Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9148–9152, December 1990
Biochemistry

cDNA clone encoding *Drosophila* transcription factor TFIID

(RNA polymerase II/TATA box)

MICHAEL L. MUHICH*, CALVIN T. IIDA*, MASAMI HORIKOSHI†, ROBERT G. ROEDER†, AND CARL S. PARKER*‡

*Division of Chemistry, 147-75, California Institute of Technology, Pasadena, CA 91125; and †Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, NY 10021

Communicated by John Abelson, August 27, 1990 (received for review June 29, 1990)

ABSTRACT **Proper initiation of transcription by RNA polymerase II requires the TATA-consensus-binding transcription factor TFIID. A cDNA clone encoding the *Drosophila* TFIID protein has been isolated and characterized. The deduced amino acid sequence reveals an open reading frame of 353 residues. The carboxyl-terminal 180 amino acids are approximately 80% identical to yeast TFIID and 88% identical to human TFIID. The amino-terminal portions of the yeast and *Drosophila* TFIID proteins lack appreciable homology, whereas the *Drosophila* and human amino termini appear qualitatively similar. In addition, the amino-terminal region of the *Drosophila* TFIID contains several sequence motifs that are found in other *Drosophila* proteins which appear to regulate transcription.**

The control of transcription initiation is an important regulatory step of eukaryotic gene expression. Initiation of transcription requires a set of general transcription factors that are shared by most genes transcribed by RNA polymerase II (refs. 1–4; for review, see ref. 5). Alterations in the rates of transcriptional initiation are thought to be mediated by the interaction of transcriptional activators and repressors with components of the general transcription apparatus (6–8). One key general transcription factor is TFIID, which binds to the TATA element in *Drosophila* (9), human (10, 11), and yeast (12–14). TFIID has been shown to be required for preinitiation complex formation, suggesting that it functions at an early critical step in the initiation process (15, 16). The yeast and human TFIID proteins function interchangeably to support a basal level of transcription *in vitro*, implying significant structural conservation between these molecules (12–14).

The recent cloning of the yeast (*Saccharomyces cerevisiae*) TFIID gene has allowed an examination of the functional and putative structural properties of this transcription factor (17–21). Deletion analysis of yeast TFIID demonstrated that a core region exists which is required for DNA binding and transcriptional activity *in vitro* (17, 22). To understand the molecular mechanisms regulating gene expression in higher eukaryotes, it is essential that analogous studies by undertaken in these organisms. In this report we describe the cloning and characterization of a cDNA that encodes the *Drosophila* TFIID protein.§

MATERIALS AND METHODS

Polymerase Chain Reaction (PCR) Amplification of *Drosophila* TFIID Sequences. The *Drosophila* TFIID hybridization probes were generated by PCR (23). A typical 100-μl preparative reaction mixture contained 1 μg of *Drosophila* Kc cell DNA, 375 pmol of each primer (phosphorylated by treatment with ATP and T4 polynucleotide kinase), 10 mM Tris·HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, and 2.5 units of *Thermus aquaticus* (*Taq*) DNA polymerase (Perkin–Elmer/Cetus). DNA amplification was achieved with 35 cycles of denaturation for 1 min at 94°C, primer annealing at 50°C or 55°C for 2 min, and extension for 2 min at 72°C. The final PCR cycle included a 10-min primer-extension step at 72°C. The PCR products were initially screened on the basis of size by agarose gel electrophoresis. Products of the correct expected size were purified by gel electrophoresis and blunt-end cloned into the *Hin*cII site of the pBluescript II KS(+) vector (Stratagene) for direct dideoxy sequence analysis (24).

Southern Analysis and cDNA Cloning. For Southern analysis of *Drosophila* genomic DNA, 5-μg samples of embryonic DNA were digested with the indicated restriction enzymes, electrophoresed through 0.8% agarose, and blotted onto nitrocellulose as described (25). Cloned PCR probes were radiolabeled with [α-^{32}P]dCTP by random hexamer priming (26) and hybridized with the DNA blot at 40°C in 6× SSPE (1× = 0.15 M NaCl/10 mM sodium phosphate, pH 7.0/1 mM EDTA)/5× Denhardt's solution (1× = 0.02% Ficoll 400/0.02% bovine serum albumin/0.02% polyvinylpyrrolidone)/0.5% $NaDodSO_4$/50% deionized formamide containing sonicated denatured salmon sperm DNA at 100 μg/ml.

A *Drosophila melanogaster* embryonic (0–24 hr) cDNA library in λZAP II (Stratagene) was screened by hybridization with a cloned PCR probe. Hybridizations were performed at 40°C in 6× SSPE/0.25% nonfat dried milk/50% deionized formamide. Filters were washed at 65°C in 2× SSC (1× = 0.15 M NaCl/15 mM sodium citrate, pH 7.0)/0.1% $NaDodSO_4$. After plaque purification, individual cDNA inserts were recovered in the form of chimeric pBluescript SK(+) phagemids by *in vivo* excision from the λ vector as described by the library supplier.

Overproduction of Recombinant *Drosophila* TFIID. The conserved region of *Drosophila* TFIID was subcloned into an *Escherichia coli* expression system as follows. The *Apa* I–*Eco*RI fragment of the *Drosophila* TFIID cDNA insert encoding the carboxyl-terminal 196 amino acid residues was inserted into the pET-8c (T7 expression) vector (27) at the *Nco* I–*Bam*HI restriction sites. A *Bam*HI site was added to the *Eco*RI site of the TFIID insert by ligation of an *Eco*RI–*Bam*HI (duplex oligonucleotide) adaptor (New England Biolabs). The modified TFIID insert was then ligated into pET-8c cut with *Bam*HI and *Nco* I. The 3′ overhang of the TFIID *Apa* I site was joined to the 5′ overhang of the *Nco* I site of pET-8c by use of an eight-base "bridging oligonucleotide" with sequence complementarity to both the *Nco* I and *Apa* I overhangs (5′-CATGGGCC-3′). The trimolecular ligation joined the initiating methionine codon of the pET-8c vector in-frame to the Gly-158 codon of the TFIID open reading frame. Plasmid DNA samples prepared from XL1-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PCR, polymerase chain reaction; IPTG, isopropyl β-D-thiogalactopyranoside.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M38388).

Blue *E. coli* (recA−, T7 RNA polymerase-deficient; Stratagene) transformed with the above construct were sequenced to confirm the correct fusion of the translational reading frames.

For expression of the recombinant TFIID protein (p22), the TFIID-pET-8c construct was transformed into *E. coli* strain BL21(DE3), which contains an isopropyl β-D-thiogalactopyranoside (IPTG)-inducible copy of the T7 RNA polymerase gene (27). Bacterial extracts were prepared from IPTG-induced cells (grown at 30°C) essentially as described (28). After guanidine solubilization, the lysate was centrifuged at 100,000 × *g* for 20 min and the supernatant was dialyzed against 100 mM KCl/25 mM Hepes·KOH, pH 7.9/10 mM $MgCl_2$/0.1 mM EDTA/10% glycerol/1 mM dithiothreitol with pepstatin A (0.7 μg/ml), leupeptin (0.5 μg/ml), and 1 mM phenylmethylsulfonyl fluoride. For *in vitro* transcription reactions, IPTG-induced bacterial lysates were cleared by centrifugation at 31,000 × *g* and used directly without the guanidine solubilization/dialysis step.

DNase I Protection Analysis and *in Vitro* Transcription. The 5′ deletion mutant, 5′Δ8, of the *Drosophila* fushi tarazu (*ftz*) promoter (29) was labeled on the coding strand at the *Hin*dIII site ≈75 base pairs (bp) upstream of the TATA site by treatment with [γ-^{32}P]ATP and T4 polynucleotide kinase. Following excision with *Eco*RI, the ≈220-bp end-labeled template was purified by gel electrophoresis. Due to the presence of nuclease activity in the bacterial extracts, the DNA binding reactions were carried out in the absence of Mg^{2+} at 4°C. In a total volume of 25 μl, ≈1 ng of template DNA was incubated with the indicated amount of bacterial extract in 60 mM KCl/25 mM Hepes·KOH, pH 7.9/1 mM EDTA/10% glycerol containing poly(dG-dC) at 100 μg/ml. DNase I and $MgCl_2$ were added to final concentrations of 400 μg/ml and 5 mM, respectively, and after 30 sec of digestion at 4°C the reactions were stopped by the addition of 100 μl of 1% sarkosyl/100 mM NaCl/100 mM Tris·HCl, pH 8.0/10 mM EDTA containing proteinase K at 100 μg/ml and calf thymus DNA at 25 μg/ml. After incubation at 55°C for 15 min, phenol/chloroform extraction, and ethanol precipitation, the end-labeled protection products were analyzed by urea/polyacrylamide gel electrophoresis.

In vitro transcription complementation assays were performed with HeLa cell nuclear extracts (30) in which endogenous TFIID activity was inactivated by heating to 47°C for 15 min (31). Transcriptional activity was measured using a *ftz* promoter/partial coding region template and specific initiations were detected by primer extension employing a 20-nucleotide primer complementary to +103 to +122 of the *Drosophila ftz* message.

Transcription reaction mixtures (25 μl) contained 20 mM Hepes·KOH (pH 7.9, 22°), 60 mM KCl, 8 mM $MgCl_2$, 0.6 mM NTP, 0.5 mM dithiothreitol, 10% glycerol, supercoiled template (25 μg/ml), 75 μg of nuclear extract protein, and the indicated amounts of added protein fractions. The mixtures were incubated at 22°C for 1 hr and the reactions were stopped by the addition of 125 μl of 1% sarkosyl/100 mM NaCl/100 mM Tris·HCl, pH 8.0/10 mM EDTA containing yeast tRNA at 100 μg/ml. After phenol/chloroform extraction and ethanol precipitation in the presence of 2.5 M ammonium acetate, the nucleic acid precipitates were washed with 75% ethanol, dried, and resuspended in 10 μl of 75 mM KCl/10 mM Tris·HCl, pH 8.0/1 mM EDTA with 200 fmol of oligonucleotide primer that had been labeled by treatment with [γ-^{32}P]ATP and T4 polynucleotide kinase. Primer annealing was carried out by incubation at 70°C for 5 min and then at 41°C for 30 min. The extension reactions were initiated by the addition of 25 μl of 75 mM KCl/50 mM Tris·HCl, pH 8.3/10 mM $MgCl_2$/5 mM dithiothreitol/1 mM

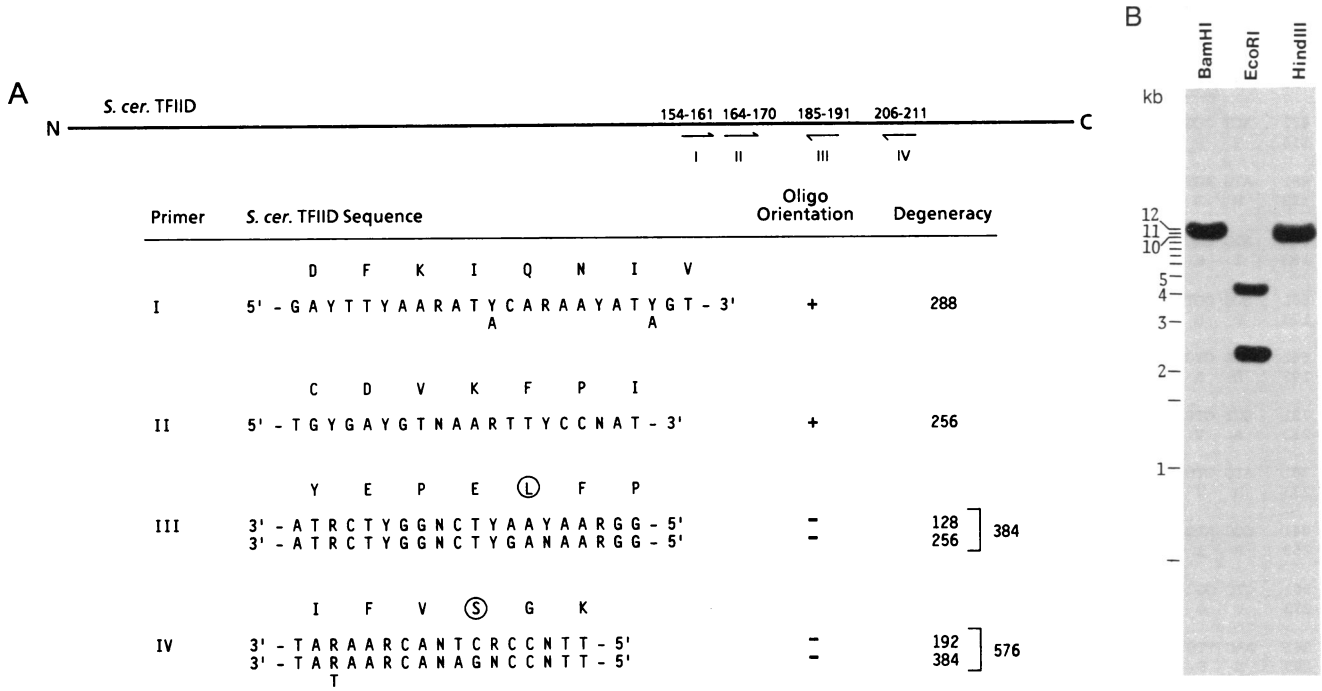

FIG. 1. PCR amplification of hybridization probes specific for *D. melanogaster* TFIID coding-region sequences. (*A*) Location of the *S. cerevisiae* TFIID sequences used to design the degenerate oligonucleotide primers for PCR. Amino acid coordinants are shown above the line drawing of the yeast TFIID open reading frame (18). The nucleotide sequences derived from the indicated *S. cerevisiae* amino acid sequences are presented along with an indication of their strandedness and degree of degeneracy. The circled amino acids are encoded by two sets of codons, which necessitated the synthesis of two oligonucleotides for priming at these sites. Nucleotide codes: R = A and G; Y = C and T; N = A, C, G, and T. (*B*) Southern blot analysis of *Drosophila* embryonic DNA digested with the indicated restriction enzymes and hybridized with ^{32}P-labeled pDmI-IV insert as described in *Materials and Methods*. The hybridization was performed in duplicate with each filter being washed at 65°C in either 2× SSC/0.1% $NaDodSO_4$ or 0.1× SSC/0.1% $NaDodSO_4$. The hybridization profiles were identical for both wash stringencies; the autoradiographic exposure of the 2× SSC-washed filter is shown. kb, Kilobases.

each dNTP containing actinomycin D (10 μg/ml) and 15 units of avian myeloblastosis virus reverse transcriptase (Life Sciences, Saint Petersburg, FL) and were terminated after 1 hr at 41°C by ethanol precipitation. The extension products were analyzed by urea/polyacrylamide gel electrophoresis.

RESULTS

A *Drosophila* TFIID cDNA clone was identified with a probe obtained by PCR amplification of *Drosophila* DNA, with primers selected from regions of low codon degeneracy within the yeast TFIID gene. The PCR products from various pairwise combinations of primers and *Drosophila* genomic DNA were initially screened by size. Products with a size corresponding to the fragment length predicted by that region of the yeast gene were isolated, cloned, and sequenced. The deduced amino acid sequences of the cloned PCR products were compared with the predicted *S. cerevisiae* sequence (18). Fig. 1 presents the primer combinations that generated PCR products bearing amino acid sequence homology to the yeast TFIID protein. Primer combinations I plus III or IV, as well as II plus III or IV, yielded products of the expected size (as well as larger uncharacterized molecules). The sequences of these products revealed a high level of amino acid sequence identity with the corresponding regions of the *S. cerevisiae* TFIID gene. The largest appropriately sized PCR product, from primer combinations I plus IV, was a 173-bp DNA fragment encoding a 57-amino acid open reading frame with 84% sequence identity to the *S. cerevisiae* TFIID protein.

Hybridization of the cloned PCR product from primer combination I plus IV (pDmI-IV) to a Southern blot of *Drosophila* embryonic DNA is shown in Fig. 1*B*. The presence of single hybridizing DNA fragments in the *Bam*HI and *Hin*dIII lanes suggests that the *Drosophila* TFIID protein is encoded by a single-copy gene. The presence of two *Eco*RI fragments (Fig. 1*B*) probably reflects allelic polymorphism, because the PCR probe, as well as the cDNAs described below, do not contain internal *Eco*RI sites.

The cloned 173-bp PCR product, pDmI-IV, was used to screen a *D. melanogaster* embryonic (0–24 hr) cDNA library. Approximately one genome equivalent was screened, from which 18 strongly hybridizing clones were obtained. Fifteen of the original 18 clones were positive through three additional rounds of plaque purification. Partial sequence analysis of the nine largest cDNA clones showed that each contained open reading frames homologous to the yeast TFIID protein. A 1.5-kb cDNA clone was selected for detailed analysis and

```
1     AAT TCC CCA AAG TTG AGC CCG TTC AGC AAT CTT GAA TTG AGT CAC TGA AAA ATT CAC CGG    60

61    AGT CCA CAA TAA ACC ATC TGT AAG ATG GAC CAA ATG CTA AGC CCC AAC TTC TCG ATT CCG   120
                                      M  D  Q  M  L  S  P  N  F  S  I  P                12

121   AGC ATC GGA ACG CCG CTC CAC CAG ATG GAA GCG GAC CAG CAG ATA GTG GCC AAT CCT GTG   180
13     S  I  G  T  P  L  H  Q  M  E  A  D  Q  Q  I  V  A  N  P  V                     32

181   TAC CAT CCT CCG GCT GTA TCG CAG CCG GAT TCG TTG ATG CCG GCA CCC GGT TCC AGT TCC   240
33     Y  H  P  P  A  V  S  Q  P  D  S  L  M  P  A  P  G  S  S  S                     52

241   GTG CAG CAC CAG CAG CAG CAA CAG CAG TCG GAC GCC AGT GGG GGA TCA GGT CTC TTT GGC   300
53     V  Q  H [Q  Q  Q  Q  Q  Q] S  D  A  S  G  G  S  G  L  F  G                     72

301   CAC GAA CCA TCG CTC CCG CTG GCG CAC AAA CAA ATG CAG AGT TAC CAG CCA TCG GCC TCC   360
73     H  E  P  S  L  P  L  A  H  K  Q  M  Q  S  Y  Q  P  S  A  S                     92

361   TAT CAG CAG CAG CAG CAG CAA CAG CAG CTC CAG AGT CAG GCG CCC GGC GGC GGT GGG AGC   420
93     Y [Q  Q  Q  Q  Q  Q  Q  Q] L  Q  S  Q  A  P  G  G  G  G  S                     112

421   ACT CCG CAG TCC ATG ATG CAG CCG CAG ACG CCG CAG AGC ATG ATG GCC CAC ATG ATG CCC   480
113    T  P  Q  S  M  M  Q  P  Q  T  P  Q  S  M  M  A  H  M  M  P                     132

481   ATG AGT GAG CGG AGT GTG GGC GGT TCG GGG GCC GGA GGT GGC GGA GAT GCC CTG AGC AAC   540
133    M  S  E  R  S  V  G  G  S  G  A  G  G  G  G  D  A  L  S  N                     152

541   ATC CAC CAG ACG ATG GGC CCC TCC ACG CCG ATG ACA CCA GCC ACA CCA GGT TCC GCT GAT   600
153    I  H  Q  T  M  G  P  S  T  P  M  T  P  A  T  P  G  S  A  D                     172

601   CCC GGT ATT GTG CCA CAA CTT CAG AAC ATC GTG TCC ACG GTT AAT CTG TGC TGC AAA CTG   660
173    P  G  I  V  P  Q  L  Q  N  I  V  S  T  V  N  L  C  C  K  L                     192

661   GAC CTC AAG AAA ATA GCA TTG CAT GCG AGA AAC GCC GAG TAC AAT CCT AAG CGA TTT GCG   720
193    D  L  K  K  I  A  L  H  A  R  N  A  E  Y  N  P  K  R  F  A                     212

721   GCT GTG ATT ATG CGA ATC CGA GAG CCC CGG ACC ACC GCC CTT ATT TTC AGC TCC GGC AAG   780
213    A  V  I  M  R  I  R  E  P  R  T  T  A  L  I  F  S  S  G  K                     232

781   ATG GTG TGC ACA GGG GCA AAG AGT GAG GAC GAC TCC AGA CTG GCA GCG AGA AAG TAT GCG   840
233    M  V  C  T  G  A  K  S  E  D  D  S  R  L  A  A  R  K  Y  A                     252

841   CGC ATC ATC CAA AAG CTC GGT TTC CCT GCA AAG TTC CTC GAC TTT AAG ATT CAA AAC ATG   900
253    R  I  I  Q  K  L  G  F  P  A  K  F  L  D  F  K  I  Q  N  M                     272

901   GTC GGC TCC TGC GAT GTC AAG TTC CCC ATA CGC TTG GAA GGC CTG GTG CTG ACC CAT TGC   960
273    V  G  S  C  D  V  K  F  P  I  R  L  E  G  L  V  L  T  H  C                     292

961   AAC TTC AGC AGC TAC GAG CCT GAG CTA TTT CCC GGC TTA ATC TAT CGT ATG GTG CGA CCT  1020
293    N  F  S  S  Y  E  P  E  L  F  P  G  L  I  Y  R  M  V  R  P                     312

1021  CGA ATC GTG CTC CTC ATC TTC GTG TCC GGA AAG GTG GTG CTC ACT GGA GCA AAG GTG CGG  1080
313    R  I  V  L  L  I  F  V  S  G  K  V  V  L  T  G  A  K  V  R                     332

1081  CAG GAG ATC TAC GAT GCC TTC GAC AAG ATA TTC CCC ATT TTA AAG AAG TTC AAG AAG CAG  1140
333    Q  E  I  Y  D  A  F  D  K  I  F  P  I  L  K  K  F  K  K  Q                     352

1141  TCA TAA ATA GGA TAG CGC TTT ATT AGT TCT GTA CGT GTA CGT TTT AAG GTC GGT AGT TCT  1200
353    S  STOP                                                                         353

1201  GGA AGT CTG ATC ATA TGA GTG GGA GCA GCC TGG CGA GCA GCT CCG ATC GAG[AAT AAA]CCA  1260

1261  CAA AGT AAT TTA GCT GTA CGG AAA AAA AAA AAA AAA AAA AA                           1301
```

FIG. 2. Sequence of the *D. melanogaster* TFIID cDNA. The nucleotide sequence was determined by the dideoxy chain-termination method, employing synthetic oligonucleotide primers at ≈150-nucleotide intervals for both strands. The nucleotide sequence is presented in groups of three bases and is numbered from the first A of the *Eco*RI cloning site. The sequence of the 353-amino acid open reading frame is presented in single-letter code. Boxes indicate the positions of the glutamine repeats. The PXT/S triplets are underlined, the glycine clusters are marked with dashes, and the methionine doublets are overlined. Brackets indicate the position of the poly(A)-cleavage/addition signal.

the DNA sequence is shown in Fig. 2. The cDNA contains an open reading frame (initiating with a methionine) of 353 amino acids, which corresponds to a calculated molecular mass of 38,407 Da. This size is considerably smaller than the apparent molecular mass of the native *Drosophila* (100–150 kDa; C.S.P., unpublished data) and human (4) TFIID proteins as determined by size-exclusion chromatography. In addition, the 3′ end of the cDNA contains both a polyadenylylation signal and a terminal poly(A) tract.

An alignment of the *Drosophila*, human (32), and yeast (18) TFIID amino acid sequences (Fig. 3) shows that the carboxyl-terminal 180 residues are highly conserved among all three species, with ≥80% sequence identity. Yeast and *Drosophila* are 88% similar in this region if chemically conserved residues are included. *Drosophila* and human are even more closely related, with 88% identity and 95% of the residues chemically conserved.

In contrast to the carboxyl terminus, the amino-terminal portions of the *Drosophila* and human proteins differ greatly from the yeast protein. The amino-terminal regions of the *Drosophila* and human TFIID proteins, however, do resemble each other in certain respects. *Drosophila* TFIID contains two blocks of glutamine repeats, consisting of 6 and 8 residues, respectively (Fig. 2). The human TFIID protein has one long stretch of 34 glutamine residues. Adjacent to the conserved region is another block of homology, which consists of a triplet repeat, PXT/S (where X is not a conserved residue). Preceding the *Drosophila* triplet repeats are two small glycine-rich blocks separated by three doublets of methionine.

To demonstrate that the cloned cDNA encodes a functional form of *Drosophila* TFIID, the conserved carboxyl-terminal 196 amino acids were inserted into a bacterial T7 expression vector (27). A protein of the expected molecular mass (≈22 kDa) was produced in the induced bacterial cells (Fig. 4*A*). The recombinant protein (p22) was tested for sequence-specific DNA binding by DNase I protection analysis using a DNA fragment derived from the *ftz* promoter (29). The recombinant protein was capable of specific recognition of the functional TATA element of the *ftz* promoter (Fig. 4*B*) and the two upstream A+T-rich, TATA-like sequences (data not shown). The protected region was limited to the TATA sequence and did not extend to the start point of transcription as was initially observed with partially purified preparations of TFIID from *Drosophila* (9) and human (11) cells.

To further demonstrate that a functional form of the *Drosophila* TFIID had been cloned, the recombinant protein was tested for its ability to reconstitute transcription *in vitro*. The assays used heat-treated HeLa cell nuclear extracts in which the endogenous TFIID activity had been inactivated by mild heat treatment (31). Addition of recombinant p22 protein to a heat-treated HeLa cell nuclear extract with the *Drosophila ftz* promoter as template restored specific initiation of transcription (Fig. 4*C*).

DISCUSSION

Drosophila TFIID can be divided into two domains based on inspection of the primary amino acid sequence. The carboxyl-terminal half of the protein is highly conserved among yeast, *Drosophila*, and human. As shown in this report the conserved portion of the *Drosophila* TFIID suffices for specific DNA binding and activation of basal-level *in vitro* transcription. In addition, features previously described for the yeast carboxyl terminus can also be applied to the *Drosophila* and human sequence (18). This includes the direct-repeat elements (20), the Myc homology (34), the high density of basic residues, and the σ homology (18, 32).

Though not as conserved as the carboxyl-terminal TFIID domains, the *Drosophila* and human amino-terminal regions do retain several similar sequence motifs—most noticeably, the long stretch of glutamine residues in human and the two clusters found in *Drosophila*. Additionally, each region contains a cluster of PXT/S triplets adjacent to their conserved carboxyl-terminal domains which may serve as potential sites for phosphorylation.

The *Drosophila* amino terminus is relatively rich in serine and proline residues in areas that surround the glutamine repeats. This type of sequence motif is also found in *Drosophila* and mammalian (35, 36) transcriptional regulatory proteins. For example, the antennapedia (37) and engrailed (38) proteins contain glutamine clusters embedded in regions of high serine and proline content. Caudal protein (39), a direct activator of *ftz* gene transcription (40), does not have glutamine clusters but rather possesses asparagine repeats, which maintain the chemical nature of the motif. The amino-terminal features characteristic of TFIID are also apparent in the hairy protein (41), which is believed to be a repressor of gene expression during early *Drosophila* development (42,

1	*Drosophila*	M D Q M L S P N F S I P S I G T P L H Q M E A D Q Q I V A N P V Y H P P A V S Q P D S L M P A P G S	50
1	Human	N N L P P Y A Q G L A S Q G A T P G I P F - S M M P Y G T G L T Q P I Q N T N S L	49
1	Yeast (*S. c.*)	- E E R L K E -	9
51	*Drosophila*	S S V Q H Q Q Q Q Q Q S D A S G G S G L F G H E P S L P L A H K Q M Q S Y Q P S A S Y Q Q Q Q Q Q Q	100
50	Human	I L E E R Q Q Q Q - - - - - - - - - - - - - - - - Q Q Q Q Q Q Q Q Q Q	84
10	Yeast (*S. c.*)	- F E A N K I V F D P N T R V W E N	29
101	*Drosophila*	Q L Q S Q A P G G G G S T P Q S M M Q P Q T P Q S M M A H M M P M S E R S V G G S G A G G G G D A L	150
85	Human	Q Q Q Q A V A A A A V Q S T S Q A T G T S G Q A P Q L F S Q T L T T A P L P T T P	134
30	Yeast (*S. c.*)	N R D G T K A T T F Q S E E D I K R A T P E S E -	54
151	*Drosophila*	S N I H Q T M G P S T P M T P A T P G S A D P G I V P Q L Q N I V S T V N L C C K L D L K K I A L H	200
135	Human	- - Y P S P T M I A E S S G T R	182
55	Yeast (*S. c.*)	- - - - - - - - - - - - - - - - - K D T T S T A T G R T V	88
201	*Drosophila*	A R N A E Y N P K R F A A V I M R I R E P R T T A L I F S S G K M V C T G A K S E D D S R L A A R K	250
183	Human	Q	232
89	Yeast (*S. c.*)	K A V K S	138
251	*Drosophila*	Y A R I I Q K L G F P A K F L D F K I Q N M V G S C D V K F P I R L E G L V L T H C N F S S Y E P E	300
233	Human	V V Q Q	282
139	Yeast (*S. c.*)	I A T I A F S R T	188
301	*Drosophila*	L F P G L I Y R M V R P R I V L L I F V S G K V V L T G A K V R Q E I Y D A F D K I F P I L K K F K	350
283	Human	I K A E E N Y G R	332
189	Yeast (*S. c.*)	T K K I Q E Q E A Y V S E R	238
351	*Drosophila*	K Q S	353
333	Human	T T	335
239	Yeast (*S. c.*)	M -	240

FIG. 3. Alignment of the deduced amino acid sequences of *Drosophila*, human (32) and *S. cerevisiae* (S. c.; ref. 18) TFIID proteins. Alignments were made using the Needleman and Wunsch algorithm (33). Sequence identities are not shown and gaps (dashes) have been inserted to maximize the alignment.

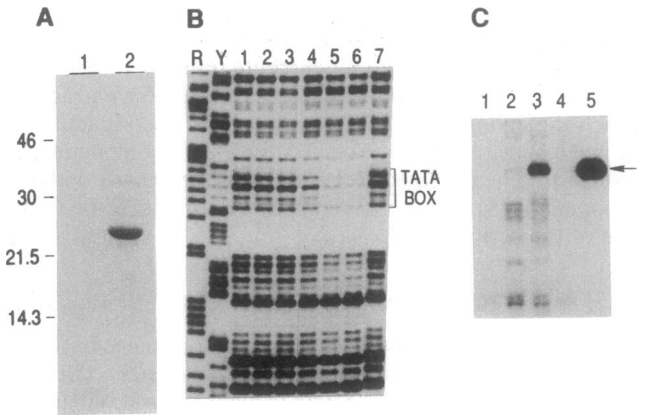

FIG. 4. TATA-consensus binding specificity and basal-level transcription activity resides within the conserved region of the *D. melanogaster* TFIID protein. (*A*) Induced expression of the carboxyl-terminal 196 amino acids of the *Drosophila* TFIID in *E. coli*. The T7 expression vector pET-8c was used to express the carboxyl-terminal 196 residues of the *Drosophila* TFIID in *E. coli* strain BL21(DE3). Cells were grown to OD_{600} of 0.5 and induced with 0.5 mM IPTG for 30 min. The cultures were then incubated for 5 min with [^{35}S]methionine and harvested by centrifugation. Cell pellets were lysed in $NaDodSO_4$ sample buffer at 100°C. Equivalent portions of the lysates were electrophoresed through a $NaDodSO_4$/15% polyacrylamide gel, which was then fixed in 10% acetic acid and dried. An autoradiographic exposure of the gel is shown. Lane 1, cells transformed with pET-8c vector minus insert; lane 2, cells transformed with the vector plus *Drosophila* TFIID insert. (*B*) DNase I footprint protection assays showing binding of the recombinant *Drosophila* TFIID protein (p22) to the *ftz* promoter TATA element. Lane R contains template molecules chemically degraded at purine residues; lane Y contains template molecules chemically degraded at pyrimidine residues; lanes 4–6, protection patterns generated by exposure of an end-labeled *ftz* promoter template to 0.7, 1.4, and 2.8 μg, respectively, of an extract from an IPTG-induced clone expressing the p22 protein. Control binding reactions: lanes 1 and 7, no protein; lanes 2 and 3, respectively, 1 μg and 2 μg of extract from induced cells transformed with the pET-8c vector minus insert. (*C*) Primer extension analysis of recombinant TFIID (p22) activation of *in vitro* transcription. Heat-treated HeLa cell nuclear extract was combined with no exogenous protein (lanes 1 and 4), with 1 μg of extract from IPTG-induced BL21(DE3) cells harboring pET-8c minus insert (lane 2) or pET-8c expressing the p22 protein (lane 3), or with partially purified HeLa cell TFIID (lane 5). Arrow indicates primer-extension products corresponding to specific (+1) initiation events.

43). These observations suggest that these domains may serve related functions in transcription.

The calculated molecular mass of the *Drosophila* TFIID protein (38,407 Da) is considerably smaller than that observed for native TFIID (100–150 kDa). In addition, the DNase I footprint observed with recombinant TFIID is considerably smaller than that observed with native *Drosophila* TFIID. The limited protection observed with the recombinant protein may result from the use of a truncated molecule. Alternatively, full-length TFIID may oligomerize or associate with other proteins to generate the extended footprint and larger observed molecular size.

The isolation of the *Drosophila* TFIID cDNA provides an important reagent for delineation of the regulatory mechanisms governing transcription in higher eukaryotes. It should now be possible with recombinant TFIID to determine the roles of proteins that may associate with TFIID in the initiation reaction, and in regulatory factor interactions.

Note. Following completion of this manuscript a similar report of the cloning of a *Drosophila* TFIID cDNA appeared (44).

We thank Alexander Hoffman (Rockefeller University) for communicating unpublished sequence data. M.L.M. was a recipient of a National Institutes of Health postdoctoral fellowship (F32 GM12244) and a fellowship from the American Cancer Society, California Division (S-38-88). C.T.I. was a recipient of a postdoctoral fellowship from the American Cancer Society, National Division (PF-03083). This research was supported by National Institutes of Health Grant GM29430 to C.S.P.

1. Matsui, T., Segall, J., Weil, P. A. & Roeder, R. G. (1980) *J. Biol. Chem.* **255**, 11992–11996.
2. Samuels, M., Fire, A. & Sharp, P. A. (1982) *J. Biol. Chem.* **257**, 14419–14427.
3. Reinberg, D., Horikoshi, M. & Roeder, R. G. (1987) *J. Biol. Chem.* **262**, 3322–3330.
4. Reinberg, D. & Roeder, R. G. (1987) *J. Biol. Chem.* **262**, 3310–3321.
5. Parker, C. S. (1989) *Curr. Opinion Cell Biology* **1**, 512–518.
6. Workman, J. L., Abmayr, S. M., Cromlish, W. A. & Roeder, R. G. (1988) *Cell* **55**, 211–219.
7. Abmayr, S. M., Workman, J. L., Cromlish, W. A. & Roeder, R. G. (1988) *Genes Dev.* **2**, 542–553.
8. Workman, J. L., Roeder, R. G. & Kingston, R. E. (1990) *EMBO J.* **9**, 1299–1308.
9. Parker, C. S. & Topol, J. (1984) *Cell* **36**, 357–369.
10. Davison, B. L., Egly, J.-M., Mulvhill, E. R. & Chambon, P. (1983) *Nature (London)* **301**, 680–686.
11. Sawadogo, M. & Roeder, R. G. (1985) *Cell* **43**, 357–369.
12. Buratowski, S., Hahn, S., Sharp, P. A. & Guarente, L. (1988) *Nature (London)* **334**, 37–42.
13. Horikoshi, M., Wang, C. K., Fuji, H., Cromlish, J. A., Weil, P. A. & Roeder, R. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2794–2798.
14. Cavallini, B., Huet, J., Plassat, J. L., Sentenac, A., Egly, J.-M. & Chambon, P. (1988) *Nature (London)* **334**, 77–80.
15. Buratowski, M., Hahn, T., Guraente, L. & Sharp, P. A. (1989) *Cell* **56**, 549–561.
16. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. (1988) *Science* **241**, 1335–1338.
17. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. (1989) *Cell* **58**, 1173–1181.
18. Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil, P. A. & Roeder, R. G. (1989) *Nature (London)* **341**, 299–303.
19. Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7785–7789.
20. Cavallini, B., Faus, I., Matthes, H., Chipoulet, J. M., Winson, B., Egly, J. M. & Chambon, P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9803–9807.
21. Eisenmann, D. M., Dollard, C. & Winston, F. (1989) *Cell* **58**, 1183–1191.
22. Horikoshi, M., Yamamoto, T., Okuma, Y., Weil, P. A. & Roeder, R. G. (1990) *Cell* **61**, 1171–1178.
23. Saiki, R. K., Gelferrd, D. M., Stoffel, S., Scharf, S., Higuchi, R. H., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
24. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
25. Muhich, M. L., Simpson, L. & Simpson, A. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4060–4064.
26. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
27. Studier, F. W. & Moffat, B. A. (1986) *J. Mol. Biol.* **189**, 113–130.
28. Hoey, T., Warrior, R., Manak, J. & Levine, M. (1988) *Mol. Cell. Biol.* **8**, 4598–4607.
29. Dearolf, C., Topol, J. & Parker, C. S. (1988) *Genes Dev.* **3**, 384–398.
30. Dignam, J. D., Martin, P. L., Shastry, B. S. & Roeder, R. G. (1983) *Methods Enzymol.* **101**, 582–598.
31. Nakajima, N., Horikoshi, M. & Roeder, R. G. (1988) *Mol. Cell. Biol.* **8**, 4028–4040.
32. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. & Roeder, R. G. (1990) *Nature (London)* **346**, 387–390.
33. Needleman, C. & Wunsch, H. (1970) *J. Mol. Biol.* **48**, 443–453.
34. Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R. G. & Chua, N.-H. (1990) *Nature (London)* **346**, 390–394.
35. Courey, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898.
36. Mitchell, P. J. & Tjian, R. (1989) *Science* **245**, 371–378.
37. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. A. (1986) *EMBO J.* **5**, 733–739.
38. Poole, S. J., Kauvar, L., Drees, B. & Kornberg, T. (1985) *Cell* **40**, 37–43.
39. Mlodzik, M., Fjose, A. & Gehring, W. J. (1985) *EMBO J.* **4**, 2961–2969.
40. Dearolf, C. R., Topol, J. & Parker, C. S. (1989) *Nature (London)* **341**, 340–343.
41. Rushlow, C. A., Hogan, A., Pinchin, S. M., Howe, K. M., Lardelli, M. & Ish-Horowitz, D. (1989) *EMBO J.* **8**, 3095–3103.
42. Howard, K. & Ingham, P. (1986) *Cell* **44**, 949–959.
43. Caroll, S. B. & Scott, M. P. (1986) *Cell* **45**, 113–126.
44. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. & Tjian, R. (1990) *Cell* **61**, 1179–1186.