IDENTIFICATION AND CHARACTERIZATION OF THE BASE J GLUCOSYLTRANSFERASE AND ITS POTENTIAL USE IN THE FIELD OF EPIGENETICS

by

WHITNEY L. BULLARD

(Under the Direction of Robert Sabatini)

ABSTRACT

Trypanosomatids, including the human infectious parasites *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* possess a unique DNA modification in their genomes. This DNA modification, known as base J, is synthesized in a two-step process in which specific thymines are hydroxylated to form hmU, then subsequently glucosylated to form base J. While the hydroxylases involved in the first step have been characterized, previous attempts to identify the glucosyltransferase has been unsuccessful. This elusive glucosyltransferase has now been identified and is known as the base J-associated glucosyltransferase, or JGT. Here we demonstrate the involvement of JGT in base J synthesis. The deletion of both alleles of JGT from the genome of *T. brucei* generates a cell line that completely lacks base J. Reintroduction of the JGT in the JGT-/- background stimulates base J synthesis. In an *in vitro* assay, recombinant JGT utilizes UDP-glucose to transfer glucose to hmU in the context of dsDNA. Further *in vitro* characterization of JGT demonstrates its ability glycosylate hmU within any sequence with no significant change in Km or kcat, suggesting that JGT possesses no DNA sequence specificity. The identification of this unique enzyme and its characterization as a DNA

sequence non-specific enzyme has led to the development of a technique using JGT to map the location of hmU throughout the genome. JGT can specifically convert hmU to base J in a genomic sample and the resulting base J can be enriched using an anti-base J antibody pull-down. This technique could be used to map the genomic location of hmU and may help to shed light on the potential functional of this oxidized DNA base. These findings presented here have ultimately led to the identification of a novel glucosyltransferase, a greater understanding of the synthesis of base J in trypansomatids, and the development of a technique that can be used to increase our understanding of epigenetic DNA modifications.

INDEX WORDS: Kinetoplastid, Trypanosome, DNA Modification, Base J, Epigenetics, Transciptional Regulation, Glucosyltransferase, Hydroxylase, DNA Modifying Enzyme

IDENTIFICATION AND CHARACTERIZATION OF THE BASE J GLUCOSYLTRANSFERASE AND ITS POTENTIAL USE IN THE FIELD OF EPIGENETICS

by

WHITNEY L .BULLARD

B.S., Berry College, 2010

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

2016

IDENTIFICATION AND CHARACTERIZATION OF THE BASE J GLUCOSYLTRANSFERASE AND ITS POTENTIAL USE IN THE FIELD OF EPIGENETICS

by

WHITNEY L. BULLARD

Major Professor:	Robert Sabatini
Committee:	David Garfinkel
	Rick Tarleton
	Michael Tiemeyer

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

ACKNOWLEDGEMENTS

I would, of course, like to thank my advisor Robert Sabatini and my committee members, David Garfinkel, Rick Tarleton, and Michael Tiemeyer, for their guidance and direction.

I would not be where I am today without the support and encouragement I've received from my family, friends, mentors, and people I've worked with over the years. I want to say thank you to all of you.

I would first like to thank my family for all of their love and support while I've been in graduate school. I especially would like to express my appreciation to my parents and grandparents for if it were not for their hard work I wouldn't have had an opportunity to continue my education. I thank them so much for the support and love that they've given me throughout my entire life and you have my utmost gratitude.

I would also like to thank certain people who were instrumental in fostering my love for science and helping me get into graduate school. I thank Mrs. Orloff for sparking my initial interest in science. A BIG thanks to Dr. Chris Hall for letting me do research in his lab as an undergraduate. I cannot thank Dr. Hall enough for being a wonderful mentor, I learned so much from working with him and I appreciate his help in jumpstarting my career and education. Another BIG thanks to Dr. Steve Hajduk for letting me be a part of his lab group and especially for giving me my own research project. I really enjoyed being in Hajduk lab and I really feel like through his mentorship I grew tremendously as a scientist.

Lastly, I've had the pleasure of working with some wonderful people over the years that I now consider some of my closest friends. Nicole Acuff has been my closest friend and scientific companion since Dr. Hall's lab and I thank her so much for being a wonderful friend. In Hajduk

lab I worked alongside some really great scientists and really great people; Natalie Stephens, Steven Sykes, David Seidman, John Harrington, and Tony Szempruch, I am so grateful to have worked with amazing people like them. I would also like to thank Laura Cliffe, Gwen Hirsch, Jessica Lopes da Rosa-Speigler, Rudo Kieft, and David Reynolds for being such wonderful labmates and friends while I was in graduate school, they have all been so kind, supportive, and helpful over the years and I will miss them dearly.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Kinetoplastids are a group of single-celled flagellate protozoan, a number of which cause serious disease in humans and other animals (1). The most common human diseases caused by kinetoplastids include African sleeping sickness, caused by an infection with the parasite *Trypanosoma brucei*; Chagas disease caused by infection with *Trypanosoma cruzi*; and Leishmaniasis, caused by infection from different species of *Leishmania* (1). It is estimated that half a billion people are at risk of contracting diseases caused by these organisms and that more than 20 million people are currently infected, resulting in more than 100,000 deaths per year (1). There are currently no cures for these diseases and drug treatments are often toxic, difficult to administer, and have varying levels of success (1).

Each of these parasites is transmitted to a human host by an arthropod vector. *T. brucei* it is transmitted by a tsetse fly, *T. cruzi* is vectored by a Triatomine insect, and *Leishmania* is passed to humans through the bite of a sandfly (1,2). These organisms must be able to respond to environmental changes associated with transmission, from an insect vector to human host, including changes in pH, temperature, osmolarity, oxygen availability, and nutrients, and initiate differentiation to new developmental stages (2). Their ability to respond to environmental cues and to adapt their gene expression to new situations is critical for their survival.

Transcriptional regulation is an important contributor to the control of gene expression in eukaryotes; however, it was thought that kinetoplastids do not regulate the expression of genes at the transcriptional level (2). Instead it is thought that regulation of gene expression in kinetoplastids is primarily through post-transcriptional mechanisms (2,3). Regulation of gene expression in kinetoplastids through post-transcriptional mechanisms is a consequence of their unusual genome arrangement. These organisms are unique from other eukaryotes in that their genes are organized into large polycistronic transcription units (PTUs) (2-4) (Figure 1.1A). Adjacent PTUs may be located on the same DNA strand, arranged in a head to tail fashion, or on different strands with two PTUs either diverging or converging (5). These sites where two PTUs converge is known as a convergent strand switch region (cSSR) and sites where two PTUs diverge is known as a divergent strand switch region (dSSR) (5) (Figure 1.1A). These PTUs contain tens-to-hundreds of functionally unrelated genes (2,3,5). The genes within a PTU are transcribed by RNA polymerase II (Pol II), yielding a polycistronic precursor mRNA (2,3) (Figure 1.1B). This polycistronic precursor mRNA is then processed into individual mRNA transcripts by trans-splicing, in which a 39nt spliced leader (SL) sequence is added to the 5' end, followed by polyadenylation of the 3' end (2,3) (Figure 1.1B). Because the genes within a PTU are transcribed at roughly the same rate, the expression of each individual transcript may be regulated post-transcriptionally (2).

The data suggesting that regulation of gene expression in kinetoplastids is entirely through post-transcriptional mechanisms has recently been challenged by the discovery of multiple epigenetic marks enriched at sites of transcription initiation and termination throughout the genome (6,7). One such modification, known as base J, is a

hypermodified DNA base that consists of a glucosylated thymidine (8). Base J represents a unique DNA modification unlike any that as been described in other eukaryotes. The unique nature of this modification and its role in gene expression in these organisms could provide us with more information about how DNA modifications and the enzymes that generate them play a role in epigenetics. Not much is known about the regulation of gene expression at the level of transcription in kinetoplastids. Research into how base J plays a role in these processes may yield information into control of transcription in these organisms and open up avenues for better genetic tools and a better understanding of polycistronic transcription. In addition to base J being a unique modification, the enzymes that synthesize it also deserve attention. Understanding the enzymes involved in base J synthesis and their regulation is critical in furthering our understanding of the role base J plays in the regulation of gene expression in these parasites. Here, a review of our current knowledge of base J, its function, and the enzymes involved in its synthesis will be presented.

BASE J

Base J (ß-D-glucopyranosyloxymethyluracil) is a hypermodified thymine residue that accounts for about 1% of thymines in the nuclear genome (9). The presence of this modified base was first indicated through analysis of *T. brucei* variant surface glycoprotein (VSG) expression sites (8). Within the bloodstream of a human host, *T. brucei* evades immune destruction through antigenic variation of the VSG protein on their cell surface (10). There are over 2,000 complete and partial VSG genes that exist within the *T. brucei* genome, however only one VSG is expressed at any given time (10)

(Figure 1.2B). The expressed VSG is located within a specialized subtelomeric transcription unit known as an expression site (ES), and while multiple ESs exist, only one of these ESs is active and all others are silent (10) (Figure 1.2A and 1.2B). Several reports observed that silent VSG genes were only partially digested by certain nucleases. However, upon ES activation, the VSG gene became fully digestible (11,12). Because certain DNA modifications can block DNA digestion by nucleases (13), it was hypothesized that a DNA modification was present on silent ES, but not active ones.

Attempts were then made to identify this potential DNA modification within the genome. Nucleotide post-labeling combined with two-dimensional thin layer chromatography revealed the presence of two modified DNA bases (14). One of these modified bases was identified as hydroxymethyluracil (hmU), however the other modified DNA base did not co-migrate with any known nucleotide (14). Mass spectrometry, chromatography, nuclear magnetic resonance spectroscopy, and other biochemical analyses later identified this unknown DNA base as ß-D-glucopyranosyloxymethyluracil and was named base J (14). No other DNA modifications have been identified within the trypanosome genome.

Base J has since been found in all kinetoplastids analyzed thus far including the human pathogens *T. cruzi*, several species of *Leishmania,* in the insect parasite *Crithidia fasciculate,* and in the fish parasite *Trypanoplasma borreli*. Base J has also been found in related organisms such as the marine flagellate *Diplonema* and in the unicellular algae *Euglena gracilis* (15,16). The presence of base J has not been identified in a variety of other protozoa, fungi, or in vertebrates (15).

LOCALIZATION OF BASE J

Base J is located at specific sites throughout the kinetoplastid genome. Base J is predominately present in repetitive sequences, such as telomeric repeats (15,17) (Figure 1.3A). The high degree of J modification in telomeric repeats has been shown for all kinetoplastids analyzed thus far (15). In *T. brucei*, base J localizes to subtelomeric regions at expression sites, which are important for VSG switching, as mentioned previously (18) (Figure 1.2A and 1.3B). Base J is enriched within other repetitive sequences throughout the *T. brucei* genome including the 50 and 70bp repeats within these expression sites and within gene sequences that exist in tandem arrays, including ribosomal RNA and SL RNA genes (19) (Figure 1.3B). In *Leishmania* about 98% of base J is restricted to telomeric repeats; none of the subtelomeric repeats possessed detectable amounts of base J (20). In *T. cruzi*, however, substantial amounts of non-telomeric base J have been found (21). This non-telomeric base J is localized to tandem gene arrays, in the subtelomeric repeat sequences, and in subtelomeric glycoprotein gene families implicated in pathogenesis (21).

Base J is not restricted to just telomeric and subtelomeric repetitive sequences, however. Recently, genome-wide enrichment of base J containing sequences and high-throughput sequencing revealed small amounts of base J at chromosomal internal regions in *T. brucei* (22). This chromosomal internal base J is specifically localized to dSSRs and cSSRs, which are sites of transcription initiation and termination, respectively (22) (Figure 1.3A). This enrichment of base J at dSSRs and cSSRs has also been shown for both *T. cruzi* and *Leishmania* as well (23-25). Not every SSR in the genome contains

base J. For example, in *T. brucei* only 50% of dSSRs and 60% of cSSRs are modified (22) and in *L. major* 64 out of 65 cSSRs are modified (23).

Base J modification *in vivo* is not only restricted to particular DNA sequences, but also to which Ts are modified within a given sequence. Nearest neighbor TLC analysis indicated sequence specific synthesis of J in telomeric repeats where only the second T in the top strand (GGTTA) is modified (17). A technique recently developed known as single-molecule, real-time (SMRT) sequencing, allows for strand-specific detection of modified bases in DNA with single base resolution (26). SMRT sequencing of episomes containing the telomeric repeat sequence from *Leishmania* confirm that only particular Ts within this sequence are base J modified (27).

It is unclear what determines the specific localization of base J throughout the genome. From analysis of DNA sequences known to contain base J, no consensus sequence or motif has emerged. There is a need to understand what regulates specific localization of base J within the genome and to determine the requirements for J insertion into a particular DNA sequence. A better understanding of the synthesis of this DNA modification and the enzymes involved is crucial for understanding the regulation and sequence specificity of base J insertion.

BASE J SYNTHESIS

Indirect evidence indicates that base J is synthesized in a two-step pathway (8). In the first step specific thymidines in the genome are hydroxylated, forming 5-hydroxymethyluracil (hmU) (Figure 1.4A). In the second step, a glucose moiety is added to the hmU intermediate forming base J (Figure 1.4A). Several lines of evidence support

this hypothesis of base J synthesis. First, the specific localization of base J in the kinetoplastid genome suggests that thymidines are modified in DNA instead of being synthesized at the nucleotide level and then incorporated in to the genome during DNA replication (15,22). Second, when *T. brucei* is grown in the presence of hmU, it is randomly incorporated into the genome and subsequently converted to base J (28,29). Furthermore, the expression of the mammalian base excision DNA repair enzyme SMUG1 leads to *T. brucei* cell death by excessive DNA repair (30). Third, it is known that hmU is freely accessible in the genome as detected by thin layer chromatography and by mass spectrometry (9).

It is now known that the first step of base J synthesis is carried out by either of two thymidine hydroxylase (TH) enzymes, J-binding protein 1 (JBP1) or J-binding protein 2 (JBP2) (29,31). JBP1 and JBP2 generate the intermediate hmU within the genome (29,31), which then becomes the substrate for a glucosyltransferase (Figure 1.4A). The glucosyltransferase involved in the second step of base J synthesis has only recently been identified (32-34) and will be discussed in greater detail in subsequent chapters.

J-Binding Protein 1

JBP1 was first identified by its ability to bind to base J containing DNA substrates (35,36). Nuclear extracts from *C. fasciculata* were incubated with synthetic base J-containing duplex DNA and protein bound to base J was detected by gel-shift assay (35). The 90kDa protein was then purified by J-DNA affinity chromatography, subjected to MS analysis, and the corresponding gene was identified (35). This gene and its *T. brucei*

and *L. tarentolae* homologues were then cloned, expressed in *E. coli*, and recombinant protein was also capable of binding J-containing duplex DNA (35). This base J binding protein eventually became known as JBP1 and was unlike any known protein identified at the time (35).

The ability of JBP1 to bind base J is dependent on the J-DNA binding domain in the C-terminal half of the protein (31,37-39) (Figure 1.4B). Gel shift assays using recombinant JBP1 and double stranded DNA (dsDNA) substrates revealed that JBP1 binds specifically to J-containing DNA and with high affinity (10-140nM) (37). Optimal DNA binding requires at least five base pairs flanking J (37). The protein does not bind to single-stranded DNA (ssDNA) containing base J or to free base J (37). No significant binding to unmodified dsDNA was observed with gel-shift assays (37), however a more sensitive fluorescence anisotropy approach revealed JBP1 can bind unmodified dsDNA albeit with ~100-fold less affinity than J –DNA (40).

DNA footprinting analysis indicate that JBP1 does not make any sequence-specific contacts with the bases surrounding base J and should therefore be able to recognize J within any sequence context (41). However it was discovered that JBP1 binds to J within the telomeric repeat sequence with higher affinity than other non-repetitive sequences (37). The preference of JBP1 for repetitive sequences correlates with the high level of J-modification in the repetitive telomeric and subtelomeric sequences.

The crystal structure of the J-binding domain revealed a novel helix-turn-helix (HTH) fold termed a 'helical bouquet' (39). Within the recognition helix of the HTH motif there resides a single aspartate residue responsible for J binding and JBP1 function

in vivo (39). Inactivation of the N-terminal TH domain did not affect J-binding (42). Based on recent fluorescent polarization measurements, JBP1 binds J-DNA and then undergoes a conformational change (43). It is thought that this conformational change upon J-binding allows the TH domain to come into proximity to DNA (43).

J-Binding Protein 2

An *in silico* search for additional JBPs in the *T. brucei* genome identified a protein with homology to the N-terminus of JBP1 now known as JBP2 (44) (Figure 1.4B). Despite the name, JBP2 does not bind the modified base directly, but is able to bind chromatin in a base J independent manner, presumably via the C-terminal SWI2/SNF2 chromatin-remodeling domain (22,44) (Figure 1.4B). Mutation of key residues within the ATPase region of the SWI2/SNF2 ablates JBP2 function *in vivo* (44), however it is unknown if ATP hydrolysis is required for JBP2 to bind/remodel chromatin to stimulate hydroxylation. It is also not known if the recognition of chromatin by JBP2 is dictated by DNA sequence or structure, or by interactions with other epigenetic marks such as histone variants. Further work is needed to fully understand the role this domain plays in base J synthesis and the requirements for proper functioning.

Thymidine Hydroxylase Activity of the JBPs

Both JBP1 and JBP2 contain a TH domain at the N-terminus that has led to the designation of these enzymes as belonging to the TET/JBP subfamily of dioxygenases that require Fe^{2+} and 2-oxoglutarate (2-OG) for activity (29,31,42,45-47) (Figure 1.4B). Members of this family are typically identified based on a structural level by the presence

of a "jelly roll" β-helix sheet that contains four key conserved residues involved in the binding of Fe^{2+} and 2-OG and are essential for catalytic activity (48,49). Mutation of these key residues in both JBP1 and JBP2 inhibits enzyme function, and affects base J synthesis *in vivo* (42). Deletion of either JBP1 or JBP2 from bloodstream form *T. brucei* results in a 20- and 8- fold reduction in the levels of base J, respectively. The simultaneous deletion of both JBP1 and JBP2 yields a cell line (JBP null) that is unable to synthesize base J unless cells are fed hmU.

To further characterize the JBPs as Fe^{2+}/2-OG-dependent dioxygenases, an *in vitro* TH assay was developed using recombinant JBP1 (42). *In vitro* JBP1 hydroxylates thymidine in the context of dsDNA and this hydroxylation activity was dependent on Fe^{2+}, 2-OG, and O_2, as expected for this class of enzymes (42). The N-terminal TH domain of JBP1 was sufficient for full activity *in vitro* (42). Mutation of residues involved in coordinating Fe^{2+} inhibited iron binding and the formation of hmU (42). A well-characterized inhibitor of dioxygenases, dimethyloxalylglycine (DMOG) inhibited the TH activity of the JBPs both *in vitro* and *in vivo* (42). These data confirm the identity of JBP1 and JBP2 as the thymidine hydroxylases catalyzing the first step of base J synthesis.

JBPs *in vivo*

While both JBP1 and JBP2 stimulate thymidine hydroxylation *in vivo*, each operates optimally under different chromatin environments (22). Re-expression of JBP2 into JBP null *T. brucei* stimulated J synthesis within the telomeres whereas JBP1 stimulates J synthesis primarily at genome internal sites (22). Optimal levels and

localization of base J only occurred upon re-expression of both JBP enzymes (22). Presumably this functional difference is due to the distinct C-terminal domains of each protein (35,38,39,44,50). It is thought that the ability of JBP1 to bind J residues in the genome may enhance TH activity (35,38,39,43,50). JBP1 may play a J-propagation function, binding to base J and spreading base J into adjacent regions. Thus, JBP1 is able to maintain and propagate J in a site-specific manner. JBP2 can bind chromatin independent of base J, presumably due to its SWI2/SNF2 domain (44,51). JBP2 may recognize specific regions of chromatin where it can remodel chromatin and allow the TH domain access to thymidine residues in DNA. However, the requirements for JBP mediated formation of hmU in specific regions of the genome remains unclear.

The ability to delete both JBP enzymes from *T. brucei* is consistent with the non-essential nature of base J in this organism (29). While deletion of either JBP1 or JBP2 in *T. cruzi* (24), and JBP2 in *L. tarentolae* (52,53), results in similar reductions in J levels as seen in *T. brucei* mutants, attempts to delete both JBPs have been unsuccessful, leading to the idea that base J is essential in these organisms. It is also possible that the JBP enzymes carry out essential functions in addition to J synthesis. Inhibiting the JBP enzymes *in vivo* through the use of DMOG can reduce the levels of base J to extremely low levels without affecting cell growth (54). This result suggests there are alternative functions of JBP enzymes outside of base J synthesis and would explain the inability to generate a JBP null in *T. cruzi* and *Leishmania*.

The Base J-Associated Glucosyltransferase

Attempts to identify the glucosyltransferase (GT) involved in the synthesis of base J had been unsuccessful until a recent bioinformatic study identified a possible candidate GT in trypanosomatids (32). Examining biochemical pathways for DNA modifications, Iyer et al. identified a GT-A-like glucosyltransferase with an operonic association to a JBP-related gene within several phage genomes (32). The authors postulated that because operons contain functionally related genes, these TET/JBP-associated glycosyltransferases may glycosylate substrates generated by the JBP-like enzymes (32). Multiple sequence alignments showed that the TET/JBP-associated glycosyltransferases, including the kinetoplastid homologs, possess many of the structural elements and catalytic residues characteristic of the Rossmannoid nucleotide-diphospho-sugar binding fold typical of the GT-A/fringe superfamily (32).

The candidate GT proposed in this study has since been confirmed as the GT involved in the synthesis of base J in kinetoplastids, called the base J-associated glucosyltransferase (JGT) (33) (Figure 1.4C). Recombinant JGT utilizes UDP-glucose to transfer glucose onto double stranded DNA substrates containing hmU *in vitro* (33). *In vivo*, deletion of both JGT alleles in *T. brucei* results in a complete loss of base J from the genome (33,34). These studies not only confirm its identity as a GT and the two-step base J synthesis model, but also indicate it is the only GT catalyzing the second step of base J synthesis.

The recent identification of the glucosyltransferase involved in base J synthesis has increased our understanding of base J synthesis and now allows further investigation into base J function in kinetoplastids. More in depth discussion of the identification and

characterization of this unique enzyme is presented in subsequent chapters (Chapters 2 and 3).

Developmental Regulation of Base J Synthesis

Base J synthesis is developmentally regulated in *T. brucei* (18). Base J is only detectable in the DNA of bloodstream form parasites, but is not present in the DNA of insect form parasites. Analysis of the two life cycle stages revealed that the developmental regulation of base J synthesis is controlled at both steps of base J synthesis. Insect form parasites down-regulate the production of both TH enzymes (JBP1 and JBP2) and the GT enzyme (JGT). However, hmU feeding or overexpression of JBP2 in insect form *T. brucei* results in base J synthesis in the genome (29,55). Therefore, developmental regulation of base J synthesis may be governed by the TH enzymes involved in the first step of base J synthesis.

Unlike *T. brucei*, both *T. cruzi* and *Leishmania* possess base J in both their insect and mammalian lifecycles, however, there appears to be some degree of developmental regulation of base J synthesis between the two lifecycles. In *L. donovani* and *T. cruzi*, there is an approximately 2-fold more base J in the DNA of bloodstream form parasites compared to insect form parasites (15,21). The specific localization of base J throughout these lifecycle stages and potential functional role for regulation of base J synthesis throughout different developmental stages is unknown.

FUNCTION OF BASE J

Because the identity of the enzymes involved in the first step of base J synthesis are known and well characterized, we have been able to exploit this step in the pathway to study the function of base J in these organisms. Although the entire picture of the functional roles of base J is still unclear it is becoming more evident that base J plays a role in the regulation of transcription in kinetoplastids.

Trypanosoma brucei

T. brucei utilizes VSGs on their cell surface to evade the host immune system by means of antigenic variation (10). The expression of VSG genes must be tightly controlled, with only one VSG expressed at a time and other VSG genes kept transcriptionally silent (10) (Figure 1.2B). As mentioned previously, in *T. brucei* base J is found at silent VSG expression sites and absent from the active VSG expression site (18) (Figure 1.3B). The association of base J with silent expression sites in BSF *T. brucei* led to the hypothesis that base J plays a role in the regulation of antigenic variation. However, no direct evidence has been put forth to support this idea.

Recent work suggests that, base J localizes at transcription termination sites within gene clusters in *T. brucei* (54). Loss of base J through inhibition of the JBPs led to read-through transcription at specific sites within the gene cluster and an increased expression of downstream genes, suggesting a role for base J in the regulation of transcription termination (54) (Figure 1.5C). While base J localizes to transcription termination sites where two PTUs converge, loss of this mark did not result in read-through transcription into the opposing gene cluster (54). These data suggests that in *T.*

brucei base J plays a role in the transcription termination at sites within a PTU and regulating the expression of specific genes downstream of that site, but is not required for termination within a cSSR.

Trypanosoma cruzi

In *T. cruzi* we are unable to generate a cell line that lacks both JBP1 and JBP2, however knockouts of the individual enzymes can be generated yielding cells that have reduced levels of base J (24). Loss of base J at transcription initiation sites led to an increased rate of RNA Polymerase II (Pol II) transcription and global changes in gene expression (24,25). Additionally these cells had increased host cell invasion efficiency and delayed egress as compared to wild type (WT) cells (24). The loss of base J at transcription termination sites did not result in any detectable defects in transcription termination, however (24,25). It was later shown that JBP1 knockout *T. cruzi* exhibited decreased nucleosome abundance, increased Pol II occupancy, and increased histone H3/H4 acetylation at sites of transcription initiation, all indicative of active transcription (25) (Figure 1.5A). These data support the claim that base J increases active chromatin marks and plays a role in regulating transcription initiation in *T. cruzi*.

Leishmania

In *Leishmania* reduction in base J levels has been accomplished through deletion of the JBP2 enzyme; attempts to create a JBP1 knockout cell line have been unsuccessful (52,53). Deletion of JBP2 from *L. tarentolae* lead to read-through of Pol II transcription termination sites with generation of antisense RNAs corresponding to the adjacent PTU

(23) (Figure 1.5B). When JBP2 knockout *L. tarentolae* is treated with the thymidine analog bromodeoxyuridine (BrdU), the levels of base J are further reduced, leading to exacerbated read-through and cell death (23). In *L. major*, reduction of base J following DMOG treatment results is a similar read-through at transcription termination sites throughout the genome, however DMOG treatment did not result in cell death (54). In *Leishmania,* base J is likely to play a role in the regulation of Pol II transcription termination within cSSRs, preventing transcription into the adjacent PTU, unlike what was seem in *T. brucei* (23,54,56). Like *T. brucei*, base J is also found within gene clusters in *L. major* (54,56). DMOG treatment of *L. major* lead to a loss of base J at these sites resulting in read-through transcription past the termination site and increased expression of downstream genes (56) (Figure 1.5C). Therefore, base J not only regulates transcription termination within cSSRs but also within gene clusters, and suppresses expression of specific genes throughout the genome.

SIMILAR DNA MODIFICATIONS AND DNA MODIFYING ENZYMES

DNA modifications are widespread in nature and have been found in eukaryotes, prokaryotes, and in bacteriophages (57,58). Nucleic acids can carry diverse modifications and have been shown play a critical role in a variety of cellular processes (58). The discovery of base J in kinetoplastids has added to this an understanding of a novel epigenetic mark and a description of the unique enzymes involved in its synthesis. While base J is, with few exceptions, found only in kinetoplastids, oxidized and glycosylated DNA bases have been described in other organisms. The enzymes involved

in the synthesis of these oxidized and glycosylated bases share many features to the enzymes involved in the synthesis of base J.

TET Enzymes

One of the best-studied DNA modifications is methylated cytosine. Methylcytosine (mC) is an important epigenetic mark within the mammalian genome that regulates gene expression of many biological processes (59). It is generally considered a repressive mark as the presence of methylated cytosines at a promoter region is directly connected with transcriptional repression of a gene (60). The presence of mC at particular regions of the genome is a dynamic and reversible process (61). Programmed alterations in the methylation landscape are crucial for regulation of genes during development (61). mC can be removed through a passive demethylation mechanism through cell replication or through an active demethylation pathway catalyzed by the ten eleven translocation (TET) enzymes (62). The TET enzymes are mC hydroxylases, which catalyze the iterative oxidation of mC to hydroxymethylcytosine (hmC), formylcytosine (fC), and carbosylcytosine (caC) (46,63,64). fC and caC can then be excised by a thymine DNA glycosylase (TDG) followed by base excision repair (BER) to replace mC with an unmodified cytosine (C) (64,65).

The TET enzymes perform a similar hydroxylation function to that of the JBP enzymes and belong to the TET/JBP superfamily of Fe^{2+}/2-OG-dependent dioxygenases (46,47). In fact, the TET enzymes were first identified based on homology to the JBP enzymes from *T. brucei* (46). The mammalian TET family contains three members, TET1, TET2, and TET3 all of which contain a conserved C-terminal catalytic domain

(47). TET1 and TET3 also have a CXXC domain, which is capable of binding mC and hmC (66). *in vitro*, recombinant full-length TET proteins and TET catalytic domains are able to convert 5mC to 5hmC in a Fe^{2+}, 2-OG, and O_2 dependent manner (66). Ectopic expression of TET proteins *in vivo* reduces mC levels and increases levels of hmC and this activity is ablated upon mutation of critical residues within the active site (62).

Increasing evidence has supported the notion that TET enzymes and mC derivatives play important roles in development. TET proteins are implicated in several stages of mouse development, particularly those in which global changes in methylation status take place, particularly the zygote, blastula, and primordial germ cells (62,66). Mice deficient in individual TET enzymes or combinations thereof display embryonic abnormalities and combined deficiency of all three TET enzymes in mESCs completely depletes hmC levels, impairs differentiation and does not support embryogenesis (66-72). Emerging evidence suggests that, in conjunction with interacting proteins, the TET oxidation products hmC, fC, and caC may themselves have unique epigenetic regulatory functions in development (73). While the TET enzymes, and the DNA modifications that they generate may be important for development it is unclear the exact mechanism, however it is speculated that they may play an epigenetic role in the regulation of gene transcription (73).

T4 β-Glucosyltransferase

The genome of the T4 bacteriophage also possesses hmC, which can further be modified by the attachment of glucose to form β-glycosyl-5-hydroxymethylcytosine (74,75). Incorporation of hmC into the phage DNA blocks cleavage by many bacterial

restriction endonucleases and glucosylated hmC further blocks attack by hmC-specific nucleases (76,77). Glucosylation of hmC has also been implicated in gene expression by influencing transcription (78). The T4 β-glucosyltransferase (β-GT) catalyzes the transfer of glucose from UDP-glucose onto hmC in duplex DNA to form glycosylated hmC (79). Like the generation of hmC by the TETs is similar to the generation of hmU by the JBPs, the T4 β-GT carries out a similar function of the glucosyltransferase involved in base J synthesis, by transferring glucose onto an oxidized DNA base.

The β-GT was the first GT x-ray crystal structure to be solved and is a founding member of the GT-B structural fold family of GTs (80). The GT-B fold consists of two domains with similar Rossman topology where the nucleotide-sugar substrate binds primarily to the C-terminal domain and the acceptor substrate binds to the N-terminal domain (80,81). As seen in other DNA modifying and repair enzymes, β-GT utilizes a base-flipping mechanism, rotating the sugar-phosphate backbone around the flipped-out target base allowing it to enter into the enzyme's activate site (82).

CONCLUSION

Base J represents a unique DNA modification unlike any that have been described in other organisms. The unique nature of base J modification and its role in gene expression in kinetoplastids could provide us with more information about how DNA modifications and the enzymes that generate them play a role in epigenetics. Not much is known about the regulation of gene expression at the level of transcription in kinetoplastids. Research into how base J plays a role in these processes may yield information into control of transcription in these organisms and open up avenues for

better genetic tools and an understanding of polycistronic transcription in general. In addition to base J being a unique modification, the enzymes that synthesize it also deserve attention. The JBPs, either themselves or through the aid of other proteins are capable of directing synthesis of base J at specific regions of the genome by hydroxylating thymidines. It was due in part to the discovery of the *T. brucei* JBPs that the identification of the mammalian homologues, the Ten Eleven Translocation (TET) proteins that hydroxylate methylcytosines within the mammalian genome, was made possible, an example of how discovery of kinetoplastid proteins impact other systems. Furthermore, the synthesis of base J represents a unique pathway only found in kinetoplastids, making it a potential drug target for treatment of these diseases. Research into this unique pathway may yield drug or peptide inhibitors specific to the enzymes of base J synthesis with little side effects to the host. Continued research into base J synthesis and function will provide insight into epigenetic regulation of transcription in kinetoplastids, may increase our overall knowledge of DNA modifying enzymes, and could become an attractive drug target to treat diseases caused by kinetoplastids.

REFERENCES

1. Stuart, K., Brun, R., Croft, S., Fairamb, A., Gurtler, R., McKerrow, J., Reed, S., and Tarleton, R. (2008) Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation*, **118**, 1301-1310.
2. Kramer, S. (2012) Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. *Molecular and Biochemical Parasitology*, **181**, 61-72.

3. Clayton, C. (2016) Gene Expression in Kinetoplastids. *Curr Opin Microbiol*, **32**, 46-51.

4. Palenchar, J., and Bellofatto, V. (2006) Gene Transcription in Trypanosomes. *Mol Biochem Parasit*, **146**, 135-141.

5. Maree, J., and Patterton, H. (2014) The epigenome of Trypanosoma brucei: a regulatory interface to an unconventional transcriptional machine. *Biochem Biophys Acta*, **1839**, 743-750.

6. Alsford, S., duBois, K., Horn, D., and Field, M. (2012) Epigenetic mechanisms, nuclear architecture and the control of gene expression in trypanosomes. *Expert Rev Mol Med*, **14**, 1-20.

7. Figueiredo, L., Cross, G., and Janzen, C. (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol*, **7**, 504-513.

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

9. Gommers-Ampt, J., Van Leeuwen, F., de Beer, AL., Vliegenthart JF., Dizdaroglu M., Kowalak, JA., Crain, PF., and Borst, P. (1993) Beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell*, **75**, 1129-1136.

10. Horn, D., and McCulloch. (2010) Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol*, **13**, 700-705.

11. Bernards, A., Van Harten-Loosbroek, N., and Borst, P. (1984) Modification of telomeric DNA in Try- panosoma brucei: a role in antigenic variation? . *Nucleic Acids Res*, **12**, 4153-4170.

12. Pays, E., Delauw, M., Laurent, M., and Steinert, M. (1984) PossibleDNAmodificationin GC dinucleotides of *Trypanosoma brucei* telomeric sequences: relationship with antigen transcription. *Nucleic Acids Res*, **12**, 5235-5247.

13. Meselson, M., Yuan, R., and Heywood, J. (1972) Restriction and Modification of DNA. *Annu Rev Biochem*, **41**, 447-466.

14. Gommers-Ampt, J., Lutgerink, J., and Borst, P. (1991) A novel DNA nucleotide in Trypanosoma brucei only present in the mammalin phase of the life-cycle. *Nucleic Acids Res*, **19**, 1745-1751.

15. Van Leeuwen, F., Taylor, MC., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., Borst, P. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci* **95**, 2366-2371.

16. Dooijes, D., Chaves, I., Kieft, R., Dirks-Mulder, A., Martin, W., and 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.

17. Van Leeuwen, F., Wijsman, E.R., Kuyl-Yeheskiely, E., Van der Marel, G.A., Van Boom, J.H. and Borst, P. (1996) The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res*, **24**, 2476–2482.

18. Van Leeuwen, F., Wijsman, E.R., Kieft, R., van der Marel, G.A., van Boom, J.H. and Borst, P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev.* , **11**, 3232–3241.

19. Van Leeuwen, F., Kieft, R., Cross, M. and Borst, P. (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by B-D-glucosyl-hydoxymethyluracil in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **109**.

20. Genest, P., ter Riet, B., Cysouw, T., van Leunen, H., and Borst, P. (2007) Telmoeric localization of the modified DNA base J in the genome of the protozoan parasite *Leishmania*. *Nucleic Acids Res*, **35**, 2116-2124.

21. Ekanayake, D., Cipriano, M., and Sabatini, R. (2007) Telomeric localization of the modified base J and contin- gency genes in the genome of the protozoan parasite *Trypanosoma cruzi*. *Nucleic Acids Res*, **35**, 6367-6377.

22. Cliffe, L.J., Siegel, T.N., Marshall, M., Cross, G.A., and 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.

23. Van Luenen, H., Farris, C., Jan, S., Genest, PA., Tripathi, P., Velds, A., Kerkhoven, R., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., Van Steensel, B., Myler, P., and Borst, P. (2013) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909-921.

24. Ekanayake, D., Minning, T., Weatherly, B., Gunasekera, K., Nilsson, D., Tarleton, R., Ochsenreiter, T., and Sabatini, R. (2011) Epigenetic Regulation of

Transcription and Virulence in *Trypanosoma cruzi* by O-Linked Thymine Glucosylation of DNA. *Mol Cell Biol*, **31**, 1690-1700.

25. Ekanayake, D., and Sabatini, R. (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi.* Modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryot. Cell* **10**, 1465-1472.

26. Roberts, R., Carneiro, M., and Schatz M. (2013) The advantages of SMRT sequencing. *Genome Biology*, **14**, 405.

27. Genest, P., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H., Korlach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P., and Borst, P. (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res*, **43**, 2102-2215.

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

29. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R.,Marshall, M., Sweeney, K., and Sabatini, R. (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.

30. Ulbert, S., Cross, M., Boorstein, R., Teebor, G., and Borst, P. (2002) Expression of the human DNA glycosylase hSMUG1 in *Trypanosoma brucei* causes DNA damage and interferes with J biosynthesis. *Nucleic Acids Res.* , **30**, 3919-3926.

31. Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* , **35**, 2107–2115.

32. Iyer, L., Zhang, D., Burroughs, M., and Aravind L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA *Nucleic Acids Res.* , **42**, 7635-7655.

33. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014) Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J Biol Chem.*, **289**, 20273-20282.

34. Sekara, A., Merritta, C., Baugha, L., Stuarta, K., and Myler, P. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, **196**, 9-11.

35. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marel, G. A., van Boom, J. H., van Leeuwen, F., and Borst, P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO*, **18**, 6573– 6581.

36. De Kort, M., Ebarhimi, E., Wijsman, E., Van der Marel, G., and Van Boom, J. (1999) Synthesis of oligodeoxynucleotides containing 5-(beta-D-glucopyranosyloxymethyl)-2′deoxyuridine, a modified nucleoside in the DNA of *Trypanosoma brucei*. *Eur J Org Chem*, **9**, 2337-2344.

37. Sabatini, R., Meeuwenoord, N., van Boom, J.H. and Borst, P. (2002) Recognition of base J in duplex DNA by J-binding protein. *J. Biol. Chem.*, **277**, 958–966.

38. Toaldo, C.B., Kieft, R., Dirks-Mulder, A., Sabatini, R., van Luenen, H.G. and Borst, P. (2005) A minor fraction of base J in kinetoplastid nuclear DNA is bound by the J-binding protein 1. *Mol. Biochem. Parasitol.*, **143**, 111–115.

39. Heidebrecht, T., Christodoulou, E., Chalmers, M. J., Jan, S., Ter Riet, B., Grover, R. K., Joosten, R. P., Littler, D., van Luenen, H., Griffin, P. R., Wentworth, P., Jr., Borst, P., and Perrakis, A. (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res.* , **39**, 5715–5728.

40. Grover, R., Pond, S., Cui, Q., Subramaniam, P., Case, D., Millar, D., and Wentworth, P. (2007) O-glycoside orientation is an essential aspect of base J recognition by the kinetoplastid DNA-binding protein JBP1. *Angew Chem Int Ed Engl*, **46**, 2839-2843.

41. Sabatini, R., Meeuwenoord, N., van Boom, J.H. and Borst, P. (2002) Site-specific interactions of JBP with base and sugar moieties in duplex J-DNA. Evidence for both major and minor groove contacts. *J Biol Chem*, **277**, 28150-28156.

42. Cliffe, L., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y., and Sabatini, R. (2012) JBP1 and JBP2 Proteins are Fe2+/2-Oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *J. Biol. Chem.* , **287**, 19886-19895.

43. Heidebrecht, T., Fish, A., von Castelmur, E., Johnson, K.A., Zaccai, G., Borst, P. and Perrakis, A. (2012) Binding of the J-binding protein to DNA containing

glucosylated hmU (base J) or 5-hmC: evidence for a rapid conformational change upon DNA binding. *J. Am. Chem. Soc.*, **134**, 13357–13365.

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

45. Reynolds, D., Cliffe, L., and Sabatini, R. (2015) 2-Oxoglutarate-Dependent Hydroxylases Involved in DNA Base J (β-D-Glucopyranosyloxymethyluracil) Synthesis. *2-Oxoglutarate-Dependent Oxygenases*, 309-323.

46. Tahiliani, M., Koh, K.P., Shen,Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.

47. Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698 –1710.

48. Schofield, C., and Zhang, Z. (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol*, **9**, 722-731.

49. Hausinger, R. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Critical Reviews in Biochemistry and Molecular Biology*, **39**, 21-68.

50. Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I. and Borst, P. (2002) J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol. Microbiol.*, **46**, 37-47.

51. Kieft, R., Brand, Verena., Ekanayake, D., Sweeney, K., DiPaolo, C., Reznikoff, W., and Sabatini R. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* , **156**, 24-31.

52. Genest PA., t.R.B., Dumas C., Papadopoulou B., van Luenen HGAM., Borst P. (2005) Formation of linear inverted repeat amplicons following targeting of an essential gene in *Leishmania*. *Nucleic Acids Res.* , **33**, 1699–1709.

53. Genest, P. (2007) Analysis of the modified DNA base J and the J-binding proteins in Leishmania. . *PhD Thesis*, **University of Amsterdam.**

54. Reynolds, D., Cliffe, L., Fo¨rstner, K., Hon, C., Siegel, N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res*, **42**, 9717-9729.

55. Van Leeuwen, F., De Kort, M., van der Marel, GA., Van Boom, JH., and Borst, P. (1998) The modified DNA base beta-D-glycosyl-hydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32P-postlabeling. *Anal. Biochem.*, **258**, 223-229.

56. Reynolds, D., Hofmeister, B., Cliffe, L., Siegel, T., Anderson B., Beverley, S., Schmitz, R., and Sabatini, R. (2016) Base J represses genes at the end of polycistronic gene clusters in Leishmania major by promoting RNAP II termination. *Mol Microbiol*.

57. Gommers-Ampt, J., and Borst, P. (1995) Hypermodified bases in DNA. *FASEB J*, **9**, 1034-1042.

58. Chen, K., Zhao, B., and He., C. (2016) Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem Biol*, **23**, 74-85.

59. Breiling, A., and Lyko, F. (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin*, **8**.

60. Jones, P. (2012) Functions of DNA methylation: islands, start sites, gene bodies, and beyond. *Nat Rev Genet*, **13**, 484-492.

61. Smith, Z., and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet*, **14**, 204-220.

62. Kohli, R.a.Z., Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**, 472-479.

63. Ito, S., Shen, L., Dia, Q., Wu, S., Collins, L., Swenberg, J., He, C., and Zhang, Y. . (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carbosylcytosine. . *Science*, **333**, 1300-1303.

64. He, Y., Li, B., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C., Zhang, K., He, C., and Xu, G. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303-1307.

65. Maiti, A., and Drohat, A. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implication for active demethylation of CpG sites. *J Biol Chem*, **286**, 35334-35338.

66. Pastor, W., Aravind, L., and Rao, A. (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev* **14**, 341-356.

67. Dawlaty, M., Ganz, K., Powell, B., Hu, Y., Markoulaki, S., Cheng, A., Gao, Q., Kim, J., Choi, S., Page, D., and Jaenisch, R. (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, **9**, 166-175.

68. Gu, T., Guo, F., Yang, H., Wu, H., Xu, G., Liu, W., Xie, Z., Shi, L., He, X., Hin, S., Iqbal, K., Shi, Y., Deng, Z., Szabo, P., Pfeifer, G., Li, J., and Xu, G. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, **477**, 606-610.

69. Li, Z., Cai, X., Cai, C., Wang, J., Zhang, W., Patersen, B., Yang, F., and Xu, M. (2011) Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*, **118**, 4509-4518.

70. Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M., Vasanthakumar, A., Patel, J., Zhao, X., Perna, F., Pandey, S., Madzo, J., Song, C., Dai, Q., He, C., Ibrahim, S., Beran, M., Zavadil, J., Nimer, S., Melnick, A., Godley, L., Aifantis, I., and Levine, R. (2011) Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, **20**, 11-24.

71. Dawlaty, M., Breiling, A., Le, T., Raddatz, G., Barrasa, M., Cheng, A., Gao, Q., Powell, B., Li, Z., Xu, M., Faull, K., Lyko, F., and Jaenisch, R. (2013) Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell*, **24**, 310-323.

72. Dawlaty, M., Breiling, A., Le, T., Barrasa, M., Raddatz, G., Gao, Q., Powell, B., Cheng, A., Faull, K., Lyko, F., and Jaenisch, R. (2014) Loss of Tet enzymes

compromises proper differentiation of embryonic stem cells. *Dev Cell*, **29**, 102-111.

73. Song, C., and He, C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends Biochem Sci*, **38**, 480-484.

74. Volkin, E. (1954) The linkage of glucose in coliphage nucleic acids. . *J. Am. Chem. Soc.*, **76**, 5892–5893.

75. Jesaitis, M. (1956) Differences in the chemical composition of the phage nucleic acids. *Nature*, **178**, 637–641.

76. Kornberg, S.R., Zimmerman, S. B., and Kornberg, A. . (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage- infected *Escherichia coli.* . *J. Biol. Chem.* , **236**, 1487−1493.

77. Josse, J., and Kornberg, A. (1962) Glucosylation of deoxyribonucleic acid. III. α- and β-glucosyltransferases from T4- infected *Escherichia coli.* . *J. Biol. Chem.* , **237**, 1968−1976.

78. Dharmalingam, K., and Goldberg, E. (1979) Restriction in vivo. III. General effects of glucosylation and restriction on phage T4 gene expression and replication. *Virology*, **96**, 393-403.

79. Terragni, J., Bitinaite, J., Zheng, Y., and Pradhan, S. (2012) Biochemical Characterization of Recombinant β-Glucosyltransferase and Analysis of Global 5-Hydroxymethylcytosine in Unique Genomes. *Biochemistry*, **51**, 1009-1019.

80. Vrielink, A., Ruger, W., Driessen, HP., and Freemont PS. . (1994) Crystal structure of the DNA modifying enzyme beta-glycosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. . *EMBO*, **13**, 3413-3422.

81. Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology*, **16**, 29-37.

82. Lariviere, L., and Morera, S. (2002) A base-flipping mechanism for the T4 Phage beta-Glucosyltransferase and identification of a transition-state analog. *J Mol Biol*, **324**, 483-490.

FIGURES

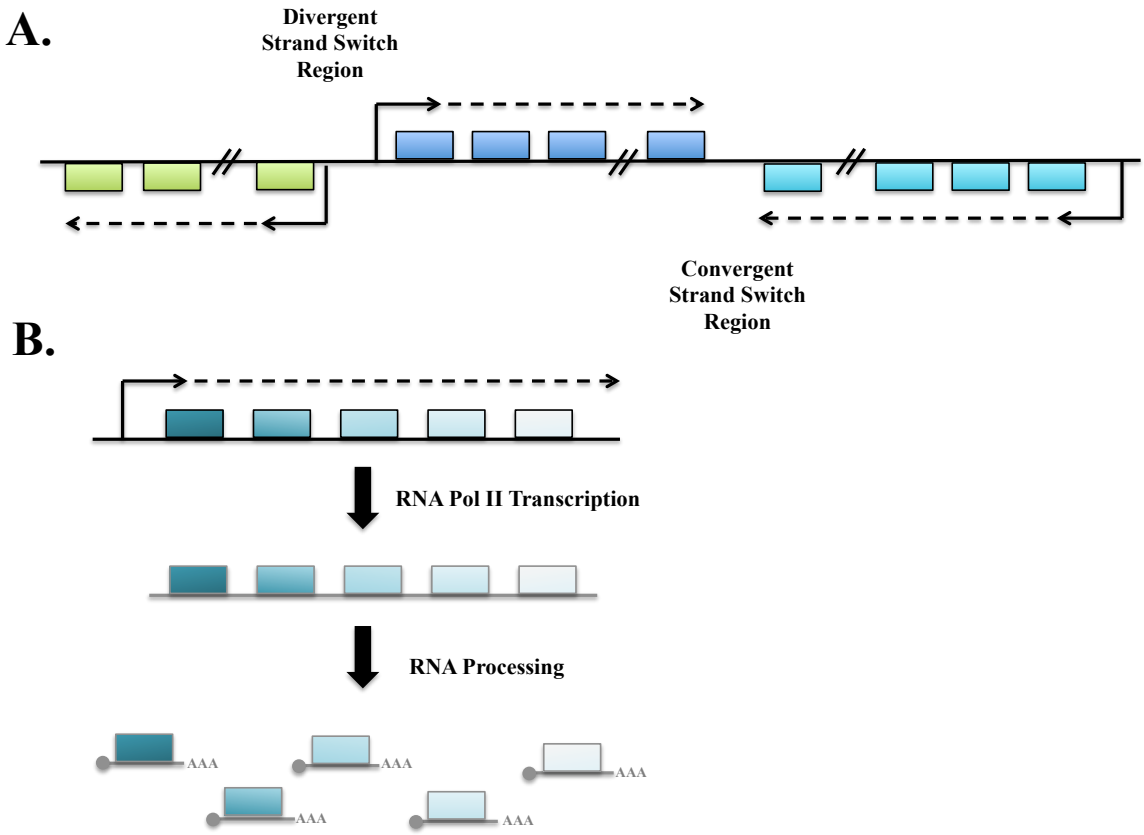


FIGURE 1.1. Genome Organization and Gene Expression in Kinetoplastids. A. Genes are arranged into large polycistronic transcription units (PTUs) containing functionally unrelated genes. Adjacent PTUs can either converge or diverge from one another. Sites where two PTUs converge are known as convergent strand switch regions (cSSRs) and sites where two PTUs diverge are known as divergent strand switch regions (dSSRs), which are sites of transcription termination and initiation, respectively. **B.** RNA polymerase II transcription proceeds through a PTU, generating a polycistronic precursor mRNA which is then processed into individual transcripts through processes such as trans-splicing and poly-adenylation.

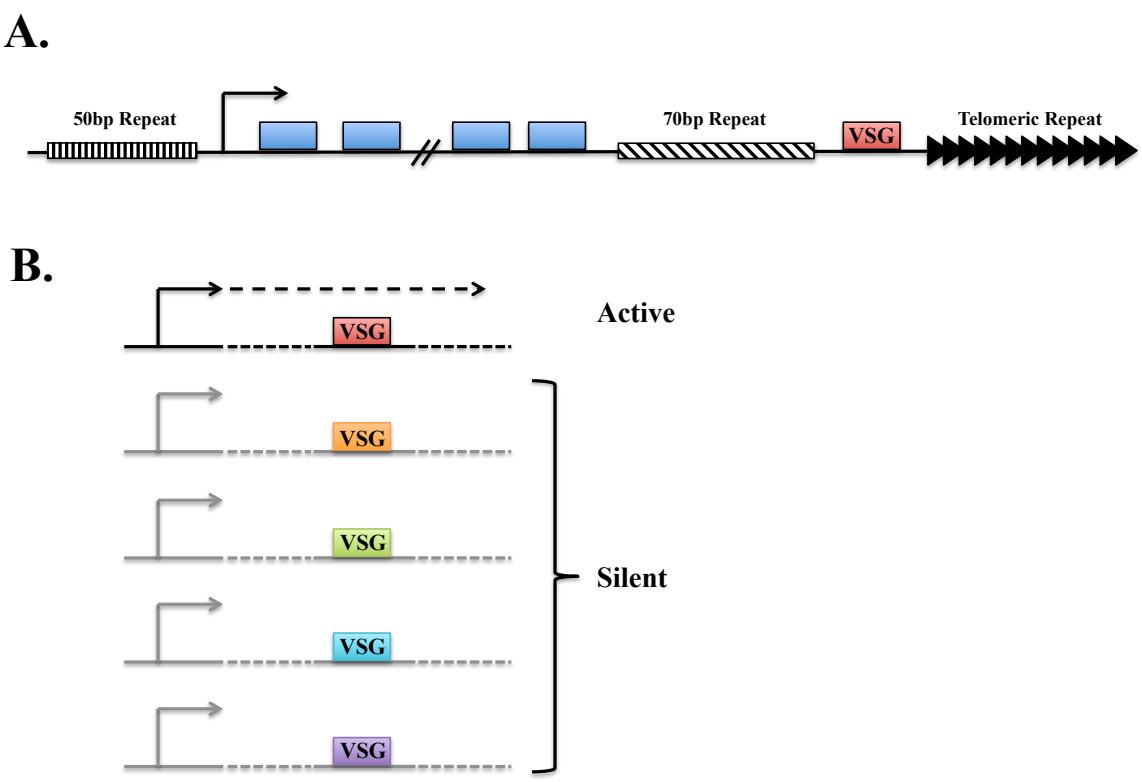


FIGURE 1.2. *T. brucei* VSG Expression Sites. A. In a *VSG* expression site, the *VSG* gene is the last one in the large polycistronic transcription unit and is located within ~2kb of the telomeric repeats. A stretch of 70bp repeats is located upstream of the *VSG* gene followed by a number of expression site associated genes (*ESAGs*). Upstream of the promoter is a stretch of 50bp repeats. **B.** Only one of several VSG expression sites is actively transcribed while the rest are silent.

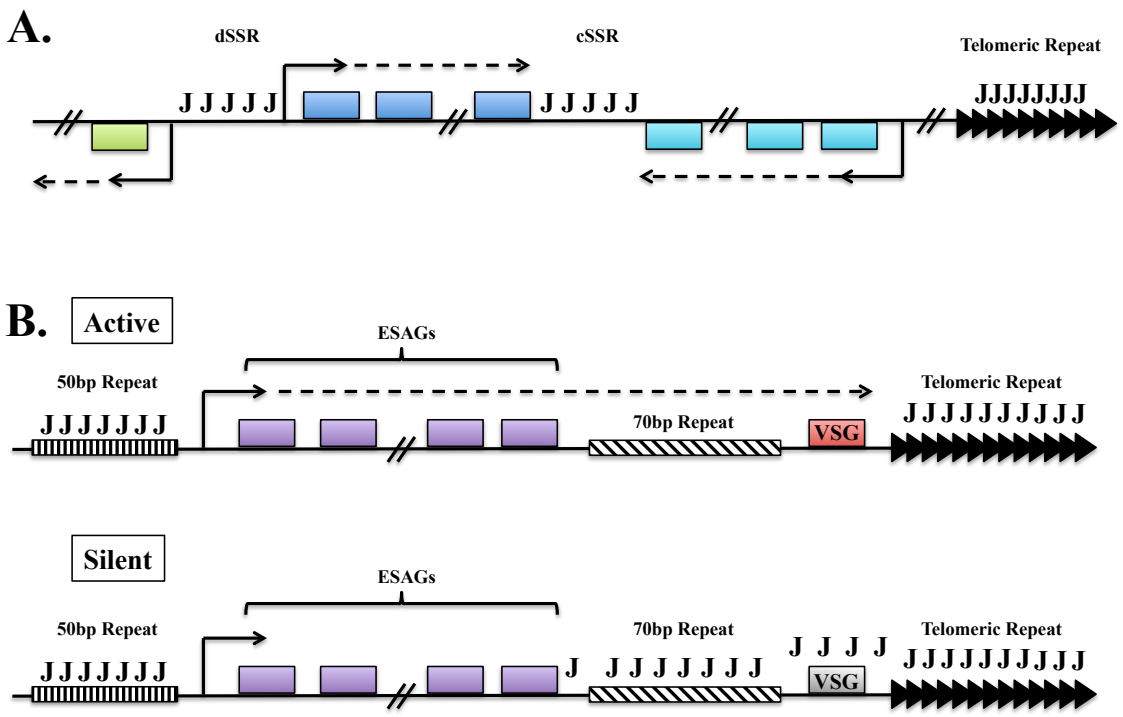


FIGURE 1.3. Localization of Base J. A. Base J is localized to specific sites throughout the genome. Base J is found at divergent and convergent strand switch regions (dSSRs and cSSRs), which are sites of transcription initiation and termination, respectively. The telomeric repeats are also modified with J. **B.** In *T. brucei*, base J is located in the inactive VSG expression site but is absent from the active one. Base J is found in the 50bp repeats upstream of the expression site and in the telomeric repeats independently of the transcriptional status of the expression site.

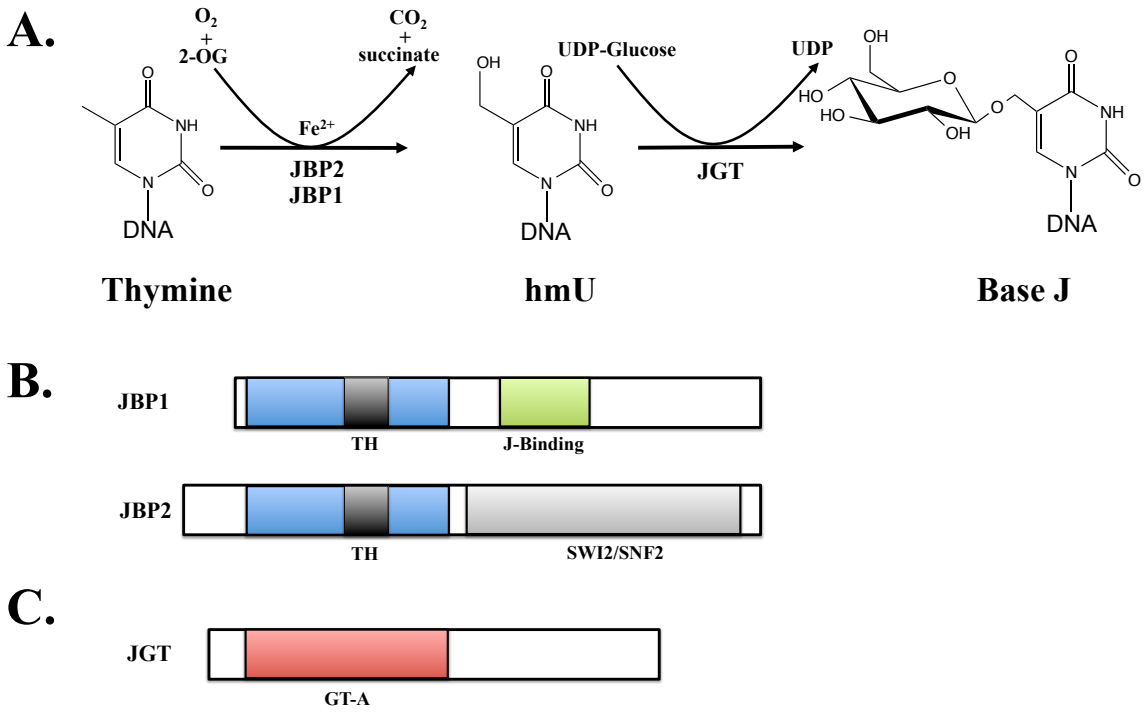


FIGURE 1.4. Base J Synthesis. A. Base J is synthesized in a two-step process. First thymine residues in the context of DNA are hydroxylated by the two thymidine hydroxylases J-binding protein 1 (JBP1) and J-binding protein 2 (JBP2) to form hydroxymethyluracil (hmU). Second, hmU becomes the substrate for a glucosyltransferase known as the base J-associated glucosyltransferase (JGT) which transfers a glucose onto hmU to form base J. **B.** Diagram of domains within JBP1 and JBP2. JBP1 and JBP2 both contain a N-terminal thymidine hydroxylase (TH) domain. JBP1 possesses a J-binding domain within the C-terminal half of the protein, while JBP2 possesses a SWI2/SNF2 chromatin-remodeling domain. **C.** Diagram of the catalytic GT-A domain within JGT.

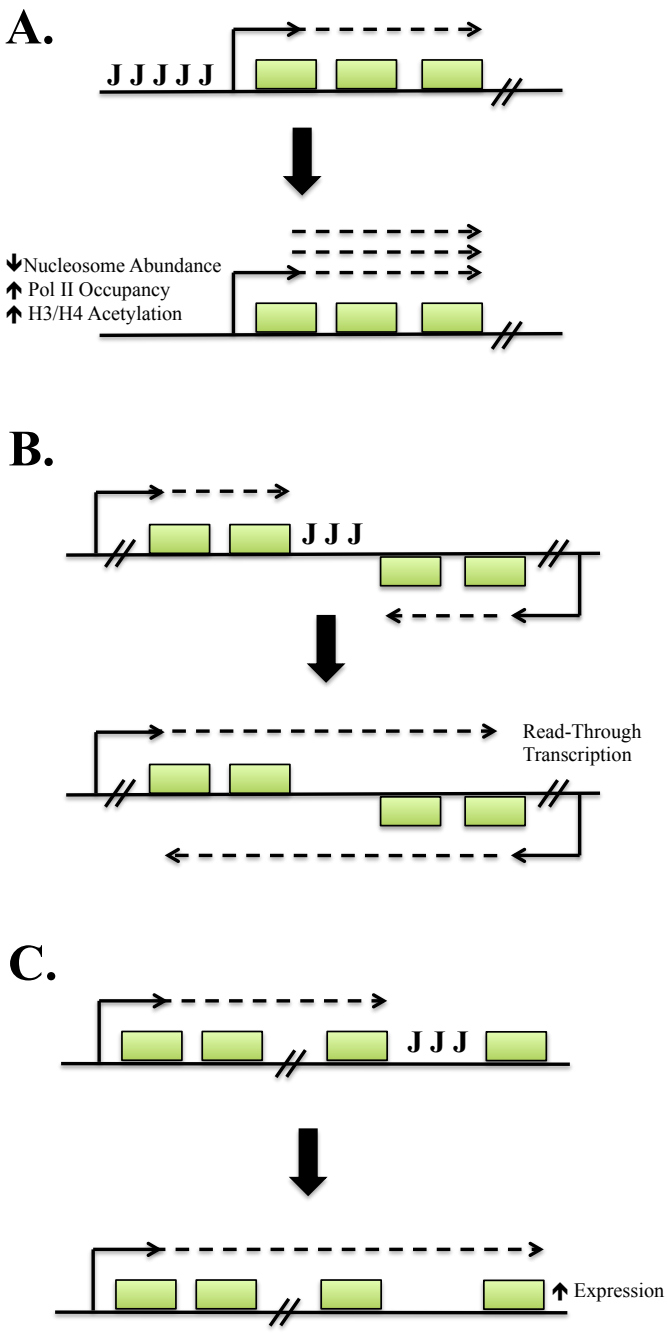


FIGURE 1.5. The role of base J in transcriptional regulation. **A.** Loss of base J from transcription initiation sites leads to decreased nucleosome abundance, increased marks of histone H3 and H4 acetylation, increased RNA Pol II occupancy, and increased transcript abundance of genes within the PTU. **B.** Loss of base J from transcription termination sites leads to read through transcription past the transcription stop sites within

a convergent strand switch region and continues into the flanking polycistronic transcription unit. **C.** Loss of base J from sites within a polycistronic transcription unit leads read through transcription and an increased expression of downstream genes.

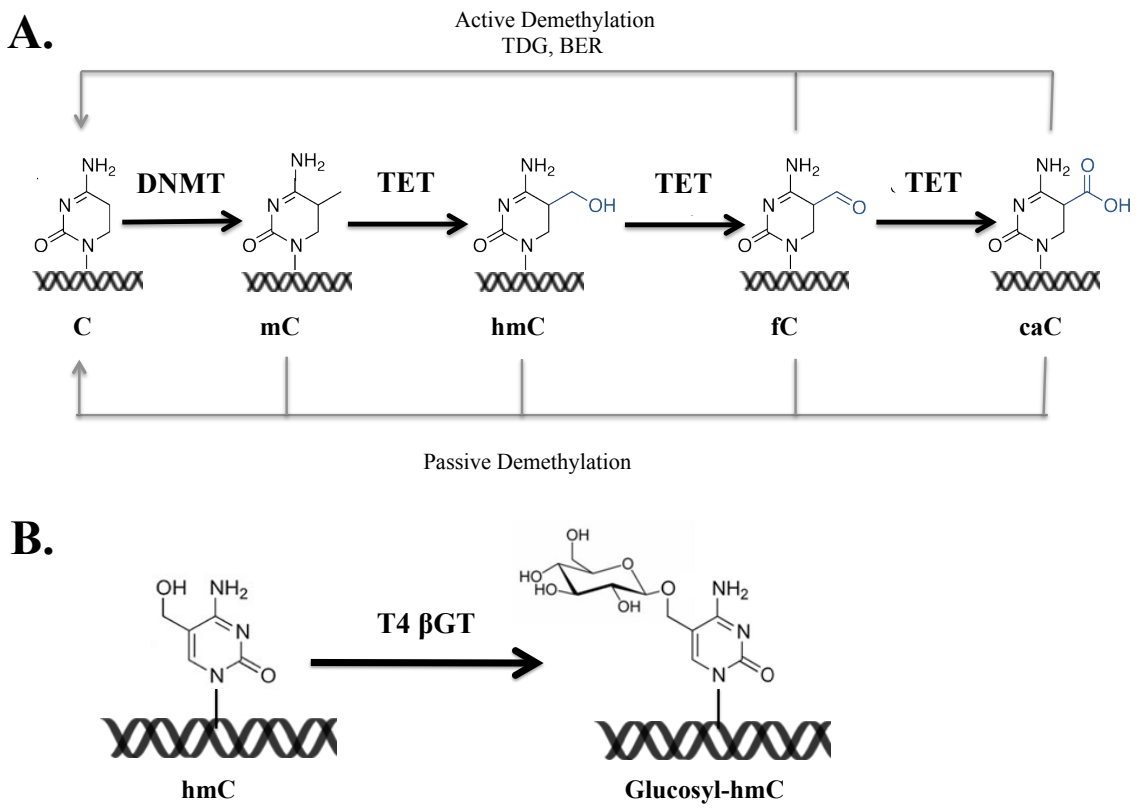


FIGURE 1.6. Oxidative and glycosylated DNA modifications in other systems. A. Iterative mC oxidation and demethylation pathways. Methylated cytosines (mC) are generated by DNA-methyltransferases (DNMTs) which can be oxidized by the ten-eleven translocation (TET) proteins to generate hydroxymethylcytosine (hmC), formylcytosine (fC), and carboxylcytosine (caC). mC, hmC, fC, and caC can be lost through passive demethylation by loss of maintainence during cell division. Active demethylation occurs when fC and caC are excised by thymine-DNA glycosylase (TDG) and replaced with cytosine (C) following base excision repair. **B.** The β-glucosyltransferase from T4 phage transfers a glucose moiety onto hydroxymethylcytosine in DNA.

CHAPTER 2

IDENTIFICATION OF THE GLUCOSYLTRANSFERASE THAT CONVERTS HYDROXYMETHYLURACIL TO BASE J IN THE TRYPANOSOMATID GENOME[1]

---

[1]Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. 2014. *J Biol Chem.*, **289**, 20273-20282. Reprinted here with permission of publisher.

ABSTRACT

O-linked glucosylation of thymine in DNA (base J) is an important regulatory epigenetic mark in trypanosomatids. β-D-glucopyranosyloxymethyluracil (base J) synthesis is initiated by the JBP1/2 enzymes that hydroxylate thymine, forming 5-hydroxymethyluracil (hmU). hmU is then glucosylated by a previously unknown glucosyltransferase. A recent computational screen identified a possible candidate for the base J-associated glucosyltransferase (JGT) in trypanosomatid genomes. We demonstrate that recombinant JGT utilizes uridine diphospho-glucose to transfer glucose to hmU in the context of dsDNA. Mutation of conserved residues typically involved in glucosyltransferase catalysis impairs DNA glucosylation *in vitro*. The deletion of both alleles of JGT from the genome of *Trypanosoma brucei* generates a cell line that completely lacks base J. Reintroduction of JGT in the JGT KO restores J synthesis. Ablation of JGT mRNA levels by RNAi leads to the sequential reduction in base J and increased levels of hmU that dissipate rapidly. The analysis of JGT function confirms the two-step J synthesis model and demonstrates that JGT is the only glucosyltransferase enzyme required for the second step of the pathway. Similar to the activity of the related Ten-Eleven Translocation (TET) family of dioxygenases on 5mC, our studies also suggest the ability of the base J-binding protein enzymes to catalyze iterative oxidation of thymine in trypanosome DNA. Here we discuss the regulation of hmU and base J formation in the trypanosome genome by JGT and base J-binding protein.

INTRODUCTION

Trypanosomatids, including the human pathogens *Trypanosoma brucei*, *Trypanosoma cruzi* , and *Leishmania* species, possess a unique DNA modification within their genome known as β-D-glucopyranosyloxymethyluracil (base J) (1,2). Base J is a hypermodified thymine residue that accounts for about 1% of thymines in the genome and is predominately present in repetitive sequences, such as telomeric repeats (2,3). In addition, base J is also found at chromosome-internal regions that coincide with sites of RNA polymerase II (Pol II) transcription initiation and termination (4). The loss of base J from these chromosome-internal regions in *T. cruzi* leads to more open chromatin and increased Pol II occupancy, ultimately leading to changes in gene expression and parasite virulence (5,6). Loss of base J in *Leishmania tarentolae* leads to transcription termination defects and generation of antisense RNAs corresponding to downstream genes (7). These data strongly support that base J represents a novel epigenetic marker involved in regulating Pol II transcription and gene expression. Understanding the enzymes involved in base J synthesis and their regulation is critical in furthering our understanding of the role base J plays in the regulation of gene expression in these parasites.

Indirect evidence, reviewed in Ref. 8, indicates a two-step pathway for the synthesis of base J (Figure 2.1A) (8). The first step of the pathway involves the hydroxylation of thymine in DNA by a thymidine hydroxylase enzyme, forming 5-hydroxymethyluracil (hmU). A glucose moiety is then attached to this intermediate, presumably by a glucosyltransferase enzyme, to form base J. Although the GT has not been identified, two thymidine hydroxylase enzymes (JBP1 and JBP2) involved in the

first step of base J synthesis have been characterized (9,10). Both JBP1 and JBP2 contain a thymidine hydroxylase domain at the N terminus that has led to the designation of these enzymes as belonging to the TET/base J-binding protein (JBP) subfamily of dioxygenases that require Fe^{2+} and 2-oxoglutarate for activity (11-15). JBP1 has a J-DNA binding domain in the C-terminal half of the protein (9,16-18). JBP2 does not bind the modified base directly, but is able to bind chromatin in a base J-independent manner, presumably via the C-terminal SWI2/SNF2 domain (10). Although both JBP1 and JBP2 stimulate *de novo* thymidine hydroxylation *in vivo*, the ability of JBP1 to bind J-DNA is thought to play a role in J propagation/maintenance (10,14,19). Deletion of either JBP1 or JBP2 from the bloodstream form *T. brucei* results in a 20- and 8-fold reduction in the levels of base J, respectively (9,14,20). The simultaneous deletion of both JBP1 and JBP2 yields a cell line that is unable to synthesize base J unless cells are fed hmU (14). Finally, using recombinant protein produced in *Escherichia coli*, we demonstrated that JBP1 stimulates the hydroxylation of thymidine in the context of dsDNA in a Fe^{2+}, 2-oxoglutarate-, and O_2 –dependent manner (15). These data confirm the identity of JBP1/2 as the thymidine hydroxylases catalyzing the first step of base J synthesis and that this step is independent from the subsequent step of glucose conjugation.

As reviewed in Borst and Sabatini (8), attempts to identify the GT involved in the synthesis of base J have been unsuccessful. However, a recent bioinformatic study identified a possible candidate GT in trypanosomatids (21). Examining biochemical pathways for DNA modifications, Iyer et al. (21) identified a GT-A-like glucosyltransferase with an operonic association to a JBP-related gene within several phage genomes. The authors postulate that because operons contain functionally related

genes, these TET/JBP-associated glycosyltransferases glycosylate substrates generated by the JBP-like enzymes. Multiple sequence alignment showed that the TET/JBP-associated glycosyltransferases, including the kinetoplastid homologs, possess many of the structural elements and catalytic residues characteristic of the Rossmannoid nucleotide-diphospho-sugar binding fold typical of the GT-A/fringe superfamily (21). We now show the first functional analysis of the *T. brucei* TET/JBP-associated glycosyltransferase homolog and refer to it as base J-specific GT (JGT). Our *in vitro* assays show that recombinant JGT utilizes UDP-Glc to transfer glucose to dsDNA substrates containing hmU. Mutation of conserved residues in the catalytic domain of the GT-A fold in JGT impairs DNA glucosylation. *In vivo*, deletion of both JGT alleles in *T. brucei* results in the complete loss of base J. Reintroduction of an ectopically expressed, HA-tagged JGT in the JGT KO restores base J synthesis. RNAi knockdown of JGT led to a sequential reduction in base J and increased levels of hmU and 5-formyluracil (fU) in the *T. brucei* genome in a time-dependent manner. The analyses of JGT function *in vivo* not only confirm its identity and the two-step base J synthesis model but also indicate that it is the only GT catalyzing the second step of base J synthesis. These studies also suggest the ability of the JBP enzymes to convert hmU to fU. Here we discuss the regulation and consequence of the oxidation and glucosylation of hmU in the trypanosome genome by JBP and JGT.

MATERIALS AND METHODS

Trypanosome Cell Culture and Generation of *T. brucei* Transfectants

The bloodstream form *T. brucei* cell line strain 427 and the 90–13 RNAi cell line were cultured in HMI9 medium supplemented with 10% heat-inactivated fetal bovine serum and 10% serum plus as described previously (10). 90-13 cells were cultured in the presence of 2.5μg/ml neomycin and 5μg/ml hygromycin to maintain the intergrated genes for T7 RNA polymerase and the tetracycline repressor, respectively. Transfections of bloodstream form *T. brucei* were essentially carried out as described previously (14).

Generation of the JGT Knockout

Constructs used to generate JGT knockout lines contain either a hygromycin phosphotransferase (Hyg) gene (pTub-Hygro) or a blasticidin S deaminase (BSR) gene (pTub-Blast). The *T. brucei* JGT-flanking regions used to target each allele were PCR-amplified from *T. brucei* 427 genomic DNA. The 611-bp fragment corresponding to the 5' flank was PCR-amplified using the sense primer 5'-GCGGCCGCCGGCACTGACGATCTTACAT-3' (the NotI site is underlined) and the antisense primer 5'-GGATCCCACATAATATAGCGCCACACATTC-3' (the BamHI site is underlined). The resulting PCR products were cloned into the pTub-Hygro and pTub-Blast vectors digested with NotI and BamHI. The 595-bp fragment corresponding to the 3' flank was PCR-amplified using the sense primer 5'-AAGCTTTGCAGATGGCGTGTTTCT-3' (the HindIII site is underlined) and the antisense primer: 5'-CTCGAGACGTGTGCCTAATACACTTACC-3' (the XhoI site is underlined) and cloned into the pTub-Hygro construct. For the pTub-Blast construct, the

3’ flank was generated using the sense primer 5’-GGGCCCTGCAGATGGCGTGTTTCT-3’ (the ApaI site is underlined) and the antisense primer 5’-GGGCCCAAGCTTACGTGTGCCTAATACACTTACC-3’ (the ApaI site is underlined). The resulting PCR product was cloned into the ApaI digested pTub-Blast vector. The pJGT-Tub-Hygro knockout construct was digested with NotI and XhoI, and the pJGT-Tub- Blast knockout construct was digested with NotI and HindIII prior to transfection. Transfectants were selected for resistance at 5μg/ml hygromycin and 5μg/ml blasticidin. Two rounds of transfection were used to replace both JGT alleles. The correct integration of the KO constructs and deletion of the JGT ORF was confirmed by PCR.

Re-expression of HA-tagged JGT

To allow ectopic expression of HA-JGT fusion, we utilized a modified ptub-phleo construct (22) that contains the enhanced green fluorescent protein ORF cloned between the tubulin flanking sequences. A 2012-bp PCR product corresponding to the *T. brucei* JGT ORF was PCR-amplified using the sense primer CCTGCAGGATGGCTTACCCATATGATGTTCCAGATTACGCTGGAGGTCCAAG GAGGGGAAG (the SbfI site is underlined and sequence coding for the HA tag is shown in boldface) and the antisense primer GGCGCGCCTTAGTCTGCCTGCGACCCTCC (the AscI site is underlined) and cloned into the ptub- GFP vector digested with SbfI and AscI. Insertion of the HAJGT PCR product in the place of EGFP now allows expression of HA-JGT N terminus fusion protein after integration into the tubulin array. The final

construct (HA-JGT-tub-phleo) was digested with XbaI and XhoI before electroporation, and transformants were selected for resistance to 2.5μg/ml phleomycin.

JGT RNAi

To allow Tet-inducible ablation of the *T. brucei* JGT mRNA, the pZJM-JGT RNA interference construct was generated by inserting a 524-bp portion of the JGT ORF into the XhoI and HindIII sites of the RNAi vector pZJM (23). A 524-bp fragment of the coding sequence of the JGT gene (tb427.10.6900) was amplified from *T. brucei* 427 genomic DNA using the sense primer CTCGAGGTGAATGTGTATGCCAACGC (the XhoI site is underlined) and the antisense primer AAGCTTTTGTTCTGCTGGCAGATGTC (the HindIII site is underlined) and cloned into the XhoI- and HindIII-digested pZJM vector. The resulting construct, pZJM-JGT, was digested with NotI prior to transfection. Transfectants were selected for resistance at 2.5μg/ml phleomycin.

Determination of the Genomic Level of J

To quantify the genomic J levels, DNA was isolated and utilized in the anti-J DNA immunoblot assay as described previously (24). Briefly, serially diluted genomic DNA was blotted to nitrocellulose, followed by incubation with anti-J antisera. Bound antibodies were detected by a secondary goat anti-rabbit antibody conjugated to HRP and visualized by ECL. The membrane was stripped and hybridized with a probe for the β-tubulin gene to correct for DNA loading.

Quantitative Reverse Transcription PCR

Total RNA was obtained using Qiagen RNeasy kits according to the instructions of the manufacturers. First-strand cDNA was synthesized from 1µg of total RNA using an iScript cDNA synthesis kit (Bio-Rad) according to the instructions of the manufacturer. Heat-inactivated cDNA reaction mixtures were finally treated with RNase H at 37°C for 45 min. Quantification of selected genes was performed on an iCycler with an iQ5 multicolor realtime PCR detection system (Bio-Rad). The reaction mixture contained 5pmol of forward and reverse primer, 2µl iQ SYBR Green Super Mix (Bio-Rad), and 2µl of template cDNA. Standard curves were prepared for each gene using 5-fold dilutions of a known quantity (250 ng/µl) of genomic DNA from WT *T. brucei* DNA. The quantities were calculated using iQ5 optical detection system software. Each sample was normalized to enolase mRNA. The primer sequences utilized in this analysis were as follows: JGT, 5'-CCTGACTGAGAACCCTTACTTC-3' (sense) and 5'-GGCACGTGTGACCATATACA-3' (antisense); JBP1, 5'-GTGTCCTAGCTGTGCTCAAA-3' (sense) and 5'-CAGGTGCGTATCGAAGAGTAAG-3' (antisense); JBP2, 5'-CCTTCCACCTTTGTGTATTCCT-3' sense) and 5'-CAACCGTCTCCTTCCTTGATAC-3 (antisense); enolase, 5' -GGCCTGCAACTCTCTTCTAC-3' (sense) and 5'-CATCACTGACCAGCCATTCT-3' (antisense).

Preparation of Recombinant JGT

The open reading frame of *T. brucei* JGT (Tb927.10.6900 TriTrypDB) was codon-optimized for *E. coli* expression and cloned into pET16b expression vector by GeneArt. The codon-optimized JGT nucleotide sequence is available upon request. The final construct was Tb-JGT-pET16b. Expression and purification of the N-terminal His10 -tagged Tb-JGT was performed with BL21-Codon-Plus(DE3)-RIL-competent cells (Agilent). Briefly, freshly transformed bacterial cell cultures were induced at A600nm 0.4–0.6 with 1mM isopropyl 1-thio-β-D-galactopyranoside for 16 hours at 16°C. Cell pellets were washed in 50mM HEPES (pH 7.5), 300mM NaCl, 10mM β-mercaptoethanol, and 100mM PMSF and then flash-frozen and stored at -70°C until purification. Cells were lysed and sonicated in lysis buffer (50mM HEPES-NaOH (pH 7.5), 300mM NaCl, 10mM β-mercaptoethanol, and 10% glycerol) with protease inhibitors (aprotinin, leupeptin, pepstatin, PMSF, and EDTA-free complete protease inhibitor; Roche). Cell lysates were clarified at 31,000xg for 20 min at 4°C. Recombinant JGT was affinity-purified with Talon resin (CloneTech) at 4°C for 1 hour, washed with 600mM NaCl lysis buffer in a batch, and eluted by gravity column with 150mM imidazole lysis buffer. Purified protein was concentrated by Centricon (Millipore) and visualized by colloidal blue-stained SDS-PAGE and anti-His Western blot analysis. JGT concentration was estimated on the basis of mass spec analysis indicating the ratio of JGT and DnaK in the purified sample and comparison to BSA standards on Coomassie-stained SDS-PAGE.

Mutagenesis of the *T. brucei* JGT was achieved using the Q5 site-directed mutagenesis kit (New England Biolabs). As a template for the PCR Tb-JGT-pET16b

construct, the following primer pairs were used: r(80)a, 5'-GAGCAAAGGTgcaTTTTATCATGAACGTGGC-3' (forward) and 5'-GGAACAAAAATCGGAACAG-3' (reverse); d241a, 5'-TGGGTTATGGcTGATGATATCG- 3' (forward) and 5'-ATACCACTGTGCTGCATG-3' (reverse); and d243a, 5'-ATGGATGATGGcTATCGCCAAATTTTTC-3' (forward) and 5'-AACCCAATACCACTGTGC-3' (reverse). Nucleotides shown in lowercase indicate altered amino acids to yield a point mutation. The mutated construct was verified by DNA sequencing and utilized as described above for expression and purification.

In Vitro Glucosylation Assay

A standard glucosylation assay consisted of 88μM UDP-[^{3}H]glucose (specific activity 45Ci/mmol), ~0.05μM recombinant JGT, and 100μM DNA containing four hmU residues in buffer (50mM potassium acetate, 20mM Tris acetate, 10mM manganese acetate, and 1mM DTT (pH 7.9)). Reaction mixtures were incubated at 37°C for 30 min, and the reaction was stopped by addition of 10μl of 400μM cold UDP-glucose and flash-freezing in liquid nitrogen for processing. The reaction mixtures were thawed and applied immediately to a 2.5-cm DE81 membrane (GE Healthcare, catalog no. 3658-325) under air pressure using a vacuum manifold. The filters where then washed in 3 x 2ml of 0.2M ammonium bicarbonate, 3 x 2 ml of water, and 3 x 2 ml of 100% ethanol. Membranes were air-dried and placed in scintillation vials containing 5mL of scintillation fluid. The solution was mixed, and tritium incorporation was measured for 1 min. All glucosylation reaction values were corrected for nonspecific binding of UDP-[^{3}H]glucose to the filters.

Background values were determined using reactions performed in the absence of enzyme but in the presence of UDP-[^{3}H]glucose. Oligos modified with hmU (H) (ACCCHAACCCHAACCCHAACCCHA) and an unmodified (thymidine) control (ACCCTAACCCTAACCCTAACCCTA) were hybridized to complementary oligo (TAGGGTTAGGGTTAGGGTTAGGGT), synthesized by Integrated DNA Technologies to generate hmU modified and unmodified dsDNA substrates, respectively. DNA duplexes were made by boiling complementary oligos for 10 min and allowing them to cool overnight.

Western Blotting

Proteins from $3x10^{7}$ cell equivalents were separated by SDS-PAGE 8% gel, transferred to nitrocellulose, and probed with anti-HA antibody (Santa Cruz Biotechnology, catalog no. SC-805) (1:1000 dilution), anti-histone H3 (Abcam, catalog no. ab8580) (1:2500 dilution) or anti-JBP2 antibody, as described previously (19). Bound antibodies were detected by a secondary goat anti-rabbit antibody conjugated to HRP and visualized by ECL.

Detection of Base J, hmdU, and fdU by Mass Spectrometry

To minimize artificial DNA oxidation, a modified procedure was used to isolate total genomic DNA for LC/MS analysis. Briefly, *T. brucei* cells were pelleted, washed in 1X PBS, and then resuspended in lysis buffer (1% SDS, 25mM EDTA, 0.4 M NaCl, 50mM Tris-HCl (pH 7.9), and 400µg/ml proteinase K) and incubated at 37 °C overnight.

5 M NaCl was then added and centrifuged to remove protein and cell debris, and DNA was precipitated with cold 100% ethanol. Purified genomic DNA from the indicated trypanosome cell lines was digested to nucleosides and subjected to LC/MS analysis. Briefly, DNA was digested to nucleosides as described previously, and isotopically labeled 5-hydroxymethyl-2'-deoxyuridine (hmdU) was added to the final mixture (25). Prior to digestion, single-stranded oligodeoxynucleotides housing 5-((β-D-glucopyranosyloxy)-methyl)-cytidine (glc-hmdC) was added to the sample to provide a surrogate internal standard for LC-MS/MS quantification of 5-((β-D-glucopyranosyloxy)-methyl)-uridine (dJ). The target nucleosides, dJ, glc-hmdC, and hmdU, were enriched via offline HPLC and collected separately for subsequent LCMS/MS measurement. LC-MS/MS analysis of dJ and glc-hmdC was conducted on an LTQ XL linear ion trap mass spectrometer equipped with a nanoelectrospray ionization source coupled to an EASY-nLC II system (Thermo Fisher Scientific, San Jose, CA). The enriched fraction containing hmdU was separated on an Agilent 1200 capillary HPLC, and the eluent was directed to an LTQ linear ion-trap mass spectrometer (ThermoFisher Scientific, San Jose, CA) following procedures reported previously (25). For LC-MS/MS/MS quantification of fdU in the genome, 250fmol of [1,3-$^{15}N_2$ -2'-D]-5-fdU was added to the enzymatic digestion mixture of 1μg of genomic DNA. The chemical synthesis of the labeled 5-hmdU and 5fdU standards is described in Ref. 26 (26). To improve the detection limit of fdU in positive-ion electrospray ionization (ESI)-MS, the digested DNA was then derivatized with Girard reagent T to form a hydrazone conjugate harboring a precharged quaternary ammonium moiety, as reported previously (27). The nucleoside mixture containing unlabeled and isotope-labeled fdU-GirT hydrazone was subsequently

extracted with chloroform to remove the enzymes, and the aqueous layer was subjected to LC-MS/MS/MS analysis.

Microscopy

The detection of HA-JGT expressed in the *T. brucei* JGT-/- was performed by anti-HA immunofluorescence analysis. Cells were fixed in 1% paraformaldehyde for 5 min on ice and then washed in HMI9 medium (without serum additives). Samples were applied to slides and allowed to air-dry. Slides were then blocked in 20% FBS in PBS for 30 min. Slides were incubated with anti-HA antibody (Covance, catalog no. MMS-101R) (1:100 dilution), followed by Alexa Fluor 594 goat anti-mouse antibody (Invitrogen, catalog no. A21125) (1:500 dilution). Images were acquired using an Axio observerZ1 equipped with an Axiocam MRm camera controlled by Axiovision version 4.6 software.

RESULTS

JGT Is a Glucosyltransferase Utilizing UDP-Glucose to Transfer Glucose to hmU

The putative base J-specific GT encoded in the trypanosomatid genome contains many of the structural elements and catalytic residues characteristic of the Rossmannoid nucleotide-diphospho-sugar binding fold typical of the GT-A/fringe superfamily (21) (Figure 2.1B). Common features of the GT-A domain that are found in JGT include a positively charged residue (Arg-80) and a "DXD" (here DDD) motif that are typically involved in nucleotide sugar binding (28-30) To test whether the JGT protein possessed glucosyltransferase activity, the *T. brucei* JGT was cloned and expressed in *E. coli* to produce recombinant protein with an N-terminal histidine tag (Figure 2.1C). Using the

purified recombinant enzyme fraction, we are able to show that JGT is able to utilize UDP-glucose to transfer glucose onto dsDNA substrates containing hmU (Figure 2.1D). We see little transfer using the same DNA substrate with unmodified Thr residues instead of hmU. Alanine substitution of Arg-80, Asp-241, and Asp-243 (within the DXD motif) caused a significant reduction in *in vitro* GT activity (Figure 2.1E). Alanine substitution of equivalent conserved (Asp) residues in other GTs resulted in similar reductions in activity (31,32). Thus, elimination of one critical residue may not necessarily result in the complete loss of substrate binding. These data not only demonstrates that the JGT possesses glucosyltransferase activity but that UDP-glucose is the sugar donor and that hmU modified DNA is a substrate for the enzyme. JGT is also predicted to be an inverting enzyme consistent with the structure of base J.

JGT Is the Glucosyltransferase Involved in Base J Synthesis *in Vivo*

To investigate whether JGT is involved in the synthesis of base J, we deleted both alleles of JGT from bloodstream form *T. brucei*. Loss of JGT mRNA was confirmed using quantitative reverse transcription PCR analysis (Figure 2.2A). J levels in the genomic DNA of JGT-/+ and JGT-/- cells were examined using an anti-base J dot blot. Although the deletion of one allele decreased the levels of J in the genome, the deletion of both alleles of JGT resulted in the complete loss of J synthesis (Figure 2.2B), as confirmed by LC-MS/MS analysis (Figure 2.2C and Figure 2.3). In this regard, we observed a peak in the selected ion chromatograms for monitoring the m/z 420è304 transition for glc-hmdC and the m/z 421è125, 143 transition for dJ (Figure 2.3). The former transition is due to the cleavage of the N-glycosidic linkage in the $[M+H]^+$ ion of

glc-hmdC (m/z 420). The collisional activation of the $[M+H]^+$ ion (m/z 421) of dJ again leads to the facile cleavage of the N -glycosidic bond to yield the ion of m/z 305, which can further eliminate a glucose or part of the glucose ($C_6H_{10}O_5$) to give fragment ions of m/z 125 and 143, respectively. To confirm that the reduction and loss of base J was caused by the loss of the JGT protein, an HA-tagged version of JGT was re-expressed in the JGT-/- cell line from the tubulin locus (Figure 2.2A and 2.2D). Upon re-expression, JGT localizes to the nucleus (Figure 2.2E) and restores J synthesis (Figure 2.2B). Taken together, these results clearly identify that JGT is the GT involved in the synthesis of base J in *T. brucei*.

According to the two-step model of base J synthesis (Figure 2.1A), the GT modifies hmU generated by the thymidine hydroxylase (JBP1/2). We have shown here that JGT can utilize hmU-DNA *in vitro* (Figure 2.1D). Therefore, we might expect an increase in hmU levels in the genome upon loss of JGT *in vivo*. However, unexpectedly, we did not detect an increase but, rather, a slight decrease in hmU levels in JGT-/- compared with WT *T. brucei* (Figure 2.2C). One possible explanation is that JGT loss leads to altered levels or activity of JBP1/2. Although we lack antisera against JBP1, we detected no change in JBP2 levels upon deletion of JGT by Western blot analysis (data not shown). Although it is possible that JGT somehow stimulates JBP hydroxylase activity, a more feasible explanation for the lack of hmU accumulation upon the loss of GT is that the intermediate is subsequently shunted to the formation of additional base analogs. The mammalian JBP/TET enzymes (TET1–3) function in the demethylation pathway in which 5-methylcytosine can be hydroxylated iteratively to 5-hydroxymethylcytosine (hmC), 5-formylcytosine, and 5-carboxycytosine (33,34).

Potentially, JBP1/2 function in a similar manner, and accumulation of hmU in trypanosome DNA is prevented via conversion to fU and 5-carboxyluracil. To test this hypothesis and more closely investigate the role of the JGT in base J biosynthesis, we took advantage of the inducible RNAi system in *T. brucei*. As expected, RNAi-mediated ablation of JGT results in decreasing levels of base J over an ~3-day time period (Figure 4A–C). LCMS/ MS/MS analyses indicated an initial increase in hmdU during the RNAi time course, followed by decreased levels by day 3.5 (Figure 4C). These results indicate that hmU does increase following the loss of JGT but that it is rapidly lost. The detection of a slight increase in fdU is consistent with the loss of hmdU occurring via conversion to fU (and other analogs).

DISCUSSION

Functional studies have established base J as an important regulatory epigenetic mark in trypanosomatids. Therefore, characterization of the enzymes and mechanisms regulating the modification of specific thymidine residues within key regulatory regions along the chromosome is critical for understanding the control of trypanosome gene expression.

Several lines of evidence support the hypothesis that base J is synthesized in DNA in two separate steps, as depicted in Figure 1A. The specific localization of J within the kinetoplastid genome provides the strongest evidence that thymidine residues are modified in DNA rather than being synthesized and then incorporated during DNA replication (2, 4). In contrast, when trypanosomes are grown in the presence of hmU, it is incorporated randomly into DNA and then converted to J. Furthermore, the expression of

the mammalian base excision DNA repair enzyme SMUG1 leads to *T. brucei* cell death by excessive DNA repair, indicating that hmU is a freely accessible intermediate in J biosynthesis (35). The hmU intermediate is also detectable in trypanosome DNA by thin layer chromatography (15) and mass spectrometry. We have recently shown the ability of JBP1 to oxidize Thr residues in dsDNA *in vitro* and regulate J synthesis in a DNA replication-independent manner *in vivo* (14, 15). We now provide conclusive evidence for the two-step base J biosynthesis mechanism by identifying the base J-associated GT and demonstrating its ability to specifically modify hmU residues in dsDNA substrate *in vitro* and regulate base J synthesis *in vivo*. In contrast with the first step involving two distinct hydroxylases, JGT is the only glucosyltransferase involved in the second step of base J synthesis.

Glycosyltransferases catalyze the transfer of monosaccharide primarily from an activated sugar donor (UDP sugars) to various substrates, including carbohydrates, proteins, and glycoproteins (36). Because they are central to all biosynthetic processes involving sugars, GTs are important targets for the development of novel drugs to help combat cancer as well as microbial infections (including parasitic diseases). JGT represents a unique glucosyltransferase in that it targets chromatin bound DNA and is localized to the nucleus. To date, no other enzyme with similar substrate (DNA) specificity has been described in eukaryotes. Interestingly, a mammalian homologue of JGT called GREB1 has also been identified in the aforementioned bioinformatic study (21). GREB1 is an estrogen-responsive gene that has a largely unknown function but has been implicated in the proliferation of estrogen receptor-positive breast cancer cells (37,38). GREB1 localizes to the nucleus and has been found bound to chromatin,

presumably functioning as a transcriptional co-activator of estrogen receptor-mediated transcription (39). On the basis of the characterization of JGT presented here, an interesting possibility is that hmC present in the mammalian genome, generated by the TET enzymes, can be glucosylated by GREB1.

Glucosyltransferases that modify hmC have been described in other organisms, namely the β-glucosyltransferase present in T-even bacteriophages. T4 β-glucosyltransferase catalyzes the transfer of a glucose residue from UDP-glucose onto hmC, converting it to β-glycosyl-5-hydroxymethylcytosine (40,41). The resulting hypermodified base prevents the phage DNA from being degraded by bacterial restriction enzymes (42,43). T4 β-glucosyltransferase was the first GT x-ray crystal structure to be solved, in 1994, and is a founding member of the GT-B structural fold family of GTs (44). Despite the broad similarity in substrate, JGT is predicted to have a GT-A structural fold, suggesting that there is no ancestral kinship between these two enzymes. A better understanding of GT enzymes, including JGT, and their mechanism of action *in vivo* and *in vitro* is essential for rational drug design as well as increasing our knowledge regarding overall glycosylation machinery. The identification of a novel nuclearly localized GT in early-branching eukaryotes will provide a critical boost in this direction. Structure-function studies are underway to fully characterize the JGT enzyme. Of particular interest is clarifying the function of regions of JGT outside of the GT-A domain (see Figure 1B).

Metazoan TET proteins serially oxidize 5-methylcytosine to hmC, 5-formylcytosine, and 5-carboxycytosine, which function as stable epigenetic marks or as potential intermediates in the DNA demethylation pathway (33, 34). In a recent study, we reported a reverse-phase HPLC coupled with tandem mass spectrometry (LC-MS/MS)

method for the accurate detection of the nucleosides dJ and hmdU in *T. brucei* DNA. As shown here, this method has indicated that a low but detectable level of fdU is present along with hmdU in the genome of *T. brucei*. This provided the first indication that JBP enzymes may catalyze three similar, iterative catalytic/oxidation cycles of thymine in trypanosome DNA. The *in vivo* analysis of JGT function in *T. brucei* further supports this hypothesis. We have demonstrated previously that both JBP1 and JBP2 can stimulate *de novo* synthesis of hmU, which is subsequently converted to base J (9, 14, 24). Upon loss of JGT activity, we would expect *de novo* activities of JBP1/2 to modify the same sites as in WT cells but without having the hmU converted to J. Trypanosomatids lack DNA glycosylases (i.e. SMUG1) that recognize 5-hmU in DNA, and no activity against 5-hmU was detected in *T. brucei* extracts (35,45). Therefore, even in the absence of any propagation activity of JBP1 (because of the loss of base J), we would expect an increased level of hmU in the JGT KO genome. However, we were able to observe the transient nature of thymine modifications because base J synthesis was inhibited in a time dependent manner. During the time course of the RNAi ablation of JGT, the levels of hmU rise initially but then decrease concomitantly with an apparent small rise in fU levels. Although we have not ruled out the possibility that JGT somehow directly regulates JBP function, the results support the idea that, in the absence of JGT, sustained hmU accumulation is prevented by subsequent conversion to fU (and possibly 5-carboxyluracil) by JBP1/2.

In vitro studies of JBP1 function support the hypothesis that JBP enzymes act as TET enzymes by performing iterative hydroxylation reactions on modified substrates. Not only is JBP1 able to bind hmU-modified DNA (17), but detection of hmU formation

by JBP1 *in vitro* peaks early during the reaction but then decreases over time. Although we acknowledge that further work is needed to conclusively demonstrate JBP-stimulated conversion of hmU to fU and 5-carboxyluracil, the data thus far allow us to propose a model where JBP and JGT compete for hmU substrate in the trypanosome genome for conversion to base J or additional base analogs (fU and 5-carboxyluracil).

The ability to genetically segment the base J synthesis pathway at the hydroxylation step and the glucosylation step provides the tools to specifically study these epigenetic markers. Interestingly, 5-formylcytosine and 5-carboxycytosine can significantly reduce the kinetics of yeast RNA polymerase II transcription, suggesting that these modifications play a role in splicing and termination (46). The coregulation of base J and hmU synthesis by JGT and JBP enzymes in trypanosomatids is intriguing because the studies demonstrating the role of base J in Pol II transcription were performed by halting hmU synthesis via altering JBP function (5–7). Future experiments utilizing the JGT KO cell line will allow us to address the specific role of hmU, fU, and base J in regulating Pol II kinetics and trypanosomatid gene expression. Finally, the ability to delete JGT or both JBP thymidine hydroxylases from *T. brucei* is consistent with the non-essential nature of base J in this organism (4, 14, 15). Although deletion of either JBP1 or JBP2 in *T. cruzi* and JBP2 in *L. tarentolae* results in similar reductions in J levels as seen in the *T. brucei* mutants, attempts to delete both JBPs have been unsuccessful, leading to the idea that base J is essential in these organisms (5, 20, 47) (5,20,47). However, chemical inhibition of JBP enzymes via a specific inhibitor (dimethyloxoglycine) reduces base J to extremely low levels without affecting cell growth (15). Therefore, it is possible that JBP1/2 have additional functions outside of

base J synthesis, explaining the apparent essential nature of these genes. With the identity of JGT, we can now directly address this conundrum in *Leishmania major* and *T. cruzi*.

ACKNOWLEDGEMENTS

We thank Dr. Lance Wells and Peng Zhao for performing MS analyses of the purified recombinant JGT preparation. We also thank members of the Steve Hajduk laboratory for discussions, Tony Szempruch for help with microscopy, and Laura Cliffe and David Reynolds for critical reading of the manuscript.

REFERENCES

1. Gommers-Ampt, J., Van Leeuwen, F., de Beer, AL., Vliegenthart JF., Dizdaroglu M., Kowalak, JA., Crain, PF., and Borst, P. (1993) Beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell*, **75**, 1129-1136.
2. Van Leeuwen, F., Taylor, MC., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., Borst, P. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci* **95**, 2366-2371.
3. Van Leeuwen, F., De Kort, M., van der Marel, GA., Van Boom, JH., and Borst, P. (1998) The modified DNA base beta-D-glycosyl-hydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32P-postlabeling. *Anal. Biochem.*, **258**, 223-229.

4. Cliffe, L.J., Siegel, T.N., Marshall, M., Cross, G.A., and 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. Ekanayake, D., Minning, T., Weatherly, B., Gunasekera, K., Nilsson, D., Tarleton, R., Ochsenreiter, T., and Sabatini, R. (2011) Epigenetic Regulation of Transcription and Virulence in *Trypanosoma cruzi* by O-Linked Thymine Glucosylation of DNA. *Mol Cell Biol*, **31**, 1690-1700.

6. Ekanayake, D., and Sabatini, R. (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi.* Modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryot. Cell* **10**, 1465-1472.

7. Van Luenen, H., Farris, C., Jan, S., Genest, PA., Tripathi, P., Velds, A., Kerkhoven, R., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., Van Steensel, B., Myler, P., and Borst, P. (2013) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909-921.

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

9. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marel, G. A., van Boom, J. H., van Leeuwen, F., and Borst, P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO*, **18**, 6573– 6581.

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

11. Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* , **35**, 2107–2115.

12. Tahiliani, M., Koh, K.P., Shen,Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.

13. Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698 –1710.

14. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R.,Marshall, M., Sweeney, K., and Sabatini, R. (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.

15. Cliffe, L., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y., and Sabatini, R. (2012) JBP1 and JBP2 Proteins are Fe2+/2-Oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *J. Biol. Chem.* , **287**, 19886-19895.

16. Toaldo, C.B., Kieft, R., Dirks-Mulder, A., Sabatini, R., van Luenen, H.G. and Borst, P. (2005) A minor fraction of base J in kinetoplastid nuclear DNA is bound by the J-binding protein 1. *Mol. Biochem. Parasitol.*, **143**, 111–115.

17. Sabatini, R., Meeuwenoord, N., van Boom, J.H. and Borst, P. (2002) Recognition of base J in duplex DNA by J-binding protein. *J. Biol. Chem.*, **277**, 958–966.

18. Heidebrecht, T., Christodoulou, E., Chalmers, M. J., Jan, S., Ter Riet, B., Grover, R. K., Joosten, R. P., Littler, D., van Luenen, H., Griffin, P. R., Wentworth, P., Jr., Borst, P., and Perrakis, A. (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res.* , **39**, 5715–5728.

19. Kieft, R., Brand, Verena., Ekanayake, D., Sweeney, K., DiPaolo, C., Reznikoff, W., and Sabatini R. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* , **156**, 24-31.

20. Vainio, S., Genest, P.A., ter Riet, B., van Luenen, H., and Borst, P. (2009) 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.

21. Iyer, L., Zhang, D., Burroughs, M., and Aravind L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA *Nucleic Acids Res.* , **42**, 7635-7655.

22. Rudenko, G., Blundell, PA., Taylor, MC., Kieft, R., and Borst P. (1994) VSG gene expression site control in insect form *Trypanosoma brucei*. *EMBO*, **13**, 5470–5482.

23. Wang, Z., Morris, J., Drew, M., and Englund, P. (2000) Inhibition of Trypanosoma brucei Gene Expression by RNA Interference Using an Integratable Vector with Opposing T7 Promoters. *J Biol Chem.*, **275**, 40174-40179.

24. Van Leeuwen, F., Wijsman, E.R., Kieft, R., van der Marel, G.A., van Boom, J.H. and Borst, P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev.* , **11**, 3232–3241.

25. Wang, J., Huachuan, C., You, C., Yuan, B., Bahde, R., Gupta, S., Nishigori, C., Niedernhofer L., Brooks, P., and Wang, Y. (2012) Endogenous formation and repair of oxidatively induced G[8-5m]T intrastrand cross-link lesion. *Nucleic Acids Res.*, **40**, 7368–7374.

26. Hong, H., Cao, H., Wang, Y., and Wang Y. (2006) Identification and Quantification of a Guanine-thymine Intrastrand Crosslink Lesion Induced by Cu(II)/H2O2/Ascorbate. *Chem. Res. Toxicol.*, **19**, 614-621.

27. Hong, H., and Wang, Y. (2007) Derivatization with Girard reagent T combined with LC-MS/MS for the sensitive detection of 5-formyl-2'-deoxyuridine in cellular DNA. *Anal. Chem.*, **79**, 322-326.

28. Lairson, L., Henrissat, B., Davies, GJ., and Withers SG. . (2008) Glycosyltransferases: Structures, Functions, and Mechanisms. . *Annu. Rev. Biochem.* , **77**, 521-555

29. Liu, J., and Mushegian A. . (2003) Three monopyletic superfamilies account for the majority of the known glycosyltransferases. *Prot. Sci.* , **12**, 1418-1431.

30. Breton, C., Snajdrova, L., Jeanneau, C., Koca, J., and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology*, **16**, 29-37.

31. Keenlevside, W., Clarke, AJ., Whitfield, C (2001) Identification of residues involved in catalytic activity of the inverting glycosyl transferase WbbE from Salmonella enterica serovar borreze. *J Bacteriol.*, **183**, 77-85.

32. Garinot-Schneider, C., Lellouch, AC., and Geremia, RA. (2000) Identification of essential amino acid residues in the Sinorhizobium meliloti glucosyltransferase ExoM. *J Biol Chem.*, **6**, 31407-31413.

33. Wu, H., and Zhang Y. (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes and Dev.* , **25**, 2436-2452.

34. Ito, S., Shen, L., Dia, Q., Wu, S., Collins, L., Swenberg, J., He, C., and Zhang, Y. . (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carbosylcytosine. . *Science*, **333**, 1300-1303.

35. Ulbert, S., Cross, M., Boorstein, R., Teebor, G., and Borst, P. (2002) Expression of the human DNA glycosylase hSMUG1 in *Trypanosoma brucei* causes DNA damage and interferes with J biosynthesis. *Nucleic Acids Res.* , **30**, 3919-3926.

36. Roseman, S. (2001) Reflections on Glycobiology. *J Biol Chem.* , **276**, 41527-41542.

37. Liu, M., Wang, G., Gomez-Fernandez, C.R. and Guo, S. . (2012) GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer. *PLoS One*, **7**, 46410.

38. Ghosh, M.G., Thompson, D.A. and Weigel, R.J. . (2000) PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. . *Cancer Res.*, **60**, 6367–6375.

39. Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L. et al. . (2013) Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. . *Cell Rep.*, **3**, 342–349.

40. Volkin, E. (1954) The linkage of glucose in coliphage nucleic acids. . *J. Am. Chem. Soc.*, **76**, 5892–5893.

41. Jesaitis, M. (1956) Differences in the chemical composition of the phage nucleic acids. *Nature*, **178**, 637–641.

42. Kornberg, S.R., Zimmerman, S. B., and Kornberg, A. . (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage- infected *Escherichia coli.* . *J. Biol. Chem.* , **236**, 1487−1493.

43. Josse, J., and Kornberg, A. (1962) Glucosylation of deoxyribonucleic acid. III. α- and β-glucosyltransferases from T4- infected *Escherichia coli.* . *J. Biol. Chem.* , **237**, 1968−1976.

44. Vrielink, A., Ruger, W., Driessen, HP., and Freemont PS. . (1994) Crystal structure of the DNA modifying enzyme beta-glycosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. . *EMBO*, **13**, 3413-3422.

45. Ulbert, S., Eide, L., Seeberg, E., and Borst P. (2004) Base J, found in nuclear DNA of *Trypanosoma brucei*, is not a target for DNA glycosylases. *DNA Repair*, **3**, 145-154.

46. Kellinger, M., Song, C., Chong, J., Lu, X., He, C., and Wang, D. . (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. . *Nat. Struct. Mol. Biol.* , **19**, 831-833.

47. Genest PA., t.R.B., Dumas C., Papadopoulou B., van Luenen HGAM., Borst P. (2005) Formation of linear inverted repeat amplicons following targeting of an essential gene in *Leishmania*. *Nucleic Acids Res.* , **33**, 1699–1709.

48. van der Wel, H., Morris, H., Panico, M., Paxton, T., Dell, A., Kaplan, L., and West, C (2002) Molecular Cloning and Expression of a UDP-N-acetylglucosamine (GlcNAc):Hydroxyproline Polypeptide GlcNAc-transferase That Modifies Skp1 in the Cytoplasm of Dictyostelium. *J Biol Chem.*, **227**, 46328-46337.

FIGURES

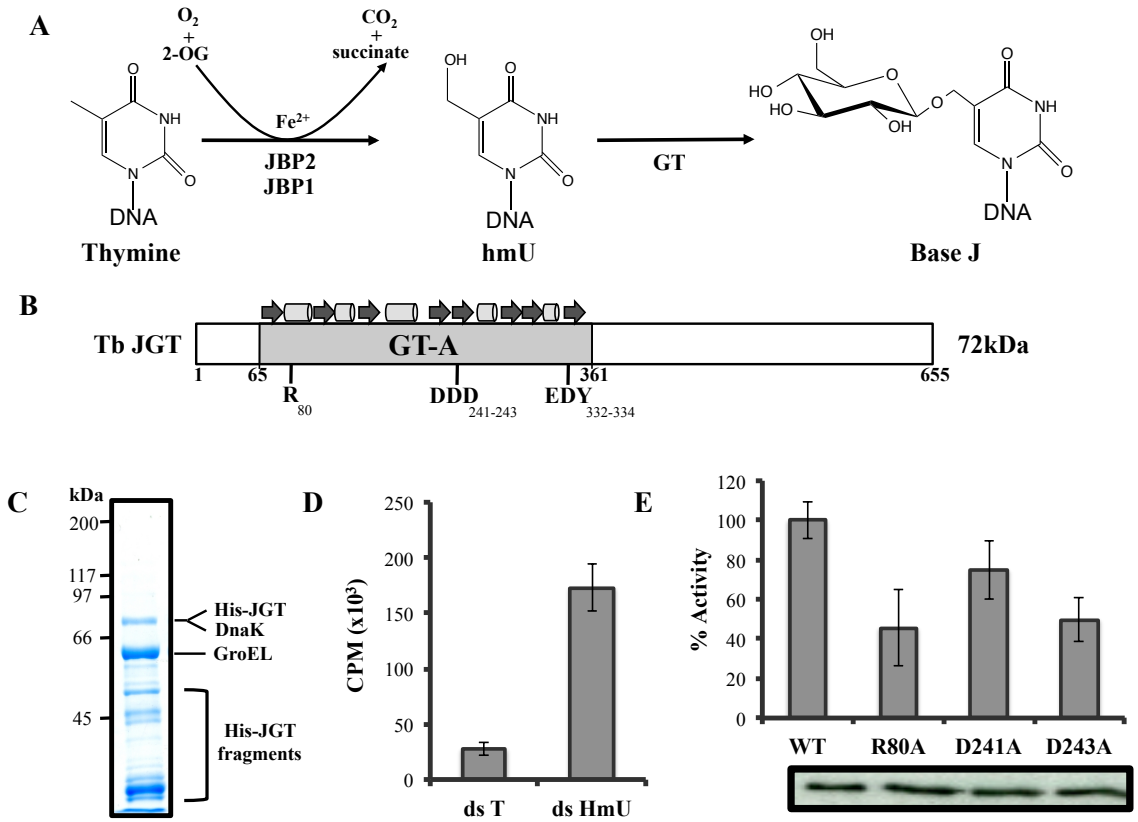


FIGURE 2.1. JGT is an hmU-DNA glucosyltransferase. A. Proposed mechanism of base J synthesis. As described in the text, base J synthesis is proposed as a two-step process in which thymines in DNA are first hydroxylated by the thymidine hydroxylases JBP1 and JBP2 to form hmU. JBP1 and JBP2 are members of the Fe^{2+}/2-oxoglutarate dioxygenase family that utilize oxygen and 2-oxoglutarate as cosubstrates and release CO2 and succinate as byproducts. The intermediate hmU is then glycosylated by a glucosyltransferase, forming base J. Presumably, the glucose donor in the second step is UDP-glucose. **B.** Schematic of the *T. brucei* (Tb) JGT protein depicting the GT domain, with the general topology shown above. Conserved residues characteristic of GT-A fold are shown below (adapted from Ref. 21). **C.** SDS-PAGE of purified JGT. Coomassie-stained gel (left panel) and anti-His Western blot analysis (right panel) are shown. LC-

MS indicates that affinity purification using Talon resin resulted in copurification of JGT and two *E. coli* molecular chaperones (DnaK and GroEL). GroEL and DnaK are common contaminants of GT enzymes following expression in *E. coli* (48). Because DnaK runs at the same location as JGT on the SDS-PAGE gel, protein concentration was measured using a combination of SDS-PAGE and a bovine serum albumin standards and anti-His Western blot analysis as described under "Experimental Procedures." The Western blot analysis indicates that the lower molecular weight proteins are degradation products of JGT. **D,** *in vitro* GT reaction. Recombinant JGT and UDP-[3H]glucose were incubated with 20-nt-long, double-stranded DNA substrate containing four hmU molecules (ds hmU) or unmodified thymidines (ds T), as described under "Experimental Procedures." CPM , counts per million, indicative of the transfer of glucose to DNA, were read for each sample. All experiments were performed in triplicate, and error bars are representative of mean ±S.D. **E.** equal amounts of recombinant WT JGT and r80a, d241a, and d243a point mutant JGT were assayed for transfer of [^{3}H]glucose from UDP-glucose to hmU-DNA as described above. The results are mean ±S.D. of triplicates and expressed as percent activity relative to the WT control reaction. Bottom panel, anti-His western blot analysis demonstrating that equal amounts of WT and mutant JGT enzyme were included in each assay.

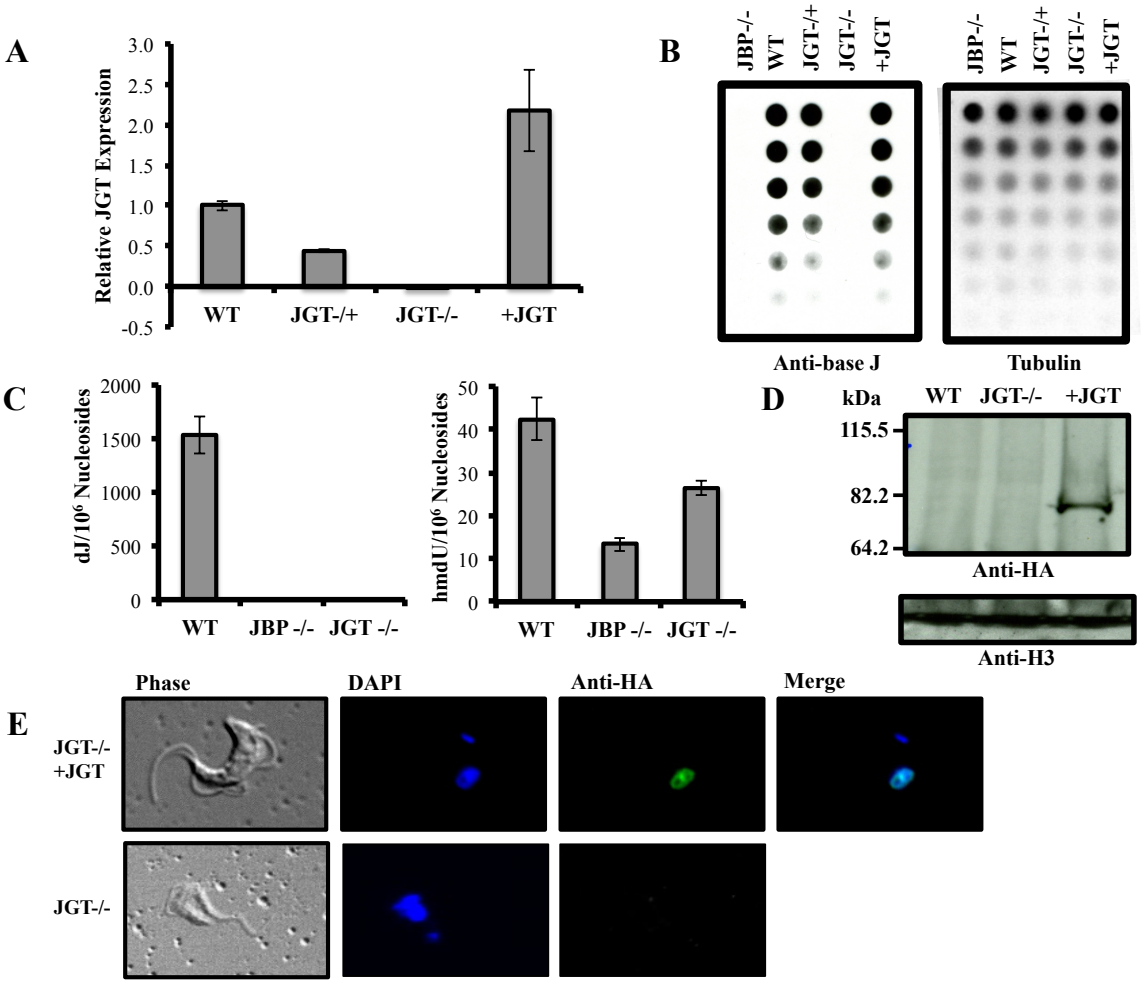


FIGURE 2.2. JGT is involved in the synthesis of base J. A. Relative expression of JGT mRNA levels. RNA was isolated from WT, JGT-/+, JGT-/-, and JGT-/- expressing an HA-tagged version of JGT (+JGT), and levels of JGT were measured by quantitative reverse transcription PCR as described under "Experimental Procedures." **B.** Anti-base J dot blot. DNA isolated from the indicated cell lines was serially diluted 2-fold and analyzed by anti-base J dot blot. The membrane was then hybridized with a radioactive tubulin probe to control for DNA loading. JBP-/- refers to the J null cell line that has all four alleles for JBP1 and JBP2 deleted from the genome (14). **C.** Quantitation of the levels of nucleosides dJ and hmdU by nano LC-MS/MS and LC-MS/MS/MS. Genomic DNA of the WT, JBP-/-, and JGT-/-cell lines was analyzed by tandem mass spectrometry

as described under “Experimental Procedures.” **D.** Anti-HA Western blot analysis of cell lysates from WT, JGT-/-, and JGT-/- cell lines expressing an HA-tagged version of JGT. **E.** Anti-HA IFA of +JGT cells. The analysis of the untransfected JGT-/- cells is provided as a negative control.

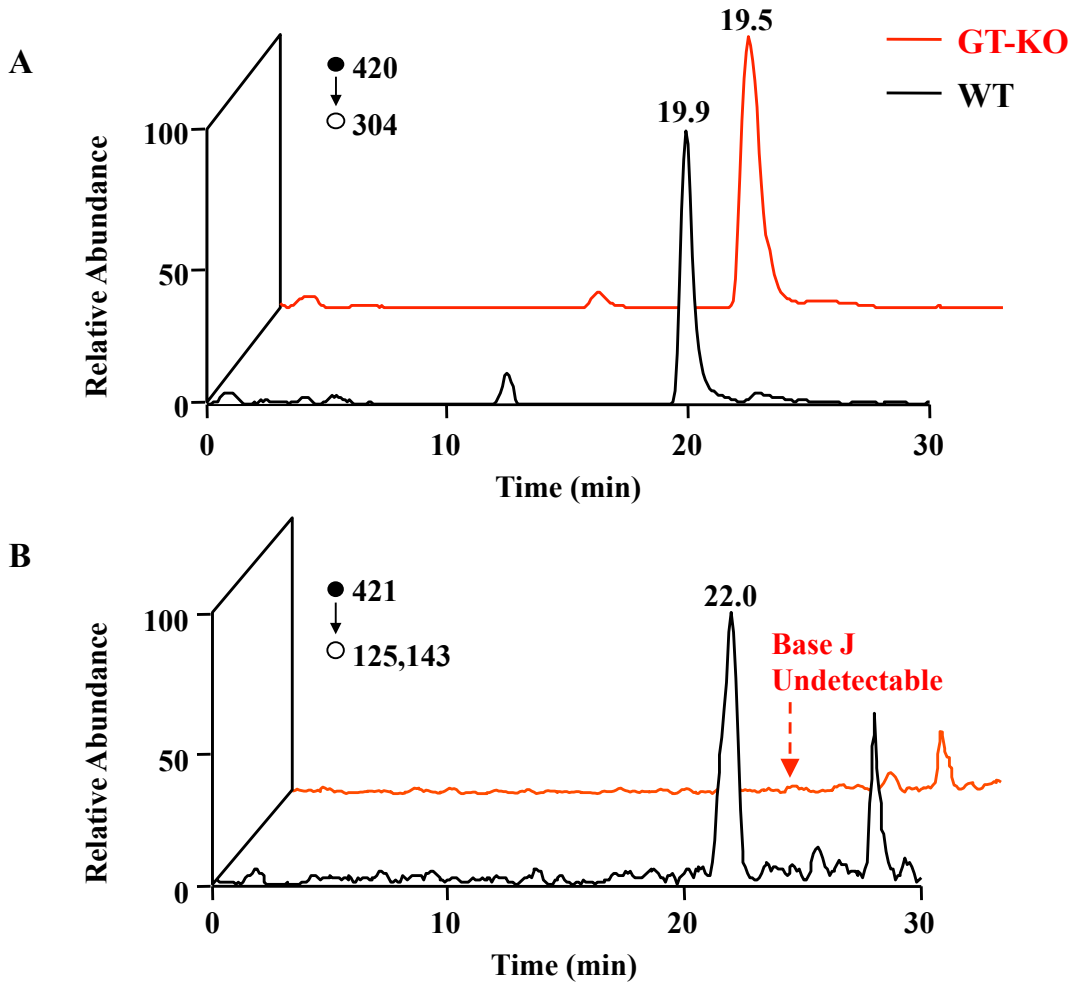


FIGURE 2.3. Representative LC-MS/MS results for the quantification of dJ. Shown are the selected ion chromatograms for monitoring the indicated transitions for the surrogate internal standard (A, glc-hmdC) and the analyte (B, dJ). The red and black lines represent samples from JGT KO and wild-type *T. brucei* cells, respectively.

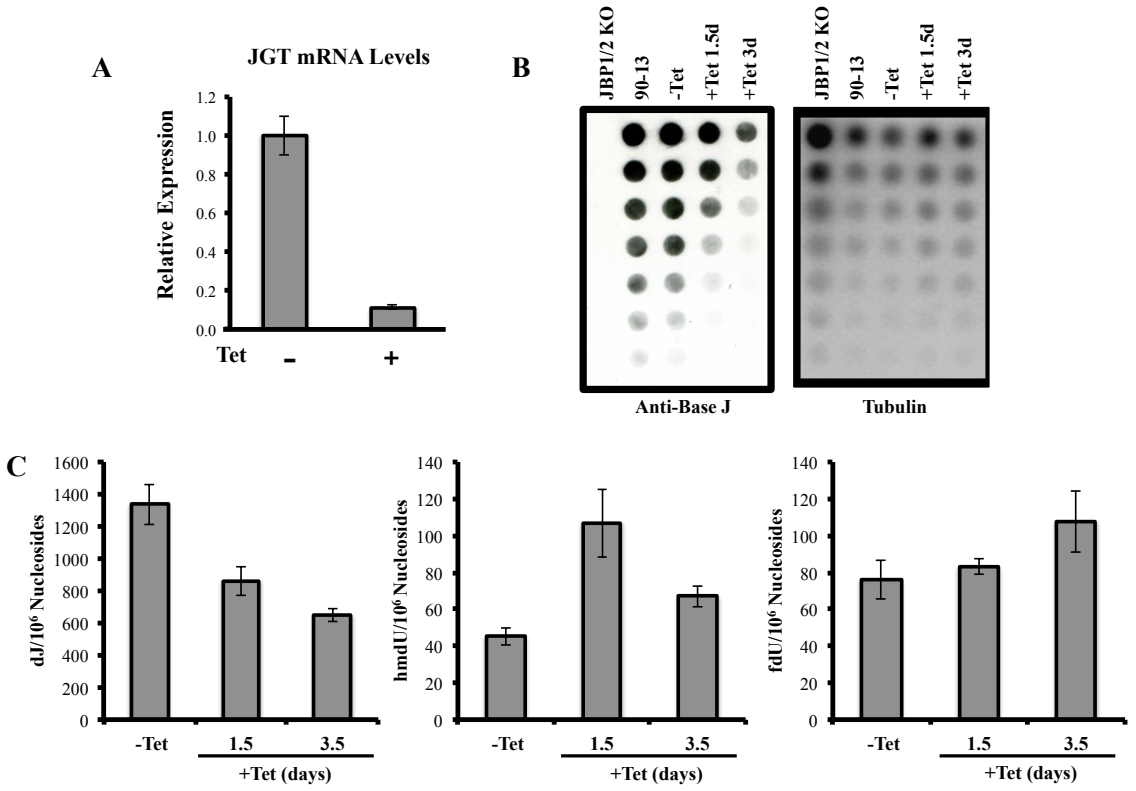


FIGURE 2.4. RNAi knockdown of JGT leads to a decrease of base J and changes in hmU and fU. A. Relative expression of JGT mRNA levels using quantitative RT-PCR in JGT RNAi cells with our without tetracycline 36 h after induction. **B.** Anti-base J dot blot. DNA isolated from the indicated cell lines was serially diluted by 2-fold and spotted onto a nitrocellulose membrane to measure base J levels as described under "Experimental Procedures." The membrane was then hybridized with a radioactive tubulin probe to control for DNA loading. 90-13, RNAi cell line untransfected with the JGT-pZJM construct; -Tet, uninduced JGT RNAi cells; +Tet, induced with tetracycline for the indicated days. **C.** Genomic DNA isolated from JGT RNAi cells uninduced (-Tet) and at the indicated days following induction with tetracycline was analyzed by LC-MS/MS as described under "Experimental Procedures." We detected a slight increase in

levels of fdU in the tetracycline-induced samples compared with the uninduced samples (day 1.5, $p = 0.5751$; day 3.5, $p = 0.1872$).

CHAPTER 3

BASE J GLUCOSYLTRANSFERASE DOES NOT REGULATE THE SEQUENCE SPECIFICITY OF J SYNTHESIS IN TRYPANOSOMATID TELOMERIC DNA[2]

---

[2]Bullard, W., Cliffe, L., Wang, P., Wang, Y., and Sabatini, R. 2015. *Mol Biochem Parasit*, **204**, 77-80. Reprinted here with premission of publisher.

ABSTRACT

Telomeric DNA of trypanosomatids possesses a modified thymine base, called base J, that is synthesized in a two-step process; the base is hydroxylated by a thymidine hydroxylase forming hydroxymethyluracil (hmU) and a glucose moiety is then attached by the J-associated glucosyltransferase (JGT). To examine the importance of JGT in modifying specific thymine in DNA, we used a *Leishmania* episome system to demonstrate that the telomeric repeat (GGGTTA) stimulates J synthesis *in vivo* while mutant telomeric sequences (GGGTTT, GGGATT, and GGGAAA) do not. Utilizing an *in vitro* GT assay we find that JGT can glycosylate hmU within any sequence with no significant change in Km or kcat, even mutant telomeric sequences that are unable to be J-modified *in vivo*. The data suggests that JGT possesses no DNA sequence specificity *in vitro*, lending support to the hypothesis that the specificity of base J synthesis is not at the level of the JGT reaction.

SHORT COMMUNICATION

Trypanosomatids, including the human pathogens *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania*, possess a unique DNA modification within their genomes known as base J (1) and (2). Base J (β-d-glucopyranosyloxymethyluracil) is a hyper-modified thymine residue predominately present in repetitive sequences, such as telomeric repeats (2,3). While the function of base J in telomeric repeats is unknown, base J is also found at chromosome-internal regions at regions flanking polycistronic transcription units called divergent strand switch regions (dSSRs) and convergent strand switch regions (cSSRs), which are sites of RNA polymerase II (Pol II) transcription initiation and termination, respectively (4). The loss of base J from these chromosome-internal sites led to alterations in transcription initiation and termination, and corresponding changes in gene expression (5-7). While it is clear that base J represents a novel epigenetic mark involved in regulating Pol II transcription and gene expression, little is understood about what regulates the specific localization of base J in the genome.

Base J is synthesized in a two-step process. First, thymine residues in the context of DNA are hydroxylated by a thymidine hydroxylase (TH) forming 5-hydroxymethyluracil (hmU). A glucose moiety is then attached to hmU by a glucosyltransferase (JGT), to form base J. Two TH enzymes, JBP1 and JBP2, have been identified in trypanosomatids (8-11). While both JBP1 and JBP2 stimulate *de novo* thymidine hydroxylation *in vivo*, the ability of JBP1 to bind J-DNA is thought to play a unique role in J propagation/maintenance (10,12-14). The simultaneous deletion of both JBP1 and JBP2 from *T. brucei* yields a cell line that is unable to synthesize base J, unless cells are fed hmU (9). Studies such as these unambiguously identified JBP1/2 as the thymidine hydroxylases catalyzing the first step of base J synthesis as well as confirm that this step is independent from the subsequent step of glucose conjugation during base J synthesis. The

glucosyltransferase involved in the second step of base J synthesis, base J associated GT (JGT), has recently been identified (15,16). Using recombinant protein we demonstrated that JGT utilizes UDP-glucose to transfer glucose to dsDNA substrates containing hmU (15). *In vivo*, deletion of both JGT alleles in *T. brucei* results in complete loss of base J synthesis in the genome (15,16). These studies further confirm the two-step mechanism of J synthesis and indicate JGT is the only glucosyltransferase involved, catalyzing the final step of base J synthesis.

Despite the recent genome-wide data sets of DNA Jaylation patterns, and elucidation of the J-biosynthetic pathway, the rules that govern the establishment of DNA J patterns in trypanosomatids remain undefined. It is unclear what determines the specific localization of base J synthesis into specific sequences in the genome. While no consensus sequence or motif is evident from the genome-wide J analysis thus far, it is clear that, for at least the telomeric repeats, there is a sequence specificity component where in the top strand (GGGTTA), only the second T is modified (17,18).

To gain systematic insight into the constraints that define endogenous Jaylation patterns, we integrated different DNA elements into episomes in *L. major*. To examine the ability of specific sequences to stimulate *de novo* J synthesis, DNA fragments were cloned into a PSP72 vector containing a hygromycin resistance gene and then transfected into *Leishmania major* and grown as episomes (Figure 3.1A). The episomes were then purified from *L. major*, digested with EcoRI and HindIII, and J synthesis assayed by anti-J IP-qPCR. This set up allows us to control for chromosomal environment and potential indirect effects such as those mediated by transcription. Furthermore, this approach allows us to measure the contributions of DNA sequence in establishing Jaylation patterns. To determine if the episome system would mimic the

specificity of J synthesis *in vivo*, sequences representing J positive and J negative regions of the *L. major* genome were cloned into the PSP72 vector and assayed for base J synthesis after growth in *L. major*. Initially, this included J positive and J negative SSRs. Approximately 1 kb genomic fragments representing a cSSR that normally contains base J (cSSR+) and a cSSR that lacks base J (cSSR-) were cloned into the PSP72 vector (6). A similar approach was employed for dSSRs that do and do not contain J *in vivo*. In both cases, only the sequences that contain J *in vivo* are able to stimulate *de novo* J synthesis when cloned in the episome (Figure 3.1B and 3.1C). Thus, even when present outside of their normal chromatin context, *i.e.* within an episome instead of within the genome, DNA sequence specificity of J synthesis is maintained. This was also seen using a similar approach in *L. tarentolae* (17).

To further address the DNA sequence specificity, we assayed the telomeric repeat sequence in the episome system. A plasmid was generated that contains six copies of the WT telomeric repeat sequence (GGGTTA) and transfected into *L. major*. As expected, WT telomeric repeat sequence (GGGTTA) was able to stimulate J synthesis (Figure 3.1D). However, mutated telomeric sequences (GGGTTT, GGGATT, or GGGAAA) were unable to support J synthesis (Figure 3.1D). These data demonstrate that some aspect of J synthesis is DNA sequence specific, at least within telomeric DNA.

Taken together, these findings indicate that DNA Jaylation is largely regulated by cis-acting sequences and is thus genetically encoded. A functional role for interplay between DNA methylation and chromatin has been well established, making it possible that chromatin could be a determining factor in establishing global J patterns in trypanosomes. While these findings do not exclude the possibility that chromatin structure or other associated proteins are crucial in

mediating local DNA Jaylation, they do show that local DNA sequence is a primary determinant of target specification for DNA Jaylation in trypanosomatids.

The current hypothesis is that the key regulatory step of J synthesis is the first step catalyzed by JBP1 and JBP2. Bypassing this first step, via feeding cells hmU, leads to J synthesis in regions of the genome that do not normally contain base J (9,19). This suggests that, regardless of where hmU is present, JGT will convert it to base J in a promiscuous (non-sequence-specific) manner. Thus, it follows that the specificity of base J localization is due to the JBP enzymes generating hmU at only specific sites throughout the genome; however, no direct evidence has confirmed this hypothesis. To address this, we tested synthetic telomeric DNA substrates in an *in vitro* GT assay to determine whether JGT can explain the sequence specificity of J synthesis *in vivo*. We took advantage of active recombinant JGT and an *in vitro* homogeneous bioluminescent UDP detection assay (UDP-Glo) that can detect the activity of glycosyltransferases that utilize UDP-sugars and release UDP as a product (see materials and methods). Using this assay we show that JGT is specific for hmU-containing dsDNA substrates and has no activity with unmodified dsDNA substrates (Figure 3.2B), consistent with assays directly measuring the transfer of glucose to DNA (Figure 3.2A), validating our use of UDP-Glo for further *in vitro* GT experiments.

To address sequence specificity, recombinant JGT was incubated with dsDNA substrates that correspond with the WT telomeric sequence (GGGTTA) as well as mutated sequences (GGGTTT and GGGATT) with an hmU positioned on the second T in the G-rich strand (WT_2, TTT, ATT; Table 3.1). Michaelis-Menten curves were generated for each substrate (Figure 3.2B and data not shown) and hmU DNA kinetic parameters were determined (Table 3.1). We found no significant difference in either affinity (Km), p value > 0.05 or turnover (kcat), p value > 0.01

of the JGT when incubated with WT or mutated telomeric sequences. We also find no significant difference in affinity or turnover for WT telomeric sequences with hmU modification at different positions, including the first T on the G-strand that is not modified *in vivo* (WT_1), or for hmU within a random (non-telomeric) DNA sequence (Random; Table 3.1). These data indicate that JGT is DNA sequence non-specific and thus, the enzyme presumably does not contribute to the telomeric DNA sequence specificity of J synthesis we characterized *in vivo*.

These studies pave the way for the elucidation of the underlying molecular mechanisms involved in regulating J synthesis *in vivo*. While the results described here strongly support the importance of the first step of J synthesis, further work is required to fully understand the formation of hmU on specific sequences by the JBPs. Similar analysis of JBP catalysis, as we performed here for JGT, will help to shed light on primary DNA sequence requirements for hmU formation. *In vitro* analysis of the sequence specificity of the JBPs has, however, proven to be quite difficult. Currently, we are only able to express recombinant protein for JBP1. All attempts to express JBP2, the key *de novo* J synthesis enzyme, have failed. While we have had success demonstrating dioxygenase activity of JBP1 *in vitro*, the reaction is extremely inefficient (20). Until we have a clear and robust assay for both hydroxylases, we are unable to fully characterize the contribution of this first step in determining the sequence specificity of J synthesis.

MATERIALS AND METHODS

Leishmania major Cell Culture and Generation of Transfectants

L. major promastigotes were cultured in MI99 media supplemented with 10% heat inactivated fetal bovine serum as described previously (21).

Generation of Plasmids

A 1560bp fragment from a cSSR on chromosome 28 that lacks base J (region 590974-592533kbp), a 1920bp fragment from a cSSR on chromosome 5 that possesses base J (region 361977-363896kbp), a 993bp dSSR on chromosome 15 that lacks base J (region 87359-88351kbp), and a 600bp fragment from a dSSR on chromosome 1 that possesses base J (region 78216-78815kbp) were generated by PCR amplification from *L. major* genomic DNA. Base J localization within the *L. major* genome is based upon J IP-seq (6) and confirmation by J IP-PCR analysis (7). Primers used to amplify the SSRs are available upon request. Fragments were then cloned into a TA 2.1 vector, grown in bacteria, and then extracted from bacteria using a Qiagen Midi Prep Kit and digested with HindIII and XbaI and gel purified. The insert DNA segment was then subcloned into HindIII and XbaI digested *Leishmania* expression vector PSP7Z α-neo-α. The 6x repeated telomeric sequences were synthesized by GENEART and cloned in the DNA vector pMA-RQ before being subcloned into HindIII and XbaI digested PSP72 α-neo-α.

Transfection of Plasmids into *L. major* and Isolation of Plasmids

Transfections of *L. major* promastigotes were essentially carried out as previously described (21). After transfection, cells were grown overnight prior to selection with neomycin (20μg/mL). The *L. major* cells were kept under selection to obtain stable cell lines. Isolation of the plasmids from *L. major* transfectants was performed using Qiagen midi prep DNA kit.

Quantification of J Levels on Plasmids

To quantify the levels of base J on plasmids extracted from *Leishmania*, DNA was digested with EcoRI and HindIII and utilized in anti-J immunoprecipitation (IP) and quantitative

PCR (qPCR) analysis as described previously (4). %IP was calculated relative to input DNA. Quantification was performed on an iCycler with an iQ5 multicolor real-time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). The reaction mixture contained 5 pmol forward and reverse primers to the α-Tubulin IR, 2x iQ SYBR green super mix (Bio-Rad Laboratories, Hercules, CA, USA), and 2 μL of template DNA.

Preparation of Recombinant JGT

Expression and purification of the N-terminal 10xHis-tagged Tb-JGT was performed with BL21-CodonPlus (DE3)- RIL competent cells as previously described with minimal changes (15). JGT was eluted from the Talon resin with increasing concentrations of imidazole elution buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 10mM β-mercaptoethanol, 5% glycerol, 10 -250 mM imidazole). Imidazole fractions 120 -250 mM were pooled and diluted 1:5 in dilution buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 10% Glycerol) and concentrated with 10,000 MWCO Centricon and visualized by blue silver-stained SDS PAGE and anti-His Western blot. The amount of recombinant JGT was estimated based on mass spectrometric analysis and visualization of Coomassie stained gel with BSA standards as previously described (15).

Generation of DNA Substrates for GT Assay

Unmodified oligodeoxyribonucleotides used in this study were purchased from Integrated DNA Technologies (Coralville, IA, USA). The 36mer hmU-bearing oligo was synthesized by TriLink Biotechnologies (San Diego, CA) and possesses the sequence CTATACCTCCTCAACTTCT**T**GATCACCGTCTCCGGCG with an hmU at the position of the bold T (hmU substrate). A matching unmodified 36mer oligo was also synthesized (T substrate).

The 15mer hmU-bearing oligos used to generate the substrates in Table I, were synthesized following previously published procedures and the identity of the modified oligodeoxyribonucleotide was confirmed by electrospray ionization-mass spectrometry (ESI-MS) and tandem MS (MS/MS) analysis (22). DNA duplexes were prepared by annealing complementary oligodeoxyribonucleotides in annealing buffer [100 mM potassium acetate; 30 mM HEPES, pH 7.5] by boiling for 10 min and allowing to cool overnight.

Radiolabeled *in vitro* Glucosylation Assay

A standard glucosylation assay consisted of 88μM UDP-[^{3}H]glucose (specific activity 45Ci/mmol), 50μM recombinant JGT and 50μM DNA in GT buffer [50mM potassium acetate, 20mM Tris acetate, 10mM manganese acetate, and 1mM DTT (pH 7.9)]. Reaction mixtures were incubated at 37°C for 30 minutes and the reaction was stopped by addition of 10μL of 400μM cold UDP-glucose and flash freezing in liquid nitrogen for processing. The reaction mixtures were thawed and applied immediately to a 2.5cm DE81 membrane (GE Healthcare, catalog no. 3658-325) under air pressure using a vacuum manifold. The filters were then washed in 3 x 2mL of 0.2M ammonium bicarbonate, 3 x 2mL of water, and 3 x 2mL of 100% ethanol. Membranes were air-dried and placed in scintillation vials containing 5mL of scintillation fluid. The solution was mixed, and tritium incorporation was measured for 1 minute.

UDP-Glo *in vitro* Glucosylation Assay

A standard glucosylation assay consisted of 1 mM UDP-glucose, 50 μM recombinant JGT and 50μM DNA in GT buffer [50 mM potassium acetate, 20 mM Tris acetate, 10mM manganese acetate, and 1 mM DTT (pH 7.9)]. Reaction mixtures were incubated at 37°C for 30

minutes. UDP-Glo Glycosyltransferase Assay (Promega, WI, USA) was used to analyze the amount of GT activity (cleaved UDP) according to the manufacturer's instructions. Reactions were stopped by addition of equal volume of UDP-Glo/Nucleotide Detection Reagent solution. Reactions were then transferred into a 384-well plate and incubated at room temperature for 1 hour. Plates were read using a Promega Glo-Max multi detection system luminometer plate reader.

Kinetic Assays

To determine the optimal enzyme concentration and reaction time, varying concentrations of recombinant JGT was incubated with a fixed concentration of DNA substrate and UDP-glucose. The glucosylation reaction remained linear with enzyme concentration between 5 and 50μM and 1-7 min. From these data, we chose an optimal enzyme concentration of 18μM and reaction time of 3 minutes at 37°C. Initial velocity studies with 5-hmU containing DNA substrates were performed by varying the concentration of DNA (0.19-25μM) while the UDP-glucose and JGT concentrations were kept at 1mM and 18μM respectively in the reaction mixture. Reactions were stopped by addition of equal volume of UDP-Glo/Nucleotide Detection Reagent solution. Reactions were then transferred into a 384-well plate and incubated at room temperature for 1 hour. Plates were read using a Promega Glo-Max multi detection system luminometer plate reader. The amount of product formed was estimated based on a standard curve of varying amounts of UDP. All kinetic data generated was then analyzed using GraphPad Prism and fit to the Michaelis-Menten equation. Kinetic constants are reported for each DNA substrate along with their standard error. Statistical analysis was performed for Km and kcat values using a T test; P-values are as follows for Km data: $WT_2 = 1.000$, $WT_1 = 0.983$, $WT_3 =$

1.000, TTT = 0.593, ATT = 0.148, Random = 0.172; P-values for kcat data: WT_2 = 1.000, WT_1 = 0.365, WT_3 = 0.012, TTT = 0.018, ATT = 0.116, Random = 0.035.

ACKNOWLEDGEMENTS

We thank Dr. Lance Wells and Rob Bridger for performing MS analyses of the purified recombinant JGT preparation. Work performed in the Wells laboratory was supported by a P41 grant from NIGMS/NIH (P41GM103490). We would also like to thank Stephanie Halmo for help with kinetic assays and analyses, and Jessica Lopes da Rosa-Spiegler, David Reynolds, and Rudo Kieft for critical reading of the manuscript. This work was supported by National Institute of Health grant (RO1AI109108).

REFERENCES

1. Gommers-Ampt, J., Van Leeuwen, F., de Beer, AL., Vliegenthart JF., Dizdaroglu M., Kowalak, JA., Crain, PF., and Borst, P. (1993) Beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell*, **75**, 1129-1136.
2. Van Leeuwen, F., Taylor, MC., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., Borst, P. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci* **95**, 2366-2371.
3. Van Leeuwen, F., De Kort, M., van der Marel, GA., Van Boom, JH., and Borst, P. (1998) The modified DNA base beta-D-glycosyl-hydroxymethyluracil confers resistance to

micrococcal nuclease and is incompletely recovered by 32P-postlabeling. *Anal. Biochem.*, **258**, 223-229.

4. Cliffe, L.J., Siegel, T.N., Marshall, M., Cross, G.A., and 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. Ekanayake, D., and Sabatini, R. (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi.* Modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryot. Cell* **10**, 1465-1472.
6. Van Luenen, H., Farris, C., Jan, S., Genest, PA., Tripathi, P., Velds, A., Kerkhoven, R., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., Van Steensel, B., Myler, P., and Borst, P. (2013) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909-921.
7. Reynolds, D., Cliffe, L., Fo ̈rstner, K., Hon, C., Siegel, N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res*, **42**, 9717-9729.
8. Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* , **35**, 2107–2115.

9. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R.,Marshall, M., Sweeney, K., and Sabatini, R. (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.

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

11. Vainio, S., Genest, P.A., ter Riet, B., van Luenen, H., and Borst, P. (2009) 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.

12. Toaldo, C.B., Kieft, R., Dirks-Mulder, A., Sabatini, R., van Luenen, H.G. and Borst, P. (2005) A minor fraction of base J in kinetoplastid nuclear DNA is bound by the J-binding protein 1. *Mol. Biochem. Parasitol.*, **143**, 111–115.

13. Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I. and Borst, P. (2002) J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol. Microbiol.*, **46**, 37-47.

14. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marel, G. A., van Boom, J. H., van Leeuwen, F., and Borst, P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO*, **18**, 6573– 6581.

15. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014) Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J Biol Chem.*, **289**, 20273-20282.

16. Sekara, A., Merritta, C., Baugha, L., Stuarta, K., and Myler, P. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, **196**, 9-11.

17. Genest, P., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H., Korlach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P., and Borst, P. (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res*, **43**, 2102-2215.

18. Van Leeuwen, F., Wijsman, E.R., Kuyl-Yeheskiely, E., Van der Marel, G.A., Van Boom, J.H. and Borst, P. (1996) The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res*, **24**, 2476–2482.

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

20. Cliffe, L., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y., and Sabatini, R. (2012) JBP1 and JBP2 Proteins are Fe2+/2-Oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *J. Biol. Chem.* , **287**, 19886-19895.

21. Kapler, G.M., Coburn, C.M. and Beverley, S.M. (1990) Stable transfection of the human parasite *Leishmania* major delineates a 30-kilobase region sufficient for extrachromosomal replication and expression. *Mol Cell Biol*, **10**, 1084-1094.

22. Wang, P., Williams, R. T., Guerrero, C. R., Ji, D., and Wang, Y. (2014) Fragmentation of electrospray-produced deprotonated ions of oligodeoxyribonucleotides containing an alkylated or oxidized thymidine. *J Am Soc Mass Spectrom*, **25**, 1167-1176.

TABLES

Substrate	Sequence	kcat (s^{-1})	Km (μM)	kcat/Km (μM^{-1} s^{-1})
WT_2	5'-TTAGGGT**T**AGGGTTA-3' 3'-AATCCCAATCCCAAT-5'	0.485 ± 0.022	0.519 ± 0.109	0.933
WT_1	5'-TTAGGG**T**TAGGGTTA-3' 3'-AATCCCAATCCCAAT-5'	0.451 ± 0.025	0.523 ± 0.137	0.862
WT_3	5'-TTAGGGTTAGGGTTA-3' 3'-AATCCCAA**T**CCCAAT-5'	0.675 ± 0.037	0.519 ± 0.134	1.302
TTT	5'-TTTGGGT**T**TGGGTTT-3' 3'-AAACCCAAACCCAAA-5'	0.703 ± 0.052	0.412 ± 0.149	1.705
ATT	5'-ATTGGGA**T**TGGGATT-3' 3'-TAACCCTAACCCTAA-5'	0.566 ± 0.034	0.267 ± 0.089	2.116
Random	5'-GTACGAG**T**CGAGTCA-3' 3'-CATGCTCAGCTCAGT-5'	0.580 ± 0.021	0.313 ± 0.059	1.854

TABLE 3.1. Oligonucleotides used as substrates for glucosyltransferase assay. Bold T indicates the position of an hmU modification. Glucosylation reactions were conducted with hmU DNA substrate concentration of 0, 0.195, 0.391, 0.781, 1.563, 3.125, 6.25, 12.5, and 25μM and fixed enzyme and UDP-glucose concentrations of 18μM and 1mM, respectively. UDP cleaved is plotted vs. substrate concentration and nonlinear regression was performed to determine kinetic parameters.

FIGURES

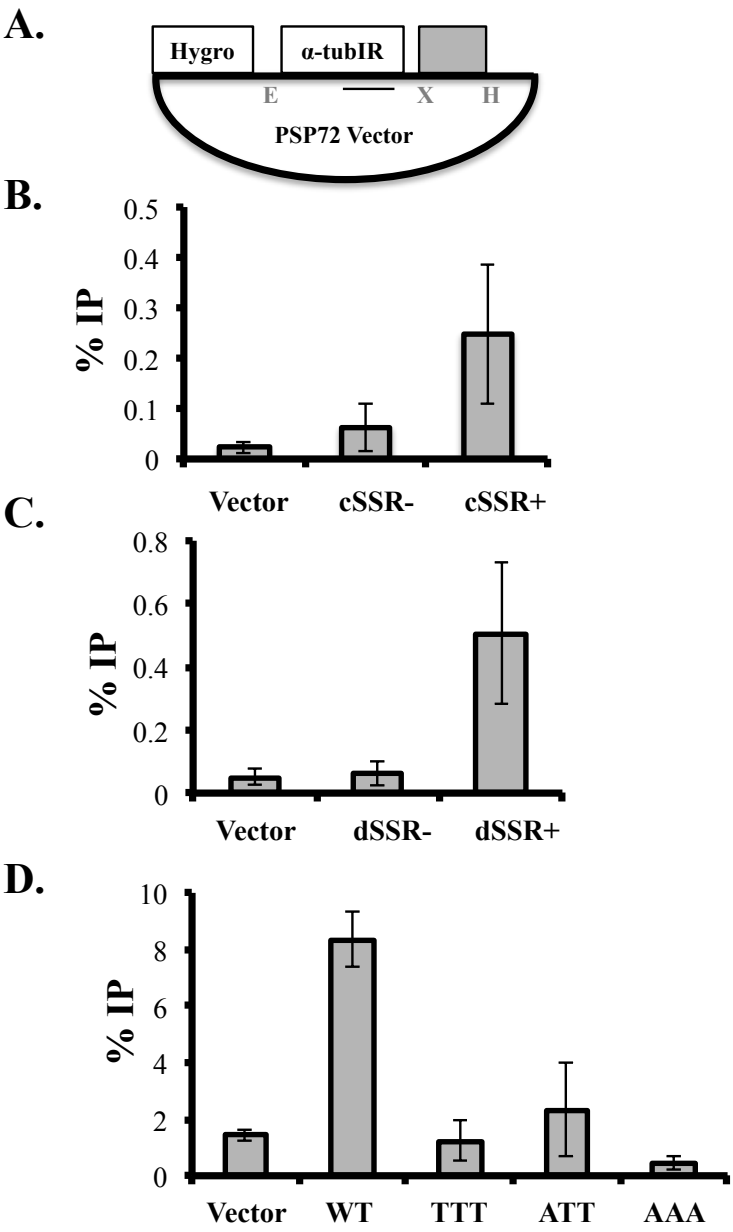


FIGURE. 3.1. Specific DNA sequences promote base J synthesis *in vivo*.

A. Schematic of the plasmids containing strand switch regions and telomeric repeats. Fragments, indicated by the grey box, corresponding to ~1 kb regions from SSRs and telomeric repeats were cloned into the XhoI (X) and HindIII (H) restriction site of the PSP72 Vector before transfection

into wild type *L. major*. The hygromycin resistance gene (Hyg) proves a selectable marker after transfection. Plasmids were digested with EcoRI (E) and HindIII (H) and J levels stimulated by the cloned DNA fragment was determined by anti-J IP-qPCR as described in Supplementary materials and methods. Solid line below tub IR indicates region amplified in qPCR. **B-D.** The percent IP of J-containing DNA from an empty PSP72 vector, and PSP72 vector containing a SSR fragment and telomeric repeats. %IP was calculated relative to input DNA. **B.** Convergent SSR that lacks base J (cSSR-) or contains base J (cSSR+) in the normal genomic context. **C.** Divergent SSR that lacks base J (dSSR-) or contains base J (dSSR+) in the normal genomic context. **D.** % J DNA IP resulting from 6copies of the wild type GGGTTA telomeric sequence; GGGTTT(TTT), GGGATT(ATT),GGGAAA(AAA). Experiments were performed in triplicate and error bars are representative of standard error.

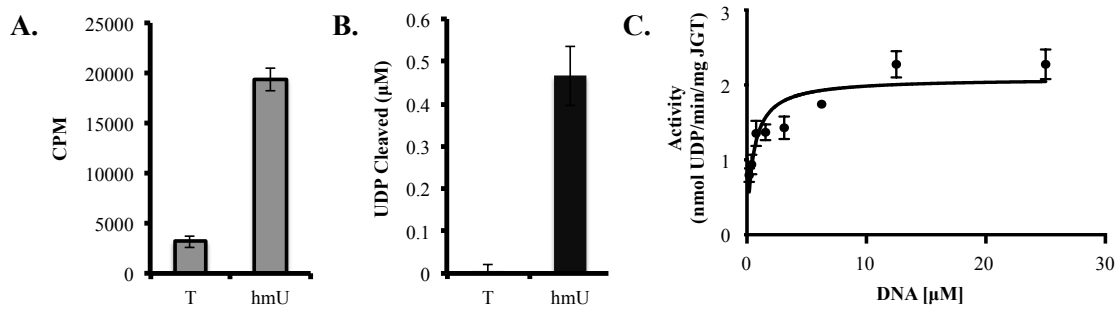


FIGURE 3.2. JGT is specific for hmU containing DNA but does not possess sequence specificity. A. Radiolabeled *in vitro* glucosyltransferase reaction. Recombinant JGT and UDP-[^{3}H] glucose was incubated with 36nt long dsDNA substrates listed in Supplementary material and methods containing one hmU modification (hmU) or unmodified dsDNA (T). CPM (counts per minute), indicative of the transfer of glucose to DNA, were read for each sample. Experiment was performed in triplicate and error bars are representative of standard error. **B.** UDP-Glo in vitro glucosyltransferase reaction. Recombinant JGT was incubated with 36nt long dsDNA substrates listed in Table 1 containing one hmU modification (hmU) or unmodified dsDNA (T). The amount of UDP Cleaved, indicative of the transfer of glucose to DNA, was estimated from a standard curve of UDP for each sample. Experiment was performed in triplicate and error bars are representative of standard error. **C.** Representative substrate−velocity curve of recombinant JGT. Recombinant JGT activity with the 15nt long hmU-containing ds DNA substrate (WT substrate) listed in Table 1. Glucosylation reactions were conducted with hmU DNA substrate concentrations of 0, 0.195, 0.391, 0.781, 1.563, 3.125, 6.25, 12.5, and 25µM and fixed enzyme and UDP-glucose concentrations of 18µM and 1mM, respectively. Kinetic experiments were performed in triplicate and error bars are representative of standard error. UDP cleaved is plotted vs. substrate concentration, and nonlinear regression was performed to determine kinetic parameters.

CHAPTER 4

A METHOD FOR THE EFFICIENT AND SELECTIVE IDENTIFICATION OF 5-HYDROXYMETHYLURACIL IN GENOMIC DNA[3]

[3]Bullard, W., Kieft, R., and Sabatini, R. 2016. *Submitted to Nucleic Acids Research.*

ABSTRACT

Recently, 5-hydroxymethyluracil (5hmU) was identified in mammalian genomic DNA as an oxidative product of thymine by the ten-eleven translocation (TET) proteins. While the biological role of this modification remains unclear, identifying its genomic location will assist in elucidating function. Here we present a rapid and robust method to selectively tag and enrich genomic regions containing 5hmU. This method involves the selective glucosylation of 5hmU residues by the base J glucosyltransferase from trypanosomes creating glucosylhydroxymethyluracil (base J). The base J can then be efficiently and selectively pulled down by antibodies against base J or by J-binding protein 1. DNA that is enriched is suitable for analysis by quantitative PCR or sequencing. We utilized this tagging reaction to provide proof of concept for the enrichment of hmU containing DNA from a pool that contains modified and unmodified DNA. Furthermore, we demonstrate that the base J pull-down assay identifies hmU at specific regions of the trypanosome genome involved in transcriptional repression. The method described here will allow for a greater understanding of the functional role and dynamics of hmU in biology.

INTRODUCTION

5-Methylcytosine (5mC) is an important epigenetic mark within the mammalian genome, regulating gene expression during many biological processes. It is generally considered a repressive mark as the presence of methylated cytosines at a promoter region is directly connected with transcriptional repression of a gene (1). The presence of 5mC at particular regions of the genome is a dynamic and reversible process. The dynamic nature of DNA methylation is crucial for regulation of genes during development (2). 5mC can be removed through a passive demethylation mechanism by loss through cell replication, or through an active demethylation pathway catalyzed by the ten-eleven translocation (TET) enzymes (3). The TET enzymes are mC hydroxylases, which catalyze the iterative oxidation of mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (4-6). 5fC and 5caC can then be excised by a thymine DNA glycosylase (TDG) followed by base excision repair (BER) to replace mC with an unmodified cytosine (C) (6,7). Deletion of the TET proteins lead to reduced 5hmC levels in mammalian genomes and defects in gene expression patterns. For example, mice deficient in individual TET enzymes display embryonic abnormalities (8-12) and combined deficiency of all three TET enzymes in mouse embryonic stem cells (mESCs) completely depletes hmC levels, impairs differentiation and does not support embryogenesis (13). In addition, knockdown of TET1 in embryonic stem cells (ESCs) leads to a loss of 5hmC at specific promoters and a corresponding increase in 5mC at transcriptional start sites of genes regulated by TET1 (14). In conjunction with interacting proteins, the TET oxidation products hmC, fC, and caC may themselves have unique epigenetic regulatory functions, presumably via regulating gene transcription (15).

Studies have recently revealed that the TET enzymes not only oxidize 5mC, but are also capable of oxidizing thymidine (T) to form hydroxymethyluracil (5hmU) (16). Isotope tracing and quantitative mass spectrometry studies indicate the majority of hmU within mESCs is produced by the mammalian TET proteins with very little 5hmU generated through 5hmC deamination or radical oxygen species (ROS) (16). Thus, the majority of hmU in the genome is matched (hmU:A) and not mismatched (hmU:G). Furthermore they found that the levels of hmU change throughout mESC differentiation, suggesting that like the other TET generated DNA modifications, hmU may serve as an important epigenetic mark (16). Presence of hmU in DNA has been shown to perturb DNA-protein interactions and transcription factor binding *in vitro* (17). To further explore the function of hmU it will be important to localize hmU derived from TET hydroxylation. Therefore, an efficient and robust method for determining the position of hmU within eukaryotic genomes is required.

New techniques were recently developed for determining the genome-wide distribution of 5hmC utilizing T4 phage β-Glucosyltransferase (βGT) to glucosylate 5hmC, which can be enriched by a specific protein (e.g. JBP1) or can be chemically modified further with biotin labeling allowing efficient pull-down using streptavidin (18-22). The addition of the glucose moiety also allows single-base-resolution (SMRT) sequencing methods for mapping hmC (23). While 5hmC and 5hmU are structurally similar, βGT is unable to modify matched hmU:A (24). Matched hmU, however, is present in trypanosomatid genomic DNA is converted to β-D-glucosyl-hydroxymethyluridine (base J), by the base J specific glycosyltransferase (JGT) (25,26) (Figure 4.1A). We have demonstrated that recombinant JGT glucosylates hmU in a DNA sequence independent manner (27), prompting us to design a similar strategy for selectively labeling hmU and mapping its location within mammalian genomes (Figure 4.1B). In the first

scheme, JGT is utilized to install a modified N_3-glucose into the hydroxyl group of hmU followed by incorporation of the biotin linker through click chemistry and capture of hmU-containing DNA fragments via streptavidin pull down. Alternatively, the hmU-containing DNA fragments can be captured by immunoprecipitation using the base J specific antibody (28-31) or by using the base J binding protein (JBP1) (18-20,31,32). The enriched fragments can be applied to deep sequencing to map the location of hmU. The resulting glucosylated hmU (base J) could also be directly mapped using SMRT-sequencing (33).

Base J has been shown to regulate transcription initiation and termination at specific sites along the trypanosomatid genome (30,31,34-36). Specific thymines in the genome are somehow targeted and modified to base J in a two-step reaction (37). The thymine base in DNA is first oxidized by one of the thymidine hydroxylases (TH), JBP1 and JBP2, to generate hmU (38,39). JGT then transfers glucose from UDP-glucose onto hmU forming base J (25). While both JBP1 and JBP2 contain a TH domain at the N-terminus that has led to the designation of these enzymes as belonging to the TET/JBP subfamily of dioxygenases (4,38-42), JBP1 has a novel C-terminal domain that allows it to bind J-DNA (32,43-45). While both JBP1 and JBP2 stimulate *de novo* thymidine hydroxylation *in vivo*, the ability of JBP1 to bind J-DNA allows the enzyme to stimulate additional hmU synthesis (and J) in the genome (39,46,47). Analysis of J synthesis *in vivo* and JGT function *in vitro* has indicated that JGT is a DNA sequence non-specific enzyme and anywhere hmU is located in the trypanosome genome it will be converted to J (27,39). Therefore, it is believed that the JBP enzymes regulate the specific localization of base J in the trypanosomatid genome. In order to test this hypothesis and the new hmU mapping procedure, we utilized the glucose tagging/J enrichment method to map the location of hmU in the

Trypanosoma brucei genome. We demonstrate that hmU synthesized by the JBP enzymes is localized at specific sites of the genome where base J is known to function.

MATERIALS AND METHODS

Trypanosome Cell Culture

The bloodstream form *T. brucei* cell lines were cultured in HMI9 medium supplemented with 10% heat-inactivated fetal bovine serum and 10% serum plus as described previously (46).

Isolation of Genomic DNA

T. brucei genomic DNA was isolated as described previously (29). Genomic mESC DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit following the manufacturer's instructions (Qiagen). All buffers were supplemented with the antioxidants 3,5-di-tert-butyl-4-hydroxytoluene, deferoxamine mesylate salt, and tetrahydrouridine at a concentration of 200μM to reduce DNA deamination and oxidation (48). Amount and purity of DNA was determined by using Nano-Drop® ND-1000 Spectrophotometer.

Preparation of Recombinant JGT

Expression and purification of the N-terminal 10xHis-tagged *T. brucei*-JGT was performed with BL21-CodonPlus (DE3)-RIL competent cells as previously described (25) with minimal changes. JGT was eluted from the Talon resin with imidazole elution buffer [50mM HEPES, pH 7.5, 300mM NaCl, 10mM β-mercaptoethanol, 5% glycerol, 300mM imidazole]. The eluted protein was diluted 1:5 in dilution buffer [50mM HEPES, pH 7.5, 50mM NaCl, 10% Glycerol], concentrated with 10,000 MWCO Centricon and visualized by blue silver-stained SDS PAGE and anti-His Western blot.

Generation of DNA Substrates

Unmodified oligonucleotides used in this study were purchased from Invitrogen. The 15bp hmU and hmC-containing oligos used in the gel-shift assay were synthesized by TriLink Biotechnologies (San Diego, CA) and possess the sequence TTAGGGT**X**AGGGTTA with either an hmU or hmC at the position of the bold X. The 36bp hmU and hmC-containing oligos used in the filter binding assays were synthesized by TriLink Biotechnologies (San Diego, CA) and possess the sequence CTATACCTCCTCAACTTC**X**GATCACCGTCTCCGGCG with either an hmU or hmC at the position of the bold X. DNA duplexes were prepared by annealing complementary oligos in annealing buffer [100mM potassium acetate, 30mM HEPES (pH 7.5)] by boiling for 10 minutes and allowing to cool overnight.

mC, hmC, and hmU modified and unmodified 75bp DNA substrates used in the anti-base J immunoprecipitations were generated using PCR. Modified deoxynucleotide triphosphates (dNTPs) were purchased from TriLink Biotechnologies and each modified dNTP completely replaced its respective unmodified dNTP in a PCR reaction mix. PCR reactions were performed using Promega GoTaq® DNA Polymerase (Promega in Green GoTaq® Reaction Buffer (Promega) using the following thermocycling conditions: 95°C 30 seconds, (95°C 15 seconds, 68°C 15 seconds) for 30 cycles, and 68°C for 5 minutes. Template DNA sequence possessed the sequence 5'-GCTATCACAGTCCTGCGCTGAGATACGAGTTGCTGCCTTGGTGCACTTAGAGGTCATGAGAAGGTTTACTGCCCG-3' with underlined portions representing the primer annealing sites.

***in vitro* Glucosyltransferase Reactions**

Filter Binding Assay. JGT glucosylation reaction consisted of 100μM UDP-[^{3}H]glucose, ~0.05μM recombinant JGT, and 100μM DNA in JGT reaction buffer [50mM potassium acetate, 20mM Tris acetate, 10mM manganese acetate, and 1mM DTT (pH 7.9)] in a 50μL reaction volume. Reactions were incubated at 25°C for 1 hour. βGT reactions consisted of 100μM UDP-[^{3}H]glucose, 5 units recombinant βGT (New England Biolabs), and 100μM DNA in T4 βGT reaction buffer [50mM potassium acetate, 20mM Tris acetate, 10mM magnesium acetate, and 1mM DTT (pH 7.9)] in a 50μL reaction volume. Reactions were incubated at 25°C for 1 hour.

Reaction mixtures were applied directly to a 2.5cm DE81 membrane (GE Healthcare) under vacuum. The filters were then washed 3 times with 2mL 0.2M ammonium bicarbonate, 3 times with water, and 3 times in 100% ethanol. Membranes were air-dried and placed in scintillation vials containing 5mL of scintillation fluid. The solution was mixed, and tritium incorporation was measured for 1 minute. Reaction values were corrected for non-specific binding of UDP-[^{3}H]glucose to the membrane. Background values were determined using reactions performed in the absence of enzyme but in the presence of UDP-[^{3}H]glucose.

*Gel Shift Assay***.** 15bp hmU or hmC containing oligonucleotides were radioactively labeled using T4 polynucleotide kinase (New England Biolabs) and [γ^{32}P]-ATP. The labeled oligonucleotide was then gel-purified and annealed with a non-labeled 27bp complimentary oligonucleotide. Radiolabeled dsDNA substrates were then incubated with JGT or T4 β-GT enzyme and either UDP-glucose or UDP-6-N_3-Glucose (Active Motif, catalog #55020) in an *in vitro* glucosyltransferase reaction. JGT glucosylation reaction consisted of 321μM UDP-glucose, ~0.05μM recombinant JGT, and 0.5pmol DNA in JGT reaction buffer [50mM potassium acetate,

20mM Tris acetate, 10mM manganese acetate, and 1mM DTT (pH 7.9)] in a 50μL reaction volume. Reactions were incubated at 25°C for 1 hour. βGT reactions consisted of 321μM UDP-glucose, 7.5 units recombinant βGT (New England Biolabs), and 0.5pmol DNA in T4 βGT reaction buffer [50mM potassium acetate, 20mM Tris acetate, 10mM magnesium acetate, and 1mM DTT (pH 7.9)] in a 50μL reaction volume. Reactions were incubated at 25°C for 1 hour. 1pmol of unlabeled DNA was added to all reactions to help minimize any DNase activity. Biotinylation reactions were then carried out according to the manufacturers instructions in the Hydroxymethyl Collector™ Kit (Active Motif, catalog #55013). Samples were then mixed with 2X formamide loading buffer and electrophoretically separated on a 20% polyacrylamide/7M urea gel in 1X TBE Buffer. The gel is fixed, dried and visualized by autoradiography

J-DNA enrichment

Base J immunoprecipitations were performed as previously described (28-30,49). Briefly, 25μL protein G beads were pre-blocked with 5μL 10mg/mL BSA and 5μL yeast tRNAs for 15 minutes. A 500μL IP reaction containing blocked beads, 1-3μg DNA, 5μL 10mg/mL BSA, 5μL yeast tRNAs, and 10μL rabbit α-base J serum were set up for each sample. Genomic DNA samples were sonicated to 0.5-3kb range prior to J-IP. IP reactions, performed in triplicate, were incubated at room temperature for 1.5 hours with rotation. DNA was eluted from beads by the addition of 400μL TE Buffer, 4μL 10mg/mL Proteinase K and 4μL 10% SDS and incubation at 37°C for 30 minutes. A phenol-chloroform extraction was then performed on each sample and DNA was precipitated with the addition of 800μL 100% Ethanol, 12μL 5M NaCl, and 4μL 20mg/mL glycogen.

Glucosylation and hmU pull down.

Spike DNA hmU enrichment assay. To determine the specificity of hmU pull-down method, JGT reactions were performed on a mixture of DNA containing 5ng modified or unmodified 75bp DNA substrate in 1µg sonicated genomic JGT KO DNA. DNA from this reaction was subjected to the J-IP protocol described above. The amount of hmU-DNA pull-down was determined by quantitative PCR (qPCR) analysis. %IP was calculated relative to input DNA. Quantification was performed on an iCycler with an iQ5 multicolor real-time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). The reaction mixture contained 5pmol forward and reverse primers, 2x iQ SYBR green super mix (Bio-Rad Laboratories, Hercules, CA, USA), and 2µL of template DNA. Primers used in qPCR analyses are listed in supplementary table S1. qPCR cycling conditions: 95°C 5 minutes, (95°C 15 seconds, 60°C 30 seconds, 72°C 30 seconds) for 40 cycles, and 95°C for 1 minute. Each standard curve generated had an R^2 value of at least 0.98, slope of -3, y-intercept of 23, and an efficiency of ~90-100%.

Genome-wide mapping of hmU. To determine the location of hmU throughout the *T. brucei* genome, JGT reactions were performed on JGT KO and JBP1/2 KO *T. brucei* genomic DNA. Reactions (50ul) containing 3µg of genomic DNA, 100µM UDP-glucose, ~0.05µM JGT and JGT reaction buffer [50mM potassium acetate, 20mM Tris acetate, 10mM manganese acetate, and 1mM DTT (pH 7.9) were incubated at 25°C for 1 hour. Reactions were then Proteinase K treated, phenol-chloroform extracted, and ethanol precipitated. 3µg of the purified glucosylated DNA was then utilized in a J-IP reaction as described above. Due to the low levels of hmU in the JGT KO and JBP1/2 KO genomes, 5 IP reactions were performed for each replicate and

pooled at the final step. The amount of hmU-DNA pull-down was determined by qPCR as described above.

DNA Dot Blots

To quantify the levels of hmU after *in vitro* JGT reactions, anti-base J immunoblots were performed as described previously (29). Briefly, genomic DNA was serially diluted, treated with 1X volume of 0.6M Sodium Hydroxide for 15 minutes, and then treated with 2X volumes cold 2M Ammonium Acetate for 5 minutes. DNA was then blotted onto a nitrocellulose membrane, followed by incubation with anti-base J antisera. Bound antibodies were detected using a secondary goat anti-rabbit antibody conjugated to HRP and visualized by ECL. The membrane was then stripped of antibody. As a loading control, mESC blots were stained with a solution of 0.02% methylene blue in 0.3M Sodium Acetate, pH 5.2.

RESULTS

JGT is an hmU-specific glucosyltransferase

We first wanted to determine if JGT is specific for hmU and if there is any activity on other TET oxidation products, specifically hmC. As expected, and consistent with previous results (25,27), we see that JGT is capable of glucose transfer onto double stranded DNA (dsDNA) substrates containing matched hmU (hmU:A) (Figure 4.2A). The other natural, but minor, form of hmU in mammalian genomic DNA is such that it is paired with G as a result of deamination of hmC (50,51). We now report that JGT has the ability to also modify substrates containing mismatched hmU:G (Figure 4.2A). JGT does not, however, show any activity on hmC containing dsDNA substrate (Figure 4.2A). In contrast, βGT is a glucosyltransferase that

acts on it primary substrate hmC, with some low level activity on hmU, but only when mis-paired to guanine (hmU:G) (24) (Figure 4.2B). βGT has no detectable activity on matched hmU DNA substrate. These data demonstrate that JGT has the unique ability to modify matched hmU in DNA, with no activity towards the other TET oxidation product, hmC.

Selective enrichment of hmU DNA using JGT and base J immunoprecipitation

Techniques to map hmC within genomes have relied on using βGT transfer of an azide-modified glucose onto hmC followed by biotin conjugation and streptavidin pull-down. As shown in Figure 4.3A, βGT is able to utilize UDP-6-N_3-Glc to transfer azido-glucose to hmC modified DNA substrate with high efficiency. However, repeated attempts with JGT have indicated the inability of the enzyme to utilize this modified glucose donor (Figure 4.3B). Since βGT and JGT belong to different glycosyltransferase structural fold families, GT-B and GT-A respectively (25,52,53), it is not surprising that the protein domain and specific interactions involved in binding UDP-glucose might be different between the two enzymes. Until we are able to synthesize other azido-linked Glc donors we will be unable to evaluate other potential substrates for use in chemical labeling of hmU modified DNA.

While JGT is unable to utilize UDP-6-N_3-Glc, the enzyme is able to utilize unmodified glucose donor to convert hmU in DNA up to 95% efficiency (Supplementary Figure S4.1). The resulting base J-DNA can be selectively enriched using base J antisera (28,30,49) or JBP1 (18-20,31,32). Both enrichment methods have been utilized in high-throughput sequencing of base J in trypanosomatid genomes (30,31). To examine whether this labeling method can discriminate between hmU and other modifications in a biological sample, we spiked in a 75bp dsDNA substrate with unmodified bases (T) or containing the modified bases 5hmU, 5mC or 5hmC into

trypanosome genomic DNA lacking base J and performed the JGT reaction and base J immunoprecipitation. The enrichment (%IP) of each substrate was analyzed by quantitative PCR (qPCR) (Figure 4.4). A significant enrichment of hmU was observed. No enrichment was observed for T, 5mC or 5hmC substrates. hmU enrichment is dependent on the JGT reaction, indicating both the strict dependence of this method on JGT and the specificity of the base J IP. These data demonstrate that incubation of hmU DNA with JGT followed by base J immunoprecipitation allows selective and efficient enrichment of hmU.

To begin to assess the feasibility of using this approach on biological samples, we performed JGT conversion of hmU in genomic DNA from mouse embryonic stem cells (mESCs). Genomic DNA was sonicated into small fragments (500-3000bp), treated with JGT in the presence of UDP-glucose to yield base J. Because of the efficiency of this reaction, this method ensures selective modification of most (if not all) hmU in genomic DNA. The conversion of hmU to base J allows accurate quantification of the amount of hmU in a genome using J antisera and HRP. Quantification of hmU levels using this approach indicate approximately 2-fold more hmU within the R1 mESC line than in the 2i mESC line (Supplementary Figure S4.2), which is consistent with previous quantitative mass spectrometry analyses (16). Taken together, these results demonstrate our ability to modify and measure hmU specifically in synthetic DNA oligonucleotides, PCR amplified DNA, and native genomic DNA.

Mapping of hmU within the *T. brucei* genome

Figure 4.4 demonstrates the ability to specifically enrich for a short hmU modified DNA substrate from a pool of DNA, however we would like to use this technique to enrich and map hmU-containing DNA within a genome. As a proof of principle, we performed JGT labeling and

J enrichment of genomic DNA from trypanosomes, subjecting the enriched fragments to qPCR analysis to allow identification of hmU-containing genomic regions. To do this, we took advantage of two different *T. brucei* cell lines which are devoid of base J; JBP1/2 KO and JGT KO. According to the J-biosynthesis model, hmU will be specifically localized in the genome of the JGT KO similar to base J profile in wild type (WT) trypanosomes, albeit at lower levels, due to JBP oxidation of thymidine (25,27). In contrast, the JBP1/2 KO lacks both of the thymidine hydroxylase enzymes involved in the first step of J-biosynthesis. Thus, the presence of hmU in this genome would presumably be due to damage during the genomic DNA isolation procedure and localized non-specifically. WT *T. brucei* genomic base J pull-down provides a positive control for J localization studies.

Base J is localized to particular sites within the trypanosome genome, including silent sub-telomeric localized variant surface glycoprotein (VSG) genes and RNA polymerase II (Pol II) transcription start sites and termination sites (30). In agreement with previous studies, we observed enrichment of base J in WT *T. brucei* at a silent VSG gene (VSG 224) versus a genome internally localized gene, ASF1 (Figure 4.5A). The no antibody control illustrates the strict dependence of the enrichment on base J antisera (Figure 4.5A). In the JGT KO genome we observed a similar specific enrichment of hmU that is dependent on the *in vitro* JGT reaction and no hmU detected in the JBP1/2 KO genome (Figure 4.5A). Further analyses also reveal that hmU is localized within regions involved in Pol II transcription termination (Figure 4.5B) and initiation (Figure 4.5C) with a profile identical to base J localization in the WT genome. No hmU is detected in these regions of the JBP1/2 KO genome. These data indicate that even with low levels of hmU in the JGT KO genome, ~25 hmU modifications per 10^6 nucleosides (25), this

method allows genome profiling of hmU. These observations also provide the first direct evidence that the JBP's dictate the specific localization of hmU (and J) in the *T. brucei* genome.

DISCUSSION

Here, we have developed a method to detect 5hmU that takes advantage of JGT-mediated glucosylation. This method ensures that only 5hmU-containing genomic regions will be precipitated and identified in subsequent assays. Given that JGT specifically catalyzes the glucosylation of 5hmU and JBP1 and J-antibody specifically recognizes the resulting glucosylated hmU base, the DNA pulled down is highly enriched in the 5hmU modification. The enriched DNA is then ready for analysis by real-time quantitative PCR or sequencing by any method, including high-throughput sequencing. Our method can be applied, with few exceptions, to any eukaryotic genome. There is no evidence for glucosylated DNA in mammalian genomes and, other than kinetoplastids, only *Diplonema* and *Euglena* have been shown to have base J (54,55). Background precipitation without JGT also provides a negative control. Furthermore, base J has been detected in synthetic DNA substrates using SMRT technology (33). Thus, our method has the potential to provide single-base resolution detection of 5hmU in a genome.

This report is, to our knowledge, the first assessment of the location of matched hmU in any genome. With the JGT tagging and enrichment procedure presented here, we observed hmU present at regions of the *T. brucei* genome where base J has been shown to regulate Pol II transcription and gene expression (30,35). The localization of hmU synthesized by the JBP enzymes at known base J sites, provides strong support for the specificity of J biosynthesis being regulated by the initial oxidation of specific thymines in the DNA rather than any bias by JGT.

We have demonstrated the ability of the JGT-mediated glucosylation method to detect and quantify levels of hmU in eukaryotic genomes, including mESCs. Future experiments will utilize this protocol for epigenetic profiling of hmU in mESC genomic DNA. While the majority of hmU in the genome is generated by the TET enzymes, a small portion of hmU is generated through deamination of cytosine yielding hmU:G mispairs (16). This mispaired hmU within the genome is rapidly removed by base excision repair and replaced with a non-modified cytosine base (50,51). Therefore, it is thought that hmU generated in this way does not contribute significantly to the steady state levels of hmU within the genome (16). While JGT can modify both matched and mismatched hmU, mismatched hmU can be identified by C-to-T mutation around the identified peaks following deep sequencing. In fact, the C-to-T mutation sites around the peak would both validate the hmU peak and mark the exact location of mismatched hmU site. These studies highlight the potential use of this method in mapping hmU localization and a powerful tool for probing the function of this newly discovered TET oxidation product.

ACKNOWLEDGEMENT

We acknowledge Steve Dalton and Ben Boward for culturing the mES cells. We thank the Borst lab for providing base J antisera. We thank Jessica Lopes da Rosa-Spiegler for critical reading of the manuscript.

REFERENCES

1. Jones, P. (2012) Functions of DNA methylation: islands, start sites, gene bodies, and beyond. *Nat Rev Genet.*, **13**, 484-492.

2. Smith, Z., and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet.*, **14**, 204-220.

3. Kohli, R.a.Z., Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**, 472-479.

4. Tahiliani, M., Koh, K.P., Shen,Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.

5. Ito, S., Shen, L., Dia, Q., Wu, S., Collins, L., Swenberg, J., He, C., and Zhang, Y. . (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carbosylcytosine. *Science*, **333**, 1300-1303.

6. He, Y., Li, B., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C., Zhang, K., He, C., and Xu, G. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303-1307.

7. Maiti, A., and Drohat, A. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implication for active demethylation of CpG sites. *J Biol Chem.*, **286**, 35334-35338.

8. Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M., Vasanthakumar, A., Patel, J., Zhao, X., Perna, F., Pandey, S., Madzo, J., Song, C., Dai, Q., He, C., Ibrahim, S., Beran, M., Zavadil, J., Nimer, S., Melnick, A., Godley, L., Aifantis, I., and Levine, R. (2011) Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, **20**, 11-24.

9. Li, Z., Cai, X., Cai, C., Wang, J., Zhang, W., Patersen, B., Yang, F., and Xu, M. (2011) Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*, **118**, 4509-4518.

10. Gu, T., Guo, F., Yang, H., Wu, H., Xu, G., Liu, W., Xie, Z., Shi, L., He, X., Hin, S., Iqbal, K., Shi, Y., Deng, Z., Szabo, P., Pfeifer, G., Li, J., and Xu, G. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, **477**, 606-610.

11. Dawlaty, M., Ganz, K., Powell, B., Hu, Y., Markoulaki, S., Cheng, A., Gao, Q., Kim, J., Choi, S., Page, D., and Jaenisch, R. (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, **9**, 166-175.

12. Dawlaty, M., Breiling, A., Le, T., Raddatz, G., Barrasa, M., Cheng, A., Gao, Q., Powell, B., Li, Z., Xu, M., Faull, K., Lyko, F., and Jaenisch, R. (2013) Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell*, **24**, 310-323.

13. Dawlaty, M., Breiling, A., Le, T., Barrasa, M., Raddatz, G., Gao, Q., Powell, B., Cheng, A., Faull, K., Lyko, F., and Jaenisch, R. (2014) Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell*, **29**, 102-111.

14. Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., and Yujiang, GS. (2011) Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol Cell*, **42**, 451-464.

15. Song, C., and He, C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends Biochem Sci.*, **38**, 480-484.

16. Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Müller, U., Spruijt, C., Vermeulen, M., Leonhardt, H., Schär, P., Müller, M., and Carell, T. (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.*, **10**, 574-584.

17. Rogstad, D., Liu, P., Burdzy, A., Lin, S., and Sowers, L. (2002) Endogenous DNA Lesions Can Inhibit the Binding of the AP-1 (c-Jun) Transcription Factor. *Biochemistry*, **41**, 8093–8102.

18. Robertson, A., Dahl, J., Vågbø, C., Tripathi, P., Krokan, H., and Klungland, A. (2011) A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **39**, e55.

19. Robertson, A., Dahl, J., Ougland, R., and Klungland, A. (2012) Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat Protoc*, **7**, 340-350.

20. Cui, L., Chung, T., Tan, D., Sun, X., and Jia X. (2014) JBP1-seq: A fast and efficient method for genome-wide profiling of 5hmC. *Genomics*, **104**, 368-375.

21. Song, C., Sun, Y., Dai, Q., Lu, X., Yu, M., Yang, C., and He, C. (2011) Detection of 5-Hydroxymethylcytosine in DNA by Transferring a Keto-Glucose by Using T4 Phage β-Glucosyltransferase. *ChemBioChem.*, **12**, 1682-1685.

22. Li, Y., Song, C., He, C., and Jin, P. (2012) Selective capture of 5-hydroxymethylcytosine from genomic DNA. *J Vis Exp.*, **68**, 4441.

23. Song, C., Clark, T., Lu, X., Kislyuk, A., Dai, Q., He, T., and Krolach, J. (2012) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nature Methods*, **9**, 75-77.

24. Yu, M., Song, C., and He, C. (2015) Detection of mismatched 5-hydroxymethyluracil in DNA by selective chemical labeling. *Methods*, **72**, 16-20.

25. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014) Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J Biol Chem.*, **289**, 20273-20282.

26. Sekara, A., Merritta, C., Baugha, L., Stuarta, K., and Myler, P. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Mol and Biochem Parasitol.*, **196**, 9-11.

27. Bullard, W., Cliffe, L., Wang, P., Wang, Y., and Sabatini, R. (2015) Base J glucosyltransferase does not regulate the sequence specificity of J synthesis in trypanosomatid telomeric DNA. *Mol Biochem Parasit*, **204**, 77-80.

28. Van Leeuwen, F., Kieft, R., Cross, M. and Borst, P. (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by B-D-glucosyl-hydoxymethyluracil in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **109**.

29. Van Leeuwen, F., Wijsman, E.R., Kieft, R., van der Marel, G.A., van Boom, J.H. and Borst, P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev.*, **11**, 3232–3241.

30. Cliffe, L.J., Siegel, T.N., Marshall, M., Cross, G.A., and 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.

31. Van Luenen, H., Farris, C., Jan, S., Genest, PA., Tripathi, P., Velds, A., Kerkhoven, R., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., Van Steensel, B., Myler, P., and Borst, P. (2013) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909-921.

32. Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marel, G. A., van Boom, J. H., van Leeuwen, F., and Borst, P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO*, **18**, 6573– 6581.

33. Genest, P., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H., Korlach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P., and Borst, P. (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res.*, **43**.

34. Ekanayake, D., and Sabatini, R. (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi.* Modulation of nucleosome abundance, histone modification, and polymerase occupancy by O-linked thymine DNA glucosylation. *Eukaryot. Cell* **10**, 1465-1472.

35. Reynolds, D., Cliffe, L., Fo¨rstner, K., Hon, C., Siegel, N., and Sabatini, R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res.*, **42**, 9717-9729.

36. Reynolds, D., Hofmeister, B., Cliffe, L., Siegel, T., Anderson B., Beverley, S., Schmitz, R., and Sabatini, R. (2016) Base J represses genes at the end of polycistronic gene clusters in Leishmania major by promoting RNAP II termination. *Mol Microbiol*.

37. Borst, P., and Sabatini, R. (2008) Base J. Discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.*, **62**, 235-251.

38. Yu, Z., Genest, P.A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G., Rigden, D. J., Sabatini, R., and Borst, P. (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.*, **35**, 2107–2115.

39. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R.,Marshall, M., Sweeney, K., and Sabatini, R. (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.

40. Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698 –1710.

41. Cliffe, L., Hirsch, G., Wang, J., Ekanayake, D., Bullard, W., Hu, M., Wang, Y., and Sabatini, R. (2012) JBP1 and JBP2 Proteins are Fe2+/2-Oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA. *J. Biol. Chem.*, **287**, 19886-19895.

42. Reynolds, D., Cliffe, L., and Sabatini, R. (2015) 2-Oxoglutarate-Dependent Hydroxylases Involved in DNA Base J (β-D-Glucopyranosyloxymethyluracil) Synthesis. *2-Oxoglutarate-Dependent Oxygenases*, 309-323.

43. Toaldo, C.B., Kieft, R., Dirks-Mulder, A., Sabatini, R., van Luenen, H.G. and Borst, P. (2005) A minor fraction of base J in kinetoplastid nuclear DNA is bound by the J-binding protein 1. *Mol. Biochem. Parasitol.*, **143**, 111–115.

44. Sabatini, R., Meeuwenoord, N., van Boom, J.H. and Borst, P. (2002) Recognition of base J in duplex DNA by J-binding protein. *J. Biol. Chem.*, **277**, 958–966.

45. Heidebrecht, T., Christodoulou, E., Chalmers, M. J., Jan, S., Ter Riet, B., Grover, R. K., Joosten, R. P., Littler, D., van Luenen, H., Griffin, P. R., Wentworth, P., Jr., Borst, P., and Perrakis, A. (2011) The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1. *Nucleic Acids Res.* , **39**, 5715–5728.

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

47. Kieft, R., Brand, Verena., Ekanayake, D., Sweeney, K., DiPaolo, C., Reznikoff, W., and Sabatini R. (2007) JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form *Trypanosoma brucei*. *Mol Biochem Parasitol.*, **156**, 24-31.

48. Taghizadeh, K., McFaline, J., Pang, B., Sullivan, M., Dong, M., Plummer, E., abd Dedon, P. (2008) Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat Protoc.*, **3**, 1287-1298.

49. Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I. and Borst, P. (2002) J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol. Microbiol.*, **46**, 37-47.

50. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L., Bartolomei, M., Rambow, F., Bassi, M., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., and Bellacosa, A. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67-79.

51. Guo, J., Su, Y., Zhong, C., Ming, G., and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423-434.

52. Vrielink, A., Ruger, W., Driessen, HP., and Freemont PS. (1994) Crystal structure of the DNA modifying enzyme beta-glycosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO*, **13**, 3413-3422.

53. Iyer, L., Zhang, D., Burroughs, M., and Aravind L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA *Nucleic Acids Res.*, **42**, 7635-7655.

54. Van Leeuwen, F., Taylor, MC., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., Borst, P. (1998) beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci.,* **95**, 2366-2371.

55. Dooijes, D., Chaves, I., Kieft, R., Dirks-Mulder, A., Martin, W., and 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.

FIGURES

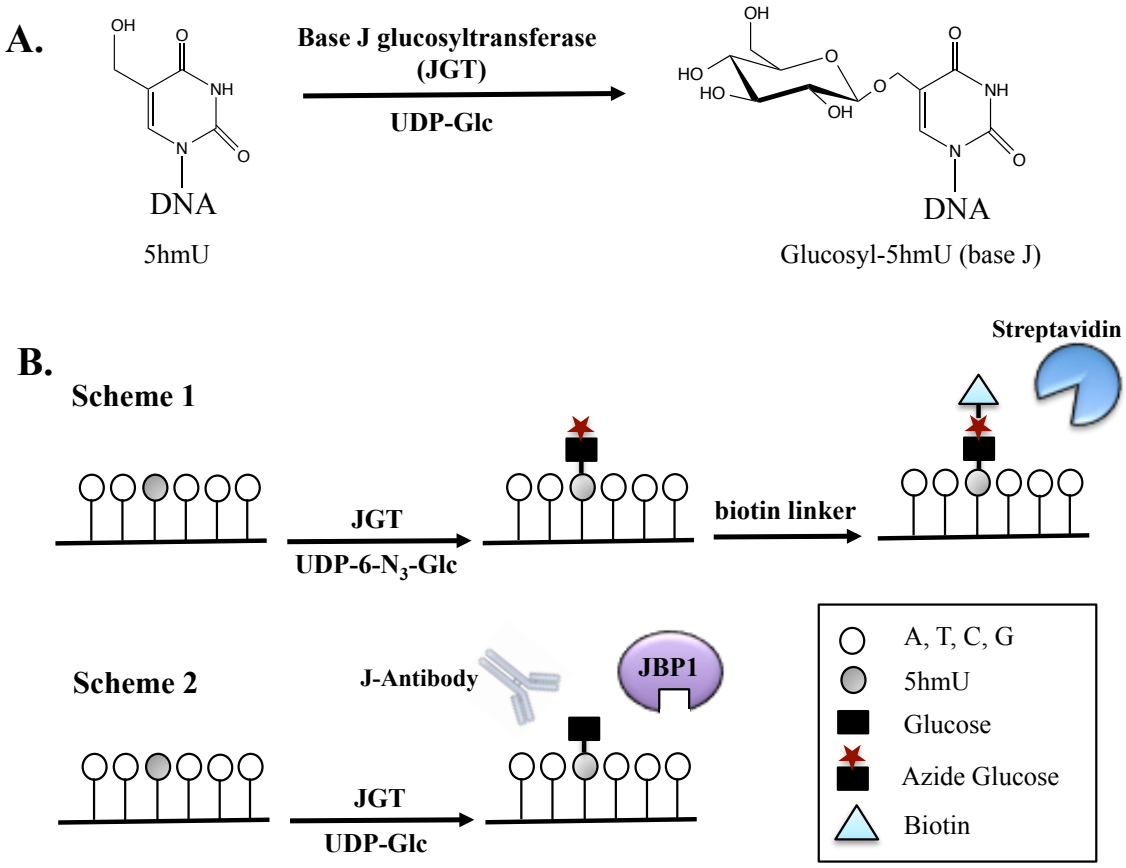


Figure 4.1. Selective labeling and enrichment of 5hmU in genomic DNA using JGT. (**A**) The hydroxyl group of 5hmU in duplex DNA can be glycosylated by JGT to form B-D-glucosylhydroxymethyluridine (glucosyl-5hmU), also called base J. (**B**) Overview of the selective labeling and enrichment strategy for 5hmU. Scheme 1. JGT is utilized to selectively label 5hmU in genomic DNA with N_3-glucose. After addition of the biotin tag through click chemistry, the hmU-containing DNA fragments can be enriched by streptavidin-coupled beads allowing detection and sequencing. Scheme 2. JGT is utilized to selectively label 5hmU with glucose allowing subsequent affinity purification using the base J-binding protein (JBP1) or antisera against base J. The glucosylated base can also be directly sequenced using SMRT-sequencing technology.

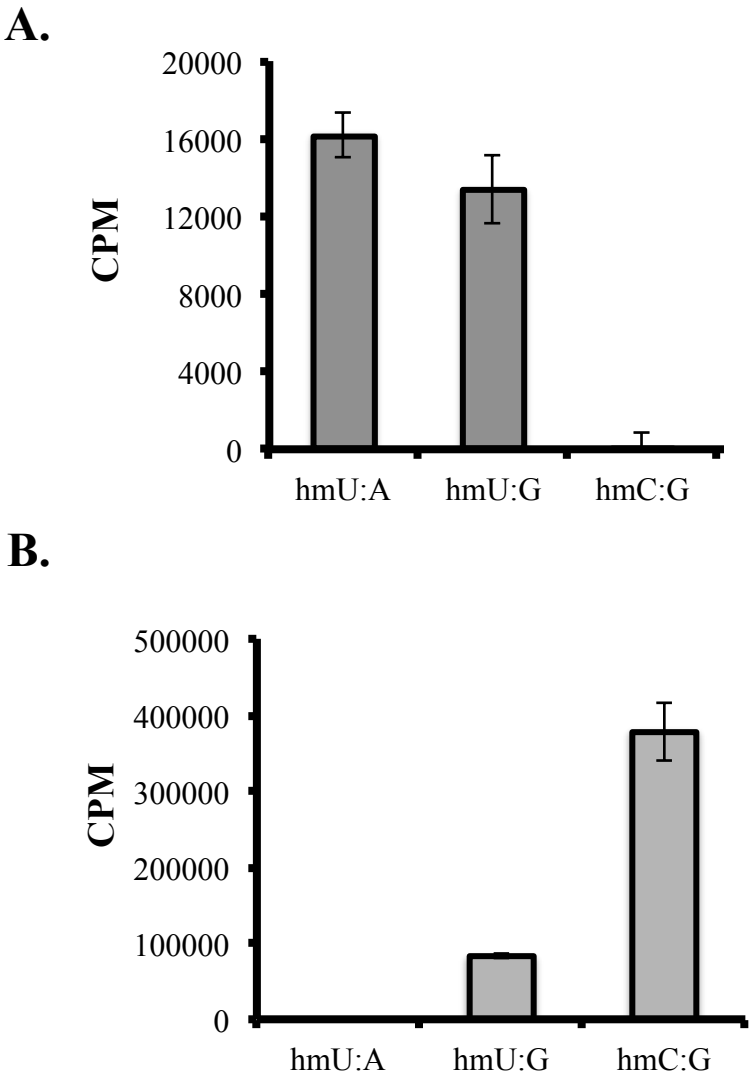


Figure 4.2. DNA substrate specificity of JGT. (**A**) Recombinant JGT or (**B**) T4 βGT and UDP-[^{3}H]glucose incubated with 36nt-long, dsDNA substrate containing one hmU or hmC molecule, as described in the materials and methods section. The modified base within the dsDNA substrate was present in the context of a matched base pair (hmU:A, hmC:G) or mismatched base pair (hmU:G). CPM, counts per minute, indicative of the transfer of glucose to DNA was measured for each sample. All experiments were performed in triplicate and error bars are representative of standard error.

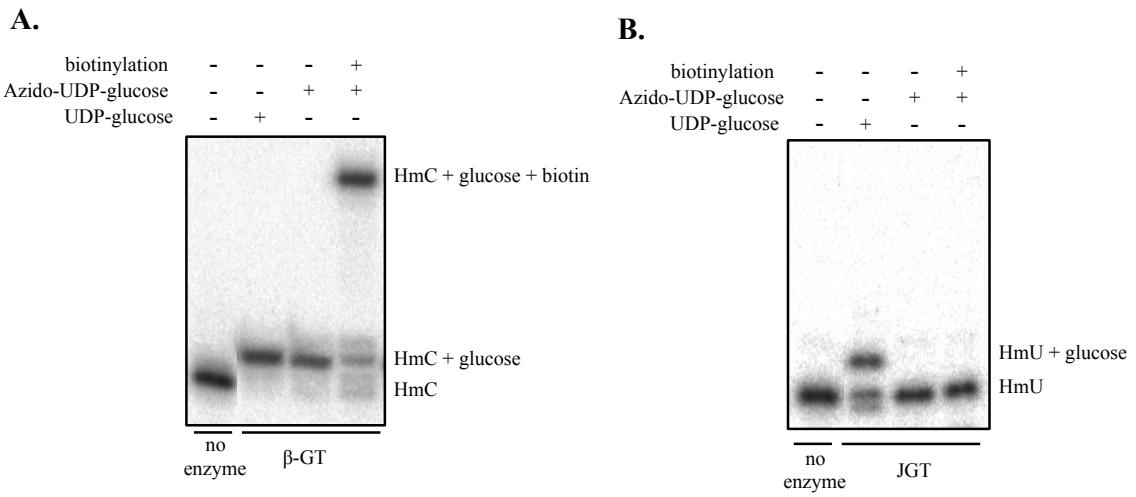


Figure 4.3. **JGT is unable to utilize UDP-6-N_3-glucose.** Denaturing polyacrylamide gel electrophoresis (PAGE) for monitoring the reaction mixtures of DNA substrates treated with (**A**) T4 βGT or (**B**) JGT and UDP-glucose or UDP-6-N_3-glucose (azido-UDP-glucose). 15nt-long ^{32}P labeled dsDNA substrate containing either an hmU or hmC were incubated with the indicated GT enzyme and nucleotide sugar as described in the materials and methods. The addition of a glucose moiety to the DNA substrate results in a visible shift on PAGE, with an even greater shift upon subsequent addition of biotin. No enzyme control indicates DNA substrate incubated without the addition of the corresponding GT enzyme.

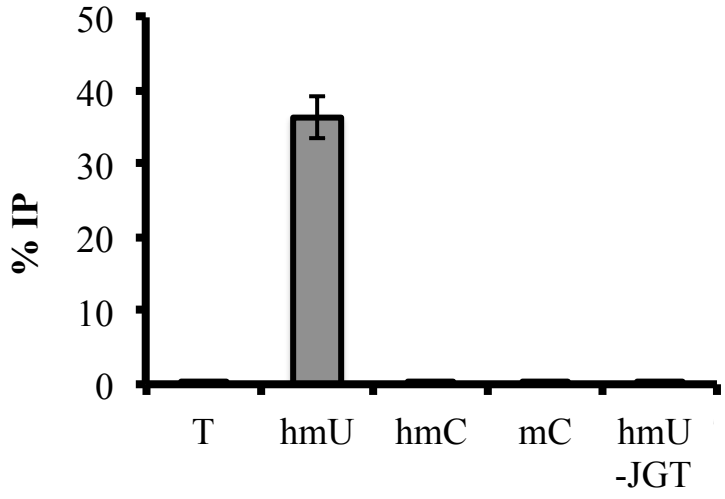


Figure 4.4. Enrichment test of the 5hmU pull-down assay. The 75nt long DNA containing T, hmU, hmC or mC was added to trypanosome DNA as spike-in controls and the JGT reaction and J-IP was performed as described in the materials and methods. hmU DNA substrate analyzed without the JGT reaction (hmU no JGT) is provided as a negative control. %IP was calculated relative to input DNA.

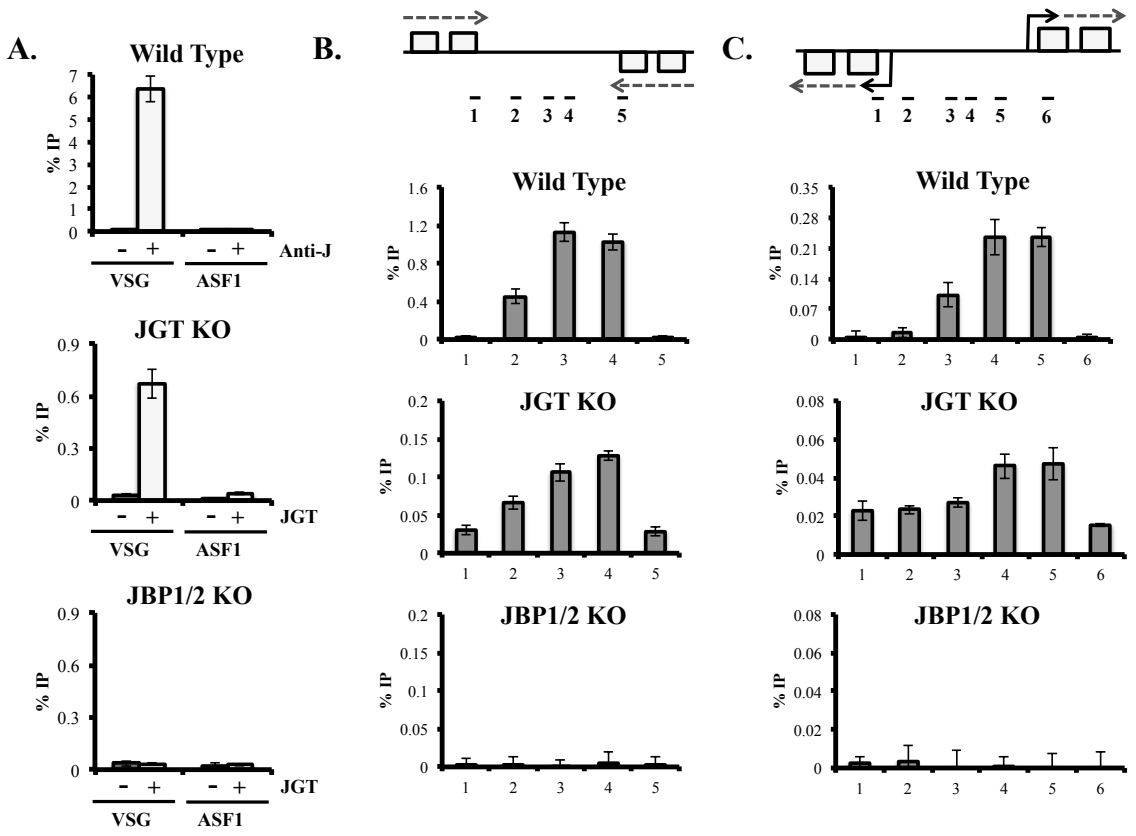


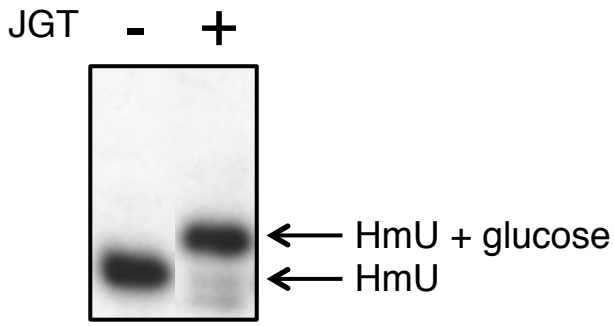
Figure 4.5. Mapping hmU in the *T. brucei* genome. Genomic DNA from the JGT KO and JBP 1/2 KO *T. brucei* cell lines was incubated with JGT and UDP-glucose and J-DNA was enriched by anti-J IP. Anti-J IP of wild type (WT) *T. brucei* DNA was used as a control to demonstrate the normal distribution of base J. qPCR analysis of J IP for the indicated regions of the genome was performed as described in material and methods. %IP was calculated relative to input DNA. (**A**) Analysis of base J and hmU at the silent 224 VSG (VSG) and the ASF1 gene. Specificity of the anti-J IP reaction is indicated by the %IP in wild type DNA with and without addition of the J antisera. Specificity of the hmU mapping method is indicated by the %IP with and without *in vitro* JGT incubation in JGT KO and JBP1/2 KO DNA. (**B** and **C**) hmU profile at a transcription termination site and a transcription initiation site. Above, diagram of a transcription termination site for two convergent polycistronic units of chromosome 10 (region 2500-2530kbp) (**B**) and a transcription start site for divergent units on chromosome 10 (region 1620-1640kbp) (**C**). Boxes

represent genes on the top and bottom DNA strand, arrows indicate direction of transcription. Location of qPCR primers spanning the known peak of base J in Wild Type cells is indicated. %IP was calculated relative to input DNA and normalized to the minus base J antisera (Wild Type) or minus JGT control (JGT KO and JBP1/2 KO).

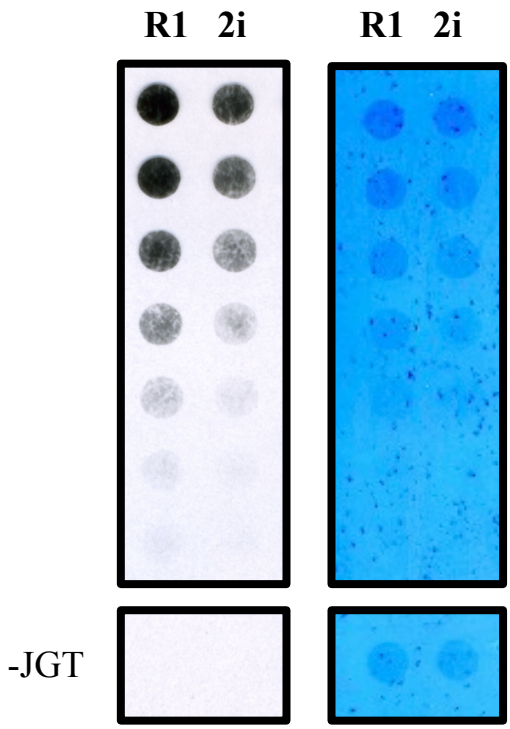
SUPPLEMENTARY TABLES AND FIGURES

Primers	Sense	Antisense
ASF1	5’-CTTTCGTGTGGGTCGGTAGT-3’	5’-CCCCTAACACTTCCTGCGTA-3’
VSG 224	5’-CGATGACGTCAATCCAGATG-3’	5’-CCGTTGGTGTCGTGTCTTC-3’
cSSR - 1	5’-GTATCACCACAGCCCGAACT-3’	5’-GGCAACCGAAAACAAAGAAA-3’
cSSR - 2	5’-AATTCGCCTACTGTCCATGCCGAT-3’	5’-TGTGCAGAACGCACATAAGGCAAC-3’
cSSR - 3	5’-GGTAAAGCTGGCGAAGTTGAAGGT-3’	5’-TTTCTTCCGGACACTCGCGATCAT-3’
cSSR - 4	5’-GGCCTTTATCCGCCGAAATTGGTT-3’	5’-CACTTTGTGGTGAATCAGCGGCAT-3’
cSSR - 5	5’-AACAACAGACTAATGGCGGG-3’	5’-TCGATGAATCTGCGCACTAC-3’
dSSR - 1	5’-CCCAATTTCACGGAAGAAAA-3’	5’-CTTGTGGACACGTGACTGCT-3’
dSSR - 2	5’-CGACCCAGCATAATGTTCCT-3’	5’-GGAAAGTGGACCGTTTTGAA-3’
dSSR - 3	5’-AAGCGGCGTCATTATTTGCAGACG-3’	5’-ATTGCTTCCACACCAACCAACGAC-3’
dSSR - 4	5’-TTCACGTGAGAGGTGCATTCCAGT-3’	5’-ACCATGCCGAATTCAGTTGTACCG-3’
dSSR - 5	5’-CACCCAATCCGTCATTCCACATCA-3’	5’-ACAGTCACAGCTCTCCTTCTCACA-3’
dSSR - 6	5’-TTCGTGTCAACAGGAGGTGCACTA-3’	5’-ACAGATGCCGTAGGTTCATTCGGT-3’

Supplementary Table S4.1. Sequences of qPCR primers used in *T. brucei* genomic analysis.



Supplementary Figure S4.1. Conversion of hmU to base J. JGT and UDP-glucose were incubated with 15nt-long ^{32}P labeled DNA substrate containing an hmU as described in the materials and methods and product visualized in PAGE as in Figure 4.3. No enzyme control indicates DNA substrate incubated without the addition of JGT enzyme.



Supplementary Figure S4.2. Quantitation of 5hmU in mESC genomic DNA. Genomic DNA isolated from two mESC lines (R1 and 2i) was incubated with JGT and UDP-glucose, spotted onto nitrocellulose in a two fold dilution series and levels of glucosyl-hmU detected by base J antisera. The dependence of the assay on the JGT labeling reaction is indicated below each blot by lack of signal from the highest DNA concentration assayed without the addition of JGT (-JGT). Methylene blue staining of the blot controls for DNA loading.

CHAPTER 5

CONCLUSIONS AND DISCUSSION

Kinetoplastids, including the human pathogens *Trypanosoma brucei*, *Trypanosoma cruzi,* and *Leishmania* possess a unique DNA modification known as base J that regulates gene expression. Base J is synthesized in a two-step pathway in which a thymidine in DNA is hydroxylated by the thymidine hydroxylases (TH) JBP1 or JBP2 to form 5-hydroxymethyluracil (hmU). A glucose moiety is then attached to the hmU intermediate by a glucosyltransferase to make base J. Understanding the enzymes involved in base J synthesis is critical to fully understand the impact of base J on gene expression in kinetoplastids.

While the JBP enzymes involved in the first step of base J synthesis have been identified and well characterized, the identity of the GT involved in the second step remained unknown, for many years. Consequently, we have been unable fully understand the synthesis of base J, its regulation, and the role this pathway plays in the regulation of transcription and gene expression in kinetoplastids.

IDENTIFYING THE BASE J GLUCOSYLTRANSFERASE

The identity of the elusive GT was first proposed from a bioinformatic study examining biochemical pathways for DNA modifications, where Avarind et al. identified a possible candidate protein that may be the glucosyltransferase involved in the synthesis of base J in *T. brucei* (1). We recently provided strong evidence showing that the candidate *T. brucei* protein is the GT involved in the synthesis of base J and is referred to as the base J associated GT, or JGT

(2,3) (Chapter 2). Using purified recombinant JGT we demonstrate that this protein utilizes UDP-glucose to transfer glucose onto double stranded DNA substrates containing hmU *in vitro* (2) (Chapter 2). Mutation of conserved residues in the catalytic domain of the GT-A fold impairs DNA glucosyltransferase activity (2) (Chapter 2). *In vivo*, deletion of both JGT alleles in *T. brucei* results in a complete loss of base J from the genome (2,3) (Chapter 2). Ectopic expression of an introduced HA-tagged JGT gene in a JGT-/- background restores base J synthesis (2) (Chapter 2). These studies not only confirm its identity as a GT and the two-step base J synthesis model, but also indicate it is the only GT catalyzing the second step of base J synthesis.

The glucosyltransferase involved in the synthesis of base J is a unique glycosyltransferase not only in that it is localized to the nucleus, but also that it is capable of modifying DNA; no other enzyme with similar substrate specificity has yet to be described in eukaryotes. A better understanding of GT enzymes, including the JGT, and their mechanism of action *in vivo* and *in vitro* is essential for rational drug design as well as increasing our knowledge regarding overall glycosylation machinery. The identification of a novel nuclear-localized GT in early branching eukaryotes could also help in the discovery of other novel glycosyltransferases. Interestingly, a mammalian homologue of JGT was also identified in the aforementioned bioinformatics study, called GREB1 (1). GREB1 is an estrogen responsive gene that has a largely unknown function but has been implicated in the proliferation of estrogen-receptor positive breast cancer cells (4,5). GREB1 localizes to the nucleus and has been found bound to chromatin, presumably functioning as a transcriptional co-activator of estrogen receptor-mediated transcription (6). Based on our characterization of JGT, an interesting possibility is that hmC present in the mammalian genome, generated by the TET enzymes, can be glucosylated by GREB1.

EXPLORING THE FUNCTIONAL ROLES OF THE JBP ENZYMES

With the discovery of JGT, we now have the ability to genetically segment the base J synthesis pathway into the hydroxylation step and the glucosylation step. While the loss of JBP leads to loss of base J, it also leads to the loss of hmU from the genome. With the identification of the JGT we can now differentiate between phenotypes due to the loss of base J versus the loss of hmU. Using the JGT we will be able to explore possible functions of hmU in the kinetoplastid genome. The presence of hmU within a DNA sequence has been show to perturb certain DNA-protein interactions, for example transcription factor-DNA interactions (7). The function of hmU in the *T. brucei* genome remains to be discovered.

The JBP enzymes may also be able to generate additional DNA modifications throughout the genome in addition to hmU. Like the mammalian TET enzymes, which can iteratively hydroxylate 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC) (8-10), the kinetoplastid JBPs may be able to function in a similar manner, converting 5hmU to 5-formyluracil (fU) and 5-carboxyluracil (caU) (2). These additional DNA modifications could themselves have some functional role within the cell. fC and caC has been shown to significantly reduce the kinetics of yeast RNA polymerase II transcription (11) and in mESCs fC and caC can recruit specific proteins, including transcription regulators (12). This data suggests that oxidized DNA bases may have regulatory functions within the genome (13). Further work will be needed to conclusively demonstrate JBP-mediated conversion of hmU to fU and caU and what function, if any, these modifications may have.

Elucidating the function of base J in regulating transcription and gene expression utilized cell lines with reduced base J levels following deletion of the TH enzymes (JBP1 or JBP2). The ability to delete JGT or both JBP genes from *T. brucei* is consistent with the non-essential nature

of base J in this organism. While deletion of either JBP1 or JBP2 in *T. cruzi,* and JBP2 in *L. tarentolae,* results in similar reductions in J levels as seen in *T. brucei* mutants, attempts to delete both JBPs have been unsuccessful, leading to the idea that base J is essential in these organisms. It is also possible that the JBP enzymes have additional functions outside of J synthesis, explaining the essential nature of the genes. Further studies analyzing the essential nature of the JGT in *L. major* and *T. cruzi* will allow direct testing of this hypothesis and will determine if base J is essential in these organisms. To determine the essential nature of this enzyme the loss or retention of an episomal copy of JGT will be tracked after deletion of both endogenous JGT alleles.

SPECIFICITY OF BASE J LOCALIZATION

While the pathway of J biosynthesis is now well understood, little is known about how J synthesis is regulated in the trypanosomatid genome. Genome-wide analysis of J localization in *T. brucei* and *L. major* indicates that J is present throughout the genome at specific regions, including SSRs involved in Pol II transcription (14,15). It is unclear what determines the localization of base J synthesis into specific sequences in the genome. While no consensus sequence or motif is evident from the genome-wide J analysis thus far, it is clear that, for at least the telomeric repeats, there is a sequence specificity component where in the top strand (GGGTTA) only the second T is modified (16,17).

One hypothesis is that the first step of J synthesis, catalyzed by JBP1 and JBP2, is the key regulatory step of J synthesis. Bypassing this first step, via feeding cells hmU, leads to J synthesis in regions of the genome that do not normally contain base J (18,19). These results support the idea that JGT will convert hmU to base J regardless of where hmU is located in the

genome and that JGT acts in a non-sequence-specific manner. Accordingly, the specificity of base J localization may be due to the JBP enzymes generating hmU at only specific sites throughout the genome. However, no direct evidence has confirmed this attractive hypothesis. To address this hypothesis and better understand the specificity of base J synthesis, an episome system in *Leishmania* was utilized to identify DNA sequences that will or will not promote base J synthesis *in vivo*. Using this system it was demonstrated that the telomeric repeat (GGGTTA) stimulates J synthesis *in vivo* while mutant telomeric sequences (GGGTTT, GGGATT, and GGGAAA) do not (20) (Chapter 3). Utilizing an *in vitro* GT assay, we find that JGT can glycosylate hmU within any sequence with no significant change in Km or kcat, even mutant telomeric sequences that are unable to be J-modified *in vivo* (20) (Chapter 3). The data suggests that JGT possesses no DNA sequence specificity *in vitro*, lending support to the hypothesis that the specificity of base J synthesis is not at the level of the JGT reaction. Taken together, it the JBP enzymes may regulate the specific localization of base J in the trypanosomatid genome.

To further test the hypothesis that the JBPs control the specificity of base J insertion, a technique was developed to map hmU, the hydroxylation product of the JBPs, within the genome of *T. brucei*. This technique relies on the conversion of hmU in genomic DNA to base J using the JGT in an *in vitro* reaction (Chapter 4). The resulting base J can then been enriched using the base J specific antibody (Chapter 4). The enriched DNA fragments can then be subjected to qPCR analysis using primers to map hmU at specific sites throughout the genome. This technique was applied to genomic DNA from JGT knockout (JGT KO) and JBP1/2 knockout (JBP1/2 KO) *T. brucei* cell lines (Chapter 4). It would be expected that in the JGT KO cell line, JBP1 and JBP2 would continue to generate hmU within the same sites that J is located in WT cells. Ablation of the JBP enzymes involved in generating hmU within the genome would result

in the loss of specifically localized hmU. Analysis of sites known to contain base J within the WT genome revealed a specific enrichment of hmU in the JGT KO genome, but no specific enrichment in the JBP1/2 KO genome (Chapter 4). This data provides the first direct evidence that the JBPs dictate the specific localization of hmU, and thus base J, in the *T. brucei* genome.

While the results described here strongly support the importance of the first step of J synthesis, further work is required to fully understand the formation of hmU at specific sequences by the JBPs. Once we have a robust *in vitro* assay for the JBP enzymes, similar analysis of JBP catalysis, as we performed for JGT, will help to shed light on primary DNA sequence requirements for hmU formation. Experiments to identify JBP1 and JBP2 associated factors may also help to explain the specificity of hmU formation *in vivo.* Immunoprecipitation of endogenously tagged JBP1 or JBP2 followed by mass spectrometry will reveal JBP-associated proteins.

DETECTION OF HYDROXYMETHYLURACIL IN GENOMIC DNA

Hydroxymethyluracil (hmU) is an oxidized DNA base found in the genomes of many organisms. In mammalian genomes, hmU was originally considered to be DNA damage produced from the oxidation of thymine by radical oxygen species (ROS), however, recent work using mouse embryonic stem cells (mESCs) suggests that hmU can also be generated though enzymatic oxidation of thymine by the ten-eleven translocation (TET) enzymes (21). The function of this enzymatically generated hmU and its location throughout the mESC genome is unknown. Studying this TET generated hmU in mESCs had been impeded by the fact that there had been no methods developed to map hmU within a genome. Therefore, we developed a technique to map the location of hmU within a genome using the JGT to convert hmU to base J

and enrichment is achieved through base J immunoprecipitation. As mentioned previously, this technique was applied to the JGT KO and JBP1/2 KO genomes with specific enrichment of hmU-DNA (Chapter 4). With few exceptions, this technique can be performed on any genome.

Although the function of hmU within the mESC genome is unknown, it is postulated that like the other TET generated DNA bases, hmU helps regulate development. Proteins that specifically recognize hmU have been identified, including transcription factors and several proteins involved in chromatin remodeling (21). In order to determine the function of this newly discovered TET oxidation product, it is crucial to map the location of hmU genome wide. As described in chapter 4, our method can be applied to the mESC genome, in conjunction with sequencing, to map hmU and potentially shed light on the function of TET-generated hmU.

SUMMARY

Studies of base J had long been impeded without the identity of the glucosyltransferase involved in the second step of its synthesis. The identification of the glucosyltransferase required for the second step in J synthesis has increased our knowledge of base J synthesis and has expanded the tools available to study this DNA modification in kinetoplastids. The discovery of this unique enzyme doesn't only impact the kinetoplastid field, but also represents a contribution to our understanding of glucosyltransferases and DNA modifying enzymes, and has potential application in the field of developmental biology. Overall, the identification and characterization of JGT has made and will continue to increase our overall understanding of DNA modifications and DNA modifying enzymes.

REFERENCES

1. Iyer, L., Zhang, D., Burroughs, M., and Aravind L. (2013) Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA *Nucleic Acids Res.* , **42**, 7635-7655.
2. Bullard, W., Lopes da Rosa-Spiegler, J., Liu, S., Wang, Y., and Sabatini, R. (2014) Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *J Biol Chem.*, **289**, 20273-20282.
3. Sekara, A., Merritta, C., Baugha, L., Stuarta, K., and Myler, P. (2014) Tb927.10.6900 encodes the glucosyltransferase involved in synthesis of base J in Trypanosoma brucei. *Molecular and Biochemical Parasitology*, **196**, 9-11.
4. Ghosh, M.G., Thompson, D.A. and Weigel, R.J. . (2000) PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. . *Cancer Res.*, **60**, 6367–6375.
5. Liu, M., Wang, G., Gomez-Fernandez, C.R. and Guo, S. . (2012) GREB1 functions as a growth promoter and is modulated by IL6/STAT3 in breast cancer. *PLoS One*, **7**, 46410.
6. Mohammed, H., D'Santos, C., Serandour, A.A., Ali, H.R., Brown, G.D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L. et al. . (2013) Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. . *Cell Rep.*, **3**, 342–349.
7. Rogstad, D., Liu, P., Burdzy, A., Lin, S., and Sowers, L. (2002) Endogenous DNA Lesions Can Inhibit the Binding of the AP-1 (c-Jun) Transcription Factor. *Biochemistry*, **41**, 8093–8102.

8. Tahiliani, M., Koh, K.P., Shen,Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.

9. Ito, S., Shen, L., Dai, Q., Wu, S., Collins, L., Swenberg, J., He, C., and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300-1303.

10. He, Y., Li, B., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C., Zhang, K., He, C., and Xu, G. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303-1307.

11. Kellinger, M., Song, C., Chong, J., Lu, X., He, C., and Wang, D. . (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. . *Nat. Struct. Mol. Biol.* , **19**, 831-833.

12. Spruijt, C., Gnerlich, F., Smits, A., Pfaffeneder, T., Jansen, P., Bauer, C., Mu¨nzel, M., Wagner, M., Mu¨ller, M., Khan, F., Eberl, H., Mensinga, A., Brinkman, A., Lephikov, K., Mu¨ller, U., Walter, J., Boelens, R., Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic Readers for 5-(Hydroxy)Methylcytosine and its Oxidized Derivatives. *Cell*, **152**, 1146-1159.

13. Song, C., and He, C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends Biochem Sci*, **38**, 480-484.

14. Cliffe, L.J., Siegel, T.N., Marshall, M., Cross, G.A., and 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.

15. Van Luenen, H., Farris, C., Jan, S., Genest, PA., Tripathi, P., Velds, A., Kerkhoven, R., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., Van Steensel, B., Myler, P., and Borst, P. (2013) Glucosylated Hydroxymethyluracil, DNA Base J, Prevents Transcriptional Readthrough in *Leishmania*. *Cell*, **150**, 909-921.

16. Van Leeuwen, F., Wijsman, E.R., Kuyl-Yeheskiely, E., Van der Marel, G.A., Van Boom, J.H. and Borst, P. (1996) The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res*, **24**, 2476–2482.

17. Genest, P., Baugh, L., Taipale, A., Zhao, W., Jan, S., van Luenen, H., Korlach, J., Clark, T., Luong, K., Boitano, M., Turner, S., Myler, P., and Borst, P. (2015) Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res*, **43**, 2102-2215.

18. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R.,Marshall, M., Sweeney, K., and Sabatini, R. (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.

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

20. Bullard, W., Cliffe, L., Wang, P., Wang, Y., and Sabatini, R. (2015) Base J glucosyltransferase does not regulate the sequence specificity of J synthesis in trypanosomatid telomeric DNA. *Mol Biochem Parasit*, **204**, 77-80.

21. Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Müller, U., Spruijt, C., Vermeulen, M., Leonhardt, H., Schär, P., Müller, M., and Carell, T. (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.*, **10**, 574-584.