The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase

Zhong Yu[1],[†], Paul-André Genest[1],[†], Bas ter Riet[1], Kate Sweeney[2], Courtney DiPaolo[2], Rudo Kieft[2],[‡], Evangelos Christodoulou[3], Anastassis Perrakis[3], Jana M. Simmons[4], Robert P. Hausinger[4],[5], Henri G.A.M. van Luenen[1], Daniel J. Rigden[6], Robert Sabatini[2],[‡] and Piet Borst[1],*

[1]Division of Molecular Biology and Centre of Biomedical Genetics, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands, [2]Global Infectious Diseases Program Marine Biological Laboratory, Woods Hole, MA 02543, USA, [3]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands, [4]Department of Biochemistry & Molecular Biology and [5]Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI 48824-4320, USA and [6]School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, UK

ABSTRACT

Trypanosomatids contain an unusual DNA base J (β-D-glucosylhydroxymethyluracil), which replaces a fraction of thymine in telomeric and other DNA repeats. To determine the function of base J, we have searched for enzymes that catalyze J biosynthesis. We present evidence that a protein that binds to J in DNA, the J-binding protein 1 (JBP1), may also catalyze the first step in J biosynthesis, the conversion of thymine in DNA into hydroxymethyluracil. We show that JBP1 belongs to the family of Fe^{2+} and 2-oxoglutarate-dependent dioxygenases and that replacement of conserved residues putatively involved in Fe^{2+} and 2-oxoglutarate-binding inactivates the ability of JBP1 to contribute to J synthesis without affecting its ability to bind to J-DNA. We propose that JBP1 is a thymidine hydroxylase responsible for the local amplification of J inserted by JBP2, another putative thymidine hydroxylase.

INTRODUCTION

Eukaryotic unicellular kinetoplastid flagellates, such as *Trypanosoma* and *Leishmania* species, contain a unique modified base in their nuclear DNA, β-D-glucosylhydroxymethyl-uracil, or J (1–4). In *Trypanosoma brucei*, J replaces 0.5–1.0% of thymine (T) in nuclear DNA (5) and somewhat lower values have been found in other trypanosomatids (2) and *Euglena* (6). J is predominantly present in repetitive DNA sequences in *T. brucei* (7–9), and most prominently in the telomeric GGGTTA repeats (7). Indirect evidence indicates that J is made in two steps (Figure 1): hydroxylation of specific Ts in DNA is followed by addition of a glucose to the hydroxymethyluracil (HMU) formed (10–12). We have tried to find the thymidine hydroxylase and the glucosyl transferase that we expect to be involved in J biosynthesis in cell extracts, by complementation, and by RNAi knockdown of candidate genes, but these attempts have failed thus far. Hence, we have not been able to generate J-less parasites to unravel the function of base J.

A protein that specifically binds to J-containing duplex DNA was found in *T. brucei*, in *Leishmania* species and in the insect trypanosome *Crithidia fasciculata* (13–15). In *Leishmania* this 90-kDa J-binding protein (JBP1) is essential (16), but in *T. brucei* it is not. The absence of JBP1 in *T. brucei* has no effect on growth, repeat stability or gene expression, but results in a 20-fold decrease in J level relative to wild-type cells (17). This decrease in J level was not due to loss of protection by JBP1 against
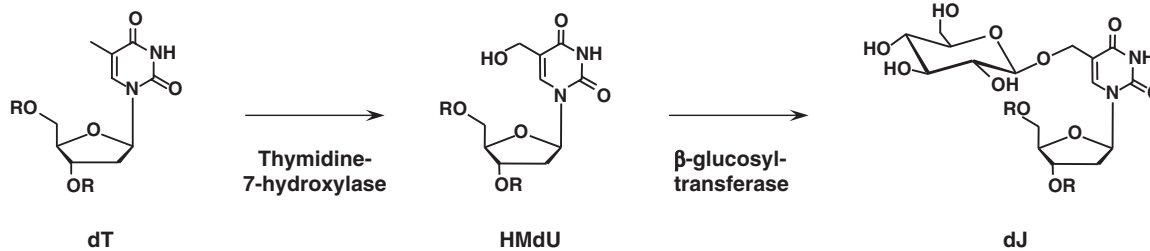
Figure 1. Proposed pathway for J biosynthesis. First, a thymidine hydroxylase catalyzes the formation in DNA of HMdU. Second, HMdU in DNA is converted into J by a putative glucosyl transferase.

removal of J from DNA and suggested that JBP1 is somehow involved in J maintenance (17).

A large protein with substantial homology with JBP1 (34% identity; 47% similarity) in its N-terminal half turned up in a database search (18). In its C-terminal half, this J-binding protein 2 (JBP2) contains a region homologous to proteins with SWI2/SNF2-like chromatin remodeling activity. Although there is no evidence that JBP2 can bind to J-DNA, DiPaulo *et al.* were able to show that JBP2 does affect J biosynthesis, because it can induce J synthesis in insect form trypanosomes, which normally do not express JBP2 (18) and are devoid of J (19,20). JBP2 is neither essential in *T. brucei*, nor in *Leishmania* (unpublished data).

As both JBP1 and JBP2 are able to increase the level of J in trypanosomes, we hypothesized that both proteins actually catalyze the first, and rate-limiting step in J biosynthesis, the hydroxylation of T in DNA. The homology of the N-terminal halves of JBP1 and JBP2 would then be due to the presence of the thymidine hydroxylase function in this part of the protein. Since the hydroxylation of T resembles the hydroxylation of methylated bases by the repair enzyme AlkB (21,22), we further hypothesized that the thymidine hydroxylase, like AlkB, would be a member of the family of dioxygenases that use Fe^{2+} and 2-oxoglutarate as cofactors (23–25). In this article, we provide evidence supporting this hypothesis for JBP1. Our results lead to a simple (but still speculative) model for the first step of J biosynthesis. In this model, JBP2 is responsible for the *de novo* region-specific synthesis of J, whereas JBP1 is responsible for the local maintenance and expansion of J.

MATERIALS AND METHODS

Bioinformatics

JBP1 and JBP2 sequences were obtained by PSI-BLAST (26) searches in the nr database combined with the genome databases of *T. brucei* (27), *T. cruzi* and *Leishmania major* (28) prior to their incorporation into nr. Their homologous regions were aligned using MUSCLE (29). Distant homologies of this common domain with other proteins were sought using the sensitive sequence comparison and fold-recognition tools of the Meta server (30). This server conveniently collects the results of analysis by various modern

sequence comparison and fold-recognition methods and applies the consensus 3D-Jury method (31) to them. The HHsearch software (32) for comparison of hidden Markov model representations of protein alignments was also used in conjunction with the domain databases Pfam (33), Smart (32), COG (34) and KOG (35). Structures were browsed in the SCOP database (36).

Mutagenesis

The mutations were made by site-directed mutagenesis using the Quik-Change site-directed mutagenesis kit (Stratagene) following the instruction of the manufacturer. A *Lt GFP-JBP1* gene fusion or a *Tb GFP-JBP1* gene fusion cloned in a *T. brucei* or a *Leishmania tarentolae* expression vector were used as a template. The different oligonucleotides used for the mutagenesis reaction as well as maps of the expression vectors used can be found in Supplementary Figure 1. Mutations were verified by sequencing using an ABI Prism 3700 DNA Analyzer (Applied Biosystems). The *GFP-JBP1* constructs were linearized and transfected into the *JBP1* null trypanosomes or transfected as episomes in *Leishmania* following standard protocols (37,38). *Leishmania* was cultured in SDM-79 medium (39) and the bloodstream form of *T. brucei* in HMI9 medium (40).

Anti-J-DNA immunoblot

Genomic DNAs of the *JBP1* null trypanosomes transfected with the various *GFP-JBP1* DNA constructs were isolated, serially diluted in steps of two, denatured, spotted on a nitrocellulose membrane and crosslinked. J levels were determined by incubation with the J antiserum as described before (8). Briefly, the membrane was blocked in skim milk, incubated with a polyclonal J antiserum, washed, incubated with a goat anti-rabbit antibody (horse-radish peroxidase conjugated) and washed, followed by enhanced chemiluminescence detection. The membrane was later hybridized with a telomeric or tubulin DNA probe in a formamide-based hybridization buffer at 42°C overnight. Washes were done with 6× SSC (0.9 M NaCl, 0.09 M Na citrate, pH 7.0), 0.1% SDS. The signal was detected by autoradiography.

Western blotting

Protein lysates were made by spinning down the parasites followed by resuspension of the pellet in Laemmli

buffer (41). Proteins were run on 8% SDS-PAGE gel, blotted according to previously described protocols (41). Membranes were blocked in skim milk and incubated with the Lt JBP1 or the Tb JBP1 antibodies as described previously (42).

Protein purification and nuclear extracts

A His tag version of the various JBP1 mutants was generated by PCR and was ligated in a pET-46 Ek/LIC or pET-16b *E. coli* expression vector (Novagen) following the instructions of the supplier. The constructs were transformed in the BL21(DE3)pLysS or in the Rosetta2(DE3)pLysS *E. coli* strains (Novagen). Protein expression was induced by addition of IPTG (isopropyl-beta-D-thiogalactopyranoside) at a final concentration of 0.5 mM in Luria-Bertani broth, for 3 h at 37°C or overnight at 15°C. Proteins were purified using His-Select Nickel Magnetic Beads (Sigma) according to the instructions given by the manufacturer and eluted with imidazole. Nuclear extracts of the *Leishmania* transfected with the Lt GFP-JBP1 mutants were made as described before (13).

Electromobility shift assay

The partially purified proteins were used in an electro-mobility shift assay following published protocols (13). Briefly, the purified proteins were incubated with radioactively labeled double-stranded J-containing DNA oligonucleotides in a buffer containing 25 mM Hepes-NaOH pH 7.9, 100 mM KCl, 5 mM $MgCl_2$, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 0.1 μg poly-dIdC per μl and 0.5 mg bovine serum albumin (BSA) per ml. Proteins were run on a 4.5% native polyacrylamide gel for 2 h in 0.5× TBE (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA). The gel was dried and the signal detected by autoradiography. The Lt His-JBP1/J-DNA complex was supershifted using an Lt JBP1 antiserum (42). The Lt GFP-JBP1/J-DNA complex was supershifted using a GFP antiserum (Clontech).

Microscopy

GFP localization of the Lt JBP1-GFP fusions in the Tb JBP1 null trypanosomes. Trypanosomes were spun down and washed in 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.3) and stained with DAPI. The samples were then embedded in low melting 2% agarose in 1× PBS, spotted on a glass slide, sealed with a cover slip and placed at 4°C for a minimum of 1.5 h to ensure little movement of the trypanosomes. Images were acquired using a motorized Zeiss Axioplan 2 and a Axiocam MRm camera which were controlled by the Axiovisons 4.4 software.

GFP localization of the Tb JBP1-GFP in the Tb JBP1 null trypanosomes. Trypanosomes were fixed in 1% formaldehyde in 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.3). They were then spotted and dried on a glass slide and embedded in 5 μM ToPro3 in Vectashield mounting medium (Vector

Laboratories, Inc.). Images were acquired using a Leica-NT confocal microscope and the Leica confocal software.

RESULTS

JBP1 and JBP2 share a conserved domain with the hydroxylase superfamily

JBP1 and JBP2 share amino acid sequence identity over a region of ∼270 residues. Iterative database searches with PSI-BLAST (26) starting from this region pick out the JBP1/2 family but nothing more. Similarly, extensive iterative PSI-BLAST (26) searches starting from nucleic-acid-modifying members of the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily, such as AlkB, (43) did not reach JBP1 or JBP2 sequences. In order to explore the possibility of a remote relationship, undetectable even by PSI-BLAST, further analyses were done. Secondary structure predictions for the 270 residue homologous region showed an all-β portion flanked by two largely α-helical parts. We worked with this all-β region, as well as with the entire JBP1/2 conserved region. Both native and derived consensus sequences were submitted for analysis to the Meta server (30). Matches to Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases were indeed present, although rarely top-scoring by individual methods. The most significant result by an individual method was the score of 3.84 (corresponding to a 90% confidence level) given to the central all-β portion of the conserved domain of *T. brucei* JBP1, in a hit to clavaminate synthase by FUGUE (44). Although scores were below confidence levels (45), the top hits by the 3D-Jury consensus method, when applied to the same sequence, were dominated by the dioxygenases and by proteins with the cupin fold. In the Structural Classification of Proteins Database (36), these architectures are united at the fold level as they share β-helix structure. At large evolutionary distances, it is no surprise that the two are confused by fold-recognition programs. Analysis with the hidden Markov model comparison program HHsearch (46) was also supportive of structural correspondence between JBP1/2 and Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases; dioxygenases monopolized the top hits in the domain databases tested, although the best E-value obtained was only 0.2.

Two arguments further supported the existence of a distant homology. The first was the excellent correspondence between JBP1/2-predicted secondary structure and the actual secondary structures of the dioxygenases in the conserved all-β core (a structural fold comprising of eight β-strands found in all Fe^{2+}/2-oxoglutarate-dependent dioxygenases) (Figure 2). The second and more telling factor was the presence in the JBP1/2 sequences of conserved residues at positions matching the key Fe^{2+} binding residues of the dioxygenases. Two His residues and one Asp residue are usually involved in Fe^{2+} binding and are invariant in the aligned JBP1/2 sequences (Figure 2). In addition, an Arg residue, also highly conserved in dioxygenases and important for binding the 2-oxoglutarate, is present in the JBP1/2 sequences
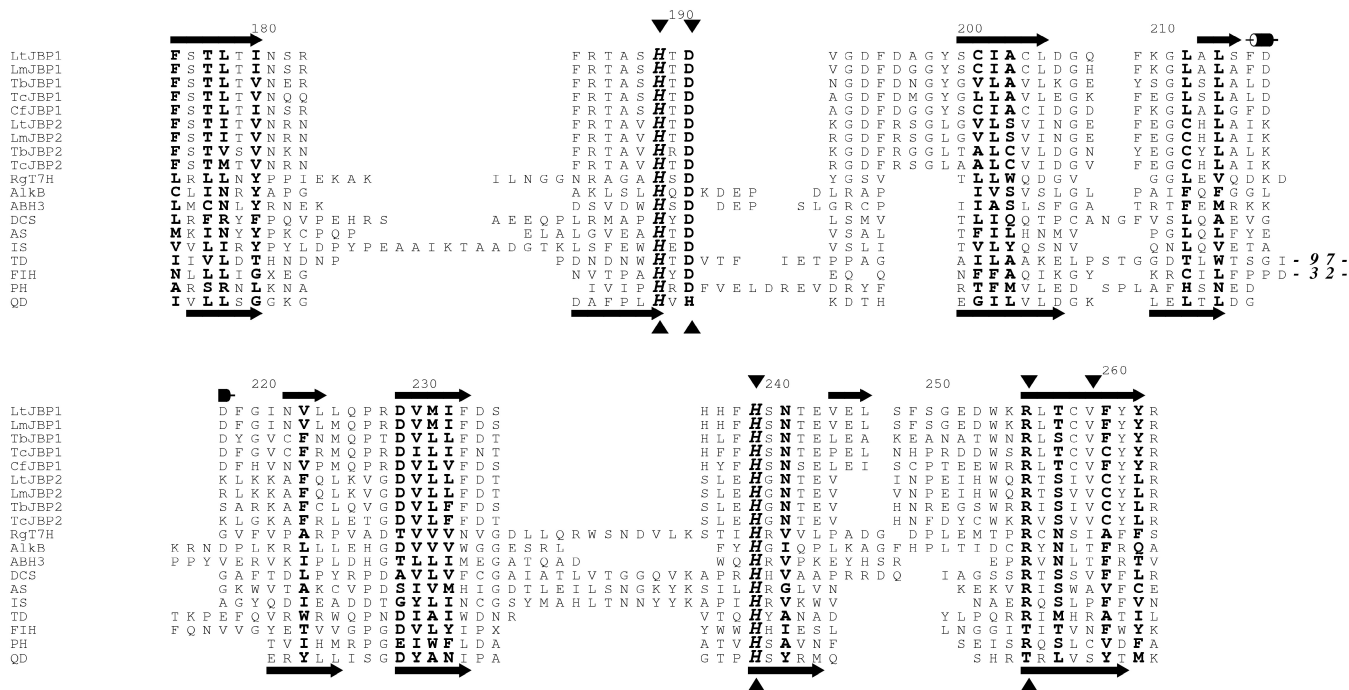
Figure 2. Sequence alignment of the conserved all-β core (see text) of JBP1 and JBP2 sequences with diverse members, of known structure, from the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily. JBP1 and JBP2 sequences are given with species as follows: Tb, *T. brucei*; Tc, *T. cruzi*; Lt, *L. tarentolae*; Lm, *L. mexicana* and Cf, *Crithidia fasciculata*. The gi codes in the nr database (Wheeler *et al.*, 2005) are: 62361410 for Lt JBP1, 68124616 for Lm JBP1, 6018041 for Tb JBP1, 71421637 for Tc JBP1, 6018043 for Cf JBP1, 68125217 for Lm JBP2, 71750205 for Tb JBP2 and 71422266 for Tc JBP2. The dioxygenase structures are as follows: DCS, deacetoxycephalosporin C synthase (PDB code 1dcs) (47); AS, anthocyanidin synthase (1gp6); IS, isopenicillin N synthase (1obn); TD, taurine dioxygenase (1os7); FIH, factor inhibiting HIF (1iz3); PH, proline 3-hydroxylase (1e5s); QD, quercetin 2,3-dioxygenase (1y3t). The better conserved positions within the conserved core are shown in bold font with invariant positions additionally italicized. Numbers in the alignment represent the positions of the amino acids in Lt JBP1 and mark the positions of large deletions that are not shown for clarity. The predicted secondary structure of Lt JBP1 is shown above the alignment and the actual secondary structure of quercetin 2,3-dioxygenase (PDB code 1y3t) is shown below the alignment. The long arrows represent β-sheets and the tubes α-helices. Arrowheads below the alignment indicate the four key residues, largely conserved throughout the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily that bind the iron (His, Asp) or 2-oxoglutarate (Arg) (23,24). The residues of Lt JBP1 that were mutated to alanines: H189, D191, H239, R255, V259, are shown by the arrowheads above the alignment. Note that there are near identical pairs of sequences for both JBP1 and JBP2 in the *T. cruzi* genome: only one of each pair is shown in the alignment.

(Figure 2). As Figure 2 and recent analyses (23,24) make clear, the all-β core with these four largely conserved residues is the only typical feature within the structures of the dioxygenase superfamily. The structures of family members outside this core cannot generally be super-imposed (data not shown), and even within the core substantial insertions are sometimes present (Figure 2). This unusual degree of structural variability probably explains the relatively poor scores for dioxygenase folds produced for the JBP1/2 common domain.

Effects of amino acid replacements in JBP1 on its ability to rescue a double inactivation of *JBP1* in *Leishmania*

If JBP1 is an Fe^{2+}-and 2-oxoglutarate-dependent thymidine hydroxylase catalyzing the first step of J biosynthesis (see Figure 1), mutation of the three residues essential for Fe^{2+}-binding, or the Arg residue important for the binding of 2-oxoglutarate, should abolish the putative hydroxylase activity. To test this, we mutated the His 189, the Asp 191, the His 239 and the Arg 255 of a GFP-tagged version of the JBP1 of *L. tarentolae* to alanines by site-directed mutagenesis (Figure 2, Supplementary Figure 1). As JBP1 is

essential in *Leishmania* (16), we could not assess the functionality of the mutants by transfecting the constructs in a *JBP1* null *Leishmania*. We therefore attempted to inactivate both wild-type chromosomal *JBP1* alleles in cell lines that expressed the wild-type, or a mutant Lt GFP-JBP1 from a rescue plasmid. This was possible with the wild type *JBP1* construct, but not with the mutant *JBP1* constructs (as seen by the generation of cell lines still expressing the endogenous wild-type protein) (Supplementary Figure 2), even though the mutant protein routed normally to the nucleus of *Leishmania* (data not shown).

Effects of amino acid replacements in JBP1 on its ability to support J biosynthesis in the *JBP1* null trypanosomes

Our *Leishmania* experiments show that replacement of amino acids critical for the putative hydroxylase function of JBP1 abrogates the ability of JBP1 to complement JBP1 null mutants, but these experiments do not prove that JBP1 converts T into HMU. To test this putative hydroxylase function more directly we turned to *T. brucei*, in which the absence of JBP1 is not lethal, but results in a 20-fold reduction of J in DNA (17).
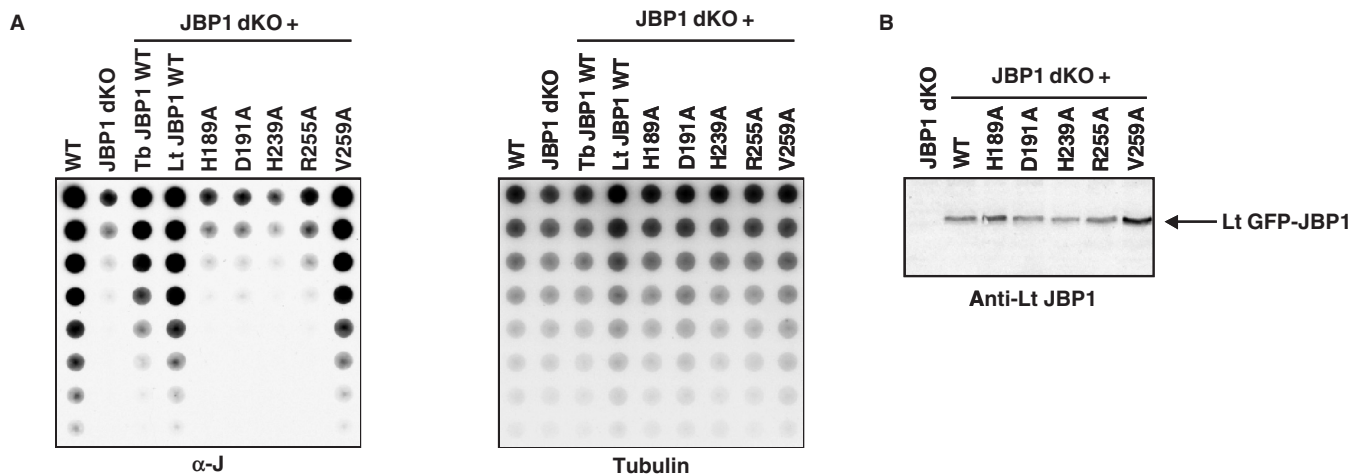
Figure 3. Determination of J levels in the DNA of JBP1 null trypanosomes expressing the Lt JBP1 mutants. (**A**) Genomic DNA of wild-type *T. brucei* bloodstream form 427 line (WT), JBP1 null trypanosomes (JBP1 dKO) and JBP1 null trypanosomes expressing the wild-type Tb GFP-JBP1 fusion (Tb JBP1 WT), the wild-type Lt GFP-JBP1 fusion (Lt JBP1 WT) or the mutant Lt GFP-JBP1 fusions (H189A, D191A, H239A, R255A, V259A) were serially diluted in steps of two, denatured, spotted on a nitrocellulose membrane and incubated with a polyclonal J antiserum (left panel). The membrane was hybridized with a tubulin DNA probe as a loading control (right panel). (**B**) Western blot on lysates of the JBP1 null trypanosomes (JBP1 dKO) and the JBP1 null trypanosomes expressing the wild-type (WT) and the mutant (H189A, D191A, H239A, R255A, V259A) Lt GFP-JBP1 fusions using an Lt JBP1 antiserum.

Table 1. Characteristics of the Lt JBP1 mutants. The position of the amino acids replaced in these mutants is indicated in Figure 2

Lt JBP1 mutants	WT	H189A	D191A	H189A + D191A	H239A	R255A	V259A
Rescues J levels in *JBP1* null *T. brucei*	Yes	No	No	No	No	No	Yes
Nuclear location in *T. brucei*	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Binds to J-DNA	Yes	Yes	Yes	Yes	Yes	Yes	Yes

We transfected wild-type and mutant *Lt GFP-JBP1* gene constructs into the *JBP1* null trypanosomes and determined the level of J in the genomic DNAs of the transfectants using the J antiserum (8). Wild-type Lt GFP-JBP1 was equally effective as the Tb GFP-JBP1 in raising the J levels of the *JBP1* null trypanosomes approximately to wild-type levels (Figure 3A), but the H189A, the D191A, the H239A and the R255A mutants were not (Figure 3A and Table 1). As a control, we replaced Val 259. This residue is conserved in all JBPs analyzed and close to the 'hydroxylase' domain, but it is not conserved in members of the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily (Figure 2). Its replacement might therefore not affect the putative hydroxylase function. Indeed, the Val 259 mutant protein did complement the *JBP1* null trypanosomes and resulted in wild-type J levels (Figure 3A and Table 1).

We also verified that the putative Fe^{2+}-binding His 213 and Asp 215 of *T. brucei* JBP1 (corresponding to the His 189 and the Asp 191 of Lt JBP1) are essential for the putative hydroxylase function. Replacement of these residues by serines resulted in mutant proteins that failed to restore the J-level in the *JBP1* null trypanosomes to wild-type levels, whereas a wild-type Tb GFP-JBP1 fusion did (Supplementary Figure 3, Supplementary Table 1). The expression levels of the wild type and mutants of the Lt and Tb GFP-JBP1 fusions were comparable, excluding the possibility that the mutants did not complement the J levels because they were poorly expressed in the parasite (Figure 3B, data not shown). All wild type and mutant GFP-JBP1 fusions of Lt JBP1 and Tb JBP1 tested were targeted to the nucleus of *T. brucei* ruling out the possibility that the mutants were not functional because they were not correctly routed to the nucleus (Table 1, Figure 4, Supplementary Table 1 and Supplementary Figure 4).

Effects of amino acid replacements in JBP1 on its ability to bind to J-DNA

We previously found that mutant versions of JBP1 that lost their J-binding property do not restore the J levels in the *JBP1* null trypanosomes (our unpublished data). The fact that the putative hydroxylase mutants did not rescue the *JBP1* inactivation in *Leishmania* and did not restore the J levels in the *JBP1* null *T. brucei* could therefore be due to their inability to bind to J-DNA. To exclude this possibility, we purified His-tagged recombinant wild-type and mutant Lt JBP1 from *E. coli* (Figure 5A) and tested these for J-binding in electromobility shift assays using radioactively labeled double-stranded J-containing DNA oligonucleotides (13,14). A J-DNA-dependent bandshift was obtained with the wild type as well as with all the mutated versions of Lt JBP1 we tested (Figure 5 and Table 1). The shifted J-DNA band, which contains a 1:1
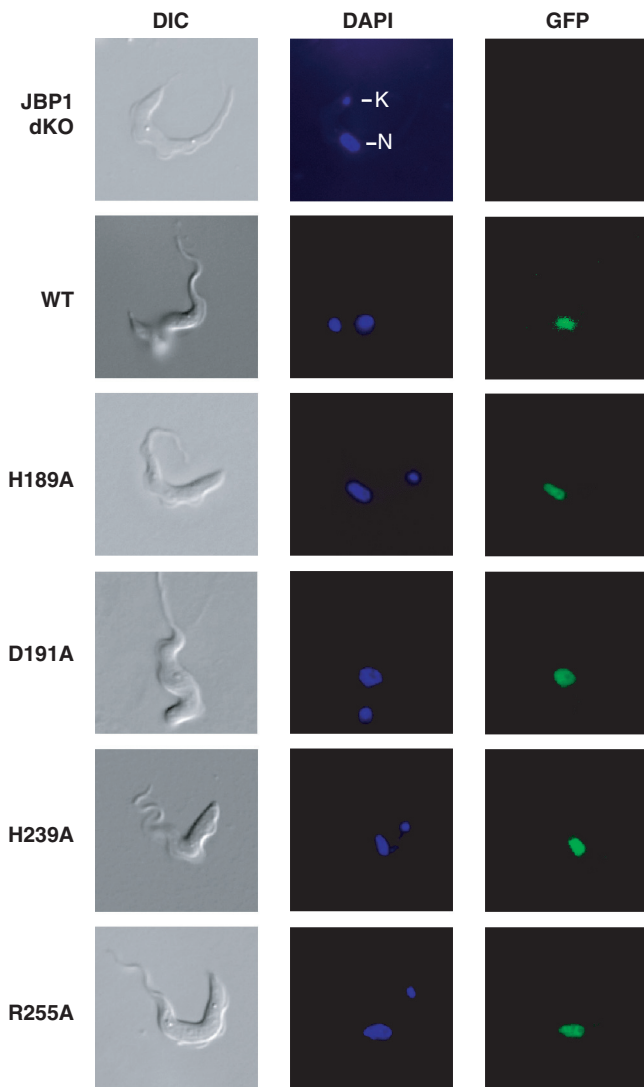
DIC	DAPI	GFP

Figure 4. Localization of mutant Lt GFP-JBP1 in JBP1 null trypanosomes. JBP1 null trypanosomes (JBP1 dKO) were transfected with the wild-type (WT) or the mutant (H189A, D191A, H239A, R255A) Lt GFP-JBP1 fusions, DAPI-stained and visualized live on agarose slides by microscopy. DAPI staining shows the location of the nucleus (N) and of the kinetoplast (mitochondrial DNA; K).

ratio of JBP1:DNA, could be supershifted using an Lt JBP1 antiserum (42), implying that the shift was indeed due to JBP1 (Figure 5B). As reported previously (13,14), purified JBP1 does not bind to DNA without J under our assay conditions (Figure 5B). The results in Figure 5 indicate that wild-type and mutant forms of Lt JBP1 have similar affinity for J-DNA. Protein titration of the double mutant (H189A plus D191A) supports this conclusion (Supplementary Figure 5).

To exclude the possibility that the GFP-tagged versions of the mutant proteins produced in *Leishmania* would not bind to J-DNA, whereas the His-tagged versions produced in *E. coli* would (Figure 5), we made nuclear extracts of the *Leishmania* cell lines expressing the H189A, the D191A and the H189A + D191A Lt GFP-JBP1 mutant proteins and

tested these for J-binding activity in electromobility shift assays. Mutant proteins bound to J-DNA are shown in Supplementary Figure 6. We conclude from these results that the H189A, the D191A, the H239A and the R255A Lt JBP1 mutants failed to rescue the J levels in the *JBP1* null trypanosomes because they lost their putative thymidine hydroxylase activity, not because they can no longer bind to J-DNA.

The relation of JBP1 and JBP2 to other members of the dioxygenase superfamily

Given the experimental evidence for the correctness of the dioxygenase fold assignment for JBP1 and JBP2, the relationships of JBP1/2 with other branches of the dioxygenase superfamily were explored further. Do JBP1/2 bear a particular similarity with other enzymes that modify nucleic acids such as AlkB and, especially, the thymine 7-hydroxylase, cloned recently from *Rhodotorula glutinis* (48), which also converts T into HMU? Alternatively, are they unexpectedly more closely related to other enzymes such as deacetoxycephalosporin C synthase (DCS) which featured in the top hits of several fold-recognition programs? These questions were addressed using PSI-BLAST and HHsearch. PSI-BLAST runs to convergence when initiated with *R. glutinis* T7H or AlkB, but failed to reach JBP1/2 sequences, indicating that there is only a distant relationship between JBP1/2 and these nucleic-acid-modifying dioxygenases. Interestingly, *R. glutinis* T7H bears only a distant relationship to AlkB, being more closely related to enzymes such as isopenicillin N synthase and gibberellin 20-oxidase (data not shown). Also, JBP1/2 sequences were not retrieved in PSI-BLAST searches starting from DCS. We then chose 22 sequences to cover the full range of the dioxygenase superfamily in the Pfam database PF03171 (33) and initiated PSI-BLAST runs with each. No run retrieved JBP1/2 sequences, confirming that only a very distant evolutionary relationship exists between JBP1/2 and all other dioxygenase superfamily members characterized to date. HHsearch-based comparison of a JBP1/2 conserved domain alignment with alignments of 55 AlkB sequences, 11 DCS proteins and 8 T7H sequences told a similar tale: E-values were similar—0.01 for T7H, 0.02 for AlkB and 0.05 for DCS—and indicative of only very distant homologies. Therefore, it appears that the AlkB, T7H and JBP1/2-catalyzed thymidine hydroxylase activities, all base modifying, represent three phylogenetically distant branches within the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily.

DISCUSSION

Our results show that JBP1 has the properties to be expected of a thymidine hydroxylase catalyzing the first step in J biosynthesis:

(1) JBP1 has all the sequence/structural hallmarks of an Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase/hydroxylase.
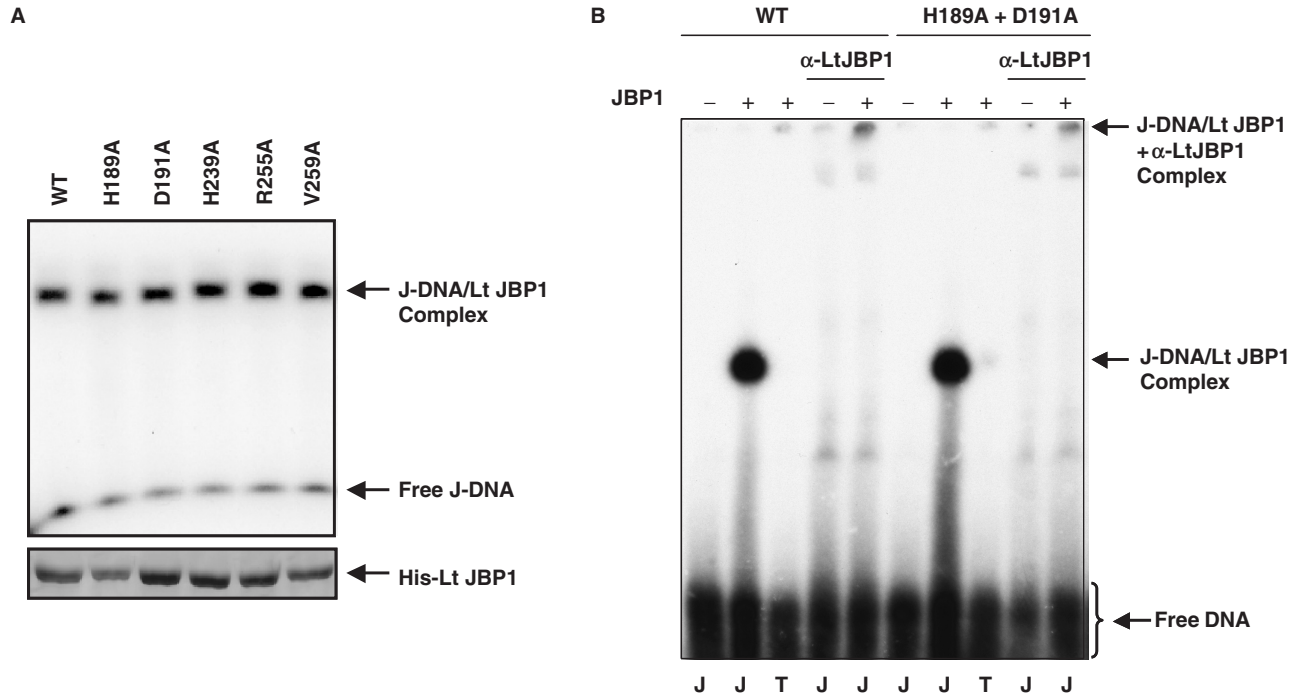
Figure 5. Determination of the J-binding activity of purified Lt JBP1 mutants by electromobility shift assay. (**A**) Upper panel: Electromobility shift assay with the purified His-tagged wild-type (WT) and mutant (H189A, D191A, H239A, R255A, V259A) Lt JBP1 proteins. Proteins (10 fmol) were incubated with a radioactively labeled double-stranded oligonucleotide containing J (J-DNA) (10 fmol) and run on 4.5% native acrylamide gel. The gel was dried and the signal detected by autoradiography. The bands corresponding to the J-DNA/Lt JBP1 and the free J-DNA are indicated. Lower panel: Segment of a Coomassie-brilliant-blue-stained acrylamide gel with the His-tagged wild-type (WT) and mutant (H189A, D191A, H239A, R255A, V259A) Lt JBP1 proteins purified from *E. coli*. Only the segment of the gel with the His-tagged Lt JBP1 band is shown. (**B**) Example of a determination of the J-binding activity of mutant Lt JBP1 by electromobility shift assay. His-tagged wild-type (WT) and H189A + D191A mutant Lt JBP1 proteins were purified from *E. coli* and incubated with a radioactively labeled double-stranded oligonucleotide containing J (J) or lacking J (T) in the presence or absence of a Lt JBP1 antibody (α-Lt JBP1). Samples were run on a 4.5% native acrylamide gel. The gel was dried and the signal detected by autoradiography. Lanes without recombinant protein are indicated by '–'. The position of the free DNA, the J-DNA/Lt JBP1 complex and the J-DNA/Lt JBP1 + α-Lt JBP1 complex is indicated.

(2) Replacement of amino acids predicted to be essential for hydroxylase function abolishes the ability of JBP1 to stimulate J synthesis. We have mutated the His and Asp residues essential for Fe^{2+} binding and the Arg important for 2-oxoglutarate binding in other members of this dioxygenase family. Replacement of these residues with Ala or Ser inactivates the putative hydroxylase function, as shown by the fact that the mutated *JBP1* is unable to complement the *JBP1* null trypanosomes. The mutations were introduced in a *GFP-JBP1* fusion gene, allowing us to show that the mutant fusion protein is made in *T. brucei* and *Leishmania* cells in similar amount as wild-type protein and that it is properly routed to the nucleus.

(3) Replacement of residues essential for the putative hydroxylase function does not affect the ability to recognize and bind J-DNA. In experiments with purified mutant and wild-type recombinant Lt JBP1 proteins, we have shown that the affinity of the JBP1 mutants for J-DNA is not substantially altered. Moreover, the GFP-tagged mutant proteins still bind to J-DNA in extracts of *L. tarentolae* as shown by bandshift of J-containing duplex oligonucleotides. The J-DNA binding by JBP1 is therefore independent of the putative hydroxylase function.

(4) Absence of JBP1 in *T. brucei* reduces the J level in nuclear DNA 20-fold (17). Although this could be an indirect effect, the result is fully compatible with a hydroxylase function for JBP1. This also holds for the ability of JBP1 to promote retention of J introduced into DNA by growing trypanosomes in medium containing HMdU (17). This nucleoside is randomly incorporated into DNA, both in bloodstream form and in insect-form trypanosomes (10), and in the presence of JBP1 this extra J is more sluggishly lost than in its absence (17).

Although these results support our hypothesis that JBP1 is an Fe^{2+}-and 2-oxoglutarate-dependent dioxygenase that catalyzes the hydroxylation of T in DNA to yield HMU, our attempts to demonstrate hydroxylase activity with purified JBP1 have failed thus far (see Supplementary Data, Materials and Methods for a detailed description of these experiments). Dioxygenases belonging to this enzyme family are often able to catalyze the oxidative decarboxylation of 2-oxoglutarate in the

absence of primary substrate (24), but we have been unable to detect succinate formation or oxygen consumption by JBP1 with or without J-containing oligonucleotide substrate, using a highly purified His-tagged JBP1 preparation from *C. fasciculata*. Negative results were also obtained in another *in vitro* assay in which the conversion of T into HMU by Cf JBP1 or Lt JBP1 (in the presence of Fe^{2+}, 2-oxoglutarate and a J-DNA oligonucleotide) was analyzed with a specific DNA base excision glycosylase for HMU (11). Finally, we looked in vain for thymidine hydroxylase activity in *E. coli* overproducing JBP1 of *L. tarentolae* using the HMU-specific glycosylase.

These negative results might raise the question whether JBP1 is really a hydroxylase or only affecting thymidine hydroxylation indirectly, e.g. by recruiting a hydroxylase to the DNA. We consider this highly unlikely. First, the possibility that a domain shared by members of the Fe^{2+}- and 2-oxoglutarate-dependent dioxygenase superfamily is present in a protein recruiting a thymidine hydroxylase is small. Second, the likelihood that replacement of conserved amino acids in this domain affects the recruitment of the hydroxylase is remote. Third, it is hard to believe that residues critical for hydroxylase function would be conserved over the long evolutionary time separating *T. brucei* and *Crithidia*, if these residues would not serve the same function as in other members of the dioxygenase family. The fact that we have not yet been able to demonstrate hydroxylase activity in our JBP1 preparations, does not mean much, like most negative results. Nothing is known yet about the nature of the DNA substrate for JBP1. It may simply require longer DNA than the oligonucleotides used thus far, DNA packaged in chromatin, or additional cofactors, etc.

The analysis of JBP2 remains to be carried out, but the available evidence supports a hydroxylase function for JBP2 as well: Its N-terminal half contains all conserved elements characteristic of Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases. Absence of JBP1 in *T. brucei* does not result in a complete loss of J, indicating the presence of a second hydroxylase, i.e. JBP2 (17). Expression of *JBP2* in insect form *T. brucei* (which does not normally contain any J) results in the formation of J (18).

On the basis of our results, we propose that the *de novo* modification of T to HMU is exclusively catalyzed by JBP2 and that it is JBP2 that determines the region, and context specificity of J in nuclear DNA (18). The low level of HMU made by JBP2 is then converted into J and locally amplified 20-fold by JBP1 (17). Although this procedure for HMU formation looks unduly complex, the need for two enzymes could stem from the requirement to insert HMU in very specific places in repetitive DNA sequences packed in chromatin. The initial precise recognition of this sequence may be a slow process, or the number of recognition sites for JBP2 might be limited. This could explain that a second enzyme is required to finish the job and raise the J level further, making use of the HMU already inserted by JBP2 (after the conversion of HMU into J). The presence of two hydroxylases, each with a unique sequence and substrate specificity, may explain why we have never been able to find a common theme in the sequences containing J, even though these sequences are highly non-random (7).

REFERENCES

1. Gommers-Ampt,J.H., Van Leeuwen,F., De Beer,A.L.J., Vliegenthart,F.G., Dizdaroglu,M., Kowalak,J.A., Crain,P.F. and Borst,P. (1993) β-D-Glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan. *Trypanosoma brucei. Cell*, **75**, 1129–1136.
2. Van Leeuwen,F., Taylor,M.C., Mondragon,A., Moreau,H., Gibson,W., Kieft,R. and 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. U.S.A.*, **95**, 2366–2371.
3. Gommers-Ampt,J.H. and Borst,P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.
4. Borst,P. and Van Leeuwen,F. (1997) Beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. *Mol. Biochem. Parasitol.*, **90**, 1–8.
5. Van Leeuwen,F., De Kort,M., van der Marel,G.A., Van Boom,J.H. and Borst,P. (1998) The modified DNA base beta-D-glucosyl-hydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by ^{32}P-postlabeling. *Anal. Biochem.*, **258**, 223–229.
6. 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.
7. 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.
8. Van Leeuwen,F., Wijsman,E.R., Kieft,R., van der Marel,G.A., Van Boom,J.H. and Borst,P. (1997) Localisation of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei. Genes Dev.*, **11**, 3232–3241.
9. Van Leeuwen,F., Kieft,R., Cross,M. and Borst,P. (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by β-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei. Mol. Biochem. Parasitol.*, **109**, 133–145.
10. 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.*, **10**, 5643–5651.
11. Ulbert,S., Cross,M., Boorstein,R.J., Teebor,G.W. 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.

12. 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.

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

14. 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.

15. 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.

16. Genest,P.A., ter Riet,B., Dumas,C., Papadopoulou,B., Van Luenen,H.G. and Borst,P. (2005) Formation of linear inverted repeat amplicons following targeting of an essential gene in *Leishmania. Nucleic Acids Res.*, **33**, 1699–1709.

17. 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.

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

19. Blundell,P.A., Van Leeuwen,F., Brun,R. and Borst,P. (1998) Changes in expression site control and DNA modification in *Trypanosoma brucei* during differentiation of the bloodstream form to the procyclic form. *Mol. Biochem. Parasitol.*, **93**, 115–130.

20. Van Leeuwen,F., Dirks-Mulder,A., Dirks,R.W., Borst,P. and Gibson,W. (1998) The modified DNA base beta-D-glucosyl-hydroxymethyluracil is not found in the tsetse fly stage of *Trypanosoma brucei. Mol. Biochem. Parasitol.*, **94**, 127–130.

21. Falnes,P.O., Johansen,R.F. and Seeberg,E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli. Nature*, **419**, 178–182.

22. Yu,B., Edstrom,W.C., Benach,J., Hamuro,Y., Weber,P.C., Gibney,B.R. and Hunt,J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879–884.

23. Aravind,L. and Koonin,E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.*, **2**, research 7.1–research 7.8.

24. Hausinger,R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxy-lases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.*, **39**, 21–68.

25. Clifton,I.J., McDonough,M.A., Ehrismann,D., Kershaw,N.J., Granatino,N. and Schofield,C.J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.*, **100**, 644–669.

26. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

27. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei. Science*, **309**, 416–422.

28. Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major. Science*, **309**, 436–442.

29. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

30. Bujnicki,J.M., Elofsson,A., Fischer,D. and Rychlewski,L. (2001) Structure prediction meta server. *Bioinformatics*, **17**, 750–751.

31. Ginalski,K., Elofsson,A., Fischer,D. and Rychlewski,L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **19**, 1015–1018.

32. Letunic,I., Copley,R.R., Schmidt,S., Ciccarelli,F.D., Doerks,T., Schultz,J., Ponting,C.P. and Bork,P. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res.*, **32**, D142–D144.

33. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.

34. Tatusov,R.L., Natale,D.A., Garkavtsev,I.V., Tatusova,T.A., Shankavaram,U.T., Rao,B.S., Kiryutin,B., Galperin,M.Y., Fedorova,N.D. *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.*, **29**, 22–28.

35. Koonin,E.V., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Krylov,D.M., Makarova,K.S., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N. *et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.*, **5**, R7.

36. Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.

37. Papadopoulou,B., Roy,G. and Ouellette,M. (1992) A novel antifolate resistance gene on the amplified H circle of *Leishmania. EMBO J.*, **11**, 3601–3608.

38. Carruthers,V.B., Van der Ploeg,L.H.T. and Cross,G.A.M. (1993) DNA-mediated transformation of bloodstream form *Trypanosoma brucei. Nucleic Acids Res.*, **21**, 2537–2538.

39. Brun,R. and Schönenberger,M. (1979) Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi defined medium. *Acta Tropica*, **36**, 289–292.

40. Hirumi,H. and Hirumi,K. (1989) Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.*, **75**, 985–989.

41. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

42. 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.

43. Falnes,P.O. and Rognes,T. (2003) DNA repair by bacterial AlkB proteins. *Res. Microbiol.*, **154**, 531–538.

44. Shi,J., Blundell,T.L. and Mizuguchi,K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substi-tution tables and structure-dependent gap penalties. *J. Mol. Biol.*, **310**, 243–257.

45. Ginalski,K. and Rychlewski,L. (2003) Detection of reliable and unexpected protein fold predictions using 3D-Jury. *Nucleic Acids Res.*, **31**, 3291–3292.

46. Soding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.

47. Deshpande,N., Addess,K.J., Bluhm,W.F., Merino-Ott,J.C., Townsend-Merino,W., Zhang,Q., Knezevich,C., Xie,L., Chen,L. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.

48. Smiley,J.A., Kundracik,M., Landfried,D.A., Barnes,V.R.Sr and Axhemi,A.A. (2005) Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a Rhodotorula glutinis cDNA library and iso-orotate decarboxylase from *Neurospora crassa. Biochim. Biophys. Acta*, **1723**, 256–264.