Cloning and sequence analysis of TFE, a helix-loop-helix transcription factor able to recognize the thyroglobulin gene promoter in vitro

Franoise Javaux, Alena Donda, Gilbert Vassart and Daniel Christophe
IRIBHN and 1Department of genetics, Université Libre de Bruxelles, Campus Erasme, 808 Route de Lennik, 1070 Brussels, Belgium

Received September 30, 1990; Revised and Accepted January 31, 1991

ABSTRACT

A cDNA that encodes a transcription factor able to recognize the thyroglobulin gene promoter in vitro was isolated from a dog thyroid cDNA expression library in λgt11. The library was screened with a multimerized 20 bp-oligonucleotide probe corresponding to the −126 to −107 bp region of the bovine thyroglobulin gene promoter. The specificity of DNA sequence recognition was demonstrated by DNA binding experiments realized with β-galactosidase-fusion protein immobilized on nitrocellulose filters and various unlabelled multimerized competing DNA fragments. The encoded protein, TFE, appears to be the canine counterpart of a recently cloned human transcription factor, ITF-2, that binds to the E5xE2 motif found in both Immunoglobulin heavy and light chains genes enhancers and belongs to the basic-Helix-Loop-Helix family of transcription factors. When TFE protein was produced in a rabbit reticulocyte lysate, it displayed the same specificity of DNA sequence recognition as the β-galactosidase fusion protein and immobilization of the translation product on nitrocellulose still appeared to be essential for detecting in vitro DNA binding activity. Functional data failed to assign a role for TFE in the control of thyroglobulin gene transcription in vitro, suggesting that the selection of TFE clone resulted from the fortuitous presence of a high affinity binding site in the probe used for screening the expression library.

INTRODUCTION

Thyroglobulin (Tg) gene expression is restricted to the follicular thyroid cells and is under the positive control of the pituitary hormone TSH (1). This control occurs, at least partly, at the transcriptional level (2,3). TSH acts via the intracellular messenger cAMP by activating adenylate cyclase (4) but none of the usual DNA consensus sequences found in the promoter of genes regulated by cAMP, cAMP Responsive Element (CRE) or the AP2 recognition sequence (5), are present in the Tg promoter. Moreover, it was shown by run-on in vitro assays that stimulation of Tg gene transcription by cAMP required active protein synthesis and followed relatively slow kinetics (a minimum of 8 hours is required)(3). This contrasts with the 'immediate early' characteristics of the control of gene expression by the CRE-binding factor, CREB (6) and AP2 (7). It is thus expected that Tg gene expression is regulated by cAMP through distinct mechanisms involving transcription factors other than CREB or AP2.

Previous experiments have shown that the first 250 bp of the Tg promoter were sufficient to direct the specific expression of a reporter-gene, the chloramphenicol acetyltransferase (CAT) gene, when transfected in thyroid cells (8,9). Factors involved in the tissue-specific expression and their DNA targets in the rat Tg promoter have been characterized (10). A specific complex between the bovine Tg gene promoter and a thyroid nuclear protein could be detected in an electrophoretic mobility shift assay (EMSA) when a 20 bp fragment of the promoter was used as a probe. This fragment which extends from −126 to −107 bp relative to the transcription start site was called bTgD (11). Point mutations in this region of the promoter decreased drastically the transcription of a linked gene, the CAT gene, in transiently transfected thyroid cells (A. Donda et al., submitted).

We attempted to clone the cDNA that encodes the factor recognizing the −126 to −107 bp fragment from the Tg promoter, using the method described by Singh et al. (12). A clone was obtained which belongs to the so-called basic-Helix-Loop-Helix (bHLH) family of transcription factors (13,14,15).

MATERIALS AND METHODS

Screening of the cDNA library

A dog thyroid cDNA library in λgt11 (16) was screened by a modification (17) of the procedure of Singh et al. (12) using 3M guanidine-hydrochloride to denature the proteins on nitrocellulose filters. The binding buffer (buffer A) was 20 mM Tris-HCl pH 7.5, 75 mM NaCl, 5 mM Na citrate, 5% glycerol, 4 mM β-mercapto-ethanol and 0.25% milk powder. The DNA probe was prepared by 5' phosphorylation of single-stranded oligonucleotides corresponding to the bTgD sequence of the bovine Tg promoter (Figure 1) using T4 polynucleotide kinase.
and radioactive ATP. The two labelled strands were then annealed and the resulting duplex ligated to allow the production of 200 to 400 bp molecules.

DNA was prepared from the isolated λTgD1 phage clone and digested with EcoR I to subclone the cDNA insert according to standard methods (18). The cDNA expression library was re-screened by DNA-DNA hybridization with the labelled λTgD1 DNA insert using a standard procedure (18). From the positive clones, the one containing the longest insert, 2.6 kb, was selected and termed λTgD3.

DNA binding competitions on nitrocellulose filters
Replica filters of confluent IPTG-induced λTgD1 phage plaques were processed as described above and cut into small squares. These were incubated with 0.1 ng of the DNA probe (bTgD or bTgE (see figure 1); prepared as described above), 0.1 µg of polydIdC-polydIdC (Pharmacia) and the indicated excess of unlabelled multimerized competing oligonucleotide in 0.5 ml of buffer A, for two hours at room temperature under slow agitation. The filters were then washed four times for three minutes in 2.5 ml of buffer A, allowed to dry and autoradiographed on Kodak Xar5 films overnight at —80°C with intensifying screen.

In vitro transcription and translation
The complete or truncated (amino acid 335 to end) open-reading frame of the 2.6 kb λTgD3 insert was amplified by PCR. A Xho I restriction site was included in the 3' primer (5' TGGCTCG-AGTCACTCTGGAGTGAG 3'). A Not I restriction site, an optimal surrounding for the translation initiation codon (19) plus an in-frame ATG in the case of the truncated sequence were contained in the 5' primer (5' TTTCGGCGGCCCACCATGTTCACCCCTCTGGA 3': complete; 5' TTTCGG-GCCGCCCCACCATGCTTTGCTTTGCTCTAGAAAAC 3': truncated). The PCR amplified products were cut with the Not I and Xho I restriction enzymes and subcloned in pSP64Tl, a modified pSP64T plasmid with Not I, Mlu I, Sfi I, Xho I sites inserted 3'of the unique Bgl II site. Several independent clones were selected in each case in view of the possible introduction of deleterious mutations during the PCR amplification step.

RNA was synthesized in a 25 µl reaction using as substrate 1 to 2 µg of linearized plasmid DNA and 30 units of SP6 RNA polymerase (BRL) for two hours at 37°C in the presence of 1 mM CAP analogue (Pharmacia).

In vitro translation of the RNAs was done in a rabbit reticulocyte lysate following the recommendations of the manufacturer (Promega). Each 50 µl reaction contained 1 µg of RNA as template.

DNA sequencing and oligonucleotides synthesis
DNA sequencing was performed using an Applied Biosystems 370 A DNA sequencer. XTgD1 and XTgD3 inserts were completely sequenced on both strands from bacteriophage M13 subclones by means of nested deletion mutants generation with exonuclease HI (18). Oligonucleotides were synthesized on an Applied Biosystems 381 A DNA synthesizer.

RESULTS
Isolation of a clone expressing DNA binding activity
A dog thyroid cDNA library in λgt11 (16) was screened with a multimerized 20 bp DNA fragment of the bovine Tg promoter, the bTgD oligonucleotide (Figure 1), according to the method described by Singh et al. (12) and modified by Vinson et al. (17). Among 700,000 recombinant phages screened, one cDNA clone was isolated which expressed a β-galactosidase fusion protein that bound to the bTgD fragment but not to a bTgB probe (Figure 1).
Figure 2. DNA binding competition experiments on immobilized recombinant protein. Replica filters of confluent IPTG-induced λTgD1 phage plaques were treated as described in material and methods and incubated with 0.1 μg of polydIdC-polydIdC and (A) 0.1 ng of bTgD probe and a 100 to 10,000 fold excess of the indicated unlabelled oligonucleotides (+bTgD and +CRE series), (B) 0.1 ng of bTgD probe and 5 μg (5,000 fold excess) of the indicated unlabelled oligonucleotides, (C) 0.1 ng of bTgD probe and increasing amounts of unlabelled bTgD or bTgE, (D) 0.1 ng of bTgD or bTgE probe and 100 ng of unlabelled bTgD or bTgE as competitors.

1. The clone termed λTgD1 was purified to homogeneity and the cDNA insert was isolated by restriction of the phage DNA with EcoRI endonuclease. The cDNA insert size was 1.1 kb.

DNA binding specificity

In order to test the specificity of the binding we developed a DNA binding competition assay with the β-galactosidase fusion protein immobilized on nitrocellulose filters. Replica filters of confluent IPTG-induced λTgD1 phage plaques were incubated with the probe in the presence of a large excess of various unlabelled multimerized oligonucleotides, then rapidly washed and autoradiographed. The use of multimerized oligonucleotides appeared essential for binding to the recombinant protein. Monomer probe or competitors were tested but neither were able to bind to the immobilized protein.

The presence of a 5000 fold excess of unlabelled bTgD allowed the detection of a competition effect on binding to the bTgD probe while even a 10,000 fold excess of an unrelated competitor (CRE sequence) had no effect (Figure 2A). A panel of multimerized fragments of the Tg promoter (Figure 1) as well as unlabelled sequences (CRE or AP2 recognition sequence) were then used as competitors in a fixed 5000 fold excess amount (Figure 2B). As expected, DNA fragments containing bTgD sequence (bTg60, bTgA and bTgD) competed with the binding of the bTgD probe, but heterologous sequences (bTgB, CRE, AP2 recognition site) did not. bTgE, which overlaps the 5' portion of bTgD, displayed high affinity for the fusion-protein while bTgC which shares the 3' portion of bTgD could not compete at all with bTgD binding. To compare the relative affinities of binding of the protein to bTgE and bTgD sequences, a competition experiment was performed using bTgD as probe and increasing amounts of the two unlabelled fragments as competitors (Figure 2C). The results showed that bTgE was able to completely abolish the binding to the bTgD probe already when present in 500 fold excess. Moreover, a bTgE probe displayed also a higher affinity for the protein immobilized on the nitrocellulose filter as is shown on figure 2D where both bTgD and bTgE were cross-competed. These experiments allowed to conclude that the cloned cDNA encoded a protein domain which could recognize with high specificity the sequence shared by bTgD and bTgE fragments.

This 10 bp segment of the bovine Tg promoter extends from −126 to −117 bp from the transcription start. However, this was in disagreement with the EMSA data obtained before with fragment bTgD as probe (11). In the latter assay, methylation interference experiments have now clearly located the binding of the thyroid nuclear factor on the GGAGGACCTC quasi-palindromic motif shared by bTgD and bTgC fragments (A.Donda et al., submitted). According to its preferential binding to the bTgE oligonucleotide, the cloned factor was called thyroid factor E (TFE).
Sequence analysis of the cDNA insert

The Agt11 library was rescreened by hybridization with the 1.1 kb cDNA insert in order to obtain a clone containing the complete sequence encoding TFE. A clone harboring a cDNA insert of 2.6 kb was isolated and subcloned into M13 vector for DNA sequencing. The entire 2666 bp EcoR I fragment sequence (Figure 3) showed a main open-reading frame of 1926 bp. No conserved polyadenylation site could be found in the 3' non-translated region.

Sequence analysis of the cDNA insert

The Agt11 library was rescreened by hybridization with the 1.1 kb cDNA insert in order to obtain a clone containing the complete sequence encoding TFE. A clone harboring a cDNA insert of 2.6 kb was isolated and subcloned into M13 vector for DNA sequencing. The entire 2666 bp EcoR I fragment sequence (Figure 3) showed a main open-reading frame of 1926 bp. No conserved polyadenylation site could be found in the 3' non-translated region.

Analysis of the sequence revealed that the protein was unusually rich in serine residues (20%). This seems to be one of the characteristics of rapidly degraded proteins according to the PEST region hypothesis (22).

The encoded 642 aa-long protein appeared to be the canine counterpart (98.4% homology in 620 amino acids overlap) of a recently cloned human protein which encodes a transcription factor of the immunoglobulin genes, ITF-2 (15). TFE/TTF-2 protein is a member of a newly described family of DNA binding proteins which share a common bHLH motif involved in both protein-protein dimerization and DNA recognition (13). Well-known representatives of this family are the immunoglobulin enhancer binding proteins E12 and E47 and the myogenic determination factor MyoD. As depicted on figure 4A, TFE/TTF-2 and E12 proteins bHLH regions are extremely well conserved. Conservation also extends on both sides of the bHLH motif, where in other respects several charged amino acids clusters can be found (Figure 4B).
In vitro production of TFE protein

The sequences encoding complete TFE protein or a truncated version containing only the carboxy-terminal region with the bHLH motif (see figure 3) were amplified by PCR and subcloned into pSP64T1 vector (see material and methods section) downstream from the SP6 RNA polymerase promoter. RNA was produced in vitro and translated in a rabbit reticulocyte lysate. The synthesized proteins displayed the expected sizes, 68 kd for the complete protein and 32 kd for the truncated version, when analyzed by SDS-PAGE (Figure 5A).

The in vitro-produced TFE protein was used in EMSA with bTgE or bTgD probes. Although various binding and electrophoretic conditions were tested no DNA-protein complex could be observed. This was not so surprising since in vitro DNA binding data for some bHLH transcription factors seem to be rather difficult to obtain. Neither E12 nor E47 proteins were able to produce a DNA methylation interference pattern (13) and no in vitro DNA binding data are known for ITF-2 (15).

Considering that the binding could be observed when the phage encoded fusion protein was immobilized on nitrocellulose filters, we decided to spot the rabbit reticulocyte lysate reaction product on nitrocellulose filters. These were then submitted to the denaturation-renaturation step (17) and incubated with a multimerized bTgE probe. A strong binding could be detected with the complete protein (Figure 5). Surprisingly, although the truncated protein contained the bHLH motif and did effectively correspond to the part of XTgD1 fusion protein encoded by the cDNA insert, it was unable to bind the probe significantly. One possible explanation is that the truncated protein when not linked to the β-galactosidase moiety is too small to bind to nitrocellulose and to DNA at the same time. DNA binding competitions experiments were then realized with the complete protein immobilized on nitrocellulose, by adding a 5,000 fold excess of various unlabelled multimerized oligonucleotides during incubation with the bTgE probe. As expected, bTgA, bTgD and bTgE behaved as strong competitors while bTgB, bTgC and the unrelated AP2 recognition sequence did not abolish the binding of the probe (Figure 5B). The in vitro produced protein displayed thus exactly the same binding specificity as the β-galactosidase fusion protein. Why immobilization on nitrocellulose was essential for the binding to occur remains however to be elucidated.

Function of TFE site in thyroglobulin promoter

The DNA binding competition experiments allowed to demonstrate that TFE was not the nuclear factor binding the bTgD sequence in the EMSA (11 and A. Donda et al., submitted). The factors belonging to the family of bHLH DNA-binding proteins seem to bind with high affinity to a CANNTG core DNA sequence (23). Figure 6 shows that such a motif is indeed found in the -126 to -117 bp fragment of the bovine Tg gene promoter and even overlaps at a G residue (considering the CAGGTG consensus core only for TFE binding) with the binding site identified in EMSA. This raises the question whether both factors could really be present on the Tg promoter at the same time in vivo. The functional relevance of this region of the Tg promoter is supported by the observation that when a GGA - TTC substitution was introduced at the -121 to -119 positions (see figure 6), it greatly (> 80%) decreased the expression of a linked gene in transient transfection experiments (A. Donda et al., submitted). This mutation destroys the quasi-palindromic structure of the 'mobility shift activity' binding site, as determined by methylation interference analysis, and indeed drastically affects the binding in EMSA (A. Donda et al., submitted). The first
one page of a document

In order to solve this ambiguity, a GTG—TCT substitution however also destroys the CAGGTG motif oligonucleotides. In order to clone the cDNA coding for a nuclear factor recognizing the bovine Tg gene promoter identified previously by EMSA (11), we have used an expression screening approach (12,17). A clone was isolated containing a cDNA insert which is likely to encode a transcription factor, TFE, that is able to bind to the −126 to −117 segment of the Tg promoter in vitro. TFE appears to be the bovine counterpart of rTF-2, a recently cloned human transcription factor which binds to E5xSE immunoglobulins genes enhancer motif (15). These factors possess a newly identified DNA binding motif, the so-called bHLH structure (13), which consists of two amphiphatic helices separated by a loop and preceded by a short basic region. It was demonstrated elsewhere that the helix-loop-helix motif enables the dimerization of this kind of factors while the basic region directs the binding to DNA and confers promoter-specific transcriptional activation (24,25,26). Homodimer formation is possible but heterodimer complexes appear to be preferred. Stable dimer formation generally involves a 'ubiquitous' factor, like E12 or E47 protein (13) and a 'cell-specific' one, like MyoD (24). Whether heterodimerization is a general characteristic of those proteins is not yet known. But if this is the case, it is likely that the factor cloned in this study, TFE, plays the role of the 'ubiquitous' partner in the dimer as it shares a higher homology with the 'ubiquitous' E12 protein than with the tissue-specific factor MyoD (24). Whether heterodimerization is a general characteristic of those proteins is not yet known. But if this is the case, it is likely that the factor cloned in this study, TFE, plays the role of the 'ubiquitous' partner in the dimer as it shares a higher homology with the 'ubiquitous' E12 protein than with the tissue-specific factor MyoD (24).

Figure 6. A: Precise location of the two known binding sites in the bovine Tg gene promoter: CANNTG sequence recognized by TFE (circled) and the 3′-neighbouring quasi-palindromic site recognized by the factor identified in EMSA (boxed). The GGA—TTC substitutions introduced in the promoter and the limits of bTgE and bTgD oligonucleotides (−136/−107 coordinates refer to transcription start site as +1) are indicated. Results of the transfection experiments: The results obtained in three separated experiments are shown. CAT activity is expressed as the amount of butyrylated 3H-chloramphenicol, in cpm, produced by cell extracts from individual dishes (relative activity of mutated Em promoter versus 14M wild-type appears in brackets). Cells were maintained either in control medium or in medium supplemented with 10 μM forskolin (Fo) after transfection. B: DNA binding to the bTgE oligonucleotide. 5 μl of the reticulocyte lysate reaction mixture were spotted on nitrocellulose filters which were subsequently treated as described in material and methods and incubated with 0.4 ng of bTgE or bTgEm probes and 0.4 μg of poly dIdC-poly dIdC. Incubation was also effected in the presence of 1.5 μg (3750 fold excess) of bTgE or bTgEm unlabelled oligonucleotides.

G — T replacement however also destroys the CAGGTG motif thought to be essential for TFE binding and indeed results in lack of binding of TFE to the mutated oligonucleotide (data not shown). In order to solve this ambiguity, a GTG — TCT substitution was introduced at the −123 to −121 positions in the Tg promoter. The corresponding bTgEm oligonucleotide (Figure 6a) was not able to compete with bTgE for binding to TFE, nor to bind to TFE when used as a probe (Figure 6b). However, transfection of primary cultured thyrocytes with the pbTgCAT Em construct, which contains the mutated Tg promoter linked to the CAT reporter gene, showed that the GTG — TCT substitution only weakly affected promoter activity (Figure 6a) (the slight reduction in promoter strength may result from a moderate effect of the last G — T replacement on the interaction detected by EMSA). According to this last result, binding of TFE does not appear to be required for proper function of the transfected Tg gene promoter.

**DISCUSSION**

In order to clone the cDNA coding for a nuclear factor recognizing the bovine Tg gene promoter identified previously by EMSA (11), we have used an expression screening approach (12,17). A clone was isolated containing a cDNA insert which is likely to encode a transcription factor, TFE, that is able to bind to the −126 to −117 segment of the Tg promoter in vitro. TFE appears to be the bovine counterpart of rTF-2, a recently cloned human transcription factor which binds to E5xSE immunoglobulins genes enhancer motif (15). These factors possess a newly identified DNA binding motif, the so-called bHLH structure (13), which consists of two amphiphatic helices separated by a loop and preceded by a short basic region. It was demonstrated elsewhere that the helix-loop-helix motif enables the dimerization of this kind of factors while the basic region directs the binding to DNA and confers promoter-specific transcriptional activation (24,25,26). Homodimer formation is possible but heterodimer complexes appear to be preferred. Stable dimer formation generally involves a 'ubiquitous' factor, like E12 or E47 protein (13) and a 'cell-specific' one, like MyoD (24). Whether heterodimerization is a general characteristic of those proteins is not yet known. But if this is the case, it is likely that the factor cloned in this study, TFE, plays the role of the 'ubiquitous' partner in the dimer as it shares a higher homology with the 'ubiquitous' E12 protein than with the tissue-specific factor MyoD (24). Whether heterodimerization is a general characteristic of those proteins is not yet known. But if this is the case, it is likely that the factor cloned in this study, TFE, plays the role of the 'ubiquitous' partner in the dimer as it shares a higher homology with the 'ubiquitous' E12 protein than with the tissue-specific factor MyoD (24).

Unfortunately, DNA binding competition experiments allowed to demonstrate that TFE was not the nuclear factor binding to the bTgD sequence in EMSA. Further, by mutating the TFE binding site in the Tg promoter we have shown that TFE binding is not required for promoter activity in vitro. We therefore suggest that the cloning of TFE resulted only from the presence of the CAGGTG (−126 to −121) sequence in the bTgD probe. Whether heterodimerization is a general characteristic of those proteins is not yet known. But if this is the case, it is likely that the factor cloned in this study, TFE, plays the role of the 'ubiquitous' partner in the dimer as it shares a higher homology with the 'ubiquitous' E12 protein than with the tissue-specific factor MyoD (24).

Considering the expected frequency of the CANNTG motif, and also the possibility of a promoting effect of the nitrocellulose blotting step on DNA binding, it may well be that selection of bHLH DNA-binding domains when using expression libraries is highly favored and leads to promiscuous cloning of this type of factors (see 27 for example). Clearly, this argues for caution when designing probes to be used in such cloning approach.
ACKNOWLEDGEMENTS

The continuous support and critical interest of Dr. J.E. Dumont are deeply acknowledged. This work was supported by grants from the Ministère de la Politique Scientifique (Sciences de la Vie), FRSM and ARBD asbl. F.J. is a fellow of the IRSIA, A.D. is a recipient of a long-term EMBO fellowship and D.C. is a Research Associate of National Fund for Scientific Research (Belgium).

REFERENCES