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 analogous sequences were detected among the few characterized genes. However, both showed an initiator element, 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 (Hiriculicis cavaicola), 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 Gongylonema 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
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 Cryptophycocinium 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 XhoI restriction site containing primer 5′-TCA CAA GCG GAT ATC TTG GAA-3′ and the Xhol restriction site containing primer 5′-TAG ATT ATA CTC GAG GGT CTT GAA CTC-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-

**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′-atactctgcgtgtcccagagcaaacgc-3′, RF1 5′-atactgctgggtcctcagccgaaga-3′; FF2 5′-agtctgatgtcttctgaagggggc-3′, RF2 5′-agtctgatgtcttactcagactc-3′; FF3 5′-gactctgcttttcggtgtagtct-3′, RF3 5′-atactgcagcttctctggcggaaa-3′; RF4 5′-ccctttgcttctctggcggaaa-3′. The DNA was amplified with Pfu polymerase. The linear PCR products were ligated overnight in the presence of 50–100 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. DNA binding reactions were performed with 20 μl 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. hsTBP and TFIIA were purified as described in Refs. 47 and 48.

**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...
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 His^{77}-Arg^{94} and Tyr^{171}-Thr^{188} pairs in the first and the second repeat of cTBP (Fig. 1, red arrows). Such a drastic amino acid substitution 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 (^{70}AEYN^{73} motif) and TFIIB (^{166}YEPE^{169} 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 × 10⁵ 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 Blastall 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, Val¹⁷¹/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 Arg²¹⁸ for the hTBP (Fig. 1), which links the two repeats, cTBP exhibits two hydrophobic amino acids (Leu¹⁰³ and Met²⁰⁶) (Fig. 1), which generated a hydrophilic 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 (His¹⁷⁷ and Arg⁲⁸⁸), which together with Ser⁰⁸⁷ and Ser²⁰⁶, 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 (Phe³⁰⁸ and Phe³⁰⁵) are replaced by Tyr⁷⁰⁷ and Thr⁸⁸⁸ (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 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 sites have 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 (TATAAAA) 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. 4. Characterization of the fusion proteins cTBP and mcTBP by PAGE analysis (A) and Western blotting (B).](image-url)
However, in the presence of TFIIA, an increase in the binding to the canonical TATA box by mcTBP was observed (Fig. 5B, 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, hsTBP 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 archaebacteria (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/TRF 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).

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 TLP/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 mcTBP 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 metazoa a few years ago revealed that the initiation of transcription was more complex, 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 protein (32). Further investigations for the presence of such unique transcription initiation factors in other dinoflagellates species and/or in other unicellular eukaryotes may exist. This emphasizes the possibility that, as the original organism—Scyphidia—was active on genes involved in specific developmental stages in metazoa a few years ago revealed that the initiation of transcription was more complex, 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 protein (32). Further investigations for the presence of such unique transcription initiation factors in other dinoflagellates 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 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

1. Armstrong, J. A., and Emerson, B. M. (1998) Curr. Opin. Genet. Dev. 8, 165–172

2. Cairns, B. R. (1998) Trends Biochem. Sci. 23, 20–25

3. Cox, J. M., Kuyk, A. R., Sanchez, J. F., and Scheperazu, A. (1998) Curr. Biol. 2, 11–17

4. Gregory, P. D., and Hore, W. (1998) Curr. Opin. Cell Biol. 10, 339–345

5. Kadonaga, J. T. (1998) Cell 92, 307–313

6. McAdam, H. H., and Arkin, A. (2000) Curr. Biol. 10, R313–R318

7. Uhmann, F. (2001) Curr. Biol. 11, R384–R387

8. Wolfe, M. L., and White, J. J. (1999) Nucleic Acids Res. 27, 711–720

9. Workman, J. L., and Kingston, R. E. (1998) Ann. Rev. Biochem. 67, 545–579

10. Wu, C. (1997) J. Biol. Chem. 272, 28171–28174

11. Berg, A. J. (1999) Curr. Op. Cell Biol. 11, 330–335

12. Buratowski, S. (2000) Curr. Op. Cell Biol. 12, 320–325

13. Klug, A. (2001) Science 292, 1844–1846

14. Ramish, J. A., and Hahn, S. (1996) Curr. Op. Genet. Dev. 6, 151–158

15. Johnson, K. M., Mitsouras, K., and Carey, M. (2001) Curr. Op. Biol. 11, R55–R513

16. Bareket-Samish, A., Cohen, I., and Haran, T. E. (2000) J. Mol. Biol. 299, 965–977

17. Burley, S. K. (1996) Curr. Op. Struct. Biol. 6, 69–75

18. Rowlands, T., Bauman, P., and Jackson, S. F. (1994) Science 264, 1326–1329

19. Cizek, J., Wurtz, J. I., Poch, O., Moras, D., and Tora, L. (1999) Trends Biochem. Sci. 24, 335–339

20. Berg, A. J. (2000) Cell 103, 5–8

21. Pugh, B. F., and Tjian, R. (1991) Genes Dev. 5, 1935–1945

22. Patkowiak, G. A., Kim, J. L., Sun, L., Yang, S.-H., Kodadek, T., and Burley, S. K. (1999) Genes Dev. 13, 3217–3230

23. Wong, J. M., Liu, F., and Bateman, E. (1999) J. Biol. Chem. 274, 28839–28847

24. Cohen, S. M., Knecht, D., Lodish, H. F., and Loomis, W. F. (1986) EMBO J. 5, 3361–3366

25. Kimmel, A. R., and Fritz, R. A. (1983) Nucleic Acids Res. 11, 541–552

26. Liston, D. R., and Johnson, P. J. (1999) Mol. Cell. Biol. 19, 2380–2388

27. McAndrew, M. B., Read, M., Sims, P. F. G., and Hyde, J. E. (1993) Gene 124, 155–171

28. Luo, H., Guling, G., Mukherjee, D., and Bellofatto, V. (1999) J. Biol. Chem. 274, 31947–31954

29. Quinn, D. V. K., Delgardillo, M. G., and Johnson, P. J. (1996) J. Mol. Biol. 43, 253–262

30. Quinn, D. V. K., Delgardillo, M. G., Khachi, A., Smale, S. T., and Johnson, P. J. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 4579–4583

31. Soyer-Gohilard, M. O., and Moreau, H. (2000) Encyclopedia of Microbiology, pp. 42–54, Academic Press, Orlando, FL

32. Raikhov, I. B. (1995) Acta Protozool. 34, 239–247

33. Ppector, D. L. (1984) Dinoflagellates, pp. 1–15, Academic Press, Orlando, FL

34. Herzog, M., and Soyer, M. O. (1981) Eur. J. Cell Biol. 23, 295–302

35. Herzog, M., De Marcillac, G. D., and Soyer, M. O. (1982) Eur. J. Cell Biol. 27, 151–155

36. Bhaud, Y., Guillebaud, D., Lennon, J. F., Defacque, H., Soyer-Gohilard, M. O., and Moreau, H. (2000) J. Cell Sci. 113, 1231–1239

37. Le, Q. H., Markovic, P., Hastings, J. W., Jouvene, R. M., and Morse, D. (1997) Mol. Gen. Genet. 355, 585–604

38. Li, L., and Hastings, J. W. (1998) Plant Mol. Biol. 36, 275–284

39. Guillebaud, D., Derelle, E., Bhaud, Y., and Moreau, H. (2001) Proteol. 152, 127–138

40. Moore, P. A., Ozer, J., Salunek, M., Jan, G., Zeryb, D., Campbell, S., and Lieberman, P. M. (1999) Mol. Cell. Biol. 19, 7610–7620

41. Piewniak, F., Thompson, J. D., and Poch, O. (2000) Bioinformatics 16, 750–759

42. Thompson, J. D., Gibson, T. F., Piewniak, K., Jeannounguang, F., and Higgins, D. G. (1997) Nucleic Acids Res. 25, 4876–4882

43. Barton, G. J. (1993) Protein Eng. 6, 37–40

44. Galtier, N., Gouy, M., and Gautier, C. (1996) Comput. Appl. Biosci. 12, 543–548

45. Aue, S., M., and Gaur, A. (1993) J. Comput. Appl. Biosci. 12, 564–580

46. Aue, S., M., and Gaur, A. (1993) J. Comput. Appl. Biosci. 12, 564–580