



HAL
open science

**A new class of transcription initiation factors,
intermediate between TATA box-binding proteins
(TBPs) and TBP-like factors (TLFs), is present in the
marine unicellular organism, the dinoflagellate
*Cryptothecodinium cohnii***

Delphine Guillebault, Souphatta Sasorith, Evelyne Derelle, Jean-Marie Wurtz,
Jean-Claude Lozano, Scott Bingham, László Tora, Hervé Moreau

► **To cite this version:**

Delphine Guillebault, Souphatta Sasorith, Evelyne Derelle, Jean-Marie Wurtz, Jean-Claude Lozano, et al.. A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Cryptothecodinium cohnii*. *Journal of Biological Chemistry*, 2002, 277 (43), pp.40881-40886. 10.1074/jbc.M205624200 . hal-03160956

HAL Id: hal-03160956

<https://hal.science/hal-03160956>

Submitted on 29 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Class of Transcription Initiation Factors, Intermediate between TATA Box-binding Proteins (TBPs) and TBP-like Factors (TLFs), Is Present in the Marine Unicellular Organism, the Dinoflagellate *Cryptothecodinium cohnii**

Received for publication, June 6, 2002, and in revised form, July 24, 2002
Published, JBC Papers in Press, August 1, 2002, DOI 10.1074/jbc.M205624200

Delphine Guillebault‡, Souphatta Sasorith§¶, Evelyne Derelle‡¶, Jean-Marie Wurtz§, Jean-Claude Lozano‡, Scott Bingham||, Laszlo Tora§, and Hervé Moreau‡**

From the ‡Observatoire océanologique, laboratoire Arago, UMR 7628 CNRS-Université Paris VI, BP 44, F-66651 Banyuls-sur-mer cedex, France, the §Institut de Genetique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP163, C. U. de Strasbourg, F-67404 Illkirch cedex, France, and the ||Department of Plant Biology, Arizona State University Main Campus, Tempe, Arizona 85287

Dinoflagellates are marine unicellular eukaryotes that exhibit unique features including a very low level of basic proteins bound to the chromatin and the complete absence of histones and nucleosomal structure. A cDNA encoding a protein with a strong homology to the TATA box-binding proteins (TBP) has been isolated from an expressed sequence tag library of the dinoflagellate *Cryptothecodinium cohnii*. The typical TBP repeat signature and the amino acid motives involved in TFIIA and TFIIB interactions were conserved in this new TBP-like protein. However, the four phenylalanines known to interact with the TATA box were substituted with hydrophilic residues (His⁷⁷, Arg⁹⁴, Tyr¹⁷¹, Thr¹⁸⁸) as has been described for TBP-like factors (TLF)/TBP-related proteins (TRP). A phylogenetic analysis showed that cTBP is intermediate between TBP and TLF/TRP protein families, and the structural similarity of cTBP with TLF was confirmed by low affinity binding to a consensus' TATA box in an equivalent manner to that usually observed for TLFs. Six 5'-upstream gene regions of dinoflagellate genes have been analyzed and neither a TATA box nor a consensus-promoting element could be found within these different sequences. Our results showed that cTBP could bind stronger to a TTTT box sequence than to the canonical TATA box, especially at high salt concentration. Same binding results were obtained with a mutated cTBP (mcTBP), in which the four phenylalanines were restored. To our knowledge, this is the first description of a TBP-like protein in a unicellular organism, which also appears as the major form of TBP present in *C. cohnii*.

In higher eukaryotes, the regulation of transcription is intimately connected to the chromatin structure, and the accessibility of the transcription factors to their recognition elements is facilitated by the chromatin-remodeling processes involving a subset of modifying machines whose properties can alter the nucleosomal structure (1–10). After the relief of repression of the chromatin, transcription is preinitiated by the interaction

of the RNA polymerase and the general transcription factors with the promoter. In mRNA synthesis, the transcription initiation step begins with the recognition of the promoter by the TFIID complex containing the TATA box-binding protein (TBP)¹ and several TBP-associated factors (11–15). The TBP, which is highly conserved among eukaryotes, was considered until recently as the universal transcription initiator factor (16–18). However, new members of the TBP family called TBP-like factors (TLF) or TBP-related proteins (TRP) were identified only in the metazoan. Many studies showed that these new factors could form a stable complex with TFIIA and TFIIB and substitute for TBP in directing transcription *in vitro* by RNA polymerase II (reviewed in Refs. 19, 20).

In higher eukaryotes, promoters do not always contain a TATA box but show an initiator element, which is loosely conserved and encompasses the transcription start site (21, 22). In protists, the TATA box is found in amoebas (*Acanthamoeba*), in slime mold (*Dictyostelium*), in ciliates (*Histriculus cavicola*), and in apicomplexa (*Plasmodium*) (23–28). In trypanosomatida (Kinetoplastidae) and trichomonadida (Parabasalia), neither the TATA box nor analogous sequences were detected among the few characterized genes. However, both showed an initiator element specific to each phylum (29–31).

Dinoflagellates are protists, which are widely distributed in the aquatic environment. These unicellular microorganisms can be free living or parasitic. Both toxic and non-toxic dinoflagellates can proliferate in seawater, causing important economic and health problems. The most prominent feature of dinoflagellate cell biology, unique among eukaryotic cells, is the lack of histones and nucleosomal organization (32–36). Moreover, conversely to other eukaryotes, the dinoflagellate chromosomes remain highly condensed during the G₁ phase, with DNA filaments protruding from the chromosome core where transcription takes place (37). The upstream gene organization is only known in the dinoflagellate species *Gonyaulax polyedra* for two genes: the peridinin chlorophyll-a-binding protein (PCP) and the luciferase genes (38, 39). These two genes exhibited a tandem repeat spaced by an intergenic region of about 1000 bp that contains a common 13-bp sequence, but no

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Both authors contributed equally to this work.

** To whom correspondence should be addressed. Tel.: 33-468-88-73-09; Fax: 33-468-88-73-98; E-mail: h.moreau@arago.obs-banyuls.fr.

¹ The abbreviations used are: TBP, TATA box-binding protein; TLF, TBP-like factor; TRP, TBP-related protein; TRP, TBP-related protein; PCP, peridinin chlorophyll-a-binding protein; RACE, rapid amplification of cDNA ends; GST, glutathione S-transferase; AdMLP, adenovirus major late promoter; EST, expressed sequence tag; TF, transcription factor; cTBP, *C. cohnii* TBP; hsTBP, *h. sapiens* TBP.

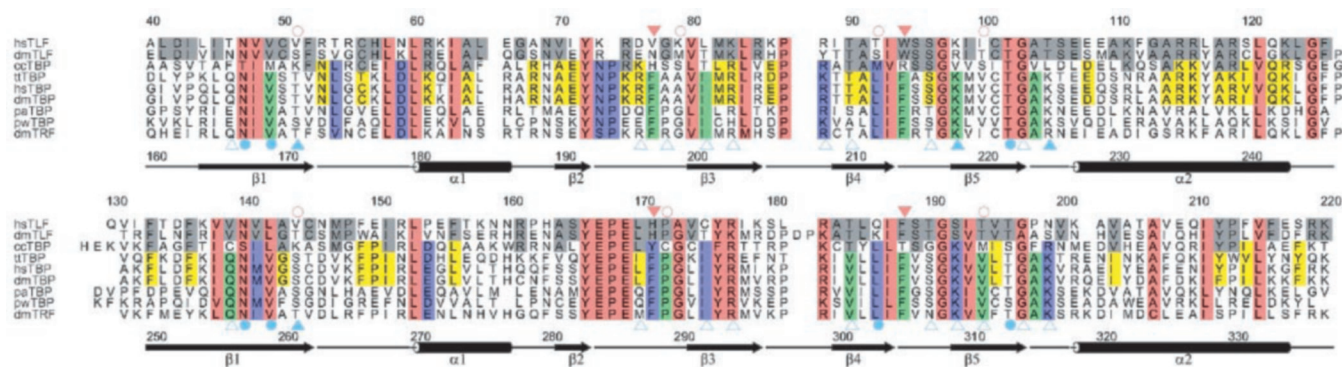


FIG. 1. **cTBP and hsTBP residue numbering are given at the top and bottom of the alignment, respectively.** The secondary structure is given according to hsTBP (49). The position of the four phenylalanine residues conserved among the TBP members are indicated with *red arrows* at the top of the alignment. In TLF, TRF, and cTBP polar or charged residues replace them. They form a complex hydrogen bond network with neighbouring residues highlighted as *red open circles*. (also Fig. 2). *Blue symbols at the bottom* show residues interacting with DNA. *Blue open triangles* indicate residues that are involved in non-specific DNA contacts (phosphate backbone and sugar), whereas the *blue-filled circles* reveal those that are implicated in specific DNA contacts (bases). *Blue-filled triangles* indicate residues where a charge mutation occurs between TBPs and cTBP. The accession number of the cTBP is AF418015. *hs*, Homo sapiens; *dm*, *D. melanogaster*; *cc*, *C. cohnii*; *tt*, *Tetrahymena thermophila*; *pa*, *Pyrobaculum aerophilum*; *pw*, *Pyrococcus woesei*.

TATA box or other known regulatory elements have been found.

To date, only two proteins involved in transcription have been described in dinoflagellate (40). The elucidation of their transcription machinery could allow these organisms to be used as powerful models for the study of eukaryotic transcription in an environment devoid of nucleosomes and provide a better understanding of the transcription network in other eukaryotes. In this work, we identified for the first time in a unicellular organism a cDNA encoding a novel TBP-like protein containing mutated key amino acids involved in DNA binding. We also analyzed the 5'-upstream part of four genes of the dinoflagellate *Cryptocodinium cohnii* and of two genes of the dinoflagellate *Gonyaulax polyedra* without any evidence of any known regulatory elements. We compared the binding of the cTBP and of a mutated form (mcTBP) to a TTTT box and to a canonical TATA box in various salt concentrations.

MATERIALS AND METHODS

Expression of Recombinant Proteins—The TBP cDNA was inserted into a pBlueScript vector and was amplified by the polymerase chain reaction using the *Nde*I restriction site containing primer 5'-TCA CAA TGT CAT ATG GCG GAT ATC TTG GAA-3' and the *Xho*I restriction site containing primer 5'-TAG ATT ATA CTC GAG GGT CTT GAA CTC CGC-3'. The PCR product was subcloned into the pGEX4T expression vector (Amersham Biosciences). The fusion protein GST-cTBP was produced in the *Escherichia coli* strain after 1 mM isopropyl-1-thio- β -D-galactopyranoside induction and purified using glutathione-Sepharose beads (Amersham Biosciences) as described elsewhere (41). The clones producing recombinant proteins were sequenced to check the absence of mutation.

Screening of the cDNA Library—The *C. cohnii* cDNA λ zap expression library was plated in N-Z-amine yeast extract medium Top agar media on NZY agar plates and incubated overnight at 37 °C. Plates were covered by nitrocellulose membranes for 2 min (Gelman, Champs sur Marne, France). The membranes were denatured for 2 min in 1.5 M NaCl, 0.5 M NaOH, neutralized for 5 min in 1.5 M NaCl, 0.5 M Tris-HCl, pH 8.0, and finally rinsed for 30 s in 2 \times SSC, 0.2 M Tris-HCl, pH 7.5. After fixation for 1 h at 80 °C, the membranes were prehybridized for 1 h at 65 °C in a solution containing 5 \times SSC, 5 \times Denhardt, 0.5% SDS, and 1 mg/ml salmon sperm DNA and were then hybridized in the same solution with a denatured ³²P-radiolabeled cTBP probe overnight at 45 °C (permissive temperature). After successive washings with the solutions 1 \times SSC/0.1% SDS, 0.2 \times SSC/0.1% SDS, and 0.1 \times SSC/0.1% SDS, the nitrocellulose filters were dried and exposed to x-ray film for 24 h. Screening was repeated until the positive clones were isolated. They were then excised from the phage, and the open reading frames were sequenced.

Sequence Alignments and Phylogenetic Analysis—Sequences were retrieved with Ballast (42) generated, aligned using ClustalX (43), and

the figure generated with Alscript 2.04 (44). The phylogenetic tree was generated using the Neighbor-joining method with the software phylowin (45).

RACE-PCR—The 5'-upstream sequences of β -tubulin (AF417567), Dip5 (AF417570), DapC (AF417569), and P80 (AF417568) *C. cohnii* genes were amplified by RACE-PCR using the universal genome walker kit (Clontech), and specific nested primers designed at the 5'-extremities of the corresponding cDNAs (40, 46). After purification and sequencing of the amplified DNA fragments, overlapping regions were used to make contigs with the 5'-untranslated regions.

Gel Mobility Shift Assay—Electrophoretic mobility shift assays were carried out with purified GST fusion proteins and double-stranded DNA. The adenovirus major late promoter sequence from -40 to -11 (relative to the start site) or the mutated AdMLP sequence in which the TATAAAA box is substituted with thymines was labeled by phosphorylation of the 5' ends using [γ -³²P]ATP (DNA 5' end-labeling system, Promega). DNA binding reactions were performed with 20 μ l of mixtures as follows: ~60 or 600 ng of GST-cTBP or GST-mcTBP, or 66 ng of human TBP were pre-incubated for 15 min at 27 °C with either buffer or human endogenous TFIIA. 20,000 cpm of probe in a solution containing 5 mM MgCl₂, 60 mM KCl, 10% glycerol, 0.5 mM EDTA, 0.05% Nonidet P-40, 1 mM DTT, 25 ng/ml BSA, 25 ng/ μ l poly(dG-dC), and 12 mM HEPES, pH 8.0, was added and incubated for a further 15 min at 27 °C. hsTBP and TFIIA were purified as described in Refs. 47 and 48. The study of the salt influence on the DNA binding of GST-cTBP was carried out in the same conditions but with an increase in the KCl concentration from 60 mM up to 800 mM. The reactions were resolved on a 4% acrylamide gel at 4 °C in 0.5 \times Tris-Borate-EDTA buffer at 160 V for the appropriate time. The gel was dried and subjected to autoradiography.

Mutagenesis of cTBP—The cTBP cDNA cloned in pBlueScript vector was successively mutated to a phenylalanine at the residues His⁷⁷, Arg⁹⁴, Tyr¹⁷¹, and Thr¹⁸⁸ by PCR mutagenesis using the following four sets of primers, respectively, for each of the four residues: FF1 5'-aatccgcgaaaatttagcagccttacg-3', RF1 5'-atactctgcgtgtccagagcaaacg-3'; FF2 5'-actgcgatggtgttctcatcgggggtc-3', RF2 5'-agctcgggggttcactagctcaactg-3', FF3 5'-gagcctgaacttttctcggcgtcatc-3', RF3 5'-atacagagcattctcagccactttgc-3', FF4 5'-cgtactctgttctctggcggaaaag-3'; RF4 5'-tgcattctgcctcgtgtgcgaaaga-3'. The cDNA was amplified with *Pfu* polymerase. The linear PCR products were ligated overnight at 4 °C and transformed in the DH5 α *E. coli* strain. The open reading frames and mutations were checked by sequencing.

RESULTS

Presence of a Novel TATA box-binding Protein in the Dinoflagellate, *C. cohnii* (cTBP)—A 5'-oriented *C. cohnii* EST library was analyzed and an EST related to the TBP family was identified using the Blast WWW-based program. The corresponding cDNA clone was completely sequenced and showed an open reading frame of 663 bp encoding for a 221-residue protein. The Blast and Prodom searches revealed that this novel

FIG. 2. Unrooted tree generated from an alignment of the core domain of representative TBP, TLF, and TRF sequences. The tree has been generated using the Neighbour Joining method (45). Numbers indicate the branch length. *pw*, *P. woesei*; *pa*, *P. aerophilum*; *dm*, *D. melanogaster*; *tt*, *Tetrahymena thermophila*; *ce*, *Caenorhabditis elegans*; *xl*, *Xenopus laevis*; *hs*, *Homo sapiens*; *ca*, *Candida albicans*; *sc*, *S. cerevisiae*; *cc*, *C. cohnii*.

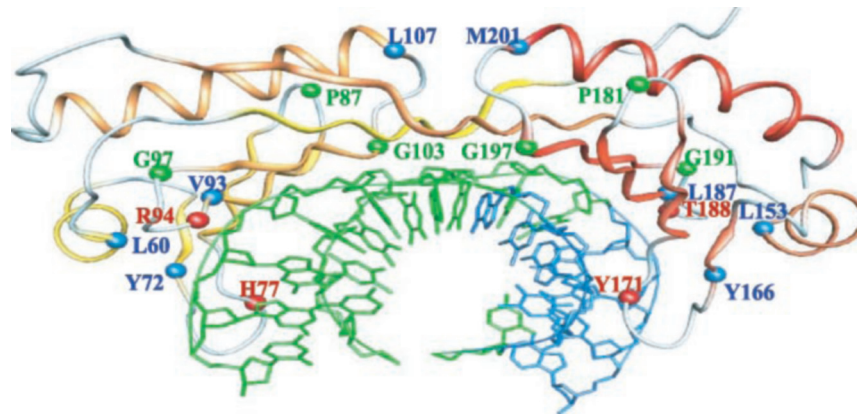
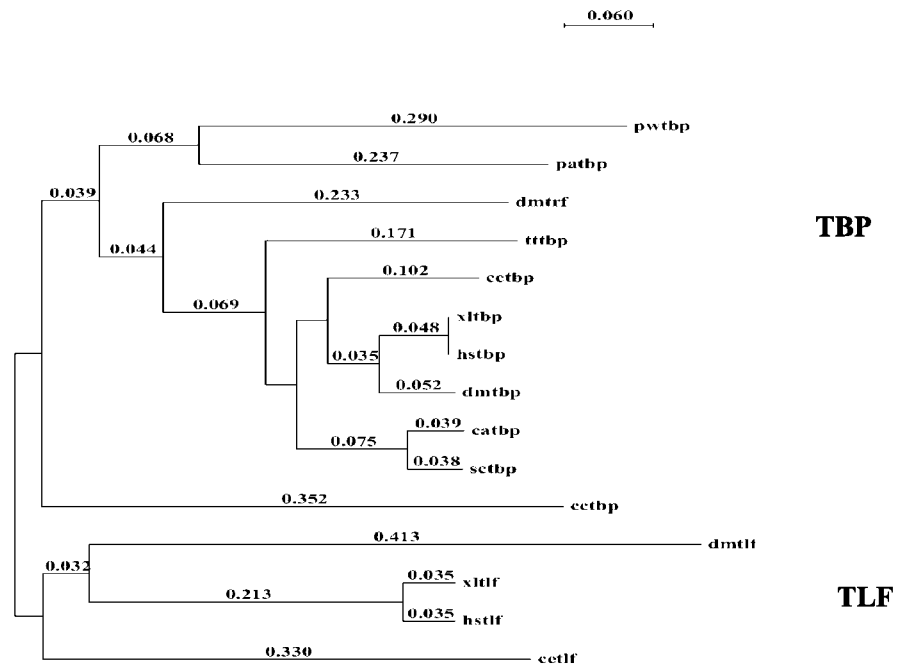


FIG. 3. Model of the core domain of cTBP bound to a TATA box element (the blue segment of the DNA represents the TATA motif). The template structure used is the hsTBP crystal structure solved in a complex with a TATA element at 1.9 Å resolution (51, 52). The protein is drawn as a backbone C- α trace. Residues involved in TBP architecture are highlighted as green and blue spheres. Red spheres indicate the position of the four phenylalanine conserved among eukaryotic TBP members. *, atoms are represented in a standard color scheme: nitrogen, blue; oxygen, red; sulphur, green. Structures have been generated by using Dino version 0.8.3, www.dino3d.org.

protein showed the typical two-repeat signature of TBP encompassing the first 180 amino acids of the C-terminal domain (Fig. 1).

This domain showed 37% identity with the C-terminal region of *Aspergillus nidulans* and *Saccharomyces cerevisiae* TBPs. The C-terminal domain encompassed two directly repeated regions, each around 80 amino acids in length, which is the typical TBP signature. The identity between these two fragments (31%) was similar to that seen in TBPs of other organisms (e.g. human, 31%; yeast 33%). The N-terminal region of the cTBP (44 amino acids) presented no homology as is usually described in other eukaryotic TBPs. Furthermore, key amino acids known to be involved in protein-protein interactions, notably with TFIIA and TFIIB, were also conserved (41).

cTBP Is Intermediate between TBP, TLF, and TRF Members—The most striking difference between cTBP and the TBPs was the replacement of the two pairs of the highly conserved phenylalanines, which are known to play a key role in the DNA kinking by minor groove intercalation, by the His⁷⁷-Arg⁹⁴ and Tyr¹⁷¹-Thr¹⁸⁸ pairs in the first and the second repeat of cTBP (Fig. 1, red arrows). Such a drastic amino acid substi-

tion was also observed in the TLF family. This particular feature could result in the recognition of a DNA element different from a TATA box (19).

Considering the sequence information, cTBP appeared closer to the TBPs (47% similarity with hsTBP) than to the TLFs (32% with hsTLF). Furthermore, the interaction surfaces between TBP and the transcription factors TFIIA (⁷⁰AEYN⁷³ motif) and TFIIB (¹⁶⁶YEPE¹⁶⁹ motif) were highly conserved both in cTBP and TBPs (50, 51). Altogether, these data suggested that cTBP was the closest resemblance to TBPs than to any TBP-like protein identified up to now. This proximity to TBP members was also revealed by phylogenetic tree analysis where cTBP clustered in a separate branch in the TBP sub-tree and was clearly distant from the TLF sub-group as revealed by bootstrap calculation (Fig. 2). In this analysis, cTBP clearly emerged as a member of a new family of transcription factors, which cannot be classified in either the TBP or TLF/TRF family.

The cTBP cDNA Is the Dominant Form of TBP mRNA in C. cohnii—cTBP was isolated after systematic sequencing of an EST library. The possibility that a more canonical TBP could

exist cannot be excluded. To ensure that this new cTBP was not a minor form of TBP, 2×10^5 plaques from a λ Zap cDNA library of *C. cohnii* were screened at low stringency (45 °C) using a probe encompassing the first C-terminal repeat of the cTBP sequence. To check if the screening conditions were optimal for the isolation of TBP as well as TLF or TRP, a hybridization of the yeast genomic DNA was carried out as its genome contains only a TBP gene (20). A signal was detected, indicating that the screening conditions allowed the detection of TBP from the *C. cohnii* cDNA library. Six positive independent clones were isolated, and after sequencing, they appeared entirely identical to the whole cTBP sequence, including the substituted residues that might be involved in DNA binding. These results clearly indicated that the identified cTBP was the major form of a potential TBP family in *C. cohnii*.

cTBP Adopts a TBP-like-fold—The alignment of TBP, TLF, and TRF sequences shown in Fig. 1 is a subset of a much larger alignment comprising 94 sequences retrieved with Ballast and aligned with ClustalX (data not shown) (42, 43, 44). Despite the low sequence conservation with the TBP members, cTBP exhibited a few remarkable amino acid conservations, and a three-dimensional homology model has been generated taking the human TBP crystal structure as a reference (Fig. 3) (19, 51) using the software Modeler 4.0 (52).

The glycine residues in the N- and C-terminal repeats of cTBP (Gly⁹⁷, Gly¹⁰³, and Gly¹⁹¹, Gly¹⁹⁷) were strictly conserved (Fig. 1). These residues, especially Gly⁹⁷ and Gly¹⁹¹, are found in all eukaryotic TBPs and are required to accommodate a particular three-dimensional structure (Fig. 3, *green spheres*), permitting a short turn between β -strands 4 and 5 in each repeat. In addition, a few other residues were highly conserved at the same positions as in the other TBPs, both in the N- and C-terminal repeat of cTBP (Leu⁶⁰/Leu¹⁵³, Try⁷²/Try¹⁶⁶, Val⁹³/Leu¹⁸⁷) (Fig. 1). These buried residues belong to the core of the TBP-fold and form a hydrophobic core in each repeat (Fig. 3, *blue spheres*). Whereas all TBPs presented a conserved salt bridge between residues Glu²²⁷ and Arg³¹⁸ for the hTBP (Fig. 1), which links the two repeats, cTBP exhibits two hydrophobic amino acids (Leu¹⁰⁷ and Met²⁰¹), which generated a hydrophobic cluster instead (Fig. 3, *blue spheres*). However, the secondary structure prediction of cTBP, calculated by the Profile network prediction of Heidelberg (PHD) (53) revealed the same organization as the one derived for the human TBP crystal structure. Altogether, these data indicated that cTBP most likely adopts a saddle-like structure similar to TBP despite some major amino acid substitutions.

In the first repeat, the two usual phenylalanine residues (Phe¹⁹⁷ and Phe²¹⁴ in human) are replaced by a histidine and an arginine in cTBP (His⁷⁷ and Arg⁹⁴), which together with Ser⁷⁹ and Ser⁹⁹, form a hydrogen bond network (Fig. 1, *red arrows and circles*). A similar pattern of interaction has already been observed in the second repeat of *Caenorhabditis briggsae* TLF with the same amino acids, which are, however, arranged differently in the structure (19). In the second repeat, the actual aromatic residues (Phe²⁸⁸ and Phe³⁰⁵) are replaced by Tyr¹⁷⁰ and Thr¹⁸⁸ (Fig. 1, *red arrows*), and to partially compensate the space left by the missing phenylalanine, a few other mutations occurred conferring a configuration that would be able to stabilize the kink through van der Waals contacts with DNA (Fig. 1, *red circles*).

Despite some major residue substitutions within the cTBP/DNA interface, the present data argue in favor of the formation of a similar complex to the one observed in the human TBP/TATA box crystal structure. However, the DNA kinking induced by this novel pattern of polar residue interactions indicates that the DNA element recognized by cTBP would

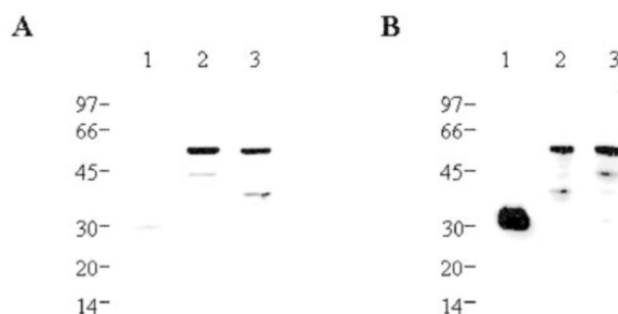


FIG. 4. Characterization of the fusion proteins cTBP and mcTBP by PAGE analysis (A) and Western blotting (B). A, lane 1, GST tag alone (500 ng), lane 2, GST-cTBP (500 ng), and lane 3, GST-mcTBP (500 ng) seen after a Coomassie Blue staining. B, Western blot of A probed with a monoclonal antibody specific to GST. Lane 1, GST tag alone (500 ng); lane 2, GST-mcTBP (500 ng); lane 3, GST-cTBP (500 ng). Molecular mass markers (kDa) are shown to the left of each figure.

probably be different from the TATA box as has already been suggested for TLFs.

No TATA Box Is Found in *C. cohnii* Upstream Gene Sequences—The characterization in *C. cohnii* of a major TBP factor exhibiting substitutions at the key amino acids involved in the TATA box binding prompted us to study the structure of the promoter region of new genes in this microorganism. We amplified and sequenced the 5'-flanking region of four new genes by RACE-PCR. One of these genes encoded the highly expressed protein β -tubulin (accession number AY117680), and the three others nuclear proteins P80, Dip5, and DapC (accession numbers AY117682, AY117683, and AY117681, respectively) (40, 46). The upstream sequences were aligned with those of the PCP and luciferase genes already published from the dinoflagellate species *G. polyedra*. Neither a TATA box nor any other known consensus promoter element could be found within the first 1000 base pairs upstream of the translation start codon (data not shown). This confirms previous observations made for the two dinoflagellate upstream coding sequences of the PCP and the luciferase already known in *G. polyedra*, where no TATA box nor any consensus promoter element could be identified (38, 39). The transcription initiation site has already been identified in the luciferase gene; however, its surrounding sequences could not be found in the promoters of the genes identified here (39).

cTBP Binds to a Mutated TATA Box Element with a Higher Efficiency Than to a Canonical TATA Box—cTBP was produced solubly as a GST-recombinant protein in *E. coli* (Fig. 4). To study in detail its DNA binding, a mutant protein (mcTBP), in which the four amino acids known to correspond to the positions of the four phenylalanines involved in the DNA binding were replaced by phenylalanines, was also produced (Fig. 4).

The cTBP-GST, mcTBP-GST, and the human TBP were incubated with the [γ -³²P]-labeled consensus (TATAAAAA) or mutated (TTTTTTTT) AdMLP oligonucleotides and subjected to polyacrylamide gel shift electrophoresis. A clear shift of the TATA fragment was observed with the hsTBP (Fig. 5A, lane 3), whereas only a very low binding was obtained in the presence of a comparable concentration of cTBP (Fig. 5A, lane 5).

The presence of TFIIA in the incubation did not change significantly the mobility of the cTBP/DNA complex (Fig. 5A, lane 6). The shift observed by incubating the cTBP with the mutated TATA was clearly stronger (Fig. 5A, lanes 7–8), compared with the shift induced by the hsTBP incubated with the same oligonucleotide (Fig. 5A, lanes 10). Interestingly the mutant mcTBP also bound to the mutated TATA (Fig. 5B, lanes 5–6 for the TATA and 7–8 for the mutated TATA) and in general showed a similar binding pattern to the wild type

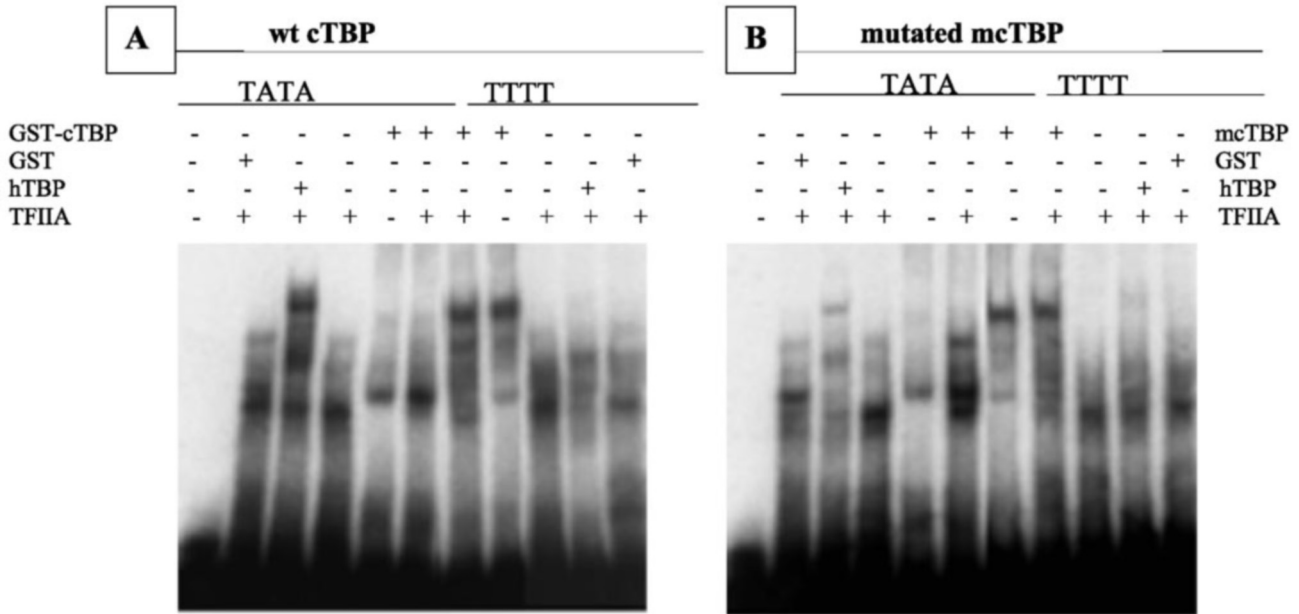


FIG. 5. Interaction of cTBP and mcTBP with DNA probes revealed by gel mobility shift assay. *A*, interaction of hTBP (60 ng) and wild type GST-cTBP (60 ng) with a double-stranded sequence for the TATA box (lanes 1–6) or the TTTT box (lanes 7–11) (oligonucleotides). *B*, interaction of hTBP (60 ng) and mutated GST-mcTBP (60 ng) with a double-stranded sequence for the TATA box (lanes 1–6) or the TTTT box (lanes 7–11) (oligonucleotides).

cTBP. However, in the presence of TFIIA, an increase in the binding to the canonical TATA box by mcTBP was observed (Fig. 5*B*, lane 6). Controls with the GST tag alone and the TFIIA were conducted to ensure that no significant binding of these components to the DNA was obtained (Fig. 5, *A* and *B*). Moreover, as described previously, hTBP did not bind to the mutated TATA box, even in the presence of TFIIA (Fig. 5, *A* and *B*, lane 10).

As cTBP is characterized by particular amino acid residues in the DNA binding site, we tested if a high salt concentration could increase its DNA binding, as reported for archaeobacteria (54). As shown in Fig. 6, the binding of the cTBP to the TATA box dramatically increased with the KCl concentration, with an optimal concentration around 300 mM. However, the high KCl concentration did not change the cTBP binding specificity. In a similar fashion to what was seen at low salt conditions, the binding was more important on the mutated than on the canonical TATA element (data not shown).

DISCUSSION

In this work we describe for the first time in a unicellular eukaryotic organism a new class of transcription initiation factors that show intermediate structural features between the TBP and TLF/TRP family of proteins. However, our DNA binding results indicated that this novel protein behaves more like TLF/TRP proteins than classical TBPs because cTBP does not bind to the classical AdMLP TATA box.

Dinoflagellates are true eukaryotes presenting the unique feature of a very low level of basic proteins linked to their chromatin and a complete absence of nucleosomal structure (32, 55). Very little is known about the molecular processes of dinoflagellate transcription, and although a RNA polymerase II activity has been described in the species *C. cohnii*, the enzyme itself has not been isolated (56). The chromosomes are highly condensed during the G₁ phase and it has been shown that transcription occurred at the periphery of the chromosomes (37, 57). Although some nuclear proteins were isolated and characterized their function in transcription in *C. cohnii* remains unclear, and the cTBP is the first transcriptional dinoflagellate homologue reported (40, 58).

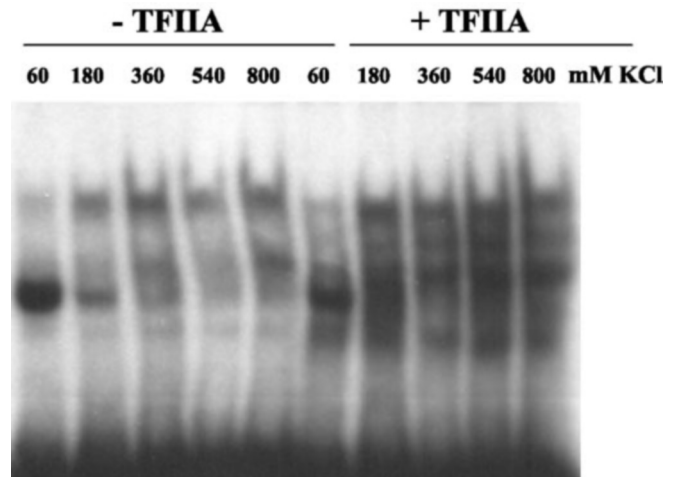


FIG. 6. Interaction of cTBP with the TATA box with increasing concentration of KCl. GST-cTBP (60 ng) was incubated with the TATA box element in the presence of KCl concentrations from 60 to 800 mM, without (lanes 1–5) or with (lanes 6–10) TFIIA.

The determination of the 5' upstream sequence of four *C. cohnii* genes confirmed the absence of a consensus TATA element as already described in two genes of another dinoflagellate species, *G. polyedra*. The six dinoflagellate promoter gene sequences showed a high variation in their global composition for each of the four nucleotides, and no potential transcription initiation motif was found from the sequence analysis. A 13-bp sequence identified in the two *G. polyedra* genes was not found in the new sequences. This 13-bp sequence is either specific to *G. polyedra* or to the highly expressed PCP and luciferase genes or more likely is not a transcription initiating sequence. In the luciferase gene this 13-bp sequence is located about 110 bp upstream of the transcription initiation start, far from the usual distance encountered for the TATA element (about 30 bp) (38, 39).

Sequence comparisons of cTBP with TBPs and TLF/TRPs revealed a probable saddle-like shape structure described in proteins belonging to the TBP family and also emphasized the

probable difference in the DNA sequence recognition. These findings correlate well with our biochemical results in which the low binding of cTBP to the TATA box in standard DNA binding conditions shows that it is functionally similar to a TLF/TRP (59, 60). A low binding to the TATA box was already observed for the TLF/TRPs, and currently no consensus sequence specifically recognized by these proteins is known (19). The effect of the increase of salt concentration on the cTBP interaction with DNA suggests a hypothetical pathway where its DNA binding would be favored by mechanisms depending on salts concentration, allowing the DNA sequences to be released in a highly condensed nuclear environment.

Little is known about how the mutation of the four phenylalanines may affect the TATA box binding. Intuitively, it would be expected that the restoration of the phenylalanines would enable the mTBP to bind the TATA box more efficiently, but this was not observed. This can be explained by a particular structure of the cTBP in which the mutations could induce a whole conformation change rendering the protein unable to bind DNA. However, in the presence of the human TFIIA, cTBP containing the four phenylalanine changes showed a significant binding to the AdMLP TATA box, suggesting that the four conventional phenylalanines may be involved in the TATA box binding specificity.

The discovery of the TLF/TRP proteins in metazoan a few years ago revealed that the initiation of transcription was more complex than initially thought. These proteins are thought to be active on genes involved in specific developmental stages in several metazoan organisms (61–65). TLFs and/or TRPs have only been reported in metazoan and not in unicellular organisms, even in *S. cerevisiae*, for which the genome is entirely sequenced and well annotated (19, 20). The expression of the cTBP as the major TBP-related protein in the unicellular organism *C. cohnii*, which does not have developmental stages, suggests that alternative mechanisms to initiate transcription can exist. This emphasizes the possibility that, as the original TBP found in dinoflagellates, the TLF/TRPs could recognize different initiation sequences that fulfill different roles in other organisms. It is tempting to propose a link between the unique structure of dinoflagellate chromatin, the absence of TATA or any consensus upstream element, and the presence of the cTBP as the major TBP protein (32). Further investigations for the presence of such unique transcription initiation factors in other dinoflagellate species and/or in other unicellular eukaryotes will be necessary to study this functional and evolutionary diversity.

Acknowledgments—We thank M. Albert, M. Groc, and C. Mary for technical assistance and Dr. West and Dr. Rebecca Jolly for correcting the manuscript. We acknowledge Dr. Gilles Crevel for critical reading of the manuscript. We are grateful to Dr. Sue Cotterill for her assistance and Dr. E. Von Baur and M. Strubin for help at a certain stage of this project.

REFERENCES

- Armstrong, J. A., and Emerson, B. M. (1998) *Curr. Op. Genet. Dev.* **8**, 165–172
- Cairns, B. R. (1998) *Trends Biochem. Sci.* **23**, 20–25
- Cox, J. M., Kays, A. R., Sanchez, J. F., and Schepartz, A. (1998) *Curr. Biol.* **2**, 11–17
- Gregory, P. D., and Hörz, W. (1998) *Curr. Op. Cell Biol.* **10**, 339–345
- Kadonaga, J. T. (1998) *Cell* **92**, 307–313
- McAdams, H. H., and Arkin, A. (2000) *Curr. Biol.* **10**, R318–R320
- Uhlmann, F. (2001) *Curr. Biol.* **11**, R384–R387
- Wolffe, A. P., and Hayes, J. J. (1999) *Nucleic Acids Res.* **27**, 711–720
- Workman, J. L., and Kingston, R. E. (1998) *Annual Rev. Biochem.* **67**, 545–579
- Wu, C. (1997) *J. Biol. Chem.* **272**, 28171–28174
- Berk, A. J. (1999) *Curr. Op. Cell Biol.* **11**, 330–335
- Buratowski, S. (2000) *Curr. Op. Cell Biol.* **12**, 320–325
- Klug, A. (2001) *Science* **292**, 1844–1846
- Ranish, J. A., and Hahn, S. (1996) *Curr. Op. Genet. Dev.* **6**, 151–158
- Johnson, K. M., Mitsouras, K., and Carey, M. (2001) *Curr. Biol.* **11**, R510–R513
- Bareket-Samish, A., Cohen, I., and Haran, T. E. (2000) *J. Mol. Biol.* **299**, 965–977
- Burley, S. K. (1996) *Curr. Op. Struct. Biol.* **6**, 69–75
- Rowlands, T., Bauman, P., and Jackson, S. P. (1994) *Science* **264**, 1326–1329
- Dantoni, J. C., Wurtz, J. M., Poch, O., Moras, D., and Tora, L. (1999) *Trends Biochem. Sci.* **24**, 335–339
- Berk, A. J. (2000) *Cell* **103**, 5–8
- Pugh, B. F., and Tjian, R. (1991) *Genes Dev.* **5**, 1935–1945
- Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S.-H., Kodadek, T., and Burley, S. K. (1999) *Genes Dev.* **13**, 3217–3230
- Wong, J. M., Liu, F., and Bateman, E. (1992) *Nucleic Acids Res.* **20**, 4817–4824
- Huang, W., and Bateman, E. (1995) *J. Biol. Chem.* **270**, 28839–28847
- Cohen, S. M., Knecht, D., Lodish, H. F., and Loomis, W. F. (1986) *EMBO J.* **5**, 3361–3366
- Kimmel, A. R., and Firtel, R. A. (1983) *Nucleic Acids Res.* **11**, 541–552
- Liston, D. R., and Johnson, P. J. (1999) *Mol. Cell. Biol.* **19**, 2380–2388
- McAndrew, M. B., Read, M., Sims, P. F. G., and Hyde, J. E. (1993) *Gene* **124**, 165–171
- Luo, H., Gilinger, G., Mukherjee, D., and Bellofatto, V. (1999) *J. Biol. Chem.* **274**, 31947–31954
- Quon, D. V. K., Delgadillo, M. G., and Johnson, P. J. (1996) *J. Mol. Evol.* **43**, 253–262
- Quon, D. V. K., Delgadillo, M. G., Khachi, A., Smale, S. T., and Johnson, P. J. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4579–4583
- Soyer-Gobillard, M. O., and Moreau, H. (2000) *Encyclopedia of Microbiology*, pp. 42–54, Academic Press, Orlando, FL
- Raikov, I. B. (1995) *Acta Protozool.* **34**, 239–247
- Spector, D. L. (1984) *Dinoflagellates*, pp. 1–15, Academic Press, Orlando, FL
- Herzog, M., and Soyer, M. O. (1981) *Eur. J. Cell Biol.* **23**, 295–302
- Herzog, M., De Marcillac, G. D., and Soyer, M. O. (1982) *Eur. J. Cell Biol.* **27**, 151–155
- Bhaud, Y., Guillebault, D., Lennon, J. F., Defacque, H., Soyer-Gobillard, M. O., and Moreau, H. (2000) *J. Cell Sci.* **113**, 1231–1239
- Le, Q. H., Markovic, P., Hastings, J. W., Jovine, R. V. M., and Morse, D. (1997) *Mol. Gen. Genet.* **255**, 595–604
- Li, L., and Hastings, J. W. (1998) *Plant Mol. Biol.* **36**, 275–284
- Guillebault, D., Derelle, E., Bhaud, Y., and Moreau, H. (2001) *Protist* **152**, 127–138
- Moore, P. A., Ozer, J., Salunek, M., Jan, G., Zerby, D., Campbell, S., and Lieberman, P. M. (1999) *Mol. Cell. Biol.* **19**, 7610–7620
- Plewniak, F., Thompson, J. D., and Poch, O. (2000) *Bioinformatics* **16**, 750–759
- Thompson, J. D., Gibson, T. F., Plewniak, K., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882
- Barton, G. J. (1993) *Protein Eng.* **6**, 37–40
- Galtier, N., Gouy, M., and Gautier, C. (1996) *Comput. Appl. Biosci.* **12**, 543–548
- Ausseau, J., Soyer-Gobillard, M. O., Géraud, M.-L., Bhaud, Y., Baines, I., Preston, T., and Moreau, H. (2000) *Protist* **150**, 197–211
- Brou, C., Chaudray, S., Davidson, I., Lutz, Y., Wu, J., Egly, J. M., Tora, L., and Chambon, P. (1993) *EMBO J.* **12**, 489–499
- Lescure, A., Lutz, Y., Eberhard, D., Jacq, X., Krom, A., Grummt, I., Davidson, I., Chambon, P., and Tora, L. (1994) *EMBO J.* **13**, 1166–1175
- Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 4862–4867
- Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) *Science* **272**, 830–836
- Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) *Nature* **377**, 119–128
- Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779–815
- Rost, B., Sander, C., and Schneider, R. (1994) *Comput. Appl. Biosci.* **10**, 53–60
- O'Brien, R., DeDecker, B., Fleming, K. G., Sigler, P. B., and Ladbury, J. E. (1998) *J. Mol. Biol.* **279**, 117–125
- Rizzo, P. J. (1991) *J. Protozool.* **38**, 246–252
- Rizzo, P. J. (1979) *J. Protozool.* **26**, 290–294
- Sigee, D. C. (1984) *Biosystem* **16**, 203–210
- Bhaud, Y., Géraud, M. L., Ausseau, J., Soyer-Gobillard, M. O., and Moreau, H. (1998) *J. Euk. Microbiol.* **46**, 259–267
- Rabenstein, M. D., Zhou, S., Lis, J. T., and Tjian, R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4791–4796
- Ohbayashi, T., Makino, Y., and Tamura, T. (1999) *Nucleic Acids Res.* **27**, 750–755
- Dantoni, J.-C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000) *Mol. Cell.* **6**, 715–722
- Holmes, M. C., and Tjian, R. (2000) *Science* **288**, 867–870
- Kaltenbach, L., Horner, M. A., Rothman, J. H., and Mango, S. E. (2000) *Mol. Cell.* **6**, 705–713
- Veenstra, G. J. C., Weeks, D. L., and Wolffe, A. P. (2000) *Science* **290**, 2312–2315
- Müller, F., Lakatos, L., Dantoni, J.-C., Strähle, U., and Tora, L. (2001) *Curr. Biol.* **11**, 282–287