Identification of a General Transcription Factor TFIIAα/β Homolog Selectively Expressed in Testis*

(Received for publication, February 9, 1999, and in revised form, March 22, 1999)

Ashok B. Upadhyaya, Sang Hyun Lee, and Jeff DeJong‡
From the Department of Molecular and Cell Biology, University of Texas at Dallas, Richardson, Texas 75083

In this paper we describe the isolation of a cDNA that encodes a human TFIIAα/β-like factor (ALF). The open reading frame of ALF predicts a protein of 478 amino acids that contains characteristic N- and C-terminal conserved domains separated by an internal nonconserved domain. In addition, a rare ALF-containing cDNA, which possesses an extended N terminus (Stoned B/TFIIAα/β-like factor; SALF) has also been identified. The results of Northern and dot blot analyses show that ALF is expressed almost exclusively in testis; in contrast, TFIIAα/β and TFIIAγ are enriched in testis but are also widely expressed in other human tissues. Recombinant ALF (69 kDa) and TFIIAγ (12 kDa) polypeptides produced in Escherichia coli form an ALF/γ complex that can stabilize TBP-TATA interactions in an electrophoretic mobility shift assay. The ALF/γ complex is also able to restore transcription from the adenovirus major late promoter using HeLa cell nuclear extracts that have been depleted of TFIIA. Overall, the data show that ALF is a functional homolog of human general transcription factor TFIIAα/β that may be uniquely important to testis biology.

The synthesis of accurately initiated messenger RNA in eukaryotic organisms requires the assembly of RNA polymerase II and the general transcription factors (TFIIA, B, D, E, F, and H) at core promoters (1, 2). Human TFIIA is composed of 35-kDa (N-terminal) and 12-kDa (C-terminal) subunits of the mature factor (3, 4) and hTFIIA α (At-1 and At-2) (39), Drosophila (TBP and TRF) (40), and humans (TBP and TLF) (41) and raise the idea that TBP homologs may contribute to tissue- and gene-specific regulation. For example, the Drosophila TRF protein is expressed in the central nervous system and gonads and is localized to a limited number of sites on polycatene chromosomes (40, 42). In this work we extend the idea of multiplicity within the general transcription factors to include human TFIIA. In particular, we identify a cDNA clone that encodes a novel human factor, ALF (TFIIAα/β-like factor), that is expressed almost exclusively in testis. In addition, ALF sequences are present as part of another cDNA, termed SALF (Stoned B/TFIIAα/β-like factor), that contains an N-terminal domain homologous to Drosophila Stoned B (43) and clathrin adapter proteins μ4 (AP47) and μ5 (AP50) (44, 45).

Previous studies have revealed the presence of two genes for TBP-like proteins in Arabidopsis (At-1 and At-2) (39), Drosophila (TBP and TRF) (40), and humans (TBP and TLF) (41) and raise the idea that TBP homologs may contribute to tissue- and gene-specific regulation. For example, the Drosophila TRF protein is expressed in the central nervous system and gonads and is localized to a limited number of sites on polycatene chromosomes (40, 42). In this work we extend the idea of multiplicity within the general transcription factors to include human TFIIA. In particular, we identify a cDNA clone that encodes a novel human factor, ALF (TFIIAα/β-like factor), that is expressed almost exclusively in testis. In addition, ALF sequences are present as part of another cDNA, termed SALF (Stoned B/TFIIAα/β-like factor), that contains an N-terminal domain homologous to Drosophila Stoned B (43) and clathrin adapter proteins μ4 (AP47) and μ5 (AP50) (44, 45).

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank‡ database (accession number(s) AF026169 (SALF) and AF108857 (ALF)).

‡ To whom correspondence should be addressed. Tel.: 972-883-6882; Fax: 972-883-2409; E-mail: dejong@utdallas.edu.

‡ The abbreviations used are: TF, transcription factor; TBP, TATA-binding protein; ORF, open reading frame; EST, expressed sequence tag; kb, kilobase pair(s); UTR, untranslated region; bp, base pair(s); PCR, polymerase chain reaction; AP, adaptor protein; AdML, adenovirus major late; ALF, TFIIAα/β-like factor; SALF, Stoned B/TFIIAα/β-like factor; NTA, nitrilotriacetic acid; TRF, TBP-related factor.
product was subcloned into the pCRII cloning vector (Invitrogen) to form pRACE4. The overlapping sequences of pRACE4 and EST ID259637 were combined to form the composite SALF sequence.

The 5′-end of the ALF cDNA was isolated by PCR (35 cycles) using 4 µl of the human ALF cDNA library (CLONTECH) with primer 2a–20 (5′-CCAGAAGTTGATTAGGGCTGCT-GATC-3′) and primer AP1 and reamplified with 2a–22 (5′-GGATGT-GAAGTGCCCAGGTCTGCTGTGG-3′) and primer AP2. The 369-bp amplification product was subcloned into pGEM-T Easy to form pRACE22. A full-length PCR product was amplified (35 cycles) from 4 µl of the testis library using primer 2a–17 (5′-GGCTGCTGCTATGCGCT-GACCCAC-3′), located within the unique 5′-end of the primer AP1. The resulting ~1.7-bp fragment was subcloned into pGEM-T Easy to form pRACE17. The overlapping sequences of pRACE22 and pRACE17 were combined to form the composite ALF sequence.

Northern and Dot Blot Analyses—Multiple tissue Northern blots containing 2 µg of poly(A) mRNA from 16 human tissues and a dot blot containing 89–514 ng of poly(A) mRNA from 50 adult and fetal tissues were obtained from CLONTECH. Gene-specific probes for hybridization experiments were as follows: ALF, a 621-bp NcoI-R2 fragment or an 899-bp HindII-BglII fragment from region II (see Fig. 1B); 5′-SALF, a 1002-bp EcoRI-EcoRI fragment from pRACE4 containing the 5′-UTR and nucleotides encoding the first 282 residues (see Fig. 1B); hTFIIA-α, a 11-kb EcoRI-EcoRI fragment from a11 (3) or a 282-bp HindIII-HaeIII fragment from region II; hTFIIAβ, a full-length 355-bp NdeI-BamHI fragment or a 262-bp NdeI-EcoRI fragment from pRSETpl2 (12); and actin and ubiquitin controls provided by CLONTECH. DNA fragments were typically labeled with [α-32P]dCTP using Ready-To-Go DNA Labeling Beads (Amersham Pharmacia Biotech). Northern blots were hybridized for 1 h in ExpressHyb solution (CLONTECH) and washed at 68 °C for 1 h. Membranes were typically exposed for 1–2 days to either XAR-5 film (Kodak) or a PhosphorImager screen (Molecular Dynamics). The results in the left hand column in Fig. 5, C and D, were obtained with different blots; other results within each column were obtained by reusing a single blot. The actin control in the left hand column is a representative result.

For experiments using the dot blot, labeled DNA was combined with 30 µg of Cot-1 DNA (Roche Molecular Biochemicals) and 100 µg of salmon sperm DNA, denatured, and allowed to reanneal in 200 µl of 5× SSC at 68 °C for 30 min prior to the addition. After hybridization in 5 ml of ExpressHyb solution at 65 °C overnight, the blot was washed in 0.1× SSC at 55 °C. Membranes were exposed as follows: Fig. 6A, 19 h; 6B, 2 h 45 min; 6C, 14 h; 6D, 25 h, and 6E, 30 min. For reprobing, Northern blots were stripped twice with 0.5% SDS, 0.2 M NaH2PO4, pH 8.0, and 0.01 M EDTA, and sonicated five times at 30 s. The denatured cell lysate (~20 ml) was then incubated with 2 ml of nickel-NTA-agarose resin (Qiagen) at room temperature for 1 h. The resin was washed successively with Buffer A containing 8 M urea at pH 8.0, 6.3, and 5.9, and bound polypeptides were eluted at pH 3.5. Preparation of expression constructs for rat TFIIAαβ and rat TFIIAγ subunits (GenBank accession numbers AF000943 and AF000944, respectively) and purification of the corresponding 55- and 12-kDa recombinant polypeptides were performed similarly. For transcription experiments the recombinant p69 and p12 polypeptides were covalized to prevent precipitation of the p12 subunit.

Human TBP was expressed from pET11d (Novagen), induced with 2 mM isopropyl-b-D-thiogalactopyranoside at A260 0.3, and purified at 4 °C from the soluble fraction of the bacterial lysate over nickel-NTA-agarose. Purification was performed by washing the resin with D700 buffer (20 mM HEPES, 20% glycerol, 0.2 mM EDTA, 10 mM b-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride, and 700 mM KCl) that contained 5, 10, and 15 mM imidazole and eluting bound polypeptides with D700 buffer that contained 100 mM imidazole. Recombinant proteins were dialyzed against Buffer C (10 mM Tris, pH 7.9, 2 mM dithiothreitol, 20% glycerol, and 0.5 mM phenylmethylsulfonyl fluoride) containing 100 mM KCl prior to use.

To express SALF, primers NN1 (5′-TACTGCTGCGACCATTTTGA-3′) and 2a–2–8 were used to generate a 2988-bp product from pRACE4. An internal 2207-bp Xhol-BglII fragment (amino acids 1–718) derived from this PCR product was then inserted into the Xhol-BglII-digested pT7T3ID vector that contained EST ID259637. Because an NdeI-BglII fragment from pRACE4 containing the 5′-end of the p12 subunit was excised during the preparation of this vector, this fragment was later reinserted in the appropriate orientation to create a full-length SALF open reading frame (ORF) (pT7T3-SALF). This construct (0.8 µg) was used to program rabbit reticulocyte lysates in the presence of [35S]methionine as described by the manufacturer (Promega). Labeled polypeptides were separated on 8% SDS-polyacrylamide gels. Mobility Shift Assays—Mobility shift assays were performed using 10 fmoles of a[32P]ATP kinase-labeled TATA-containing oligonucleotide (5′-GGCTGCTGCTATGCGCTGACCCAC-3′) and a 29-mer spanning the initiation site just beyond the G-free cassette prior to use. Reactions (20 µl) of 25 final volume were performed in 10 mM HEPES (pH 7.9), 2% polyethylene glycol-8000 (w/v), 60 mM KCl, 5 mM dithiothreitol, 0.2 mM EDTA, 5 mM ammonium sulfate, 4 mM MgCl2, and 8% glycerol essentially as described (3, 47). Recombinant rat p55 (30 ng, 29 nm), rat p12 (1.1 µg, 3.5 µm), human ALF (180 ng, 137 nm), and human TBP (125 ng, 133 nm) were added to reactions as indicated. Reactions were incubated for 30 min at room temperature, and complexes were separated on native 5% polyacrylamide gels containing 0.5% TBE and 50% urea. Labeled polypeptides contained either cold AdML TATA or SP1 (5′-TGTCGCGCGGCCGGCCAG-3′) oligonucleotides, and antibody supershift reactions contained 2–4 µl of rabbit polyclonal antiserum raised against the 55-kDa hTFIIA/β polypeptide (3).

In Vivo Transcription—HeLa cell nuclear extracts were depleted of TFIIA essentially as described (3–5, 7). In brief, 200 µl of extract was incubated with 100 µl of nickel-NTA-agarose resin for 30 min at 4 °C in the presence of 400 mM KCl. Control extracts were processed similarly, except that no nickel-NTA-agarose was present. After microcentrifugation for 5 min, supernatants were removed and dialyzed for 3 h against Buffer C that contained 100 mM KCl. Transcription reactions were performed using a template (pMLC_AT) that contains the AdML promoter upstream of a G-free cassette (48). The template was linearized at a Smal site just beyond the G-free cassette prior to use. Reactions (20 µl) contained 8 µl of nuclear extract (~60 µg protein), 2 µl (550 ng) of recombinant p69 (0.22 µm) and p12 (0.9 µm) proteins, 1 µg of pMLC_AT, 10 mM HEPES (pH 7.5), 25 mM KCl, 6 mM MgCl2, 625 µM UTP, 625 µM ATP, 35 µM CTP, 200 µM O-methyl-GTP, 3% glycerol, 0.7 µl of stop solution (8 M urea and 37.3 units of RNaseA (Amersham Pharmacia Biotech). After incubation at 30 °C for 45 min, reactions were terminated by adding 270 µl of stop solution (0.25 µl Nac), 1% SDS, 20 mM Tris, pH 7.5, 5 mM EDTA, and 66.7 µg/ml RNA) and extracted with an equal volume of phenol/chloroform (1:1). Ethanol-precipitated transcripts were resuspended in formamidine-containing loading dye and electrophoresed on 5% acrylamide gels containing 1× TBE and 8% urea.

Other Procedures—Sequencing reactions were performed at The University of Texas Southwestern Medical Center. Sequence comparisons and alignments were done with Lasergene (DNASTAR). Custom oligonucleotides were obtained from Operon.
RESULTS

Isolation of SALF—Computer searches of expressed sequence tags (dbEST) with a human TFIIAβ-like query identified a homologous placental cDNA sequence (I.M.A.G.E. Consortium Clone ID 259637). Sequence analysis showed that this 1885-bp clone (Fig. 1A) encodes 471 amino acids similar in sequence and organization to TFIIAβ (gray area) domains. The locations of primers 2a2–1, 2a2–6, and 2a2–8 (black boxes) and the corresponding PCR products (single lines) are indicated. The bold line indicates UTR sequences. B, the composite SALF cDNA includes TFIIAβ-like sequences and an upstream Stoned B-like region and is shown with selected restriction enzyme sites. The 5′-end of SALF was isolated in PCR reactions using gene-specific primers 2a2–1 and adapter primers AP1 and AP2. The resulting clone (pRACE22) is shown as a single line that is interrupted by a 141-bp internal deletion. C, the composite ALF sequence consists of TFIIAβ-like sequences only. The 5′-end of ALF was identified using gene-specific primers 2a2–6 and 2a2–8 and library-specific primers AP1 and AP2. The resulting clone (pRACE4) is shown as a single line that is interrupted by a 141-bp internal deletion. D, a PCR product that spans the entire ALF sequence (pRACE17) was obtained using the gene-specific primer 2a2–17 and the library-specific primer AP1. D, PCR products that contain the junction between the Stoned B- and TFIIAβ-like domains of SALF were amplified from human placenta and liver cDNA libraries. The primers used are indicated above each lane (2a2–1, 2a2–6, and 2a2–8). Lanes 5 and 6 are control reactions to which no cDNA template was added. M, marker.

To determine whether this EST reflects the actual structure of a corresponding human mRNA, we performed PCR analysis. Primers within the TFIIAβ-like sequence (2a2–6 and 2a2–8) and the upstream Stoned B-like sequence (2a2–1) amplified products of expected size (1.1- and 0.9-kb) from both human placenta and liver cDNA libraries (Fig. 1, A and D, lanes 1–4). Similar results were obtained using a human testis cDNA library (data not shown). Analysis of the liver-derived products showed that they contain sequences identical to those present in EST ID259637, confirming that the EST is a real, albeit incomplete, human cDNA clone.

To isolate the full-length 5′-end of this species, we performed PCR reactions on a human placental cDNA library using gene-specific primers within the TFIIAβ-like sequence (2a2–6 and 2a2–8) and adapter primers AP1 and AP2. The resulting PCR product (pRACE4) (Fig. 1B) contains 3′-sequences that are identical to EST ID259637, except for a 141-bp deletion that removes 47 amino acids (787–833) within region II of the TFIIAβ-like domain. pRACE4 also contains a 1968-nucleotide 5′-end that extends the region of Stoned B homology by 344 amino acids and continues for 274 amino acids further upstream.

The results of these experiments reveal a novel human mRNA, SALF, that is composed of both Stoned B/clathrin AP-like and TFIIAβ-like sequences. The 3853-bp composite sequence contains a 114-nucleotide 5′-UTR and a 161-nucleotide 3′-UTR with a poly(A) addition signal and a 29-nucleotide poly(A) tract (Fig. 1B). The deduced ORF commences with a putative start codon (AUG) that is preceded by an in-frame stop codon 27 nucleotides upstream and predicts a 1182-residue polypeptide (Fig. 2) with a molecular mass of 132 kDa and a pI of 5.1.

Isolation of ALF—Northern blot analysis of human mRNA using a probe from the TFIIAβ-like region of SALF revealed a 1.8-kb mRNA in testis (see Fig. 5), indicating that this sequence is also present in an mRNA that is substantially shorter than that predicted by SALF. To isolate the 5′-end of the corresponding cDNA, we performed PCR on a human testis cDNA library using gene-specific primers 2a2–20 and 2a2–22 and adapter primers AP1 and AP2. The resulting clone (pRACE22) (Fig. 1C) contains 298 bp that are identical to SALF and a 35-bp 5′-end that is unique. This new 5′-end consists of
a 15-bp UTR, a putative initiation codon (GTCATGG) that conforms to the Kozak consensus ((A/G)NNATGG) (51), and 17 bp downstream of the ATG that predict six amino acids (ACLNPV) not present in SALF. These residues include the conserved valine found at the N terminus of all other TFIIA large subunits (Fig. 3B). To isolate a complete cDNA, we then performed PCR with the gene-specific primer 2a2–17 and primer AP1. The sequence of the resulting clone (pRACE17) (Fig. 1C) is identical to the composite SALF sequence except for its unique 5' end and a longer poly(A) tail (90 nucleotides), which begins four nucleotides downstream of the poly(A) tail in SALF. Together, pRACE22 and pRACE17 form a 1617-bp composite cDNA, ALF, that predicts a 478-amino acid polypeptide (Fig. 2) with a molecular mass of 52 kDa and a pI of 4.4.

Two ESTs that contain partial ALF sequences connected at nucleotide 1344 to an alternative 261-bp 3' end were also identified (I.M.A.G.E. Consortium Clone ID 785133 and 1657721). These clones both predict a C terminus in which the last 35 amino acids of ALF are replaced with the residues "AFPRRTSFNT" followed by a stop codon and a 3' UTR that contains a poly(A) addition signal and a poly(A) tail. PCR analysis has verified that both ALF and SALF cDNAs that contain this alternative 3' end can be amplified from human cDNA libraries (data not shown). However, because the conserved C terminus is required for TFIIA activity, it is not clear whether these clones will encode functional polypeptides.
Description of the Predicted ORFs—A schematic comparison of ALF and other TFIIA large subunit sequences from humans (3, 4), Drosophila (9), Arabidopsis (GenBank Accession number X98861), and yeast (8) is shown in Fig. 3A. These sequences share a common organization consisting of conserved regions I and IV, acidic region III, and an internal nonconserved region II. ALF is similar to its human TFIIAα/β counterpart in region I (amino acids 1–54, 67%), and region IV (amino acids 417–478, 73%), as shown in Fig. 3B, and in the negatively charged region III (amino acids −340–414, 42% D/E residues). In contrast, region II shares no homology with the corresponding region in hTFIIAα/β (or other TFIIA large subunits) and is approximately 100 residues longer. The importance of the conserved regions I, III, and IV for TFIIA structure and function has been demonstrated by mutagenesis (52) and by X-ray crystallographic studies (11, 12). The presence of similar domains in ALF (and in SALF) suggests that these sequences encode functional TFIIA large subunits.

The unique N terminus of SALF is 711 amino acids in length (Fig. 3C) and contains a region between amino acids 44 and 150 that is rich in proline (20%), serine (21%), and threonine (9%) residues. Residues between 275 and 692 display 47% similarity to the Drosophila Stoned B protein (43) and 46% similarity to an uncharacterized Stoned B-like ORF in Caenorhabditis elegans, C27H6.1 (53) (Fig. 3D). The Drosophila stoned locus was first identified as a class of mutations that caused neurological defects such as temperature-sensitive paralysis (54), and it has been suggested that Stoned B functions in membrane trafficking in neurons (43). In addition, residues from 410 to 692 within the Stoned B homology region are 33 and 37% similar to the mouse μ1 (AP47) and rat μ2 (AP50) clathrin APs, respectively (Fig. 3D) (44, 45). The μ1 (AP47) and μ2 (AP50) clathrin APs are subunits of the AP-1 and AP-2 complexes associated with the trans-Golgi and plasma membranes, respectively, and function in the internalization, sorting, and recycling of receptors and other membrane proteins (49, 50). Thus, the N terminus of SALF is related to a family of proteins involved in membrane trafficking.

Genomic DNA Blot Analysis—To confirm that the ALF sequences were derived from a distinct human gene, we performed genomic DNA hybridization (Fig. 4). Hybridization with an ALF probe revealed bands of 8.6, 6.9, 5.0, and 1.0 kb (Bgl II, lane 1) or 11.5, 8.4, 6.0, and 4.5 kb (EcoRI, lane 2). In contrast, hybridization with a TFIIAα/β probe revealed bands of 15, 8.0, and 4.2 kb (Bgl II, lane 3) or 12.0, 8.0, 5.4, and 4.2 kb (EcoRI, lane 4). Hybridization with a TFIIAγ probe showed bands of 4.2 and 3.8 kb (Bgl II, lane 5) or 10.0, 8.0, 6.6, 2.5, and 1.7 kb (EcoRI, lane 6). The presence of multiple bands in some lanes (e.g., lane 6) is due in part to the presence of restriction enzyme sites in the corresponding cDNAs. These results show that sequences complimentary to ALF and TFIIAα/β are present on different sized human genomic DNA fragments and indicate that the corresponding mRNAs are encoded by separate genes.

Northern Blot Analysis—To evaluate the abundance and distribution of the ALF and SALF mRNAs, we performed Northern blot analysis of mRNA from 16 different human tissues. Hybridization with a probe from the TFIIAα/β-like region of SALF revealed a 1.8-kb mRNA that was present in testis but not in other tissues (Fig. 5A, lane 12). The isolation of the ALF cDNA that corresponds to this species is illustrated in Fig. 1C. Surprisingly, the predicted 3.8-kb SALF mRNA was not visible in mRNA from any of the tissues examined, including placenta, liver, and testis from which SALF can be amplified by PCR. These results indicate that ALF and TFIIAα/β are the major transcripts encoding human TFIIA large subunits and that SALF is relatively rare.

Hybridization with a probe specific for the 5′-end of SALF (5′-SALF) revealed a 6.5-kb species that was present at the highest levels in heart, placenta, kidney, prostate, and uterus (Fig. 5B, lanes 1, 3, 7, 11, and 13) and at lower levels in other tissues. This transcript, termed RNA6.5, was not detected using the ALF-specific probe (Fig. 5A), indicating that it does not contain a downstream ALF domain. Thus, RNA6.5 is an independent human transcript that contains sequences similar or identical to those present at the 5′-end of SALF.

Together with the cloning data, these experiments reveal the existence of three related mRNA transcripts: 1) a 1.8-kb testis-specific ALF mRNA encoding a human TFIIAα/β homolog, 2) a ubiquitously expressed 6.5-kb mRNA (RNA6.5) whose exact function is unknown but may encode a protein involved in membrane trafficking, and 3) a rare ~3.8-kb SALF mRNA that contains sequences that are present in and possibly derived from the more abundant ALF and RNA6.5 transcripts (see “Discussion”).

Northern experiments were also performed using probes for human TFIIAα/β and TFIIAγ. Hybridization with TFIIAα/β revealed transcripts at 6.5 and 7.0 kb, as described earlier (4). These species were present alone or together in all tissues and were enriched in placenta (10-fold), skeletal muscle (9-fold), and testis (13-fold) (Fig. 5C, lanes 3, 6, and 12). Hybridization with a TFIIAγ-specific probe revealed a 1.0-kb transcript that was present in all tissues and was enriched 32-fold in testis (Fig. 5D, lane 12). In contrast to the restricted expression of ALF in human tissues, the widespread distribution of the TFIIAα/β and TFIIAγ mRNAs is consistent with their role as generally required basal transcription factors.

RNA Dot Blot Analysis—We also examined mRNA from 50 human adult and fetal tissues for the presence of ALF, SALF, and RNA6.5. Using an ALF-specific probe, we observed a strong signal in testis that is due to the 1.8-kb ALF transcript (Fig. 6A, position D1). In addition, weak signals were observed in approximately 24 of the remaining tissues, including the small intestine, bladder, uterus, and prostate (positions E3, C5, C6, and C7). These weak signals indicate that ALF or SALF is
expressed at low levels in non-testis tissues, and their detection in this experiment likely reflects the greater sensitivity of the dot blot. When this blot was stripped and reprobed with the 5'-SALF probe, signals were detected in all tissues. A short exposure (shown in Fig. 6B) showed highest levels in placenta, uterus, spinal cord, fetal kidney (positions F4, C6, B7, and G3), and several others and lower levels in the remaining tissues. Because this probe detected high levels of RNA6.5 (but not SALF) in Northern blot analysis, we believe the signals in Fig. 6B are primarily due to expression of RNA6.5.

Further inspection of the data reveals that the signals detected with the ALF-specific probe in Fig. 6A are present in a range of tissues that are nearly identical to those observed in Fig. 6B. This holds true for approximately 20 tissues, including bladder, uterus, prostate, ovary, placenta (positions C5, C6, C7, D2, and F4), and others but not for testis (position D1). Likewise, the absence of signals in Fig. 6A correlates with the absence of signals in Fig. 6B. It is unlikely that this similarity is because of nonspecific hybridization or incomplete removal of the earlier probe because: 1) the signals in Fig. 6B are much stronger than background, and 2) the experiment in Fig. 6B was performed after stripping the relatively faint signals in Fig. 6A and not vice versa. Although the molecular basis for this observation is not yet clear, the results suggest a relationship between the expression of RNA6.5 with ALF-containing transcripts (possibly SALF) present at low levels in non-testis tissues.

Hybridization with human TFIIAα/β- and TFIIAγ-specific probes (Fig. 6, C and D) shows that the corresponding mRNAs are expressed in all tissues. Quantitation of the results confirms that ALF (50-fold), TFIIAα/β (4-fold), and TFIIAγ (10-fold) are enriched in testis tissue (see "Discussion").

Expression and Reconstitution of ALF/γ Complexes—To prepare recombinant ALF protein for functional assays, we overexpressed a 479-amino acid histidine-tagged polypeptide that spans residues Val7 to Trp 478. The predicted size of this polypeptide is 53 kDa, but the mobility on SDS-polyacrylamide gel electrophoresis is 69 kDa (Fig. 7A, lane 2). The predicted size of this polypeptide is 53 kDa, but the mobility on SDS-polyacrylamide gel electrophoresis is 69 kDa (Fig. 7A, lane 2). The predicted size of this polypeptide is 53 kDa, but the mobility on SDS-polyacrylamide gel electrophoresis is 69 kDa (Fig. 7A, lane 2). The predicted size of this polypeptide is 53 kDa, but the mobility on SDS-polyacrylamide gel electrophoresis is 69 kDa (Fig. 7A, lane 2). The predicted size of this polypeptide is 53 kDa, but the mobility on SDS-polyacrylamide gel electrophoresis is 69 kDa (Fig. 7A, lane 2).
counterparts (GenBank Accession numbers AF000943 and AF000944).

We then tested the activity of these polypeptides in electrophoretic mobility shift assays using human TBP. Under the conditions of this assay, TBP alone is unable to bind DNA (Fig. 7B, lane 1). However, the presence of TFIIAα/β (p55) and TFIIAγ (p12) stabilizes the TBP-DNA interaction via TFIIAα/βγ-TBP-DNA complex formation (lane 2). Likewise, the recombinant ALF (p69) polypeptide, in conjunction with the TFIIAγ (p12) subunit, is able to form an ALFγ-TBP-DNA complex (lane 5). Formation of this complex depends on the presence of both ALF and TFIIAγ (data not shown). Interestingly, although ALF (p69) is 102 amino acids longer than TFIIAα/β (p55) and migrates as a larger species in SDS-polyacrylamide gel electrophoresis, the ALFγ-TBP-DNA complex migrates slightly faster than the TFIIAα/βγ-TBP-DNA complex (lanes 2 and 5). These reactions were run side-by-side on the same gel using ALF and TFIIA subunits that had been purified and renatured using the same procedure. In these experiments approximately 5-fold higher ALF concentrations were used to generate shifted complexes equivalent to that observed with TFIIAα/β. This fact may reflect differences in the activities of these two factors or, alternatively, differences in their ability to be purified and renatured in active form. The specificity of the ALFγ-TBP-DNA complex is similar to the TFIIAα/βγ-TBP-DNA complex, as judged by competition with specific TATA (lanes 3 and 6) and non-specific Sp1 site (lanes 4 and 7) oligonucleotides. In addition, both complexes are supershifted to the well when co-incubated with antiserum against hTFIIAα/β (lanes 8 and 9), indicating that ALF and TFIIAα/β are immunologically related and are present in their respective complex.

To further support the idea that ALF is a functional polypeptide, we performed TFIIA-dependent in vitro transcription assays. For this purpose, we took advantage of the fact that TFIIAα/β contains an intrinsic seven-histidine region that allows for the efficient removal of TFIIA from HeLa cell nuclear extracts using nickel-NTA-agarose (3–5, 7). Depleted extracts are transcriptionally active but can be restored to normal activity by the addition of TFIIA. As shown in Fig. 7C (lanes 1–5), control (undepleted) extracts produced a [α-32P]CTP-labeled G-free RNA transcript expressed under the control of the AdML promoter (pMLC γAT) (52). The addition of recombinant ALF (p69) and TFIIAγ (p12) to these extracts did not enhance transcription (lane 2). TFIIA-depleted extracts were transcriptionally inactive and were not affected by the readdition of either ALF (p69) or TFIIAγ (p12) alone (lanes 3–5). However, the addition of both ALF (p69) and TFIIAγ (p12) at concentrations similar to those reported for recombinant TFIIAα/β and TFIIAγ (3, 4, 6, 7) restored transcription to the level observed with control extracts (lane 6). Based on the results of the electrophoretic mobility shift and in vitro transcription assays shown in Fig. 7B and C, we conclude that ALF is a functional homolog of TFIIAα/β and that both ALF and TFIIAα/β require the TFIIAγ subunit for activity.

To determine whether a full-length SALF cDNA construct was capable of directing the translation of an intact protein, we performed in vitro transcription-translation reactions. As shown in Fig. 7D, rabbit reticulocyte lysates programmed with pT7T3-SALF produced a [35S]methionine-labeled polypeptide that migrated at 170 kDa (compared with a predicted size of 132 kDa). Lysates programmed with pT7T3-SALF truncated at an internal EcoRI site at nucleotide position 960 (see Fig. 1B) produced a 36-kDa product similar to the predicted size of 32 kDa (data not shown). These results demonstrate that although SALF contains a suboptimal ATG initiation codon (AA-GATGΔ) and encodes a large ORF composed of two distinct regions, it can be translated efficiently in this assay. However, we have not produced full-length SALF proteins in quantities sufficient to test whether it is transcriptionally active.
DISCUSSION

This report describes the isolation and characterization of human cDNA clones that encode a TFIIAα/β homolog, ALF, and a related factor, SALF, and provides evidence for multiplicity and tissue specificity among the human general transcription factors.

Multiplicity among Human TFIIA-like Factors—The predicted amino acid sequence of ALF is organized into four regions (I-IV) that are characteristic of all TFIIA large subunits (Fig. 3A), including the conserved N- and C-terminal domains (Fig. 3B). In addition, the recombinant ALF protein, together with the TFIIAγ (p12) subunit, can stabilize TBP-DNA interactions in an electrophoretic mobility shift assay (Fig. 7B) and can restore RNA polymerase II-dependent transcription from the AdML promoter using TFIIA-depleted HeLa cell nuclear extracts (Fig. 7C). These results show that there are at least two genes for functional TFIIAα/β-like subunits in humans, namely ALF and TFIIAα/β. Likewise, multiple genes that encode the general transcription factor TBP have been described in Arabidopsis (At-1 and At-2) (39), Drosophila (TBP and TRF) (40), and humans (TBP and TLF) (41). Together, these studies reveal a complexity among the basal transcription factors that has not been evident from the biochemical characterization of fractionated nuclear extracts.

As TFIIA is composed of large (α/β) and small (γ) subunits, the question is raised as to whether there are additional human genes for TFIIAγ. The results of computer data base searches did not identify homologous sequences, and the genomic Southern blotting experiments suggest that TFIIAγ could be a single copy gene (Fig. 4) but will require genomic characterization to verify. Moreover, the ubiquitous co-expression of TFIIAα/β and TFIIAγ (Figs. 5 and 6) and the especially high levels of both ALF and TFIIAγ mRNAs in testis (Fig. 5, A and D) suggest that TFIIAγ is a common subunit for both factors. Interestingly, induction of the Drosophila dTFIIA-S gene (but not dTFIIA-L) is required for development of photoreceptors (55), implying that the formation or function of TFIIA-like complexes (such as TFIIAα/βγ and ALF/γ) might be regulated by the availability of the small subunit.

Expression of General Transcription Factors in Testis—At least five of the RNA polymerase II general transcription factors are expressed to higher levels in testis than in other tissues. First, as reported here, mRNA for both TFIIAα/β and TFIIAγ are more abundant (4–13-fold and 10–32-fold, respectively) in human testes (Figs. 5 and 6). Second, mRNA for TFIIIB is enriched 6–11-fold in rat testis (56). Third, mRNA for TBP is enriched 50–80-fold in rodent testis (56, 57) and is estimated to be 1000-fold higher in haploid round spermatid cells of mice (57). Finally, protein levels of the TFIIEγ subunit, as well as the RNA polymerase II large subunit, are several-fold higher in rodent testis (56, 57).

In addition, two homologs of general transcription factors are testis-specific or nearly so. ALF is expressed to 50-fold higher levels in human testis than in other tissues, as seen in the dot blot analysis (Fig. 6), and is the first example of a human tissue-specific general transcription factor. The level of ALF expression might be even higher if restricted to a particular cell type, as noted for TBP. Likewise, the Drosophila TBP-related factor TRF is expressed only in the central nervous system and primary spermatocytes of adult flies (40, 42). Interestingly, TRF is associated with a distinct set of neuronal TRF-associated factors and has been localized to a limited number of sites on salivary gland polytene chromosomes, implying that it may function at the promoters of relatively few genes (42).

The observations above suggest that general transcription factors and their homologs have a unique role in testis gene expression (see Refs. 59 and 60 and the references therein). The main function of this organ is the production of haploid spermatozoa from undifferentiated stem cells, a process called spermatogenesis (reviewed in Ref. 61). It is notable that mutations in the Drosophila trf gene result in immotile sperm and male sterility (40), indicating that TRF might be selectively involved in the transcription of genes whose products are required for spermatogenesis. In addition, transcription of some genes in testis occurs from promoters that are not used in somatic cells, giving rise to mRNAs with novel 5’-end sequences (59, 60). For example, the mouse tbp gene initiates from at least six promoters in testis, five of which are testis-specific, and produces 10 different transcripts (58). These observations suggest an unusual flexibility in core promoter recognition by the basal factors or a greater accessibility to core promoter-like sequences in genomic DNA that may occur during spermatogenesis. Finally, the overall level of transcription in testis is relatively high, possibly consistent with the enrichment of the RNA polymerase II machinery in this tissue. As our understanding of the relationship between testis gene expression and the general transcription factors is incomplete, ALF may be an ideal factor on which to focus future studies of this problem.

Functional Significance of ALF—The results of in vitro transcription assays show that ALF/γ and TFIIAα/βγ complexes can restore RNA polymerase II-dependent transcription in TFIIA-depleted HeLa cell nuclear extracts (Fig. 7C). Although these results are limited to the AdML promoter, they raise the possibility that ALF may be functionally identical with TFIIAα/β. If so, the importance of the human ALF gene may be to transcribe high levels of TFIIA large subunit mRNA in testis that cannot be obtained by transcription of the TFIIAα/β gene alone. This idea is similar to the suggestion that the increase in phosphoglycerate kinase PGK-2 gene expression during spermatogenesis compensates for the loss of PGK-1 transcripts, in this case because of the inactivation of the PGK-1 gene (62).

An alternative, and more interesting, possibility is that ALF, like TRF, may be required in vivo for the expression of a subset of class II genes. For instance, ALF/γ complexes might be selectively required to stabilize TBP (TFIIB)/DNA interactions on the core promoter sequences of testis-specific genes or to help mediate the function of testis-specific transcriptional activators. These activities may further depend on whether ALF is co-expressed in the same cell types as TFIIAα/β and whether it is post-translationally processed into ALFα and ALFβ subunits analogous to TFIIAα and TFIIAβ. A more detailed biochemical and functional characterization of the recombinant and endogenous factors will be required to address these issues.

Structure and Functional Significance of SALF—This paper also describes the isolation of a 3.8-kb cDNA, SALF, that predicts a nearly intact ALF protein connected to an N-terminal domain homologous to Drosophila Stoned B and clathrin APs μ1 (AP47) and μ2 (AP50). The conclusion that this cDNA represents a naturally occurring human mRNA is based on: 1) the identification of a partial human EST (I.M.A.G.E. Consortium Clone ID 259637) (Fig. 1A); 2) PCR amplification of products that span the Stoned B-TFIIAαβ junction from placenta, liver, and testis cDNA libraries (Fig. 1D); and 3) the isolation of contiguous upstream Stoned B-like sequences using primers within the downstream TFIIAαβ-like region (pRACE4) (Fig. 1B). It is important to note that novel ALF-containing transcripts are not limited to SALF, as two additional ESTs (I.M.A.G.E. Consortium Clone Ids 785133 and 1657721) that predict an alternative C terminus were identified in a data base search and verified by PCR analysis (data not shown).

The structure of SALF raises several issues. First, it is likely
significant that sequences similar or identical to the 5’-end of SALF are present in an abundant and widely distributed 6.5-kb mRNA (RNA6.5) (Fig. 5B) and that expression of RNA6.5 and ALF-containing RNAs overlaps in non-testis tissues (Fig. 6, A and B). These observations suggest that SALF might be derived from a gene (or adjacent genes) that contains RNA6.5- and ALF-like sequences or, alternatively, that it is formed from post-transcriptional processing of RNA6.5- and ALF-containing transcript(s). Second, the fact that SALF contains a domain potentially involved in the clathrin-dependent trafficking of membrane proteins (49, 50) suggests that, in this context, it might affect the subcellular distribution or activity of the C-terminal ALF domain. In any event, a \textit{bona fide} role for SALF in gene regulation would probably require the protein to be active even at low levels in the cell or that expression is higher in other cell types or under conditions not considered here.

In summary, the identification of ALF (and SALF) shows that multiple genes encoding TFIIB large subunits are present in human. Furthermore, the testis-specific expression of ALF, together with the overexpression of several other general transcription factors in this tissue, points to a special importance in testis biology and perhaps spermatogenesis.

Acknowledgments—We thank Dr. Robert G. Roeder for HeLa cell nuclear extracts, Dr. JingHong Mu for helpful comments on the manuscript, and Natalia Nikiforova and Melanie Marketon for assistance with \textit{in vitro} transcription-translation reactions.

REFERENCES

1. Roeder, R. G. (1991) Trends Biochem. Sci. 31, 327–335
2. Orphanides, G., Lagrange, T., and Reinberg, D. (1996) Genes Dev. 10, 2657–2663
3. De Jong, J., and Roeder, R. G. (1993) Genes Dev. 7, 2220–2234
4. Ma, D., Watanabe, H., Merzmel, F., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D., and Handa, H. (1993) Genes Dev. 7, 2246–2257
5. Ozer, J., Moore, P. A., Bolden, A. H., Lee, A., Rosen, C. A., and Lieberman, P. (1994) Genes Dev. 8, 2324–2335
6. Sun, X., Ma, D., Sheldon, M., Yeung, K., and Reinberg, D. (1994) Genes Dev. 8, 2336–2348
7. De Jong, J., Bernstein, R., and Roeder, R. G. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3313–3317
8. Ranish, J. A., Lane, W. S., and Hahn, S. (1992) Science 255, 1127–1129
9. Yokomori, K., Admon, A., Goodrich, J. A., Chen, J.-L., and Tjian, R. (1993) Genes Dev. 7, 2235–2245
10. Yokomori, K., Zeidler, M. P., Chen, J.-L., Verrijzer, C. P., and Tjian, R. (1994) Genes Dev. 8, 2313–2323
11. Bernstein, R., De Jong, J., and Roeder, R. G. (1994) J. Biol. Chem. 269, 24361–24366
12. Tan, S., Hunziker, Y., Sargent, D. F., and Richardson, T. J. (1996) Nature 381, 59–60
13. Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) Science 272, 830–836
14. Baratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989) Cell 56, 549–561
15. Lee, D. K., De Jong, J., Hashimoto, S., Horikoshi, M., and Roeder, R. G. (1992) Mol. Cell. Biol. 12, 5189–5196
16. Imbalzano, A. N., Zaret, K. S., and Kingston, R. E. (1994) J. Biol. Chem. 269, 8280–8286
17. Weideman, C. A., Netter, R. C., Benjamini, R. M., McAllister, J. J., Schmiedekemp, L. A., Coleman, R. A., and Pugh, B. F. (1997) J. Mol. Biol. 271, 61–75