tvcp12: a novel *Trichomonas vaginalis* cathepsin L-like cysteine proteinase-encoding gene

Claudia R. León-Sicairos, Josefina León-Félix and Rossana Arroyo

Departmento de Patología Experimental, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV-IPN), AP 14-740, Mexico City, 07000, Mexico

*Trichomonas vaginalis* is the causative agent of trichomoniasis, one of the most common sexually transmitted diseases in humans. This protozoan has multiple proteinases that are mainly of the cysteine proteinase (CP) type, some of which are known to be involved in the parasite’s virulence. Here, a novel *T. vaginalis* CP-encoding gene, *tvcp12*, was identified and characterized. *tvcp12* is 948 bp long and encodes a predicted 34.4 kDa protein that has the characteristics of the papain-like CP family. TvCP12 does not appear to have a signal peptide, suggesting that this is a cytoplasmic CP. By Southern blot assays, the *tvcp12* gene was found as a single copy in the *T. vaginalis* genome. Remarkably, Northern blot experiments showed a single transcript band of ~1.3 kb in the mRNA obtained from parasites grown in low iron conditions and no transcript was observed in the mRNA from parasites grown in high iron conditions. By RT-PCR assays, a 270 bp band was amplified from the cDNA of parasites grown in low iron medium, which was very faint when cDNA from parasites grown in high iron conditions was used. Transcripts of the 3’ region obtained in both iron conditions presented differences in their poly(A) tail length. These data suggest that *tvcp12* is another gene that is negatively regulated by iron and that the length of the poly(A) tail may be one of the factors involved in the iron-modulated protein expression.

The many roles that proteinases play in these organisms are starting to be understood. Central roles have been proposed for proteases in diverse processes such as host cell invasion and exit, encystation, exocytosis, catabolism of host proteins, differentiation, cell cycle progression, cytoadherence and both stimulation and evasion of the host immune responses (Klemba & Goldberg, 2002; Sajid & McKerrow, 2002).

Many CPs play important roles in the virulence of parasites (García-Rivera et al., 1999; Mottram et al., 1996). In *T. vaginalis*, some CPs are involved in cytoadherence (Arroyo & Alderete, 1989, 1995; Mendoza-López et al., 2000), haemolysis (Dailey et al., 1990), cytotoxicity (Álvarez-Sánchez et al., 2000), disruption of the host cell membrane skeleton (Fiori et al., 1997), degradation of the secretory leukocyte protease inhibitor (Draper et al., 1998) and trichomonal immune evasion mechanisms by degradation of human immunoglobulins (Provenzano & Alderete, 1995). Interestingly, the expression, proteolytic activity and surface localization of certain trichomonad CPs are modulated by the presence of iron in the medium (Alderete et al., 1995; C. R. León-Sicairos, J. León-Félix & R. Arroyo, unpublished results).

Iron is an essential nutrient for growth, metabolism and expression of virulence factors for many pathogens (Gorrell, 1985; Griffith, 1985; Wilson & Britigan, 1998).
In trichomonads, iron is acquired from the host via highly specific receptor-mediated mechanisms (Peterson & Alderete, 1984; Lehker et al., 1990; Lehker & Alderete, 1992). In addition, iron controls the expression of hydrogenosomal proteins such as pyruvate:ferredoxin oxidoreductase and ferredoxin (Gorrell, 1985), and virulence genes, such as those encoding some of the T. vaginalis adhesin proteins (Lehker et al., 1991; Arroyo et al., 1995; Engbring et al., 1996; Alderete et al., 1998; Garcia et al., 2003). Tsai et al. (2002) identified an iron-responsive promoter in the ap65-1 gene which could be involved in the transcriptional regulation of some of these genes. Iron also mediates T. vaginalis resistance to complement lysis due to degradation of the C3 on the trichomonal surface (Alderete et al., 1995) and negatively regulates the expression of a fibronectin-like gene, flp-1, reported recently (Crouch & Alderete, 2001).

Here we report the identification and analysis of a novel tvcp12 complete gene that encodes a new papain-like CP of T. vaginalis and is negatively regulated by iron.

**METHODS**

**Parasites and culture conditions.** T. vaginalis CNCD 147 (CNCD, Centro Nacional de Clinicas de Displasias) clinical isolate was cultured throughout this work (Alvarez-Sanchez et al., 2000). Parasites were grown in trypticase/yeast extract/maltose medium supplemented with 10% (v/v) heat-inactivated horse serum as described by Diamond (1957). Exponential-phase parasites were grown in high iron (250 μM ferrous sulfate) and in iron-depleted medium with 100 mM 2,2′-dipyridyl (Sigma) (Lehker et al., 1991; Crouch & Alderete, 2001).

**Cloning of tvcp12 and sequencing.** Previously, we obtained a 3035 bp DNA fragment (T. vaginalis locus 1.2.1.1) by hybridization of a genomic library (Espinosa et al., 2001) with a conserved region of papain-like CPs as a probe. This genomic clone contains a gene fragment and a complete gene separated by 1386 bp. One of the sequences (765 bp) corresponded to a fragment of a CP-encoding gene that we have called tvcp12 (Leon-SicaIros et al., 2003). We amplified by PCR this 765 bp fragment using the sense primer CP5-1.2.1.1 (5′-CGACTCTCAATGAAACATC-3′; nucleotide positions 184–202) and the antisense primer CP3-1.2.1.1 (5′-TTATTAACGGTGCGGGACG-3′; nucleotide positions 933–948) (Fig. 1b), corresponding to the ends of the tvcp12 fragment present in the genomic clone. The PCR product was cloned into the vector pCRII-TOPO (Invitrogen). The 5′ and 3′ ends of the tvcp12 gene were obtained by using the rapid amplification of cDNA ends (RACE) system (Invitrogen), according to the manufacturer’s instructions. Briefly, for the synthesis of cDNA, 1 μg total RNA was first dephosphorylated, the cap structure was removed from the mRNA, the oligo-RNA was ligated into the cap-less RNA, and reverse transcription of the trichomonad mRNA was performed. To obtain the missing 5′ end of the tvcp12 gene, we used 1 μl cDNA, the GeneRacer 5′ nested primer (5′-GGACACTGACATGGCTAGGAGATT-3′) from the kit and the antisense primer RACE 5′-1.2.1.1 (5′-GGATGAAGGTGGCGTGGTCTTT-3′), designed based on the known gene sequence at nucleotide positions 415–438 (Fig. 1b). To amplify the 3′ end of the tvcp12 gene, a PCR was performed with 2 μl cDNA using as a sense primer RACE 3′-1.2.1.1 (5′-GACGGATTTGCTGCTACAG-3′; nucleotide positions 625–647) (Fig. 1b) and as an antisense primer GeneRacer 3′ (5′-GGCAGATGACGATCGG-3′). The 3′ TVC products amplified by PCR were cloned into the vector pCRIV-TOPO (Invitrogen) and both strands were sequenced independently by using the dyeoxy chain-termination method (Sanger et al., 1977) in an ABI377 Applied Biosystems Automatic Sequencer (UNAM, Department of Cell Physiology, Mexico City, Mexico). The DNA sequences were analysed by using the BLAST (http://www.ncbi.nlm.nih.gov/blast/), EXPASY (http://us.expasy.org) and WORKBENCH (http://workbench.sdsc.edu) search engine tools.

**Southern and Northern blot assays.** High-molecular-mass DNA from T. vaginalis was obtained as described by Espinosa et al. (2001). For Southern blot hybridization analysis, genomic DNA (10 μg) was digested with restriction endonucleases that cut out of the tvcp12 gene (EcoRI, NotI and XbaI). Digested DNA was electrophoresed in agarose gels and transferred to a Zetaprobe nylon membrane (Bio-Rad) according to the manufacturer’s instructions. The membranes were pre-hybridized with 50% (v/v) formamide, 5× SSC, 50 mM Tris/HCl pH 7.4, 1% (w/v) SDS, 5× Denhardt’s solution for 2 h at 42°C and hybridized in the same solution at 42°C for 16 h with the 765 bp tvcp12 gene probe labelled with [α-32P]ATP [specific activity 3000 Ci mmol⁻¹ (111 TBq mmol⁻¹); Life Technologies] by random priming. The filters were washed three times for 20 min in 0.5× SSC, 1% (w/v) SDS at 65°C.

Total RNA was prepared using TRIzol reagent. The poly(A)+ RNA was obtained by oligo(dT) cellulosic chromatography. For Northern blot analysis, the mRNA samples were separated on 1% (w/v) formaldehyde agarose gels and transferred to Zetaprobe nylon membranes using 20× SSC. The filters were pre-hybridized, hybridized and washed as described for Southern blot assays. Each lane on the gel was loaded with 3–4 μg of mRNA as quantified by absorbance at 260 nm (Ausubel et al., 1999).

**RT-PCR assays.** These were done using the Superscript RNase H⁻ Reverse Transcriptase kit (Stratagene) according to the manufacturer’s instructions. Total RNA was reverse-transcribed using AMV reverse transcriptase and the oligo(dT) primer. Then, a fragment of 270 bp of the tvcp12 gene was amplified by PCR using the sense primer vPROCP-1.2.1.1 (5′-GATTTCACTTGCTCCGGCCAT-3′; nucleotide positions 8–30) and the antisense primer vRACE 5′-12 (5′-CTTGAACGTTGCGCGCATAACAT-3′) and vRACE 5′-12 (5′-CAACGTTAGGGAACAGCTAT-3′), as an internal control.

**2-DS polycrylamide gels co-polymerized with gelatin.** Parasite proteinases were analysed by 2-DS PAGE under reducing conditions. The polycrylamide gels were co-polymerized with 0–2% gelatin as a substrate (substrate gels) as described previously. After electrophoresis, proteins were renatured with 2.5% Triton X-100 and activated with 100 mM sodium acetate buffer pH 4–5 for 18 h at 37°C. Gels were Coomassie blue-stained and de-stained for analysis (Alvarez-Sanchez et al., 2000).

**RESULTS**

tvcp12 is a novel T. vaginalis gene encoding a CP of the papain family C1 in clan CA

Here we report the identification and analysis of a novel 1020 bp CP-encoding gene, tvcp12, that was initially detected as a 765 bp fragment in T. vaginalis locus 1.2.1.1 (Leon-SicaIros et al., 2003). By RACE assays, we obtained the full sequence of the gene (Fig. 1a). The tvcp12 948 bp open reading frame encoded a predicted 34×4 kDa poly-peptide (Fig. 1b). This putative protein shares 70–73%
homology with the products of tvcp1, tvcp2, tvcp3, tvcp4 and tvcp25, *T. vaginalis* CP-encoding genes (León-Sicairos et al., 2003; Mallinson et al., 1994). At nucleotide 21 of the 5' untranslated region (UTR), we found an AU-rich sequence that did not have the typical Inr promoter element described for other trichomonad genes (Liston & Johnson, 1999). Noteworthy is the observation that the putative cleavage site AAAUU is located 18 nt downstream of the putative UAAA polyadenylation signal reported by Espinosa et al. (2002). A U-rich region downstream of the putative cleavage site was also observed (Fig. 1b).

All *T. vaginalis* CPs characterized to date have a similar structure, with the most important motifs typical of cathepsin L-like CPs located in the same position (Fig. 2a). The signal peptide was not detected in the amino acid sequence of TvCP12. However, this sequence is present in TvCP2 and TvCP3 but not in TvCP1 and TvCP25, while the amino terminus of TvCP4 is unknown (Fig. 2a). The ERFNIN-like [Ex3Rx2(Ile/Val)Fx2Nx3Ix3N] motif, which is present in all CPs analysed so far, is located at amino acid residues 37–56 in TvCP12, whereas the potential or known start site of the mature TvCP12 is located at residue 99. All CP sequences, except that of TvCP3, have the three residues that form the catalytic triad (Cys-25, His-159 and Asn-175) located at amino acids 124, 263 and 282. Also, all CPs except TvCP3 have the six conserved cysteine residues (Cys-22/Cys-63, Cys-56/Cys-95 and Cys-153/Cys-200, papain numbering system) that form three disulfide bonds to stabilize the tertiary structure (Fig. 2a).

A phylogenetic tree constructed using the deduced amino acid sequences corresponding to the catalytic domains between the catalytic Cys and His residues of the six reported trichomonad CP-encoding genes (Mallinson et al., 1994; León-Sicairos et al., 2003) was obtained by using the CLUSTAL W (workbench 3.2) tool. The dendrogram (PHYLIP format) generated showed that the six *T. vaginalis* CPs
shared a high degree of relatedness, suggesting that these CPs evolved from a common ancestor probably at a very early stage of the parasite’s evolution. It also showed that the six CPs are grouped into three clusters. The first group corresponds to TvCP1, TvCP4 and TvCP25, with TvCP4 and TvCP25 being more closely related to each other than to TvCP1. The second group is formed by TvCP2 and TvCP3, which showed high identity (60%). Lastly, TvCP12 belongs to an independent group that has diverged very early in evolution and shows less identity with the other five T. vaginalis CPs (Fig. 2b). The high identity among the six CPs suggests that their respective genes arose as a result of gene duplication as proposed by Mallinson et al. (1994).

tvcp12 is a single copy gene transcribed as a 1·3 kb mRNA

To investigate the copy number of the tvcp12 gene in the T. vaginalis genome, we performed Southern blot assays using EcoRI-, NotI- and XbaI-digested genomic DNA and the PCR-amplified 765 bp fragment as a probe. In all cases, a single DNA band was observed (Fig. 3a) suggesting that a single copy of the tvcp12 gene is present in the T. vaginalis genome, as is the case for the tvcp1, tvcp2 and tvcp3 genes (Mallinson et al., 1994). By Northern blot experiments using the poly(A) + RNA and the 765 bp probe, we detected a single transcript band of 1·3 kb (Fig. 3b). The length of this mRNA is consistent with the length of the complete tvcp12 gene sequence, of 948 bp.

Expression of tvcp12 is down-regulated by iron

Iron modulates the expression of distinct T. vaginalis genes (Alderete et al., 1995, 1998; Arroyo et al., 1995; Engbring et al., 1996; Crouch & Alderete, 2001; Lehker & Alderete, 1992). To determine the effect of iron on tvcp12 transcription, trichomonads were grown under iron-replete and iron-depleted conditions (Arroyo et al., 1995; Crouch & Alderete, 2001). Northern blot analysis of mRNA from parasites grown under different iron conditions was performed using the 32P-labelled 765 bp probe. After hybridization, a single band of 1·3 kb appeared in low iron conditions. The transcript did not appear in this experiment when the parasites were cultured in an iron-replete medium (Fig. 4a). To confirm our results, we performed RT-PCR assays using primers specific for the tvcp12 gene. The expected 270 bp product was obtained from parasites grown in a low iron medium, while a faint band appeared in the RNA of parasites grown in high iron conditions.
conditions, indicating that tvcp12 is also transcribed in the presence of high iron concentrations but probably in a lower amount. Nevertheless, we can not discard the possibility that other post-transcriptional mechanisms could be participating in the transcript stability. As a control, we amplified the \( \beta \)-tubulin gene from the same samples as used above. An identical band of 112 bp from the \( \beta \)-tubulin gene was amplified using the RNA from both iron growth conditions (Fig. 4c). Densitometric analysis of the RT-PCR products confirmed the differences in tvcp12 transcription when \( T. \) vaginalis was grown in the presence of low and high concentrations of iron (Fig. 4d).

Next, we obtained several cDNA clones by 3' RACE and PCR amplification to analyse the 3' end of the tvcp12 transcripts generated from parasites grown in the presence of low and high concentrations of iron. PCR assays gave products with different sizes that were cloned and sequenced. Analysis of the cDNA sequences showed the presence of a 19 nt 3' UTR in both iron conditions (Fig. 4e, f). Five nucleotides downstream of the stop codon, this region contains an AU-rich element (AURE), represented by the RNA consensus sequence AUUUA (Fig. 4f), which is known to regulate eukaryotic mRNA stability (Ross, 1995, 1996; Shaw & Kamen, 1986; Decker & Parker, 1995). Interestingly, clones obtained from parasites grown in the presence of a low iron concentration presented 24 As in their cDNA, whereas those from trichomonads grown in the presence of a high iron concentration presented only 17 As (Fig. 4f).

**Proteolytic activity of the 30 kDa region is modulated by iron**

Lastly, we investigated whether the proteolytic activity of mature CPs with a predicted molecular mass of \( \sim 30 \) kDa, which may be the molecular mass of the mature TcCP12 proteinase, was differentially modulated in parasites grown in low and high concentrations of iron. We ran 2-D substrate gels (zymograms) using extracts from \( T. \) vaginalis cultured in both iron conditions (Fig. 5). Differences in the proteolytic activity patterns of the trichomonad's 25–35 kDa region were observed among the zymograms of parasite extracts from the different iron concentrations. A total of seven spots with proteolytic activity was observed in this region in the pI range from 4.5 to 6.0.
Spots 1, 3, 4, 5 and 6 were observed in the extracts from parasites grown in low and high iron concentrations. Interestingly, spots 2 and 7 were only observed in extracts from parasites grown in low iron medium (Fig. 5, L), whereas spot 5 activity was more visible in parasites grown in high iron medium (Fig. 5, H). These results suggest that one of these spots observed only in low concentrations of iron may correspond to TvCP12, but more experiments are necessary to confirm this.

DISCUSSION

T. vaginalis is a rich source of CPs in nature (Barrett et al., 1998). CPs are important virulence factors in T. vaginalis that participate in adherence to host cells, cellular destruction and IgG digestion (Arroyo & Alderete, 1989, 1995; Mendoza-López et al., 2000; Álvez-Sánchez et al., 2000; Provenzano & Alderete, 1995). In this report, we have described the nucleic acid and deduced amino acid sequences for a novel T. vaginalis gene, tcpc12. tcpc12 is the sixth CP-encoding gene cloned from T. vaginalis. It shares 70–73 % homology with the CPs of the papain family and contains the main motifs of the members of this family, including the conserved catalytic triad residues (Cys, His, Asn), the ERFINN-like motif and the Cys residues necessary for the protein’s tertiary structure. The putative TvCP12 protein has no signal peptide, suggesting that this is a cytoplasmic protein or that it is delivered to the plasma membrane by another mechanism, such as the membrane glycosylphosphatidylinositol (GPI) anchor mediated by the Leishmania mexicana GPI protein transamidases (Hilley et al., 2000). However, the hypothesis that TvCP12 is a cytoplasmic protein is supported by the lack of predicted membrane sequences on TvCP12. The cell location of TvCP12 will be investigated with antibodies against recombinant TvCP12.

The phylogenetic data suggest that TvCP12 appears to be part of a new cluster of CPs that shows high sequence divergence from the two other groups. Explanation for this early divergence will be found once the biochemical and functional characterization of this CP are completed. In addition, these should be considered preliminary phylogenetic data for the T. vaginalis CPs, since only a few of the multiple CP-encoding genes of this parasite have been cloned. The ongoing T. vaginalis genome project could help to confirm or modify this phylogenetic analysis. The pattern of the phylogenetic tree is similar to that of the 20 Entamoeba histolytica CP-encoding genes, which are grouped in two distinct clades suggesting that these amoeba CP-encoding genes evolved from two separate ancestors at a very early stage of parasite evolution (Bruchhaus et al., 2003).

As in other genes, the 5’ UTR of tcpc12 is only 21 nt long, but tcpc12 does not have a typical Inr consensus sequence located up to 20 nt upstream of the initiation codon for T. vaginalis genes (Liston & Johnson, 1999). Since no typical TATA box has been described, we do not yet know how this gene is transcribed. However, we know that tcpc12 is expressed as a 1.3 kb mRNA in low iron conditions. In addition, RT-PCR assays evidenced that it is expressed when there is a low concentration of iron in the medium, whereas in high iron concentrations the transcript amount is very low.

The tcpc12 gene is iron-regulated; this is not surprising, because it has been established that different iron concentrations in the medium induce expression of different genes (Lehker & Alderete, 1992). The flp1 and flp2 genes of T. vaginalis, which encode proteins with homology to fibronectin, are induced under low iron growth conditions (Crouch & Alderete, 2001), whereas ap65 and the ap51 and ap33 genes are overexpressed in high iron concentration media (Arroyo et al., 1995; Enghbring et al., 1996; Alderete et al., 1998).

The stability of the mRNA could be influenced by the presence of destabilizing elements such as the AUREs and also due to the length of the poly(A) tail. In mammals, if an AURE from the 3’ UTR of an unstable mRNA, for example, one encoding granulocyte macrophage colony stimulating factor (GM-CSF), is placed within the 3’ UTR of the β-globin stable mRNA, the chimeric transcript decays with a half-life of less than 30 min (Caput et al., 1986). β-Globin mRNA lacking the AURE is stable for well over 2 h in the same transfected cells. In T. vaginalis, mRNAs have been detected that contain AUREs such as two of the transcripts for the adhesin AP65 genes (Engbring et al., 1996) and the flp-1 and the flp-2 genes with
homology to fibronectin (Crouch & Alderete, 2001) that could be important for destabilization of these mRNAs. However, studies need to be done to demonstrate if the AUREs found in the tvcp12 mRNAs from two different iron conditions are functioning as destabilizing elements (work in progress).

The stability of the mRNA due to variations in the length of the poly(A) tail could partially explain the differences observed in the transcript of the tvcp12 gene in the two iron conditions. The poly(A) tail length is important in mRNA stability in many genes studied such as the EhPgp1 gene involved in the MDR phenotype of E. histolytica (López-Camarillo et al., 2003). This could also be a regulatory mechanism involved in the low amount of transcript detected in T. vaginalis grown in high iron. These data suggest that iron concentrations during growth are affecting the tvcp12 transcript stability possibly through the size of the poly(A) tail.

The UAAA sequence in the 3′ UTR of T. vaginalis mRNAs was proposed as the polyadenylation signal necessary but not sufficient to direct the polyadenylation of transcripts. This signal needs to be positioned at the proper distance between 11 and 30 nt from the cleavage site (AAAAUUU) followed by a U-rich motif (Espinosa et al., 2002). We found this putative signal and the cleavage sequence in the genomic sequence of the tvcp12 gene (León-Sicaíros et al., 2003) and our data from the cDNA sequence in Fig. 4(e) strongly suggest that this could be working as the putative polyadenylation signal of the tvcp12 mRNA.

Lastly, assuming that the pro-region is cleaved in all six cathepsin L-like T. vaginalis CPs (Mallinson et al., 1994; León-Sicaíros et al., 2003) as in other cathepsin L CPs, these proteinases would have a molecular size of between 23 and 30 kDa. However, within this size range only three spots with proteolytic activity have been observed previously in trichomonad extracts (Neale & Alderete, 1990; Mendoza-López et al., 2000). These data suggest that not all of the CP-encoding genes identified from this region are expressed in the parasite nor are they activated under the same environmental conditions, and processed in the same way. In addition, other proteinases with narrow specificity, able to degrade certain substrates, could exist in T. vaginalis.

After analysing the proteolytic activity under low and high iron concentrations, we were able to detect up to seven spots with different isoelectric points in the 25–35 kDa region; whether these spots belong to the activity of all cathepsin L-like CPs awaits to be determined. However, the size and the down-regulation by iron of tvcp12 suggest that the TvCP12 protein may be one of the papain-like proteolytic spots that are only present in low iron conditions, such as spots 2 and 7.

It is noteworthy that proteinases with the ability to degrade specific substrates such as spectrin (Fiori et al., 1997), human immunoglobulins (Provenzano & Alderete, 1995) and the secretory leukocyte protease inhibitor (Draper et al., 1998) have been found in T. vaginalis. Recently, we have also found two legumain-like CPs in the 25–35 kDa region with pI 6-3 and 6-5 that do not belong to the papain-like CP family in clan CA (Sajid & McKerrow, 2002). These legumains appear to be part of a new subfamily of CPs within the family C13 of clan CD (León-Felix et al., 2004) that has high substrate specificity (Sajid & McKerrow, 2002).

ACKNOWLEDGEMENTS

This work was partially supported by grants 25572-N and 33044-M from the Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico (to R.A.). C.R.L.-S. and J.L.-F. were scholarship recipients also from CONACyT. We thank Drs Esther Orozco and Jaime Ortega for their critical reviews of the manuscript. We also thank Alfredo Padilla for his help with the artwork and Leticia Avila and Alberto García for their technical assistance.

REFERENCES


