Volume 301, number 3, 294–298 FEBS 10989 April 1992

Two different cDNAs encoding TFIID proteins of maize

Michael M. Haaß and Günter Feix

Institut für Biologie III, Albert-Ludwigs-Universität Freiburg, Schänzlestraße 1, D-7800 Freiburg i. Br., Germany

Received 18 March 1992

Two different complementary DNAs (cDNAs) encoding maize TFIID proteins were isolated from a maize leaf cDNA. Both cDNA sequences reveal two types of TFIID, each encoding an open reading frame of 200 amino acids. The two cDNAs are 76% identical at the DNA level and their putative amino acid sequences differ at only three amino acids. Like TATA box binding proteins from other organisms they show a bipartite structure containing a specific N-terminal region and a highly conserved C- terminal domain expected to be necessary and sufficient for the essential TFIID functions in transcriptional initiation.

Transcription initiation; TFIID; TATA box binding protein; *Zea mays*

1. INTRODUCTION

The general transcription factor IID (TFIID) or TATA box binding protein (TBP) is a key element in the assembly of functional transcription initiation complexes. By its specific binding to the TATA box, to other general transcription factors and to the RNA polymerase II, it assists in the assembly of the initiation complex at the transcription start site. Due to this important role in the establishment of pre-initiation complexes, TFIID has been considered to be the target of interaction with upstream regulatory proteins that modify the specificity or activity of transcription [1]. In the case of the transcriptional activators SP1 and CTF for example, co-activators that co-purify with the TFIID complex were proposed to be involved in the transcriptional regulation [2]. TFIID was first isolated and identified from human tissue culture cells as a fraction (fraction D) of a phosphocellulose column [3–5]. Previous findings suggest that TFIID is a multifunctional complex that includes co-activators and the TATA box binding protein [6,7]. Two different TFIID complexes, containing the 38 kDa hTBP could be identified in HeLa cells revealing different binding and transcription activating activities [8].

Since then cDNA clones of the TFIID have been isolated from yeast, mammals and *Drosophila* [9–21]. Amino acid sequence comparisons of the TFIIDs reveal a bipartite structure of the proteins with a carboxy-terminal region of 180 residues, which contains highly conserved structural motifs necessary and sufficient for the binding of TFIID to the TATA box and to recruit TFIIA and TFIIB to the promoter [15,20,22–24]. The divergent amino-terminal regions of the various TFIIDs share little sequence similarity, is sometimes rich in glutamine residues and is poorly characterized in its function [15,17,21].

More recently, cDNA clones encoding TFIID have been isolated from the dicotyledonous plant, *Arabidopsis thaliana*. In contrast to mammals, *Drosophila* and yeast where only one TFIID gene has been found, *Arabidopsis* contains two genes encoding slightly different forms of the TFIID. Both factors are able to confer basal transcription but exhibit a different binding pattern in electrophoretic mobility shift assays [25]. We report here the isolation and characterization of two different types of cDNA clones from a maize leaf cDNA library encoding TATA box binding proteins. The proteins derived from the longest open reading frames of the different cDNAs reveal the highly conserved structural motifs common to the C-terminal region of TFIID and display a N-terminal domain which is different from the TFIID clones of other organisms analysed so far.

Correspondence address: G. Feix, Institut für Biologie III, Albert-Ludwigs-Universität Freiburg, Schänzlestraße 1, D-7800 Freiburg, Germany. Fax: (49) (761) 203 2745.

2. MATERIALS AND METHODS

2.1. *Screening of a maize cDNA library for TFIID encoding cDNAs*

A maize leaf cDNA library in lambda ZAP (Stratagene) was screened with a 538 bp *Hin*dII/*Eco*RI fragment of the TFIID cDNA AT1 from *Arabidopsis thaliana* [25] by standard procedures [26]. Hybridization with the DIG-labeled probes (Boehringer) was performed at 42°C for at least 16 h in a buffer containing 5× SSC, 20% formamide, 0.1% sarcosyl, 0.02% SDS and 2% blocking powder (Boehringer) followed by washes at 65°C for 15 min each in 2× SSC, 0.1% SDS and 0.1 × SSC, 0.1% SDS. Hybridizing cDNAs were detected by the addition of an alkaline phosphatase-conjugated anti-DIG antiserum (Boehringer) followed by the BCIP/NBT colour reaction. Positive phage clones were purified by further rounds of screening.

Published by Elsevier Science Publishers B.V.

2.2. *Nucleotide sequence analysis*

Recombinant pBluescript plasmids were isolated from purified phage clones by in vivo excision with helper phage R408 as outlined by the manufacturer (Stratagene). After amplification of the plasmids, the cDNA inserts were subjected to restriction enzyme analysis. Nucleotide sequence analysis was carried out on an EMBL automated fluorescent DNA sequencer [27] according to a modified didesoxy method using fluorescent labeled primers [28]. The cDNA sequences were analysed using the PC/Gene program (IntelliGenetics and Genofit).

2.3. *Southern blot analysis*

20 μg of high molecular weigth total DNA, isolated from maize seedling leaves [29], were digested with restriction enzymes, separated in 0.8% agarose gels, and transferred to a ZetaProbe nylon membrane (Bio-Rad) by capillary blotting [26]. Hybridizations with two TFIID type-specific cDNA fragments labeled with ^{32}P by random priming were performed at 63°C in a buffer containing 4× SSPE, 1% SDS, 0.5% Blotto, 0.5 mg/ml salmon sperm DNA for at least 12 h followed by washes at 63°C for 15 min with 2× SSC, 0.1% SDS and 1× SSC, 0.1% SDS. The central *Pst*I fragment of the cDNA clone 13/1 and the C-terminal *Xho*I fragment of the clone 3/1 were used as gene specific probes (Fig. 1). Hybridizing bands were visualized by autoradiography on Fuji X-ray films.

3. RESULTS

3.1. *Isolation of two different maize TFIID cDNA clones*

Twenty-seven hybridizing plaques were isolated from the screening of a maize leaf cDNA library with the *Arabidopsis* TFIID cDNA probe and purified by rescreening. Restriction enzyme mapping of the cDNA inserts from the recombinant phages demonstrated the presence of two different cDNAs (Fig. 1). The two types, ZM TFIID-1 and ZM TFIID-2, were equally represented in the cDNA library. As yet only overlapping cDNA clones of the second TFIID type cDNA have been obtained. The size of the ZM TFIID-1 type cDNA is comparable to the RNA of 1,400 nucleotides length detected in Northern blot analysis (results not shown).

Fig. 1. Restriction enzyme maps and schematic alignment of three of the independently isolated cDNA clones revealing the two TFIID types in maize. The dotted boxes represent the coding regions. The positions of restriction sites are indicated.

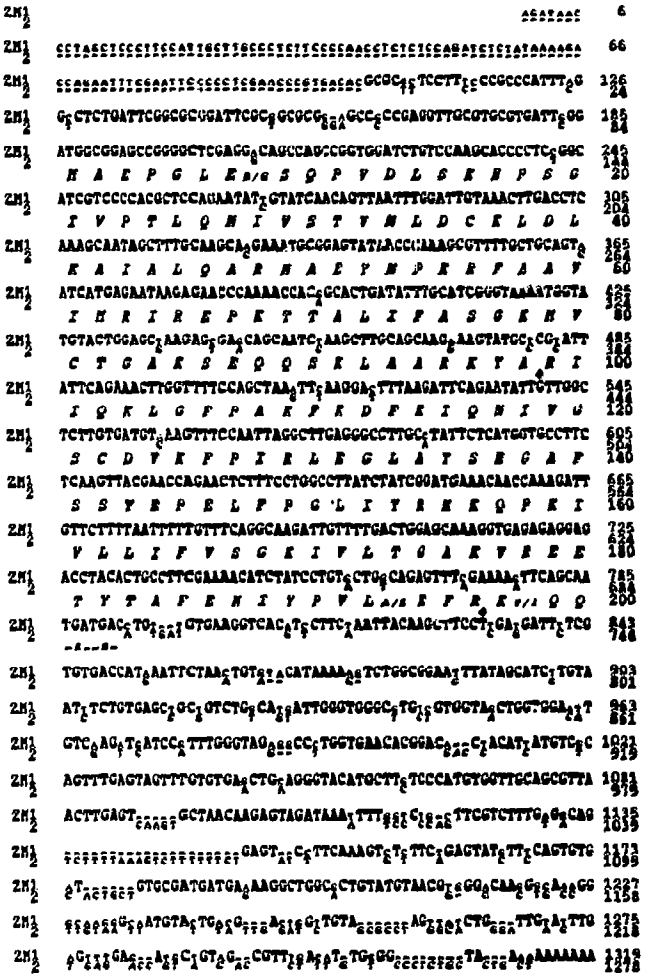

Fig. 2. Nucleotide sequence and deduced amino acid sequence (in single letter code in italics) of the cDNAs encoding two TFIID proteins of *Zea mays*. The termination codons are marked by asterisks. Differences between the two maize TFIID clones are indicated by small letters. The diamonds delineate the region of overlap between the cDNA clones 7/2 and 3/1 of the ZM TFIID-2 type.

3.2. *Comparative nucleotide sequence analysis of the two cDNAs*

The nucleotide sequences of the two TFIID cDNAs confirmed the differences seen in the restriction enzyme analysis (Fig. 2). The two maize TFIID cDNA clones display only 76% sequence identity at the DNA level over their whole length. However, while the cDNA coding regions share 96% identity, the 5′- and 3′-non-cod-

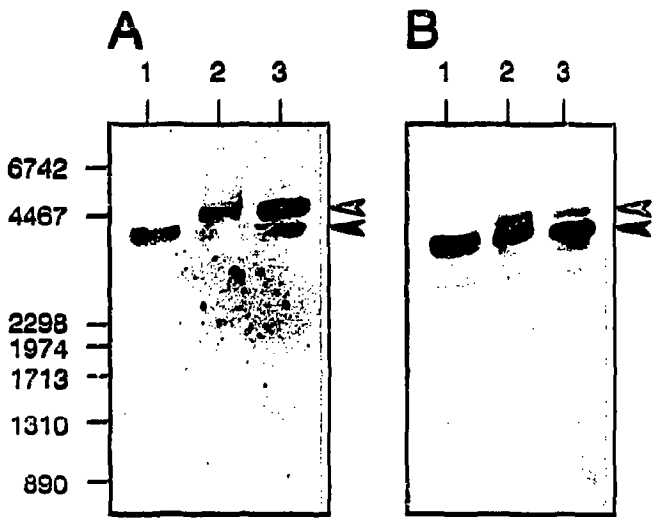

Fig. 3. Southern blot analysis with 20 μg of total maize DNA digested with *Hin*dIII, *Bam*HI or *Eco*RI (lanes 1–3, respectively). Hybridization was performed with different ^{32}P labeled TFIID cDNA fragments: (A) central *Pst*I fragment of the cDNA clone 13/1 and (B) C-terminal *Xho*I fragment of the clone 3/1. The arrows indicate specific hybridization bands in both Southern experiments (open arrows for TFIID-1, filled arrows for TFIID-2). The DNA markers on the left (*Hin*dIII-digested lambda-DNA) are indicated in bp.

ing regions are less than 60% identical. At the protein level the two cDNAs are 98% identical. Both cDNAs contain an open reading frame of 200 amino acids which code for proteins of relative molecular mass of 22,300. The sequences differ in only three amino acid positions, 8, 193 and 198, of which only the valine/isoleucine[198] difference represents a conservative amino acid change. While the sequences of the N-terminal 20 amino acids of both maize proteins differ from the amino acid sequences of the other known TFIID proteins, the remaining 180 amino acids show a strong sequence conservation.

3.3. *Southern analysis of the two TFIID genes*

The differences between the two maize TFIID cDNA sequences suggest the presence of at least two genes. This is supported by the results of Southern blot analysis of genomic DNA. Hybridizations with probes from the cDNA 13/1 clone (Fig. 3A) and the cDNA 3/1 clone (Fig. 3B) show clear differences in the banding pattern between the two groups suggesting that different genomic fragments contain the TFIID genes. Some bands are present in both Southern experiments and probably result from cross hybridizations between regions of homology in the two genes. The weaker hybridizing bands may also reflect other TFIID-related sequences in the maize genome. A further indication for the presence in the genome of different genes for the TFIID protein is obtained by performing polymerase chain reactions on genomic DNA with oligonucleotide primers common to both maize TFIID clones. This resulted in the synthesis of multiple fragments under stringent annealing conditions (data not shown).

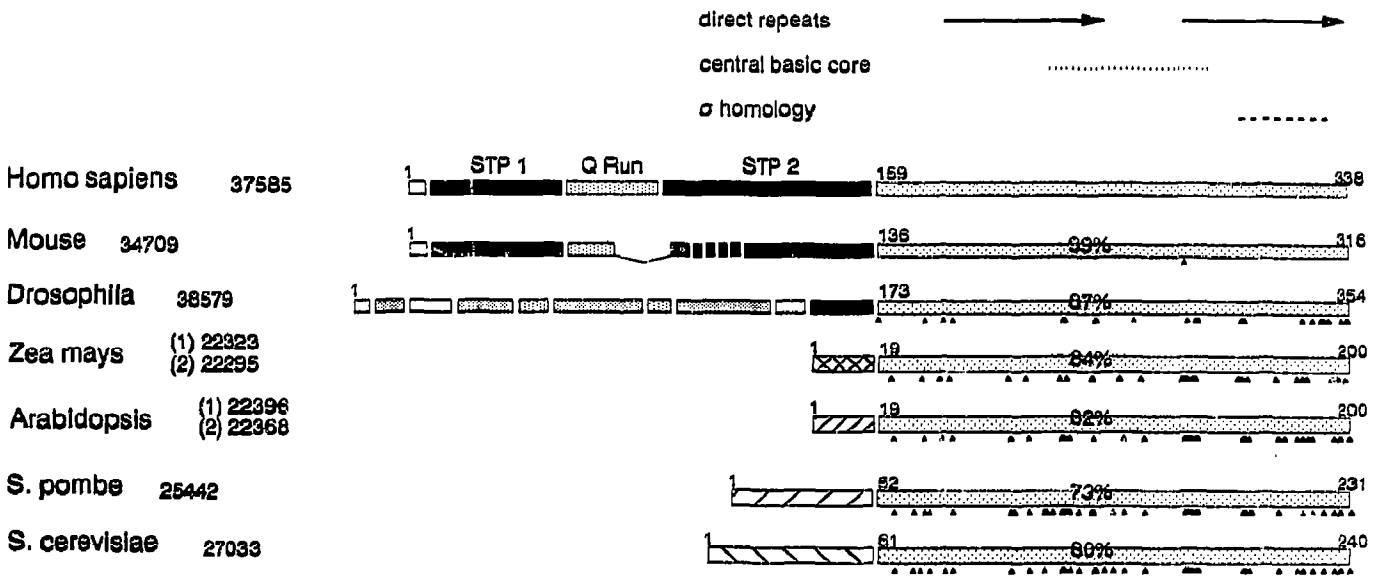

Fig. 4. Schematical representation of the TFIID primary structure showing the evolutionary conservation of the TATA box binding factor. The relative molecular weights of the different TFIID proteins are given following the species names. The numbered positions display the boundaries of the N- and C-terminal domains. The N-terminal regions of these proteins differ both in length and amino acid composition. The black and heavily dotted boxes represent sequences with stronger and weaker sequence similarities to, respectively the human STP stretches and Q-runs. The lightly dotted boxes represent the conserved C-terminal core. Structural motifs within the C-terminal region are outlined above. Small triangles under the C-terminal core indicate positions of amino acid differences from the human TFIID sequence and the degree of sequence homology (relative to human protein) within the conserved C-terminal domains of the TFIID's is summarized.

4. DISCUSSION

DNA binding proteins, such as TFIID, involved in establishing transcription initiation complexes have recently gained increasing importance. In addition to its role in guiding the transcription machinery to the initiation site, the TFIID protein is also involved apparently in mediating the reception of specific signals from upstream activator proteins by protein–protein interactions [1]. In order to interact directly with the TATA box region and with other proteins in active complex formation, the TFIID protein is composed of several domains which potentially have different functions. All of the TFIID genes characterised to date consist of a divergent N-terminal and a conserved C-terminal region of 180 amino acids (Fig. 4).

The C-terminal region contains highly conserved sequence motifs, namely the two direct repeats, a domain rich in basic residues and a region of similarity to prokaryotic sigma-factors. The high sequence conservation suggests that the C-terminal region of the maize TFIID protein is also involved in specific reactions like TATA box binding and interactions with the general transcriptional machinery, as has been shown for yeast and human TFIIDs [20,22–24]. It is interesting to note that most of the differences lie in the central basic core region that has the potential to form an alpha-helix and might be involved in protein–protein interactions [23,30].

The N-terminal region of the plant TFIID is shorter than its animal and yeast counterparts and does not display sequence motifs such as uninterrupted glutamine residues (Q-runs) or regions enriched in serine, threonine and proline (STP-stretches). It is furthermore of interest that the N-terminal part of the plant TFIID proteins analysed so far shows also distinct sequence differences to each other. Since it is assumed that the N-terminal domain is involved in mediating promoter regulation by gene- and cell-specific activators by protein–protein interactions [2,20], the identification of two cDNAs types with specific N-terminal domains in maize and *Arabidopsis* might reflect a special regulation mechanism in plants.

A major difference between plants and other eukaryotes is the presence of at least two TFIID genes in *Arabidopsis* and maize, while in animals and yeast only a single copy has been described. The two maize TFIID proteins differ in only three amino acid positions. The changes lie close to the N-terminal and C-terminal ends and two out of the three are non-conservative changes. The two *Arabidopsis* TFIID cDNAs differ in thirteen amino acid positions, the majority of which also lie mainly at the N- and C-terminal ends. Two of the *Arabidopsis* differences occur in identical positions to the two C-terminal differences in the maize cDNAs. Both *Arabidopsis* and maize TFIIDs contain the valine/isoleucine[198] difference but while the *Arabidopsis* clones show a serine/arginine[193] difference the maize clones contain either serine or alanine at this position. Since it appears that both dicotyledonous and monocotyledonous plants contain at least two copies of TFIID the possibility exists that in plants some of the functions of TFIID may be divided up between the two proteins.

Experiments are in progress towards an analysis of the functional relevance of the identified structural domains of the maize TFIID proteins and, in particular, in search for functional differences between the two types of maize TFIID proteins.

Acknowledgements: We would like to thank Mrs. Elke Brutzer for excellent technical assistance, Dr. G. Igloi of this department for DNA sequence determination and Dr. J.W.S. Brown of the University of Dundee, Scotland for critical reading of the manuscript. We thank Drs. A. Gasch and N.-H. Chua of the Rockefeller University, New York for generously supplying us with a TFIID clone of *Arabidopsis thaliana*. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie.

REFERENCES

[1] Greenblatt, J. (1991) Cell 66, 1067–1070.
[2] Pugh, B.F. and Tjian, R. (1990) Cell 61, 1187–1197.
[3] Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980) J. Biol. Chem. 255, 11992–11996.
[4] Samuels, M., Fire, A. and Sharp, P.A. (1982) J. Biol. Chem. 257, 114419–114427.
[5] Davison, B.L., Egly, J., Mulvihill, E.R. and Chambon, P. (1983) Nature 301, 680–686.
[6] Tanese, N., Pugh, B.F. and Tjian, R. (1991) Genes Dev. 5, 2212–2224.
[7] Pugh, B.F. and Tjian, R. (1992) JBC 267,679–682.
[8] Timmers, H.Th.M. and Sharp, P.A. (1991) Genes Dev. 5,1946–1956.
[9] Cavallini, B., Faus, I., Matthes, H., Chipoulet, J.M., Winsor, B., Egly, J. and Chambon, P. (1989) Proc. Natl. Acad. Sci. USA 86, 9803–9807.
[10] Eisenmann, D., Dollard, C. and Winston, F. (1989) Cell 58, 1183–1191.
[11] Hahn, S., Buratowski, S., Sharp, P.A. and Guarente, L. (1989) Cell 58, 1173–1181.
[12] Horikoshi, M., Wang, C.K., Fuji, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G. (1989) Nature 341, 299–301.
[13] Schmidt, M.C., Kao, C.C., Pei, R. and Berk, A.J. (1989) Proc. Natl. Acad. Sci. USA 86, 7785–7789.
[14] Fikes, J.D., Becker, D.M., Winston, F. and Guarente, L. (1990) Nature 346, 291–294.
[15] Hoey, T., Dynlacht, B.D., Peterson, M.G., Pugh, B.F. and Tjian, R. (1990) Cell 61, 1179–1186.
[16] Hoffmann, A., Horikoshi, M., Wang, C.K., Schroeder, S., Weil, P.A. and Roeder, R.G. (1990) Genes Dev. 4, 1141–1148.
[17] Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. and Roeder, R.G. (1990) Nature 346 , 387–390.
[18] Kao, C.C., Liebermann, P.M., Schmidt, M.C., Zhou, Q., Pei, R. and Berk, A.J. (1990) Science 248, 1646–1650.
[19] Muhich, M., Iida, C., Horikoshi, M., Roeder, R.G. and Parker, C. (1990) Proc. Natl. Acad. Sci. USA 87, 9148–9152.
[20] Peterson, M.G., Tanese, N., Pugh, B.F. and Tjian, R. (1990) Science 248, 1625–1630.
[21] Tamura, T., Sumita, K., Fujino, I., Aoyama, A., Horikoshi, M., Hoffmann, A., Roeder, R.G., Muramatsu, M. and Mikoshiba, K. (1991) Nucleic Acids Res. 19, 3861–3865.
[22] Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P.A. and Roeder R.G. (1990) Cell 61, 1171–1178.

[23] Buratowski, S. and Zhou, H. (1992) Science 255, 1130–1132.
[24] Strubin, M. and Struhl, K. (1992) Cell 68, 721–730
[25] Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R.G. and Chua, N.-H. (1990) Nature 346, 390–394.
[26] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[27] Ansorge, W., Sproat, B.S., Stegemann, J. and Schwager, C. (1986) J. Biochem. Biophys. Methods 13, 315–323.
[28] Voss, H., Schwager, C., Kristensen, T., Duthie, S., Olsson, A., Erfle, H., Stegemann, J., Zimmermann, J. and Ansorge, W. (1989) Methods Mol. Cell. Biol. 1, 155–159.
[29] Quayle, T.J.A., Brown, J.W.S. and Feix, G. (1989) Gene 80, 249–257.
[30] Horikoshi, N., Maguire, K., Kralli, A., Maldonado, E., Reinberg, D. and Weinmann, R. (1991) Proc. Natl. Acad. Sci. USA 88, 5124–5128.