ments designed to test this three-cell model for tolerance induction are under way. At present, none of the above models can be ruled out.

Three other aspects of the data warrant comment. First, the minute quantities of cells ($CD8^+$ cells) required to induce Mls^a tolerance contrast with the relatively massive numbers of cells (5×10^7 spleen cells) needed to induce MHC tolerance in the classic system of Billingham *et al.* (*2*). The simplest explanation for this discrepancy is that tolerance to Mls^a and MHC antigens is controlled by different cells: whereas Mls^a tolerance is controlled by a common cell type ($CD8^+$ cells), full induction of MHC tolerance probably requires contact with dendritic cells, a relatively rare cell type (*1*). Whatever the explanation, the ~1000-fold difference in cell numbers required for Mls^a versus MHC tolerance indicates that the rules governing MHC tolerance induction do not necessarily apply to other antigens. Second, the finding that $CD8^+$ cells are highly effective at inducing Mls^a tolerance but are nonstimulatory for mature T cells indicates that the cell types controlling tolerogenicity and immunogenicity can be distinctly different. Third, the potency of $CD8^+$ cells in mediating Mls^a tolerance supports the view (*19*) that these cells may play a critical role in inducing tolerance to various self components.

REFERENCES AND NOTES

1. J. Sprent and S. Webb, *Adv. Immunol.* **41**, 39 (1987).
2. R. E. Billingham, L. Brent, P. B. Medawar, *Nature* **172**, 603 (1953).
3. H. Festenstein, *Transplant. Proc.* **8**, 339 (1976); C. A. Janeway, Jr., and M. E. Katz, *J. Immunol.* **134**, 2057 (1985); R. Abe and R. J. Hodes, *Annu. Rev. Immunol.* **7**, 683 (1989).
4. S. R. Webb and J. Sprent, *Immunol. Rev.* **107**, 141 (1989).
5. J. W. Kappler, U. Staerz, J. White, P. C. Marrack, *Nature* **332**, 35 (1988); H. R. MacDonald *et al.*, *ibid.*, p. 40; R. Abe, M. Vacchio, B. Fox, R. Hodes, *ibid.* **335**, 827 (1988); A. M. Pullen, P. Marrack, J. W. Kappler, *ibid.*, p. 796; A. M. Fry and L. A. Matis, *ibid.*, p. 830.
6. O. Kanagawa, E. Palmer, J. Bill, *Cell. Immunol.* **119**, 412 (1989).
7. H. von Boehmer and J. Sprent, *Nature* **249**, 363 (1974); A. Ahmed, I. Scher, A. H. Smith, K. W. Sell, *J. Immunogenet. (Oxford)* **4**, 201 (1977); I. J. Molina *et al.*, *J. Immunol.* **143**, 39 (1989).
8. S. R. Webb, J. Hu Li, D. Wilson, J. Sprent, *Eur. J. Immunol.* **115**, 92 (1985).
9. S. R. Webb, A. Okamoto, Y. Ron, J. Sprent, *J. Exp. Med.* **169**, 1 (1989).
10. S. MacPhail, S. T. Ishizaka, M. J. Bykowsky, E. C. Lattime, O. Stutman, *J. Immunol.* **135**, 2967 (1985); D. J. Waite, R. A. Miller, G. Sunshine, *Cell. Immunol.* **117**, 70 (1988); H. R. MacDonald *et al.*, *J. Exp. Med.* **167**, 2005 (1988).
11. D. J. Waite and G. H. Sunshine, *Cell. Immunol.* **117**, 78 (1988).
12. S. R. Webb and J. Sprent, unpublished data.
13. D. M. Lubaroff and W. K. Silvers, *J. Immunol.* **111**, 65 (1973); P. J. Morrissey, S. O. Sharrow, Y. Kohno, J. A. Berzofsky, A. Singer, *Transplantation* **40**, 68 (1985).
14. A. M. Pullen, J. W. Kappler, P. Marrack, *Immunol. Rev.* **107**, 125 (1989); D. E. Speiser, R. Schneider, H. Hengartner, H. R. MacDonald, R. Zinkernagal, *J. Exp. Med.* **170**, 595 (1989); F. Ramsdell, T. Lantz, B. J. Fowlkes, *Science* **246**, 1038 (1989).
15. L. Hudson and J. Sprent, *J. Exp. Med.* **143**, 444 (1976); E. Nagy, M. Nabholz, P. H. Krammer, B. Pernis, *ibid.*, p. 648; I. Melchers and H. O. McDevitt, in *Regulatory T Lymphocytes*, B. Pernis and H. J. Vogel, Eds. (Academic Press, New York, 1980), pp. 91–117.
16. P. J. Fink, R. P. Shimonkevitz, M. J. Bevan, *Annu. Rev. Immunol.* **6**, 115 (1988).
17. S. R. Webb, unpublished observations.
18. M. E. Katz and C. A. Janeway, Jr., *J. Immunol.* **134**, 2064 (1985).
19. R. P. Shimonkevitz and M. J. Bevan, *J. Exp. Med.* **168**, 143 (1988).
20. M. Sarmiento, A. L. Glasebrook, F. W. Fitch, *J. Immunol.* **125**, 2665 (1980); R. Ceredig, J. Lowenthal, M. Nabholz, H. R. MacDonald, *Nature* **314**, 98 (1985); J. Bruce, F. W. Symington, T. J. McKearn, J. Sprent, *J. Immunol.* **127**, 2496 (1981); G. Dennert, R. Hyman, J. Leoley, I. Trowbridge, *Cell. Immunol.* **53**, 300 (1980).

21. We thank O. Kanagawa for providing the RR47 MAb, B. Marchand for preparation of the manuscript, and D. McQuitty, M. Rehse, and W. Wiltse for flow cytometry. Supported by grants CA 25803, CA 38355, and CA 41993 from the National Cancer Institute and AI 21487 from the National Institute of Allergy and Infectious Diseases. This is publication no. 6251-IMM from the Research Institute of Scripps Clinic.

2 February 1990; accepted 27 April 1990

Cloning of a Transcriptionally Active Human TATA Binding Factor

C. CHENG KAO, PAUL M. LIEBERMAN, MARTIN C. SCHMIDT, QIANG ZHOU, RUI PEI, ARNOLD J. BERK*

Transcription factor IID (TFIID) binds to the TATA box promoter element and regulates the expression of most eukaryotic genes transcribed by RNA polymerase II. Complementary DNA (cDNA) encoding a human TFIID protein has been cloned. The human TFIID polypeptide has 339 amino acids and a molecular size of 37,745 daltons. The carboxyl-terminal 181 amino acids of the human TFIID protein shares 80% identity with the TFIID protein from *Saccharomyces cerevisiae*. The amino terminus contains an unusual repeat of 38 consecutive glutamine residues and an X-Thr-Pro repeat. Expression of DNA in reticulocyte lysates or in *Escherichia coli* yielded a protein that was competent for both DNA binding and transcription activation.

INITIATION OF TRANSCRIPTION IN EUkaryotes by RNA polymerase II is a complex process that requires the orchestrated function of several factors (*1*). The binding of a protein, TFIID, to the TATA box promoter element is the first step in the assembly of the transcription complex at the promoter (*2*). Because it has been difficult to purify, the mammalian TFIID protein has been poorly characterized. However, the gene from *S. cerevisiae* that encodes the yeast TFIID protein (YIID) has been cloned (*3–6*), and the YIID protein can substitute for mammalian TFIID in transcription assays conducted in vitro in HeLa nuclear extracts (*7*).

YIID consists of 240 amino acids with a molecular size of ~27 kilodaltons (kD). The carboxyl terminus has a partially repeated sequence between amino acid residues 67 to 131 and 157 to 222 (*6, 8*), which may form a helix-turn-helix structure (*5*). Nearly perfect direct repeats occur between amino acid residues 109 to 127 (PKTTALIFASGKMVVTGAK) and 200 to 218 (PKIVLLIFVSGKIVLTGAK) (*3, 9*).

Although YIID can substitute for mammalian TFIID in an in vitro transcription assay, it is unclear why YIID responds to upstream activator proteins for some promoters (*10*), but not others (*11*). To understand fully the function of TFIID in regulating transcription in mammalian cells, we cloned and expressed the gene that encodes a mammalian TFIID protein.

Our approach for isolating the human TFIID gene was to identify regions in the protein that are highly conserved in evolution by determining the sequence of the TFIID gene from several species. We then designed oligonucleotide primers based on the conserved regions and used these primers to amplify a fragment of the human TFIID gene by the polymerase chain reaction (PCR). YIID DNA was used to probe digested human, mouse, *Drosophila*, and *Lemna* genomic DNA (*12*), and no specific hybridizing sequences were observed. We next looked for cross-hybridizing sequences in more closely related species of budding yeast. All 13 species that we examined had

C. C. Kao, P. M. Lieberman, M. C. Schmidt, Q. Zhou, A. J. Berk, Department of Microbiology, Molecular Biology Institute, University of California, Los Angeles, CA 90024–1570.

R. Pei, Department of Biological Chemistry, Molecular Biology Institute, University of California, Los Angeles, CA 90024–1570.

*To whom correspondence should be addressed.

Fig. 1. The sequence of *S. cerevisiae* (S. c.) TFIID residues 120 to 213 (*3*). The arrows indicate sequences to which oligonucleotide primers were synthesized as well as the orientation of the primers. The residues from *K. lactis* (K. l.), *P. heedii* (P. h.), *N. crassa* (N. c.), *D. melanogaster* (D. m.) between BK10 and BK12 that have diverged from YIID are indicated. DNA sequences and clones are available upon request.

```
       (BK10)                                                             (BK12)             (BK11)
      ------->                                                         <-------             <-------
S. c. KMVVTGAKSEDDSKLASRKYARIIQKIGFAAKFTDFKIQNIVGSCDVKFPIRLEGLAFSHGTFSSYEPELFPGLIYRMVKPKIVLLIFVSGKIV
K. l.                              S
P. h.                 A         L  N                           YA
N. c.                           L  N                I          SK HN                I
D. m.              R  A   RA    L  P   L  T   M         H     VHA CNL                R R
```

cross-hybridizing sequences, but the filamentous fungus *Neurospora crassa* did not. The hybridizing sequences from *Pichia heedii* and *Kluyveromyces lactis* were cloned and sequenced (*13*). The predicted amino acid sequences revealed that the NH_2-terminal 60 amino acids of *Pichia* and *Kluyveromyces* TFIID proteins diverged considerably from the *S. cerevisiae* YIID gene and from each other. However, the COOH-terminal 180 amino acids were greater than 95% identical. We designed oligonucleotides to several absolutely conserved regions. Two oligonucleotides, BK10 and BK11 (Fig. 1) (*14*), designed from the amino acid sequences of the direct repeats were included because these sequences were conserved among the three yeast species. We focused our efforts on BK10 and BK11 after we learned that the direct repeat is conserved between the human and yeast TFIID proteins (*15*).

Oligonucleotides BK10 and BK11 were used in PCR reactions (*16*) to amplify DNA fragments from *Neurospora* genomic DNA and from a *Drosophila* cDNA library. The PCR fragments were cloned and sequenced (*17*), and the predicted amino acid sequences had more than 80% identity to YIID (Fig. 1). However, with BK10 and BK11, we could not amplify by PCR a similar fragment from human cDNA. Comparison of the amino acid sequences encoded by the *Neurospora* and *Drosophila* PCR fragments and the three yeast TFIID genes revealed a highly conserved region to which we designed a third oligonucleotide, BK12 (Fig. 1). PCR reactions (*16*) were performed with BK10, BK11, and human cDNA. We then amplified a portion of the product with BK10 and BK12. The result was a 233–base pair (bp) fragment whose sequence encoded a polypeptide that was highly similar to YIID.

The 233-bp fragment was used to identify six plaques from a lambda ZAP HeLa cDNA phagemid library (*18*). The inserts from four of these plaques had similar restriction patterns, and sequencing confirmed that they contained overlapping inserts. The plasmid with the largest insert, pKB104, was sequenced (*17*). Its 1.7-kb insert contained a 1068-bp open reading frame (Fig. 2). Starting with the first methionine, the open reading frame encoded a protein of 339 amino acids with a calculated molecular size of 37,745 daltons.

To confirm that the gene we isolated encoded a functional TFIID protein, we expressed the protein in an in vitro translation system (*19*). Transcription of mRNA with bacteriophage T3 polymerase from linearized pKB104 plasmid was translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. A labeled protein with an apparent molecular size of 38 to 40 kD was observed (Fig. 3A, lane 2), which agreed with the predicted molecular size (37.8 kD). This polypeptide was not observed in a control extract (lane 1).

The TFIID protein translated in vitro was assayed for its ability to bind to a TATA box–containing oligonucleotide in a gel mobility shift assay (*20*). A nonspecific retarded band was observed regardless of the oligonucleotide probe used and whether or not the reticulocyte lysate contained in vitro–translated TFIID (Fig. 3B). However, a specific complex (B) was observed when reticulocyte extract programmed with TFIID message was incubated with ^{32}P-labeled oligonucleotides containing the E1B TATA box (T_{1B}) or the E4 TATA box (T_4), both from adenovirus type 2 (Fig. 3B, lanes 2, 4, and 7). The specific retarded complex was greatly reduced when we used an oligonucleotide with a single base pair change in the E1B TATA box (T_{16C}), converting the sequence TATATAA to TACATAA (lane 6). Furthermore, the specific complex with T_{1B} was competed by a large molar excess of unlabeled T_{1B} (lane 8) and T_4 (lane 9), but

A

```
1
MDQNNSLPPY AQGLASPNGA MTPGIPIFSP MMPYGTGLTP QPIQNTNSLS Hs

51
ILEEQQRQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQAVAAA Hs
                                                    M  Sc

101
AVQQSTSQQA TQGTSGQAPQ LFHSQTLTTA PLPGTTPLYP SPMTPMTPIT Hs
DEERLKEFKE ANKIVFDPNT RQVWENQNRD GTKPA TFQS EEDIKRATPE Sc

151
PATPASESSG IVPQLQNIVS TVNLGCKLDL KTIALRRRNA EYNPKRFAAV Hs
SEKDT AT      T     A   T   R      V  HA               Sc

201
IMRIREPRTT ALIFSSGKMV CTGAKSEEQS RLAARKYARV VQKLGFPAKF Hs
       K       A      V      DD  K  S ·   I I  I  A    Sc

251
LDFKIQNMVG·SCDVKFPIRL EGLVLTHQQF SSYEPELFPG LIYRMIKPRI Hs
T      I                 AFS GT                  V  K  Sc

301
VLLIFVSGKV VLTGAKVRAE IYEAFENIYP ILKGFRKTT            Hs
         I       Q E    Q   A    V SE  ·M              Sc
```

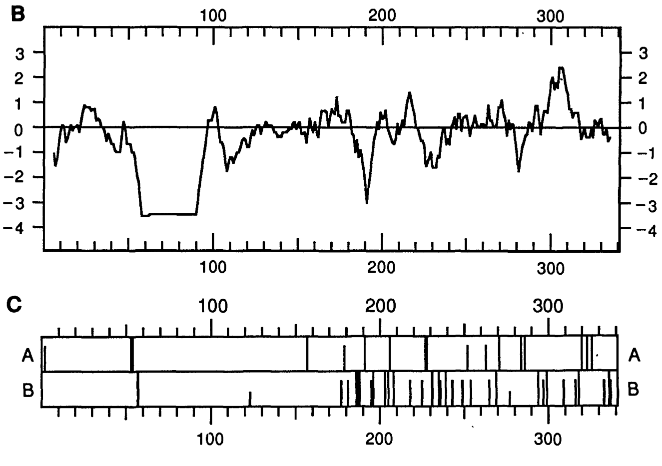

Fig. 2. Analysis of the human TFIID protein sequence. (**A**) Comparison of the amino acid sequence of YIID (Sc) and the predicted amino acid sequence of human (Hs) TFIID. The initiating methionine for YIID is at position 99. Positions at which YIID diverges from human TFIID are shown. The nucleotide sequence of the human TFIID gene can be found in GenBank under the accession number M34960. The sequence is also available upon request. (**B**) A linear plot of the hydrophobic and hydrophilic regions of the human TFIID protein by the algorithm of Kyte and Doolittle in the DNA Stryder sequence analysis program (*32*). (**C**) The amino acid charge profile of human TFIID. A, distribution of acidic residues; B, distribution of basic residues. Short bars denote amino acids D or K, while tall bars denote E or R (*9*).

was competed to a lesser extent with T_{16C} (lane 10). Thus, the polypeptide encoded by our human cDNA clone bound with high specificity to the E1B and E4 TATA boxes.

We next found that human recombinant TFIID expressed in *Escherichia coli* (*21*) can footprint a TATA box. A specific region of the E1B promoter that contained the TATA box was protected from deoxyribonuclease (DNase) I digestion by both YIID (Fig. 3C, lanes 2 to 4) and recombinant human TFIID (lanes 5 to 7). This protected region was not observed with the LS−30/−23 E1B promoter probe, which contains a linker-scanning mutation in the E1B TATA box (lanes 9 to 16). Unlike YIID, human TFIID showed a hypersensitive DNase I cleavage site at approximately nucleotide 73 (arrow) and more extensive protection 5′ to the TATA box (lanes 3 and 6). Because human TFIID has an extended amino terminus, we speculate that the amino terminus of the protein protects the sequence 5′ to the TATA box. At high YIID and human TFIID concentrations, protection of the region near −90 was also observed on both the wild-type E1B and LS−30/−23 probes (lanes 4, 7, 12, and 15). This footprint is centered over the AATAAA cleavage-polyadenylation signal of the upstream E1A transcription unit and was observed previously at high YIID concentrations (*3*). Also at high protein concentration, human TFIID partially protected sequences upstream of the LS−30/−23 linker insertion (lane 15). The DNase I footprint results demonstate that the polypeptide translated from pKB104 can bind specifically to the E1B TATA box sequence.

The reticulocyte-translated TFIID protein was functional in in vitro transcription reactions (Fig. 3D). Phosphocellulose B and C fractions of a HeLa cell nuclear extract that contained RNA polymerase II and general transcription factors and was depleted of TFIID activity were used to reconstitute transcription from the adenovirus E1B promoter (*22*). Specific transcription was not detectable in the absence of exogenously added TFIID. However, transcription was restored by addition of a partially purified human TFIID fraction (*3*), purified YIID (Fig. 3D, lanes 1 and 2), or in vitro–translated recombinant human TFIID (lanes 3 and 4), but not by addition of reticulocyte lysate without TFIID protein (lanes 5 and 6). These data demonstrate that human TFIID made in reticulocyte lysates is transcriptionally active. We also observed that human TFIID expressed in *E. coli* was active in in vitro transcription reactions with the E1B promoter (*23*).

The E1B promoter used in the in vitro transcription reactions (Fig. 3D) has a TATA box and a single GC box, which is a binding site for the transcription factor, Sp1 (*9, 24, 25*). The pAd1BLS−48/−39 plasmid (*9*), which contains a substitution in the GC box, was also used as template for the transcription assay (Fig. 3D). When transcription from the mutant was compared to the wild-type E1B template, we consistently observed a higher amount of transcription from the wild-type template. Transcriptional stimulation by the GC box was observed whether we used partially purified human TFIID fractions (*9*), highly purified YIID, or the in vitro–translated or *E. coli*–expressed human TFIID. It seems likely that the higher level of transcription observed with the wild-type promoter was a result of activation by Sp1 present in the partially purified preparations of RNA polymerase II and general transcription factors. However, it is also possible that the higher transcription observed with the wild-type template resulted from other factors interacting with the GC box.

The predicted amino acid sequence of human TFIID was compared to that of YIID (Fig. 2A). The NH_2-terminal 158 amino acids in the human TFIID protein were not similar to the YIID protein. The last 181 amino acids, however, were 80% identical to YIID. The divergent NH_2-terminal 158 amino acids of the TFIID protein contained several interesting motifs. First, there are only a few charged amino acids in the hydrophobic NH_2-terminus (Fig. 2, B and C). Between amino acids 55 and 95 is a glutamine-rich region, which contains 38 consecutive glutamine residues. The glutamine-rich region is followed by a string of seven residues which are either alanine or valine. The sequence between residues 125 and 160 is rich in serines, threonines, and prolines. The sequence X-Thr-Pro, where X represents methionine, isoleucine, or alanine, is repeated four consecutive times be-

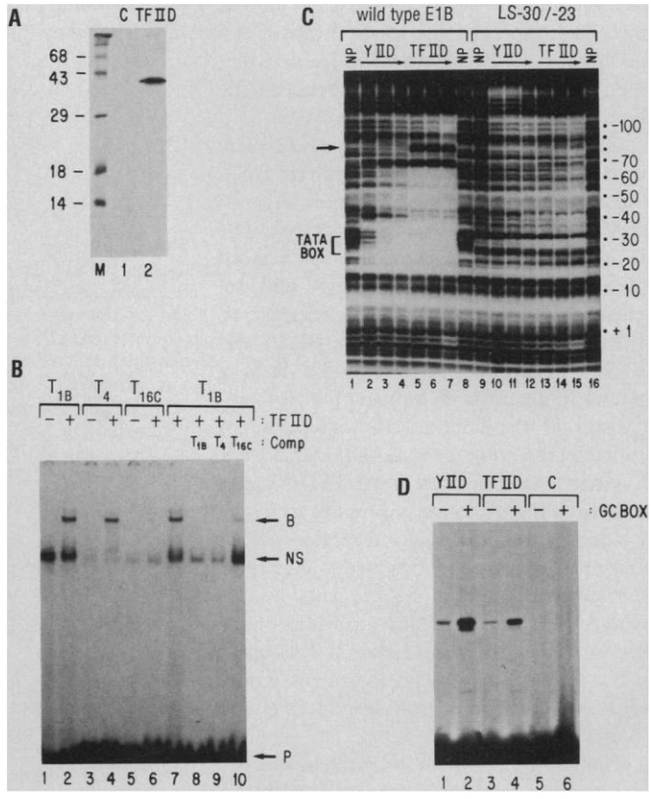

Fig. 3. Functional activity of in vitro–translated human TFIID. (**A**) Rabbit reticulocyte lysate programmed with in vitro–transcribed RNA from plasmid pKB104 (lane 2) or with no exogenous RNA (lane 1) were compared on a 15% SDS-polyacrylamide gel. [^{14}C]molecular size markers are indicated in kilodaltons. (**B**) Gel mobility shift assay with the in vitro–translated human TFIID. Rabbit reticulocyte lysates programmed with RNA from pKB104 (+) or with no exogenously added RNA (−) were incubated either with T_{1B}, T_4, or T_{16C}. A 500-fold molar excess of unlabeled competitor (Comp) T_{1B}, T_4, or T_{16C} were added as indicated. The free probe (P) and specifically bound (B) TFIID-oligonucleotide complex and a nonspecific complex (NS) are indicated by arrows. (**C**) DNase I footprint analysis of human TFIID. Either the wild-type E1B or the LS−30/−23 probes were used. YIID and TFIID denote reactions where highly purified YIID or partially purified recombinant human TFIID, respectively, were incubated with probes. YIID and TFIID were added at three concentrations of 1/2, 1, and 2 footprint units. NP, no protein added. The position of the TATA box is indicated. The numbers denote the nucleotide sequence of the E1B promoter and were derived by a DNA sequence ladder run in the same experiment (*33*). The arrow indicates a hypersensitive DNase I cleavage site in reactions containing human TFIID. (**D**) Primer extension analysis of RNA synthesized by in vitro transcription conducted with human TFIID and YIID. The adenovirus wild-type (+) E1B and pAd1BLS−48/−39 (−) templates were used to assay TFIID activity. Reaction contained either highly purified YIID protein (6 ng) expressed in *E. coli*, rabbit reticulocyte lysate (2 μl) programmed with RNA transcribed from pKB104 (TFIID), or control reticulocyte lysate (2 μl). An arrow indicates the expected extension product.

tween residues 143 and 154. In YIID, an X-Thr-Pro sequence is found only once at the junction between the conserved amino acids and the divergent NH_2- terminus (*3–6*). It is possible that the X-Thr-Pro motif functions in the folding and activity of TFIID. Because the NH_2- terminus of human TFIID has diverged from that of YIID, it may interact specifically with the mammalian transcriptional machinery.

In the conserved COOH-terminal half of TFIID, only 36 of the 181 amino acids diverged between the yeast and human proteins, and most of the differences do not alter the charge of the amino acid. This high degree of conservation explains why YIID can substitute for the human TFIID protein in in vitro transcription assays.

Curiously, the structure of the human TFIID protein resembled Sp1. For example, both proteins have serine- and threonine-rich domains. The most striking similarity is that human TFIID has one long glutamine-rich region while Sp1 has two short glutamine-rich domains in its NH_2-terminus (*25*). Deletion of both polyglutamine-rich regions in Sp1 results in an inability to activate transcription (*26*). It is thus tempting to postulate a similar role for the TFIID glutamine-rich region. However, it is also possible that the glutamine-rich sequence may serve some other regulatory function.

Glutamine repeats have been noticed in the *opa* repeat family of genes, which includes the *Drosophila notch* (*27*) and *zeste* (*28*) genes. Several homeotic genes also have an *opa* repeat, including the *fushi tarazu*, *bithorax*, and *antennapedia* genes from *Drosophila*, and the mouse *mopa* fragment (*27, 29*). Other than the glutamine-rich repeats, computer-assisted DNA sequence analysis did not reveal extensive sequence similarity between TFIID and the *opa* repeat gene family.

We used restriction fragments of pKB104 insert to probe digested human chromosomal DNA and purified HeLa cell polyadenylated RNA (*30*) (Fig. 4). A highly enriched 2.2-kilobase (kb) RNA was detected (Fig. 4B). At moderate stringency washes (*12*) only one restriction fragment was detected in human genomic DNA digested with two different sets of restriction enzymes (Fig. 4A, lanes 2 and 3). In genomic DNA derived from K562 cells, the c-*abl* gene is present at approximately four to eight copies per genome (*31*). A parallel blot hybridized with a c-*abl* probe yielded a band that was five times as intense as the TFIID band (Fig. 4A, lane 1), indicating that there is one copy of the TFIID gene per genome.

In summary, we have cloned a gene that encodes a transcriptionally active TFIID protein. Biochemical characterization of the human TFIID protein will aid in understanding the complex process of gene regulation in mammalian cells.

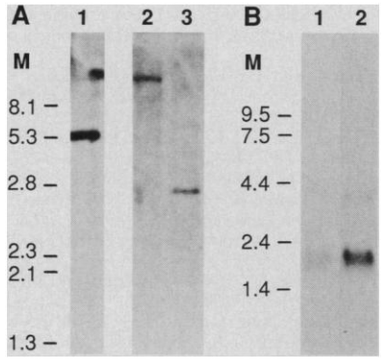

Fig. 4. Genomic Southern and Northern analyses of the TFIID gene. (**A**) DNA (10 μg) from a human lymphocytic cell line, K562, digested with Eco RI and Hind III (lanes 1 and 3) or with Eco RI (lane 2). Lane 1 was probed with a c-*abl* probe (2×10^7 cpm). Lanes 2 and 3 were probed with the 233-bp PCR fragment (2×10^7 cpm). M denotes the position of the molecular size markers in kilobases. (**B**) Total cytoplasmic RNA (20 μg) purified from Hela cells (lane 1) and poly(A)$^+$-enriched RNA (10 μg) (lane 2). The probe was a 1-kb internal restriction fragment of pKB104 that does not contain the poly(A) or the polyglutamine sequences. The faint band of approximately 4 kb may be partially processed TFIID mRNA.

REFERENCES AND NOTES

1. T. Matsui, J. Segall, P. A. Weil, R. G. Roeder, *J. Biol. Chem.* **255**, 11992 (1980); M. Samuels, A. Fire, P. A. Sharp, *ibid.* **257**, 14419 (1982).
2. B. L. Davison, M. J. Egly, E. R. Mulvihill, P. Chambon, *Nature* **301**, 680 (1983); M. W. Van Dyke, R. G. Roeder, M. Sawadago, *Science* **241**, 1335 (1988); S. Buratowski, S. Hahn, L. Guarente, P. A. Sharp, *Cell* **56**, 549 (1989).
3. M. C. Schmidt, C. C. Kao, R. Pei, A. J. Berk, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7785 (1989).
4. S. Hahn, S. Buratowski, P. A. Sharp, L. Guarente, *Cell* **58**, 1173 (1989); D. M. Eisenmann, C. Dollard, F. Winston, *ibid.*, p. 1183.
5. M. Horikoshi *et al.*, *Nature* **341**, 299 (1989).
6. B. Cavallini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9803 (1989).
7. S. Buratowski, S. Hahn, P. A. Sharp, L. Guarente, *Nature* **334**, 37 (1988); B. Cavallini *et al.*, *ibid.*, p. 77.
8. J. H. J. Hoeijmakers, *ibid.* **343**, 417 (1990); K. Nagai, *ibid.*, p. 418.
9. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
10. M. C. Schmidt, Q. Zhou, A. J. Berk, *Mol. Cell. Biol.* **9**, 3299 (1989).
11. S. Smale, M. C. Schmidt, A. J. Berk, D. Baltimore, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
12. Southern hybridization conditions were as described (*3*), except that 20% formamide was used. The blots shown in Fig. 4A were washed in a solution containing 0.9 M NaCl, 0.09 M sodium citrate, and 0.1% SDS at 68°C.
13. C. C. Kao and A. J. Berk, in preparation.
14. The sequence of the oligonucleotides (Midland Certified Reagent Co.) used for PCR: BK10, 5′-ATAGGATCCAAPATGGTNGTNACNGGNGCNAA-3′; BK11, 5′-ATAGGATCCAC(T/G/A)ATYTTNCCN(G/C)(A/T)NACPAA(T/G/A)AT-3′; BK12, 5′-ATCGGATCCCCNGGPAANAPYTCNGGYTCP-TA-3′. Y, pyrimidine; P, purine; N, all four nucleotides. Where indicated by a slash, combinations of nucleotides were inserted. All of the oligonucleotides contained at their 5′ end a nine-nucleotide sequence that included the recognition site for Bam HI restriction enzyme.
15. R. G. Roeder, presented at the ICN-UCLA Keystone meeting, Colorado, 28 January 1990.
16. PCR with genomic DNA or cDNA was performed in 100-μl reactions that contained the appropriate DNA template, 2.5 μg of each oligonucleotide, 0.25 mM deoxyribonucleoside triphosphates (dNTP), and 2.5 units of Taq polymerase with its supplied buffer (New England Biolabs). The reaction was first allowed to cycle five times among three temperatures: 94°C (1.5 min), 45°C (1.5 min), and 72°C (2.5 min). The reaction was then cycled 35 times at 94°C (1.5 min), 52°C (1.5 min), and 72°C (2.5 min). In the PCR reaction with human cDNA, the first round was performed with BK10 and BK11, then 2 μl of the product was used as the template for a second round of reactions with BK10 and BK12. cDNA was synthesized from cytoplasmic RNA in a 40-μl reaction that contained 3 μg of purified cytoplasmic RNA, 100 pmol of random hexamer oligonucleotide primers (Boehringer Mannheim), 33 units of RNAsin (Promega), 0.4 mM dNTP, 3.8 mM dithiothreitol (DTT), and 20 units of M-MuLV reverse transcriptase (Stratagene) in a buffer containing 50 mM tris (pH 7.5), 75 mM KCl, 10 mM DTT, and 3 mM $MgCl_2$. The reaction was mixed on ice, incubated at 37°C for 1 hour, then heated to 95°C for 5 min prior to ethanol precipitation. The precipitated material was used for PCR amplification.
17. DNA sequencing (Sequenase, US Biochemicals) on plasmids purified by alkaline lysis [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983] and gel filtration chromatography on S1000 (Pharmacia). Both strands of pKB104 were sequenced from subcloned restriction fragments and synthesized internal oligonucleotides. Plasmid subclones were propagated in the *E. coli* strain DH5α [*endA*−, *hsd*R17(r_k−, m_k+), *rec*A1].
18. The lambda Zap HeLa cDNA library (Stratagene) was screened as described [W. D. Benton and R. W. Davis, *Science* **196**, 180 (1977)]. Approximately 750,000 plaques were probed with ^{32}P-labeled 233-bp PCR-amplified DNA. The plasmid was excised by coinfecting a male strain of *E. coli* (XL1 Blue) with M13 helper phage.
19. The in vitro transcription and translation reactions used to express the TFIID polypeptide were performed according to the manufacturer (Promega) with T3 RNA polymerase (Stratagene). pKB104 (3 μg), linearized at a Kpn I site downstream of the open reading frame, was transcribed with 50 units of T3 RNA polymerase. One-fifth of the RNA was used in the in vitro translation reaction.
20. For gel mobility shift assays, complementary double-stranded oligonucleotides with 4-nt extensions at the 5′ end were labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP to a specific activity of 1.3×10^4 cpm/fmol. The top strand sequence of the E1B 29-bp oligonucleotide (T_{1B}) was 5′-TCGACTTAAAGGGTATATAATGCGCCGTG-3′. The T_{16C} top strand sequence was: 5′-TCGACTTAAAGTAATAATGCGCCGTG-3′. The T_4 top strand sequence was 5′-TCGACTTAAAGCTATATACTCGCCGTG-3′. Oligonucleotide probe (2 fmol) was incubated with 2 μl of rabbit reticulocyte lysate and 24 μg/ml of poly(dGdC)–poly(dGdC) nonspecific competitor in a total volume of 12 μl. Unlabeled specific competitor, when included, was at 20 ng per reaction. Binding reactions contained 12 mM Hepes, pH 7.9, 10% glycerol, 5 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, 50 μg/ml bovine serum albumin, 0.5 mM EDTA, and 0.05% NP40. After incubation for 30 min at 30°C the binding reactions were loaded onto 5% polyacrylamide gels (0.045 M tris, 0.045 M borate, 0.001 M EDTA). The 14-cm gel was subjected to electrophoresis for 30 min prior to sample addition and then subjected to further electrophoresis at 100 V for 2 to 3 hours.
21. Human TFIID protein was expressed from a bacte-

riophage T7 promoter in plasmid pET3B as described [W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986); A. H. Rosenberg *et al.*, *Gene* **56**, 125 (1987)]. Soluble *E. coli* extract containing the recombinant TFIID was fractionated by passage through heparin agarose in 0.1 M KCl, followed by step elution with 0.25 M, 0.5 M, and 0.8 M KCl in 20 mM Hepes, pH 7.9, 20% glycerol, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride, 0.2 mM EDTA. TFIID was assayed by SDS polyacrylamide gel electrophoresis of column fractions. Peak TFIID in the 0.5 M KCl fraction was used in footprinting assays. Empirically determined 1/2, 1, and 2 minimum footprint units were used for footprinting as described (*10*). Recombinant YIID expressed in *E. coli* (*3*) was prepared (*7*) and used at 1/2, 1, and 2 footprint units.

22. Phosphocellulose fractions were prepared as described [J. D. Dignam, P. L. Martin, B. S. Shastry, R. G. Roeder, *Methods Enzymol.* **101**, 582 (1983)]. The in vitro transcription assays were performed as described (*10*) with 1 μg of template plasmid and 2 μl of reticulocyte extracts. Primer extension assays (*10*) were performed with the B1750 oligonucleotide.
23. Q. Zhou and A. J. Berk, unpublished data.
24. L. Wu, D. Rosser, M. C. Schmidt, A. J. Berk, *Nature* **326**, 512 (1987).
25. J. Kadonaga, K. C. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987).
26. A. J. Courey and R. Tjian, *ibid.* **55**, 887 (1988).
27. K. A. Wharton, B. Yedvobnick, V. G. Finnerty, S. Artavnis-Tsakonas, *ibid.* **40**, 55 (1985).
28. C. Mariani, V. Pirrotta, E. Manet, *EMBO J.* **4**, 2045 (1985).
29. D. Duboule, M. Haenlin, B. Galliot, E. Mohier, *Mol. Cell. Biol.* **7**, 2003 (1987).
30. Cytoplasmic RNA for Northern analysis was purified from HeLa cells as described [A. Jacobson, *Methods Enzymol.* **152**, 254 (1987)]. Poly(A)$^+$ RNA was enriched by chromatography on oligo (dT)-celluose (Collaborative Research). RNA was purified on a gel consisting of 1% agarose, 6.2% formaldehyde, 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, transferred to nitrocellulose UV–crosslinked, and probed with a ^{32}P-labeled restriction fragment from the TFIID cDNA.
31. S. J. Collins and M. T. Groudine, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4813 (1983).
32. W. R. Pearson and D. J. Lipman, *ibid.* **85**, 2444 (1988).
33. A. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
34. We thank R. Gaynor and colleagues for the lambda Zap library; M. Gishinsky for human genomic DNA and cytoplasmic RNA; S. Crews for a *Drosophila* cDNA library and discussions; C. Guthrie and E. Shuster for numerous yeast strains; and P. O'Farrell for discussions. Supported by American Cancer Society (ACS), California division, postdoctoral fellowship and ACS postdoctoral fellowships J-3-89 (to C.C.K.), PF3380 (to P.M.L.), and PF2715 (to M.C.S.) and by National Cancer Institute grant CA25235.

18 April 1990; accepted 22 May 1990

A cDNA for a Protein That Interacts with the Human Immunodeficiency Virus Tat Transactivator

PETER NELBOCK, PATRICK J. DILLON, ANN PERKINS, CRAIG A. ROSEN*

The human immunodeficiency virus (HIV) *tat* protein (Tat) is a positive regulator of virus gene expression and replication. Biotinylated Tat was used as a probe to screen a λgt11 fusion protein library, and a complementary DNA encoding a protein that interacts with Tat was cloned. Expression of this protein, designated TBP-1 (for Tat binding protein–1), was observed in a variety of cell lines, with expression being highest in human cells. TBP-1 was localized predominantly in the nucleus, which is consistent with the nuclear localization of Tat. In cotransfection experiments, expression of TBP-1 was able to specifically suppress Tat-mediated transactivation. The strategy described may be useful for direct identification and cloning of genes encoding proteins that associate with other proteins to modulate their activity in a positive or negative fashion.

UNLIKE MOST OTHER RETROVIRUSES, HIV gene expression and replication is dependent on the expression of both viral-encoded and cellular transacting regulatory proteins. One such viral regulatory protein, referred to as Tat (*1*), is a positive regulator of HIV gene expression. Tat functions through a sequence known as TAR (*2*), which is located between nucleotides +1 to +43 in the long terminal repeat (LTR) (*3*). In this context, TAR is present in both DNA and the 5′ untranslated region of all viral messages. Although the mechanism for transactivation remains unclear, with both transcriptional (*2–5*) and posttranscriptional (*6, 7*) components being considered, recent studies indicate that Tat interacts with TAR RNA (*8*). Whether this interaction, in itself, is sufficient to bring about transactivation remains unclear, in part because of the lack of an in vitro model to examine Tat function.

Since Tat interacts with TAR RNA, and also elicits a clear transcriptional effect on gene expression, it has been suggested that TAR may function as an RNA enhancer (*9*). We therefore speculated that cellular factors interact with Tat, which in turn mediates transcriptional effects through other interactions within the LTR. With this in mind, we investigated whether a human cDNA encoding a protein that interacts with Tat could be identified. Our approach was similar to that employed for in situ detection of sequence-specific DNA binding proteins (*10*) or fusion proteins that interact with a specific antibody (*11*). Specifically, purified Tat protein was used as a probe to screen a cDNA library. *Escherichia coli*–derived Tat, which was previously shown to be fully functional (*12*), was biotinylated and used as a probe to screen a λgt11 fusion cDNA library prepared from human Jurkat T-cell cDNA (Clonetech). Proteins expressed from the phage were adsorbed onto nitrocellulose, and phage-encoding proteins able to interact with biotinylated Tat were identified by subsequent addition of a streptavidin-coupled alkaline phosphatase conjugate plus a color indicator (*13*). Approximately 2×10^6 phage were screened, and ten were identified that expressed proteins that interacted with Tat. Tat did not bind to a filter containing a random mix of phage from the library (Fig. 1), but showed preferential binding to one of the ten enriched phage isolates. After enrichment of the ten clones, their cDNA inserts were amplified by the polymerase chain reaction (PCR) (*14*), with flanking primers derived from the λgt11 sequence, and cloned into the appropriate expression vectors.

Examination of the PCR products by Southern (DNA) blot analysis revealed that six of the independent isolates actually represented two different clones (that is, each of these two clones was independently isolated three times) demonstrating the specificity and reproducibility of the screening strategy. A clone from one group, designated TBP-1, for Tat binding protein–1, was subjected to further analysis.

The 1341-nucleotide cDNA was found to contain an open reading frame (ORF) of 404 amino acids (Fig. 2). Visual inspection of the TBP-1 ORF revealed no obvious features common to previously characterized transcriptional or translational regulatory factors. The sequence information suggested that TBP-1 protein was produced from its own AUG codon present at nucleotide 63, suggesting that the full-length cDNA was obtained. Reverse transcription of cellular TBP-1 mRNA followed by primer extension of the cDNA product provides further support that the full-length cDNA was obtained (*15*). The coding capacity of

Department of Molecular Oncology and Virology, Roche Institute of Molecular Biology, Hoffmann–La Roche Inc., Roche Research Center, Nutley, NJ 07110.

*To whom correspondence should be addressed.