BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **234,** 95–100 (1997)
ARTICLE NO. RC976583

Cloning and Sequence Analysis of the DNA Polymerase α Gene of *Leishmania donovani:* Comparison with the Human Homologue

Kim Luton and Alan. M. Johnson[1]
Molecular Parasitology Unit, Department of Cell and Molecular Biology, University of Technology, Sydney, Australia

Received March 17, 1997

The gene encoding the DNA polymerase α catalytic subunit of the kinetoplastid parasite *L. donovani* has been isolated, sequenced and compared with other eukaryotic homologues. The coding region is 4020 bp in length and specifies an inferred protein sequence of 1339 amino acids (aa). There is a high level of variability between the human and *L. donovani* gene sequences, but functional substrate-binding residues identified in humans and yeast appear to also be conserved in this parasite. The discovery of a cysteine-rich region located in the midst of the active sites of the enzyme, which appears to be unique to the Kinetoplastids, and aa differences found between some of the conserved regions implicated in catalytic function, may aid in drug design. The putative DNA binding Zn finger at the C-terminus of the protein appears highly species specific and may have potential as a drug target for blocking enzyme catalysis in the parasite.

Leishmania donovani is a flagellated heteroxenous protozoan parasite which causes a visceral infection in humans, usually resulting in death in 2 to 3 years if untreated. Approximately 200 million people are at risk of visceral leishmaniasis annually in 47 countries (1). Current chemotherapeutic treatments are often toxic to the host, are expensive and difficult to administer, and drug resistance is increasing worldwide due to the ability of protozoan parasites to develop alternate metabolic pathways for negating the blocking effect of drugs. In some endemic areas up to 70% of patients may have mucocutaneous or visceral leishmaniasis that is unresponsive to certain drugs (2). There is clearly a need for new safe, effective drugs which can treat the whole range of *Leishmania* infections.

A strategy to target one of the important replicative enzymes in DNA synthesis for the design of an antiparasitic drug was recently suggested by Biñas and Johnson (3). Given that DNA polymerase α is the most abundant cellular DNA polymerase and is essential for cellular multiplication and proliferation, it is an ideal target for rational drug design. The enzyme comprises one large catalytic subunit and three small subunit polypeptides and is responsible for lagging strand DNA synthesis. To date only two protozoan parasite DNA polymerase α genes have been sequenced (4,5). Additional gene sequences need to be determined to better characterise the differences between the human host and parasite enzymes. A recent study of *Leishmania mexicana* DNA polymerase found the major activity to be from an α-like DNA polymerase which exhibited different responses to both specific mammalian DNA polymerase α inhibitors and antisera (6). This suggested that *L. donovani* would have a similar pattern of polymerase activity. If specific inhibitors to the DNA polymerase α catalytic subunit could be designed for *L. donovani* then parasite proliferation could be prevented by selectively inhibiting DNA replication of the parasite without affecting the human host.

Here we report that the complete gene which codes for the catalytic subunit of DNA polymerase α of *L. donovani* has been sequenced and compared with the homologous gene sequences of other eukaryotes. Differences between the *L. donovani* and human DNA polymerase α sequences which may aid in the design of specific inhibitors of the parasite enzyme have been identified.

MATERIALS AND METHODS

Genomic DNA isolation and PCR amplification of probe sequence. Leishmania donovani donovani MHOM/IN/80/DD8 strain, obtained from the World Health Organisation *Leishmania* cryobank, was routinely cultured in HO-MEM medium (7). Genomic DNA was prepared

[1] Correspondence to: Molecular Parasitology Unit, Department of Cell and Molecular Biology, University of Technology, Sydney, Westbourne St, Gore Hill, NSW 2065, Australia. Fax: (02) 9514 4026; E-mail: Alan.Johnson@uts.edu.au.

0006-291X/97 $25.00

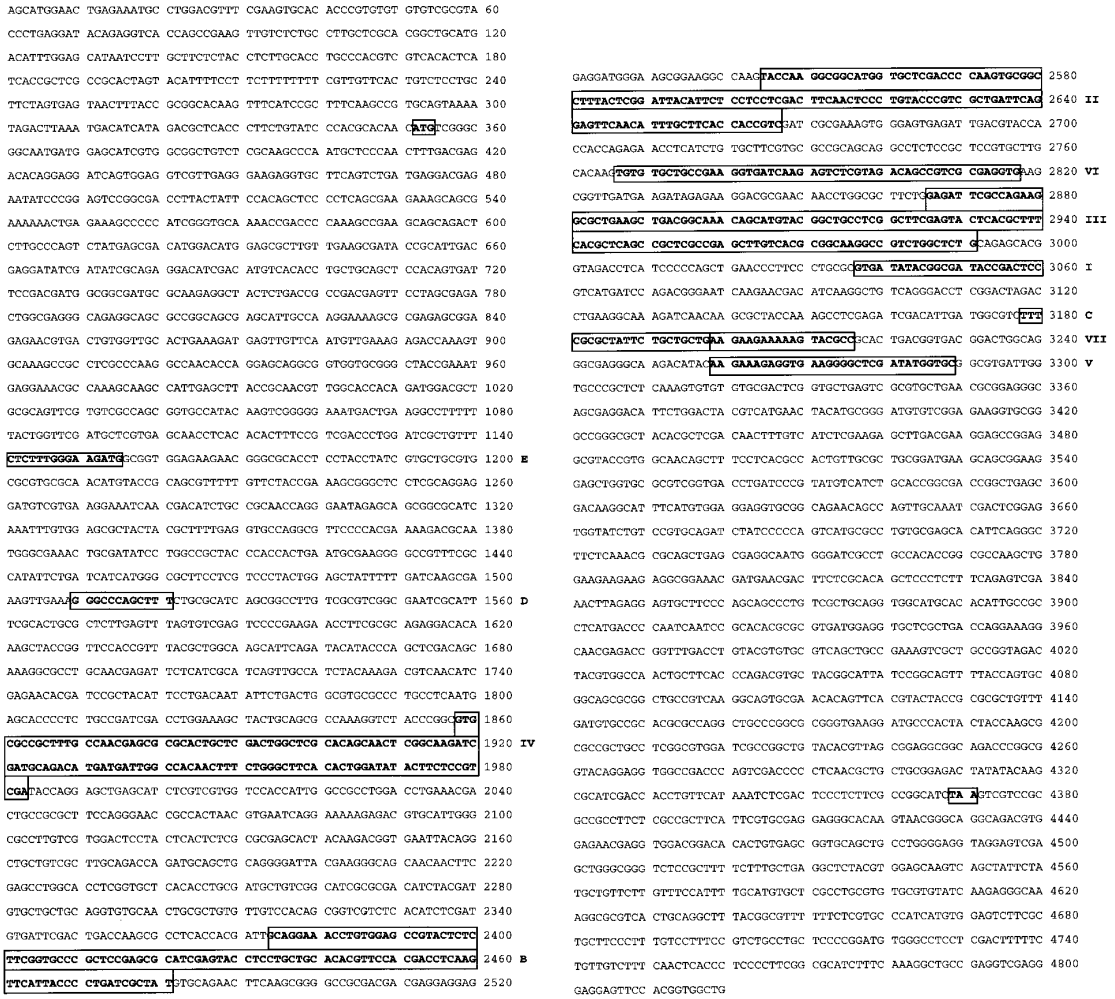

FIG. 1. Nucleotide sequence of the *L. donovani* DNA polymerase α gene. The start codon beginning at position 352 and the stop codon ending at position 4371 are indicated in boxes. Conserved regions I to VII and B to E are indicated in bold and outlined, as described by Mansour et al. (25).

using a standard protocol (8). A PCR primer was designed using the aligned DNA polymerase α nucleotide sequences of *Trypanosoma brucei* (GenBank accession number X60951), *Oxytricha nova* (GenBank accession number U02001) and *Plasmodium falciparum* (GenBank accession number L18758) with a *L. donovani* codon usage table (9) to limit the level of degeneracy. Forward primer Lpol2 ($^{5'}$-CTG GAC TTC AAT A/T G/C C CT G/T TAT CC-$^{3'}$) was used in conjunction with reverse primer R1Ro ($^{5'}$-ATC ACT GAG TCC GTG TCG CCG TA-$^{3'}$), a primer designed for amplifying *Toxoplasma gondii* DNA polymerase α (Biñas and Johnson, unpublished), to amplify a PCR product spanning conserved region II to conserved region I. The reaction mixtures (67 mM Tris-HCl pH8.8, 3.5 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 0.5 mM KCl, 0.45% Triton X-100, 0.2 mg/ml gelatin, 200 μM each dNTP, 50 pmol each primer and 100 ng genomic DNA in 100 μl final volume) underwent 26 cycles of denaturing at 93°C for 1 min, annealing at 40°C for 2 min and extension at 72°C for 2 min, with the initial denaturation for 3 min and the final extension for 5 min. The PCR product was cloned into pUC19 vector and sequenced on a LiCor 4000L automated DNA sequencer (USA) using the LiCor Sequitherm Long Read kit to perform cycle sequencing reactions.

Construction and sceening of library. A genomic library was made by ligating partially *Sau*3AI digested genomic DNA from *L. donovani* into λ GEM11 *Xho*I half site arms according to the manufacturer's instructions (PROMEGA). The cloned PCR product was radiolabelled with ^{32}P-dATP using a Megaprime Labelling kit (Amersham,UK) and used as a probe for screening the library.

Sequence analysis. Restriction enzyme fragments from the phage clones were subcloned into pUC19 vector and sequenced on a LiCor automated sequencer as described above. Contiguous sequence was compiled by walking out into the gene with fragments overlapping the highly conserved sequence of the PCR product. The complete sequence was analysed for open reading frames (ORF) to determine the coding region of the gene using MacVector™ Version 3.5. The sequence has been submitted to the GenBank database under the accession number U78172.

RESULTS AND DISCUSSION

Primers Lpol2 and R1Ro produced a PCR product of the predicted size of 460 bp. After cloning and sequencing, the product was compared with the *T. brucei* DNA polymerase α gene (data not shown). This revealed 69% identity at the nucleotide level and 79% identity at the

TABLE 1

Percentage Identity[a] between the Conserved Regions of DNA Polymerase α Sequences (I to VII and B to E)

	1	2	3	4	5
1. *L. donovani*	—	81.5	62.9	48.7	68.5
2. *T. brucei*	69.5	—	62.1	49.1	68.1
3. *O. nova*	55.2	56.9	—	50.0	65.9
4. *P. falciparum*	39.3	44.0	51.9	—	53.0
5. *H. sapiens*	59.8	60.1	61.9	46.7	—

Note. Above diagonal shows aa comparisons, below diagonal shows nucleotide comparisons.

[a] Both the nucleotide sequence and the inferred protein sequence of the coding region were aligned with the corresponding *T. brucei*, human, *O. nova* and *P. falciparum* DNA polymerase α sequences using Clustal W Version 1.4 (24). Percentage identity results were derived from the genetic distances calculated using PAUP version 3.1.1. Introduced alignment gaps were not counted.

aa level. The strong homology to *Trypanosoma* supported the use of this clone as a probe to screen the genomic library for the remainder of the gene.

Three positive phage clones were isolated and rescreened to purity. Restriction enzyme mapping revealed that the 16, 18 and 20 kb clones completely overlap one another. Southern blotting of the restriction enzyme digested clones and hybridisation with the PCR product indicated that all three clones contained the entire DNA polymerase α gene (data not shown). The 16 kb phage clone was further mapped, with overlapping fragments subcloned and sequenced.

All seven conserved regions characteristic of DNA polymerases were identified in the *L. donovani* gene sequence (Figure 1). Conserved regions exclusively characteristic of DNA polymerase α, designated as regions B to E (10), were also identified. Like *T. brucei* and *P. falciparum*, *L. donovani* does not have the conserved A region at the 5′ end of the gene. When the aligned sequences of *L. donovani*, *T. brucei*, *O. nova*, *P. falciparum* and humans were compared for the conserved regions only, they showed a high level of homology to one another (Table 1). *L. donovani* and *T. brucei* were found to have 81.5% identity at the aa level and 69.5% identity at the nucleotide level, supporting the identification of the *L. donovani* gene sequence as DNA polymerase α. Calculations of both nucleotide and aa sequence variation for the entire gene demonstrate the close relationship of *Leishmania* to *Trypanosoma* but indicate low levels of similarity overall (Table 2). The *P. falciparum* sequence has been previously shown to be the most divergent DNA polymerase α yet reported (5), and this is confirmed by the new comparison with *L. donovani* (Table 2).

An article on the polymerase activity of *L. donovani*, published towards the end of this sequence analysis, has indicated that the major activity peak was not DNA polymerase α but was actually unlike any other DNA polymerase yet characterised (11). This is surprising given that the major activity peak isolated from *L. mexicana* was α-like and that the gene sequence reported here contains all the regions thought to be exclusively characteristic of DNA polymerase α. It is possible that one of the other two activity peaks not described in the *L. donovani* study (11), may in fact contain DNA polymerase α activity.

The sequence contains only one ATG start codon followed by a long ORF (Figure 1). The ORF begins with aa sequence which contains strong homology to the N-terminal region of *T. brucei* DNA polymerase α (data not shown). There are no TATAA, CCAAT or GC boxes in the 352 bp of 5′ leader sequence, and there is no poly(A) addition signal (5′-AATAAA) in the 448 bp of 3′ trailer sequence. This is similar to the pattern for human DNA polymerase α which has been reported to have no TATA box in the 5′ leader although a CCAAT box was found on the opposite strand (12). The *L. donovani* ORF is 4020 bp long including the stop codon, and specifies a putative polypeptide of 1339 aa.

The alignment of the inferred aa sequences for *L. donovani* and the human DNA polymerase α gene is shown in Figure 2. Although sequence comparisons showed human and *Leishmania* to have only 28.3% identity across the entire gene (Table 2), a comparison of their conserved regions shows 68.5% identity (Table 1). It has been suggested that many of these regions are essential for catalytic function due to their conservation across a diverse range of organisms (12). Putative functional domains have been identified in the *L. donovani* sequence based on aa sequence alignments with human and with *T. brucei*. Region D (alignment positions 504-508) has been shown to be conserved in humans, *T.brucei* and yeast at approximately the same position in relation to the C-terminus suggesting that similar protein-protein interactions may take place in each of these (4). Mutation studies on yeast DNA polymerase α have implicated the highly conserved glycine residue at alignment position 504 (Figure 2) in primase subunit interaction (13), suggesting that this domain may be involved in primase binding in *Leishmania*.

TABLE 2

Percentage Identity[a] between Sequences of the Entire DNA Polymerase α Coding Region

	1	2	3	4	5
1. *L. donovani*	—	57.3	29.9	21.0	28.3
2. *T. brucei*	59.0	—	31.2	21.9	29.0
3. *O. nova*	38.9	39.9	—	21.8	31.3
4. *P. falciparum*	32.3	36.0	47.5	—	22.1
5. *H. sapiens*	42.9	44.6	42.3	38.4	—

Note. Above diagonal shows aa comparisons, below diagonal shows nucleotide comparisons.

[a] See footnote *a* under Table 1.

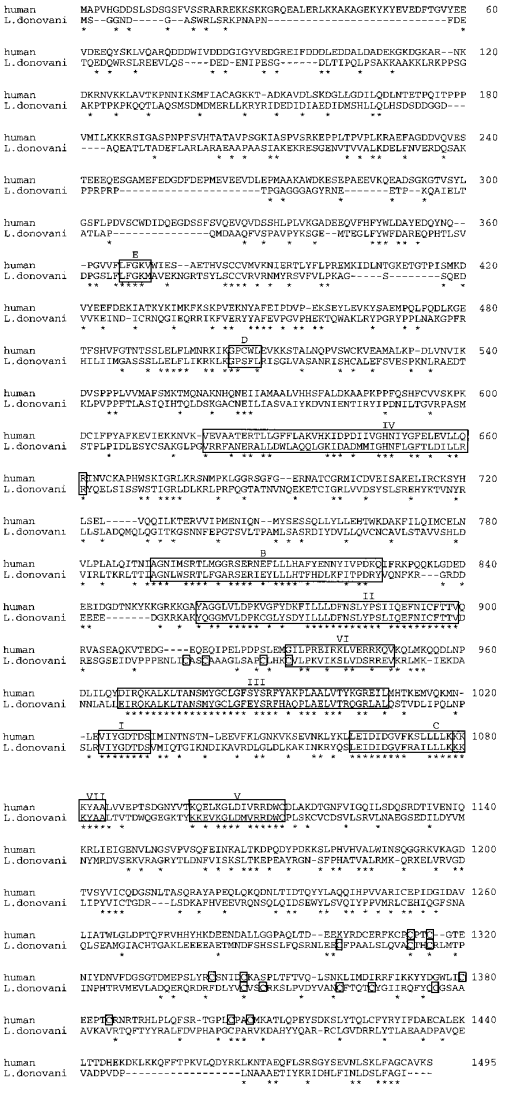

FIG. 2. Alignment of *L. donovani* and human DNA polymerase α inferred protein sequences. Conserved regions I to VII and B to E are indicated in boxes, as described in Mansour et al. (25). Cysteine residues implicated in DNA binding zinc fingers are outlined. Sequence identity is indicated by *. *L. donovani* nucleotide sequence was translated to aa sequence using MacVector Version 3.5 and aligned with the human aa sequence (GenBank accession number X06745) using Clustal W Version 1.4.

Although the glycine residue is present in *L. donovani*, unlike *P. falciparum* where it has been replaced by leucine (5), the remainder of region D is less well conserved.

There are also several aa differences between the *L. donovani* and human sequences in regions II and III (Figure 2). However, specific aa residues in regions I, II and III of human DNA polymerase α that have been implicated in metal activator binding (14), dNTP and primer binding (15,16) and dNTP binding (17) respectively, are all conserved in the *L. donovani* gene sequence. Several of these aa residues were also identified as being conserved between human DNA polymerase α and the corresponding active regions of *E. coli* pol I, HIV-1 RT and T7 RNA pol, whose crystalline structures are known (18). Although no crystalline structure has been determined for the polymerase α class of DNA polymerases, they have been predicted to form the "cupped right hand" tertiary structure found for *E. coli* pol I, HIV-1 RT and T7 RNA pol, where the polymerase domain forms a "fingers", "palm" and "thumb" subdomains (19). Homology of the predicted secondary structure of human DNA polymerase α regions I, II and III to the known structures of the active sites supports this (18), though there may be some topological changes in the "fingers" subdomain due to differences in spacing between these regions. The predicted secondary structures for regions I to III of *L. donovani* closely match those of the human sequence (Figure 3), indicating that regions I and II are in the "palm" subdomain of the active site and region III forms an α helix in the "fingers" subdomain. The conservation of these residues between human and *Leishmania* supports the hypothesis that these are critically active sites for all DNA polymerase α enzymes.

Cysteine-rich regions homologous to those in *T. brucei* have been located in *L. donovani* close to the C-terminal end. It has been suggested that the closely spaced cysteine (Cys) residues, conserved in humans, yeast and *T. brucei*, may form DNA binding zinc fingers (4). *T. brucei* was described as having three Cys pairs (4), but comparison with *L. donovani* suggests that these kinetoplastids may both have four Cys pairs, shown at alignment postions 1301-1376 in *L. donovani* ($^{5'}$-C-X_{10}-C-X_2-C-X_{30}-C-X_2-C-X_{11}-C-X_4-C-X9-C$^{3'}$). An aa sequence alignment of the protein sequences of humans, yeast, *Oxytricha* and *Plasmodium* shows that the Cys-rich region of these organisms also contains four Cys pairs but although most of the spatial arrangement of the eight Cys's is well conserved (Figure 4a), the level of conservation among intervening residues is poor (Figure 4b). Different iterations of the GCG (19) alignment programme Pileup consistently showed that only four of the Cys sites appear to be conserved among all the species, suggesting that these may be the residues involved in metal ion coordination (Figure 4b).

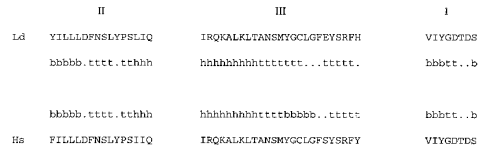

FIG. 3. Comparison of secondary structure predictions for the three most conserved regions of *L. donovani* (Ld) and human (Hs) DNA polymerase alpha, where a = α helix; b = β sheet; t = turn; . = undefined. The secondary structure Chou-Fasman/Garnier-Osguthorpe-Robson predictions were made using the GCG (19) program Peptidestructure.

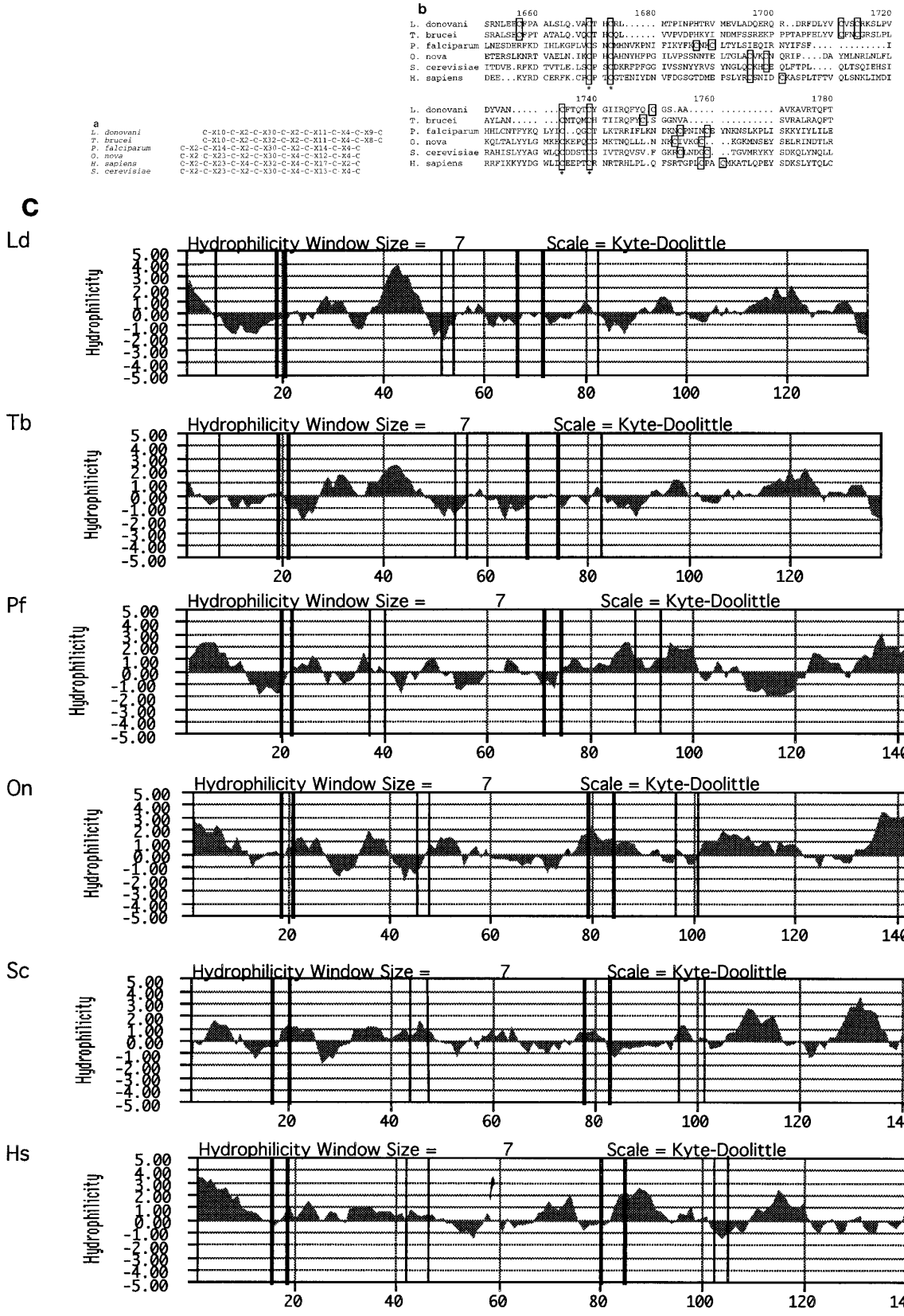

FIG. 4. Analysis of putative DNA binding Zinc finger. (a) Spatial arrangement of cysteine residues in C-terminal region. C = cysteine; X = any aa. (b) Multiple sequence alignment of cysteine-rich region taken from GCG-Pileup alignment (19) of entire predicted aa sequences. Cysteine residues involved in putative Zn-finger are boxed (4,5,25,12,21). * indicates possible conserved sites. (c) Hydrophilicity plots of cysteine-rich region determined using MacVector™ 3.5. Vertical lines indicate position of cysteine residues; bold lines indicate possible conserved sites. Ld, *L. donovani;* Tb, *T.brucei;* Pf, *P. falciparum;* On, *O. nova;* Sc, *S. cerevisiae;* Hs, *H. sapiens.*

This is supported by the fact that the region between these Cys's shows a high number of hydrophilic residues (Figure 4c), which have the potential to interact with nucleic acids (21). The conservation of these Cys-sites is also supported by an earlier study which aligned all available DNA polymerase sequences (22). In fact, the postion of these four Cys residues is also conserved in all the DNA polymerase δ sequences (22). The last four Cys residues in the human sequence were earlier proposed to form a DNA binding loop (23), but with many more sequences now available for alignment it appears that the conserved Cys positions shown on Figure 4b have better potential as a DNA binding Zn finger. There is a high level of variability among species in this region, which would make sequence alignment difficult without the fact that areas either side of this exhibit greater homology. If the region between the conserved Cys sites is important for DNA binding, the high level of sequence variation could have implications for the stringent species specificity that DNA polymerase α shows for intiation and replication of chromosomal DNA (24).

It is interesting to note that a second Cys cluster found in *T. brucei*, between conserved regions II and VI, is also present in *L. donovani* at alignment positions 917 - 933 ($^{5'}$-C-X_2-C-X_8-C-X_3-C-$^{3'}$). It is not present in the human, yeast or *Oxytricha* polymerases. This Cys-rich region contains mostly hydrophobic aa's however, which means it is less likely to be a nucleic acid binding motif. The DNA binding role of these domains in DNA polymerase α genes needs to be experimentally tested to determine whether the putative zinc fingers are important for catalytic activity. If either of the unique Cys-rich regions in *L. donovani* are used for DNA binding as the primary structure seems to suggest, then they may be useful targets for drug design.

REFERENCES

1. Ashford, R. W., Desjeux, P., and deRaadt, P. (1992) *Parasitol. Today* **8,** 104–105.
2. Ouellette, M., and Papadopoulou, B. (1993) *Parasitol. Today* **9,** 150–153.
3. Biñas, M., and Johnson, A. M. (1994) *Appl. Parasitol.* **35,** 157–168.
4. Leegwater, P. A., Stating, M., Murphy, N. B., Kooy, R. F., van der Vliet, P. C., and Overdulve, J. P. (1991) *Nucleic Acids Res.* **19,** 6441–6447.
5. White, J. H., Kilbey, B. J., de Vries, E., Goman, M., Alano, P., Cheesman, S., McAleese, S., and Ridley, R. G. (1993) *Nucleic Acids Res.* **21,** 3643–3646.
6. Nolan, L. L., Rivera, J. H., and Khan, N. N. (1992) *Biochim. Biophys. Acta* **1120,** 322–328.
7. Berens, R. L., and Marr, J. J. (1978) *J. Parasitol.* **64,** 160.
8. Kelly, J. M. (1993) Isolation of DNA and RNA from *Leishmania. In* Methods in Molecular Biology (Hyde, J. E., Ed.), Vol 21, Protocols in Molecular Parasitology, Humana Press, Totowa, NJ.
9. Alvarez, F., Robello, C., and Vignali, M. (1994) *Mol. Biol. Evol.* **11,** 790–802.
10. Damagnez, V., Tillit, J., de Recondo, A.-M., and Baldacci, G. (1991) *Mol. Gen. Genetics* **226,** 182–189.
11. Sen, S., Khalkho, N. V. M., and Majumder, H. K. (1996) *Biochem. Biophys. Res. Comm.* **221,** 662–669.
12. Wang, T. S.-F. (1991) *Ann. Rev. Biochem.* **60,** 513–552.
13. Pizzagalli, A., Valsasnini, P., plevani, P., and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3772–3776.
14. Copeland, W. C., and Wang, T. S.-F. (1993) *J. Biol. Chem.* **268,** 11028–11040.
15. Dong, Q., Copeland, W. C., and Wang, T. S.-F. (1993a) *J. Biol. Chem.* **268,** 24163–24174.
16. Dong, Q., Copeland, W. C., and Wang, T. S.-F. (1993b) *J. Biol. Chem.* **268,** 24175–24182.
17. Dong, Q., and Wang, T. S.-F. (1995) *J. Biol. Chem.* **270,** 21563–21570.
18. Copeland, W. C., Dong, Q., and Wang, T. S.-F. (1995) *Methods Enz.* **262,** 294–303.
19. Sousa, R. (1996) *TIBS* **21,** 186–190.
20. GCG suite of programs described *in* Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, WI, USA.
21. Berg, J. M. (1986) *Science* **232,** 485–486.
22. Braithwaite, D. K., and Ito, J. (1993) *Nucleic Acids Res.* **21**(4), 787–802.
23. Wang, T. S.-F., Wong, S. W., and Korn, D. (1989) *FASEB J.* **3**(1), 14–21.
24. Francesconi, S., Copeland, W. C., and Wang, T. S.-F. (1993) *Mol. Gen. Genet.* **241,** 457–466.
25. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
26. Mansour, S. J., Hoffman, D. C., and Prescott, D. M. (1994) *Gene* **144,** 155–161.