J24249

MLC/Homology/Probes/Compare

ELSEVIER

Biochimica et Biophysica Acta 1261 (1995) 304–306

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding protein (T/EBP) gene: extensive identity of the deduced amino acid sequence with the human protein

Hidenori Oguchi, Yu-Ting Pan, Shioko Kimura *

Laboratory of Molecular Carcinogenesis, National Cancer Institute, Bethesda, MD 20892, USA

Received 20 January 1995; accepted 27 January 1995

Abstract

A mouse thyroid-specific enhancer-binding protein (T/EBP) gene and its flanking regions have been cloned and completely sequenced. The gene consists of 2 exons and exhibits high similarity (83–97%) to the rat sequence throughout the coding region and including an intron and up to 1.3 kbp upstream to the ATG initiation codon. A cDNA clone encoding human T/EBP has been also isolated and sequenced. Comparison of the deduced amino acid sequence of T/EBP revealed an extensive identity of 98% between mouse and the human protein.

Keywords: Thyroid-specific enhancer-binding protein (T/EBP); cDNA cloning; Gene cloning; Gene structure; Sequence comparison

The thyroid-specific enhancer-binding protein (T/EBP) controls thyroid-specific transcription of the human thyroid peroxidase gene through binding to an enhancer element located 5.5 kilobase pairs (kbp) upstream of the gene's transcription start site [1,2]. T/EBP is also known as thyroid-specific transcription factor-1 which binds to a specific sequence within promoter of the rat thyroglobulin gene and confers thyroid specificity to the expression of the gene [3]. It appears that T/EBP is a common DNA-binding protein that governs the expression of thyroid-specific genes [4–6]. This is further suggested by recent reports that the thyroid-specific expression of rat thyrotropin (TSH) receptor gene is also controlled by T/EBP [7,8]. T/EBP is a homeodomain-containing DNA-binding protein [9] and is expressed in thyroid and lung [2,9]. In situ hybridization and immunohistochemical analyses have shown that during rat embryogenesis, T/EBP is expressed in the thyroid and lung rudiments and in restricted area of the brain, suggesting that T/EBP may play an important role in the development of these organs [10]. Further, it has been recently reported that T/EBP is involved in lung-specific surfactant protein B gene expression [11]. In order to understand the mechanism of regulation of T/EBP gene expression, characterization of the gene structure seems to be inevitable. To this end, we isolated a gene encoding mouse T/EBP. We also cloned a human T/EBP cDNA and present here an extensive identity in amino acid sequence between mouse T/EBP and the human protein.

A mouse genomic library in the vector λFix II (Stratagene Cloning Systems) was screened by using rat T/EBP cDNA coding sequence as a probe [2]. A positive phage clone was isolated and the DNA purified. An 11 kbp *Bam*HI fragment which gave a positive hybridization signal with the probe, was subcloned into pUC 9 and sequenced using fluorescent-tagged dideoxynucleotides on an Applied Biosystems 373A DNA sequencer. A human lung cDNA library was constructed in λZIPLOX (Life Technologies) using poly(A) RNA isolated from human lung H441 cells (ATCC HTB 174). Positive clones that hybridized to the rat T/EBP cDNA coding sequence were subcloned into pUC 9 which were then subjected to sequencing. The nucleotide sequences of the mouse T/EBP gene and its flanking region and the human T/EBP cDNA coding region have been submitted to the EMBL/GenBank databases under accession numbers U19755 and U19756, respectively.

The 11 kbp *Bam*HI mouse genomic fragment was completely sequenced and was found to contain a se-

* Corresponding author. E-mail: shioko@helix.nih.gov. Fax: +1 (301) 4968419.

SSDI 0167-4781(95)00033-X

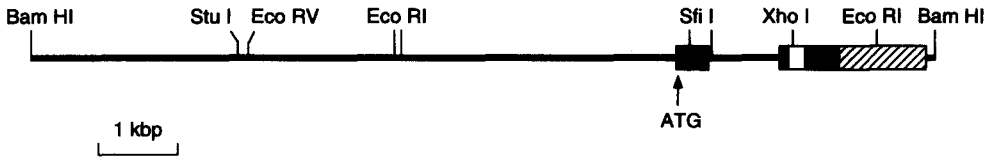

Fig. 1. Schematical diagram of the mouse T/EBP gene. The gene consists of 2 exons. The T/EBP coding sequence and the 3′ untranslated region are shown by closed and hatched boxes, respectively. Location of the homeodomain is indicated by an open box. Unique restriction enzyme sites are shown.

quence corresponding to an entire rat T/EBP cDNA sequence [2,9] (Fig. 1). The mouse T/EBP gene sequence consists of 2 exons by comparison with the rat cDNA sequence and a 907 bp intron. Location of the exon-intron splice junction is identical to the predicted one based on the T/EBP-3 cDNA sequence in our previous report [2] which appeared to be a product of incomplete pre-mRNA processing or alternative splicing of the rat T/EBP gene's primary transcript. The presence of two exons has also been recently reported in the rat T/EBP gene [12]. Sequence of the splice junction in the mouse T/EBP gene confirmed the 'GT-AG' rule [13]. Nucleotide sequence of the mouse T/EBP cDNA coding and the 3′ untranslated region has 97% and 83% similarities to those of rat T/EBP, respectively. The nucleotide sequence similarity between mouse and rat is further extended to approx. 1.3 kbp upstream to the ATG initiation codon and the 907 bp intron where 94 and 84% similarities were respectively obtained [12]. The similarities were calculated based on the mouse sequence since the total number of nucleotides is different between mouse and rat. There is no consensus TATA sequence found in this region and the T/EBP transcriptional start site remains unknown.

The deduced amino acid sequence derived from the mouse T/EBP cDNA coding region was compared with those of rat [9], human, and dog [14] (Fig. 2). Mouse and rat T/EBP protein consists of 372 amino acids whereas human and dog have one amino acid less (371 residues). The difference is seen within the 9 glycine repeat at residue 234 through 242 in mouse and rat where only 8 repeats are observed in human and dog. The T/EBP amino acid sequence is extremely conserved among four species. Between mouse and rat, only 2 residues differ (99% similarity). Human and dog exhibit only 6 and 5 amino acid residues which differ from mouse (98% similarity). The position and pattern of amino acid substitutions seem to be also conserved among them. Especially of interest is that the rat sequence appears to be more similar to those of human and dog than mouse. This can be seen at amino acid residues 285 and 340. All four contain the

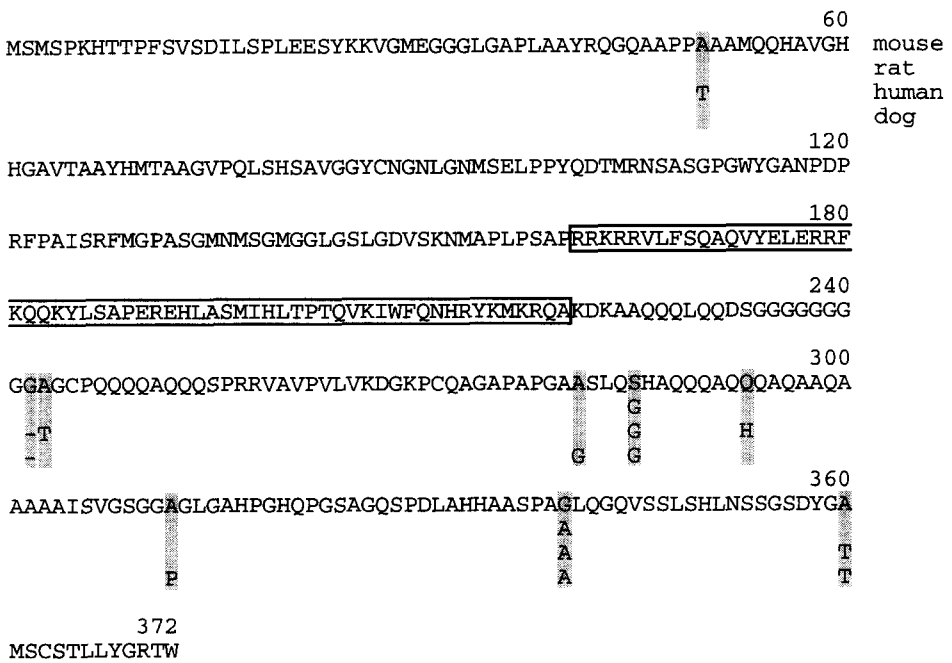

Fig. 2. Alignment of T/EBP amino acid sequences from rat, mouse, human, and dog. Complete mouse T/EBP amino acid sequence is presented and only amino acid residues in rat, human, and dog which are different from mouse are shown in shaded areas. One glycine residue within the glycine repeat that is not present in the human and dog sequences is also indicated by '–' in a shade (residue 242). The homeodomain sequence is boxed. Sequence alignment was carried out by using MacVector™ program (International Biotechnologies). Rat and dog T/EBP amino acid sequences are taken from Refs. [9] and [14], respectively.

identical homeodomain sequence. This level of homology throughout a protein, however, does not seem to be extraordinary for homeobox genes. For instance, Hox-D3 and D-10 protein sequences are 95% and 97% identical between mouse and human [15]. This high level of homology between different species further suggests an important role that T/EBP may have in maintenance and possibly development of these mammals.

We thank Dr. Frank Gonzalez for advice and discussion.

References

[1] Kikkawa, F., Gonzalez, F.J. and Kimura, S. (1990) Mol. Cell. Biol. 10, 6216–6224.
[2] Mizuno, K., Gonzalez, F.J. and Kimura, S. (1991) Mol. Cell. Biol. 11, 4927–4933.
[3] Civitareale, D., Lonigro, R., Sinclair, A. and Di Lauro, R. (1989) EMBO J. 8, 2537–2542.
[4] Abramowicz, M.J., Vassart, G. and Christophe, D. (1992) Eur. J. Biochem. 203, 467–473.
[5] Javaux, F., Bertaux, F., Donda, A., Francis-Lang, H., Vassart, G., Di Lauro, R. and Christophe, D. (1992) FEBS Lett. 300, 222–226.
[6] Francis-Lang, H., Price, M., Polycarpou-Schwarz, M. and Di Lauro, R. (1992) Mol. Cell. Biol. 12, 576–588.
[7] Civitareale, D., Castelli, M.P., Falasca, P. and Saiardi, A. (1993) Mol. Endocrinol. 7, 1589–1595.
[8] Shimura, H., Okajima, F., Ikuyama, S., Shimura, Y., Kimura, S., Saji, M. and Kohn, L.D. (1994) Mol. Endocrinol. 8, 1049–1069.
[9] Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.-G. and Di Lauro, R. (1990) EMBO J. 9, 3631–3639.
[10] Lazzaro, D., Price, M., De Felice, M. and Di Lauro, R. (1991) Development 113, 1093–1104.
[11] Bohinski, R.J., Di Lauro, R. and Whitsett, J.A. (1994) Mol. Cell. Biol. 14, 5671–5681.
[12] Endo, T., Ohta, K., Saito, T., Haraguchi, K., Nakazato, M., Kogai, T. and Onaya, T. (1994) Biochem. Biophys. Res. Commun. 204, 1358–1363.
[13] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–472.
[14] Van Renterghem, P.H.G., Vassart, G. and Christophe, D. (1994) EMBL/GenBank accession No. X77910, Seq ID 457488.
[15] Brown, W.M., Zhou, L. and Taylor. G.R. (1994) Biochim. Biophys. Acta 1219, 219–222.