The TATA-box binding protein of *Entamoeba histolytica*: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy

Juan P. Luna-Arias,1,3 Rosaura Hernandez-Rivas,1 Guadalupe de Dios-Bravo,3 Job Garcia1, Leobardo Mendoza3 and Esther Orozco2,3

Author for correspondence: Esther Orozco. Tel: +52 5 747 7000 ext. 5642. Fax: +52 5 747 7108. e-mail: esther@mail.cinvestav.mx

A 309 bp DNA fragment from *Entamoeba histolytica* was amplified by PCR using primers derived from the Acanthamoeba castellanii consensus TATA-box binding protein amino acid sequence. The amplified fragment was used to isolate cDNA and genomic DNA clones containing an ORF encoding the complete *E. histolytica* TATA-box binding protein (Ehtbp, 702 bp, 234 aa, molecular mass 26 kDa). The EhTBP functional domain showed 55% sequence identity to that of *Homo sapiens*, 54% to *A. castellanii* and 37% to *Plasmodium falciparum* TBPs. In Southern blot experiments we detected a single Ehtbp band, which was transcribed as a 1.3 kb mRNA containing a 420 nt 5′ untranslated region. However, the probe hybridized with the 0.8 and 1.5 Mb chromosomes, suggesting that this sequence is diploid. In situ PCR assays showed two signals in 95% of trophozoites, one located in the nucleus and another in EhKO, the novel DNA-containing organelle recently reported. The recombinant *E. histolytica* TATA-box binding protein was expressed in *Escherichia coli*. Antibodies against it recognized two proteins of 26 and 29 kDa in *E. histolytica* nuclear extracts. Confocal microscopy immunofluorescence analysis located the protein in both the nucleus and EhKO.

Keywords: *Entamoeba histolytica*, TATA-box binding protein, *tbp* gene expression

INTRODUCTION

*Entamoeba histolytica* is a protozoon responsible for human amoebiasis. Knowing its DNA structure and its mechanisms for regulating gene transcription is a step toward understanding virulence. *E. histolytica* has an A+T rich genome (67 mol%) (Sánchez et al., 1994; Tannich & Horstmann, 1992) and its codon usage is unusual, with 85% A+U in the third codon position (Bruchhaus et al., 1993). Putative TATA boxes with TATTTAAA, TATAAA (Sánchez et al., 1994) and TATTTAAA sequences (Bruchhaus et al., 1993; Li et al., 1992) have been located 27 to 50 bp upstream from the ATG start codon. The ATTCA and ATCA sequences have been described as consensus transcription start sites (Bruchhaus et al., 1993). Remarkably, protein-encoding gene transcription is insensitive *in vitro* to 1 mg α-amanitin ml⁻¹, suggesting that the RNA polymerase differs from those of other eukaryotic organisms (Lioutas & Tannich, 1993), but little is known about the transcription machinery.

Studies of promoter structure and function began recently with the characterization of the *hgl2* and *hgl5* gene promoters, which encode the slightly different 170 kDa heavy subunits of the galactose-inhibitable lectin (Buß et al., 1995; Purdy et al., 1996; Singh et al., 1997). The TATA box, the CCAAT box (*hgl2*) and the GAAC element (*hgl5*) were identified as regulators for full promoter activity *in vivo*. We have studied factors involved in expression of the EhPgp1 and EhPgp5 genes, both responsible for the multidrug resistance phenotype.

Abbreviations: TAFE, transverse alternating field electrophoresis; TBP, TATA-box binding protein; UTR, untranslated region.

The EMBL accession number for the sequence reported in this paper is 248307.
in *E. histolytica* (Orozco et al., 1995). The *EhPgp5* gene is induced by the presence of emetine in the medium and has a typical TATA box element that was not found in the promoter of the constitutively expressed *Ehgp1* gene (Gómez et al., 1998; Pérez et al., 1998).

The TATA-box binding protein (TBP) is associated with several transcription factors to make up the multiprotein complexes SL1 (Comai et al., 1992), TFIID (Dynlacht et al., 1991) and TFIIH (Kassavetis et al., 1992; Taggart et al., 1992), which are used by a specific RNA polymerase (White & Jackson, 1992). TBP is the first molecule in the assembly of TFIID, indicating its central role in transcription (Chen et al., 1994). Its functional domain, a 180 aa sequence located at the C-terminus, is highly conserved in most *tbp* genes (Hori koshi et al., 1989; Nikolov et al., 1992; Struhl, 1994). However, the *Plasmodium falciparum* *tbp* sequence diverges significantly from other *tbp* genes. Due to the unusual A+T richness of the *P. falciparum* and *E. histolytica* genomes (Pollack et al., 1982; Bruchhaus et al., 1993; Mc Andrew et al., 1993), the transcription factors in these parasites may differ from those reported in humans, making them a good target for drug design. In this paper we report the cloning, molecular characterization and expression of the *E. histolytica* *tbp* gene (*Ehtbp*). The localization of the protein by immunofluorescence and confocal microscopy was also performed.

**METHODS**

**E. histolytica cultures.** Trophozoites of *E. histolytica* clone A, strain HM1:IMSS were cultured in TYI-S-33 medium and harvested during the exponential growth phase (Diamond et al., 1978).

**Cloning and sequencing of the Ehtbp gene.** Two oligonucleotide primers (sense: 5' GTCTTACATGCTAGAGATTGCT 3' and antisense: 5' TGGAATAATTCTGTATTTCAM 3') corresponding to the nucleotide sequence in conserved regions in the C-terminal domain of *A. castellanii* TBP (Actbp, bp 354–374 and 651–671) (Wong et al., 1992a) were synthesized (model 381A, Applied Biosystems) using *E. histolytica* codon preferences (Tannich & Hortsmann, 1992).

![Image](image)

**Table 1.** Percentage amino acid identities in the C-terminal domain of EhTBP and other TBPs

<table>
<thead>
<tr>
<th></th>
<th>C terminus*</th>
<th>DR1†</th>
<th>DR2†</th>
<th>B domain‡</th>
<th>σ region§</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. castellanii</em></td>
<td>54</td>
<td>65</td>
<td>62</td>
<td>47</td>
<td>59</td>
</tr>
<tr>
<td><em>P. falciparum</em></td>
<td>37</td>
<td>43</td>
<td>41</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
<td>55</td>
<td>63</td>
<td>64</td>
<td>44</td>
<td>59</td>
</tr>
</tbody>
</table>

* C terminus is the functional domain for each TBP (residues 48–234 in EhTBP).
† DR1 and DR2 correspond to direct repeats 1 (residues 57–116 in EhTBP) and 2 (residues 147–207 in EhTBP), respectively.
‡ B domain is the basic domain (residues 109–145 in EhTBP).
§ σ region (residues 170–201 in EhTBP).
For PCR, 100 ng E. histolytica genomic DNA was used as the template. Denaturation was carried out at 94 °C for 1 min, annealing at 45 °C for 1 min and extension at 72 °C for 2 min, for a total of 35 cycles. The 309 bp PCR-amplified fragment was sequenced with Sequenase version 2.0 (US Biochemicals). After labelling with [α-32P]dATP and a Random Primed DNA Labelling Kit (Boehringer), the fragment was sequenced to screen E. histolytica cDNA and genomic DNA libraries constructed in ZAP II (Desscoteaux et al., 1992; Sánchez et al., 1994). Both strands of cDNA and DNA clones containing the full length Ehtbp gene were sequenced. The predicted amino acid sequence of EhTBP was compared with other TBP sequences in GenBank using the algorithm of Altschul et al. (1990). For hydrophagy analyses of EhTBP, the algorithm of Kyte & Doolittle (1982) was used.

Southern and Northern blot analysis. Total DNA from E. histolytica was isolated by discontinuous caesium chloride gradients (Sánchez et al., 1994) and digested with various endonucleases. The digested DNA (5 μg) was electrophoresed through 0.8% (w/v) agarose gels and transferred to nylon membranes (DuPont NEN). On the other hand, E. histolytica DNA was separated by transverse alternating field electrophoresis (TAFE) in 1% (w/v) agarose gels and transferred to nylon membranes (Orozco et al., 1993). Hybridization was performed at 42 °C and filters washed at 65 °C (Orozco et al., 1993). For Northern blot experiments, total RNA from E. histolytica was isolated using 4 M guanidine isothiocyanate and caesium chloride centrifugation as described (Ausubel et al., 1994; Sánchez et al., 1994). Poly(A)+ RNA was purified through poly(U)-Sephadex columns (Gibco-BRL) according to the manufacturer's instructions. Poly(A)+ RNA (3–6 μg) was electrophoresed through formaldehyde 1% (w/v) agarose gels, transferred to nylon membranes and hybridized with Ehtbp DNA using the conditions described above.

In situ PCR. This was performed using a GeneAmp In situ PCR System 1000 (Perkin Elmer), following the manufacturer’s instructions. As primers we used the TBPl (5’ CCGGGGATCCATGTCACACCGAGAGA 3’) and TBP2 (5’ CCGGGAATTCTTATGGATATATCTTC 3’) oligonucleotides.

Determination of the Ehtbp transcription start site. Primer extension experiments were performed as described by Gómez et al. (1998) and Pérez et al. (1998) with some modifications. Total RNA (50 μg) was used to extend the pexttbp (5’ ACCTCAGTTGAACTGACATA 3’) or ptpbn-01 (5’ TCTTCTTGAATGGCTCA 3’) primers. 100 U SuperScript II (Gibco BRL) were used per reaction; the conditions used were 94 °C for 1 min, 42 °C for 10 min and 37 °C for 10 min.

Expression and purification of EhTBP. The full length Ehtbp was cloned in-frame into the pRSETA expression vector (Invitrogen) after PCR amplification of the E. histolytica genomic clone with the oligonucleotides TBPl and TBP2. A recombinant plasmid was isolated and used to transform E. coli BL21(DE3) allowing expression of the recombinant EhTBP as a fusion polypeptide at the N-terminus, containing 4aa (MRGS) followed by a 6 H tag and the sequence GMASMTGQGQMGRLDYDDDDKDRWGS, in which there is an enterokinase recognition sequence. Expression of recombinant EhTBP was induced by 1 mM IPTG for 1.5 h. Cells were harvested by centrifugation and the recombinant EhTBP was partially purified under denaturing conditions through a Ni2+-NTA affinity column (QIAGEN). Further purification was obtained by 12% SDS-PAGE and electrophoration of the protein (Ausubel et al., 1994).

Production, purification of anti-EhTBP rabbit polyclonal antibodies and Western blot analysis. Recombinant EhTBP (100 μg) was electroeluted from a 12% SDS-polyacrylamide gel and subcutaneously injected with 500 μl Freund’s complete adjuvant into a New Zealand rabbit (2–4 months old). Before the serum was obtained, three additional immunizations were done at 2–3 week intervals using Freund’s incomplete adjuvant. Anti-EhTBP-specific antibodies were affinity purified using the recombinant EhTBP immobilized on a nitrocellulose membrane (Ausubel et al., 1994). Western blot assays were done using the purified recombinant TBP and nuclear extracts (Gómez et al., 1998; Pérez et al., 1998) and the ECL detection system (Amersham).

Confocal and immunofluorescence microscopy. Trophozoites were grown overnight on sterile coverslips, fixed with 3:8% (v/v) paraformaldehyde, permeabilized with 0.1% (v/v) Triton X-100 and blocked with 1% (w/v) BSA. The fixed trophozoites were incubated with a 1:10 dilution of rabbit anti-EhTBP polyclonal antibody and then with a 1:1000 dilution of goat anti-rabbit polyclonal antibody labelled with Cy2 Dye (Amersham). Finally, cells were counterstained with 0.5 mg propidium iodide ml-1 in PBS for 5 min and examined through a Nikon microscope attached to a laser confocal scanning system MRC 1024 (Bio-Rad).

RESULTS AND DISCUSSION

Cloning the Ehtbp gene

A 309 bp fragment of E. histolytica DNA was amplified by PCR using primers from bp 354–374 and 651–671 of the A. castellanii tbp gene (not shown). This fragment was used to isolate two identical genomic DNA and cDNA clones, indicating the absence of introns in the Ehtbp gene, as also reported for other E. histolytica genes (Descoteaux et al., 1992; Sánchez et al., 1994). The genomic DNA clone contained a 2190 bp insert with a

Fig. 2. Southern and Northern blot analysis of the E. histolytica tbp gene. (a) Southern blot of E. histolytica genomic DNA digested with EcoRI and Accl enzymes. (b) Southern blot of high molecular mass DNA separated by TAFE. (c) Northern blot of poly(A)+ RNA.
702 bp ORF encoding a predicted 234 aa (26 kDa) polypeptide with a pI of 9.27. The predicted amino acid sequence of this ORF showed 55% identity to the *H. sapiens* TBP, 54% to the *A. castellanii* TBP and 37% to the TBP of *P. falciparum* (Fig. 1a). *Ehtbp* contained an ATCA consensus sequence 8 bp upstream of the ATG start codon, which is similar to the sequence described as a transcription start site (Bruchhaus et al., 1993). In addition, downstream of the stop codon we located three TAATT sequences at 48, 75 and 180 nt, proposed as the polyadenylation signal for several genes (Bruchhaus et al., 1993), and at 76 nt an AATTAA sequence proposed as the polyadenylation signal for the *Ehatubl* gene (Sánchez et al., 1994). As reported for other genes (Tannich & Horstmann, 1992), the *Ehtbp* gene has 68% A+U base preference in the coding regions.

**Characterization of the EhTBP sequence**

The EhTBP C-terminal domain (residues 48–229) comprises 182 residues and its alignment with other TBP functional domains showed 55% and 54% identity with the *H. sapiens* and *A. castellanii* TBPs, respectively, and only 37% identity with the TBP from *P. falciparum* (Fig. 1a; Table 1), which is a protozoon with a similar A+T content in its genome.

The EhTBP C-terminal domain (Fig. 1b) contains all the structural elements identified in other TBPs. It has two
TATA-box binding protein of Entamoeba histolytica

Fig. 5. Purification and immunodetection of the recombinant EhTBP. (a) 12% SDS-PAGE. Lanes: 1, molecular mass markers; 2 and 3, E. coli BL21(DE3) cells transformed with pRSETA (lane 2) or pRSETA/Ehtbp plasmids (lane 3), induced with 1 mM IPTG as described in Methods; 4, recombinant EhTBP purified by Ni²⁺-NTA affinity chromatography. (b). Western blot analysis of molecular mass markers (lane 1), 100μg E. histolytica nuclear extracts (lane 2) and 500 ng purified recombinant EhTBP (lane 3).

direct repeats (DR1 and DR2, residues 57–116 and 147–207, respectively; Fig. 1b) described in other organisms as responsible for protein–DNA binding (Nikolov et al., 1992). The EhTBP repeats show 41–65% identity to those of A. castellanii, P. falciparum and H. sapiens TBPs (Table 1). A region with homology to the sigma factor (σ factor 2.3 and 2.4 domains) (Horikoshi et al., 1989), located at residues 170–201, is contained in DR2 (Fig. 1b). A highly basic region (residues 109–145) connects the two direct repeats in EhTBP (Fig. 1b). This core has eight basic residues, seven of which are conserved in all TBPs studied (Fig. 1a). We found 40–47% identity with the A. castellanii, P. falciparum and H. sapiens TBPs in this region of EhTBP (Table 1).

The EhTBP N-terminal domain is 47 aa long and rich in hydrophobic residues (Fig. 1a). In contrast, the human TBP N-terminus is 158 aa long, but is also rich in hydrophobic residues, with a segment of 38 consecutive glutamines (Q-run) (Peterson et al., 1990) that was not

Fig. 6. Confocal immunofluorescence microscopy of EhTBP. E. histolytica trophozoites were exponentially grown and treated with anti-EhTBP rabbit polyclonal antibodies and counterstained with propidium iodide. (a) Control cells treated only with the goat anti-rabbit polyclonal antibodies coupled to Cy2 Dye and observed in the green channel. (b) Cells stained with propidium iodide and observed in the red channel. (c) Cells treated with anti-EhTBP rabbit polyclonal antibodies and goat anti-rabbit polyclonal antibodies coupled to Cy2 Dye and observed in the green channel. (d) Cells treated as in (b) and (c) observed simultaneously in both channels. N, nucleus; EhkO, EhkO organelle.
found in the EhTBP N-terminal domain. One of the challenges in the study of human parasites is to find differences between host and parasite molecules that can be used as drug targets, and to design vaccine targets. Differences in the TBP N-terminal domains could be used to identify E. histolytica-specific transcription factors or other molecules involved in gene-expression regulation, like the DNA-dependent protein kinase that hyperphosphorylates the N-terminal domain of Xenopus TBP (Labhart, 1996). This phosphorylation was seen to be fully dependent on the presence of the C-terminal core domain.

Copy number of the Ehhtp gene

To assess the copy number of the Ehhtp gene, we carried out Southern blot experiments using both DNA digested with several restriction enzymes and high molecular mass DNA separated by TAFE (Orozco et al., 1993). We used the full-length gene as a probe. Single bands of 7 and 20 kb were observed in DNA digested with EcoRI and AccI, respectively, enzymes that cut outside the ORF (Fig. 2a). Analysis with DNA digested with several other enzymes also revealed a single band (data not shown), suggesting that there is a single copy of the Ehhtp gene in the genome. The PCR fragments, and the cDNA and genomic DNA clones contained the same sequence, also suggesting the presence of a unique gene. However, in TAFE-separated DNA, Ehhtp hybridized with 0.8 and 1-5 Mb chromosomes (Fig. 2b). These two bands, reproducibly obtained in TAFE, could correspond to two copies of the same gene located on different chromosomes, suggesting diploidy of Ehhtp or cross-hybridization with other sequences in the genome.

In situ PCR

To investigate the cellular location of the Ehhtp gene we carried out in situ PCR experiments. The TBP1 (sense) and TBP2 (antisense) oligonucleotides used for the in situ PCR amplified the 702 bp full length tpb gene. In 100% of the cells the Cy5 fluorochrome detected the tpb gene in the nuclei (Fig. 3). In contrast, 95% of the trophozoites also presented small fluorescent spots outside the nucleus, probably corresponding to the EhKo organelle. This strongly suggests that there is more than one copy of the tpb gene in trophozoites. The two copies of the tpb gene must show similar sequence, because a single restriction band was found in Southern blot experiments (Fig. 2a) using the full length gene as a probe, total DNA and enzymes cutting outside the gene. Thus, the two DNA bands detected in TAFE experiments using whole trophozoites may correspond to similar tpb genes located on different DNA molecules, one in the nucleus and another in EhKo. However, some reports indicate that there is a TBP-related factor in Drosophila that is capable of replacing TBP for in vitro transcription (Buratowski, 1997; Hansen et al., 1997). We have not yet investigated the presence of a TBP-related factor in E. histolytica, whose gene could cross-hybridize with Ehhtp.

Transcription of the Ehhtp gene

Northern blot experiments revealed a 1.3 kb transcript in poly(A)+ RNA (Fig. 2c). This mRNA is larger than the Ehhtp ORF, which is 702 bp long. These results agree with other reports indicating that several tbp genes are transcribed as larger mRNAs than their ORFs. The ORFs of the Drosophila (Hoey et al., 1990), human (Peterson et al., 1990) and Plasmodium (McAndrew et al., 1993) tpb genes are 1059, 1017 and 684 bp long, respectively, while their transcripts are 1-6, 1-8 and 1-8 kb, respectively. As in other organisms, Ehhtp mRNA contains untranslated regions (UTRs), probably important for its stability or localization. Human and Drosophila tpb genes have long 5' UTRs (241 and 173 nt, respectively) (Peterson et al., 1990; Hoey et al., 1990), while A. castellanii tpb contains a 67 nt 5' UTR (Wong et al., 1992b).

To confirm that the tpb gene has a long 5' UTR and that it was transcribed as a single transcript, we carried out primer extension experiments using the pexttbp oligonucleotide located 100 bases downstream of the ATG start codon (not shown) or the ptbpn-01 oligonucleotide located at 130 bases upstream of the ATG. In all cases, primer extension results showed a single transcription start site for the tpb gene at the ACGC sequence located 420 bases upstream of the ATG codon (Fig. 4). The Ehhtp mRNA 3' UTR may be 180 bases long to give the 1-3 kb transcript. Most E. histolytica genes have short 5' UTRs (Bruchhaus et al., 1993), although the EhPAK and EhMCM3 mRNAs have 265 and 126 nt 5' UTRs (Gangopadhyay et al., 1997). The Ehhtp 5' UTR is the longest reported for amoebic genes. In these three cases, the transcription start site sequences (CAATT, TTAA and ACGC for EhPAK, EhMCM3 and Ehhtp mRNA, respectively) differed from the ATTCA or ATCA sequences reported as the consensus start sites (Bruchhaus et al., 1993). A putative TATA box (ATTAATT) was located at 28 bp upstream of the transcription start site.

Expression of EhTBP

We cloned the Ehhtp ORF in-frame into the pRSETA expression vector and its expression was IPTG induced. Transformated bacteria produced a 30 kDa polypeptide (Fig. 5a), 4 kDa larger than the molecular mass deduced from the Ehhtp ORF. This was due to the small fusion polypeptide encoded by the expression vector (described in Methods). Rabbit polyclonal antibodies against the recombinant EhTBP were generated and they detected two 26 and 29 kDa polypeptides in nuclear extracts (Fig. 5b). The 29 kDa protein could be the phosphorylated form of the 26 kDa TBP, as has been reported for other TBPs (White et al., 1995; Ghavidel & Schultz, 1997; Chibazakura et al., 1997), although protein degradation can not be disregarded.

Intracellular location of EhTBP

Confocal microscopy and immunofluorescence analysis revealed that EhTBP is located in the nucleus and in
subcellular structures corresponding to EhkO, the novel organelle recently reported (Orozco et al., 1997). In agreement with our in situ PCR experiments (Fig. 3), EhTBP colocalized with propidium-iodide-stained material in EhkO and in the nucleus, confirming that this novel organelle contains genetic material, which may be transcriptionally active. As reported, the diameter of EhkO varied from 0.5 to 5 μm (Orozco et al., 1997). We are currently studying EhkO function and the nature of its genetic material.

In summary, Ehtbp encodes a 234 aa protein with a molecular mass of 26 kDa. It contains a 182 aa conserved C-terminal domain with all the structural elements that have been found in other TBPs (two direct repeats, a basic domain and a α-factor homology domain). As in other organisms, the 47 aa N-terminal domain is rich in hydrophobic residues and is not conserved, differing totally from human TBP. Antibodies against the recombinant protein detected two polypeptides of 26 and 29 kDa in nuclear extracts and immunostained both the nucleus and EhkO. In situ PCR also revealed that these organelles contained at least one copy of Ehtbp.

ACKNOWLEDGEMENTS

Dr E. Orozco is an International Research Scholar from the Howard Hughes Medical Institute (USA). This work was also supported by CONACyT (Mexico).

REFERENCES


Received 12 May 1998; revised 20 August 1998; accepted 28 August 1998.