Cloning and bioinformatic analysis of full-length novel pepper (*Capsicum annuum*) genes *TAF10* and *TAF13*

J.F. Wen[1]*, J.L. Huo[2]*, H.X. Chen[3]*, C.H. Ma[2], H. Jiang[3], H.S. Zhu[2], H. Zhou[2] and M.H. Deng[2]

[1]Kunming University of Science and Technology, Kunming, China
[2]Yunnan Agricultural University, Kunming, China
[3]College of Horticulture, Hunan Agricultural University, Changsha, China

*These authors contributed equally to this study.
Corresponding author: M.H. Deng
E-mail: dengminghua2013@163.com

ABSTRACT. We isolated two TATA-binding protein-associated factor (TAF) genes, *TAF10* and *TAF13*, from pepper (*Capsicum annuum*). The complete coding sequences were amplified using reverse transcriptase-PCR on the basis of conserved sequence information of eggplant and several other plant species. Nucleotide sequence analysis of these two genes revealed that the pepper *TAF10* gene encodes a protein of 103 amino acids that belongs to the *TAF10* superfamily. The pepper *TAF10* gene was highly expressed in the pericarp and placenta, moderately expressed in the stems, flowers, seeds and leaves, and weakly expressed in roots. The *TAF13* gene was found to encode a protein of 130 amino acids that belongs to the *TAF13* superfamily. The *TAF13* gene was highly expressed in the stems, flowers and pericarp, moderately expressed in the leaves, placenta and seeds, and weakly expressed in roots.

Key words: *Capsicum annuum* L.; Gene expression profile; *TAF10*; *TAF13*; Bioinformatic analysis

INTRODUCTION

The first step of gene expression is transcription, and this step can be strictly regulated to ensure appropriate levels of the gene product (Sawadogo and Sentenac, 1990; Orphanides et al., 1996). It is critical for a complete understanding of gene regulation to know which factors work together to achieve correct regulation of transcription (Woychik et al., 1990; Young, 1991). Pre-mRNA synthesis involves a series of transcription pre-initiation complex, which comprises the general transcription factors and RNAP II on the promoter. Pre-initiation complex includes the class II transcription factors (TFII) that include TFII-A, TFII-B, TFII-D, TFII-E, TFII-F, and TFII-H (Chiang and Roeder, 1995; Hampsey, 1998).

TFII-D plays a critical role in transcription initiation. TFI-D is a multiprotein complex composed of the TATA-binding protein (TBP) and a series of evolutionarily conserved TBP-associated factors (TAFIIs) (Samuels et al., 1982; Conaway et al., 1990; Tora, 2002). TBP specifically binds to the TATA element (Kao et al., 1990; Peterson et al., 1990).

TAFIIs are involved in promoter recognition and act as specific transcriptional coactivators *in vitro* and transfected mammalian cells (Pugh and Tjian, 1990, 1991; Sharp, 1992; Burley and Roeder, 1996).

In recent years, TAFs have been intensively studied. TAFs have been identified and analyzed in many organisms such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens*, and initial results have shown strong conservation (Tora, 2002).

Biochemical experiments and X-ray crystallography have revealed that a subset of TAFs contain the histone fold domain, a sequence motif homologous to a motif found on histones (Birck et al., 1998). The histone fold-containing TAFs form heterodimers, such as TAF6/9, TAF4/12, TAF8/10, TAF3/10, and TAF11/13 (Gangloff et al., 2001).

TAF10 is associated with a subset of TFIID complexes. Studies in human and mammalian cells have shown that this subunit is required for transcriptional activation by the estrogen receptor, for progression through the cell cycle, and may also be required for certain cell differentiation programs (Jacq et al., 1994; Collazo et al., 2005).

TAF13 is one of several TAFs that bind TBP and is involved in forming the TFIID complex (Meyers and Sharp, 1993).

In our study, two novel pepper genes, *TAF10* and *TAF13*, were isolated from pepper (*Capsicum annuum* L.), and their expression patterns were determined. This provides preliminary information for further understanding the biochemical functions of *TAF10* and *TAF13* in *C. annuum* L.

MATERIAL AND METHODS

Sample collection

All plants were obtained from the College of Horticulture and Landscape, Yunnan Agricultural University. Pepper tissues (root, stem, leaf, flower, pericarp, placenta, and seed) were instantly frozen in liquid nitrogen and stored at -80°C before use.

Total RNA extraction and first-strand cDNA synthesis

Total RNA from pepper was extracted with Trizol (TaKaRa, China), and cDNA was synthesized with the High Fidelity PrimeScript RT-PCR kit (TaKaRa, China) according to the manufacturer protocol.

PCR amplification

PCR was performed to isolate the pepper genes from pooled cDNAs from different tissues. Reactions were performed as described previously (Deng et al., 2012; Huo et al., 2012). The PCR program of the *TAF10* and *TAF13* genes consisted of 94°C for 3 min, followed by 34 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, and a final extension at 72°C for 10 min and 4°C hold.

The mRNA and amino acid sequences for *TAF10* and *TAF13* from various plant species archived at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) were used to locate conserved regions by multiple sequence alignment with CLUSTALW 1.8. The following primers were designed: TAF10-F: 5'-AGATGATGCTGCACTCGC-3', TAF10-R: 5'-TCTATCAACGGTGGAAAC-3'; TAF13-F: 5'-GGATACCCTAATCTTCGC-3', TAF13-R: 5'-GTCCTATTCTGCCGCTCA-3'.

Reverse transcriptase (RT)-PCR expression profile

RT-PCR primers for pepper *TAF10* and *TAF13* used for expression profile analysis were the same as those used for PCR. PCR was optimized to ensure sufficient product intensity within the linear phase of amplification.

Bioinformatics

PCR amplification was repeated 5 times. Amplification products were cloned into pMD18-T (TaKaRa) and bidirectionally sequenced. At least 10 independent clones were sequenced for each PCR product. Sequencing data were edited and aligned in DNASTAR (DNAStar Inc., USA). cDNA sequences were predicted with GenScan (http://genes.mit.edu/GENSCAN.html).

Putative protein theoretical molecular weight (Mw), isoelectric point (pI), amino acid composition, atomic composition, extinction coefficients, estimated half-life, instability index, aliphatic index, grand average of hydropathicity (GRAVY) were predicted with ProtParam (http://web.expasy.org/protparam/). Signal peptide, subcellular localization, membrane-spanning regions, and PFAM domains were identified with the SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP/), PSort (http://psort.hgc.jp/), TMpred (http://www.ch.embnet.org/software/TMPRED_form.html), and SMART (http://smart.embl-heidelberg.de). The BLASTP program and Conserved Domain Architecture Retrieval Tool were used to search for similar proteins and conserved domains (http://www.ncbi.nlm.nih.gov/Blast). Alignment of the nucleotide sequences and deduced amino acid sequences was performed with ClustalX, and phylogenetic trees were constructed in ClustalX and Mega 4.0 with standard parameters. Secondary structures of deduced amino acid sequences were predicted in SOPMA (http://

npsa-pbil.ibcp.fr/). The 3-D structures were predicted based on existing 3-D structures derived from amino acid homology modeling on Swiss Server (http://swissmodel.expasy.org/).

RESULTS

RT-PCR of pepper *TAF10* and *TAF13*

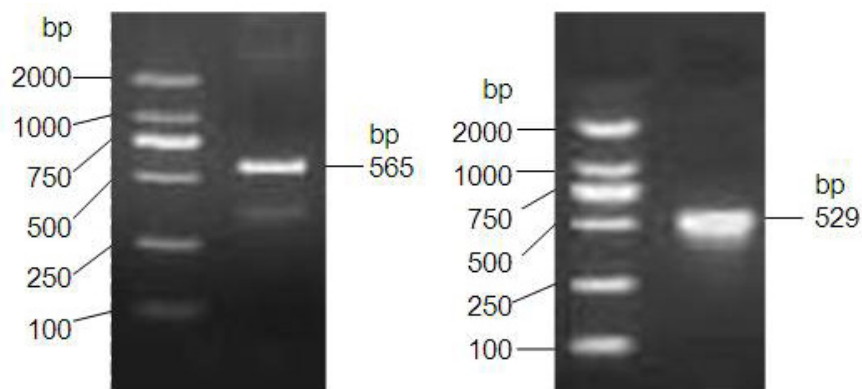RT-PCR with pooled tissue cDNAs for pepper *TAF10* and *TAF13* genes yielded products of 565 and 529 bp (Figure 1).

Figure 1. RT-PCR of pepper *TAF10* and *TAF13* genes. DL2000 DNA marker. Left = PCR product for *TAF10*; right = PCR product for *TAF13*.

cDNA nucleotide sequence analysis revealed that the genes are not homologous to any known pepper genes. Sequence prediction showed that the 565- and 529-bp cDNAs represent 2 genes that encode 103- and 130-amino acid proteins, respectively.

Protein properties

The complete coding sequences (CDS) and the encoded amino acids are presented in Figures 2 and 3.

Putative protein Mw, pI, amino acid composition, atomic composition, extinction coefficients, estimated half-life, instability index, aliphatic index, and GRAVY were computed.

pI of the putative proteins was 8.46 and 5.73, Mw was 11,709 and 14,605 Da, total number of negatively charged residues (Asp + Glu) was 15 and 22, and total number of positively charged residues (Arg + Lys) was 17 and 20, respectively. Formulas were $C_{510}H_{820}N_{150}O_{156}S_5$ and $C_{632}H_{1028}N_{180}O_{202}S_7$, total number of atoms was 1641 and 2049, extinction coefficient (assuming that all pairs of Cys residues form cystines) was 10,095 and 2980, extinction coefficient (assuming that all Cys residues are reduced) was 9970 and 2980, estimated half-life (mammalian reticulocytes, *in vitro*) was 30 h, instability index was 46.10 and 48.29, indicating unstable, aliphatic index was 77.67 and 75.77, and GRAVY was -0.621 and -0.617, respectively.

```
ATGAATTGGTGTGAGCATTACTTGGGTAAA
  M   N   W   C   E   H   Y   L   G   K
AGTGGATTTCAATGTCCCGACGTTAGATTA
  S   G   F   Q   C   P   D   V   R   L
ATAAGGTTGGTGGCAGTTGCTACACAGAAG
  I   R   L   V   A   V   A   T   Q   K
TTTATTGCAGATGTTGCCACTGATGCTCTC
  F   I   A   D   V   A   T   D   A   L
CAGCATTGCAAGGCAAGGCAGTCAGCTGTG
  Q   H   C   K   A   R   Q   S   A   V
GTCAAGGATAAAAGAGATAAGCAGCAAAAG
  V   K   D   K   R   D   K   Q   Q   K
GACAAACGCCTTATCTTGACGATGGATGAT
  D   K   R   L   I   L   T   M   D   D
CTTTCAAAATCTTTACGAGAGTATGGTGTA
  L   S   K   S   L   R   E   Y   G   V
AACGTCAAACATCAAGAATACTTTGCTGAT
  N   V   K   H   Q   E   Y   F   A   D
AGCCCATCTGCTGGGTTAGATCCTGCTTCA
  S   P   S   A   G   L   D   P   A   S
AGAGAGGAATAA
  R   E   E   .
```

Figure 2. Complete CDS and amino acid sequence of TAF10 in pepper. Asterisk indicates the stop codon.

```
atgaacaactcttctgcaggaccttcatct
  M   N   N   S   S   A   G   P   S   S
aaagctagagcaggagcatcacaaccttca
  K   A   R   A   G   A   S   Q   P   S
gaaagttcacttaaaaggaaacgtggcatg
  E   S   S   L   K   R   K   R   G   M
tttcagaaagacttgcaacacatgatgtat
  F   Q   K   D   L   Q   H   M   M   Y
ggttttggagatgatagcaatccacttcct
  G   F   G   D   D   S   N   P   L   P
gaaactgtcgcacttgtggaagacattgtg
  E   T   V   A   L   V   E   D   I   V
gtggattatgtcactgatatggtgcacaaa
  V   D   Y   V   T   D   M   V   H   K
gctcaggatgttgcaaccaaaagggggaag
  A   Q   D   V   A   T   K   R   G   K
cttttgacagaggattttctgttcttaatt
  L   L   T   E   D   F   L   F   L   I
agaaaggacctgccgaaacttaaccggtgt
  R   K   D   L   P   K   L   N   R   C
acagaactgttgtccatgaatgaggagttg
  T   E   L   L   S   M   N   E   E   L
aaacaagctcggaaggctttcgaggttgat
  K   Q   A   R   K   A   F   E   V   D
gaagagaaattggctcgtgaggggggatgtg
  E   E   K   L   A   R   E   G   D   V
tga
  .
```

Figure 3. Complete CDS and amino acid sequence of TAF13 in pepper. Asterisk indicates the stop codon.

The putative TAF10 and TAF13 proteins were also analyzed. The TAF10-conserved domains were identified as low-complexity regions (positions 52 to 63) and TAF10 super-family. TAF13-conserved domains were identified as coiled-coil region (positions 100 to 128) and TAF13 superfamily.

The results of secondary structure prediction showed that the deduced pepper TAF10 protein contained 43.7% alpha helices, 8.7% extended strands, 5.8% beta turns, and 41.7% random coils. Deduced pepper TAF13 protein contained 56.9% alpha helices, 5.4% extended strands, 2.3% beta turns, and 35.4% random coils.

The pepper *TAF10* and *TAF13* genes did not contain putative signal peptides (Bendtsen et al., 2004; Deng et al., 2009) and were not potential membrane proteins (Moller et al., 2001). Pepper TAF10 was probably located with 55% likelihood in the endoplasmic reticulum and TAF13 was probably located with 30% likelihood in the nucleus (Nakai and Horton, 1999).

Homology modeling

The 3-D structure of TAF13 (30 to 74) was based on template 1bh9A (2.60 A) with 51.11% sequence identity; no suitable templates were found to predict the 3-D structure of TAF10 (Guex and Peitsch, 1997; Schwede et al., 2003; Arnold et al., 2006; Yu et al., 2010). The 3-D structure analysis may provide a basis for understanding the relationship between structure and function of TAF13.

Sequence analysis and evolutionary relationships

The homology of the pepper *TAF10* gene was determined and analyzed. The results showed that the pepper TAF10 protein had high homology with this protein of four other plant species (Figure 4). *TAF13* shared homology with the TAF13 protein in the species shown in Figure 5.

Figure 4. Multiple amino acid sequence alignment of TAF10.

Figure 5. Multiple amino acid sequence alignment TAF13.

Phylogenetic trees were constructed on the basis of the alignments as shown in Figure 6. Pepper *TAF10* and *TAF13* genes were most closely related to *TAF10* and *TAF13* of eggplant.
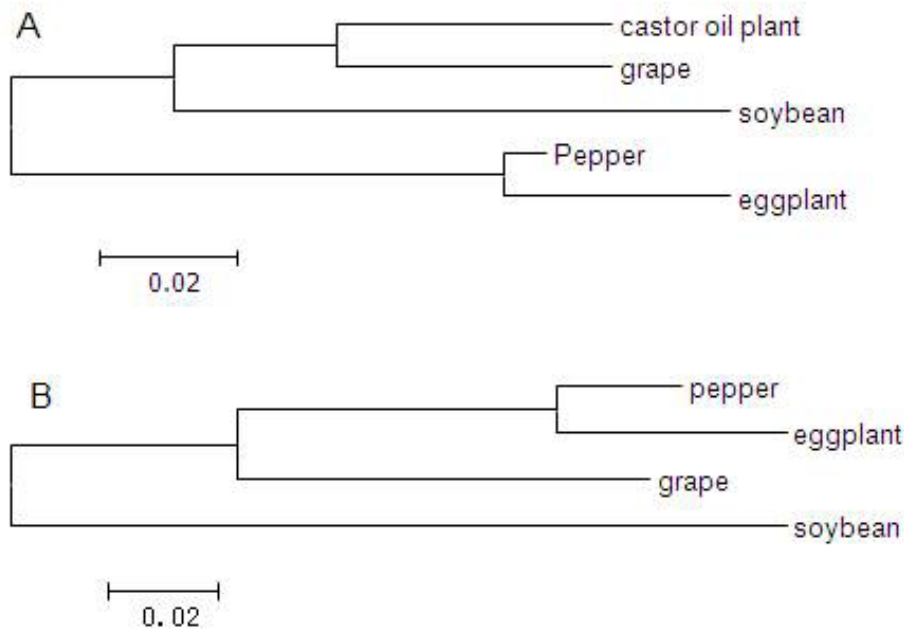

Figure 6. Phylogenetic tree of *TAF10* and *TAF13* genes. **A.** *TAF10*. **B.** *TAF13*.

Tissue expression

The pepper *TAF10* gene was highly expressed in the pericarp and placenta, moder-

ately expressed in the stems, flowers, seeds, and leaves, and weakly expressed in roots (Figure 7). The *TAF13* gene was highly expressed in the stems, flowers and pericarp, moderately expressed in the leaves, placenta and seeds and weakly expressed in roots.
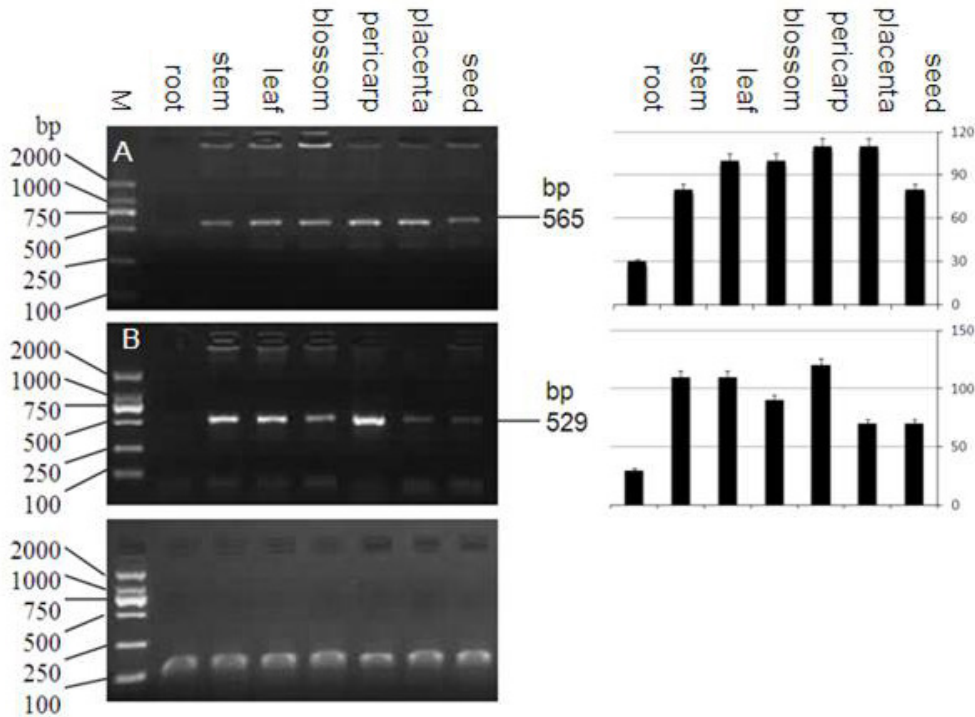

Figure 7. Tissue expression of pepper *TAF10* and *TAF13* genes. Actin served as the internal control.

DISCUSSION

Comparative genomics determines the relationship of genome structure and function of different species (Hardison, 2003). Several researchers have learned that TAF10 or TAF13 proteins from different species are highly conserved (Moqtaderi et al., 1996; Martinez et al., 1998; Mohan Jr. et al., 2003; Indra et al., 2005; Soutoglou et al., 2005). Therefore, we can use eggplant as a model organism to study the pepper *TAF10* and *TAF13* or use pepper as a model organism to study the tomato, potato or tobacco *TAF10* and *TAF13* genes.

In this study, pepper *TAF10* and *TAF13* genes were isolated and characterized. Many genes encoding *TAF10* and *TAF13* from several plant species have been isolated and characterized at the genetic, chemical, and enzymological levels (Moqtaderi et al., 1996; Birck et al., 1998; Martinez et al., 1998; Kouskouti et al., 2004; Indra et al., 2005; Soutoglou et al., 2005). However, cloning of the *TAF10* and *TAF13* genes from pepper has not been previously reported. This study verifies for the first time the presence of *TAF10* and *TAF13* in pepper.

The isolated *TAF10* cDNA was 312 bp long and encoded 103 amino acids. Our com-

parison of its amino acid sequence showed high homology to eggplant, soybean, grape, and castor-oil plant. The isolated *TAF13* cDNA was 393 bp long and encoded 130 amino acids. Sequence analysis revealed that the pepper *TAF13* has high homology with the proteins of 3 species - eggplant, grape and soybean. From the alignment analyses for pepper TAF10 and TAF13 proteins, we also found that pepper TAF10 and TAF13 proteins did not show complete identity with eggplant or other plants. This implied that pepper *TAF10* and *TAF13* genes have some different functions compared to eggplant and other plants.

In this study, we not only cloned the CDS sequences of the pepper *TAF10* and *TAF13* genes but also performed a sequence analysis and determined the tissue expression profile. The tissue expression profile analysis indicated that the genes were obviously differentially expressed in various tissues. Since we did not study functions at the protein level, we did not look into the many possible reasons for differential expression of these pepper genes. The suitable explanation for this under current conditions is that the biological activities associated with the functions of the genes were required to a different extent in different tissues at the same time.

In summary, we first isolated the pepper *TAF10* and *TAF13* genes and performed necessary functional analysis and tissue expression profile analysis. The cDNA clone, sequence information and functional analysis of the pepper *TAF10* and *TAF13* genes will be extremely important in elucidating the molecular mechanism of gene expression.

ACKNOWLEDGMENTS

REFERENCES

Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.

Bendtsen JD, Nielsen H, von Heijne G and Brunak S (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340: 783-795.

Birck C, Poch O, Romier C, Ruff M, et al. (1998). Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* 94: 239-249.

Burley SK and Roeder RG (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65: 769-799.

Chiang CM and Roeder RG (1995). Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science* 267: 531-536.

Collazo E, Couture JF, Bulfer S and Trievel RC (2005). A coupled fluorescent assay for histone methyltransferases. *Anal. Biochem.* 342: 86-92.

Conaway JW, Reines D and Conaway RC (1990). Transcription initiated by RNA polymerase II and purified transcription factors from liver. Cooperative action of transcription factors tau and epsilon in initial complex formation. *J. Biol. Chem.* 265: 7552-7558.

Deng M-H, Wen J-F, Huo J-L, Zhu H-S, et al. (2012). Relationship of metabolism of reactive oxygen species with cytoplasmic male sterility in pepper (*Capsicum annuum* L.). *Sci. Hortic.* 134: 232-236.

Deng W, Tan Y, Wang X, Xi D, et al. (2009). Molecular cloning, sequence characteristics, and polymorphism analyses of the tyrosinase-related protein 2 / DOPAchrome tautomerase gene of black-boned sheep (*Ovis aries*). *Genome* 52: 1001-1011.

Gangloff YG, Romier C, Thuault S, Werten S, et al. (2001). The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* 26: 250-257.

Guex N and Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.

Hampsey M (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* 62: 465-503.

Hardison RC (2003). Comparative genomics. *PloS Biol.* 1: e58.

Huo J, Wang P, Wei H and Zeng Y (2012). Molecular cloning, sequence characterization and tissue transcription profile of a novel gene PQBP1 from Chinese Banna Mini-pig inbred line (BMI). *J. Anim. Vet. Adv.* 11: 1674-1680.

Indra AK, Mohan WS, Frontini M, Scheer E, et al. (2005). *TAF10* is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. *Dev. Biol.* 285: 28-37.

Jacq X, Brou C, Lutz Y, Davidson I, et al. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* 79: 107-117.

Kao CC, Lieberman PM, Schmidt MC, Zhou Q, et al. (1990). Cloning of a transcriptionally active human TATA binding factor. *Science* 248: 1646-1650.

Kouskouti A, Scheer E, Staub A, Tora L, et al. (2004). Gene-specific modulation of *TAF10* function by SET9-mediated methylation. *Mol. Cell* 14: 175-182.

Martinez E, Kundu TK, Fu J and Roeder RG (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. *J. Biol. Chem.* 273: 23781-23785.

Meyers RE and Sharp PA (1993). TATA-binding protein and associated factors in polymerase II and polymerase III transcription. *Mol. Cell Biol.* 13: 7953-7960.

Mohan WS Jr, Scheer E, Wendling O, Metzger D, et al. (2003). *TAF10* (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. *Mol. Cell Biol.* 23: 4307-4318.

Moller S, Croning MD and Apweiler R (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17: 646-653.

Moqtaderi Z, Yale JD, Struhl K and Buratowski S (1996). Yeast homologues of higher eukaryotic TFIID subunits. *Proc. Natl. Acad. Sci. U. S. A.* 93: 14654-14658.

Nakai K and Horton P (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24: 34-36.

Orphanides G, Lagrange T and Reinberg D (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* 10: 2657-2683.

Peterson MG, Tanese N, Pugh BF and Tjian R (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. *Science* 248: 1625-1630.

Pugh BF and Tjian R (1990). Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 61: 1187-1197.

Pugh BF and Tjian R (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* 5: 1935-1945.

Samuels M, Fire A and Sharp PA (1982). Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J. Biol. Chem.* 257: 14419-14427.

Sawadogo M and Sentenac A (1990). RNA polymerase B (II) and general transcription factors. *Annu. Rev. Biochem.* 59: 711-754.

Schwede T, Kopp J, Guex N and Peitsch MC (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381-3385.

Sharp PA (1992). TATA-binding protein is a classless factor. *Cell* 68: 819-821.

Soutoglou E, Demeny MA, Scheer E, Fienga G, et al. (2005). The nuclear import of *TAF10* is regulated by one of its three histone fold domain-containing interaction partners. *Mol. Cell Biol.* 25: 4092-4104.

Tora L (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* 16: 673-675.

Woychik NA, Liao SM, Kolodziej PA and Young RA (1990). Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev.* 4: 313-323.

Young RA (1991). RNA polymerase II. *Annu. Rev. Biochem.* 60: 689-715.

Yu H, Chen S, Xi D, He Y, et al. (2010). Molecular cloning, sequence characterization and tissue transcription profile analyses of two novel genes: *LCK* and *CDK2* from the Black-boned sheep (*Ovis aries*). *Mol. Biol. Rep.* 37: 39-45.