View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Volume 311, number 2, 124–127 FEBS 11642 October 1992
© 1992 Federation of European Biochemical Societies 00145793/92/$5.00

Characterization of a multi-copy gene for a major stage-specific cysteine proteinase of *Leishmania mexicana*

Augustine E. Souza[a], Sheila Waugh[a], Graham H. Coombs[b] and Jeremy C. Mottram[a]

[a]*Wellcome Unit of Molecular Parasitology, Institute of Genetics, University of Glasgow, Church Street, Glasgow G11 5JS, Scotland, UK* and [b]*Laboratory for Biochemical Parasitology, Department of Zoology, University of Glasgow, Glasgow G12 8QQ, Scotland, UK*

Received 1 September 1992

lmcpb, a gene from *Leishmania mexicana* that encodes a major cysteine proteinase in the parasite, has been cloned and sequenced. LmCPb is related more to cysteine proteinases from *Trypanosoma brucei* and *Trypanosoma cruzi* than to a previously characterized cysteine proteinase, LmCPa, of *L. mexicana*. It contains a long C-terminal extension characteristic of similar enzymes of *T. brucei* and *T. cruzi*. The gene is multi-copy and tandemly arranged. *lmcpb* RNA levels are developmentally regulated with steady state levels being high in amastigotes, low in metacyclic promastigotes and undetectable in multiplicative promastigotes. This variation correlates with and may account for the stage-specific expression of LmCPb enzyme activity.

Leishmania mexicana; Cysteine proteinase; Amastigote-specific; Developmental regulation

1. INTRODUCTION

Parasites of the genus *Leishmania* cause visceral and cutaneous disease in humans and other animals [1]. In its life-cycle, *Leishmania* must infect two hosts, a mammal and an insect. The environments encountered in the two hosts impose differing adaptive requirements on the organism and the three main forms of the parasite, the amastigote in the mammalian macrophage, the multiplicative promastigote in the insect mid-gut, and the metacyclic promastigote in the insect mouthparts, are morphologically and biochemically very distinct [1]. One feature of the amastigotes of *L. mexicana* is the very high levels of cysteine proteinases (CPs) most of which are stage-specific and are thought to be involved in enabling survival within macrophages [2]. As part of our studies to assess the role of CPs in survival, and possibly virulence, we have cloned and sequenced a cDNA encoding LmCPb, a CP of *L. mexicana* primarily expressed in amastigotes [3].

2. EXPERIMENTAL

Details of the culture and harvesting of *L. m. mexicana* (MNYC/BZ/62/M379) and purification of *L. mexicana* CPs are described elsewhere [4,5]. A 450 bp CP gene fragment was PCR-amplified from *L. mexicana* genomic DNA by using two degenerate oligonucleotide primers: no. 1 (5′ CCGAATTCCARGGICARTGYGGIWSITGYTGG 3′); and no. 2 (5′ CCAAGCTTCCAISWRTTYTTIACDATCCARTA 3′) which correspond to highly conserved regions of eukaryotic CP genes [6,7]. The fragment was cloned and sequenced and found to encode part of a CP gene different from another recently identified *L. mexicana* CP gene, *lmcpa* [4]. A cDNA library of *L. mexicana* amastigotes constructed in λZAP II (Stratagene) was screened with the cloned PCR fragment and several hybridizing plaques were detected. Phagemids from several positive plaques were rescued, subcloned and sequenced by the Sanger method [8] with Sequenase (USB) using plasmid and single-stranded DNA as templates; both strands of the largest cDNA, pSWb1a, were completely sequenced over the region encompassing the open reading frame. The DNA sequence (deposited in the EMBL nucleic acid database with accession number Z14061) was analysed using the University of Wisconsin GCG package [9]. The entire insert was labelled with [^{32}P]dCTP by random priming and used as described previously [4,6] to probe Southern blots of *L. mexicana* DNA digested with various restriction enzymes and Northern blots of total RNA isolated from logarithmic phase promastigotes (which comprise mainly multiplicative promastigotes), stationary phase promastigotes (which contain a high proportion of metacyclic promastigotes), and from amastigotes. A ^{32}P end-labeled oligonucleotide complementary to the small subunit ribosomal RNA (5′ TACAATGGTCTCTAATCATCTTC 3′) [10] was used to probe the Northern blot as a control for RNA loading.

Correspondence address: J.C. Mottram, Institute of Genetics, Church Street, University of Glasgow, Glasgow G11 5JS, Scotland, UK. Fax: (44) (41) 330-5994.

3. RESULTS AND DISCUSSION

A 2.2 kb cDNA clone, pSWb1a, was isolated from an amastigote cDNA library by screening with a PCR fragment amplified using degenerate primers based on highly conserved regions of CPs. The cDNA was full-length since the 5′ end carried 14 nt of the splice leader sequence which is present at the 5′ ends of nuclear mRNAs in *Leishmania* [11] and the 3′ end had a poly(A) stretch. The 5′ untranslated region (UTR) of the cDNA, defined by the sequence between the 3′ end of the splice leader and the first in-frame ATG codon, was 144 nt long. We did not sequence all of the 3′ UTR for

Published by Elsevier Science Publishers B.V.

```
                   -125                       -100
LmCPb              MATSRAALCAVAVVCVVLAA..ACAPARAIHVGTPAAALFEEFKRTYGRAYETLAEEQQRLA
TcCP                MSGW-RALSL-A-L--M.-CLVP-ATASL-AEETL-SQ-V---QKH--V--SA---RF--S
TbCP               MPR-EMVRFVRLP--LLAM--CL-SVALGSL--EESLEMR-AA--KK--KV-KDAK--AF-FR
LmCPa             MARRNPLLFAIV-TILF--CYGSALI-QTPPPVDNFV-S-HYGS--KRH-K-FGGD---GH-FN
DdCP1                         MK-ILLFVLAVFTVFVSSRGIPPEEQSQ-L--QDKFNKK-.SHE-YLE-FE
HsCPL                         MNPTLI--FCL-GIASATLTFDHSLE-QWTKW-AMHN-L-.GMN--GW-R-

                        -50
LmCPb        NFERNLELMREHQARNPHAQFGIT....KFFDLSEAEFAARYLNGAAYFAAAKRHAAQHYRKARADLSA
TcCP         V-RE--F-A-L-A-A----T--V-...P.-S--TRE--RS--H----H----QER-RVPVNVEVVG...
TbCP         A--E-M-QAKIQA-A--Y-T--V-...P.-S-MTRE--R---R---S-----QKRLRKTVNVTTGR...
LmCPa        A-KQ-MQTAYFLNTQ----HYDVS...G--A--TPQ---KL---.PD-Y-RHLKNHKEDVHVDDS....
DdCP1        I-KS--GKIE-LNLIAINHKADTKFGVN--A---SD--KNY---NKE..-IFTDDLPVADYLDDEFINS
HsCPL        VW-K-MKMIEL-NQEYREGKHSF-MAMNA-G-MTSE--RQV....MN..GFQN-KPRKGKVFQEPLFYE

             1                                                  50
LmCPb        VPD...AVDWREKGAVTPVKDQGACGSCWAFSAVGNIEGQWYLAGHELVSLSEQQLVSCD..........
TcCP         APA...-----AR----A-----Q----------I--V-C--F----P-TN----M-----..........
TbCP         A-A...---------------V--Q--------TI-------QV--NP-------M-----..........
LmCPa        A-SGVMS----D--------N--L---------I-------AAS--S-------M-----..........
DdCP1        I-T...-F---TR-------N--Q-----S--TT--V---HFISQNK-------N--D--HECMEYEGEE
HsCPL        A-R...S-------Y-----N--Q---------T-AL---MFRKTGR-I-----N--D-........SGP

                                                       100
LmCPb        DMDNGCSGGLMLQAFDWLLQNTNGHLHTEDSYPYVSGNGYVPECSNSSELVVGAQIDGHVLIGSSEKAMA
TcCP         KT-S--G----NN--E-IV-EN--AVY-------A--E-IS-P-TT-GH.T---T-T---ELPQD-AQI-
TbCP         TI-S--N----DN--N-IVNSNG-NVF--A---------EQ-Q-QMNGH.EI--A-TD--DLPQD-D-I-
LmCPa        NI-E--N----D--MN-IM-SH--SVF--A----T--G-TR-P-HDEG..E---K-T-FLSLPHD-ERI-
DdCP1        AC-E--N---QPN-YNYIIK..--GIQ--S----TAET-TQ..-NFN-A.NI--K-SNFTM-PKN-TV--
HsCPL        QGNE--N----DY--QYVQD..--G-DS-E----EA.TEES..-KYNPK.YSV-NDT-F-D-PKQ---LM

                                150
LmCPb        AWLAKNGPIAIALDA..SSFMSYKSGVL..TACIGKQLNHGVLLVGYDMTG.....EVPYWVIKNSWGGD
TcCP         -CV-V---V-V-V--..--W-T-TG--M..-S-VSE--D--------NDSA.....A----I-----TAQ
TbCP         -Y--E---L---V--..E---D-NG-I-..-S-TS---D--------NDNS.....NP---I-----SNM
LmCPa        E-VE-R--V-V-V--..TTWQL-FG--V..SL-LAWS------I--FNKNA.....KP---IV-----SS
DdCP1        GYIVST--L---A--VE..WQF-IG--.FDIP-NPNS-D--I-I---SAKNTIFRKNM---IV-----A-
HsCPL        KAV-TV---SV-I--GHE--LF--E-IYFEPD-SSEDMD----V---GFESTESDNNK.--LV-----EE

                      200                    +1
LmCPb        WGEQGYVRVVMG.VNACLLSEYPVSAHVRESAAPGTSTSSETPAPRP...VVVEQVICFDKNCRRGCRKT
TcCP         ---D--I-IAK-.S-Q--VK-EAS--V-GGPGPTPEP-TTT-TSAPGPSPSYFV-MS-T-AA-IV--ENV
TbCP         ---D--I-IEK-.T-Q--MNQAVS--V-GGPTP-PPPPPPPSATFT........-DF-EG-G-TK--SHA
LmCPa        ---K--I-LA--.S-Q-M-KN-----T-ESPHTPHV-T-TA
DdCP1        ------IYLRR-.K-T-GV-NFVSTSII
HsCPL        --MG---KMAKDRR-H-GIASAASYPT-

                      50                                                 100
LmCPb        LIKANECHKNGGGGASMIKCSPQKVTMCTYS.NEFCVGGGLCFETHDGKCSPYFFGSIMNTCHYT
TcCP         TLPTGQ-LLTTS-VSAIVT-GAETL-EEVFFTSTH-S-PSVRSSVPLNQ-NWLLR--VEFF-GSS(+24)
TbCP         TFPTG--VQTT-V-SVIAT-GASNL-QII-PLSRS-S-PSVPITVPLD--I-ILI--VEYH-STN(+12)
```

Fig. 1. Alignment of LmCPb with selected cysteine proteinases. TcCP=*T. cruzi* (CYSP_TRYCR); TbCP=*T. brucei* (CYSP_TRYBR); LmCPa=*L. mexicana* (LCPA_LEIME); DdCP1=*Dictyostelium discoideum* CP1 (CYS1_DICDI); HsCPL=human cathepsin L (CATL_HUMAN). Annotations in brackets refer to accession names in Swiss-Prot 22. Numbers above the LmCPb sequence are described in the text. Dashes in other sequences indicate identities to LmCPb. Dots indicate gaps made to maximize alignments. Doubly underlined residues in the mature regions of some sequences represent the N-terminal amino acids of the mature CP as determined by amino acid sequencing.

pSWb1a, but, as for several other mRNAs of *Leishmania*, the *lmcpb* UTR appears relatively large (ca. 800 nt).

The largest open reading frame predicted a 443 amino acid, 47.8 kDa protein which was aligned to trypanosomatid and other eukaryotic CP sequences using GAP [9] and manually (Fig. 1). From this alignment, and from N-terminal amino acid sequencing of *L. mexicana* CPs (see below), residues −125 to −1 of LmCPb would be expected to form the pre- and pro-regions which are post-translationally removed to produce the mature

Table I

Percent identity between LmCPb and other CPs for the different domains

	Pre/Pro	Central	C-terminal
TcCP	44	59	22
TbCP	36	59	26
LmCPa	27	56	–
DdCP1	20	48	–
HsCPL	19	46	–

proteinase with a predicted size of 34 kDa. Residues 1–218 comprise the central domain of the proteinase while residues +1 to +100 make up a C-terminal extension similar to those present in CPs of *T. brucei* (TbCP) [6] and *T. cruzi* (TcCP) [12,13]. As summarized in Table I, LmCPb is more related to TcCP and TbCP than it is to LmCPa or the non-trypanosomatid CPs.

The end of the central domain (defined in [2,4,6]) and the start of the C-terminal extension was defined by comparing the LmCPb sequence with the two other trypanosomatid CP sequences which possess lengthy extensions. In TbCP, a run of 9 Pro residues is associated with the beginning of the extension and in TcCP, 6 consecutive Thr residues are present. In LmCPb, however, the demarcation between the central domain and the extension is less clear: at the analogous position is a stretch rich in Ala, Gly, Pro, Ser, and Thr. The only other *L. mexicana* CP cDNA sequenced so far (*lmcpa*) lacks a comparable C-terminal extension but does extend a short way beyond the *Dictyostelium* and human sequences with the extra amino acids again being predominantly Pro and Thr [4]. Of the identities common to LmCPb, TcCP, and TbCP in the C-terminal extension, 8 are Cys residues. An examination of the extensions of alpha- and beta-oryzains [14] indicate that they are also rich in cysteines but these residues do not appear align with those of LmCPb, TbCP, or TcCP. Potential roles for the extension include targeting, stability, or interaction with protein substrates.

The CP content of *L. mexicana* is complex [2,3,15]. Amastigotes contain 3 main classes (A, B and C) that are distinguished on the basis of differences in apparent size, substrate specificity, and glycosylation [3]. Interestingly, direct amino acid sequence analysis of purified CPs of these classes yielded the same N-terminal sequence for each (Robertson and Coombs, unpublished results). This sequence is identical to that present in the *lmcpb* open reading frame (Fig. 1). The observed differences among these *L. mexicana* CPs could be due to variable post-translational processing of a single CP gene product. For example, a potential *N*-linked glycosylation site is present at position 103 in the central domain where the consensus VPECSNSSELVVGA occurs: glycosylation at this site could account for the difference between types B and C CPs which do not bind to concanavalin-A Sepharose columns and type A CPs which do [3]. Alternatively, a number of closely related but distinct genes could exist.

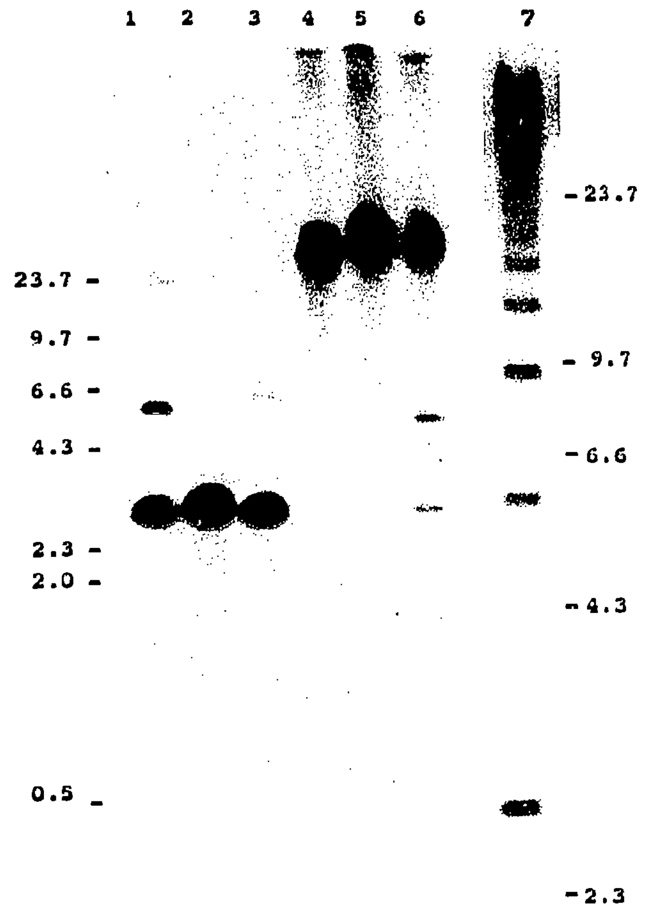

Fig. 2. Genomic organization of *lmcpb*: total genomic DNA (10 μg) was digested to completion with restriction enzymes, electrophoresed through 1% agarose, blotted, probed with the pSWb1a cDNA probe. Lane 1–6: *Eco*RI, *Sal*I, *Sac*I, *Hin*dIII, *Xba*I, *Bam*HI. Marker sizes are shown on the left. Lane 7 shows a partial digest of total DNA digested with *Eco*RI and probed as above.

To characterize the genomic arrangement of *lmcpb*, a Southern blot of total *L. mexicana* genomic DNA cut to completion with various restriction enzymes was probed with the *lmcpb* cDNA (Fig. 2; lanes 1–6). Three enzymes (*Eco*RI, *Sal*I and *Sac*I) gave a major band at 2.8 kb whereas 3 other enzymes (*Hin*dIII, *Xba*I and *Bam*HI) gave a major band of >23 kb. A blot of DNA partially digested with *Eco*RI gave a hybridization ladder beginning at 2.8 kb (Fig. 2; lane 7) implying that *lmcpb* is a multicopy gene of 2.8 kb unit size with at least 10 tandemly repeated copies per locus. The minor bands seen in some of the digests may represent flanking sequences, polymorphisms or possibly copies located at other loci. This organization in which the majority of genes is present as a tandem array is similar to that reported for the *T. brucei* CP [6] and *T. cruzi* CP [12,13]

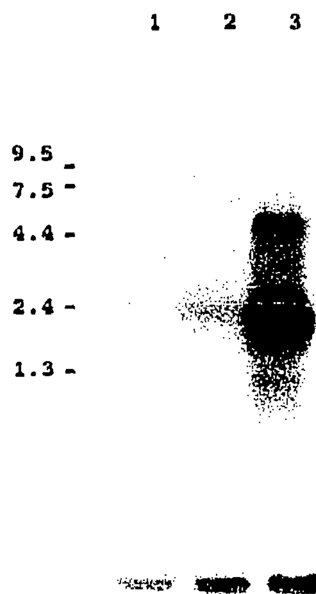

Fig. 3. Developmental differences in *lmcpb* RNA levels: Ten μg of total RNA from multiplicative promastigotes (lane 1), stationary phase promastigotes (lane 2) and amastigotes (lane 3) were probed with the *lmcpb* probe (upper panel) or with an oligonucleotide complementary to the small subunit rRNA (lower panel). BRL RNA ladder sizes are shown on the left for the upper panel.

but different from that of the other *L. mexicana* CP gene, *lmcpa*, which exists as a single copy gene [4].

The *lmcpb* cDNA was also used to probe Northern blots of total RNA from the three major developmental stages of the parasite (Fig. 3). Strong hybridization was seen for the *lmcpb* RNA from amastigotes at 2.3 and weaker signals were seen at ~5 and ~8 kb. A weak signal at 2.3 kb was detected in metacyclic promastigote RNA. *lmcpb* RNA was undetectable in multiplicative promastigotes. The 2.3 kb transcript corresponds well with the size of the cDNA and the larger signals could be the result of inefficient post-transcriptional events such as 5′ cleavage for trans-splicing or 3′ end formation for polyadenylation [13]. The steady state levels of *lmcpb* RNA correlate very well with the activities of the types A, B and C CPs that occur in different forms of *L. mexicana*: very high in amastigotes, low in metacyclic promastigotes and undetectable in multiplicative promastigotes [5].

Acknowledgements: This work was supported by the MRC and The Wellcome Trust. Colin Robertson is thanked for sharing unpublished results and critical reading of the manuscript. David Laughland and John Laurie are thanked for technical assistance.

REFERENCES

[1] Chang, K.P. and Chaudhuri, G. (1990) Annu. Rev. Microbiol. 44, 499–529.
[2] North, M.J., Mottram, J.C. and Coombs, G.H. (1990) Parasitol. Today 6, 270–275.
[3] Robertson, C.D. and Coombs, G.H. (1990) Mol. Biochem. Parasitol. 42, 269–276.
[4] Mottram, J.C., Robertson, C.D., Coombs, G.H. and Barry, J.D. (1992) Mol. Microbiol. 6, 1925–1932.
[5] Robertson, C.D. and Coombs, G.H. (1992) FEMS Microbiol. Lett. 94, 127–132.
[6] Mottram, J.C., North, M.J., Barry, J.D. and Coombs, G.H. (1989) FEBS Lett. 258, 211–215.
[7] Eakin, A.E., Bouvier, J., Sakanari, J.A., Craik, C.S. and McKerrow, J.H. (1990) Mol. Biochem. Parasitol. 39, 1–8.
[8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Devereux, H., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[10] Briones, M.R.S., Nelson, K., Beverley, S.M., Affonso, H.T., Camargo, E.P. and Floeter-Winter, L.M. (1992) Mol. Biochem. Parasitol. 53, 121–127.
[11] Borst, P. (1986) Annu. Rev. Biochem. 55, 701–732.
[12] Campetella, O., Henriksson, J., Aslund, L., Frasch, A.C.C., Pettersson, U. and Cazzulo, J.J. (1992) Mol. Biochem. Parasitol. 50, 225–234.
[13] Eakin, A.E., Mills, A.A., Harth, G., McKerrow, J.H. and Craik, C.S. (1992) J. Biol. Chem. 267, 7411–7420.
[14] Watanabe, H., Abe, K., Emori, Y., Hosoyama, H. and Arai, S. (1991) J. Biol. Chem. 266, 16897–16902.
[15] Coombs, G.H., Robertson, C.D. and Mottram, J.C. (1991) in: Biochemical Protozoology (Coombs, G. and North, M. eds.) pp. 208–220, Francis & Taylor Ltd., London.