Functional Domains and Upstream Activation Properties of Cloned Human TATA Binding Protein

MICHAEL GREGORY PETERSON, NAOKO TANESE, B. FRANKLIN PUGH, ROBERT TJIAN

The TATA binding protein, TFIID, plays a central role in the initiation of eukaryotic mRNA synthesis. Here, we present a human cDNA clone for this factor. Comparison of its predicted protein sequence with those from *Drosophila* and yeast reveals a highly conserved carboxyl-terminal 180 amino acids. By contrast, the amino-terminal region of TFIID has diverged in both sequence and length. A striking feature of the human protein is a stretch of 38 glutamine residues in the NH_2-terminal region. Expression of human TFIID in both *Escherichia coli* and HeLa cells produces a protein that binds specifically to a TATA box and promotes basal transcription; the conserved COOH-terminal portion of the protein is sufficient for both of these activities. Recombinant TFIID forms a stable complex on a TATA box either alone or in combination with either of the general transcription factors, TFIIA or TFIIB. Full-length recombinant TFIID is able to support Sp1 activated transcription in a TFIID-depleted nuclear extract, while a deletion of the NH_2-terminal half of the protein is not. These results indicate the importance of the NH_2-terminal region for upstream activation functions and suggest that additional factors (co-activators) are required for mediating interactions with specific regulators.

INITIATION OF TRANSCRIPTION OF EUKARYOTIC PROTEIN-coding genes is a complex process requiring RNA polymerase II (Pol II) as well as a number of other auxiliary transcription factors. These factors can be divided into two classes based on their function: the general factors that are required for transcription of all Pol II genes, and sequence specific (or promoter selective) factors that are required for optimal transcription of only a subset of these genes. The general transcription factors are thought to form a complex with Pol II at or near the transcriptional start site by interacting with core promoter elements (*1*). The most common core element is the TATAAA sequence or TATA box, which is typically located 25 to 30 base pairs (bp) upstream of the transcriptional start site or initiator element in many, but not all, Pol II promoters. The presence of these core elements alone appears to be sufficient for the general factors and Pol II to form a productive complex that directs a low unregulated (basal) level of transcription. DNA sequence elements located proximal and distal to the core promoter elements can dramatically enhance or repress transcription. These cis-control elements are DNA binding sites for specific upstream transcription factors that have been the focus of intense study (*2*).

Structural analysis of these sequence-specific factors has defined two separable functional domains, one that binds DNA and another that activates transcription (*2*). Structurally, three types of transcriptional activation domains have been described: those that have net negative charge, those that contain a high density of proline residues, and those that contain a preponderance of glutamine residues. How these activation domains function to increase transcription initiation is not understood. Conceptually, they could function by recruiting one or more of the general transcription factors to the promoter region, or they could act after the initiation complex has assembled. Another unresolved issue is whether the upstream transcription factors and the general factors interact directly or whether there exists a third class of molecules that serve as intermediary targets for the upstream factors.

In vivo reconstituted basal level transcription from Pol II promoters requires at least four distinct general factor activities, TFIIA, TFIIB, TFIID, and TFIIE/F, in addition to Pol II, all of which have been partially purified (*3–5*). The TFIID fraction contains an activity that binds specifically to the TATA box (*3, 6*). This TATA binding factor is thought to play a pivotal role in orchestrating transcription initiation from Pol II promoters. In vitro, the binding of TFIID to the TATA box is a prerequisite for the subsequent binding of the other factors and Pol II to form an active transcription complex (*3, 6*). The TFIID protein is also thought to be the target for several upstream transcription factors including USF, ATF, and GAL4 (*7–9*). The slow rate of TFIID binding to its target sequence suggests that some regulatory factors might function by interacting directly with TFIID to enhance its binding. However, the relationship between upstream regulators and TFIID remains unclear.

A TATA binding protein has been purified and cloned from yeast (*10, 11*). Although this polypeptide binds specifically to TATA sequences and potentiates basal level transcription in a HeLa cell reconstituted transcription system, it is unable to support transcriptional activation by upstream regulators such as the human GC box binding factor, Sp1 (*12*). We have also found the glutamine-rich activation domains of Sp1 to be nonfunctional when introduced into yeast cells (*13*). Moreover, there is a marked discrepancy in the apparent size of the endogenous yeast and human TFIID activity as measured by gel filtration. The yeast TFIID activity chromatographs as a 23- to 27-kilodalton polypeptide (*10*), in accordance with the

Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720.

molecular size predicted from its amino acid sequence (*11*), while the human TFIID activity chromatographs at 120 to 140 kD (*4*). Given the apparent structural and functional differences between TFIID from yeast and humans, it was important to isolate a human TFIID complementary DNA (cDNA) clone to study the mechanism by which TFIID functions with mammalian transcription factors such as Sp1 to activate transcription.

Molecular cloning of the human TATA binding protein. For cloning human TFIID, we used sets of degenerate oligonucleotide primers selected on the basis of the yeast protein sequence to amplify fragments in a polymerase chain reaction (PCR). First, we amplified *Drosophila* TFIID sequences as an intermediate step to the human

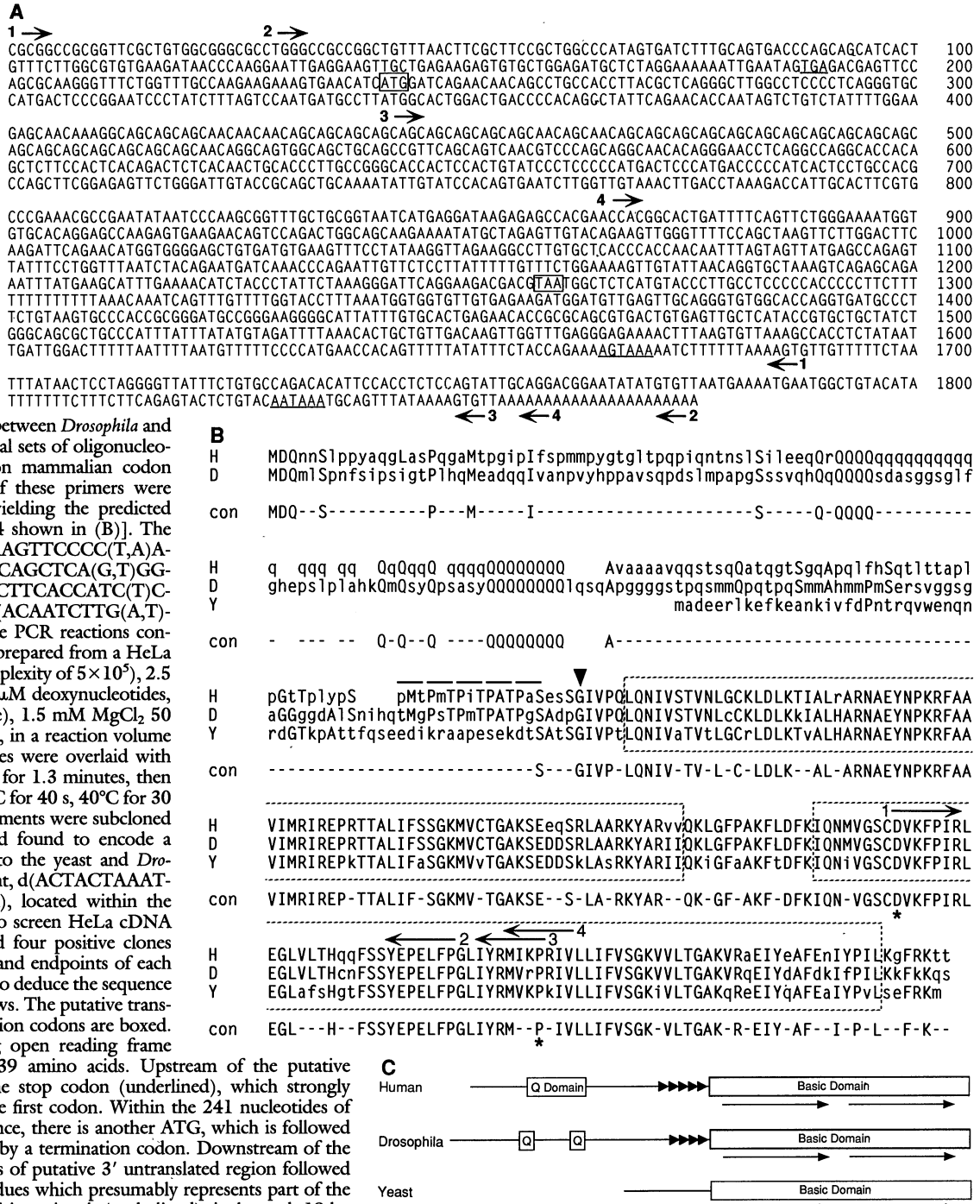

Fig. 1. DNA sequence and amino acid comparisons of human TFIID. (**A**) Sequence of human TFIID cDNA. The first step in cloning human TFIID involved the use of PCR with sets of degenerate oligonucleotide primers to amplify a fragment of the *Drosophila* TFIID gene (*14*). A fragment of the correct size, as predicted from the yeast sequence, was sequenced and found to encode a polypeptide sharing a high degree of similarity with yeast TFIID (asterisks indicate the extent of this sequence) (*14*). After the short amino acid stretches sharing amino acid identity between *Drosophila* and yeast were found, five additional sets of oligonucleotide primers (20 nt) based on mammalian codon usage were designed. Four of these primers were successfully used with PCR yielding the predicted size fragments [primers 1 to 4 shown in (B)]. The primers were (i) d(GATGTGAAGTTCCCC(T,A)ATC(T)C(A)G), (ii) d(GGGAACAGCTCA(G,T)GGC(T)TCA(G)TA), (iii) d(GGCTTCACCATC(T)CG(T)G(A)TAG(A)AT), (iv) d(ACAATCTTG(A,T)GGC(T)TTC(G)ACCAT). The PCR reactions contained 100 ng of phage DNA prepared from a HeLa cell λGT10 cDNA library (complexity of 5×10^5), 2.5 units of Taq polymerase, 100 μM deoxynucleotides, 1.5 μM each primer (each one), 1.5 mM $MgCl_2$ 50 mM KCl, 10 mM tris, pH 8.4, in a reaction volume of 50 μl. The reaction mixtures were overlaid with mineral oil, incubated at 94°C for 1.3 minutes, then taken through 35 cycles at 94°C for 40 s, 40°C for 30 s, and 72°C for 15 s. PCR fragments were subcloned into M13 then sequenced and found to encode a polypeptide sequence similar to the yeast and *Drosophila* TFIID. A 28-nt fragment, d(ACTACTAAATTGTTGGTGGGTGAGCACA), located within the amplified sequence was used to screen HeLa cDNA libraries (2×10^6 plaques), and four positive clones were isolated. The beginning and endpoints of each of the four cDNA clones used to deduce the sequence are denoted by numbered arrows. The putative translational initiation and termination codons are boxed. The sequence predicts a long open reading frame encoding a polypeptide of 339 amino acids. Upstream of the putative initiation codon is an in-frame stop codon (underlined), which strongly supports our assignment of the first codon. Within the 241 nucleotides of putative 5′ untranslated sequence, there is another ATG, which is followed three nucleotides downstream by a termination codon. Downstream of the stop codon there are 593 bases of putative 3′ untranslated region followed by a stretch of 22 adenine residues which presumably represents part of the poly(A) tail. A poly(A) addition signal (underlined) is located 19-bp upstream of the adenine stretch. Clone 1 ends in 15 adenine residues that may represent a poly(A) tail utilizing a variant signal (underlined) located 17-bp upstream. (**B**) Comparison of the deduced protein sequences of human (H), *Drosophila* (D) (*14*), and yeast (Y) (*11*) TFIID. Identities across all species are noted as a consensus (con). Where two sequences are identical, they are shown in uppercase. The region amplified by PCR from *Drosophila* is marked by two asterisks. The positions of the four oligonucleotide primers successfully used to amplify fragments of human TFIID cDNA are shown as numbered arrows. The second codon in the truncated human TFIID clone bhTFIID-180C is denoted by a caret (a methionine residue was added). The two copies of an imperfect 77–78 amino acid direct repeat are boxed. The five copies of an imperfect tripeptide repeat Pro-Met-Thr are overlined. (**C**) Schematic comparison of TFIID from human, *Drosophila*, and yeast. Q domain represents uninterrupted stretches of Glu residues. Basic domain represents the conserved COOH-terminal 180 amino acids, which is rich in basic residues. The two arrows represent the imperfect 77–78 amino acid residue direct repeat. Tandem arrow heads represent copies of the imperfect tripeptide repeat Pro-Met-Thr.

clone because it offers the advantages of a lower DNA complexity and a stronger codon usage bias (*14*). We obtained and sequenced a fragment of the *Drosophila* TFIID gene (asterisks in Fig. 1B), then used the amino acid sequences conserved between *Drosophila* and yeast to amplify fragments of human TFIID cDNA (primers in Fig. 1B). An oligonucleotide corresponding to 28 nucleotides located within the amplified human sequence was subsequently used to screen HeLa cell cDNA libraries, and four positive clones were isolated. The DNA sequence obtained from these cDNA clones is shown in Fig. 1A. RNA blot analysis indicates that the human TFIID mRNA is 1.8 to 2.0 kb in length, this heterogeneity presumably residing in the poly(A) tail. Therefore, the cDNA sequence that we obtained appears to encode most, if not all, of the mRNA. Low stringency DNA blot analysis indicates that TFIID, as in yeast (*11*) and *Drosophila* (14), appears to be encoded by a single gene in humans.

The DNA sequence of the human TATA binding protein predicts a long open reading frame encoding a polypeptide of 339 amino acids (aa) with a molecular size of 37.7 kD. Comparison of the deduced polypeptide sequences of the human, yeast (*11*), and *Drosophila* (*14*) genes (Fig. 1B) reveals a distinct bipartite structure composed of a highly conserved carboxyl-terminal domain and a divergent amino-terminal domain (Fig. 1C). The COOH-terminal 180 amino acids of the human and *Drosophila* sequences are 90 percent identical, while the yeast sequence shares 80 percent identity with either human or *Drosophila* proteins. Interestingly, most of this conserved region is composed of an imperfect 77 or 78 amino acid direct repeat (Fig. 1B).

In contrast to the conservation observed in the COOH-terminal 180 amino acids, the NH_2-terminal regions of the TFIID molecules have diverged significantly. The human and *Drosophila* sequences are more similar to each other than to the yeast, and share 24 percent identity, although the *Drosophila* sequence is longer by 14 amino acids. The human sequence is unusual in containing an uninterrupted stretch of 38 Glu residues, whereas the *Drosophila* sequence contains two blocks of 6 and 8 Glu residues each, which are separated by 32 amino acids. These Glu stretches are located in approximately the same region of the molecule. The NH_2-terminal sequence of the human TFIID, like that of *Drosophila*, is also rich in proline, methionine, and serine or threonine. Moreover, this region contains very few charged residues, in contrast with the COOH-terminal end which is rich in basic residues. There are also five copies of an imperfectly repeated tripeptide, Pro-Met-Thr, which shows some conservation in *Drosophila*. The yeast NH_2-terminal sequence is much shorter, highly charged, and lacks a glutamine-rich region, and therefore bears no obvious similarity to the human and *Drosophila* sequences.

Biochemical properties of recombinant TFIID. Buratowski *et al.* (*3*) have demonstrated by bandshift analysis with purified endogenous yeast TFIID and other HeLa cell fractions that TFIID requires TFIIA to form a stable complex with the adenovirus major late promoter (AdMLP) TATA box. TFIIB was shown to bind subsequently to the TFIID-TFIIA-DNA complex. In contrast, Horikoshi *et al.* (*11*) showed that recombinant yeast TFIID alone could indeed form a stable complex with the AdMLP TATA box. To characterize the human TFIID polypeptide biochemically we effected overexpression in *Escherichia coli* with a T7 polymerase vector (*15*) and in HeLa cells with a vaccinia virus expression vector (*16*). The T7 polymerase expressed protein (bhTFIID) was purified by heparin affinity chromatography. The vaccinia expressed protein (vhTFIID) was purified by chromatography on phosphocellulose, S-300 gel filtration, and heparin-agarose. The resulting protein fractions were analyzed on a silver-stained gel (Fig. 2). Both the vhTFIID and bhTFIID migrate at 40 kD, close to their predicted molecular size of 37.7 kD (Fig. 2, lanes 1 and 2). Endogenous HeLa TFIID activity chromatographs in the excluded volume of an S-300 gel filtration column, behaving as a large molecular entity (*4*), whereas the overproduced recombinant protein chromatographed within the range of 22 to 40 kD, in accordance with its predicted molecular size. This difference may be due to complexing of the endogenous protein with other components present in limiting amounts in HeLa cells. We also overexpressed the COOH-terminal 180 amino acids of human TFIID (bhTFIID-180C) in *E. coli* to analyze the biochemical properties of this conserved subdomain. The truncated hTFIID protein migrates at 22 kD which is slightly larger than its predicted size of 20.4 kD (lane 3).

The vhTFIID and bhTFIID proteins were tested for their ability to bind to the AdMLP TATA box using a deoxyribonuclease (DNase) I protection assay. The vhTFIID and bhTFIID proteins display essentially identical DNase I protection patterns and bind with similar affinities, protecting a 20-nucleotide region centered on the TATA box (Fig. 3, lanes 2 to 7). When the same amounts of recombinant protein were used, neither hTFIID shows a specific protection pattern on a fragment containing a nonfunctional mutant TATA box (Fig. 3, lanes 12 to 17). The truncated bhTFIID-180C protein shows a very similar protection pattern to the full-length proteins and binds with a similar affinity, indicating that the COOH-terminal region of the protein conserved among human, *Drosophila*, and yeast is sufficient to confer sequence-specific DNA binding to the TATA box (Fig. 3, lanes 19 and 21).

Endogenous yeast TFIID has been reported to form a stable complex on DNA with other components of the general transcriptional apparatus (*3*). We therefore tested the ability of the recombinant product to bind DNA in a bandshift assay and to interact with HeLa TFIIA and TFIIB. The vhTFIID protein interacts with the AdMLP TATA box to form a specific shifted complex (Fig. 4, lane

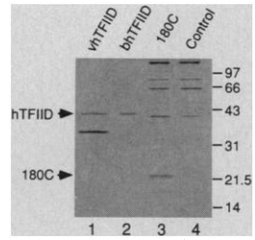

Fig. 2. Silver-stained SDS-polyacrylamide gel of recombinant human TFIID proteins. (Lane 1) Vaccinia virus–expressed vhTFIID (30 ng); (lane 2) *E. coli*–expressed bhTFIID (30 ng); (lane 3) *E. coli*–expressed bhTFIID-180C (30 ng); (lane 4) control *E. coli* pAR3038 fraction lacking TFIID (equivalent fraction and volume to lane 3). Molecular standards set are in kilodaltons. Fractions were placed on the gel to show similar amounts of recombinant hTFIID protein. The purity of each fraction is estimated to be 30 percent for vhTFIID, 50 percent for bhTFIID, and 10 percent for bhTFIID-180C. The complete coding region and a fragment encoding the COOH-terminal 180 amino acids were generated by PCR with primers containing Nde I sites and subcloned into the Nde I site of the T7 polymerase vector pAR3038 (*15*). Expression of the bacterial clones was induced in *E. coli* BL21 (*14*). Three 500-ml cultures were processed in parallel, cells containing pAR3038, pARhTFIID, and pARhTFIID-180C, by the purification scheme of Hoey *et al.* (*14*). The fractions were eluted from heparin-Sepharose at 0.2 to 0.4 M KCl and used in all subsequent experiments. The PCR fragment containing the complete coding region of hTFIID was also subcloned into the Sma I site of the vaccinia virus expression vector pAbT4537 and a recombinant virus was generated as described (*16*). Nuclear extracts were prepared from HeLa cells infected with either recombinant or control virus (New York City Board of Health strain of vaccinia) as described (*16*). The vhTFIID protein was purified by phosphocellulose chromatography, with the nuclear extract in 0.1 M KCl, 5 mM Hepes-KOH, pH 7.9, 2 mM $MgCl_2$, 10 percent glycerol, 1 mM dithiothreitol (DTT). The activity peak (0.1 to 0.45 M KCl fraction) was fractionated by S-300 gel filtration chromatography in 0.5 M KCl, 20 mM tris-HCl pH 7.9, 0.1 mM EDTA, 10 percent glycerol, 0.05 percent NP-40, 1 mM DTT. The activity peak was pooled, dialyzed in 0.1 M KCl, 25 mM Hepes-KOH, pH 7.9, 12.5 mM $MgCl_2$, 0.1 mM EDTA, 20 percent glycerol, 0.1 percent NP-40, 1 mM DTT, and fractionated by heparin-agarose chromatography. The heparin 0.25 to 0.4 M KCl fraction is shown and was used in all experiments unless otherwise indicated.

1). This specific complex is competed efficiently with excess wild-type TATA box sequences (Fig. 4, lanes 2 and 3). In contrast, a mutant TATA box sequence or an unrelated nonspecific DNA fragment are less efficient in competing for binding of the specific complex (Fig. 4, lanes 4 to 7). The minor nonspecific bands visible in each lane are competed equally with the different competitor DNA's. When HeLa TFIIA fraction was added with vhTFIID, two additional slower migrating DNA protein complexes are observed (arrows in Fig. 4, lane 10). Because TFIIA fails to form a stable complex alone (Fig. 4, lane 9), these shifted complexes presumably contain both TFIIA and vhTFIID. As expected, the TFIID-TFIIA complexes are specifically competed with excess AdMLP TATA box sequence but not with a nonspecific sequence. Since TFIID is thought to form stable complexes with both TFIIA and TFIIB, we also tested the ability of TFIIB to participate in complex formation. The HeLa TFIIB fraction alone does not form a stable complex with the AdMLP TATA box sequence (Fig. 4, lane 11). Addition of TFIIB along with vhTFIID results in two additional slower migrating complexes (arrows, lane 12), which are specifically competed with excess AdMLP TATA box sequence. Our results confirm that the TATA binding factor is indeed able to interact directly with TFIIA and TFIIB. Furthermore, bhTFIID-180C also forms similar complexes with TFIIA and TFIIB, an indication that the conserved COOH-terminal 180 residues are sufficient for interaction with the other general transcription factors.

Activation of basal and Sp1-dependent transcription by recombinant TFIID. The ability of vhTFIID and bhTFIID to functionally substitute for the endogenous HeLa cell TFIID fraction was tested with partially purified HeLa TFIIA, TFIIB, TFIIE/F, and Pol II in a reconstituted transcription reaction. Reconstituted transcription from the AdMLP is dependent on the addition of an endogenous HeLa cell TFIID fraction to the reconstituted reaction (Fig. 5A, lanes 1 and 2). Addition of a phosphocellulose fraction prepared from HeLa cells infected with the control virus shows no activity (Fig. 5, lane 3), while the corresponding fraction prepared from vhTFIID expressing HeLa cells promotes a high level of activity (lane 4). Moreover, 5 ng of highly purified vhTFIID (Fig. 5, lane 5) also directs efficient basal level transcription. Similarly, addition of bhTFIID (5 ng) or bhTFIID-180C (2.5 ng) stimulates basal transcription to the same extent as that achieved with vhTFIID. Since the truncated bhTFIID-180C has transcriptional properties similar to those of the full-length vhTFIID and bhTFIID proteins, we conclude that the COOH-terminal 180 amino acids are sufficient for both TATA binding and interacting with the other general factors to promote basal transcription.

The endogenous human TFIID fraction is capable of supporting Sp1-dependent transcriptional activation, whereas the bacterially expressed recombinant yeast TFIID and the endogenous yeast TFIID fraction are not (*13*). The failure of the yeast protein to

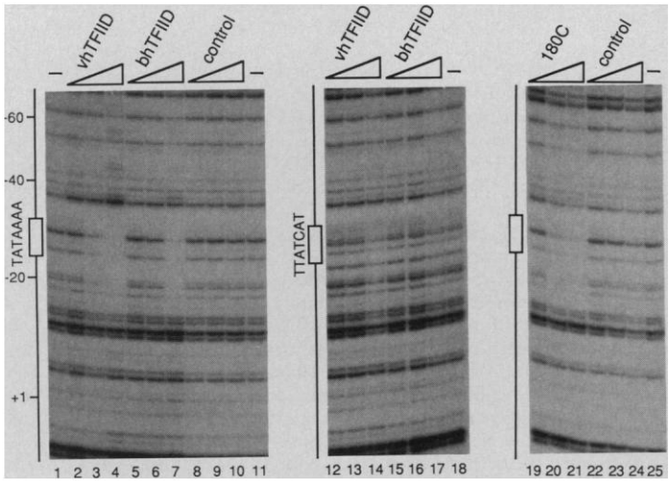

Fig. 3. DNAse I protection patterns of recombinant human TFIID on TATA elements. Labeled promoter fragments containing lanes 1 to 11 and 19 to 25, AdMLP TATA box (TATAAAA); (lanes 12 to 18) mutant TATA box (TTATCAT), were incubated with increasing amounts of the following: (lanes 2 to 4, and 12 to 14) vhTFIID protein; (lanes 5 to 7 and 15 to 17), bhTFIID protein; (lanes 8 to 10) the equivalent control *E. coli* pAR3038 fraction lacking TFIID. In each case, 1.5 ng of the TFIID protein (highest amount shown) gave a 20-bp blank-out protection pattern centered around the AdMLP TATA element but not the mutant element. Lanes (19 to 21) show binding of the truncated bhTFIID-180C protein to the AdMLP TATA box with similar affinity (lane 21 contains 4 ng of TFIID protein); (lanes 22 to 24), control *E. coli* pAR3038 fraction lacking TFIID. The experiment was performed essentially as described (*14*), with the Xho I (+60) to Hpa I (−205) fragment from pS-G_6TI (*12*) that was 5′-labeled on the noncoding strand. The mutant TATA box fragment differs only in that the TATAAA sequence has been replaced by TTATCAT. The difference in protein concentration between each step in the titration is three to four times.

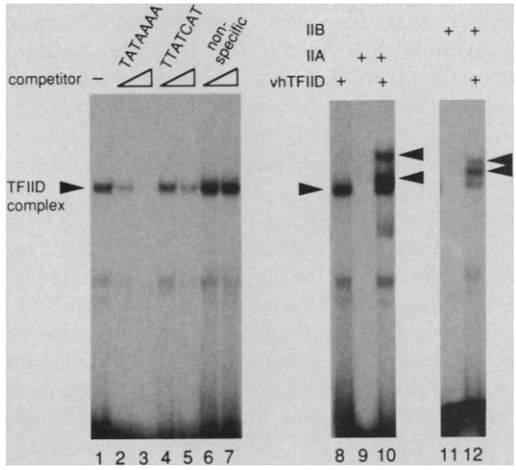

Fig. 4. Electrophoretic mobility shift assay of recombinant human TFIID. Probe DNA (Sac I–Eco RI, 28 bp fragment from the plasmid pS-G_6TI (*12*) containing the AdMLP TATA box) was incubated with 1.5 ng of vhTFIID protein. (Lane 1) No specific competitor DNA; (lanes 2 and 3) increasing amounts of competitor DNA containing the AdMLP TATA box (TATAAAA); (lanes 4 and 5) mutant TATA sequence (TTATCAT); (lanes 6 and 7) oligonucleotide of unrelated sequence; the TFIID-specific complex is indicated by the arrow on the left; (lane 10) 100 ng of HeLa TFIIA (Mono Q fraction) added; (lane 12) 0.9 ng of TFIIB (phenyl fraction) added to the reaction containing vhTFIID gave rise to protein-DNA complexes that migrated more slowly on the gel (arrows on right). Reactions containing the equivalent amount of (lane 9) TFIIA alone; (lane 11) TFIIB alone, did not show any visible complexes. Binding assays contained 20 mM Hepes-KOH pH 7.9, 25 mM KCl, 2 mM spermidine, 0.1 mM EDTA, 0.025 percent NP-40, 10 percent glycerol, 0.5 mM DTT, bovine serum albumin at 100 μg/ml, 100 ng of poly(dG-dC) and 0.12 ng of probe DNA and for (lanes 1 to 7) 2 mM $MgCl_2$, (lanes 8 to 12) 5 mM $MgCl_2$. A molar (10 to 40 times more) excess of the competitor DNA was added where indicated. The sequence of the nonspecific competitor oligonucleotide was d(TCGAGCAG-GACCGGTTCAA). Binding reactions were performed for 40 minutes at room temperature, placed on a native polyacrylamide gel containing 0.5 TBE (tris, borate, EDTA), 1 mM EDTA, and 0.05 percent NP-40 and subjected to electrophoresis at 4°C. The TFIIA S-300 fraction (*12*) was further purified by dialysis in 0.1 M KCl in 20 mM tris-HCl, pH 7.9, 2 mM $MgCl_2$, 10 percent glycerol, 0.03 percent lauryldiamethylamide oxide, 1 mM DTT (TMGL), and FPLC Mono Q (Pharmacia) chromatography. The TFIIA activity eluted between 0.35 to 0.5 M KCl-TMGL. The TFIIB S300 fraction (*12*) was further purified by dialysis in 0.1 M KCL-TMGL and Cibacron blue-agarose chromatography. The activity eluted between 1.0 and 2.0 M KCl-TMGL and was further fractionated by HPLC 5PW phenyl chromatography with a 1.0 to 0.0 M KCl-TMGL gradient. The peak activity eluted at 0.5 M KCl-TMGL.

support upstream activation by Sp1 could be the result of species-specific differences or the lack of additional factors, or both. In order to address these questions, we tested the ability of recombinant human TFIID to support Sp1-dependent activation in a HeLa cell derived reconstituted transcription reaction with a synthetic promoter containing multiple Sp1 binding sites upstream of the AdMLP TATA box. In the presence of partially purified HeLa TFIID fractions, Sp1 typically enhances transcription four- to eightfold on this promoter (Fig. 5B, lanes 1 and 2). In contrast, Sp1 is unable to activate transcription in a reconstituted reaction where either bhTFIID (Fig. 5B, lanes 3 and 4) or vhTFIID (lanes 5 and 6) is substituted for the TFIID fraction. Instead, we consistently observe an inhibition of transcription in the presence of Sp1, which is also seen with the recombinant yeast TFIID and bhTFIID-180C proteins. We do not understand the reason for this inhibition. These results suggest that the partially purified TFIID fraction contains one or more activities, which we refer to as co-activators, necessary to mediate Sp1 activation in a reconstituted transcription reaction. Alternatively, it is possible that recombinant hTFIID, even when expressed in HeLa cells, is not folded correctly to direct upstream activation.

The inability of the cloned TFIID product to support upstream activation by Sp1 prompted us to find out whether the cloned product is fully functional when supplemented with the hypothetical coactivator function. Since our attempts to separate TFIID chromatographically from the putative coactivators in the partially purified fraction had so far been unsuccessful, we turned to an alternative assay. On the basis of the well-documented heat labile nature of HeLa cell TFIID activity (*17*), we believed that it might be possible to heat inactivate the endogenous TFIID activity resident in the transcriptionally active nuclear extract without destroying other essential activities. If so, then the addition of recombinant TFIID in the presence and absence of Sp1 should establish the actual capability of recombinant TFIID to mediate an Sp1 response. To destroy the endogenous TFIID activity so that transcription potentiated by the recombinant hTFIID may be observed, the crude HeLa cell nuclear extract was heated to 47°C for 15 minutes (*17*). This treatment appeared to destroy more than 90 percent of the endogenous TFIID activity (Fig. 6, lane 1). Basal level transcription could be restored by the addition of either recombinant bhTFIID or bhTFIID-180C, or a partially purified endogenous TFIID fraction (Fig. 6, lanes 3, 5, and 7). When Sp1 was added to the heat-treated nuclear extract, a modest level of transcription was observed, presumably resulting from the small amount of endogenous TFIID that survived the heat inactivation (Fig. 6, lane 2).

As expected, addition of Sp1 along with the partially purified endogenous TFIID fraction, resulted in a typical enhancement in transcription (fivefold) over that achieved with the TFIID in the absence of Sp1 (Fig. 6, lanes 7 and 8). When Sp1 was added in combination with the cloned product, bhTFIID, transcription was also significantly increased (3.5-fold) over that observed in the absence of Sp1 (Fig. 6, lanes 3 and 4). This result was striking because of the inhibition of transcription observed in a system reconstituted from partially purified HeLa cell fractions. These results indicate that in the heat-treated nuclear extract, bhTFIID is able to support an Sp1-dependent activation of transcription, presumably because resident coactivators remain functional in the extract. In contrast to the full-length TFIID protein, when the version (bhTFIID-180C) truncated at the NH_2-terminus was added to the heat-treated nuclear extract in the presence of Sp1, no Sp1-dependent activation was observed (Fig. 6, lanes 5 and 6). The lack of an Sp1 response observed with the truncated protein, suggests that the NH_2-terminal region of hTFIID is necessary to achieve activation of transcription by Sp1.

Differences between recombinant and endogenous TFIID. It is necessary to know why the recombinant human TFIIDs fail to support an Sp1-dependent activation of transcription in a reconstituted system, whereas the endogenous HeLa cell TFIID fraction supports such a response. Since we used recombinant proteins expressed in both *E. coli* and HeLa cells with identical results, it is unlikely that posttranslational modification differences can account for the lack of upstream stimulation. Instead, we favor the hypothesis that partially purified TFIID fractions contain novel transcription factors, which we term coactivators, that serve as adaptors for mediating upstream activation by TFIID. Purification of TFIID over five different chromatography columns fails to separate the Sp1 responsive activity from the basal transcription activity. Also, the endogenous TFIID fraction typically gave rise to a DNAse footprint on the AdMLP that spanned approximately 75 nucleotides from −40 to +35 relative to the transcriptional start site (*8, 9, 17*). In contrast, we find that the DNAse I footprint obtained with purified

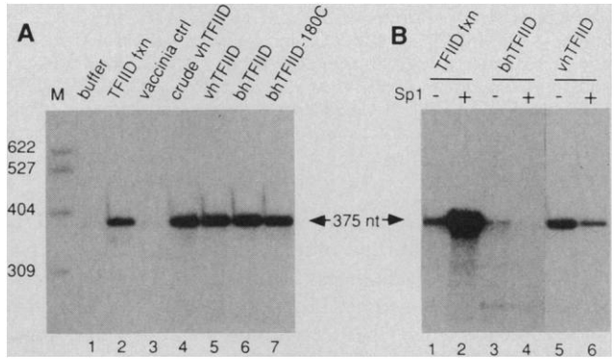

Fig. 5. Transcriptional promoting properties of cloned hTFIID in a reconstituted system. (**A**) Basal transcription properties of cloned hTFIID. Transcription reactions were reconstituted on the AdMLP with HeLa general factors TFIIA, TFIIB, TFIIE-F-pol II (*12*), and the indicated TFIID. Transcripts initiating and terminating at the correct positions produce a 375-nt fragment. Reactions were performed as described (*12*); they contained 2.8 μg of TFIIA (S-300 fraction), 1.8 μg of TFIIB (S-300 fraction), 2.3 μg of TFIIE-F-pol II (S-300 fraction), 0.2 μg of MLC_2AT template (*22*), 50 μM ATP, 50 μM CTP, 10 μM UTP, 5 μCi of [α-^{32}P]UTP, 20 mM Hepes-KOH, 1.5 percent polyvinyl alcohol, 100 mM potassium glutamate, 50 mM KCl, 10 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 3 μM $ZnSO_4$, 1 mM DTT, and additions as indicated for each lane. (Lane 1) no added TFIID; (lane 2) 1.2 μg of endogenous HeLa TFIID (Mono Q fraction); (lane 3) 25 ng of (total protein) wild-type control (0.1 to 0.45 M KCl, phosphocellulose fraction); (lane 4) 25 ng (total protein) vhTFIID (0.1 to 0.45 M KCl, phosphocellulose fraction); (lane 5) 5 ng of vhTFIID; (lane 6) 5 ng of bhTFIID; (lane 7) 2.5 ng (equimolar) of bhTFIID-180C. In these reactions, a cassette that lacked G's (*22*) is transcribed and addition of the ribonuclease (RNase) T1 and 3′-*O*-methyl GTP were not necessary. Reactions were incubated at 30°C for 30 minutes, then stopped with 80 μl of a solution containing 3.125 M ammonium acetate and tRNA at 125 μg/ml. The RNA transcripts were purified by extraction with phenol and chloroform and precipitation with ethanol. The transcripts were analyzed by electrophoresis on denaturing 6 percent polyacrylamide-urea gels, with subsequent autoradiography. Molecular markers (M) are shown on the left in nucleotides. The endogenous TFIID fraction was purified by phosphocellulose, DEAE-Sepharose, S-300 gel filtration, and Mono Q chromatography (*12*). (**B**) Sp1 mediated upstream activation. Transcription reactions were reconstituted on an Sp1-responsive promoter containing six GC-boxes from the SV40 early promoter and a TATA box and initiator from the AdMLP [pS-G_6TI (C_2AT)] (*12*). (Lanes 1 and 2) Fourteen micrograms (total protein) of endogenous HeLa TFIID (S-300 fraction); (lanes 3 and 4) 0.5 ng of bhTFIID; (lanes 5 and 6) 5 ng of vhTFIID. Lanes 1 to 4 contain 200 ng of poly(dG-dC) and 1.5 ng (5 fmol) of a 510 bp Sph I fragment from pS-G_6TI (C_2AT) that spans the promoter region. Lanes 5 and 6 contain 100 ng of poly(dG-dC) and 100 ng of pS-G_6TI (C_2AT) (50 fmol). Lanes 2 and 4 contain 50 ng of purified Sp1 (*12*); lane 6 contains 200 ng of purified Sp1. A different set of conditions were used in lanes 3 and 4 compared to lanes 5 and 6 to illustrate the range over which Sp1 inhibition of recombinant hTFIID transcription is observed.

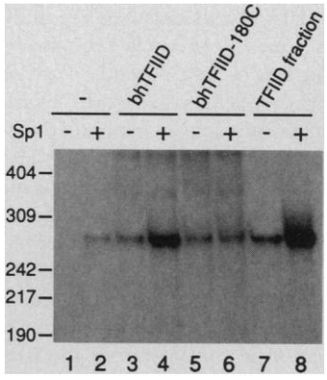

Fig. 6. Sp1 activation of recombinant hTFIID transcription in heat-treated nuclear extract. Runoff transcription reactions with heat-treated nuclear extract were performed on an Sp1-responsive promoter containing six GC boxes from the SV40 early promoter and a TATA box from the adenovirus E4 promoter (GCE_4T). (Lanes 1 and 2) No added TFIID; (lanes 3 and 4) 30 ng of bhTFIID; (lanes 5 and 6) 40 ng bhTFIID-180C; (lanes 7 and 8) 9 μg (total protein) of endogenous HeLa cell TFIID (S-300 fraction). Lanes 2, 4, 6, and 8 contained 1 μg of purified Sp1 (see *12*). Construct GCE_4T was made by digesting $pS\text{-}G_6TI$ (*12*) with Eco RI and producing blunt ends with mung bean nuclease. After digestion with Sph I, the fragment containing the GC boxes was ligated into Sph I–Hinc II digested pΔ-38 (*24*). For transcription reactions GCE_4T was digested with Eco RI giving a runoff transcript of 268 nucleotides. Sp1 depleted HeLa cell nuclear extracts were prepared (*12*) and heat-treated at 47°C for 15 minutes (*17*). Reactions were performed in 20 mM Hepes-KOH, pH 7.9, 4 mM spermidine, 4 mM $MgCl_2$, 40 mM KCl, and 10 percent glycerol with 170 μg of heat-treated nuclear extract and 200 ng of GCE_4T. Reactions were incubated for 30 minutes at 30°C, then 0.5 mM each of ATP, CTP, GTP, 10 μM UTP, and 5 μCi ^{32}P-UTP was added and the reactions were incubated at 30°C for 10 minutes. The total reaction volume was 20 μl. The reaction products were processed as in Fig. 5.

human TFIID is confined to a 20-bp region centered on the TATA box. Taken together with the observation that endogenous HeLa cell TFIID activity fractionates as a much larger protein, we postulate that the endogenous TFIID is present as a multiprotein complex with tightly associated coactivators and that Sp1-directed transcriptional activation is dependent on coactivator proteins present in this complex.

Recombinant hTFIID can support Sp1-dependent activation of transcription in a heat-treated nuclear extract system. Although heat treatment destroys nearly all of the TFIID activity, the other molecules necessary for mediating an Sp1 response apparently survive this treatment. If these Sp1 coactivator molecules are associated with the endogenous TFIID as we have suggested, then this complex may be dissociated upon heating, allowing the coactivator molecules to associate with recombinant TFIID. As bhTFIID-180C does not support an Sp1 response, the divergent NH_2-terminal domain of hTFIID is implicated in Sp1 activation of transcription. Perhaps there are direct protein-protein interactions between the Sp1 coactivator and the glutamine-rich NH_2-terminal domain of TFIID. It is interesting to note a structural similarity between the contiguous glutamine stretch in the NH_2-terminal domain of hTFIID and the glutamine-rich activation domains of Sp1. The glutamine-rich domains of Sp1 are thought to form protein interfaces that direct synergistic interactions to enhance transcription (*18*). Thus, the glutamine rich region of Sp1 may represent a structural framework for protein-protein interaction, and perhaps the stretch of glutamines in hTFIID also functions in an analogous fashion to serve as a region of contact with other proteins.

Interestingly, transcriptional activation by acidic activators also appears to be mediated by the NH_2-terminal region of TFIID. The strong acidic activator GAL4-VP16 (*19*) is able to activate transcription in a heat-treated TFIID deficient HeLa nuclear extract, when supplemented with recombinant human TFIID. However, the NH_2-terminally truncated molecule bhTFIID-180C, cannot support a GAL4-VP16 response in a heat-treated HeLa nuclear extract (*20*). Deletion analysis should reveal whether the same regions of the NH_2-terminal domain mediate both the Sp1 and GAL4-VP16 responses.

We have found independent evidence for the existence of an Sp1 coactivator protein associated with the endogenous *Drosophila* TFIID fraction (*12*). Moreover, the divergent NH_2-terminal region of *Drosophila* TFIID also appears to be required for Sp1 coactivator function. Initial experiments suggest that recombinant human TFIID cannot productively interact with the *Drosophila* Sp1 coactivator, indicating that there may be species-specific protein-protein interactions involved in Sp1 activation. The divergence of the NH_2-terminal regions of the human and *Drosophila* proteins may form the basis of this species-specificity.

Are coactivators necessary for the action of other classes of upstream activators such as the acidic activators? GAL4-VP16 activates transcription in yeast and HeLa cell nuclear extracts (*21–23*). While it also activates transcription in a heat-treated TFIID-deficient HeLa cell nuclear extract supplemented with recombinant human TFIID, we have found that it does not activate transcription when supplemented with recombinant yeast TFIID (*20*). We interpret these results to indicate that GAL4-VP16 cannot function when provided with purified TFIID and the other general factors, but that it also requires additional species-specific (coactivator) molecules to activate transcription. Similarly, Pugh and Tjian (*12*) have described experiments suggesting that CTF, which has a proline rich activation domain, requires a coactivator that is distinct from the Sp1 coactivator. Thus, coactivators may represent a new class of transcription factors that mediate signals from the upstream transcription factors to the general transcription machinery. The availability of purified recombinant hTFIID provides us with a critical tool to detect and isolate other components of the initiation complex necessary for Sp1 activation of transcription. Reconstituting transcription reactions with homogeneously purified factors should allow us to unravel the mechanisms governing regulated transcription.

REFERENCES AND NOTES

1. A. G. Saltzman and R. Weinmann, *FASEB J.* **3**, 1723 (1989).
2. P. J. Mitchell and R. Tjian, *Science* **245**, 371 (1989); N. C. Jones, P. W. J. Rigby, E. B. Ziff, *Genes Dev.* **2**, 267 (1988).
3. S. Buratowski, S. Hahn, L. Guarente, P. A. Sharp, *Cell* **56**, 549 (1989).
4. D. Reinberg, M. Horikoshi, R. Roeder, *J. Biol. Chem.* **262**, 3322 (1987).
5. D. Reinberg and R. G. Roeder, *ibid.*, p. 3310.
6. M. W. Van Dyke, R. G. Roeder, M. Sawadogo, *Science* **241**, 1335 (1988).
7. M. Sawadogo and R. G. Roeder, *Cell* **43**, 165 (1985).
8. M. Horikoshi, M. F. Carey, H. Kakidani, R. G. Roeder, *ibid.* **54**, 665 (1988).
9. M. Horikoshi, T. Hai, Y. S. Lin, M. R. Green, and R. G. Roeder, *ibid.*, p. 1033.
10. S. Buratowski, S. Hahn, P. A. Sharp, L. Guarente, *Nature* **334**, 37 (1988).
11. B. Cavallini *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9803 (1989); D. M. Eisenmann, C. Dollard, F. Winston, *Cell* **58**, 1183 (1989); S. Hahn, S. Buratowski, P. A. Sharp, L. Guarente, *ibid.*, p. 1173; M. Horikoshi *et al.*, *Nature* **341**, 299 (1989); M. C. Schmidt, C. C. Kao, R. Pei, A. J. Berk, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7785 (1989).
12. B. F. Pugh and R. Tjian, *Cell*, in press.
13. G. Gill, E. Pascal, R. Tjian, unpublished result.
14. T. Hoey, B. D. Dynlacht, M. G. Peterson, B. F. Pugh, R. Tjian, *Cell*, in press.
15. A. H. Rosenberg, B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, F. W. Studier, *Gene* **56**, 125 (1987); F. W. Studier and B. A. Moffat, *J. Mol. Biol.* **189**, 113 (1986).
16. B. D. Dynlacht, L. D. Attardi, A. Admon, M. Freeman, R. Tjian, *Genes Dev.* **3**, 1677 (1989).
17. N. Nakajima, M. Horikoshi, R. G. Roeder, *Mol. Cell. Biol.* **8**, 4028 (1988).
18. A. J. Courey and R. Tjian, *Cell* **55**, 887 (1988); A. J. Courey, D. A. Holtzman, S. P. Jackson, R. Tjian, *ibid.* **59**, 827 (1989).
19. I. Sadowski, J. Ma, S. Triezenberg, M. Ptashne, *Nature* **335**, 563 (1988).
20. B. F. Pugh, unpublished data.
21. M. Carey, J. Leatherwood, M. Ptashne, *Science* **247**, 710 (1990).
22. M. Sawadogo and R. G. Roeder, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4394 (1985).
23. B. F. Pugh and G. Gill, unpublished data.
24. Y-S. Lin, M. F. Carey, M. Ptashne, M. R. Green, *Cell* **54**, 659 (1988).
25. We thank T. Williams for pointing out the amino acid sequence duplication in the yeast TFIID sequence, B. Luscher for providing a HeLa cell cDNA library, G. Gill for promoter construct GCE_4T, S. H. S. Bell, B. Dynlacht, G. Gill, and R. Turner for their comments on the manuscript. Supported in part by a grant (to R.T.) from the National Institutes of Health; a Leukemia Society of America special fellowship (M.G.P.) and a Leukemia Society of America fellowship (B.F.P.).

9 May 1990; accepted 28 May 1990