genes within a PTU, the regulation of trypanosome gene expression was thought to occur primarily via differential mRNA decay or other post-transcriptional mechanisms [3]. Until very recently, there was no evidence for regulation of gene expression at the level of Pol II transcription in kinetoplastids [4]. Furthermore, very little is understood about the DNA sequences and proteins involved in transcription initiation and termination in kinetoplastids. With the exception of the spliced-leader RNA promoter, Pol II promoters and associated factors have not been identified. It has been suggested that histone modifications and histone variants are involved in transcription initiation and termination in trypanosomes, but no direct evidence exists to support this [2,5]. Thus, how Pol II transcription is initiated/regulated, including the role of chromatin structure, in trypanosomes is unclear.

We have recently localized the novel hyper-modified DNA base, β-D-glucosyl-hydroxymethyuracil or base J, at regions flanking PTUs throughout the *T. brucei* genome

[6]. These PTU flanking regions primarily correspond to divergent and convergent strand-switch regions (SSR) involved in Pol II transcription initiation and termination, respectively. This specific localization of base J suggested a role of the modified DNA base in regulating Pol II transcription initiation and termination and therefore gene expression. To directly address this, we reduced the levels of the modified base in the genome by deleting enzymes involved in J biosynthesis [4]. Base J is synthesized in two steps where a thymidine hydroxylase converts a thymidine residue into hydroxymethyldeoxyuridine (HOMedU), which is subsequently converted into base J by the addition of glucose by a glucosyltransferase [7,8,9]. Previous work has identified two distinct thymidine hydroxylases involved in J biosynthesis, JBP1 and JBP2 [9,10,11,12,13]. Deletion of JBP1 and JBP2 in *T. cruzi* resulted in 20- and 8-fold reduction in base J levels, respectively [4]. While the remaining J in these knockout (KO) trypanosomes is located in telomeric DNA, base J is lost from the divergent and convergent strand switch regions. The loss of base J at promoter regions led to an increased rate of Pol II transcription and global changes in gene expression and parasite virulence [4]. In contrast, the loss of base J at convergent strand-switch regions involved in Pol II transcription termination did not result in any detectable defects in termination. While these studies provided the first direct evidence for epigenetic regulation of gene expression in kinetoplastids at the level of Pol II transcription, they do not explain how base J regulates transcription.

Transcription in eukaryotes is influenced by chromatin structure. In particular, the packaging of DNA into chromatin imposes significant obstacles to transcription initiation by Pol II [14,15]. The structure of chromatin can regulate the binding of

proteins/complexes to promoter regions including the recruitment of Pol II and formation of the preinitiation complex. Thus, chromatin is highly dynamic and continuously shifting between an open, transcriptionally active conformation and a compact, silenced one [14]. This fluid nature of chromatin is tightly regulated through multiple mechanisms, including histone modification, nucleosome density, and DNA methylation [15]. Global mapping of nucleosomes in yeast demonstrates that upon gene activation nucleosomes are evicted at many promoters [16,17,18] and are reassembled upon gene silencing [19,20]. Histone acetylation is associated with an increase in the accessibility of DNA to transcriptional machinery [21]. Acetylation of lysine residues at the N-terminal domain of histone H3 and H4 weaken the interactions with DNA and result in a destabilization of nucleosomal structures and activation of gene transcription [22,23]. Accordingly, increased histone acetylation at the promoter region has been linked to activate transcription, where the level of acetylation is proportional to the transcription rate [15,24].

Genome-wide chromosome immuno-precipitation (ChIP-seq) studies of *Trypanosoma brucei* and L*eishmania major* found twin peaks of acetylated H4K10 enrichment at every divergent SSR, marking initiation [5,25]. Similarly, in *T. cruzi*, acetylated H3K9/H3K14, H4K10 and methylated H3K4 are enriched at divergent SSR with a bimodular profile demarcating the bidirectional transcription initiation sites [2]. Therefore, the presence of these multiple post-transcriptional modifications of histones has been implicated in the regulation of transcription initiation in kinetoplastids.

Modulation of transcription initiation sites of many organisms is regulated by cooperative interaction of epigenetic modifications of DNA and histone [26,27]. The

presence of 5-methylcytosine within chromatin insures robust transcriptional silencing, as it can directly inhibit transcription factor binding [28] and act as a binding platform for recruiting histone modifying and chromatin remodeling proteins, including the histone deacetylase co-repressor complex [29,30]. Regulation of nucleosome density, histone modification and DNA modification are therefore important mechanisms to enable the transcription machinery to access the DNA.

Similar to the mechanism of epigenetic modifications in other eukaryotes, we predicted that presence of base J at promoter regions alters chromatin structure and changes the accessibility of DNA binding proteins, including RNA polymerase II, thereby regulating transcription initiation and trypanosome gene expression. Therefore, we analyzed structure and abundance of modified and unmodified histones of *T. cruzi* chromatin following the loss of base J. We find that the loss of base J coincides with decreased nucleosome abundance, increased histone H3/H4 acetylation and increased Pol II occupancy at promoter regions. In contrast, no changes in chromatin structure are detected at termination regions in the JBP1dKO, consistent with the absence of defects in transcription termination in these cells. These studies present the first detailed look at the role of epigenetics and chromatin structure in the regulation of Pol II transcription in kinetoplastids. Furthermore, we now provide a mechanism for the novel hyper-modified base J regulating Pol II transcription initiation in *T. cruzi*.

MATERIALS AND METHODS

Growth of cells. Y strain *T. cruzi* wild type and JBP1 dKO epimastigotes were grown in Liver Infusion Tryptose media containing 10% fetal bovine serum as previously described [31]. Epimastigotes from both WT and KOs were harvested at mid log phase by centrifugation for subsequent analysis.

FAIRE analysis. Formaldehyde assisted regulatory elements (FAIRE) analysis was performed as previously described [32]. Briefly, $1x10^9$ *T. cruzi* cells were fixed in 1% formaldehyde for 20 min in LIT media and reaction was terminated by adding 2.5M glycine to a final concentration of 125mM. Fixed cells were then lysed by adding a cell lysis solution containing 10mM EDTA, 50mMTris-HCL, 1% SDS and protease inhibitors. DNA was sonicated with a sonic amplitude sonicator for 10min (30s on and off cycles) in order to obtain chromatin fragments of average 500bp in length. Debris was removed by centrifugation. No cross-linked sample was obtained for each replicate as total DNA control. DNA from both cross-linked and non cross-linked samples was extracted by two consecutive phenol-chloroform extractions and DNA was ethanol precipitated in the presence of 20µg/ml of glycogen after RNA and proteinase K treatment. Each experiment was performed in triplicate and quantification of the FAIRE and total DNA samples were performed by real-time PCR as described below.

ChIP analysis for Histone and Pol II. Chromatin immuno-precipitation (ChIP) was performed as previously described [32]. Briefly, DNA was cross-linked to protein using

formaldehyde as in FAIRE analysis. Sonicated DNA extract was pre cleared using protein A agarose beads and incubated with or without relevant antibodies. Chromatin from 7X 10^7 cells equivalent was used in each immuno-precipitation reaction. Lysate was incubated with commercially prepared anti-histone H3 (AbCam) at a concentration of 2μg per IP reaction or anti – AcH4 (AbCam) at a concentration of 5 μg per IP reaction or anti-AcH3 at a concentration of 5μg per IP reaction overnight at 4^0C. *T. cruzi* RNA Pol II CTD antibodies (a kind gift from Sergio Schenkman) were used at 10μg per IP reaction. Protein-DNA complexes were incubated with protein agarose A beads for 2hrs and washed 3 times using wash buffers containing 0.1% SDS, 1% TritonX-100, 2mM EDTA, 20mM Tris, 500-150mM NaCL and protease inhibitors. DNA was eluted from beads using elution buffer containing 0.1% SDS and 0.1M $NaHCO_3$. Cross-linking was then reversed by adding NaCl to a final concentration of 325mM and incubated at 65°C overnight. DNA was then extracted using phenol-chloroform after RNAse and proteinase K treatments. Each ChIP experiment was performed in triplicate and analyzed by quantitative RT-PCR.

Quantitative RT-PCR. Input and immunoprecipitated DNA was obtained from three biological replicates from WT and KO cells as described above. Quantification was performed on an iCycler with an iQ5 Multicolor Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA). Primer sequences used in the analysis are available upon request. Reaction mixtures contain 5 pmole forward and reverse primer, 2x iQ SYBR Green Super Mix (Bio-Rad Laboratories, Hercules, CA) and 2μl of template cDNA. Standard curves were prepared for each gene using ten folds dilutions of known

quantity (15ng/μl) of WT DNA. Quantifications were calculated using iQ5 optical detection system software. Each sample was normalized to the 24S rRNA values. The average ratios were calculated for each gene using 3 stage and 3 reference sample comparisons.

RESULTS

Base J regulates nucleosome abundance and histone acetylation at divergent Pol II promoter regions. Most of the trypanosome genome is arranged in bi-directional polycistronic transcription units (PTUs) flanked by divergent and convergent strand switch regions. We, and others, have demonstrated that divergent SSRs in *T. cruzi* contain bi-directional promoters [2,4]. Our previous studies demonstrated the presence of base J at the divergent and convergent SSRs and the loss of base J from these regions in JBP1dKO cells lead to increased Pol II transcription with no defects in termination [4]. We concluded that base J specifically regulated chromatin structure at Pol II promoter regions. To investigate this possibility, we first looked at the nucleosome occupancy in SSRs using FAIRE analysis of WT and JBP1dKO cells. FAIRE identifies nucleosome depleted or naked regions such as highly transcribed promoter regions and has previously been utilized to characterize the chromatin landscape of Pol I, II and III loci in trypanosomes [32,33]. When formaldehyde cross-linked and sheared chromatin is extracted with phenol-chloroform, fragments of DNA that were naked or contained loosely associated histones or other proteins are enriched in the aqueous phase. We initially scanned a SSR by qPCR to amplify ~200bp long fragments covering the

divergent SSR as well as regions extending into the adjacent polycistronic unit (Fig. 1A). Our previous studies mapped the approximate location of the divergent promoters in this SSR and demonstrated that the loss of J leads to increased Pol II transcription rate of the adjacent PTUs [34]. We now demonstrate increased FAIRE enrichment in the JBP1dKO, corresponding to increased soluble chromatin, as we progress into the SSRs with a peak approximately in the center and adjacent to the mapped promoters. We then analyzed the center of a three additional divergent SSRs within the *T. cruzi* genome (Fig. 1B and C). We have previously demonstrated that the loss of base J in these SSRs led to increased rate of PTU transcription [4]. Overall, we found that divergent SSRs were significantly more enriched in the aqueous phase in the transcriptionally activated JBP1dKO cell line than WT, consistent with the depletion of nucleosomes at promoter regions following the loss of base J (Fig. 1B). In contrast, there was no change at convergent SSRs between JBP1dKO and WT (Fig. 1C). The low level of soluble chromatin at convergent SSRs compared with divergent promoter regions (20-60 fold less enrichment of convergent regions in WT cells), indicates an increased levels of protein bound to DNA involved in transcriptional termination than regions involved in initiation.

We next performed chromatin immunoprecipitation (ChIP) using formaldehyde cross-linked chromatin and anti histone H3 antibodies followed by qPCR to directly characterize the nucleosome density within various SSRs. As reported previously in trypanosomes, histone H3 content is a good indicator of nucleosome density [32,35,36]. Consistent with FAIRE, we see a significant reduction in H3 occupancy at divergent SSRs of JBP1dKO cells compared to WT and no change at convergent SSRs (Fig. 1D and E). Overall these data suggested a significant change in chromatin structure of Pol II

promoter regions upon loss of base J. However, the loss of base J has no discernable effect on the protein rich chromatin representing transcription termination sites. Since the levels of nucleosomes is similar within divergent and convergent regions of WT cells, the low FAIRE enrichment of divergent termination regions is due to increased non-nucleosomal DNA-bound factors.

Histone acetylation is almost invariably associated with activation of transcription. Previous studies showed an enrichment of acetylated histone H3 and H4 at Pol II transcriptional start sites in kinetoplastids [2,5]. Furthermore, acetylated histones H3 and H4 are components of active promoters in the protozoan parasite *T. gondii* [37,38]. In order to investigate if base J affects the acetylation of core histones within promoter regions, we performed a ChIP assay using anti-H3Ac and anti-H4Ac antibodies in WT and JBP1dKO cells. Although nucleosome abundance is reduced at divergent SSRs, the fraction of acetylated H3 and H4 histones has significantly increased in the JBP1 dKO cell line (Fig. 1F and H). Convergent termination regions contain lower levels of acetylated histones than divergent promoter regions with no change in levels upon the loss of base J (Fig. 1G and I). Lower levels of acetylated H3 and H4 in termination regions compared to divergent SSRs in *T. cruzi* has been characterized previously [2].

Together with the FAIRE and histone ChIP experiments described above, we conclude that the loss of base J from promoter regions of Pol II transcribed gene clusters lead to decreased nucleosome abundance and increased H3 and H4 acetylation of the remaining nucleosomes. Furthermore, acetylated nucleosomes were preferentially localized at divergent promoter regions compared with termination sites. These results are consistent with our earlier conclusion that the modulation of chromatin structure by

base J regulates Pol II transcription initiation and genome-wide changes in gene expression [4]. The finding that loss of base J does not lead to any detectable changes in chromatin structure at convergent termination sites is also consistent with our inability to detect defects in Pol II transcription termination in the JBP1dKO [4].

Chromatin changes at the SL RNA promoter/locus upon the loss of base J. In trypanosomes, maturation of polycistronic mRNAs requires the addition of a 39-nucleotide spliced leader (SL) RNA onto the 5'-end of every mRNA [39]. In order to provide the large amount of SL RNA required for trans-splicing, the trypanosome has up to 200 SL RNA transcription units that are transcribed by Pol II. The SL RNA gene contains the only defined Pol II-dependent promoter in trypanosomes [40]. In fact, all knowledge of proteins and DNA sequences essential for Pol II-dependent transcription in kinetoplastids comes from SL RNA gene studies. Therefore, to fully understand the role of base J in regulating Pol II transcription, we analyzed the SL RNA gene array. Figure 2A depicts a SL RNA gene and promoter and associated regions within a representative array. A non-transcribed spacer region (nts) extends ~1 kb downstream (and upstream) of the promoter and CZAR represent a non-transcribed retrotransposable element which has integrated exclusively within the SL RNA array ~1 kb downstream of the SL gene. Using anti-J IP, we have localized base J within the SL gene array of WT *T. cruzi* (Fig. 2 B). As expected, base J was absent from these regions in the JBP1dKO cell line (Fig. 3B). This loss of base J leads to a 2.33-fold increase in rate of Pol II transcription of the SL RNA gene (6). In order to characterize changes in chromatin environment of a defined Pol II

promoter resulting from the loss of base J and explain the increased SL transcription, we performed FAIRE and ChIP analyses of the SL array as described above. We found an increase in abundance of the soluble fraction of chromatin at the SL promoter and non-transcribed spacer region of JBP1dKO cells compared to WT. Consistent with inactive regions of chromatin, there was approximately 10 fold less soluble chromatin present at the CZAR repeats of WT cells compared to the SL promoter and no significant change in the JBP1 dKO cells (Fig. 2C). Consistent with FAIRE, we found approximately 2-fold less histone H3 present at the SL promoter and non-transcribed spacer regions (Fig. 2D). The initial low abundance of histone H3 at the SL promoter/RNA gene in WT cells is consistent with previous studies in Leishmania, *T. brucei* and *T. cruzi* [2,35,41]. We also detected a significant increase in the acetylated histone H3 and H4 abundance at the SL gene and non-transcribed spacer region in the JBP1dKO cell line (Fig. 2E and F). More significant changes in H3 and H4 acetylation are evident at the SL gene/promoter region than ~1 kb away in the nts. Taken together, we can conclude that loss of base J leads to changes in chromatin structure and increased H3 and H4 acetylation at the defined SL Pol II promoter as well as undefined promoters within SSRs between divergent gene clusters throughout the *T. cruzi* genome.

Increased RNA polymerase II occupancy following the loss of base J. Eukaryotic gene expression begins with recruitment of the transcription machinery to a gene promoter and formation of a preinitiation complex composed of Pol II and general transcription factors [42]. To evaluate the impact of base J on the formation of the pre-initiation complex *in vivo*, we investigated the changes in Pol II promoter occupancy

upon depletion of JBP1 by ChIP using specific antibodies to the *T. cruzi* Pol II unique carboxy-terminal domain [43]. Initially, we analyzed Pol II occupancy within a divergent SSR using ChIP/qPCR as previously performed with FAIRE (Fig. 3A). We found that the highest levels of Pol II within the SSR flank the previously mapped bidirectional promoter regions, similar to the nucleosome depletion profile (Fig. 1A). Loss of base J in the JBP1 dKO lead to a 2-4 fold increase in Pol II occupancy at these promoter regions within the divergent SSR (Fig. 3A), as well as a 3-6 fold increase within three additional SSRs (Fig. 3B).

We then extended our analysis to determine the Pol II accumulation at the defined Pol II-dependent SL promoter in *T. cruzi*. As expected, in both WT and JBP1 dKO cell lines we see higher levels of Pol II at the promoter than the SL gene and almost complete absence of the polymerase at the non-transcribed spacer and CZAR retrotransposon. However, we detected a ~3-fold increased accumulation of Pol II at the SL promoter and gene in the JBP1dKO (Fig. 3C). Overall, the changes in Pol II occupancy correlated with the increased rate of nascent mRNA formation within the adjacent polycistronic units and SL RNA gene following the loss of base J [4]. These results suggest that the base J regulates chromatin structure and the formation of the Pol II pre-initiation complex.

DISCUSSION

In eukaryotes, transcription initiation is a key regulatory point in controlling the level of gene expression. Transcription initiation involves a series of events including protein modification (i.e. histone), chromatin remodeling, binding of specific proteins to DNA (i.e. activators), and recruitment of Pol II polymerase. However, previous studies of trypanosome gene expression indicated that genes are not activated at the transcriptional level [3]. The arrangement of functionally unrelated genes in polycistronic units and the absence of classic Pol II promoters led to the belief that transcription initiation was not a key factor in regulating trypanosome mRNA production. Localization of specific modified and variant histones at SSRs in trypanosomes [2,5] and visible changes in chromatin architecture during *T. cruzi* development [44] suggests a role for epigenetics in control of gene expression. However, no direct evidence has been provided. We recently described a novel epigenetic modification of DNA, consisting of the hydroxylation and glucosylation of the thymine base of DNA, called base J, at Pol II promoter regions and which, upon removal, leads to an increased rate of mRNA transcription. Accordingly, we now demonstrate that the loss of base J at Poll II promoter regions leads to distinct changes in chromatin structure/modification and increased occupancy of Poll II polymerase. These studies provide the first evidence that Poll II transcription initiation, and key regulatory mechanisms including his tone modification, chromatin remodeling and recruitment of polymerase, are regulated in an epigenetic fashion in trypanosomes.

Chromatin structure needs to be dynamically modulated in order to allow transcription and regulate cellular function. This is especially true for single celled

organisms that need to rapidly adapt to changing environmental growth conditions (i.e. host environment) during its lifecycle, reflected in changes in gene expression. Chromatin structure has been shown to regulate all stages of transcription initiation in eukaryotes, including the binding of activators upstream of the core promoter, recruitment of complexes that facilitate the binding of general transcription factors and recruitment of Pol II and formation of the preinitiation complex. However, very little is understood regarding the regulation of Pol II initiation, including the role of chromatin, in trypanosomes. Here, we identify that even in these early-divergent eukaryotes, DNA modification, histone modification and nucleosome remodeling/abundance work together to regulate chromatin structure and Pol II transcription.

According to the histone code hypothesis, modification of histone serves as a recognition platform for binding of factors such as chromatin remodeling proteins [15]. It has been demonstrated in yeast that covalent modification of histones may precede their dissociation from DNA [45,46]. Furthermore, histone acetylation serves as a mark for actively transcribed chromatin where changes in the net charge of the histone due to the acetyl group presumably leads to unfolding of the nucleosomal fiber, rendering regulatory sites more accessible [47]. We demonstrate here that the loss of base J is followed by the reduction of nucleosomes and increased acetylation of histone H3 and H4 at Pol II promoter regions. Regulation of nucleosome density and histone modification are therefore highly conserved mechanisms to enable the transcription machinery to access the DNA. Furthermore, it follows that in the case of trypanosomes, the acetylation of promoter-associated nucleosomes occurs prior to their disassembly. We propose that the loss of base J leads to increased accessibility of DNA binding proteins

including histone modifying factors. The modification of histone tails, including acetylation of H4/H3 at Pol II promoters eventually leads to decreased abundance of nucleosomes and increased recruitment of Pol II resulting in increased rate of transcription initiation. In contrast, the loss of J at tightly closed chromatin, such as regions involved in Pol II transcription termination, fails to affect the recruitment of histone modifying factors. FAIRE indicates increased localization/binding of non-nucleosomal proteins within the convergent SSRs. Therefore, the removal of base J in these regions had no measurable affect on chromatin structure or the ability to properly terminate Pol II transcription [4].

While we clearly demonstrate epigenetic mechanisms that regulate Pol II occupancy and transcription initiation in trypanosomes, it is difficult to propose a model in which transcription of the PTUs are tightly regulated rather than constitutively transcribed. Unlike prokaryotic operons, there is no clear functional relationship among genes within a PTU in trypanosomes. Therefore, there is no obvious benefit to the parasite to differentially regulate transcription of individual PTUs. However, we have previously shown that the loss of base J at Pol II promoter regions and increased transcription of PTUs lead to significant up- and down- regulation of gene expression in *T. cruzi* [34]. Apparently, the regulation of transcription via base J works in conjunction with established post-transcriptional regulatory mechanisms. Increased transcription throughout the entire PTU still results in reduced steady-state concentrations of certain mRNAs. We have demonstrated in *T. cruzi* that the loss of JBP2 function did not lead to the loss of J from every PTU Pol II promoter region and the loss of each JBP enzyme led to differential changes in gene expression [34]. Therefore, a combination of differential

regulation of JBP function and levels of J at PTU promoters and post-transcriptional regulatory mechanisms may provide a mechanism to fine-tune trypanosome gene expression, and to potentiate genes for further or future activation and/or inactivation.

In summary, we have provided the first direct evidence for epigenetic regulation of Pol II transcription initiation via DNA modification and chromatin structure in early branching trypanosomes. Future studies are required to identify how the unique hyper-modified DNA base J regulates chromatin structure and nucleosome modifications and why these organisms need to regulate transcription initiation at large polygene clusters throughout the genome.

ACKNOWLEDGMENTS

The authors would like to thank Sergio Schenkman for providing the Pol II antisera and Luisa Figueiredo for assistance with FAIRE analysis. We also thank members of the Hajduk lab for helpful discussions; Rudo Kieft, Laura Cliffe, Steve Hajduk and Piet Borst for critical reading of the manuscript. This work was supported by grant number RO1AI063523 from the National Institute of Allergy and Infectious Disease (NIAID) of the National Institutes of Health (NIH) and by a pre-doctoral grant 09PRE2290187 from the American Heart Foundation (AHF).

REFERENCES

1. Tyler KM, Engman DM (2001) The life cycle of *Trypanosoma cruzi* revisited. Int J Parasitol 31: 472-481.
2. Respuela P, Ferella M, Rada-Iglesias A, Aslund L (2008) Histone acetylation and methylation at sites initiating divergent polycistronic transcription in *Trypanosoma cruzi.* J Biol Chem 283: 665-673.
3. Clayton C, Shapira M (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 156: 93-101.
4. Ekanayake DK, Minning T, Weatherly B, Gunasekera K, Nilsson D, et al. (2011) Epigenetic regulation of transcription and virulence in *Trypanosoma cruzi* by O-linked thymine glucosylation of DNA. Mol Cell Biol 31: 1690-1700.
5. 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 Dev 23: 1063-1076.
6. Cliffe LJ, Siegel TN, Marshall M, Cross GA, 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 Res 38: 3923-3935.
7. van Leeuwen F, Kieft R, Cross M, Borst P (1998) Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. Mol Cell Biol 18: 5643-5651.

8. Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62: 235-251.

9. Sabatini R, Cliffe L, Vainio S, Borst P (2009) Enzymatic Formation of the Hypermodified DNA Base J (β-D-Glucopyranosyloxymethyluracil) DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.

10. Vainio S, Genest PA, ter Riet B, van Luenen H, Borst P (2008) Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. Mol Biochem Parasitol 164: 157-161.

11. 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 Res 37: 1452-1462.

12. Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, et al. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res 35: 2107-2115.

13. Borst P, van Leeuwen F (1997) beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. Mol Biochem Parasitol 90: 1-8.

14. Hager GL, ., McNally JG, Misteli T (2009) Transcription dynamics. Mol Cell 35: 741-753.

15. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707-719.

16. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 36: 900–905.

17. Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL (2004) Global nucleosome occupancy in yeast. Genome Biol 5: R62.

18. Reinke H, Hörz W (2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell: 1599-1607.

19. Erkina TY, Erkine AM (2006) Displacement of Histones at Promoters of Saccharomyces cerevisiae Heat Shock Genes Is Differentially Associated with Histone H3 Acetylation. Mol Cell Biol 26: 7587-7600.

20. Schermer UJ, Korber P, Hörz W (2005) Histones are incorporated in trans during reassembly of the yeast PHO5 promoter. Mol Cell 19: 279-285.

21. El-Osta A, Wolffe AP (2000) DNA methylation and histone deacetylation in the control of gene expression: basic biochemistry to human development and disease. Gene Expression 9, 63–75. Gene Expression 9, 63–75 9: 63-75.

22. Ura K, Kurumizaka H, Dimitrov S, Almouzni G, Wolffe AP (1997) Histone acetylation: influence on transcription, nucleosome mobility and positioning, and linker histone-dependent transcriptional repression. EMBO J 16: 2096-2107.

23. Mizzen CA, Allis CD (1998) Linking histone acetylation to transcriptional regulation. Cell Mol Life Sci 54: 6-20.

24. Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, et al. (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 22: 517-527.

25. Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ (2009) Histone acetylations mark origins of polycistronic transcription in Leishmania major. BMC Genomics 8: 152-158.

26. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31: 89-97.

27. Quina AS, Buschbeck M, Di Croce L (2006) Chromatin structure and epigenetics. Biochem Pharmacol 72: 1563-1569.

28. Watt F, Molloy PL (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 2: 1136-1143.

29. Wade PA, Jones PL, Vermaak D, Wolffe AP (1998) Amultiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. Curr Biol 7: 843-846.

30. Nan, Ng HH, Johnson CA, Laherty CD, Turner BM, et al. (1988) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-389.

31. Camargo EP (1964) Growth and Differentiation in *Trypanosoma cruzi.* I. Origin of Metacyclic Trypanosomes in Liquid Media. Rev Inst Med Trop Sao Paulo 6: 93-100.

32. Figueiredo LM, Cross GAM (2010) Nucleosomes Are Depleted at the VSG Expression Site Transcribed by RNA Polymerase I in African Trypanosomes. Eukaryotic Cell 9: 148-154.

33. Nagy PL, Cleary ML, Brown PO, Lieb JD (2003) Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. Proc Natl Acad Sci 100: 6364-6369.

34. Ekanayake DK, Cipriano MJ, Sabatini R (2007) Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite *Trypanosoma cruzi.* Nucleic Acids Res 35: 6367-6377.

35. Stanne TM, Rudenko G (2010) Active VSG Expression Sites in *Trypanosoma brucei* Are Depleted of Nucleosomes. Eukaryotic Cell 9: 136-147.

36. Lowell JE, Cross GA (2004) A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. J Cell Sci: 5937-5947.

37. Saksouk N, Bhatti MM, S. K, Smith AT, Musset K, et al. (2005) Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma gondii*. Mol Cell Biol 25: 10301-10314.

38. Gissot M, Kelly KA, Ajioka JW, Greally JM, Kim K (2007) Epigenomic modifications predict active promoters and gene structure in Toxoplasma gondii. PLoS Pathog 3: 77.

39. Liang XL, Haritan A, Uliel S, Michaeli S (2003) trans and cis Splicing in Trypanosomatids: Mechanism, Factors, and Regulation. Eukaryotic Cell 2: 830-840.

40. Nunes LR, Carvalho MR, Shakarian AM, Buck GA (1997) The transcription promoter of the spliced leader gene from *Trypanosoma cruzi*. Gene 188: 157-168.

41. Hitchcock RA, Thomas S, Campbell DA, N.A. S (2007) The promoter and transcribed regions of the Leishmania tarentolae spliced leader RNA gene array are devoid of nucleosomes. BMC Microbiol 7: 44.

42. Roeder RG (2005) Transcriptional regulation and the role of diverse coactivator in animal cells. FEBS Letters 579: 909-915.

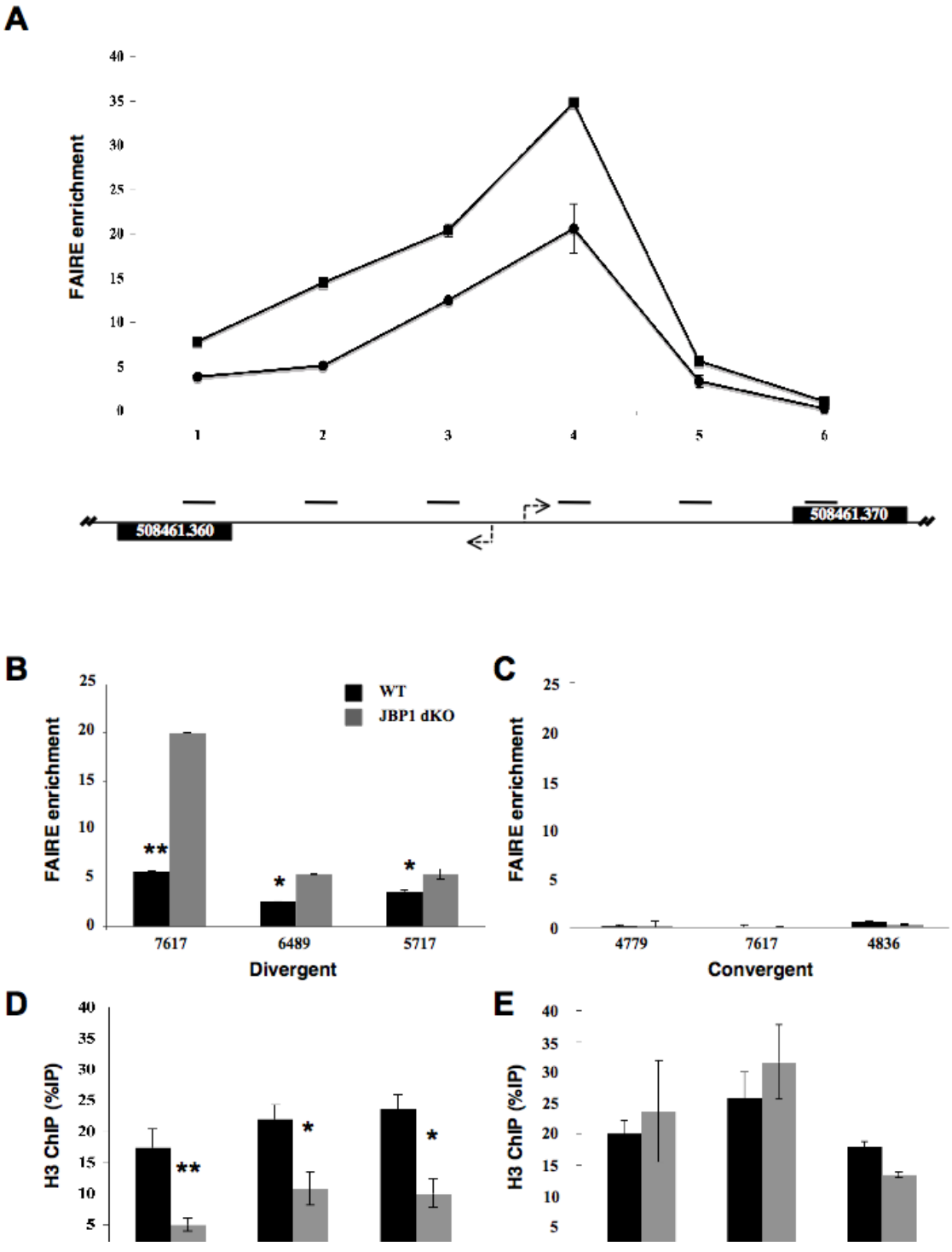
43. Dossin FM, Schenkman S (2005) Actively Transcribing RNA Polymerase II Concentrates on Spliced Leader Genes in the Nucleus of *Trypanosoma cruzi.* Eukaryotic Cell 5: 960-970.

44. Elias MC, Marques-Porto R, Freymuller E, Schenkman S (2001) Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. Mol Biochem Parasitol 112: 79-90.

45. Zhao J, Herrera-Diaz J, Gross DS (2005) Domain-wide displacement of histones by activated heat shock factor occurs independently of Swi/Snf and is not correlated with RNA polymerase II density. Mol Cell Biol 25: 8985–8999

46. Reinke H, Horz W (2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell 11: 1599-1600.

47. Eberharter A, Becker PB (2002) Histone acetylation: a switch between repressive and permissive chromatin. EMBO Rep 3: 224–229.


A
FAIRE enrichment
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
508461.360
508461.370
B
FAIRE enrichment
WT
JBP1 dKO
**
*
*
7617
6489
5717
Divergent
C
FAIRE enrichment
4779
7617
4836
Convergent
D
H3 ChIP (%IP)
**
*
*
E
H3 ChIP (%IP)

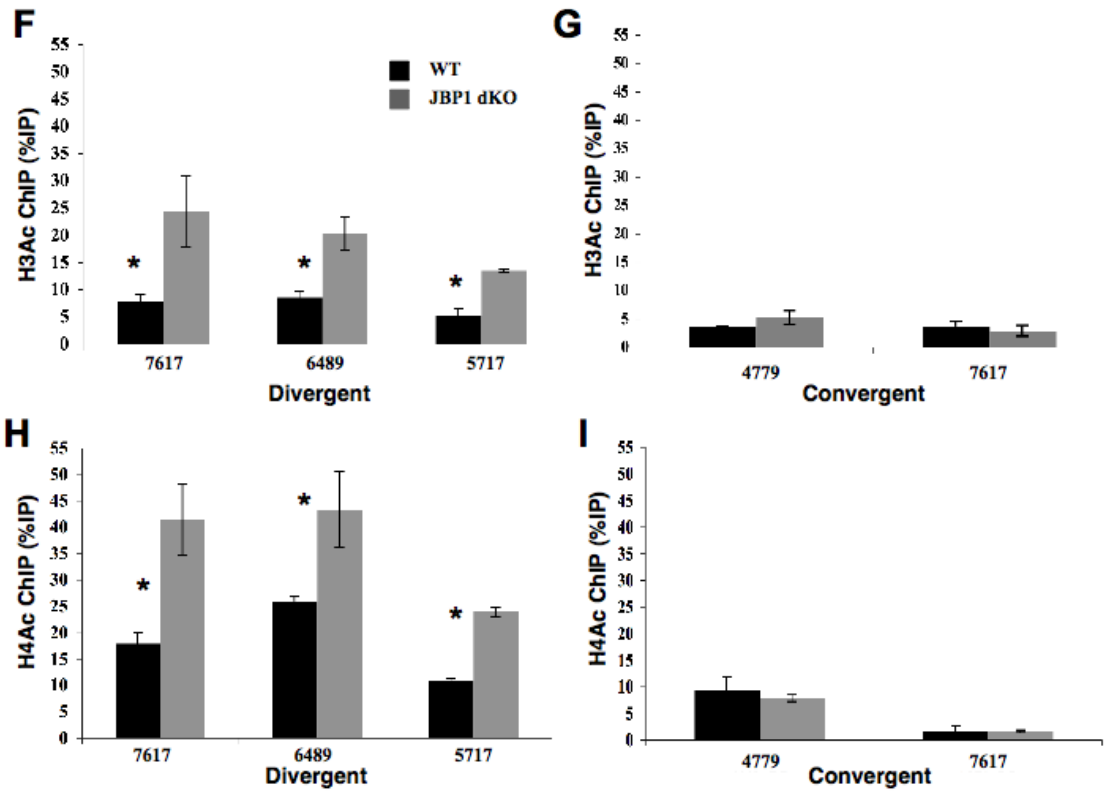


Figure 4.1: FAIRE and histone IP analysis shows that divergent SSRs are depleted of nucleosomes and undergo increased histone acetylation upon the loss of base J. (**A**) FAIRE analysis spanning a divergent SSR and adjacent polycistronic unit of WT and JBP1dKO cells. Sonicated DNA fragments (200-500bp) were quantified by real time PCR. FAIRE enrichment is shown as the fold difference between the amounts of DNA quantified in the aqueous phase in the cross-linked samples relative to the non-cross-linked sample normalized to the 24S gene. Solid circles; WT trypanosomes and solid boxes; JBP1dKO cells. The schematic represents the 7739 divergent SSR where horizontal lines indicate the approximate positions of amplicons. The arrowheads show the approximate location of bidirectional promoter sites as determined previously (6). (**B-**

C) FAIRE analysis of three divergent SSRs (**B**), and three convergent SSRs (**C**) of WT and JBP1dKO cells. Black horizontal bars, WT; grey horizontal bars, JBP1dKO cells. The numbers refer to DNA sequence contigs that correspond to the SSR in the *T. cruzi* genome analyzed. However, contig 7617 contains the entire PTU and the flanking divergent and convergent SSR. (**D-E**) Distribution of histone H3 over three divergent SSR (**D**) and three convergent SSRs (**E**) as determined by ChIP using histone H3 antibody (no antibody as negative control) in WT and the JBP1dKO. (**F-I**) ChIP using anti-acetylated histone H3 and anti-acetylated histone H4 for three divergent SSRs (F and H) and two convergent SSRs (**G** and **I**) of WT and the JBP1dKO. DNA was quantitated by real-time PCR and compared to the total input material. No antibody control was performed for all experiments. Results show the average from three independent experiments with the standard deviations indicated with error bars (**$P<0.001$, *$P<0.05$).

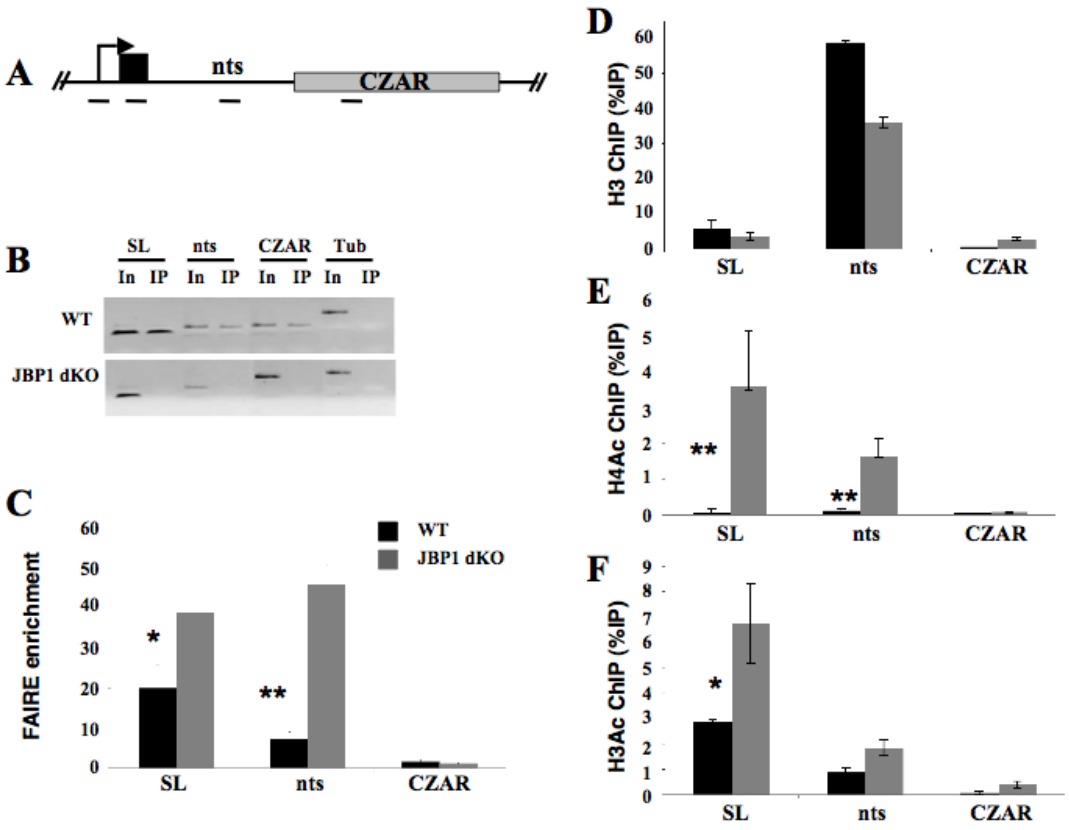


Figure 4. 2. **The spliced leader transcription unit becomes more organized in an open chromatin structure upon the loss of base J.** (**A**) Schematic of genomic region containing the SL RNA transcription unit. SL promoter is indicated by the arrow, the SL gene by the black box, and the retrotransposon element (CZAR) by the grey box. Regions analyzed by real-time PCR are indicated with horizontal lines. **(B**) Anti-base J IP/PCR analysis of the SL transcription unit. Tubulin is used as negative control. In for input DNA and IP for immuno-precipitated DNA (**C**) FAIRE analysis of the SL gene array for WT (black bars) and JBP1dKO (grey bars). (**D-F**) Histone ChIP analysis. H3 (**D**), actylated H4 (**E**) and acetylated H3 (**F**) immunoprecipitations in WT (black bars) and JBP1dKO cell lines (grey bars). DNA was quantified by real time PCR and compared to the total input material. The data shown are the average of three independent experiment

with the standard deviations indicated with error bars (**P<0.001, *P<0.05). No antibody was used as negative control.

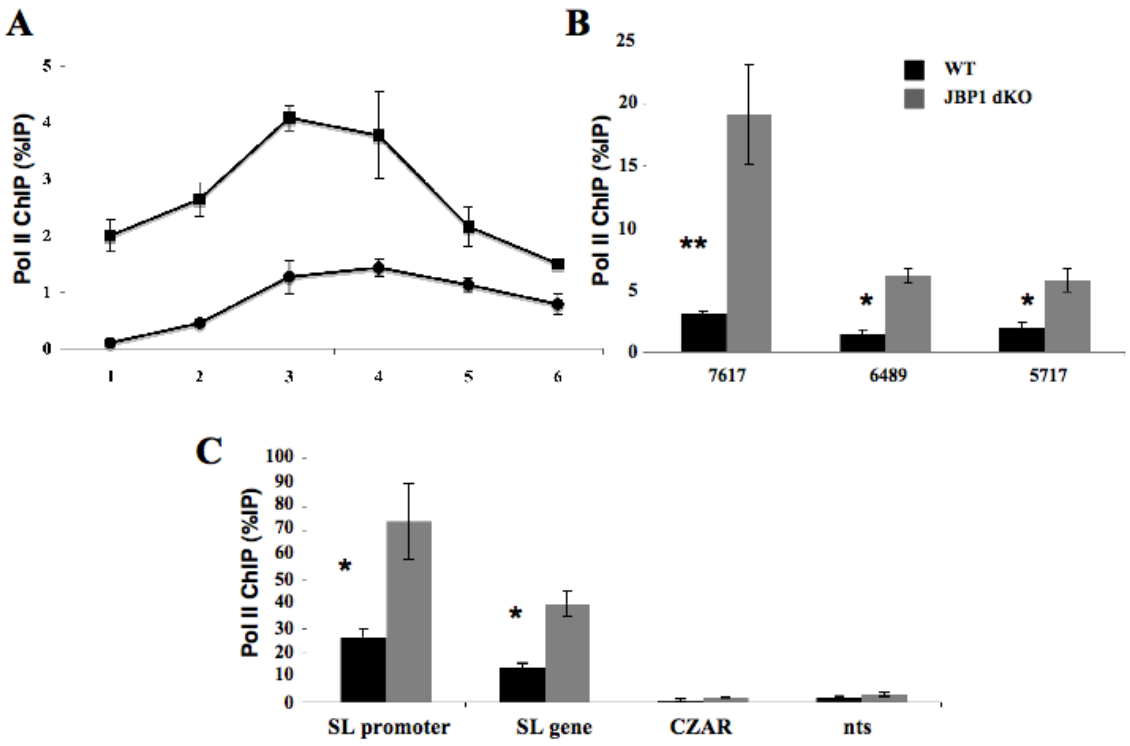


Figure 4.3: The loss of base J at Pol II promoters leads to increased Pol II occupancy. Increased RNA Pol II occupancy at Pol II promoters of the JBP1 dKO. (**A**) Pol II occupancy within a divergent SSR of WT and JBP1dKO cells as determined by ChIP using anti Pol II antibody. Solid circles, WT; solid boxes, JBP1 dKO. (**B**) Pol II accumulation at three different divergent SSR using anti-Pol II ChIP/qPCR. Black boxes, WT; grey boxes, JBP1dKO. (**C**) Distribution of Pol II across the SL gene array. Pol II ChIP/qPCR analysis of the indicated regions within the SL array. All the data are normalized to 24S gene and no antibody was used as negative control. Data represent three independent experiments. Standard deviations are indicates with error bars (**P<0.001, *P<0.05).

CHAPTER 5

CONCLUSIONS, DISCUSSION, AND FUTURE DIRECTIONS

Base J is a novel DNA modification first detected in the African trypanosome *T. brucei.* Telomeric localization of base J in *T. brucei* suggested a role of the base in regulation of VSG gene expression [1]. However, no clear function of base J has been previously recognized.

In this study, we used *T. cruzi* as a model system to characterize the role of base J in kinetoplastids. We demonstrated that base J localizes to regions of the *T. cruzi* genome that are important for Pol II transcriptional regulation. This observation was surprising in view of the prevailing dogma that gene expression may not be regulated at the level of transcription initiation in kinetoplastids. Accordingly, it was unclear why a potential epigenetic regulatory mark might be enriched at putative Pol II promoter regions.

We have now demonstrated that depletion of base J from transcription initiation sites leads to increased Pol II occupancy at these regions, and results in an increased rate of Pol II-dependent transcription and corresponding changes in gene expression. We thus conclude that RNA Pol II-mediated transcription initiation is regulated by base J. Our analysis of base J function in *T. cruzi* therefore provides the first example of epigenetic regulation of Pol II-dependent transcription initiation in trypanosomes.

Mechanism of base J in regulation of gene expression

In order to determine the mechanism by which base J regulates transcription, we characterized changes in chromatin structure and Pol II recruitment to promoter regions following loss of base J. Base J presents a large glucose moiety residing in the major groove of DNA. Thus, two non-exclusive potential mechanisms for transcriptional regulation are that base J physically interferes with recruitment of the transcription machinery (*i.e.* Pol II), or exerts its influence indirectly at the level of chromatin structure. In support of the latter hypothesis, we observed significant changes in chromatin architecture where loss of base J corresponds with decreased nucleosome occupancy and increased levels of histone acetylation.

We propose that loss of base J leads to recruitment of chromatin remodeling factors and histone modifying enzymes that facilitate access of the transcription machinery to promoters. Accordingly, we speculate that the more open chromatin environment which accompanies base J depletion increases the accessibility of promoter regions to Pol II recruitment, and thereby promotes the formation of the pre-initiation complex. This, in turn, leads to increased basal transcription rates and corresponding changes in gene expression.

An alternative possibility is that base J itself directly interferes with transcription by physically impeding the association of Pol II and associated transcription factors. This occurrence presents a "chicken-and-egg" phenomenon of sorts, in that it is difficult to understand the causal order of events driving the mechanism of transcription regulation by base J. However, while there is clear evidence that base J depletion leads to changes in local chromatin environments, there is no evidence that base J can directly interfere with

transcription through physical inhibition of Pol II association. Future experiments may address this possibility by performing *in vitro* DNA-protein binding assays using recombinant Pol II and J modified DNA as substrate. Regardless of the specific mechanisms involved, these studies present the first direct evidence for epigenetic regulation of Pol II transcription initiation via DNA modification and chromatin structure in kinetoplastids.

Biological significance of J regulation of gene expression

We have clearly demonstrated that loss of base J at Pol II promoter regions and increased transcription of PTUs leads to significant changes in gene expression in *T. cruzi*. Trypanosome PTUs are not arranged in the manner of prokaryotic operons, where functionally related genes are expressed simultaneously. Accordingly, it is difficult to propose a model that accounts for regulation of transcription in trypanosomal PTUs. However, different levels of J between the JBP1 and JBP2 knockouts led to differences in gene expression. This observation suggests that the genomic occupancy of base J may correlate with differential expression of multiple genes, including those encoding surface glycoproteins, allowing for their selective expression throughout the *T. cruzi* lifecycle.

Further studies are needed to examine the biological significance of base J-dependent regulation of expression of PTUs at specific stages of the *T. cruzi* lifecycle. However, genome-wide expression profiling has revealed that base J depletion clearly impacts surface protein gene expression along with other Pol II transcribed genes. In addition, we have also detected changes in parasite phenotype *in vitro.* JBP knockout *T. cruzi* exhibit significant defects in mammalian cell egress and invasion efficiency. The

direct link between these specific phenotypes and J regulation remains unclear. Furthermore, the phenotype of JBP knockout *T. cruzi* remains to be evaluated *in vivo*. Thus, in future studies, it will be interesting to examine tissue-specific parasite burden and pathology in murine models to determine if and how gene expression changes impact *in vivo* infectivity. It might also interesting to look at the changes in magnitude and frequencies of the immune response in murine models using J-deficient strains of *T. cruzi.* The analysis of J function *in vivo* may also uncover additional functions of the modified base in trypanosome biology, as discussed in the sections below.

T. cruzi presents multiple targets for innate and cell mediated recognition, and elicits a strong immune response during infection. Effector mechanisms include induction of antibody expression, activation of phagocytic cells, and the CD4+ T and CD8+ T cell responses. *T. cruzi* releases proteins that are presented to host cell surface class 1 MHCs, which are recognized by CD8+ T cells in order to destroy *T. cruzi* infected cells. CD8+ T cells also recognize GPI anchored surface epitopes, particularly trans-sialidases [2]. Strain dependent differences have been observed in the frequency with which CD8+ T cells recognize dominant *T. cruzi* epitopes [3]. Altered gene expression in *T. cruzi* cells with reduced J levels affected the expression of thousands of genes, including trans-sialidases. An array of available MHC class I tetramers generated against different TS epitope sequences [3] can be used to determine whether changes of gene expression JBP knockout *T. cruzi* has direct effects on the immune response. However, isolating a single event or factor to study during mammalian infection is a difficult task, since many factors are interconnected during the disease process.

Other possible functions of base J

In addition to its enrichment at Pol II promoters, base J was found to localize to convergent SSRs involved in transcription termination. This genome-internal base J accounts for only ~5-10% of total base J in the trypanosme genome [4]. A more substantial fraction of base J in *T. cruzi* is localized to the telomeric region of the genome. Subtelomeric localization of base J is exceptional from that found at SSRs, in which base J localizes to surface protein genes and retrotransposons within PTUs. We speculate that base J may have additional functions in regulating Pol II transcription termination, DNA rearrangements, retrotransposition and telomeric repeat stability in trypanosomes. However, the localization of base J in each of these regions will need to be investigated in greater detail to understand its respective roles.

Pol II termination

Similar to transcription initiation sites, there are no known sequences determining transcription termination in *T. cruzi*. The transcription of a poly-U stretch typically serves as a universal pause signal for eukaryotic RNAPs [5]. Pol II transcribed spliced leader (SL) gene transcription in trypanosomes is terminated at a T-rich region downstream of the SL gene. However, Pol II transcription of protein coding genes is not terminated at T-rich regions [6]. Base J is localized to transcription termination regions, suggesting its potential role in proper termination. In support of this, a recent finding of Piet Borst's lab showed that loss of base J at transcription termination sites in *Leishmania spp* led to massive transcriptional read-through without significant changes in transcription initiation (Piet Borst, unpublished data). However, in *T. cruzi*, we observe that loss of

base J leads to massive changes in transcription initiation with no detectable defect in transcription termination. The viability of JBP1 and JBP2 null mutants of *T. brucei* without detectable base J indicates that base J may not be absolutely essential in regulating global Pol II transcription initiation or termination in this organism. One potential explanation for these conflicting observations is that there may be differences among the various organisms in the function of base J. Transcription termination may be tightly regulated in *T. cruzi,* with many other factors being involved, and the absence of base J may have minimal effect in proper termination. According to our FAIRE and H3 ChIP analysis, chromatin structure is tightly regulated at transcription termination sites in T. cruzi by non-nucleosomal factors and loss of base J at those regions has a minimal effect. However, chromatin structure at transcription termination sites of *Leishmania spp* is significantly changed upon loss of base J, indicating that there may be differences in the chromatin environments of transcription termination sites in these two organisms (Sabatini unpublished data). Thus, it would be interesting to examine the extent to which functional differences in the role of base J are observed in different kinetoplastids.

Recombination

One exciting implication of subtelomeric base J localization is that base J might influence the generation of antigenic diversity in trypanosomes through control of subtelomeric gene rearrangement and recombination [7,8]. In many parasitic protozoans, genes encoding surface proteins reside near telomeres. It has previously been suggested that telomeric regions represent the most favorable region for homologous and ectopic recombination [7]. Hence, the positioning of surface protein genes in trypanosome

subtelomeres may facilitate the creation of new surface protein gene variants in a short time span.

In *T. brucei*, every chromosome end consists of an expression site for VSGs, with the exception of mini-chromosomes. Monopolization of one active VSG for the process of antigenic variation is a distinct phenomenon in this organization. VSG expression occurs either through *in situ* activation of a silent VSG gene by a DNA rearrangement that involves insertion of a VSG from a silent expression site, or by insertion of an internal silent copy of a VSG into an active expression site. These DNA rearrangements are thought to be facilitated by the presence of homologous repeat sequences between expression sites and mini-chromosomes, as well as by sequence similarities between the VSG expression sites. Base J was identified in 50 bp repeats of active and silent expression sites and 70 bp repeats and VSG of silent expression sites [9]. Analysis of J null blood stream forms of *T. brucei* shows an upregulation of VSG DNA rearrangement events, suggesting a function for base J in regulating recombination and antigenic variation (Sabatini unpublished data).

DNA methylation has long been recognized as a repressor of telomeric recombination [10]. Loss of this epigenetic mark in telomeres and subtelomeres is linked to abnormally high telomeric and subtelomeric recombination frequency [11]. Possibly, base J may play a similar role in kinetoplastids. Although there is no experimental evidence for DNA rearrangements between surface protein encoding genes of *T. cruzi,* it is possible that base J localization may facilitate such events to promote antigenic diversity. Further studies will thus be needed to evaluate the consequences of base J and surface protein gene co localization at *T. cruzi* subtelomeres.

Retrotransposition

Retrotransposons are mobile genetic elements that can threaten the structure and regulated expression of the host genome. Transposition can disrupt genes by insertional mutagenesis, and recombination between non-allelic repeats can cause translocations and other unfavorable gene rearrangements. Mobilization of retroelements occurs by target-primed reverse transcription, through which RNA encoded by the retroelement is formed into a DNA copy that is subsequently integrated at a new site in the genome [10]. Some organisms appear to have evolved effective nuclear host defense systems that shield their genomes from transposable elements. *De novo* methylation of retrotransposons and RNAi-induced silencing are two extensively studied mechanisms by which retrotransposon movement is suppressed [12].

Retrotransposons that contain long terminal repeats are called LTR retrotransposons, and retrotransposons that lack LTRs include non-LTR retrotransposons such as LINEs. Both classes of retrotrotransposons have been identified in trypanosomes [13]. *T. brucei* siRNAs are largely complementary to retrotransposons, and deletion of the Argonaute family protein TbAGO1 involved in RNAi leads to increased copy numbers of the RIME non-LTR element, and possibly others [14,15]. A large fraction of m5C in *T. brucei* also co localizes with retrotransposons, suggesting that it may serve a function in suppressing retrotransposon activity [16]. Alternatively, the relatively benign behavior of retrotransposons in many organisms may be due to the inactivation of retrotransposable elements by direct mutation. This may be the case in *L. major*, where the majority of retrotransposable elements have been mutated, and only nonfunctional remnants can be detected [17].

The *T. cruzi* genome is enriched for retroelements such as L1Tc, VIPER, SIRE, DIRE, and CZAR [18]. L1Tc is the best represented LINE of *T. cruzi,* in which several functional copies exist. This retroelement encodes all of the machinery involved in its own retrotransposition, has been found in most if not all chromosomes and its elements are actively transcribed in all three *T. cruzi* life stages [19]. RNAi appears to be absent from *T. cruzi*. We have shown that subtelomeric members of the VIPER, DIRE and L1Tc gene families contain base J. Since some retrotransposable elements have been shown to harbor their own Pol II dependent promoters [20], base J may have a function in regulating these promoters, much as it appears to regulate other putative Pol II promoters within *T. cruzi*. Thus, the presence of base J at subtelomeric retroelements may reflect a specific role for base J in regulating subtelomeric retrotransposition. However, the internal counterparts of these subtelomeric retrotransposons lack detectable base J modification, and it is not clear how their transcription may be regulated. Accordingly, it would be interesting to examine whether internal retroelements lack functional promoters or retrotransposition activity, and to compare this with the functionality of subtelomeric retroelements with which base J associates.

Telomeric repeat stability

Telomeres play important roles in maintaining eukaryotic genome integrity, and are crucial for controlling cell division and longevity. With successive rounds of replication, chromosomes gradually lose terminal DNA repeat sequences at telomeres. Telomere length is a crucial regulator of replicative senescence. Thus, eukaryotes have multiple means to maintain telomere length and preserve genome integrity [21].

A major fraction of base J is localized to the telomeric repeats in all kinetoplastids. Roughly 50 % of total genomic base J is present within the telomeric repeats of *T. brucei,* whereas approximately 90% of genomic base J is localized to telomeres in *L. major*. In *T. cruzi,* roughly 75% of genomic base J is present at telomeres [22]. Thus, base J may have a critical function in maintaining telomere stability, length or integrity. However, preliminary analysis of J null *T. brucei* failed to provide evidence for changes in telomere length, consistent with the nonessential nature of base J within this organism (Sabatini unpublished data). Significant differences exist between *T. brucei* and *T. cruzi* genome organization, and it would be interesting to analyze the function of base J in *T. cruzi* telomeres specifically. However, our failure to generate J-null *T. cruzi* has rendered it difficult to test this phenomenon. Base J can be significantly reduced in *T. cruzi* by feeding BrdU to JBP1dKO cells (Ekanayake and Sabatini unpublished data). This technique, alone or in conjunction with genomic deletion of JBP2, might someday prove useful for evaluating the role of base J at *T. cruzi* telomeres.

Summary

Previous studies on in our laboratory have led to an understanding of the function of the JBP enzymes in base J biosynthesis. In this study, we have taken advantage of the established roles of JBP1 and JBP2 in base J biosynthesis to elucidate the function of the modified base in *T. cruzi.* As a consequence, we present the first direct evidence for epigenetic regulation of chromatin structure and Pol II transcription initiation via direct DNA modification in kinetoplastids. Finally, we provide a mechanism for regulation of trypanosome gene expression by the hyper-modified base J.

REFERENCES

1. Borst P, van Leeuwen F (1997) beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. Mol Biochem Parasitol 90: 1-8.
2. Tarleton R (2007) Immune system recognistion of *Trypanosoma cruzi.* Current Opinion in Immunol 19: 430-434.
3. Martin DL, Weatherly B, Laucella SA, Cabinian MA, Crim MT, et al. (2006) CD8+ T-cell response to *Trypanosoma cruzi* are highly focused on strain-variant trans-sialidase epitopes. PLOS Pathogens 2: 77.
4. Cliffe LJ, Siegel TN, Marshall M, Cross GA, 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 Res 38: 3923-3935.
5. Grohmann D, Werner F (2010) Hold on!: RNA polymerase interactions with the nascent RNA modulate transcription elongation and termination. RNA Biol: 310-315.
6. Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE (2010) Gene expression in trypanosomatid parasites. J Biomed Biotechnol 2010: 525241.
7. Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33: 29-45.
8. Ekanayake DK, Cipriano M, Sabatini R (2007) Telomeric localization of the modified base J and contingency genes in the genome of the protozoan parasite *Trypanosoma cruzi*. Nucleic Acids Res 35: 6367-6377.

9. van Leeuwen F, Kieft R, Cross M, Borst P (1998) Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. Mol Cell Biol 18: 5643-5651.

10. Kazazian HHJ (2004) Mobile elements: drivers of genome evolution. Science 303: 1626-1632.

11. Capkova Frydrychova R, Biessmann H, Mason JM (2008) Regulation of telomere length in Drosophila. Cytogenet Genome Res 122: 356-364.

12. Kaneko-Ishino T, Ishino F (2010) Retrotransposon silencing by DNA methylation contributed to the evolution of placentation and genomic imprinting in mammals. Dev Growth Differ 62: 533-543.

13. Bringaud F, Berriman M, Hertz-Fowler C (2009) Trypanosomatid genomes contain several subfamilies of ingi-related retroposons. Eukaryot Cell 8: 1531-1542.

14. Durand-Dubief M, Absalon S, Menzer L, Ngwabyt S, Ersfeld K, et al. (2007) The Argonaute protein TbAGO1 contributes to large and mini-chromosome segregation and is required for control of RIME retroposons and RHS pseudogene-associated transcripts. Mol Biochem Parasitol 156: 144-153.

15. Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. RNA 7: 1522–1530.

16. Militello KT, Wang P, Jayakar SK, Pietrasik RL, Dupont CD, et al. (2008) African Trypanosomes Contain 5-Methylcytosine in Nuclear DNA. Eukarotic Cell 7: 2012-2016.

17. Bringaud F, Ghedin E, Blandin G, Bartholomeu DC, Caler E, et al. (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. Mol Biochem Parasitol 145: 158-170.

18. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of T*rypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005 Jul 15;309(5733):409-15 309: 409-415.

19. Lorenzi HA, Robledo G, Levin MJ (2006) The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-enconding retrotransposons. Mol Biochem Parasitol 145: 184-194.

20. Heras SR, López MC, Olivares M, Thomas MC (2007) The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. Nucleic Acids Res 35: 2199-2214.

21. Jain D, Cooper JP (2010) Telomeric strategies: means to an end. Annu Rev Genet 44: 243-269.

22. Borst P, Sabatini R (2008) Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol 62: 235-251.