Received 15 July; accepted 30 September 1991.

1. Kawakami, K., Scheidereit, C. & Roeder, R. *Proc. natn. Acad. Sci. U.S.A.* **85,** 4700–4704 (1988).
2. Baeuerle, P. A. & Baltimore, D. *Genes Dev.* **3,** 1689–1699 (1989).
3. Lenardo, M. J. & Baltimore, D. *Cell* **58,** 227–229 (1989).
4. Kieran, M. *et al. Cell* **62,** 1007–1018 (1990).
5. Ghosh, S. *et al. Cell* **62,** 1019–1029 (1990).
6. Nolan, G. *et al. Cell* **64,** 961–969 (1991).
7. Ruben, S. M. *et al. Science* **251,** 1490–1492 (1991).
8. Riviere, Y. *et al. Nature* **350,** 625–626 (1991).
9. Munro, S. & Pelham, H. R. B. *Cell* **48,** 899–907 (1987).
10. Field, J. *et al. Molec. cell. Biol.* **8,** 2159–2165 (1988).
11. Lenardo, M. J., Fan, C.-M., Maniatis, T. & Baltimore, D. *Cell* **57,** 287–294 (1989).
12. Fan, C.-M. & Maniatis, T. *EMBO J.* **8,** 101–110 (1989).
13. Hershko, A. *J. biol. Chem.* **263,** 15237–15240 (1988).
14. Rechsteiner, M. *Cell* **66,** 615–618 (1991).
15. Haskill, S. *et al. Cell* **65,** 1281–1289 (1991).
16. Ballard, D. W. *et al. Cell* **63,** 803–814 (1990).
17. Inoue, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **88,** 3715–3719 (1991).
18. Tanaka, M. & Herr, W. *Cell* **60,** 375–386 (1990).
19. Harlow, E. & Lane, D. *Antibodies, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1988).

ACKNOWLEDGEMENTS. We thank A. Michelson for discussions and comments on this work and S. Abmayr and L. Whittemore for critical reading of the manuscript. This work was supported by an NIH grant to T.M.

Structural motifs and potential σ homologies in the large subunit of human general transcription factor TFIIE

Yoshiaki Ohkuma, Hideki Sumimoto, Alexander Hoffmann, Shunichi Shimasaki*, Masami Horikoshi & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, New York 10021, USA
* Department of Molecular Endocrinology, The Whittier Institute for Diabetes and Endocrinology, La Jolla, California 92037, USA

THE general transcription factor TFIIE has an essential role in eukaryotic transcription initiation together with RNA polymerase II and other general factors[1,2]. Human TFIIE consists of two subunits of relative molecular mass 57,000 (TFIIE-α) and 34,000 (TFIIE-β)[3–5] and joins the preinitiation complex after RNA polymerase II and TFIIF (ref. 5). Here we report the cloning and structure of a complementary DNA encoding a functional human TFIIE-α. TFIIE-α is necessary for transcription initiation together with TFIIE-β, and recombinant TFIIE-α can fully replace the natural subunit in an *in vitro* transcription assay. The sequence contains several interesting structural motifs[6] (leucine repeat, zinc finger and helix–turn–helix) and sequence similarities to bacterial σ factors that suggest direct involvement in the regulation of transcription initiation.

Partial amino-acid sequences of human TFIIE-α was determined from five cyanogen bromide peptides (underlined regions in Fig. 1). With primers based on these sequences, polymerase chain reaction (PCR) and plaque hybrydization techniques were used to isolate partial TFIIE-α cDNA clones from human placenta[7] and Namalwa cell[8] cDNA libraries. One (p2EYC) was found to contain a 1.7-kilobase (kb) insert with all of the peptide sequences derived from the purified TFIIE-α present in a long open reading frame (ORF). The putative ORF encodes a 439-amino-acid polypeptide with a calculated relative molecular mass (M_r) of 49,562 (Fig. 1).

Northern analysis with the TFIIE-α cDNA probe revealed a 3.2-kb transcript (data not shown), indicative of a unique TFIIE-α messenger RNA with long untranslated regions. Genomic Southern analysis showed single hybridizing DNA fragments (in *Bam*HI and *Pst*I digests), indicating that the human TFIIE-α is encoded by a single-copy gene (data not shown).

To demonstrate that the cloned cDNA encodes a functional TFIIE-α, the basal transcription activity of expressed proteins was tested with a reconstituted transcription system[3,9]. A polypeptide of apparent M_r 57K, which shows the same mobility as purified TFIIE-α, was generated by *in vitro* translation of cDNA-derived transcripts in a reticulocyte lysate[10] (Fig. 2*a*) or by expression of the cDNA in bacteria[11] (data not shown). In previous studies[3,5] natural TFIIE-α displayed basal level transcription in a system containing partially purified human native TFIIB and TFIIG. But when assayed with the recombinant TFIIB (ref. 12) and HPLC-purified TFIIG (ref. 13), both the bacterially expressed and the natural α subunit showed transcription activity in the presence of the natural β subunit (Fig. 2*b*, lanes 4 and 6), whereas neither the α subunit (lanes 3 and 5) nor the β subunit (lane 7) alone showed any activity. These results demonstrate that the isolated clone encodes a functional TFIIE-α and that both α and β subunits are necessary for TFIIE function in transcription initiation.

As it is believed that TFIIE acts at a very late step in transcription initiation[1,2,5], it is instructive to consider sequences or motifs that might be involved in protein–protein or DNA–protein interactions. The deduced amino-acid sequence of TFIIE-α contains

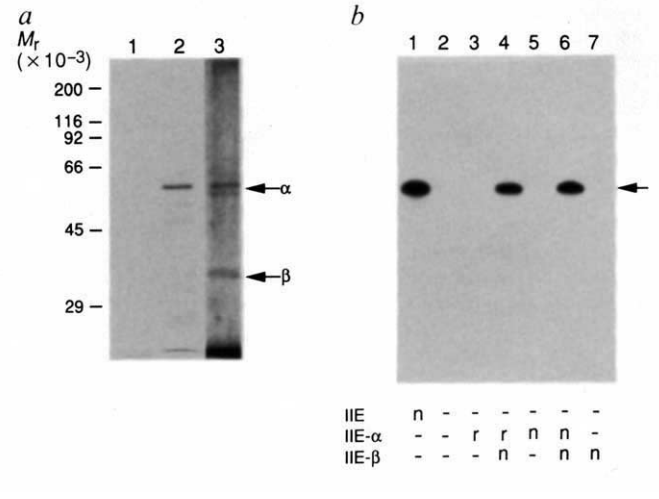

FIG. 2 Expression and function of cloned TFIIE-α. *a*, SDS-PAGE of [^{35}S]methionine-labelled TFIIE-α produced in a reticulocyte lysate. *In vitro* translation was done with no RNA (lane 1) or cloned TFIIE-α-derived mRNA (lane 2). Heparin-HPLC-purified TFIIE-α was run at the same time and silver-stained (lane 3). Arrows, positions of TFIIE-α and β. *b*, *In vitro* transcription by expressed recombinant TFIIE-α. Reactions containing the adenovirus major late promoter template, pML(C_2AT)Δ-50 (ref. 13) and general transcription factors TFIIB, TFIID, TFIIF, TFIIG and RNA polymerase II (refs 3, 13) were complemented with natural TFIIE (lane 1), buffer only (lane 2), bacterially expressed recombinant TFIIE-α (lane 3), recombinant α and renatured natural TFIIE-β (lane 4), renatured natural α (lane 5), renatured natural α and β (lane 6), and renatured natural β (lane 7). *r*, *n*, Recombinant and natural proteins, respectively. Arrow, position of the specific transcript. METHODS. A human TFIIE-α cDNA (nucleotides 1–1,599) in pBluescript SK(−) was transcribed with T7 RNA polymerase (Stratagene) and derived RNA was translated with rabbit reticulocyte lysate (Promega). The TFIIE-α expression vector was constructed as follows. An *Nde*I site (5′-CATATG) was made at the translation start site by site-directed mutagenesis of the p2EYC cDNA. The *Nde*I and *Bam*HI fragment was cloned into the pET3a plasmid[11]. TFIIE-α was expressed in E. *coli* BL21(DE3)pLysS (ref. 11) by induction with IPTG (isopropylthiogalactoside). Lysates were prepared by sonication of cells in the buffer containing 20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 20 mM 2-mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride, 20 μg ml^{-1} leupeptin, 20 μg ml^{-1} pepstatin and 10% glycerol. TFIIE-α was partially purified through two columns (DEAE-cellulose and heparin-Sepharose). Transcription assay was performed as described[3], except bacterially expressed human TFIIB (ref. 12) (20 ng) and TFIID (ref. 30) (30 ng), HPLC-purified TFIIG (ref. 13) and a saturating amount of recombinant TFIIE-α (1 ng) were used.

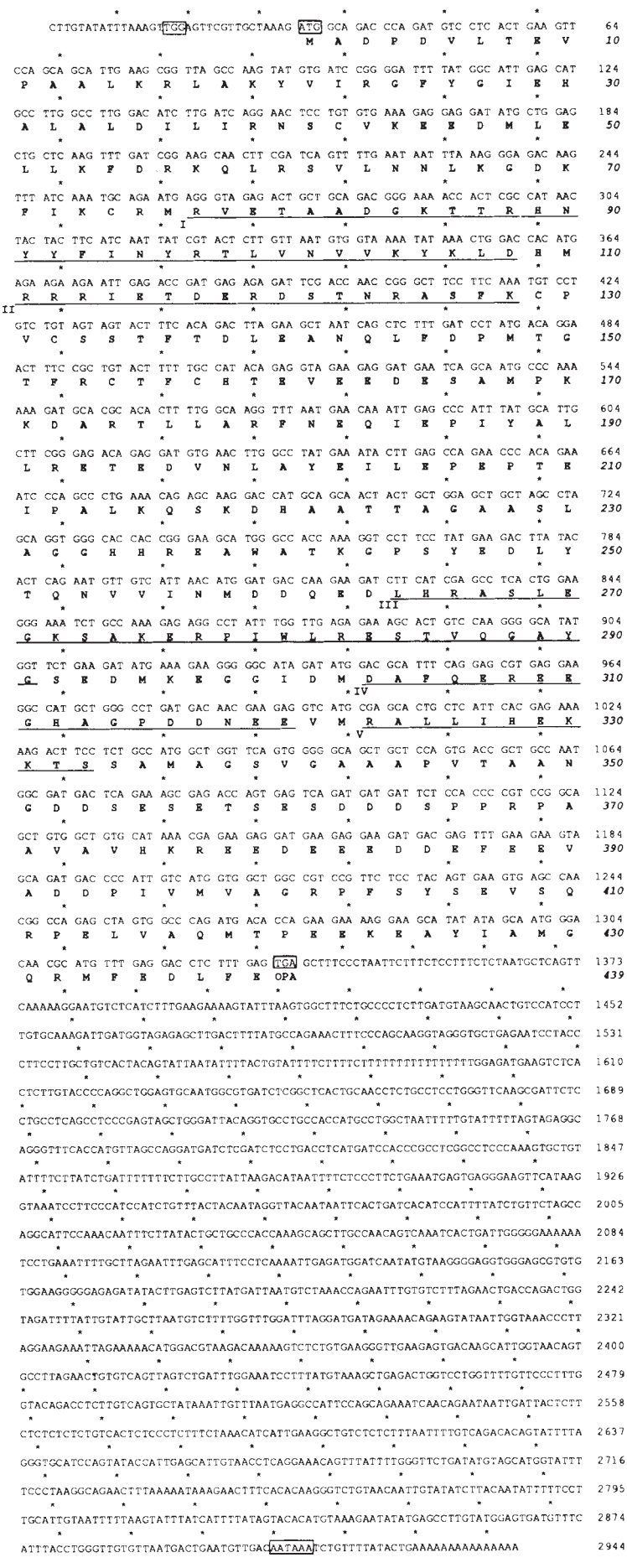

FIG. 1 Nucleotide sequence of TFIIE-α cDNA and its deduced amino-acid sequence (single-letter code). Amino-acid sequences matching CNBr-polypeptides (I–V) are underlined. Translation start, stop codons and a potential polyadenylation signal sequence are boxed. The translation start sequence matched the Kozak consensus (nucleotide A in position −3 and G in +4)[29].

METHODS. Reversed-phase HPLC-purified TFIIE-α (ref. 3) (40 μg) was digested with CNBr and some of the derived fragments (I–V) were sequenced. These sequences (with uncertain residues indicated by X or in parentheses) were: I, NH_2-RVETAADGKT(T/S)RHNYYFINYRTLVNVVKYKLD-COOH; II, NH_2-(R)X(R)IETDERD(S/T)XNXASX(K)-COOH; III, NH_2-(L/E)-HRASLEGKSAKERPIXLRE (S)(T)(V)(Q)(G)(A)(Y)(G)-COOH; IV, NH_2-DAFQEREEGHAGPDDNEE-COOH; V, NH_2-RALLIHEKKTS-COOH. For the first PCR, the sense and antisense primers based on fragment IV residues 1–6 and 13–18 were 5′-GA(C/T)GC(A/G/C/T)TT(C/T)CA(A/G)GA(A/G)(A/C)G-3′ and 3′-GG(A/G/C/T)CT(A/G)TT(A/G)CT(C/T)CT(C/T)-5′, respectively. A 54-bp PCR fragment was obtained. The internal antisense sequence (3′-CCCAGCATGGCCTTCCTC-5′), which corresponds to fragment IV residues 7–12, was used as a primer for the second PCR together with the sense primer 5′-GA(A/G)GG-IAA(A/G)(A/T)(G/C)IGCIAA(A/G)GA-3′ based on fragment III residues 6–12. Inosine (I) was used instead of a four-nucleotide mixture. A 135-bp fragment was obtained by PCR amplification using a human Namalwa cDNA library[8] (both oligo(dT) and random primed). This fragment was used to screen a human placenta cDNA library[7] (oligo(dT)-primed). A partial cDNA (p2EYA) was obtained whose 5′-end (a 232-bp *Eco*RI–*Nsi*I fragment) was used to isolate clones which contained the N-terminal region of TFIIE-α from the Namalwa cDNA library. One clone (p2EYC) was found to encode all of the peptide sequences. Screening was done under standard high-stringency conditions[26].

N-terminal regions with sequence similarity to bacterial σ factors[14] (Fig. 3*b*), as well as three potential structural motifs[6] (leucine repeat, zinc finger, and helix-turn-helix (HTH)) that are characteristic of many transcription factors (Fig. 3*a*). For each σ factor, the region with the greatest similarity to TFIIE-α is the 2.1 region, which has been suggested to be involved in single-stranded DNA binding[14]. The basic and aromatic residues suggested to be important for this σ-factor function are well conserved (and even enriched) in TFIIE-α; TFIIE-α contains a tyrosine at the position where a serine to tyrosine mutation enhances the activity of σ^{70} (residue 6 of the 2.1 domain)[14]. σ^{70} sequences N-terminal to and including the 2.1–2.2 region, as well as a homologous region in RAP30 (the small subunit of TFIIF), have a bacterial core RNA polymerase-binding activity[15]. TFIIE-α also has sequence similarity to these regions, consistent with the finding that partially purified TFIIE can bind directly to RNA polymerase II (ref. 16). A region of TFIIE-α that overlaps both the σ 2.1–2.2 similarity region and part of the leucine repeat (residues 13–60) also has a high sequence similarity to the σ 4.1–4.2 region (Fig. 3*c*). Residues 33–60 have sequence similarity to the σ 4.2 region and contain a putative helix-turn-helix (HTH) structure. This σ 4.2 region (especially the HTH region) has been postulated to contact the −35 region of bacterial promoter sequences[14]. Although the extent of sequence similarity in this HTH region is not so high (especially in the second helix region, which is usually important for DNA-binding specificity), this HTH region might have a different DNA-binding activity from that of the σ 4.2 region. The overall σ-similarity region might have either DNA-binding activity (to either single-stranded or duplex DNA) or RNA polymerase-binding activity.

C-terminal to the putative σ homology there is a putative amphipathic α helix with a leucine repeat (residues 38–66) that could be involved in homomeric and/or heteromeric interactions

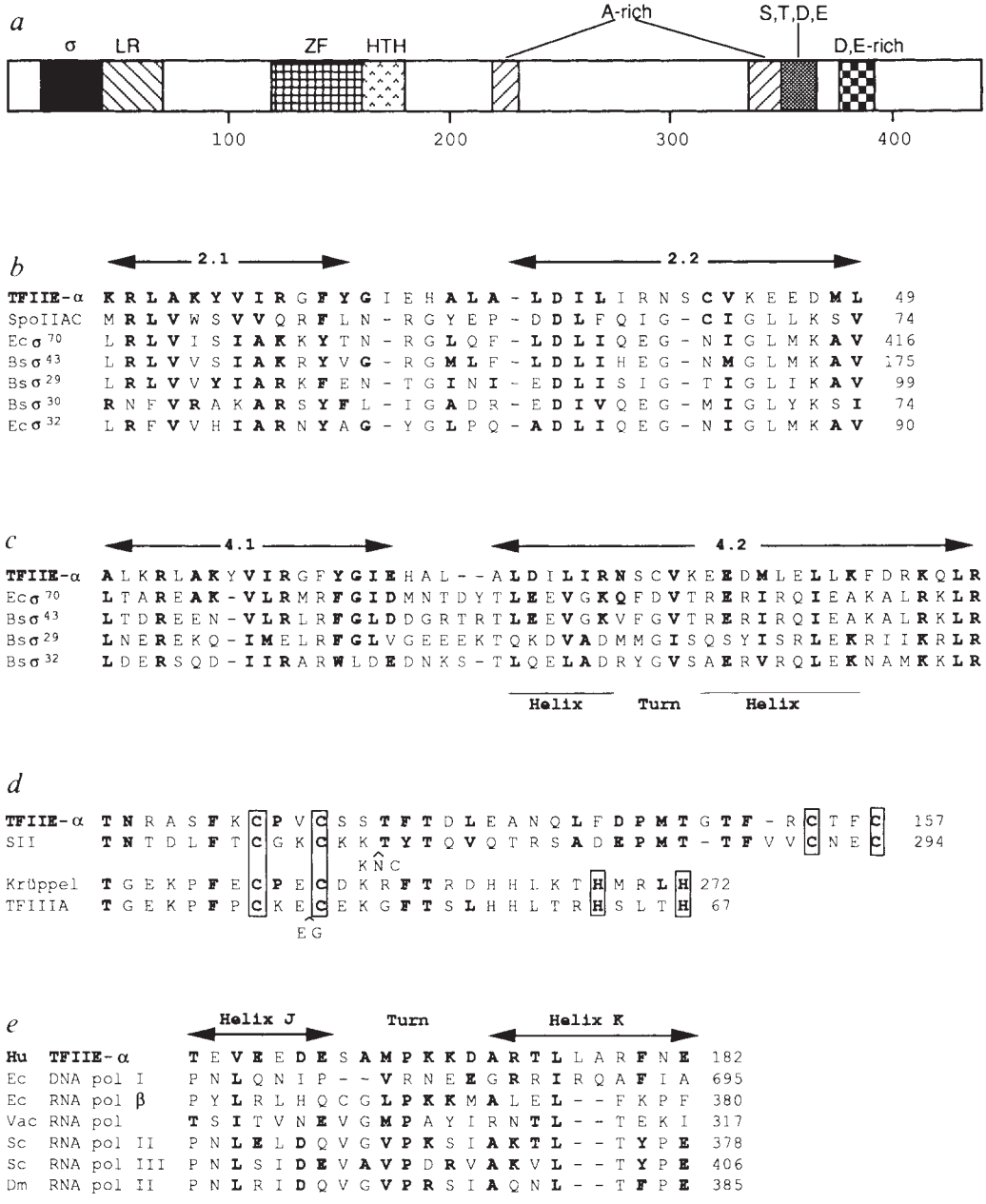

FIG. 3 Structural features of TFIIE-α. *a*, Schematic diagram of representative structures found in TFIIE-α. σ, Similarity to bacterial σ factors; LR, leucine repeat; ZF, zinc-finger; HTH, helix-turn-helix; A-rich, alanine-rich; S, T, D, E, region consists of serine, threonine, aspartic acid and glutamic acid; D, E-rich, acidic region. *b*, Alignment with bacterial σ-factor regions 2.1–2.2. *E. coli* (Ec) and *Bacillus subtilis* (Bs) σ-factor sequences are from ref. 13. SpoIIAC is from a *B. subtilis* phage. The position in the protein sequence of the last residue in each line is indicated in the last column. Amino acids identical (or conserved) to those in TFIIE-α are in bold according to the grouping (E, D), (K, R, H), (N, Q), (G, P), (A, I, L, V, M), (F, W, Y), (C) and (S, T). The most highly conserved domain has similarity to the σ-factor 2.1 region and the first half of the 2.2 region. *c*, Alignment with bacterial σ-factor regions 4.1–4.2. The region which completely overlaps with the above mentioned σ 2.1–2.2 region and partially overlaps with the leucine-repeat region also has good similarity to the σ-factor regions 4.1–4.2. The HTH region is underlined. *d*, Zinc-finger. Transcription elongation factor S-II (ref. 19) has the strongest similarity to TFIIE-α and both polypeptides possess the C_2C_2 motif (boxed). TFIIE-α also contains well conserved residues found in the 5S gene-specific TFIIIA and the *Drosophila* developmental protein Krüppel (boxed C_2H_2 motifs). *e*, HTH. Helix J and helix K are helices found initially in *E. coli* DNA polymerase I. Hu, human; Ec, *E. coli*; Vac, vaccinia virus; Sc, *Saccharomyces cereviciae*, Dm, *Drosophila melanogaster*.

with other leucine repeat-containing transcription factors[6] (Fig. 3*a*)[13]. The second motif is a putative zinc-finger containing several residues that are conserved in the C_2H_2 (two cysteine and two histidine) zinc-finger structures present in the 5S gene-specific TFIIIA (ref. 17) and the *Drosophila* Krüppel protein family[18] (Fig. 3*d*). But unlike these proteins, which contain multiple fingers of the C_2H_2 class, TFIIE-α contains a single putative zinc-finger of the C_2C_2 (two cysteine pairs) class. Thus the structure and function of the putative TFIIE-α zinc-finger might be more similar to that of the single zinc-finger in transcription elongation factor S-II (ref. 19), suggesting an involvement either in DNA-strand opening or protein–protein interactions. The third motif is a HTH with sequence similarity (and probable homology) to the largest subunit of both prokaryotic and eukaryotic RNA polymerases and *Escherichia coli* DNA polymerase I (refs 20, 21) (Fig. 3*e*). This region has been proposed as a DNA-binding region which interacts through the major groove of duplex DNA[20,22], although there is little similarity to the 'classical' HTH motif which is contained, for instance, in homeodomain proteins.

All four of these conspicuous motifs reside in the N-terminal 182 residues of TFIIE-α and can be grouped in two domains. The first domain (residues 13–66), consisting of the 'σ' and 'leucine repeat' regions (Fig. 3*a*, *b*, *c*), is quite basic (pI 9.5). This domain might form a novel basic region/leucine zipper structure, with the σ region replacing the usual basic region, which might recognize (perhaps alternately) either duplex or single-stranded DNA. The second domain (residues 121–182), containing the adjacent 'zinc-finger' and 'HTH' motifs (Fig. 3*a*, *d*, *e*), could bind to double-stranded DNA. Both domains might be important for transcription initiation by aiding the DNA-melting reaction. In contrast, no known motifs were identified in the C-terminal half of TFIIE-α, which might therefore contain novel functional interaction domains.

Note that TFIIE-α (mainly the C-terminal 300 residues) is highly acidic (pI 4.5) with several acidic clusters[23]. By contrast, TFIIE-β, the smaller subunit of TFIIE, is highly basic (pI 9.5)[24], suggesting that the two subunits could easily complex by ionic interactions. TFIIE-α may also interact with other positively charged transcription factors such as initiation factors TFIID (refs 25, 26) and TFIIB (ref. 27).

Sequence similarities of TFIIE to structural motifs found in the essential regions of RNA polymerases[2,28] and other transcription factors, regions of similarity with bacterial σ factors, and the presence of acidic clusters suggest various potential interactions with other transcription factors. With the above structural motifs in mind, it will be intriguing to examine how TFIIE functions, through protein–protein and DNA–protein interactions, during preinitiation complex formation and activation (including transitions from a closed to an open complex and from initiation to elongation). □

Received 5 August; accepted 14 October 1991.

1. Saltzman, A. G. & Weinmann, R. *FASEB J.* **3,** 1723–1733 (1989).
2. Sawadogo, M. & Sentenac, A. *A. Rev. Biochem.* **59,** 711–754 (1990).
3. Ohkuma, Y., Sumimoto, H., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87,** 9163–9167 (1990).
4. Conaway, J. W., Hanley, J., Garrett, K. P. & Conaway, R. C. *J. biol. Chem.* **266,** 7804–7811 (1991).
5. Inostroza, J., Flores, O. & Reinberg, D. *J. biol. Chem.* **266,** 9304–9308 (1991).
6. Johnson, P. F. & McKnight, S. L. *A. Rev. Biochem.* **58,** 799–839 (1989).
7. Hirata, M. *et al. Nature* **349,** 617–620 (1991).
8. Scheidereit, C. *et al. Nature* **336,** 551–557 (1988).
9. Sawadogo, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **82,** 4394–4398 (1985).
10. Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P. A. & Roeder R. G. *Cell* **61,** 1171–1178 (1990).
11. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185,** 60–89 (1990).
12. Malik, S. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
13. Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **87,** 9158–9162 (1990).
14. Helmann, J. D. & Chamberlin, M. J. *A. Rev. Biochem.* **57,** 839–872 (1988).
15. McCracken, S. & Greenblatt, J. *Science* **253,** 900–902 (1991).
16. Flores, O., Maldonado, E. & Reinberg, D. *J. biol. Chem.* **264,** 8913–8921 (1989).
17. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4,** 1609–1614 (1985).
18. Rosenberg, U. B. *et al. Nature* **319,** 336–339 (1986).
19. Hirashima, S., Hirai, H., Nakanishi, Y. & Natori, S. *J. biol. Chem.* **263,** 3858–3363 (1988).
20. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. *Cell* **42,** 599–610 (1985).
21. Broyles, S. S. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83,** 3141–3145 (1986).
22. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steiz, T. A. *Nature* **313,** 762–766 (1985).
23. Ptashne, M. *Nature* **335,** 683–689 (1988).
24. Sumimoto, H. *et al. Nature* **354,** 401–404 (1991).
25. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. *Cell* **54,** 665–669 (1988).
26. Horikoshi, M. *et al. Nature* **341,** 299–303 (1989).
27. Lin, Y.-S. & Green, M. R. *Cell* **64,** 971–981 (1991).
28. Woychik, N. A. & Young, R. A. *Trends biol. Sci.* **15,** 347–351 (1990).
29. Kozak, M. *Cell* **44,** 283–292 (1986).
30. Hoffmann, A. *et al. Nature* **346,** 387–390 (1990).

ACKNOWLEDGEMENTS. We thank R. Kovelman for critical reading of the manuscript, F. Katagiri, B. Kirschbaum, H. Murakami, H. Sabe and J. Sukegawa for technical advice, Y. Hayashi, R. Kageyama and S. Nakanishi for the human placenta cDNA library, S. Malik and K. Hisatake for recombinant human TFIIB and J. Wang for DNA sequencing. Y.O. is supported by a fellowship from the Human Frontiers Scientific Program Organization and A.H. by a fellowship from the Boehringer Ingelheim Fund. M.H. is an Alexandrine and Alexander L. Sinsheimer Scholar. This study was supported by NIH grants (to R.G.R. and M.H.) and by the Pew Trusts to The Rockefeller University.

Conserved sequence motifs in the small subunit of human general transcription factor TFIIE

Hideki Sumimoto, Yoshiaki Ohkuma, Eric Sinn, Hiroyuki Kato, Shunichi Shimasaki*, Masami Horikoshi & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology,
The Rockefeller University, New York, New York 10021, USA
* Department of Molecular Endocrinology,
The Whittier Institute for Diabetes and Endocrinology,
La Jolla, California 92037, USA

A GENERAL initiation factor, TFIIE, is essential for transcription initiation by RNA polymerase II in conjunction with other general factors[1,2]. TFIIE is a heterotetramer containing two subunits of relative molecular mass 57,000 (TFIIE-α) and two of 34,000 (TFIIE-β)[3,4]. TFIIE-β is required in conjunction with TFIIE-α for transcription initiation. Here we report the cloning and expression of a complementary DNA encoding a functional human TFIIE-β. Recombinant TFIIE-β could replace the natural TFIIE-β for transcription in conjunction with TFIIE-α. Amino-acid sequence comparisons reveal regions with sequence similarities to: subregion 3 of bacterial σ factors[6]; a region of RAP30 (the small subunit of TFIIF) with sequence similarity to a σ-factor subregion implicated in binding to RNA polymerase[7]; and a portion of the basic region–helix–loop–helix motif found in several enhancer-binding proteins[8–10]. These potential homologies have implications for the role of TFIIE in preinitiation complex assembly and function.

Four cyanogen bromide (CNBr)-digested peptides (I–IV) of TFIIE-β purified from HeLa cells[4] were separated by reversed-phase HPLC and their amino-acid sequences determined. With primers based on these sequences, amplification of a human cDNA fragment by polymerase chain reaction (PCR) yielded a 289-base-pair fragment containing the expected sequences. This fragment was used to screen a human cDNA library derived from Namalwa cells[11] and several overlapping clones were isolated and sequenced. Sequences encoding the four CNBr-derived peptides were found in an open reading frame of 873 nucleotides which encodes a polypeptide of 291 amino acids with a calculated relative molecular mass of 33,040 (Fig. 1). An RNA blot indicated that the human TFIIE-β messenger RNA is around 1.6 kilobases (kb), close to the length of the cDNA, and no difference in size was found between the Namalwa and HeLa cell mRNAs (H.S. and E.S., unpublished results). Only one major cross-hybridizing restriction fragment was detected in human genomic DNA digested with *Hin*dIII (H.S. and E.S., unpublished results), suggesting that there is one copy of the TFIIE-β gene per genome.

To assess the function of the cDNA-encoded protein, it