Telomeric co-localization of the modified base J and contingency genes in the protozoan parasite *Trypanosoma cruzi*

Dilrukshi K. Ekanayake, Michael J. Cipriano and Robert Sabatini*

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

ABSTRACT

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

INTRODUCTION

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

For many pathogenic organisms, telomeric regions of the genome are involved in homologous and ectopic recombination events and acquisition of heterologous gene sequences to diversify, as well as regulate the expression of, genes involved in virulence (4). For example, the telomere and subtelomeric regions of the blood protozoan parasites *Trypanosoma brucei* and *Plasmodium falciparum* are specialized in the expression of specific variant-surface genes that allow the parasite to evade the host immune system (5). In the case of *T. brucei*, which reside primarily within the mammalian bloodstream, evasion of the host immune response is achieved by regularly changing their variant surface glycoprotein (VSG) coat in a process termed antigenic variation. While each trypanosome contains ~1000 VSG genes, the cell only expresses one at a time. The active VSG is found within one of about 20 telomeric expression sites (ESs). In order to exclusively express a particular VSG, only one ES is active at a time, while the others are silenced. The preferred subtelomeric region in the genome of these organisms is believed

*To whom correspondence should be addressed. Tel: +1 706 542 9806; Fax: +1 508 457 4727; Email: rsabatini@bmb.uga.edu

to facilitate gene switching and expression, and the generation of new variants.

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

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

MATERIALS AND METHODS

Maintenance of parasite cultures

Trypanosoma cruzi epimastigotes of the Y strain was grown in LIT media containing 10% fetal calf serum at 28°C as described earlier (15). *In vitro* metacyclogenesis was performed according to the previously described methods. Briefly, epimastigotes growing in LIT were harvested by centrifugation and resuspended in triatome urine media (TAU) to a concentration of 3×10^5/ml and incubated for 2 h at 28°C. Parasites are then transferred to 150 cm^2 culture flasks containing TAU3AAG media

(supplemented with 10 mM proline, 50 mM L-glutamate and 50 mM L-aspartate and 10 mM glucose). Parasites are then incubated for 2–7 days at 28°C (16).

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

Anti J immunoblotting of terminal restriction digests and dot blots

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

In order to quantitate the amount of J outside telomeres in *T. cruzi*, J-containing fragments were detected on Southern blots as described in (10). Briefly, 200 ng of epimastigotes and trypomastigotes DNA was digested using frequently cutting restriction enzymes and size fractionated in a 0.5% agarose gel. Fragments were then blotted onto a nylon membrane (Hybond N+, Amersham Bioscience, Uppsala, Sweden) and blocked for 2 h in TBST with 5% milk. Immunodetection was performed using HRP conjugated goat-anti-rabbit antibodies in combination with enhanced chemiluminescence (Amersham Bioscience) after incubation with anti-J antisera. The membrane was then hybridized with a ^{32}P end-labeled telomeric oligonucleotide $(TTAGGG)_5$ probe in hybridization buffer at 55°C and washed three times with $6 \times SSC$. The resulting signals from the telomeric Southern blot and the J-immunoblot were then compared. The regions of the blot reacting with J antisera but not with the telomeric probe were quantitated by phosphoimager analysis. The optical density of the non-telomeric localized J was divided by the optical density of the whole lane following correction for background.

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

Trypanosoma cruzi CL Brener strain scaffold (CH473309–CH473946) and contig (AAHK01000001–AAHK01032746) information was downloaded from Genbank genome project database (AAHK00000000). A total of 32 746 singleton contigs and 29 495 scaffolds were imported into a local database. Telomeric repeats were annotated by searching the *T. cruzi* genome using known hexameric repeat sequence (5'-CCCTAA-3'). It searches for the known telomere repeat sequence, though requiring at least three copies of the sequence in a row in order to annotate that area. Original gene calls were imported from the Genbank entries. No additional gene calls were automatically made. Contig entries were reverse complemented from their original Genbank entries in order for scaffolds to show correct contig alignment. We refer to these regions as S-00 to denote the

representative super-contig of the *T. cruzi* sequencing database. Any questionable overlapping regions during the assemblies were confirmed by PCR. Gene calls were annotated by blastp searching against nr, Swiss-Prot, mitop and HMM searching against Pfam _ls (17). All contigs between 100 bp and 10 kb were submitted to nr and Swiss-Prot via blastx algorithm. Intergenic regions were obtained and searched with blastx against nr and Swiss-Prot. TMHMM and SignalP algorithms were applied to all gene calls (18,19). Interpro searches were also performed on all gene calls (20). The whole assembly was run through nucmer searching for intra-genomic matches of size 40 or greater and annotated within the genome.

J-immunoprecipitation

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

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

RESULTS

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

In all kinetoplastids analyzed so far, base J is abundantly found within telomeric repeat sequences (7). In the African trypanosome, J biosynthesis is restricted to the mammalian bloodstream life stage where 50% of the total amount of J is localized within the telomeric repeats. Previous analysis of J in *T. cruzi* identified the modified base in the telomeric repeats and suggested its synthesis is developmentally regulated (7). However, it was unclear from these studies whether the DNA samples representing the different life stages were from the same strain of *T. cruzi*. It has since been shown that telomere length varies significantly between different strains of
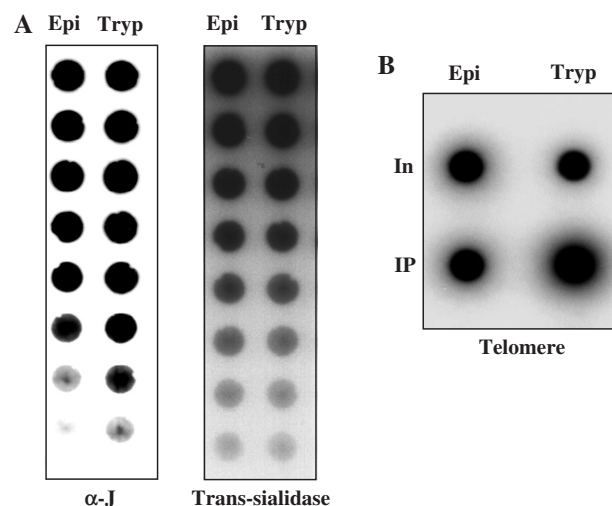

Figure 1. Developmental regulation of J-biosynthesis in *T. cruzi*. (**A**) A total of 200 ng of DNA from each life stage was serially diluted, spotted onto nitrocellulose and incubated with α-J antibodies. Signals were detected using enhanced chemiluminescence with a secondary antibody linked to HRP. Left lane represents the dilution series for the epimastigotes (Epi) and right lane represents the trypomastigotes (Tryp) life stage. The same blot was stripped and hybridized with a *T. cruzi*-specific trans-sialidase (TS) gene probe to check the DNA loading. (**B**) Epimastigotes and trypomastigotes DNA was sonicated and immunoprecipitated with α-J antibodies. Immunoprecipitated DNA (IP) along with 10% input DNA (In) was used in dot-blot assay and hybridized with an oligonucleotide probe of telomeric repeats.

T. cruzi (21). Thus, the level of J between different strains may be difficult to compare if the amount of J is proportional to the length of the telomere. In order to alleviate this problem, we isolated DNA from Y strain cultures of *T. cruzi* representing epimastigote and trypomastigote life stages and analyzed the levels of base J by dot-blot hybridization using anti-J antibodies. DNA loading was checked by Southern hybridization with a *T. cruzi*-specific TS gene probe. As shown in Figure 1A, there is about a 2-fold increase in the amount of base J in the mammalian life stage. Similar results were found with the CL-Brener strain (data not shown).

In order to examine the regulated J-synthesis in telomeric sequences, genomic DNA fragments immunoprecipitated with the J-specific antisera were analyzed by dot-blot hybridization (Figure 1B). In agreement with the previous study, we found that there is a similar ∼2-fold up-regulation of the level of J in the telomeric repeats of trypomastigotes compared to epimastigotes. These results confirm the regulation of base J synthesis in *T. cruzi* life cycle and suggest that telomeric localized J accounts for a significant fraction of the regulated J synthesis.

In order to quantitate the fraction of J localized outside the telomeric repeats, we performed terminal restriction digestions of genomic DNA followed by J immunoblot and Southern hybridization. To do this, we used a series of four base cutters (RsaI, AluI and MspI restriction enzymes) to separate out telomeric repeat regions from the rest of the chromosome. Telomeric repeats are not digested by these enzymes and result in a telomeric fragment of ∼500 bp, corresponding to the mean telomeric
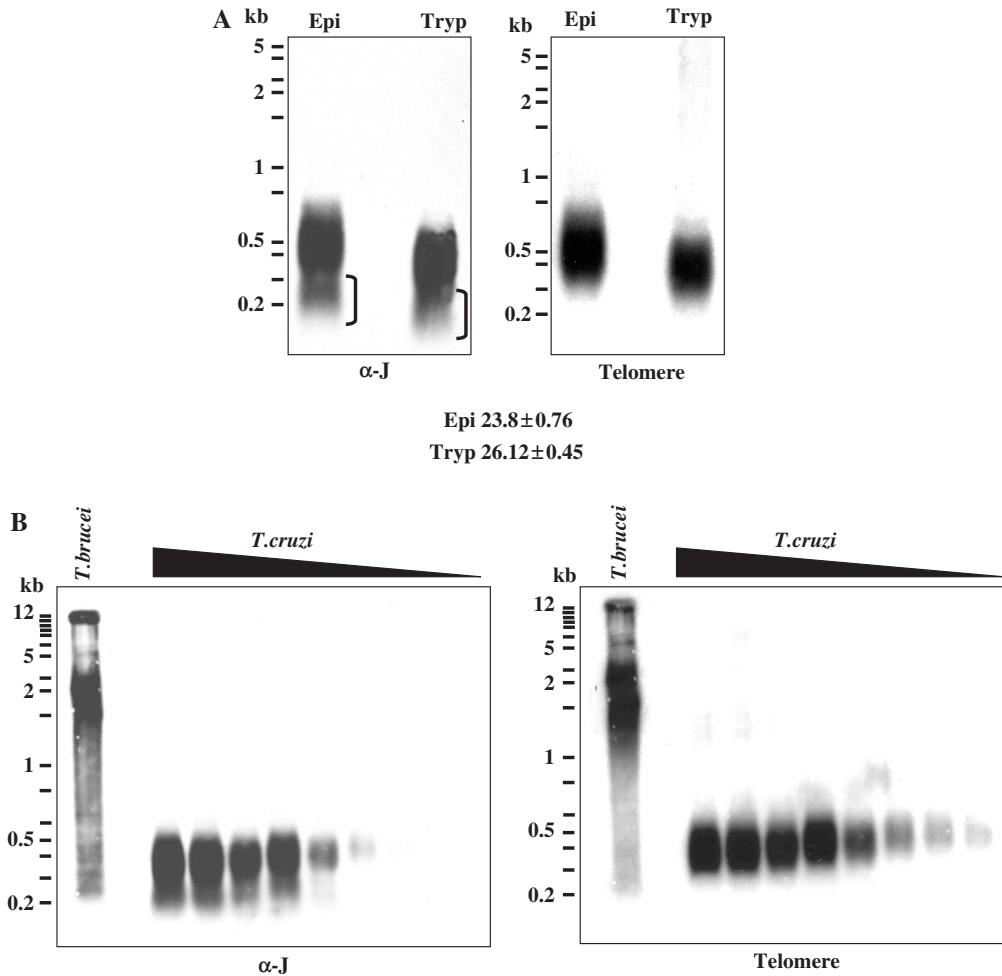
Figure 2. (**A**) Quantitating the fraction of base J present outside the telomeric repeat regions of *T. cruzi*. Southern blot and J immunoblot of genomic DNA from Y strain *T. cruzi* epimastigotes and trypomastigotes digested with frequently cutting restriction enzymes. DNA was digested with a mix of terminal restriction enzymes (MspI, AluI and RsaI) to generate the 500 bp telomeric repeat fragment. This digested DNA was size fractionated by gel electrophoresis and blotted onto a nylon membrane. The blot was incubated with the α-J antibodies and signals were detected by ECL (J Blot). The stripped blot was then hybridized with radioactively labeled telomeric repeat oligonucleotide probe (Telomere). Brackets indicate DNA fragments that contain base J but do not correspond to telomeric sequence. Quantitation of the percent J outside the telomeric repeats from three separate experiments is provided below. (**B**) Determination of the sensitivity of the Southern blot. Two-fold serially diluted genomic DNA from *T. cruzi* digested with a mix of AluI, MspI and RsaI restriction enzymes (starting with 250 ng digested DNA) and 4 μg of DNA of *T. brucei* digested with mix of AluI, AvaII, HinfI, RsaI and SspI restriction enzymes was size fractionated in 0.5% agarose gel and blotted onto nylon membrane. The blot was incubated with the α-J antibodies and signals were detected by ECL. The stripped blot was then hybridized with radioactively labeled telomeric repeat probe.

length in the Y strain (21). The digested DNA was size fractionated by electrophoresis in an agarose gel, blotted and incubated with the J-antisera followed by hybridization with a telomeric probe. As shown in Figure 2A, most of the telomeric fragments co-migrated with the J DNA suggesting that the telomeric repeats contain high quantities of the modified base. However, a fraction of the digest did not hybridized with the telomeric probe but did react with anti-J antibody. To examine the sensitivity of our assays, we increased the amount of digested genomic DNA in 2-fold dilution series (Figure 2B). By this approach, we found that even with increased amounts of DNA we did not detect smaller telomeric fragments within 100–250 bp range whereas, the smaller J-DNA fragments were readily detectable by the J immunoblot. Furthermore, we see equivalent levels of detection of

telomeric fragments of *T. cruzi* (∼500 bp) and *T. brucei* (∼10 kb). Thus, the lack of detectable shorter telomeric DNA fragments in *T. cruzi* is not due to any differential ability to detect 500 bp versus 250 bp of telomeric DNA on the Southern blot.

Quantification of the J-blot in Figure 2A indicates ∼23% and ∼26% of the total J was found outside the telomeric repeats in the epimastigote and trypomastigote life stage, respectively. These results demonstrate that there is a similar and significant level of base J present outside the telomeric repeat regions of the insect and mammalian life stages of *T. cruzi*.

Localization of base J in tandem repeat arrays

In *T. brucei* DNA base J was found in tandem repeat sequences that are primarily subtelomeric and some repeat

Table 1. Distribution of base J in tandem repeats of *T.cruzi*

Sequence	Unit size	Copy numbers/array	% IP	±SE	*n*	Density[a]	Ref[b]
VIPER	2.3 kb	70	10.49	0.12	3	0.065	(3–22)
DGF-1	10 kb	~100	12.47	0.91	3	0.012	(3–22)
LITc	5 kb	300–500	4.25	0.46	3	0.002	(3–22)
DIRE	260 bp	150	10.56	0.99	3	0.270	(3–22)
Tubulin	3.6 kb	10	0.003	0.004	3	NA	(3)
LS-rRNA	18 kb	80	9.39	0.55	3	0.006	(3)
SS- rRNA	740 bp	100	8.58	0.80	3	0.115	(3)
SL RNA coding	120 bp	~1–200	0.92	0.32	3	0.076	(24–25)
SL RNA nts	200 bp–5 kb[c]	ND	5.55	0.41	3	ND	(24–25)
Sat repeats	195 bp	20000	0.02	0.03	3	NA	(22)

[a]Density = %IP/(unit size × copy number).
[b]Reference for repeat organization.
[c]Based on Southern hybridization in references and analysis performed here.
NA-Not applicable since%IP indicates the absence of base J in these regions.
ND-Not determined/not published.

arrays that are chromosomal internal. This includes the subtelomeric 50 and 70 bp repeats, the 177 bp repeats (satellite repeats) in the minichromosomes, and the chromosome internal long arrays of 5S RNA and SL RNA gene repeats (10). Many of these J-containing sequences in *T. brucei* are lacking in *T. cruzi* (22). Therefore, we tested several of the known repetitive sequences of *T. cruzi*, such as the satellite sequences (195 bp repeats), DIRE, L1Tc, VIPER and DGF-1. We also looked for the presence of J in repetitive genes like the SL RNA, the 18S and 24S rRNA as well as the α-tubulin gene cluster. The results of these experiments are presented in Table 1. Using sonicated DNA, we find significant immunoprecipitation of all the repetitive sequences tested with the exception of the 195 bp satellite repeats. The 0.003% immunoprecipitation of tubulin fragments is consistent with background immunoprecipitation observed with DNA without J (7,10,11).

All trypanosomatid ribosomal gene sequences and SL RNA genes exist in tandem arrays (24). In order to determine whether the ribosomal genes are modified, we used J-immunoprecipitation of sonicated DNA followed by dot-blot hybridization. As shown in Table 1, the small and large subunit rRNA genes are significantly immunoprecipitated. The trans-spliced 39-mer region of SL RNA gene is highly conserved between species while the 3′ non-spliced portion varies in size and sequences. At least 10% of SL RNA genes in *T. cruzi* are interrupted by a site-specific retrotransposon element, called cruzi-associated retrotransposon (CZAR), between nucleotides 11 and 12 of the SL 39-mer (25). To determined whether SL RNA gene units are modified, we used sonicated DNA and immunoprecipitated with anti-J antibodies, dot blotted onto a nylon membrane and hybridized using probes representing the SL RNA transcribed region and non-transcribed region. As indicated in Table 1, the SL RNA transcribed region showed a slight immunoprecipitation whereas the non-transcribed region showed a significant immunoprecipitation. These results were confirmed by Southern blot analysis (data not shown) and are comparable with the analysis of J within the SL RNA coding and non-transcribed region in *T. brucei* (10). However, the CZAR site-specific retro element of *T. cruzi* did not contain base J (data not shown).

Subtelomeric localization of J

As shown in Table 1, we obtained significant immunoprecipitation from repeat arrays, including DGF-1, DIRE, L1Tc and VIPER. These repeat sequences are known to be adjacent to telomeric repeats in some chromosomes and may have been pulled down in sonicated DNA because of their linkage to the J-containing telomeric repeats. To more closely examine the presence of J within these sequences, we needed a restriction map of representative subtelomeric arrays and re-examine the ability to pull-down specific repeat sequences that have been cleaved from the telomeric repeat arrays. Furthermore, sequence assemblies representing telomeric end of several chromosomes would allow the detailed localization of base J throughout the subtelomeric region. In order to assemble large telomeric scaffolds (20–400 kb in size), we pulled out 70 *T. cruzi* telomeric contigs from *T. cruzi* DB using the known hexameric repeat sequence (5′-CCCTAA-3′). Contigs of these subtelomeres were further assembled into scaffolds to comprehensively expand the telomeric details. These assembled telomeres were annotated using Genbank annotations and GMOD and GBrowse was used to visualize the genes, pseudogenes and repeat regions. All of the subtelomeric assemblies have the varying length of telomeric repeats followed by telomere associated 189 bp telomeric junction sequences specific to *T. cruzi* (22). At least 25 of them consisted of TS genes or pseudogenes and many contain retro elements like VIPER, DIRE and L1Tc. VIPER sequences, DGF-1 and retrotransposon hot spot genes are particularly abundant in these subtelomeres. Ten percent of the subtelomeres we assembled have mucin genes, pseudogenes and strand switch regions within 40 kb of the telomeric repeats. Since we performed our initial assemblies, *T. cruzi* DB has updated their database to include similar scaffold assemblies including telomeric details. Our assemblies are consistent with the scaffold details now available from *T. cruzi* DB.
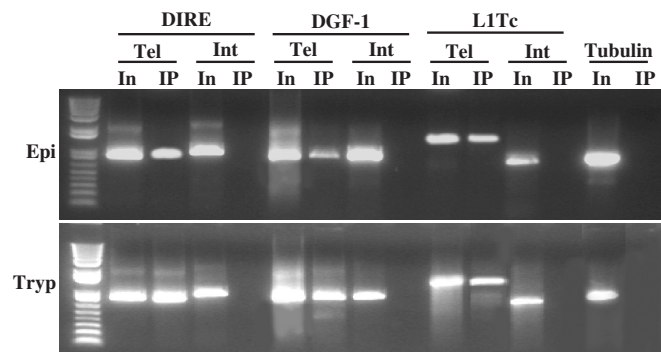
Figure 3. Telomeric-localized repetitive regions contain base J. IP and PCR analysis of the internal versus subtelomeric repeat arrays of *T. cruzi*. Genomic DNA from both life stages was digested using different restriction enzymes and immunoprecipitated using α-J antibodies. PCR was performed for immunoprecipitated DNA (IP) along with 10% of input DNA (In). Specific primers were designed to distinguish internal versus subtelomeric copy of the same repeat based on the *T. cruzi* database. The same digested DNA sample was used in the analysis of the internal and subtelomeric copy of a given repeat. The tubulin gene was used as a control.

Table 2. Telomeric versus internal repeat scaffold

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

[a]Represents the distance of the end of the repeat sequence to the telomeric repeat.
[b]Represents the distance of the end of the repeat sequence to the closest end of the scaffold. Thus indicating the minimal distance of the repeat to the telomeric repeat. Based on http://www.tcruzidb.org.

Utilizing this telomeric contig database, we were able to examine the localization of J within internal versus telomeric-localized repeat arrays. To do this, we designed specific primers to distinguish internal versus subtelomeric members of the same repeat sequences (see Supplementary Figure 1A and B for more detail). Furthermore, we cloned and sequenced all the PCR products to confirm the specific identity of these regions. Internal repeat sequences were chosen based on their location in the genome scaffolds and the presence of adjacent internal gene arrays (i.e. ribosomal genes). See Table 2 for a summary of the assemblies analyzed representing telomeric versus internal repeat regions. Digested DNA was immunoprecipitated and analyzed by PCR using primers against the internal or subtelomeric-localized sequences. As demonstrated in Figure 3, internal members of DGF-1, DIRE and L1Tc did not contain DNA base J but their subtelomeric counterparts were readily immunoprecipitated with anti-J antibodies. This result was consistent between epimastigotes and trypomastigote life stages. All PCR products were cloned and sequenced to confirm identities. VIPER sequences are generally found in subtelomeres, and thus, we were able to get a positive reaction for subtelomeric VIPER sequences (Figure 5 and data not shown). However, due to the fact that we were unable to distinguish internal versus subtelomeric members of VIPER, we are unable to conclude that only subtelomeric VIPER sequences are modified by base J. These results demonstrate that DNA base J is present in the subtelomeric-localized repeat arrays of *T. cruzi* chromosomes throughout the life cycle.

To determine the distribution of base J in more detail, we generated restriction maps for four subtelomeric regions (Figures 4 and 5). Genomic DNA digested with the indicated restriction enzymes was immunoprecipitated with J antibodies and used in a PCR reaction with specific primers designed to amplify defined regions of the subtelomere using a similar approach described for the analysis of the internal and telomeric repeat arrays (Figure 3 and Supplementary Figure 1). As shown in Figure 4B, all regions analyzed, up to 30 kb from the telomeric repeats, in contig 197 are immunoprecipitated using DNA from the trypomastigote life stage. In contrast, while J is present up to 30 kb from the telomeric repeats, two regions of the subtelomere in epimastigote DNA lacks J. As a negative control, tubulin gene was not immunoprecipitated with J antibodies. Specificity of the PCR assay was confirmed by direct sequencing of the PCR products (data not shown) and by Southern analysis (Figure 4C–E). In the Southern blot analysis, immunoprecipitated DNA was size fractionated and hybridized using the PCR fragments as probes. In many cases, complete digestion and the developmental regulation of J in particular regions is clearly verified by the Southern blot analysis (for example see Figure 4E). The subtelomeric regions analyzed are rich in sequences representing members of large gene families (i.e. TS and RHS). Therefore, while the PCR primers are specific, the amplified products may cross-hybridize to various fragments on the Southern blot. Interestingly, only a select fraction of the cross-hybridized fragments are immunoprecipitated. Whether these cross-hybridized fragments localize to telomeric regions of other chromosomes remains to be tested.

In order to further examine the developmental regulation and subtelomeric localization of J, we analyzed three additional telomeric contigs. The results from the PCR analysis of epimastigote and trypomastigote J-DNA immunoprecipitation in these regions, is presented in Figure 5. All results were confirmed by Southern hybridization (data not shown). The lack of J in regions immediately adjacent to telomeric repeats provides additional support that the digests were complete and false positive IP due to the linkage to telomeric repeats is unlikely. Therefore, our analysis of four telomeric contigs indicates that subtelomeric regions of *T. cruzi* extending to at least 50 kb from the telomeric repeats, including TS genes and pseudogenes and repeat regions such as VIPER and DGF-1, are modified by base J. Interestingly, some regions containing TS genes and pseudogenes (S-197, S-673, S-651) showed a developmental regulation where base J was absent in epimastigote stage. This suggests that there is not only localization pattern but also developmentally regulated distribution pattern of J biosynthesis in this region of the *T. cruzi* genome.
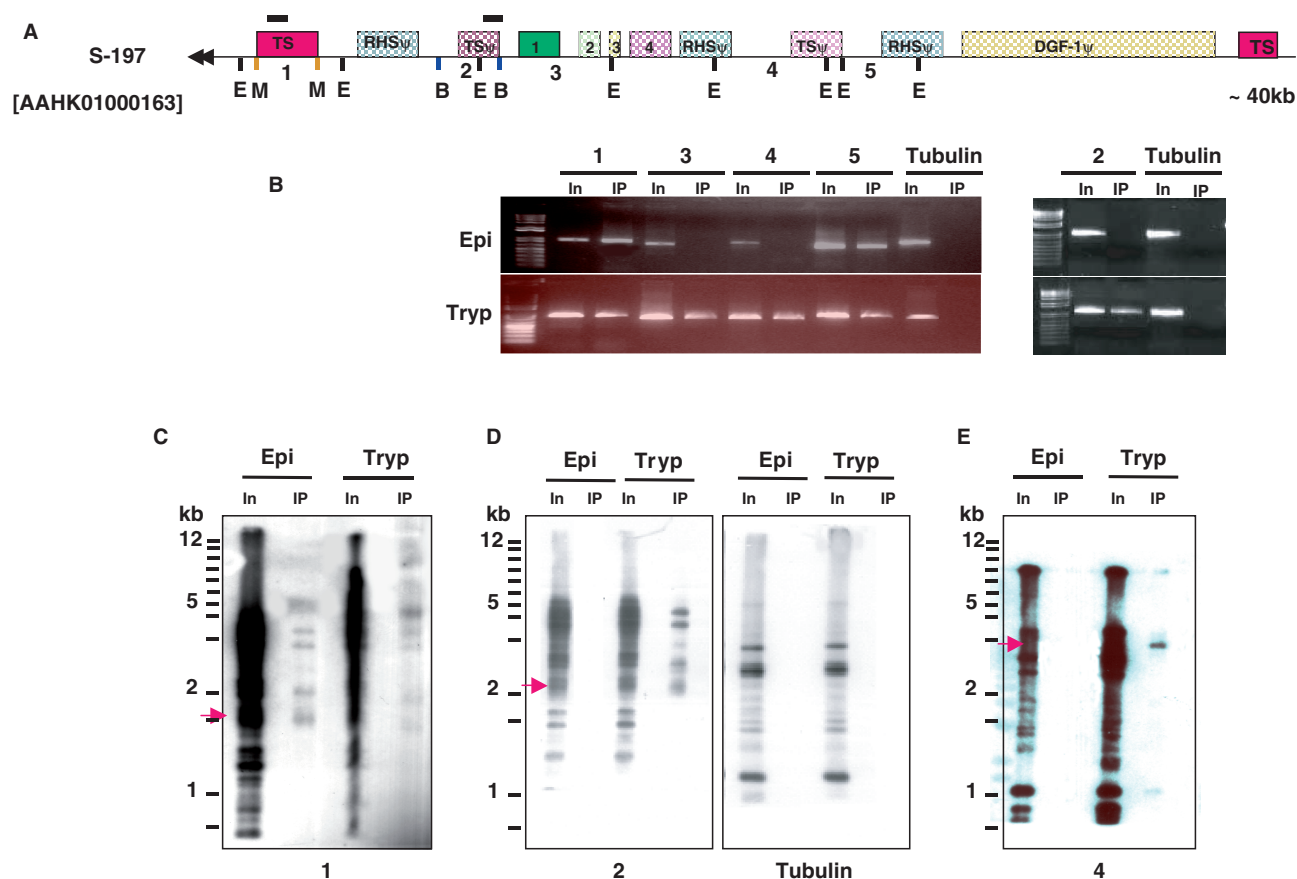
Figure 4. Localization of J in the subtelomeric region. (**A**) Schematic diagram and restriction map of the selected subtelomere using contig assembly of *T. cruzi* database. Vertical lines of the diagram of the subtelomere indicate the restriction site for the enzyme used for PCR and Southern hybridization; EcoRI (E), MscI (M) and BsrGI (B). The nomenclature used here for the subtelomeres and genes is according to the *T. cruzi* database and Genbank annotations. The Genbank scaffold/contig accession number is indicated in parenthesis. Arrowheads indicate the repetitive regions and horizontal lines represent the PCR amplified regions (also used as probes for Southern hybridization). Numbers represent the regions of interest. (**B**) PCR analysis. EcoRI-digested DNA was immunoprecipitated with α-J antibodies. Ten percent input and immunoprecipitated DNA was used in PCR reactions along with specific primers to amplify the indicated regions. Tubulin gene primers were used as a control. Upper panel represents the IP/PCR reactions for DNA of epimastigotes and lower panel for trypomastigotes. IP/PCR reaction for fragment 2 was from BsrGI digestion and amplified using the same oligos used for the analysis of fragment 3 following EcoRI digestion. It is shown separately with the corresponding Tubulin control for both epimastigotes and trypomastigotes. (**C–E**). Southern blot analysis. DNA was digested using different restriction enzymes (C, MscI; D, BsrGI and E, EcoRI) to isolate fragments of the regions analyzed above. The digested DNA was immunoprecipitated, size fractionated in an agarose gel and blotted onto a nylon membrane along with 10% input DNA. PCR products, representing the indicated regions, were used as the probe for Southern hybridization. Left lanes demonstrate the hybridization signals for epimastigotes and right lanes for trypomastigotes. Small arrows are used to point out the expected restriction digested fragments. The tubulin control hybridization is shown for the BsrGI digest/IP reaction.

DISCUSSION

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

It is becoming increasingly clear that members of the stage-specific surface glycoprotein gene families of *T. cruzi* that have been implicated in pathogenesis (i.e. TS and mucins) are localized to telomeric regions of the chromosome. For many pathogenic organisms, the telomeric regions are involved in homologous and ectopic
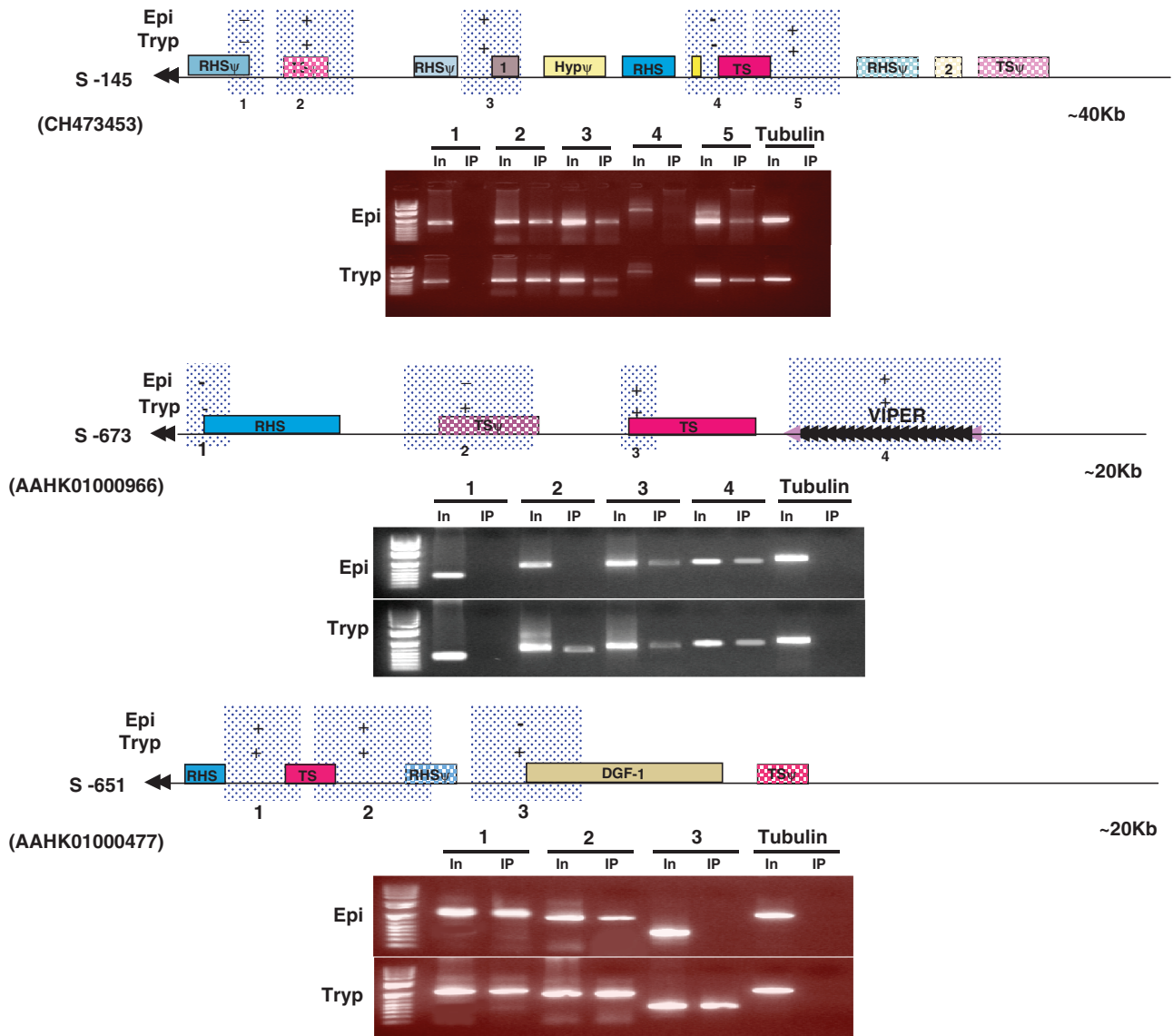
Figure 5. J localization within three additional subtelomeres. PCR data for epimastigotes and trypomastigotes is presented along with the schematic diagram of the representative subtelomeres. The nomenclature used here is according to the *T. cruzi* database and Genbank annotations and as described in Figure 5. Patch boxes show the restriction digested fragment analyzed and plus/minus signs indicate the presence and absence of base J for respective life stage.

recombination events and acquisition of heterologous gene sequences to diversify contingency gene families involved in virulence (4). In the case of *T. cruzi*, it has been speculated that telomeric localization of these surface glycoproteins may play a role in the evolution of their genetic diversity and thus increasing the parasites ability to endure host defenses (4,14,28,29). In order to begin to explore the role of J in telomeric gene expression/ diversity in *T. cruzi*, we have characterized the distribution of the modified base in more detail.

We find that base J synthesis is developmentally regulated in *T. cruzi*, with 2-fold up-regulation during the infective mammalian life stage, and ∼25% of the total J localized outside of the telomeric repeats. Similar to the distribution of J in *T. brucei*, we find J primarily within simple repeat sequences in *T. cruzi*. These include the

SL RNA and the 24S and 5S RNA gene repeats. While the amount of J may be low in these internal repetitive gene arrays, as indicated by the % IP, telomeric-localized repeat sequences (i.e. VIPER, DIRE) contain significant levels of J. However, when these same repeat sequences are examined when they are present within the chromosome, they do not contain the modified base. The lack of J within these sequences when they are localized within the chromosome, as well as the lack of J within the 195 bp satellite repeats, suggests that it is not the sequence or repetitive nature *per se* that determines whether a particular DNA region contains J, only that it is localized near the telomeric end of the chromosome. The ability to analyze the same DNA sequence in different locations in the *T. cruzi* genome has allowed us, for the first time, to directly examine the effect of genome context on the

regulation of J biosynthesis. Presumably, some aspect of the telomeric context is able to optimally recruit JBP2; the key thymine hydroxylase involved in stimulating *de novo* J synthesis (30,31).

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

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

The results presented here suggest that like in *T. brucei*, base J is mainly a modification of telomeric-localized repetitive DNA sequences in *T. cruzi*. Base J is thought to be involved in generalized repression of transcription as it was initially found in inactive ESs of *T. brucei* suggesting a role in antigenic variation. The significant fraction of J within DNA repeats in *T. brucei* also suggested that base J is involved in repression of recombination between repetitive sequences or retro elements and thus, stabilizing the genome. In fact, the majority of VSG switching events are due to DNA rearrangements within the subtelomeric repetitive DNA sequences. Furthermore, manipulation of the levels of J in the genome of *T. brucei* has supported its role in regulating telomeric DNA rearrangements (9). It would appear that *T. cruzi* adopted a different mechanism to survive in the host environment. Rather than the mono-allelic expression of a single VSG, *T. cruzi* cells simultaneously expresses large number of surface glycoproteins including TSs and mucins (13,34). Although there is no evidence for telomeric ESs in *T. cruzi*, we frequently find base J in these subtelomeric contingency gene families along with repetitive sequences/retro elements. Therefore, it is tempting to propose a role for base J in regulating recombination in the subtelomeric region of *T. cruzi*.

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

Leishmania species, the phylogenetically distant kinetoplastid of trypanosomatids, do not undergo antigenic variation and base J is restricted to the telomeric repeat sequences (27). They exhibit only a limited genetic variability in comparison with other kinetoplastids and do not contain active non-LTR-retro elements in their genome (40,41). Therefore, telomeric recombination may not be a vital event for its survival. This may explain the apparent lack of J outside the telomeric repeats

in *Leishmania*. In contrast, base J is abundant in telomeric-localized repeat arrays and essential gene ORFs in *T. brucei* and, as we demonstrate here, in *T. cruzi*. We have recently generated a *T. brucei* blood-stream form cell line that is unable to synthesize base J. Consistent with the proposed role of J in regulating DNA recombination versus transcription, preliminary analysis of the J-null trypanosome indicates a significant increase in telomeric exchange based VSG switching events and no transcriptional de-repression of the 19 silent telomeric ESs (Sabatini *et al.*, unpublished data). Future studies will focus on the consequence of a lack of J on the subtelomeric-localized contingency genes and repeat arrays of *T. cruzi*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. WHO, (2002) Control of Chagas disease. *WHO Technical Report Series*, **905**, 82–83.
2. Tyler,K. and Engman,D.M. (2001) The life cycle of *Trypanosoma cruzi* revisited. *Int. J. Parasitol.*, **31**, 3916–3923.
3. El-Sayed,N., Myler,P., Bartholomeu,D., Nilsson,D., Aggarwal,G., Tran,A.-N., Ghedin,E., Worthey,A.E., Delcher,A.L. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, **309**, 409–415.
4. Barry,J., Ginger,M.L., Burton,P. and McCulloch,R. (2003) Why are parasite contingency genes often associated with telomeres? *Int. J. of Parasitol.*, **33**, 29–45.
5. Voss,T., Healer,J., Marty,A.J., Duffy,M.F., Thompson,J.K., Beeson,J.G., Reeder,J.C., Crabb,B.S. *et al.* (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature*, **439**, 1004–1008.
6. Borst,P. and van Leeuwen,F. (1997) β-D-Glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. *Mol. Biochem. Parasitol.*, **80**, 1–8.
7. van Leeuwen,F., Taylor,M.C., Mondragon,A., Moreau,H., Gibson,W., Kieft,R. and Borst,P. (1998) β-D-Glucosyl-hydroxymethyluracil is a conserved DNA modification in kineto-plastid protozoans and is abundant in their telomeres. *Proc. Natl Acad. Sci. USA*, **95**, 2366–2371.
8. 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.
9. 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.
10. van Leeuwen,F., Kieft,R., Cross,M. and Borst,P. (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **109**, 133–145.
11. 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*. *Gen. Develop.*, **11**, 3232–3241.
12. Pereira-Chioccola,V. and Schenkman,S. (1999) Biological role of *Trypanosoma cruzi* trans-sialidase. *Biochem. Soc. Trans.*, **27**, 516–518.
13. Buscaglia,C., Campo,V., Frasch,A. and Di Noia,J. (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat. Rev. Microbiol.*, **4**, 229–236.
14. Kim,D., Chiurillo,M.A., El-Sayed,N., Jones,K., Santos,M.R., Porcile,P.E., Andersson,B., Myler,P., da Silveira,J.F. *et al.* (2005) Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. *Gene.*, **346**, 153–161.
15. Contreras,V., Morel,C. and Goldenberg,S. (1985) Stage specific gene expression precedes morphological changes during *Trypanosoma cruzi* metacyclogenesis. *Mol. Biochem. Parasitol.*, **14**, 83–96.
16. Contreras,V., Salles,J., Thomas,N., Morel,C. and Goldenberg,S. (1985) In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol. Biochem. Parasitol.*, **16**, 315–327.
17. Finn,R., Mistry,J., Schuster-Bockler,B., Griffiths-Jones,S., Hollich,V., Lassmann,T., Moxon,S., Marshall,M., Khanna,A. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, D247–D251.
18. Bendetsen,T., Nielsen,H., Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: signal P 3.0. *J. Mol. Biol.*, **340**, 783–795.
19. Krogh,A., Larsson,B., van Heijne,G. and Sonnhammer,E. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genome. *J. Mol. Biol.*, **305**, 567–580.
20. Quevilon,E., Silventoinen,V., Pillai,S., Hark,N., Mulder,N., Apweiler,R. and Lopez,R. (2005) Interproscan: Protein domains identifier. *Nucleic Acids Res.*, **33**, W116–W120.
21. Freitas-Junior,L., Porto,R., Pirrit,L., Schenkman,S. and Scherf,A. (1999) Identification of the telomere in *Trypanosoma cruzi* reveals highly heterogeneous telomere lengths in different parasite strains. *Nucleic Acids Res.*, **27**, 2451–2456.
22. Wickstead,B., Ersfeld,K. and Gull,K. (2003) Repetitive elements in genomes of parasitic protozoa. *Microbiol. Mol. Biol. Rev.*, **45**, 360–375.
23. Aksoy,S., Lalor,T.M, Martin,J., Van der Ploeg,L.H. and Richard,F.F. (1987) Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosomes, *Trypanosoma gambience*. *EMBO J.*, **6**, 3819–3826.
24. Villanueva,M.S., Williams,S.P., Beard,C.B., Richards,F.F. and Aksoy,S. (1991) A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. *Mol. Cell Biol.*, **11**, 6139–6148.
25. 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 hypermodi-fied base J in both strands. *Nucleic Acid Res.*, **24**, 2476–2482.
26. Genest,P., Riet,B., Cijsouw,T., Luenen,H.V. and Borst,P. (2007) Localization of the modified DNA base J in the genome of the protozoan parasite *Leishmania*. *Nucleic Acids Res.*, **35**, 2116–2124.
27. Ruef,B., Dawson,B.D., Tewari,D., Fouts,D.L. and Manning,J.E. (1994) Expression and evolution of members of the *Trypanosoma cruzi* trypomastigote surface antigen multigene family. *Mol. Biochem. Parasitol.*, **63**, 109–120.
28. Bogliolo,A., Lauria-Pires,L. and Gibson,W. (1996) Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. *Acta Trop.*, **61**, 31–40.

29. 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.

30. Yu,Z., Genest,P.-A., Riet,B.T., Sweeney,K., DiPaolo,C., Kieft,R., Christodoulou,E., Perrakis,A., Simmons,J.M. *et al.* (2007) The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acid Res.*, **35**, 2107–2115.

31. Briones,M.R., Egima,C.M. and Schenkman,S. (1995) *Trypanosoma cruzi* trans-sialidase gene lacking C-terminal repeats and expressed in epimastigote forms. *Mol. Biochem. Parasitol.*, **70**, 9–17.

32. Chaves,L.B., Briones,M.R. and Schenkman,S. (1993) Trans-sialidase from *Trypanosoma cruzi* epimastigotes is expressed at the stationary phase and is different from the enzyme expressed in trypomastigotes. *Mol. Biochem. Parasitol.*, **61**, 97–106.

33. Frasch,A.C.C. (2000) Functional diversity in the trans-sialidase and Mucin families in *Trypanosoma cruzi. Parasitol. Today*, **16**, 282–286.

34. Bhattacharya,S., Bakre,A. and Bhattacharya,A. (2002) Mobile genetic elements in protozoan parasites. *J. Genetics.*, **81**, 73–86.

35. Kazazian,H.J. (2004) Mobile elements: drivers of genome evolution. *Science*, **303**, 1626–1632.

36. Mieczkowski,P., Lemoine,F.J. and Petes,T.D. (2006) Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair*, **5**, 1010–1020.

37. Lander,E., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

38. International Human Genome Sequencing Consortium. (2001). *Nature*, **409**, 860–921.

39. Bringaud,F., Ghedin,E., Blandin,G., Bartholomeu,D.C., Caler,E., Levin,M.J., Baltz,T. and El-Sayed,N.M. (2006) Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol. Biochem. Parasitol.*, **145**, 158–170.

40. Kissinger,J. (2006) A tale of three genomes: the kinetoplastids have arrived. *Trends Parasitol.*, **22**, 240–243.