Published in final edited form as:

Mol Biochem Parasitol. 2014 August ; 196(1): 9–11. doi:10.1016/j.molbiopara.2014.07.005.

Tb927.10.6900* encodes the glucosyltransferase involved in synthesis of base J in *Trypanosoma brucei

Aarthi Sekar^a, Christopher Merritt^a, Loren Baugh^a, Kenneth Stuart^{a,c}, and Peter J. Myler^{a,b,c,*}

^aSeattle Biomedical Research Institute, 307 Westlake Avenue, Seattle WA 98109-5219, USA

^bDepartment of Biomedical Informatics & Medical Education, University of Washington, Seattle 98195 WA, USA

^cDepartment of Global Health, University of Washington, Seattle 98195 WA, USA

Abstract

Base J is a DNA modification found in the genome of *Trypanosoma brucei* and all other kinetoplastids analyzed, where it replaces a small fraction of Ts, mainly in telomeric and chromosome-internal transcription initiation and termination regions. The synthesis of base J is a two-step process whereby a specific T is converted to HOMedU (hydroxymethyldeoxyuridine) and subsequently glucosylated to generate J. The thymidine hydroxylases (JPB1 and JBP2) that catalyze the first step have been characterized, but the identity of the glucosyltransferase catalyzing the second step has proven elusive. Recent bioinformatic analysis by Iyer *et al.* [*Nucleic Acids Res.* 2013; **41**, 7635] suggested that *Tb927.10.6900* encodes the glucosyltransferase (HmdUGT) responsible for converting HOMedU to J in *T. brucei*. We now present experimental evidence to validate this hypothesis; null mutants of *Tb927.10.6900* are unable to synthesize base J. Orthologues from related kinetoplastids show only modest conservation, with several insertion sequences found in those from *Leishmania* and related genera.

Keywords

DNA modification; epigenetics; glucosyltransferase; transcriptional read-through; trypanosomatids

The only modified DNA base identified to-date in the nuclear genome of kinetoplastid parasites is β -d-glucopyranosyloxymethyluracil (base J), which replaces $\sim 1\%$ of T residues and is predominantly localized within repetitive DNA sequences at telomeres [1-5]. However, a minor fraction of J is also located at chromosome-internal regions that coincide with RNA polymerase (RNAP) II transcriptional initiation and termination sites [6,7], and is

© 2014 Elsevier B.V. All rights reserved.

*Corresponding author: peter.myler@seattlebiomed.org, Phone: (206) 256-7332, Fax: (206) 256-7229.

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

required to prevent transcription read-through, at least in *Leishmania* [7]. Interestingly, J is present only in the bloodstream form of *T. brucei* [1,2], where it is found in transcriptionally silent variant surface glycoprotein (VSG) expression sites (ES), but not in the single active ES, suggesting a role in regulating antigenic variation [8,9]. Base J synthesis occurs in two steps: hydroxylation of a specific thymine to form the intermediate, hydroxymethyldeoxyuridine (HOMedU), and glucosylation of HOMedU to form J [10,11]. Two thymidine hydroxylases, JBP1 and JBP2, which belong to the TET/JBP superfamily of Fe²⁺/αKG enzymes [12], carry out the first step, but the glucosyltransferase (HmdUGT) that catalyzes the second step has proven elusive. A recent bioinformatic analysis by the Aravind laboratory [13] identified a putative HmdUGT gene immediately downstream of a TET/JBP gene in several phage genomes, and they proposed that the most closely related gene in *T. brucei* (*Tb927.10.6900*) encodes the long sought-after glucosyltransferase involved in J synthesis. Here, we provide genetic evidence that validates this hypothesis.

In anticipation that the HmdUGT would not be essential in *T. brucei* (since *JBP1* and/or *JBP2* null mutants are viable) [11], we generated *Tb927.10.6900* double knockout (dKO) transfectants of bloodstream forms, which normally contain J. We employed an efficient and rapid PCR fusion method [14] to construct DNA fragments containing selectable markers and ~500 nt from the 5' and 3' regions of the *Tb927.10.6900* coding sequence. The selectable markers were amplified from vectors that included 5' and 3' untranslated regions flanking the EP/PARP (procyclin) and ribosomal protein L4 genes, respectively, to ensure high expression. Transfections were carried out using *T. brucei* SM427 cells [15]. Single knock-out (sKO) and dKO cell lines were generated by sequential selection for hygromycin (*HYG*) and/or blasticidin (*BSD*) resistance. Successful replacement of *Tb927.10.6900* was confirmed by PCR of genomic DNA using primers flanking the gene. The *Tb927.10.6900* null mutants grew normally and did not display significant phenotypic differences from wild-type (WT). Genomic DNA was isolated from *Tb927.10.6900* sKO and dKO cell lines obtained from two independent transfections, digested with *DdeI*, and J levels were quantified using dot-blot analysis with anti-J antiserum [16]. A slight reduction (~10-30%, based on fluorescence intensity) in J levels in the sKO *versus* WT was observed, although this was only marginally statistically significant for one clone (Figure 1). However, all dKO clones appeared to lack J, since they showed only background levels of fluorescence (<20% of WT) at all DNA concentrations tested (Figure 1). An additional blot that included procyclic (J null) *T. brucei* gDNA as a negative control, and that used a different blocking agent (BLOTTO/TBS + 0.2% Tween 20 rather than 5% powdered skim milk in TBS + 0.2% Tween 20), showed no residual signal in the HmdUGT dKO or procyclic cell gDNA (Supplementary Figure S1). These results indicate that *Tb927.10.6900* encodes the HmdUGT responsible for the second step in base J synthesis in *T. brucei*, consistent with previous analyses showing that *Tb927.10.6900* mRNA is more abundant in bloodstream form *T. brucei* (which have J) than in procyclic forms (which lack J) [17-19].

Tb927.10.6900 has an orthologue (in a syntenic location) in all other trypanosomatid genomes analyzed, so the corresponding amino acid sequences were retrieved from GenBank and aligned using ClustalW [20] (Supplemental Figure S2). The HmdUGT orthologues showed only modest sequence conservation, ranging from ~42% identity between *T. brucei* and *T. cruzi*, ~20% between *Trypanosoma* and *Leishmania* or *Crithidia*,

and 14-20% between *Trypanosoma* and *Phytomonas*, *Angomonas* or *Strigomonas* (Supplemental Table S1). Indeed, there was only 80-90% identity between the various *Leishmania* species and ~54% identity between *Leishmania* and *Crithidia*. In addition, the *Leishmania* proteins (as well as *Crithidia*, *Phytomonas*, *Angomonas* or *Strigomonas*) contain several insertion sequences (of ~25-100 amino acids) that are absent in *Trypanosoma*. It remains to be seen whether these differences have any functional significance in terms of enzyme activity and/or specificity. We (and our collaborators) are currently expressing HmdUGT protein from several species for x-ray crystallography, as well as attempting to generate conditional knockouts in *Leishmania* (for which J is essential) to further elucidate the role of HmdUGT in these parasites.

After submission of this manuscript, a study published elsewhere [21] confirmed the complete loss of J after knockout of *Tb927.10.6900* in *T. brucei* and that reintroduction of the gene into *Tb927.10.6900*-null *T. brucei* restored J synthesis. That study also showed that reduction of HmdUGT mRNA by RNAi causes reduced J and increased HOMedU levels and that the glucosyltransferase uses uridine diphosphoglucose to transfer glucose to HOMedU.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Professor Piet Borst at the Division of Molecular Oncology, Netherlands Cancer Institute in Amsterdam for helpful discussions about this work. Research reported in this publication was supported, in part, by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number 1R01AI103858.

References

1. van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, Kieft R, et al. β -D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci USA*. 1998; 95:2366–2371. [PubMed: 9482891]
2. Gommers-Ampt JH, van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA, et al. β -D-glucosylhydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell*. 1993; 75:1129–1136. [PubMed: 8261512]
3. van Leeuwen F, Wijsman ER, Kuyl-Yeheskiely E, van der Marel GA, van Boom JH, Borst P. The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res*. 1996; 24:2476–2482. [PubMed: 8692684]
4. van Leeuwen F, Wijsman ER, Kieft R, van der Marel GA, van Boom JH, Borst P. Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev*. 1997; 11:3232–3241. [PubMed: 9389654]
5. van Leeuwen F, Kieft R, Cross M, Borst P. Tandemly repeated DNA is a target for the partial replacement of thymine by β -D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol Biochem Parasitol*. 2000; 109:133–145. [PubMed: 10960172]
6. Cliffe LJ, Siegel TN, Marshall M, Cross GA, Sabatini R. 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*. 2010; 38:3923–3935. [PubMed: 20215442]

7. van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, Velds A, et al. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell*. 2012; 150:909–921. [PubMed: 22939620]
8. Bernards A, Harten-Loosbroek N, Borst P. Modification of telomeric DNA in *Trypanosoma brucei*; a role in antigenic variation? *Nucleic Acids Res*. 1984; 12:4153–4170. [PubMed: 6328412]
9. Pays E, Delauw MF, Laurent M, Steinert M. Possible DNA modification in GC dinucleotides of *Trypanosoma brucei* telomeric sequences; relationship with antigen gene transcription. *Nucleic Acids Res*. 1984; 12:5235–5247. [PubMed: 6087278]
10. Borst P, Sabatini R. Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol*. 2008; 62:235–251. [PubMed: 18729733]
11. Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, et al. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res*. 2009; 37:1452–62. [PubMed: 19136460]
12. Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*. 2009; 8:1698–1710. [PubMed: 19411852]
13. Iyer LM, Zhang D, Maxwell Burroughs A, Aravind L. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res*. 2013; 41:7635–7655. [PubMed: 23814188]
14. Merritt C, Stuart K. Identification of essential and non-essential protein kinases by a fusion PCR method for efficient production of transgenic *Trypanosoma brucei*. *Mol Biochem Parasitol*. 2013; 19:44–49. [PubMed: 23685343]
15. Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol*. 1999; 99:89–101. [PubMed: 10215027]
16. Genest PA, ter Riet B, Cijssouw T, van Luenen H, Borst P. Telomeric localization of modified DNA base J in the genome of the protozoan parasite *Leishmania*. *Nucleic Acids Res*. 2007; 35:2116–2124. [PubMed: 17329373]
17. Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons P. Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. *BMC Genomics*. 2009; 10:482. [PubMed: 19840382]
18. Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GA. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res*. 2010; 38:4946–4957. [PubMed: 20385579]
19. Nilsson D, Gunasekera K, Mani J, Osteras M, Farinelli L, Baerlocher L, et al. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog*. 2010; 6:e1001037. [PubMed: 20700444]
20. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22:4673–4680. [PubMed: 7984417]
21. Bullard W, Lopes da Rosa-Spiegler J, Liu S, Wang Y, Sabatini R. Identification of the glucosyltransferase that converts hydroxymethyluracil to base J in the trypanosomatid genome. *JBC*. 201410.1074/jbc.M114.579821

Highlights

- In trypanosomatid genomic DNA base J replaces ~1% of T nucleotides
- Bloodstream forms of *Tb927.10.6900* null mutants lack base J
- *Tb927.10.6900* encodes the hydroxymethyldeoxyuridine glucosyltransferase (HmdUGT)
- HmdUGT orthologues in other trypanosomatids show only modest sequence conservation

Fig. 1. *Tb927.10.6900*-null bloodstream forms lack base J

(A) Genomic DNA from *T. brucei* SM427 wild-type (WT), *Tb927.10.6900::HYG* (sKO1), *Tb927.10.6900::Tb927.10.6900::BSD* (sKO2), *Tb927.10.6900::BSD/ Tb927.10.6900::HYG* (dKO1A and dKO1B), and *Tb927.10.6900::HYG/ Tb927.10.6900::BSD* (dKO2) bloodstream forms was digested with DdeI, spotted onto a nitrocellulose membrane, and probed with anti-J antiserum and secondary antibody (IRDye® 680RD goat anti-rabbit IgG (H+L)). (B) Fluorescence intensities from three independent dot-blot measurements were measured using LI-COR Image Studio software, normalized as a percentage of the signal obtained for 100 ng WT DNA, and the results plotted as mean plus one standard deviation. Samples that showed statistically significant reduction from the corresponding WT sample for each DNA concentration are indicated by asterisks (* p<0.05, ** p<0.01, *** p<0.001 by two-sample, one-tailed t-test, assuming equal sample variance). dKO1B was present on only one dot-blot and was omitted from this analysis.