The TBP gene from Aspergillus nidulans—structure and expression in Saccharomyces cerevisiae

Robert Kucharski1† and Ewa Bartnik1,2

Author for correspondence: Ewa Bartnik. Tel: +48 22 6597072 ext. 2247. Fax: +48 22 6584754. E-mail: ebartnik@ibbrain.ibb.waw.pl

Department of Genetics, University of Warsaw1, and Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawłowskiego 5a, 02-106 Warsaw, Poland

The genomic and cDNA copy of the TATA-binding protein (TBP) gene from the filamentous fungus Aspergillus nidulans have been cloned. The gene is interrupted by four introns, one of which is in the long 5' untranslated region of 615 bp. The transcription initiation site was established and the levels of mRNA were analysed under diverse growth conditions and found to vary severalfold. The gene encodes a protein of 268 amino acids composed of an N-terminal domain of 88 amino acids with no significant homology to other TBPs and a C-terminal domain of 180 amino acids with about 95% homology to other fungal TBPs. A cDNA clone under the yeast ADH1 promoter was able to substitute for the yeast TBP gene in vivo; however, the transformants obtained grew poorly at 35 °C and on galactose and glycerol at 30 °C, though they could grow in the presence of copper ions or aminotriazole at this temperature. This phenotype may be the result of altered function of A. nidulans TBP in certain yeast transcription activation pathways.

Keywords: Aspergillus nidulans, TBP

INTRODUCTION

Since the cloning of the first gene encoding TATA-binding protein (TBP) from Saccharomyces cerevisiae, the TBP genes (or more often CDAs) have been cloned from a number of organisms. General comparisons indicate a strong conservation of the C-terminal region of the TBP protein with a considerable variation of the N-terminal region. The crystallographic (three-dimensional) structure of the protein–DNA complex has been analysed and a lot is known about the interactions of TBP with various general and specific transcription factors such as TAFs, TFII B, E1A and others (for a review see Hernandez, 1993).

For many years, the filamentous fungus Aspergillus nidulans has served as a model organism for studying gene structure and expression. The detailed regulation of many genes has been analysed, and the genes for a number of specific transcriptional regulators have been cloned and sequenced. So far, however, there are no data on structure and interactions of the components of its basic transcriptional apparatus. Moreover, no TBP gene sequence from a filamentous fungus has been published. We therefore started the analysis of A. nidulans transcription initiation complexes from the isolation and characterization of the TBP gene.

METHODS

Plasmids and DNA libraries. pBluescript KS(+) was from Stratagene and the S. cerevisiae shuttle vector pL361 was obtained from Dr T. Vernet (McGill University, Canada). The pIC19R plasmid containing a fragment of the A. nidulans actin gene was obtained from Dr R. Bradshaw (Massey University, Palmerston North, New Zealand).

The A. nidulans genomic DNA library in pEMBL3 was obtained from Dr T. Vernet (McGill University, Canada). It contains 7–20 kb DNA fragments obtained by partial Sau3A digestion cloned in the BamHI site of the vector. The A. nidulans cDNA library in λ ZAP was purchased from the Fungal Genetics Stock Center (University of Kansas Medical Center, USA).

Organisms and growth conditions. The following strains were used: Escherichia coli DH5α (Sambrook et al., 1989) and XL-1 Blue (Stratagene), S. cerevisiae SHY67 (Hahn et al., 1989), A. nidulans proA6 pabaA9 biA1 suC6 and the proto-
trophic y strain (Department of Genetics, University of Warsaw).

The organisms were grown on solid and liquid media: for E. coli, LB and 2 × TY; for S. cerevisiae, YPD or YNB (with supplements as required; Sambrook et al., 1989). Medium containing 5-fluoroorotic acid for plasmid shuffling was according to Boeke et al. (1987). Carbon sources were used at 2%. Copper sulphate was added up to a 0.25 mM concentration and 50 mM 3-amino-1,2,4-triazole was used. A. nidulans was grown at 37°C on CM or MM medium (Pontecorvo et al., 1953) supplemented with a carbon source (glycerol or glucose) at 1% and nitrogen sources (glutamine, sodium nitrate or arginine) at 10 mM. Heat shock was performed at 55°C for 15 min in standard MM.

**Molecular biology techniques.** E. coli was transformed by the standard calcium chloride method (Sambrook et al., 1989). Yeast cells were transformed according to Chen et al. (1992).

Nucleic acid isolation and manipulation were performed according to Sambrook et al. (1989) unless stated otherwise.

A. nidulans DNA was prepared according to Yelton et al. (1985).

RNA from A. nidulans was prepared using the TRIzol reagent (BRL). poly(A) RNA was purified by single-step chromatography on oligo-dT cellulose (Sigma).

Restriction endonucleases and ligase were from Amersham and MBI Fermentas and were used according to the manufacturers’ instructions.

DNA fragments were purified after excision from gels using electroderection in the Bio-Rad model 422 apparatus.

DNA electrophoresis was performed in standard Tris-acetate buffer using 0.3–3% agarose (Sigma type 11).

DNA fragments were labelled using [α-32P]dATP (Amersham) and the Megaprime kit (Amersham). Nested deletions for sequencing by the dideoxy method were made using the Pharmacia double-stranded Nested Deletions kit. Sequencing was done using Sequenase version 2.0 (USB) or Taq polymerase (MBI Fermentas or Promega). Sequencing and primer extension products were analysed on 6% polyacrylamide gels containing 50% urea.

Primer extension was performed according to de Graaff et al. (1994) with some modifications, using Superscript II RTase (BRL) and 10 µg total RNA.

**PCR.** PCR reactions were performed in a Perkin Elmer Thermocycler. The primers were degenerate 20-mers corresponding to NAENPK (5’ primer ‘TBPFOR’) and FVSGLIV (3’ primer ‘TBPREV’). They were added at 20 pmol each to 1 µg genomic DNA. Initial denaturation was for 5 min at 94°C. Cycling conditions were five cycles of low stringency (denaturation for 1 min at 92°C; annealing at 35°C; ramping for 2.5 min at 72°C; elongation for 3 min at 72°C) followed by 45 moderate stringency cycles (denaturation for 1 min at 92°C; annealing for 1 min at 42°C; elongation for 3 min at 72°C). Each cycle was 5 s longer than the previous cycle. The reaction products were purified on a Centricon 30 concentrator (Amicon). Products were cloned in the T-tailed vector prepared from pBluescript KS(+) according to Marchuk et al. (1991).

**Hybridization.** Hybond N (Amersham) was used for all blotting procedures and hybridizations were performed in a Hybaid mini-oven using standard solutions. The Southern genomic hybridizations were performed for 24 h at 65°C in 3 × SSC with 0.25% lyophilized skim milk. After hybridization, the filters were washed four times for 10 min in 3 × SSC, 0.1% SDS at room temperature and then for 3 min at 65°C in 0.1 × SSC, 0.1% SDS. After hybridization either standard autoradiography was performed or the filters were used to expose phosphor storage screens and the images were analysed using a PhosphorImager SF (Molecular Dynamics).

**Sequence analysis.** Sequence homologies were analysed using BLAST (Altschul et al., 1990). Multiple sequence alignments were performed with the clustal v package (Higgins & Sharp, 1988). Phylip (Felsenstein, 1989) was used to derive a phylogenetic tree. The algorithm described by Zuker (1989) was used to analyse potential RNA conformation. The ImageQuant version 3.3 program (Molecular Dynamics) was used for quantification of hybridization signals.

**RESULTS AND DISCUSSION**

The TBP C-terminal domain of 180 amino acids is strongly conserved and the degenerate oligonucleotide primers for PCR were designed on the basis of alignment of sequences from seven organisms. We did not attempt to use a heterologous probe as preliminary data (P. P. Stepień, personal communication) indicated that no hybridization of the S. cerevisiae TBP gene to A. nidulans DNA would be obtained. The PCR primers (degenerate 20-mers) corresponded to NAENPK (5’ primer ‘TBPFOR’) and FVSGLIV (3’ primer ‘TBPREV’) sequences of the C-terminal domain.

**Fig. 1.** Genomic hybridization of DNA restriction fragments of A. nidulans with the AnTBPG probe - a fragment of the TBP coding sequence of A. nidulans obtained using PCR. Lanes 1–9 show a Southern blot of genomic DNA digested with: 1, EcoRV; 2, BamHI; 3, BglII; 4, EcoRI; 5, HindIII; 6, PstI; 7, SalI; 8, XbaI; 9, XhoI. The molecular mass marker is the 1 kb ladder (BRL).

1264
The designed primers were used for PCR of *A. nidulans* genomic DNA and a fragment larger than expected was obtained (approx. 440 bp instead of 368 bp). The genome-amplified PCR fragment (designated AnTBPG) was cloned and sequenced. A high homology to known TBP coding fragments had been amplified, and that there is only one copy of the TBP coding gene in the genome. This is typical of most eukaryotes, although two copies have been found in *Arabidopsis* and in other plants (Vogel et al., 1993).

The AnTBPG probe was used to screen an *A. nidulans* genomic library in the zEMBL3 vector. One clone, zTBPGS.11, was subjected to detailed restriction and hybridization analysis. A 3.7 kb XbaI–SalI fragment containing the TBP sequence was subcloned into pBluescript KS(+) plasmid. The plasmid obtained was designated pAnTBPG1. The cDNA library was screened and in other plants (Vogel et al., 1993). The AnTBPG probe was used to screen an *A. nidulans* genomic library in the zEMBL3 vector. One clone, zTBPGS.11, was subjected to detailed restriction and hybridization analysis. A 3.7 kb XbaI–SalI fragment containing the TBP sequence was subcloned into pBluescript KS(+) plasmid. The plasmid obtained was designated pAnTBPG1. The cDNA library was screened using the AnTBPG probe and after preliminary restriction mapping of plasmids rescued from 18 clones the largest plasmid was selected for further analysis; it

The designed primers were used for PCR of *A. nidulans* genomic DNA and a fragment larger than expected was obtained (approx. 440 bp instead of 368 bp). The genome-amplified PCR fragment (designated AnTBPG) was cloned and sequenced. A high homology to known TBP coding fragments had been ampliﬁed, and that there is only one copy of the TBP coding gene in the genome. This is typical of most eukaryotes, although two copies have been found in *Arabidopsis* and in other plants (Vogel et al., 1993).
The structure of the TBP gene is typical of most genes of filamentous fungi. It is composed of five exons and four introns. All introns have consensus sequences and exon-intron junctions similar to the consensus occurring in introns of filamentous fungi (5' WG/GTRNGGY...YGCTAAC...YAG/RY 3') (Ballance, 1986; Gurr et al., 1987). Three of the introns have a size which is typical of most A. nidulans introns (up to 80 nucleotides). The first intron is slightly longer (92 nucleotides) and is moreover located in the 5' UTR of the transcript. This is the second known example of an intron in an A. nidulans sequence which is not translated, as a 140 nucleotide intron has been found in the 5' UTR of the facA gene encoding acetyl-CoA synthetase (Connerton et al., 1990). The strong conservation of introns in areas which do not undergo translation may play a regulatory role in the expression of genes in plants and vertebrates (Pearson & Meagher, 1990) but the mechanism has only been described in detail for the leader of the gene encoding human β-actin containing an enhancer which binds a factor required for efficient transcription (Kawamoto et al., 1988). The second and the third introns are found in strongly conserved positions, as in other TBP genes, which may reflect the structure of the ancestral gene (compare Nakashima et al., 1995).

The mature TBP mRNA in A. nidulans has a very long 5' UTR of 615 nucleotides which contains an ORF encoding a peptide of 26 amino acids at the very 5' end (Fig. 2). In general, the length of the leader sequences in A. nidulans mRNAs does not exceed 200 nucleotides and sequences longer than 500 nucleotides are exceptional (Gurr et al., 1987). Three other examples of this type of long leader sequence are known: in the stuA gene (1088 nucleotides; Miller et al., 1991), the creA gene (950 nucleotides; Dowzer & Kelly, 1991) and the bimG gene (885 nucleotides; Doonan & Morris, 1989). All these genes encode regulatory proteins. Moreover, the 5' UTR of the TBP gene shows a significant degree of sequence homology to the leader of the bimG gene, which also contains an uncharacterized short ORF encoding a potential peptide of 34 amino acids (data not shown). In addition, computer analysis revealed that the 5' UTR has regions with a strong tendency to form secondary structure, an especially stable structure [ΔG298 = -27.7 kcal mol-1 (-116.3 kJ mol-1)] can form on the segment 245-450 nucleotides downstream from the transcription initiation site.

The function of the ORF in the mRNA leader (uORF, upstream open reading frame) is unknown. According to Kozak (1987), the presence of two cistrons in eukaryotic mRNA does not prevent translational reinitiation in the second one if the first cistron has a limited length and both are sufficiently far apart. The uORF may decrease the rate of initiation of translation and cause increased transcript degradation (Oliviera & McCarthy, 1995). Two models explaining the role of these uORFs exist: the first proposes a decrease of the efficiency of translation of the proper protein by a decrease in the rate of ribosome movement due to uORF translation (ribosome stall model; Damiani & Wessler, 1993) and the second postulates a repressor function of the uORF-encoded peptide, which, by a specific in-

---

**Fig. 3.** Primer extension analysis of the transcription initiation site. Digital image from PhosphorImager. The analysis of products of reverse transcription of total A. nidulans RNA using a primer complementary to nucleotides 644–663 is shown. The products of PCR sequencing of the pAnTBP-G1 plasmid using the same primer are shown in the first four lanes. The letters denote the nucleotides used for termination in sequencing reactions. tsp, transcriptional initiation site; arrows denote potential minor start sites.
teraction with the ribosome, a termination factor or RNA, causes a blocking of the mRNA 5' end (peptide repressor model; Hill & Morris, 1993). The secondary structure can additionally act as a negative translation regulation signal since in fungi such structures occurring near the initiation codon may stop the small ribosome subunit from scanning (Vega Loso et al., 1993).

Analysis of codon usage was performed using the relative synonymous codon usage (RSCU) index (Sharp et al., 1986). In the TBP gene, optimal codons are the most common, but for valine, asparagine, glutamic acid, aspartic acid and serine, codons which very rarely occur in A. nidulans genes are used. On the basis of our analysis, the TBP mRNA may be considered as a transcript with a moderate translational efficiency. Most genes encoding A. nidulans regulatory proteins show a lack of preference in codon usage (Lloyd & Sharp, 1991).

The promoter of the analysed gene contains many elements characteristic of genes of higher eukaryotes. These include the CAAT box, a polyypyrimidine tract, the Sp1 recognition sequence and a consensus TATA box (Fig. 2). This distinguishes the A. nidulans TBP gene from most fungal genes, which rarely contain all these elements simultaneously (Gurr et al., 1987). The TATA box occurs at the -25 position characteristic of the promoters of most eukaryotic genes transcribed by RNA polymerase II. Even though relatively few transcription factors have been analysed from A. nidulans, in fungi in general many homologues of mammalian transcription factors are known (Dhawan & Lane, 1993). The presence of the above mentioned regulatory sequences in the TBP gene promoter may therefore be considered significant for transcriptional control. Three sequences found in the vicinity of the transcription initiation site are known to be recognized by specific factors involved in general regulation of A. nidulans metabolism (CreA — carbon repression; AreA — nitrogen repression; and AbaA — involved in regulation of conidiophore development).

The presence of both potential binding sites for positive transcriptional regulators (AreA, etc.) in the vicinity of the TBP transcriptional initiation site and of negative regulatory elements in the 5' end of the mRNA suggests that a complicated interplay may occur between processes at the level of initiation of transcription and translation to maintain an optimal TBP level and prevent the formation of non-functional complexes between excess TBP protein and other components of the transcriptional machinery.

Regulation of TBP transcription by main regulatory circuits such as carbon and nitrogen repression was analysed, as were the effects of heat shock and stage of development. The amount of the specific transcript per cell was found to be very low, and poly(A) RNA had to be used for all the experiments. A probe for detecting A. nidulans actin mRNA was used to standardize the amounts of RNA present in each sample. TBP mRNA levels were found to vary depending on the culture conditions, especially depending on the nitrogen source used. The transcript level also increased after incubation at 55 °C. We were unable to isolate sufficient amounts of poly(A) RNA from mycelium subjected to starvation to analyse the level of transcripts for TBP and actin. Analysis of the levels of the TBP transcript in developing mycelium indicated a rapid increase in transcript amounts between 6 and 8 h after swelling the conidia in minimal medium. Insufficient material for detailed analysis was obtained from conidia and during germination (4 h after swelling). The regulation of mRNA levels by carbon and nitrogen source correlates well with the occurrence of potential binding sites for CreA and AreA near the transcription initiation site (data not shown).

In the 3' region of the TBP gene sequences resembling the yeast 'far upstream' transcription termination element (Zaret & Sherman, 1982; nucleotides 2662–2699), the mammalian polyadenylation signal (nucleotides 2735–2740) and a sequence 70% homologous to the efficiency element (EE) of Schiz. pombe (Humphrey et al., 1994; nucleotides 2882–2897) have been found. On the basis of the genomic DNA sequence as well as that of the eight cloned cDNAs analysed, the existence of four alternative polyadenylation sites in the vicinity of various signals can be postulated. The polyadenylation site after position 2701 of the gene is used most often and is located beyond the sequence homologous to one of the S. cerevisiae transcription termination signals. The existence of multiple polyadenylation sites is characteristic of many genes of filamentous fungi (Gurr et al., 1987).

The initiation codon for TBP was determined on the basis of the consensus sequences determined for various organisms (Kozak, 1991) and A. nidulans (R. Kucharski, unpublished data). Only the AUG codon at position 1258 is located within a consensus sequence with an adenine at -3 (the A. nidulans consensus sequence is 5' YCANATGNNY 3'). The presence of the consensus sequence and the distinct homology of the amino acid sequence to the N-terminus of the Schiz. pombe TBP allowed the assignment of the first TBP codon in A. nidulans with a high degree of certainty.

The A. nidulans C-terminal TBP domain of 180 amino acids has 95% identity to corresponding domains in other fungi. In the N-terminal domain of 88 amino acids, only one similar block can be found in another N-terminal domain — that of Schiz. pombe (PATAQ block; Fig. 4). In addition, there are two short arrays of simple sequences (CAR) encoding blocks of two and three glutamines. Comparison of the A. nidulans and S. cerevisiae TBP C-terminal domain sequences indicates that most differences are located in polymorphic positions. A number of amino acid substitutions can be found in the second repeat, especially helices H1' and H2'. As these amino acids are exposed on the protein surface (Chasman et al., 1993) they may participate in protein–protein interactions. The putative regulatory role of TBP N-terminal domains is unclear. They
undergo an independent evolution and differ both in length and sequence between different organisms (Hernandez, 1993). The existence of stretches of glutamines encoded by simple repetitive sequences, e.g. \((\text{YRR})_n\), is characteristic of the TBP N-terminal domains of higher eukaryotes (Hancock, 1993) and may be the effect of a base slippage phenomenon. A phylogenetic tree based on the N-terminal domain sequence is shown in Fig. 5.

As there is no easy system of analysing \textit{in vivo} or \textit{in vitro} expression of \textit{A. nidulans} genes (no \textit{in vitro} transcription system has been established and there are essentially no non-integrating vectors for transformation), we introduced the TBP cDNA under the control of the \textit{S. cerevisiae ADH1} promoter in the shuttle vector pL361 into \textit{S. cerevisiae} strain SHY87, which allows for the selection of transformants bearing only the recombinant \textit{A. nidulans} TBP sequences and deprived of the yeast TBP functions. However, in

**Fig. 4.** Alignment of N-terminal (a) and C-terminal (b) domains of fungal TBPs: \textit{A. nidulans} (A.n.), \textit{Schiz. pombe} (S.p.; Fikes \textit{et al}., 1990), \textit{S. cerevisiae} (S.c.; Hahn \textit{et al}., 1989) and \textit{Pneumocystis carinii}, prototype (P.c.; Sunkin & Stringer, 1995). The PATAQ block and similar amino acids in the C-terminal domains are emphasized in bold.

**Fig. 5.** Phylogenetic tree of amino acid sequences of C-terminal TBP domains obtained by using the following programs: CLUSTALW (sequence comparison), PROTDIST (creation of distance matrices), NEIGHBOR (obtaining tree structure by neighbour joining) and DRAWGRAM (drawing of a rooted tree). The lengths of the branches do not correspond to the frequencies of amino acid substitutions.

**Fig. 6.** Growth tests of \textit{S. cerevisiae} expressing the \textit{A. nidulans} TBP compared to the control strain grown at 30°C on YPD media containing galactose or glycerol and on YNB mineral media containing aminotriazole and copper ions and YPD/glucose medium at 35°C. Digital camera picture.
contrast to cells containing the S. cerevisiae TBP, cells with A. nidulans TBP grew poorly at 35 °C (Fig. 6). So far, Schiz. pombe (Fikes et al., 1990) and two variants of maize TBP (Vogel et al., 1993) have been found to substitute for the S. cerevisiae TBP, as did human TBP after changing the 231st amino acid from arginine to lysine (Cormack et al., 1994).

S. cerevisiae cells containing the A. nidulans TBP grew very poorly on glycerol and galactose as a source of carbon but were insensitive to 0-25 mM copper sulphate and 50 mM aminitrozole (Fig. 6). These results indicate that the A. nidulans protein probably cannot participate effectively in at least one process – that of Gal4 activation. On the other hand, resistance to copper ions and aminitrozole is normal, indicating that genes stimulated by Ace1 and Gcn4 act normally. The lack of activation by Gal4 and normal activation by Ace1 and Gcn4 may indicate that sequence differences between A. nidulans and S. cerevisiae TBP may cause incorrect behaviour of TBP during stimulation by Gal4. The result concerning the growth on galactose and copper resistance is of particular interest, as mutations analysed so far in yeast TBP (Stargell & Struhl, 1995) caused a simultaneous lack of activation by three acidic activators – Gal4, Ace1 and Gcn4. This may indicate that there are distinct domains in TBP which are responsible for the interaction with different coactivators or for direct interaction with activating domains of Ace1 and Gal4. Even though a function in interaction with transcription factors has been ascribed to a number of the amino acids which are different between the two organisms, the effects of other differences occurring in C-terminal domains and also the influence of the N-terminus of the A. nidulans TBP on the formation of fully functional preinitiation complexes in S. cerevisiae cells cannot be excluded. Our results may indicate the rapid coevolution of components of the transcriptional apparatus in eukaryotic micro-organisms.

ACKNOWLEDGEMENTS

We would like to thank Dr Steven Hahn for kindly providing yeast strain SHY67. This work was supported by grant nos BW-1154 and BW-1203/27 from the Faculty of Biology, University of Warsaw, and grant 6P04A04509 from the State Committee for Scientific Research. Help from the Foundation for Polish Science under the BIMOL Programme is also acknowledged.

REFERENCES


