Cloning, expression and characterization of the human transcription elongation factor, TFIIS

OokJoon Yoo, HoSup Yoon, KwangHee Baek, ChoonJu Jeon, Kenichi Miyamoto, Akemichi Ueno and Kan Agarwal*

Departments of Biochemistry, Molecular Biology and Chemistry, The University of Chicago, Chicago, IL 60637, USA

Received November 28, 1990; Revised and Accepted February 1, 1991

EMBL accession no. X57198

ABSTRACT

The cDNA for the human elongation factor, TFIIS, has been cloned and expressed in *E. coli* with the T7 expression system. This 280-amino acid TFIIS protein is shorter by 21 residues than that of the mouse. The missing 21 residues are located in the amino-terminal region, which is not thought to be required for transcriptional stimulation. Apart from this gap, human and mouse proteins reveal 96% overall identity and 98.5% sequence similarity if conservative substitutions are taken into account. The bacterially expressed human protein and the purified calf thymus proteins are indistinguishable in their ability to stimulate transcript elongation by purified RNA polymerase II. Estimation of the native molecular size of the human protein in solution indicates that it exists as a dimer.

INTRODUCTION

The underlying mechanisms responsible for eukaryotic transcriptional regulation are being deciphered with the isolation and functional characterization of the factors that govern defined steps in RNA synthesis (1, 2). Numerous factors that promote initiation and subsequent activation of RNA polymerase II transcription have been isolated from a variety of sources, and biochemical studies of the relationships between their structure and functions are beginning to reveal the mechanism of their function (2). In contrast, little is known about regulation at transcript elongation and eventual termination. A transcription factor, TFIIS (previously known as S-II), was first discovered as a factor stimulating transcription by RNA polymerase II 2- to 3-fold (3) and has now been shown to act during transcript elongation (4). In addition, this factor exerts influence on the intrinsic ability of polymerase to recognize some pause sites or premature termination sites, by allowing the polymerase to read through at these sites (5). These biochemical functions of TFIIS emerged from *in vitro* studies that employed well-defined purified components, but the true role of TFIIS *in vivo* remains unknown.

The wide distribution of this factor in yeast (6), *Drosophila* (7), mouse (8), calf thymus (9), and human (4), indicates that this factor might play an essential role in the transcript elongation process. The functional specificity of this factor for RNA polymerase II, as opposed to polymerase I and II and *E. coli* polymerase, further suggests that it plays an important but well-defined role in transcription elongation (4). Its ability to function effectively with interspecies RNA polymerase II (and vice versa) suggests strong conservation in structural and functional domains (4, 5).

Biochemical analysis of mouse, calf thymus, and yeast TFIIS proteins revealed that these proteins were capable of binding to purified RNA polymerase II, independent of DNA and NTP's, and that the complexes could be readily isolated by glycerol gradient centrifugation for analysis (4). This binding (dissociation constant $5 \times 10^{-8} \text{M}$), in the case of yeast protein (p. 37), was determined to be salt-sensitive indicating involvement of ionic interactions (6). Interestingly, yeast protein binds specifically to the RNA polymerase II form that contains phosphorylated CTD (6). Binding of TFIIS to an additional site on the largest subunit has also been suggested by an experiment that employed a fusion protein containing a fragment lacking CTD from the largest subunit of RNA polymerase II (10). While the functional significance of TFIIS binding to the internal region of polymerase is unclear, the interaction between CTD and TFIIS is significant in light of the suggestion that CTD may play an important role in the transcript elongation process (11, 12).

The small amount of TFIIS in eukaryotic cells and the difficulty in generating mutant forms of the protein *in vivo* were in part, the reasons why the underlying biochemical mechanisms of its function remained poorly understood. However, the recent molecular cloning of mouse TFIIS now allows ready access to wild-type and mutant forms of proteins for biochemical analysis (13).

In this paper we describe the isolation and characterization of human TFIIS cDNA and the purification and biochemical characterization of the bacterially expressed protein. In contrast

* To whom correspondence should be addressed

to mouse protein, the human protein lacks a stretch of 21 amino acids from the amino terminal region. The amino acid sequences of human and mouse proteins reveal a remarkable overall identity of 96%, and all of the amino acid substitutions are located in the 100-amino acid amino terminal region. The *E. coli*-expressed TFIIS protein is as active as calf thymus protein in stimulation of the transcription of purified RNA polymerase II. *E. coli* and calf thymus proteins share similar antigenic determinants as demonstrated by the finding that antibodies against *E. coli* TFIIS are able effectively to react with calf thymus protein. In addition, we show that TFIIS exists as a dimer in solution.

MATERIALS AND METHODS

Enzymes. Calf thymus RNA polymerase II was purified to > 90% homogeneity with the procedures reported previously (14, 15). One unit of RNA polymerase II was defined as the amount of the enzyme required to incorporate 1 μ mol of CMP or UMP per minute in a 10 min reaction at 37°C. Calf thymus TFIIS was purified to > 85% homogeneity with the modified procedure described previously (9). One unit of TFIIS was defined as the amount of enzyme required to stimulate 1 unit of RNA polymerase II to 2 units (4).

Preparation of C-tailed DNA. The pUST1 clone that contained the *Ava*II-BstNI fragment of the gastrin gene 3' flanking region cloned into the pUC plasmid (16) was linearized by digestion with *Sma*I (this site is located 5' to the *Ava*II site). The resulting 3' ends were extended with deoxynucleotidyl terminal transferase and [³H] dCTP by the procedure described previously (17). An average of 60–100 dC-residues were added per 3' end. The tailed plasmid was then digested with *Hind*III (this site is located 3' to the BstNI region), and the resulting uniquely tailed fragment was purified with gel electrophoresis.

Assay of TFIIS activity. Two transcription assays were used. In one of these calf thymus DNA was the template, and in the other, dC-tailed DNA was the template.

Assay employing calf thymus DNA. The transcription reaction mixture contained 62.5 mM Tris HCl, pH 7.9, 1.69 mM MnCl₂, 600 mM each of ATP, GTP, CTP, and 100 μ M UTP, along with 2 μ Ci of [5, 6-³H] UTP (40 Ci/mmol), 8.25% glycerol, 25 μ M EDTA, 125 mM DTT, 50 mM (NH₄)₂SO₄, and 330 μ g/ml of calf thymus DNA in 30 μ l. Reactions were carried out at 37°C for 10 min in the presence of 0.1–0.5 units of purified RNA polymerase II alone or in combination with an appropriate amount of TFIIS. Workup of the reaction followed the procedure described previously (9).

Assay employing dC-tailed DNA. The dC-tailed DNA (2 nM) in 25 mM K-HEPES, pH 7.9, 0.5 mM EDTA, 0.5 mM DTT, 6.25 mM MgCl₂, 5 mM spermidine, 150 mM K-glutamate, 10% glycerol was incubated with 1 unit RNA polymerase II alone or in combination with 7 to 10 units TFIIS at 30°C for 30 min. The reaction was initiated with the addition of 500 μ M each of ATP, GTP, UTP, and 100 μ M CTP plus 10 μ Ci (α -³²P)CTP (3,000 Ci/mmol) in a final volume of 12.5 μ l and was incubated further at 37°C for 10 min. The reaction was stopped with 1 μ l of 0.5M EDTA in a one-tenth volume of 3M Na-acetate containing 2 μ g of tRNA, followed by extraction with phenol/CHCl₃/isoamyl alcohol. The nucleic acids were precipitated with ethanol and analyzed with 8% polyacrylamide 8M urea gels.

Cloning of human TFIIS cDNA. A human kidney cDNA library in λ gt 10, generously supplied by Dr. Graeme Bell, was screened for the TFIIS cDNA clone with the following oligonucleotide probes: CTGCACCATTTTGRCCATCTTCTTGGCAATGCG-AACCACCTCGTCCTCCAT (this sequence corresponds to the 5' region of the mouse TFIIS mRNA (13); residues 82–133), and CCACCGATTTCACATTTCATTACATACAAC-AAATGTTGTCATTGGTTCATCAGCAC (this probe represents the 3' region of the mouse mRNA; residues 921–976). The library was first screened with a mixture of these probes; duplicate screening of the positive clones with individual probes followed. Screening of the phage library was performed according to the procedure described previously (18). The nucleotide sequences of the cDNA inserts of the positive clones were determined by subcloning the insert into M13 phage DNA followed by dideoxysequencing (19). Both strands were sequenced.

Overproduction of the TFIIS protein in *E. coli*. The cDNA encoding TFIIS (nucleotides 125–2,357) was excised from the plasmid, pHIIS44, by double digestion with *Nco*I and *Kpn*I, and the resulting fragment was isolated and then subcloned into the *Nco*I and *Kpn*I sites of the plasmid, pT/CEBP (20) by a fragment replacement strategy. In this construction (pEIIS), the TFIIS coding sequence (residues 125–2,357) was directly under the control of the T7 RNA polymerase promoter, and translation was initiated at the TFIIS Met. The clone, pT/CEBP was kindly donated by Dr. B. Landschulz of the Carnegie Institute of Washington (20). Overproduction of TFIIS from pEIIS was achieved by adding 0.5 mM IPTG to BL21 (DE3)/pEIIS at an optical density of 1.0 at 600 nm followed by harvesting of the cells 2 hours later.

Purification of TFIIS. The procedures described for the isolation of TFIIS from mouse ascites cells (8) and calf thymus tissue (9) for isolation of TFIIS could not be adopted for the BL21(DE3) strain. A large proportion of native TFIIS protein was degraded to a truncated, inactive protein even though several protease inhibitors were present at all times during purification. The procedure to be briefly described resulted in the isolation of a highly enriched and fully active TFIIS protein. Thirty grams of BL21(DE3) cells were suspended in 200 ml of buffer A (500 mM Tris HCl, pH 7.9, 5% glycerol, 0.5 mM DTT, 1 mM PMSF, 1 mM aprotinin, 5 mg/ml leupeptin, 2 mg per ml pepstatin) containing 233 mM NaCl, and 2 mM EDTA and treated with 50 mg lysozyme for 20 min followed by addition of 4% DOC to a final concentration of 0.5%, and then allowed to stand for 20 min at 4°C. To the clear supernatant, generated after centrifugation for 45 min at 9,000 rpm, a 5% streptomycin sulfate solution was added to a final concentration of 1.2%; the precipitate was removed by repeating the centrifugation. To the supernatant was added 89.2 gm of powdered ammonium sulfate to 55% saturation. After 1.0 hr of gentle stirring, the precipitate was collected by centrifugation (9,000 rpm for 30 min). The precipitate was suspended in 50 ml of buffer B (buffer A containing 50 mM NaCl and 0.2 mM EDTA), and the solution was dialyzed against buffer B with three changes of buffer. The dialysate was loaded on to a preequilibrated DEAE-Sepharose Column (4.4 cm $\times$ 13.2 cm) and the resin washed extensively with buffer B. Most of the TFIIS protein was found in the flow-through solution as judged with SDS-PAGE and the RNA polymerase II stimulatory activity assay. The fractions containing TFIIS

protein were combined and treated with solid ammonium sulfate to 55% saturation. The precipitate was collected by centrifugation and then suspended in buffer B containing 100 mM instead of 50 mM NaCl and loaded onto a phosphocellulose (P-11) Column (2×15 cm). After the column was washed with buffer B containing 100 mM NaCl, it was developed with 200 ml of a linear gradient of 100 mM-500 mM NaCl in the same buffer. The TFIIS protein was eluted at approximately 300 mM NaCl and was judged by SDS-PAGE to be contaminated with a smaller protein. Further purification was accomplished with chromatography on a Sephacryl S-200 column (2.2 cm×96 cm) in buffer B containing 500 mM NaCl. This purification step resulted in a > 85% homogeneous TFIIS protein as judged with SDS-PAGE. The purified protein samples were stored in buffer B containing 300 mM NaCl at -20° C.

Preparation of ³⁵S-labeled TFIIS protein. Cells were grown in M9 minimal medium to an optical density of 0.5 at 600 nm, and then induced by the addition of IPTG to a concentration of 0.3 mM, and allowed to grow for an additional 30 min. Rifampicin was then added to a concentration of 200 µg/ml, and cell growth was continued for another 90 min. ³⁵S Met (20 µCi/ml of culture) was added, and the cells were grown for an additional 5 min. Cells were centrifuged for 10 min at 5,000, rpm and the cell pellet was washed by resuspending it in 25 mM HEPES, pH 7.6, 1 mM EDTA. The cells were repelleted with centrifugation and were resuspended in one tenth of the culture volume and allowed to freeze in liquid nitrogen and stored at -80° C. The labeled TFIIS was isolated with the procedure described above. Purity of the labeled protein was assessed with 12% SDS-PAGE, and the protein bands were made visible with 0.1% coomassie blue stain and autoradiography.

Physical properties of TFIIS protein. The molecular weight of the native form of TFIIS was estimated with two methods: (1) chromatography on a Sephacryl S-300 Column along with proteins of known molecular weight, and (2) chemical crosslinking of TFIIS in solution. Chromatography of ³⁵S Met-labeled TFIIS on a Sephacryl S-300 Column was performed as recommended by the manufacturer. Fractions containing radioactivity were analyzed with SDS-PAGE.

Chemical crosslinking of ³⁵S Met-labeled TFIIS was accomplished by treatment of a 160 µg/ml solution of TFIIS protein in 60 mM Na₃PO₄, pH 7.0, 200 mM NaCl, 10% glycerol, with 0.5 mM bismaleimidoethane (BMH, Pierce) for 40 min at room temperature. The reaction was stopped with the addition of 5 mM 2-mercaptoethanol and 5 mM of ethanolamine as described previously (21). The reaction mixture was analyzed with 12% SDS-PAGE followed by autoradiography.

Preparation of antibodies against *E. coli* expressed human TFIIS. One mg of purified TFIIS in Freund complete adjuvant was injected into rabbits which followed (after 3 weeks) booster injection of 100 µg of the protein in incomplete Freund adjuvant. The animals were bled one week after the booster injection, and antisera were prepared and examined with immunodiffusion method (22). IgG was purified from the sera by (NH₄)₂SO₄ precipitation followed by DEAE-cellulose chromatography (23). The isolated IgG was specific for the TFIIS protein was determined by its ability to block the stimulatory activity of TFIIS in an RNA polymerase II transcription assay. Western blot analysis also showed binding of IgG to purified TFIIS protein.

RESULTS

Isolation and characterization of human TFIIS clones

A human kidney cDNA library was screened with two oligonucleotide probes whose sequences were derived from the mouse TFIIS mRNA and corresponded to the amino- and carboxyl-terminal regions of the protein (9). Five clones that hybridized to each of these probes were isolated; they contained

1	GCCTAGCCCG	GGGTGCGGTG	GTGGGGGTTT	GTGCGCGCGG	GGGGTGGCTC	CTGCTGTGTC
61	TTCCCGTCCA	CGTTGCCCCA	CTTCCCTTTG	CCAGCGGGGT	GGGGCGGGAG	AAGACTGCGG
121	GAGCCATGGA	GGACGAAGTG	GTCCGCTTTG	CCAGAAGATG	GGACAAGATG	GTGCAGAAGA
181	AGAAAGCGGT	CACAAGAATC	GGAATGTCAG	TTAATGCTAT	TGCGAAGCAG	AGTACAGATG
241	AGGAAGTTAC	ATCTTTGGCA	AAGTCTCTCA	TCAAATCCTG	GAAGAAATTA	TTAGATGGGC
301	CATCAACTGA	GAAAGACCTT	GAGCAAAAGA	AGAAAGAAC	TGCAATTTACA	TCCGAGAACA
361	GCCTCTAGGC	AAGAGAAGAA	AGTACTTTCA	GCGGCAATGT	AAGCAACAGA	AAGGATGAGA
421	CAATGCTGCG	AGATACCTAT	GTTCATCTCT	TTCTCTGGGC	ACCAAGCACT	TCTGACATTT
481	ATTCTGTGCTG	GTGGAAGTGT	AGGGAATGTC	TTGCTGCAGC	TCTTCCGACA	GGGGATGACT
541	CAATGTGAGC	TGATGAGGAA	GAATTAGGAT	CTCAATTTGA	AGAAGCTATA	TATCAAGAAA
601	TAAGGAATAC	AGACATGAAA	TACAAAATAA	GAGTACGAAG	TAGGATATCA	AACTTTAAAG
661	ATGCAAAAAA	TCCAAATTTA	AGGAAAATAG	TCCTCTGTGG	GAAATTTCTT	CCTGACTTAT
721	TTGCTAGAAT	GACAGCAGAG	GAAATGGCTA	GTGATGAGCT	GAAAGAGATG	CGGAAAAACT
781	TGACCAAGA	AGCCATCAGA	GAGCATCAGA	TGGCCAAGAC	TGGTGGGACC	CAGACTGACT
841	TGTTCAATGT	TGGCAAAATG	AAAAAGAGTA	ATTCAGCTTA	CACACAGGAT	CAGACCCGTA
901	GTGCTGATGA	ACCAATGACA	ACATTGTGTT	TCTGTAATGA	ATGTGGAATG	CGATGGGAAT
961	TCTGTGAGT	TGGAAGAATT	GGCAAAATAT	CTGGACCAAT	AAGAAAACGG	ATTTTGTAA
1021	TAGCTTTAAA	CTAGGCCAAG	CAACTAGTGT	TCCTGCAAA	CAAAATTTTA	AAGCAACTTG
1081	GGTTAGACTT	TGTTTTTGAC	CTAATCTCCC	TCCTTTAAAT	GCCTTCTGTA	GTCTCAGATC
1141	AGTAGGGAGA	CCATATAAAT	ATTTTATGTT	ACCTGTTTCA	AAACATATTT	TTTCTGTTTT
1201	TATGATTAAG	TATGATTTAA	TAAACCTCTT	GGCAATATTT	CTTCTTTCTT	AAAGGAATAT
1261	ATACCTTAAC	TTTTTTTCTT	TACACTGTG	AAACATACAC	AGTAGAAAT	CTGTACTCT
1321	CTGTTATTAA	TACATAAATG	AAAATACATT	TTTTTCCATA	TTGGCATGTA	GCTACAATAA
1381	TAAAGGAGC	AGAAAAGGTA	ATATAATTTT	AGGTTTACCA	AATATGGTGT	GATTTCAAA
1441	AATACCTTGC	GACCTTATCT	AAAATGTACA	TAATTTTGAG	TAGTCTTATG	AATTTGATT
1501	TAATTTATAT	GTTCACAAGC	TGGAATATTT	AGATATTTAT	TTGATCTGTT	AGCTAACCGT
1561	GATCATCAT	TCTTGTAAAT	TCTGTATACT	GTATATTACT	TGTTCTTAAT	AGATTTTGTG
1621	AAACAAGACT	TTATTGAGAT	CAGTTGGTAT	TTCTGTTTAA	TTTACCTGTT	TGACTTTATA
1681	ATGTGTTTAA	GTTTTGCAGA	AGAACACTGT	TGATGTTTAA	AGAGCTTTTC	ATAAATCCCC
1741	TCATAGGCAA	AGATGAAAAC	TTCCCACTAT	TTTTTCCCC	TCTTAGGAAG	ACATACCTGA
1801	AAGAAAATAT	TAGCATCTT	AGTGTAGTAT	AGCTATTGTA	AACAGTTTAT	GACTAGTACT
1861	TGATTCGGAA	ATCTATACCT	ACCAAGGATT	AATCTTAAAG	ATTGTATTAAT	TCATTAAGCG
1921	TGTGTTGCTT	CCATGTGGAG	ACTGATAGAA	AATAATTTTG	TCCCAAGTCT	TATTTGCTGA
1981	CTTTTCTGCT	CATGAGTGAG	ATTGTTGAAC	AAACGTAATA	TATGGGCTAT	AGCAAGTAGC
2041	TTTACAGTTT	GATCTTTACA	ATTGATGTTT	GCTTTTGTGA	AAAGTTGTAC	CATTTTCTCT
2101	GTGTTGAGTA	AGACAAAAAT	TGTTTTGACA	TAGGTTCCCT	AGGGTACACT	TGCTCTAGCA
2161	TACTTTAAAG	GCCACTGTGT	CAAGCTTATC	ATTTTATGCT	GAATCTGAGT	TCTGTGAGCA
2221	ACCCGTAGAA	AGACCTCAGT	ACATGCTTTG	CACCTCTCCT	TGCTCCCTTT	TTCTTAGAAG
2281	AATTTCTTAT	TGCATATCAT	TTGTTGTAAT	ACAGAAAGCA	GCAATTTTAA	ATGCTCCGTG
2341	TTGGCCCACT	GGTACCAACT	CACCTCTATT	TTGTCAGTTC	ATAGTTGAAG	ATTTGTTTAA
2401	TTTCAAAAAC	AAAGTACATT	TTTGAATAAA	TGTTTCAGAA	TAAATAATCT	TCACTTTAA
2461	GTGATCCTTT	TAAAAATTTG	AATTCATAAA	AGTTTTTTTT	GTGTTTAAAC	TAAACG

Figure 1. Primary structure of human TFIIS mRNA. The nucleotide sequence of the mRNA deduced from the 2.5-Kbp cDNA insert in clone pHIS44 is shown. Nucleotide 1 represents the first base of cDNA. The sites of translation initiation and termination and of polyadenylation are underlined.

Human (pHIS78)	AAC	GCG	GCT	GGA	GCA	TTG	GAT	TTG	CTA	AAG	GAG	CTT	AAG	AAT
Mouse	AAT	GCG	GCT	GGA	GCA	TTG	GAT	TTG	CTG	AAA	GAG	CTT	AAG	AAT
Human (pHIS44)	AAT	GCG	-	-	-	-	-	-	-	-	-	-	-	-
	ATT	CCT	ATG	ACC	CTG	GAA	TTA	CTG	CAG	TGT	GCA	GCT	GAC	CCC
	ATT	CCC	ATG	ACC	CTA	GAA	TTG	CTA	CAG	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	TAA	GCG	TGC	TTA	ATG	CAC	TCT	CCA	CCT	TTC	CAG	GCT	TCC	ACA
	-	-	-	-	-	-	-	-	-	-	-	-	TCC	ACA
	-	-	-	-	-	-	-	-	-	-	-	-	TCC	ACA
	-	-	-	-	-	-	-	-	-	-	-	-	TCC	ACA

Figure 2. Comparison of the human and mouse sequences in the region that predicts the presence of an intron in the human TFIIS cDNA is shown. Comparison of human and mouse sequences in the region of differences is shown. The missing nucleotides are shown with dashes. The numbering of the nucleotides corresponds to that of the respective clones. The proposed intron/extron boundaries are indicated by an inverted triangle. The dinucleotides generally found at these boundaries are underlined. The pyrimidine-rich sequence frequently found upstream of AG dinucleotide is underlined. The stop codon is boxed.

Human	1	MEDEVVRF	AKKMDKMVQK	NA.....	STRIGMSV	29
		:				
Mouse	1	MEDEVVRI	AKKMDKMVQK	NAAGALDLLKEL	KNPMTLELLQ	STRIGMSV 50
	30	NAIRKQST	DEEVTS	LAKSLIKSWK	LLDGPSTEKDL	DEKKKEPAITSQNS 79
		:		:		
	51	NALRKQST	DEEVTS	LAKSLIKSWK	LLDGPSTDKD	PEKKKEPAISSQNS 100
	80	PEAREEST	SSGNVSNR	KDETNRD	TYVSSFP	PRAPSTSDSVRLKCREMLAA 129
		:		:		
	101	PEAREESS	SSSNVSR	KDETNRD	TYVSSFP	PRAPSTSDSVRLKCREMLAA 150
	130	ALRTGDD	YIAIGADE	EELGSQIEE	AIYQEI	RNTDMKYKNRVRSRISNLKD 179
		:		:		
	151	ALRTGDD	YVAIGADE	EELGSQIEE	AIYQEI	RNTDMKYKNRVRSRISNLKD 200
	180	AKNPNLR	KNVLCGN	IPPDFL	FARMTAE	EMASDELKEMRKNLTKEAIREHQM 229
		:		:		
	201	AKNPNLR	KNVLCGN	IPPDFL	FARMTAE	EMASDELKEMRKNLTKEAIREHQM 250
	230	AKTGGTQ	TDLFTCG	KCKKKN	CTYTVQ	QVTRSADEPMTTFVVCNCGNRWKF 279
		:		:		
	251	AKTGGTQ	TDLFTCG	KCKKKN	CTYTVQ	QVTRSADEPMTTFVVCNCGNRWKF 300
	280	C	280			
	301	C	301			

Figure 3. Amino acid alignment of human and mouse TFIIIS proteins. The vertical lines indicate identity, and colons indicate conservative amino acid replacements. The gap is shown with dots.

inserts of approximately 2.5 Kbp. The restriction-endonuclease-mapping pattern indicated the existence of two classes of clones; one class contained an additional pVulII site and was designated pHIS75, while the other was lacking this site and was designated as pHIS44. Analysis with detailed restriction mapping indicated that the pVulII site was in the NcoI-XhoI region (nucleotides 124–427; Fig. 1); this region was estimated to be 100–200 nucleotides longer than the corresponding region in the pHIS44 clone (data not shown).

Nucleotide sequence analysis revealed virtual identity between the clones (the sequence of the pHIS44 clone is presented in Fig. 1), except that the pHIS44 clone contained a deletion of 112 nucleotides. This insert appeared to be an unspliced intron sequence for the following reasons: 1) analysis of the coding ability of the pHIS44 clone indicated the potential generation of a 280-amino acid polypeptide similar to the mouse 301 amino acid protein, whereas pHIS75 coded for only an 89 amino acid peptide, and 2) examination of the 112 nucleotide sequence revealed the presence of the dinucleotides, GC and AG, and a pyrimidine-rich region upstream of the AG dinucleotide; these are features common to functional introns (Fig. 2). It was of interest, however, that the mouse mRNA contained 63 of the 112 human nucleotides (these 63 exhibited 91% identity) and encoded for 21 amino acids that made the mouse protein larger than the human protein. This 63 nucleotide region, although it contained features of a functional intron, apparently did not undergo splicing, perhaps because it did not meet the critical-size requirement. In human mRNA, the presence of a stop codon (TAA) may exert an evolutionary pressure to generate an mRNA that encodes for a functional protein, as a consequence of which a spliceable intron is created. Whatever the mechanism, one is curious how this human intron region evolved among different species.

Comparison of human and mouse TFIIIS DNA and protein sequences

Human and mouse TFIIIS mRNAs shared a remarkable 92% identity in the nucleotide sequences of their coding region,

Figure 4. Expression of TFIIIS in *E. coli* and its purification. The conditions of cell growth and protein purification are described in the text. A 10% polyacrylamide gel was stained with coomassie blue after electrophoresis in the presence of SDS. The uninduced lane contains *E. coli* BL21(DE3)/pT5 cells (lacks cDNA insert). The induced lane contains BL21(DE3)/pEIIIS cells induced with IPTG. The purified TFIIIS protein from the induced-cell extract is shown in the purified TFIIIS lane.

indicating a high level of conservation during the evolution of this gene. The sequence was less conserved in the 5' (46%) and 3' (55%) untranslated regions, however. Identity between human and mouse amino acid sequences was equally striking at 96% with a gap of 21 amino acids and 98.5% if conservative amino acid substitutions are considered (Fig. 3). Almost all of the substitutions were located in the 100-residue N-terminal, a phosphorylated region (24). The 180-residue C-terminal region was 99% identical, with a conservative substitution of isoleucine with valine (Fig. 3).

Sequence similarity of TFIIIS to other proteins

A search of the NBRF data bank with the human sequence identified its similarity to the yeast PPR2 protein. These two proteins (the 128 amino acid PPR2 and the 280 amino acid human TFIIIS) shared an overall identity of 40%, and in the C-terminal 44 amino acids they shared an identity of 70%. Similar homology was recently noted between mouse TFIIIS and PPR2 protein (25). While this manuscript was in preparation, Ahn *et al.* reported a sequence of the vaccinia virus RNA polymerase subunit that was somewhat similar to mouse TFIIIS (26). Approximately 23% of the residues were identical over a 180 amino acid overlap between the mouse TFIIIS and vaccinia rpo30. While it remains to be shown whether rpo30 is a true functional homolog of TFIIIS, it is striking that there exists significant identity near the TFIIIS C-terminal region that is thought to be involved in zinc binding. A high degree of sequence similarity (72%) has recently been reported for *Drosophila* and mouse TFIIIS proteins as well (29).

Figure 5. Physical and immunological characterization of TFIIS. (A) Purified *E. coli* expressed TFIIS protein and purified calf thymus (C.T.) protein were analyzed with 10% SDS-PAGE followed by staining with coomassie blue. Results of the analysis of the mixture of human and calf thymus proteins is also shown. (B) Results of western-blot analysis of human and calf thymus proteins using antihuman antibodies is shown. Conditions for the blot analysis are described in the text.

Expression and purification of human TFIIS

The human TFIIS cDNA was expressed in *E. coli* with the T7 RNA polymerase expression system developed by Studier and coworkers (27). Cloning of cDNA into the expression plasmid is described in the Methods section. In the expression plasmid, pEIIS, transcription of the cDNA was under the direct control of the T7 promoter and the translation of the mRNA was initiated at the TFIIS amino terminal methionine. As shown in Figure 4, a protein of 35.8 Kd was expressed in the cells harboring the pEIIS plasmid on induction with IPTG (lane 2), while this protein was not expressed in the control cells containing the expression plasmid that lacked the TFIIS cDNA insert (lane 1). The amount of TFIIS expressed was estimated at 3% to 5% of the total *E. coli* proteins.

The moderately high expression of the TFIIS protein allowed its isolation in highly enriched form. However the native protein was susceptible to truncation in the crude extract during the early purification steps. A purification scheme, although laborious, allowed the isolation of a protein that contained roughly 85% native form and 15% truncated form (Fig. 4, lane 3). Several-milligram quantities of TFIIS were isolated with this procedure for structural and biochemical characterization.

Characterization of *E. coli* expressed TFIIS

The purified protein was characterized physically, immunologically, and by its ability to stimulate RNA polymerase II transcription. The denatured human protein (35.8 Kd) is smaller (Fig. 5A, lane 2) than the purified calf thymus protein (38 Kd; lanes 1 and 3). This suggests that calf thymus and mouse proteins (38 Kd) are of similar size and the human protein is smaller because of the gap of 21 amino acids. The structural relation between human and calf thymus proteins was examined by determining whether anti-human-protein-specific antibodies interacted with calf thymus protein. As shown in Figure 5B, human antibodies effectively interacted with calf thymus protein, indicating that both proteins share immunological epitopes. Interestingly, calf thymus protein contained a low-molecular-

Figure 6. Results of analysis of the effect of human and calf thymus TFIIS proteins on RNA polymerase II transcription. The dC-tailed DNA template used is illustrated in the upper part of the figure. Its preparation is described in the text. Sizes of the full-length and terminated transcripts are indicated. The lower part of the figure shows results of the analysis of transcription of dC-tailed DNA by RNA polymerase II (pol II) alone, pol II + calf thymus (C.T.) TFIIS, and pol II + human TFIIS. Details of the transcription reaction are given in the text. Similar amounts of calf thymus and human proteins were used in the assay. Positions of the full-length and terminated transcripts are indicated by arrows.

weight protein similar to human protein (Fig. 5B; compare lanes 1 and 3) further indicating structural similarity between these proteins. This observation also suggested the existence of a protease in both *E. coli* and mammalian cells that can digest TFIIS.

The ability of *E. coli* expressed human protein to stimulate RNA polymerase II transcription of dC-tailed DNA was examined and compared with that of purified calf thymus protein. As shown in Figure 6, both proteins stimulated transcription to similar levels, as evidenced by the nearly equal amounts of full-length RNA produced (compare lanes 2 and 3 with lane 1). The level of stimulation in each case was approximately 5- to 6-fold. In addition to stimulation of transcript elongation, human protein allowed RNA polymerase II to read through the gastrin gene termination site to the same extent as calf thymus protein (compare lanes 2 and 3). This suggested that *E. coli* expressed human protein was functionally identical to purified calf thymus protein.

Estimation of the native molecular size of human TFIIS

Analysis of the elution behavior of human protein with chromatography on a Sephacryl S-300 Column indicated that the human protein might exist as a dimer of 72Kd, (Figure 7A). However, considering the pitfalls of the gel-filtration procedure in estimating native molecular size, we employed a chemical crosslinking procedure that allows estimation of native molecular size, independent of shape. Crosslinking by glutaraldehyde was uninformative because the protein products rapidly precipitated out of solution. This may have been due to formation of crosslinked networks with lysine residues. The problem was circumvented, however, by employing a mild thiol-specific bismaleimidehexane (BMH) reagent (21). Under controlled

Figure 7. Results of the analysis of molecular size of the native TFIIIS protein. (A) Results of the analysis of the protein on Sephacryl S-300 Column is shown. Sizes of the protein markers analyzed on the same column are indicated. Estimated molecular size of TFIIIS protein is also indicated. (B) Results of chemical crosslinking of [³⁵S] Met-labeled TFIIIS protein with BMH (see text) is shown. The reaction products were analyzed by 10% SDS-PAGE followed by autoradiography. Positions of the monomer and dimer are shown by arrows.

reaction conditions with a protein concentration of 160 μ g/ml, about 5%–10% of the protein crosslinked to a dimer form, as shown in Figure 7, lanes 1 and 2. This suggested that human TFIIIS exists as a dimer in solution. On the basis of virtual identity among human, mouse, and calf thymus proteins, it is reasonable to suggest that mouse and calf thymus proteins may also exist as dimers in solution. It is not known, however, whether TFIIIS acts as a dimer. Studies to probe this aspect are currently under way.

DISCUSSION

The cDNA clone described in this report was identified by screening a human kidney cDNA library with oligonucleotide probes derived from the mouse mRNA. The goal was to generate a cDNA clone for structural and functional studies. The mouse clone would have served the desired purpose, but since it was not in hand, we embarked upon isolation and characterization of the human TFIIIS clone. Interestingly, two classes of clones were identified; the mature form and presumed pre-mRNA form. The protein sequence derived from the mature form of mRNA encoded a 280 amino acid polypeptide that is shorter by 21 amino acid than the mouse protein (13). The additional 21 amino acids of the mouse protein are encoded by the sequence that is part of the intron and that is highly homologous to the human sequence. One wonders how a nonspliceable potential mouse intron of 63 nucleotides evolved into a 112 nucleotide spliceable intron. Some useful information relating to the evolution of this gene may emerge from comparison of the human, mouse, *Drosophila*, and yeast nucleotide sequences.

The amino acid sequences of human and mouse are remarkably conserved, indicating their essential role in transcript elongation by RNA polymerase II. It is likely that TFIIIS proteins from other eukaryotic species also exhibit high conservation of their amino acid sequences because of the essential role of these sequences in RNA polymerase II transcription and the fact that proteins identified from other eukaryotic species are of similar molecular size. The absence of 21 amino acids in the amino terminal region, the clustering of amino acid substitutions in this region, and the functional similarity between human and calf thymus proteins indicate that the amino terminal region may not be required for transcript elongation by RNA polymerase II. Indeed, deletion of the N-terminal 100 amino acid produced a mutant form of

TFIIIS that is fully functional (28). This is in agreement with the observation that the mouse 21 Kd C-terminal fragment is capable of stimulating RNA polymerase II transcription in a manner similar to that of wild-type protein (24).

The functional significance of the 180 residue C-terminal is indicated by the requirement of this entire region for both binding to RNA polymerase II and stimulation of transcript elongation (28). In this segment, the polymerase binding ability resides in the residues 100–230 (28). While this region is highly conserved in human and mouse proteins, the level of conservation may vary among species depending on evolutionary relationships exhibited by the cognate RNA polymerases. The low homology between vaccinia virus rpo30 protein and the human/mouse protein proposed as an elongation factor for polymerase, tends to support this idea (26). However, there is no biochemical evidence that the rpo30 protein serves as an elongation factor for vaccinia virus RNA polymerase, and hence caution should be exercised in accepting this proposal (26).

Stimulation of transcript elongation is exerted by the C-terminal Cys₂/Cys₂-motif zinc finger (28). This region shows absolute identity between the amino acid sequences of human and mouse proteins (residues 230–280), and in addition shows highly significant homology with the vaccinia virus rpo30 and yeast PPR2 proteins (25, 26). This high level of homology underscores the functional significance of zinc finger region in transcriptional stimulation. Assuming that the vaccinia virus rpo30 and yeast PPR2 proteins exert their influence at the level of transcript elongation, it should be possible to exploit this homology data in the precise definition of structure-function relationships. Studies along these lines are now in progress.

Like many transcription factors, the human TFIIIS protein exists as a dimer in solution. On the basis of the nucleic acid-binding property of the zinc finger region, we have recently proposed a model for transcription stimulation with involvement of monomer form of TFIIIS (28). It is conceivable that the region used in dimer formation is also used in the binding to polymerase. Studies along these lines are also in progress.

ACKNOWLEDGEMENTS

The authors thank Marsha Rosner and Herbert Friedmann for critical reading of the manuscript and Hoen Kim for preparation of the figures. They also thank Diane Kahng for preparation of the manuscript. This research was supported by the National Institute of Health grant DK 21901.

REFERENCES

- Johnson, P.F. and McKnight, S.L. (1989). *Ann. Rev. of Biochem.* **58**, 799–839.
- Lewin, B. (1990). *Cell* **61**, 1161–1164.
- Natori, S. (1982). *Mol. Cell. Biochem.* **46**, 173–187.
- Reinberg, D., and Roeder, R.G. (1987). *J. Biol. Chem.* **262**, 3331–3337.
- SivaRaman, L., Reines, D., and Kane, C.M. (1990). *J. Biol. Chem.* **265**, 14554–14560.
- Sawadogo, M., Sentenac, A., and Fromageot, P. (1980). *J. Biol. Chem.* **255**, 12–15.
- Sluder, A.E., Greenleaf, A.L., and Price, D.H. (1989). *J. Biol. Chem.* **264**, 8963–8969.
- Sekimizu, K., Nakanishi, Y., Mizuno, D., and Natori, S. (1979). *Biochemistry* **18**, 1582–1588.
- Rappaport, J., Reinberg, D., Zandomeni, R., and Weinmann, R. (1987). *J. Biol. Chem.* **262**, 5227–5232.
- Rappaport, J., Cho, K., Saltzman, A., Prenger, J., Golomb, M., and Weinmann, R. (1988). *Mol. Cell. Biol.* **8**, 3136–3142.
- Sawadogo, M., and Sentenac, A. (1990). *Ann. Rev. Biochem.* **59**, 711–754.

12. Corden, J.L. (1990). *TIBS* **15**, 383–387.
13. Hirashima, S., Hirai, H., Nakanishi, Y., and Natori, S. (1988). *J. Biol. Chem.* **263**, 3858–3863.
14. Hodo, H.G., and Blatti, S.P. (1977). *Biochemistry* **241**, 2306–2310.
15. Kerppola, T.K., and Kane, C.M. (1988). *Mol. Cell. Biol.* **8**, 4389–4394.
16. Baek, K.H., Sato, K., Ito, R., and Agarwal, K. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 7623–7627.
17. Kadesch, T.R., and Chamberlain, M.J. (1982). *J. Biol. Chem.* **257**, 5286–5295.
18. Sambrook, J., Fritsch, F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
19. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J., and McKnight, S.L. (1988). *Genes. Dev.* **2**, 786–800.
21. Partis, M.D., Griffiths, D.G., Roberts, C.G., and Beechy, R.B. (1983). *J. Prot. Chem.*, **2**, 263–277.
22. Ouchterlony, O. (1958). *Prog. Allergy* **5**, 1–78.
23. McCauley, R., and Racker, E. (1973). *Mol. Cell. Biochem.* **1**, 73–81.
24. Horikoshi, M., Sekimizu, K., and Natori, S. (1985). *J. Biol. Chem.* **259**, 608–611.
25. Davies, C.J., Trgovcich, J., and Chamberlain, M.J. (1990). *Nature* **345**, 298.
26. Ahn, B-Y., Gershon, P.D., Jones, E.V., and Moss, B. (1990). *Mol. Cell. Biol.*, **10**, 5433–5441.
27. Studier, F.W., and Moffatt, B.A. (1986). *J. Mol. Biol.*, **189**, 113–130.
28. Agarwal, K., Baek, K-H., Jeon, C-J., Miyamoto, K., Ueno, A., and Yoon, H-S. (1991). *EMBO J.* In press.
29. Marshall, T.K., Guo, H., and Price, D.H. (1990). *Nucleic Acids Res.* **18**, 6293–6298.