Isolation and characterization of a cDNA encoding *Drosophila* transcription factor TFIIB

(transcription initiation/*in vitro* transcription/direct repeat/basic repeat/ σ sequence similarity)

SHINYA YAMASHITA*, KEIJI WADA†, MASAMI HORIKOSHI‡, DA-WEI GONG*, TETSURO KOKUBO*, KOJI HISATAKE‡, NOBORU YOKOTANI†, SOHAIL MALIK‡, ROBERT G. ROEDER‡, AND YOSHIHIRO NAKATANI*§

*Laboratory of Molecular Biology, National Institute of Neurological Disorders and Stroke, and †Laboratory of Neurochemistry, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892; and ‡Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, NY 10021

Contributed by Robert G. Roeder, December 20, 1991

ABSTRACT A *Drosophila* cDNA encoding a human transcription factor TFIIB homologue was isolated by PCR methods. The deduced amino acid sequence indicates 85% sequence similarity with human TFIIB, and the corresponding cDNA product expressed in *Escherichia coli* is interchangeable with human TFIIB for both basal and GAL4-VP16-induced transcription. Structural motifs including the direct repeats, basic repeats, and σ sequence similarities are well conserved among *Drosophila*, human, and *Xenopus* TFIIB. However, the N-terminal region of each direct repeat is less conserved among the three species, suggesting the presence of two structural subdomains in the direct repeat. Moreover, the amino acid changes in the N-terminal subdomain produce altered positions of the conserved amino acids between the direct repeats. An overall similarity in general structural features between TFIIB and TFIID τ (the TATA-binding subunit of TFIID) was previously noted. However, in contrast to the sequence divergence reported for the N-terminal domains of TFIID τ from different species, the N-terminal sequence of TFIIB was highly conserved among the species. This suggests that TFIIB has a more rigid structure, consistent with its function as a “bridging” protein between TFIID and RNA polymerase II. Further implications of the TFIIB structure are discussed.

Drosophila is one of the best organisms for studies of morphogenesis, differentiation, and development. Innovative techniques have been used to isolate many regulatory genes required for development, such as maternal, segmentation, and homeotic genes (e.g., reviewed in ref. 1). Although many of the gene products are thought to be transcription factors because of their structural motifs, their biochemical functions have not been well characterized. Our goal is to understand the molecular mechanisms by which such gene products regulate RNA polymerase II activity. To analyze functional interactions between sequence-specific transcription factors, cofactors, general transcription initiation factors, and RNA polymerase II (for reviews, see refs. 2 and 3), it is important to establish an *in vitro* transcription system reconstituted with purified factors. As a first approach, we have begun to isolate cDNAs for the corresponding general transcription initiation factors from *Drosophila*.

In the human system, initiation of transcription on protein-encoding genes is a multistep process that involves the participation of at least six transcription factors (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIG) in addition to RNA polymerase II (refs. 4–8; reviewed in ref. 2). The first step in preinitiation complex formation is TFIID binding to the TATA box in the promoter region, a process that might be stimulated by TFIIA. Subsequently, TFIIB binds to the

promoter-TFIID complex and recruits RNA polymerase II and the associated TFIIF into the complex. Further association of TFIIIE (and possibly TFIIG) completes formation of the preinitiation complex, which is then capable of correctly initiating transcription. Although the level of transcription by these factors is regulated by sequence-specific DNA binding factors, the exact mechanism of action of the latter is unclear. However, considerable evidence suggests that TFIID and TFIIB are targets for specific activators (refs. 9–13; reviewed in refs. 2 and 14). These interactions may alter the rate or extent of preinitiation complex formation and thus regulate transcription. Therefore, analyses of structure-function relationships for TFIID and TFIIB are extremely important for understanding the molecular mechanisms of transcriptional regulation.

Although *Drosophila* homologues for all the factors have not been identified, at least the TFIID, TFIIB, and TFIIIE/F fractions and RNA polymerase II are functionally interchangeable between the *Drosophila* and human systems (15, 16). This functional similarity suggested the presence of evolutionarily conserved structural motif(s) that might make it possible to isolate cDNAs for *Drosophila* factors on the basis of sequence information for the human factors. In the latter case, cDNAs for TFIID τ (the TATA binding subunit of TFIID; refs. 17–19), the small subunit of TFIIF (RAP 30; refs. 20 and 21), TFIIB (22, 23), TFIIIE α (24, 25), and TFIIIE β (25, 26) have been isolated. However, in *Drosophila*, only the TFIID τ cDNA has been cloned (27, 28). Here, we describe the isolation of a cDNA encoding *Drosophila* TFIIB (dTFIIB)[¶] and the functional characterization of its product.

MATERIALS AND METHODS

Cloning of cDNA for dTFIIB. The PCR reaction mixture was heated at 92°C for 10 min, and 2.5 units of *Taq* DNA polymerase (Perkin-Elmer/Cetus) were added to each tube. Amplification of cDNA was achieved with 30 cycles of 1 min at 92°C, 2 min at 35°C, and 2 min at 72°C. The primers, which yielded a fragment homologous to human TFIIB (hTFIIB), were as follows: DB3, 5'-AC(A/C/G/T)TT(C/T)AA(A/G)GA(A/G)AT(A/C/T)TG(C/T)GC-3'; and DB6, 5'-AT(A/G)TC(A/C/G/T)CC(A/G/T)AT(C/T)TC(C/T)TT(C/T)TG-3'. The PCR DNA fragment was used as a probe to screen a *Drosophila* embryo (3–12 hr) cDNA library (29).

Bacterial Expression. An *Nde* I site was introduced at the deduced initiation methionine of dTFIIB (see Fig. 1) by

Abbreviations: dTFIIB, hTFIIB, and xTFIIB, *Drosophila*, human, and *Xenopus* TFIIB, respectively.

[§]To whom reprint requests should be addressed at: National Institutes of Health, Building 36, Room 3D02, Bethesda, MD 20892.

[¶]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M88164).

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.

site-directed mutagenesis (30). The resulting 1.4-kilobase (kb) *Nde* I–*Eco*RI fragment containing the entire open reading frame was subcloned into *Nde* I–*Eco*RI-cut 6His-pET5a, which was constructed by replacing an 82-base-pair (bp) *Xba* I–*Bam*HI fragment of pET5a with a 105-bp *Xba* I–*Bam*HI fragment of pET15b (ref. 31; Novagene, Madison, WI). The resulting plasmid was transformed into *Escherichia coli* BL21(DE3) (ref. 31; Novagene). Induction and purification of the recombinant protein were according to published protocols (22).

In Vitro Transcription. For *in vitro* transcription, a system was reconstituted with recombinant dTFIID γ or a Sephacryl S-300 fraction of partially purified native dTFIID, an hTFIIE/F fraction, partially purified RNA polymerase II, and either an hTFIIB fraction (for human factors, see ref. 9) or recombinant dTFIIB protein as noted. Standard transcription mixtures (25 μ l) contained 35 mM Hepes-KOH (pH 8.0), 60 mM KCl, 6% (vol/vol) glycerol, 3 mM MgCl₂, 100 μ M ATP, 100 μ M UTP, 25 μ M [α -³²P]CTP, 100 μ M 3'-O-methylguanosine triphosphate, RNasin (Promega) at 160 units/ml, 4 μ M ZnCl₂, 10 mM dithiothreitol, and each template DNA (4 μ g/ml) indicated in the figure legends.

RESULTS

Isolation of a *Drosophila* cDNA Encoding the hTFIIB Homologue. The PCR method was used to obtain a probe to isolate a *Drosophila* cDNA encoding the hTFIIB homologue. We designed degenerate oligonucleotide primers derived from the direct-repeat region because the overall structural organization of TFIIB resembles that of TFIID γ , in which the direct-repeat region is well conserved among various species. A combination of primers DB3 and DB6 (see *Materials and Methods*) gave a fragment of the expected size based on the human sequence. The DNA sequence of the PCR fragment and its translated sequence showed 66% and 84% identity, respectively, with the corresponding DNA and protein regions of the hTFIIB sequences. Five positive clones were isolated from a screen of 1×10^6 clones with this fragment as a probe. The complete nucleotide and deduced amino acid sequences of the longest cDNA are shown in Fig. 1. The cDNA contains an open reading frame encoding a polypeptide of 315 amino acids, corresponding to a calculated molecular mass of 34.4 kDa.

Functional Analysis of the Cloned cDNA Product. The recombinant protein was expressed in *E. coli* as a histidine fusion protein and purified (Fig. 2A); it migrated on an SDS/polyacrylamide gel with an apparent molecular mass of 38 kDa. The ability of the purified protein to substitute for hTFIIB in basal transcription in a complementation assay containing other general factors (Fig. 2B) was observed. The recombinant protein was also interchangeable with hTFIIB in transcription induced by the chimeric acidic activator GAL4-VP16 (Fig. 2C). We thus conclude that the cDNA product is a functional dTFIIB.

dTFIIB Is Encoded by a Single Gene. RNA blot experiments with a complete cDNA-derived probe showed that dTFIIB is expressed in Schneider cells as two distinct RNAs of 1.3 kb and 1.6 kb (Fig. 3A). A similar ratio between the two bands was observed under both low [$2 \times$ standard saline citrate (SSC)/0.1% SDS at 25°C for 30 min] and high (0.1 $\times$ SSC/0.1% SDS at 65°C for 30 min) stringency conditions. A "middle" probe (nucleotide positions 574–847, Fig. 1) produced the same results as the entire cDNA probe (data not shown), whereas the "3' end" probe (positions 1332–1470, Fig. 1) hybridized only with the 1.6-kb RNA. These results suggest that the two distinct RNAs are derived from alternative splicing at the 3' untranslated region and/or from alternative polyadenylation. In support of this view, PCR amplification of *Drosophila* cDNA by using five different

```

1  GACCAGTTGATGCTGGTATTAAAAATATATAGCTTACAGAGTACAAATGGCATCGACA
1  M A S T
61  TCGAGACTGGACAACAAGGTGTGCTGCTACGCACACCCGAATCTCCGCTCATCGAG
5  S R L D N N K V C C Y A H P E S P L I E
121 GACTACAGGGCTGGCGATATGATTGCTCCGAGTGTGGTCTGGTGGTGGGACCGGTA
25 D Y R A G D M I C S E C G L V V G D R V
181 ATTGATGTGGGCTCCGAGTGGCGTACCTTTAGCAACGAGAAGCGCGGTGGATCCAGT
45 I D V G S E W R T F S N E K S G V D P S
241 CGTGTGGTGGACCGGAGAAATCCGCTGCTAGTGGCGGCACTTCCACCATAATTGGT
65 R V G G P E N P L L S G D L S T I I G
301 CCGGGCACAGGATCCGCATCCTTTGACGCCCTTTGGAGCACCAAGTACCAAAACAGACGC
85 P G T G S A S F D A F G A P K Y Q N R R
361 ACCATGAGCAGCTCGGATCGCTCCCTCATATCAGCCTTTAAGGAGATATCTCGATGGCC
105 T M S S S D R S L S L A K K E I S L I G
421 GATCGCATTAACTTGCCAAAACCATTTGCGATCGGGCCAACTTGTTCACACAGTT
125 D R I N L P K T I V D R A N N L F K Q V
481 CATGATGGAAGAAGCTCAAGGGTCTGCTCAATGATGCCAAGGACATCCGCTGTCTGTAT
145 H D G K N L K R G L S D G A K A S I I G
541 ATTGCTGTGCGCAAGAAGGAGTCTCGCACCATTCAAGGAATATGTGTGTGTCAGCAAG
165 I A C R Q E G V P R T F K E I C A V S K
601 ATCAGCAAAAAGAGATGTGGCGTTGCTTTAAGCTTAACATGAAGGCTCTCGAGCAAGT
185 I S K K E I G R C F T L K L E T S
661 GTGGATCTCATAACCACTGCGGACTTTATGTGCCGTTCTGCGCAACTTGGACCTGCCG
205 V D L I T T A D F M C R F C A N L D L P
721 AACATGGTCCAACGCGCTGCCACGCATTCGAAAGAAGCCGTGAGATGGATTTGTA
225 N M V Q R A A T H I A K A V E M D I V
781 CCGGAGCGTTCGCCAATTTCTGTGGCGCGACCGCAATCTATATGGCCTCACAAGCATCC
245 P G R S P I S V A A A A I Y M A S Q A S
841 GAGCATAAGAGAAGCAGGAAGAGATCGGTGATAGCGCGGTGTGGCCAGTCAACCATC
265 E H K R S Q K E I G D I A G V A D V T I
901 AGACAGTCTCACAACCTTATGTATCCGACGCGAGCTTCTCCCGAGGAGCTTTAAG
285 R Q S Y K L M Y P H A A K L F P E D F K
961 TTTACCACTCCCATTTGATCAGTTACCAGAGATGCTTTCTTTTCTGTAGATCCGATGCTCA
305 F T T P I D Q L P Q M
1021 ATACTAACGCTAAGCTTAAGTGAAGGAAAAATCTTAATTAATCACAATACGATTTTA
1081 TTAATATAAGTTAGCATAACTAATATTCAATAGAAGCTTTGGACCAATCAGAACATAAT
1141 TAAATCAATGTTAAATTAATATCTGAGTTAAACCCCTTCAGTTTGTGAATTTATTCCA
1201 ATGATTCGTAATACTTTGCTTGGTTGATTTATTTTACATTTTGTGTATCTTTGGTAT
1261 AATATAAGCTAATAACCAAGAAACAAAATTCGACACTACGTTTATTGGCAGTTATTGGC
1321 GAAATAAGTTTAAACAAACCGTACTCTGTTTATACATAAGCTTTAAGTAATTAGTAAT
1381 AATATAAATGGATAAAAAATCGCAATAAGACGCAATCTGTAACCTTCAACATTATC
1441 AATTGCAATAAAGCGGGCGTTTAAACGTAAAAAATAAAAAAAAAA

```

FIG. 1. Nucleotide and deduced amino acid sequence of a *Drosophila* cDNA encoding an hTFIIB homologue. The open reading frame is defined by translation start and stop codons (boxed in the nucleotide sequence). The putative polyadenylation signals (5'-AATAAA-3' and 5'-ATTAAA-3') are underlined.

combinations of primers derived from the dTFIIB coding sequence gave only a single band in each combination (data not shown). Moreover, genomic DNA blotting analysis showed a single band (Fig. 3B). Additional potential polyadenylation signals were found in the 3' untranslated region (Fig. 1), although it is unknown which signal is functional in the cell.

Structure of dTFIIB. Comparison of the predicted amino acid sequence of the dTFIIB with hTFIIB and *Xenopus* TFIIB (xTFIIB) sequences revealed 79% identity and 85% similarity, respectively, with complete continuity, except that hTFIIB and xTFIIB (33) contain one extra amino acid between positions 9 and 10 of the *Drosophila* sequence (Fig. 4A). Like the other species of TFIIB, dTFIIB also contains an imperfect direct repeat (Fig. 4A). However, the N-terminal portions (structural subdomain 1) of each repeat show less sequence similarity than do the C-terminal portions (structural subdomain 2) (Fig. 4A and B). Interestingly, in subdomain 1, the amino acid changes produce altered positions of conserved amino acids between the direct repeats (Fig. 4B). These results suggest that the indicated subdomains of the direct repeat may have distinct functions and that the altered subdomain 1 could reflect variable functions or interactions of TFIIB in the different species. All basic amino acids in the basic repeat regions are conserved among the three species (Fig. 4A and C). The areas with sequence similarity to σ regions 2.1/2.2 or 2.4 (34) are well conserved among the three species (Fig. 4D).

FIG. 2. Expression and functional characterization of the cDNA product. (A) Expression and purification of the recombinant protein. Fifty micrograms of uninduced (lane 1) and isopropyl β -D-thiogalactoside-induced (lane 2) soluble protein fractions and 1 μ g of protein eluted from the Ni-agarose column (lane 3) were analyzed and stained with Coomassie brilliant blue R-250. The position of the purified protein is indicated by an arrow. (B) Basal transcription mediated by the recombinant protein. A plasmid containing the adenovirus major late core promoter attached to the 380-bp G-less cassette (32) was transcribed in a complementation assay in the presence of either hTFIIB (lane 1), no TFIIB (lane 2), or increasing amounts of dTFIIB (lane 3, 0.1 ng; lane 4, 1 ng; lane 5, 10 ng). The position of the accurately initiated transcript is shown by the arrow. (C) Activator-induced level of transcription mediated by the recombinant protein. A plasmid with GAL4 binding sites and the human immunodeficiency virus core promoter attached to the 380-bp G-less cassette (G6HIV) and a control plasmid with the adenovirus major late core promoter attached to the 232-bp G-less cassette (AdMLP) were cotranscribed in a reconstituted system with dTFIIB (lanes 1 and 2) or no TFIIB (lanes 3 and 4). GAL4-VP16 (44) was added to lanes 1 and 3. The positions of the accurately initiated transcripts from each template are shown by arrows.

DISCUSSION

Overall Structure of dTFIIB. TFIIB has an overall structural organization similar to that of TFIID τ (22, 23). Both contain similar motifs, which include an imperfect direct repeat, basic repeats, and sequence similarities to σ ; moreover, both are likely to function as monomers (22, 23, 35). The various species of TFIID τ contain almost perfectly conserved C-terminal cores of 180 residues and diverse N-terminal sequences (reviewed in ref. 17). Between the *Drosophila* and human factors, the sequence conservation in the C-terminal conserved region of TFIID τ (88% identity and

95% similarity) is considerably higher than that in TFIIB (79% identity and 85% similarity). However, whereas the N-terminal sequences of TFIID τ diverge markedly in different species, the N-terminal sequences of TFIIB are highly conserved among the species (Fig. 4A). This conservation of structure is consistent with the observed requirement of the hTFIIB N terminus for basal transcription (K.H., S.M., M.H., and R.G.R., unpublished results).

Unlike TFIID τ , which contains 2–5 cysteine residues, dTFIIB contains 10 cysteine residues, and the positions of 8 cysteine residues are conserved among the three species. Mutagenesis studies showed that the cysteine residues in *Saccharomyces cerevisiae* TFIID τ are not required for basal transcription activity (35), whereas the functional importance of the sulfhydryl groups in TFIIB was suggested by the sensitivity of partially purified hTFIIB to *N*-ethylmaleimide treatment (36). These conserved cysteine residues might play a role in sensing the intracellular redox potential (37) and/or a role in effecting a rigid TFIIB structure that could be related to the function of TFIIB as a “bridging” factor between TFIID and RNA polymerase II/TFIIF, which interact at the TATA box and the transcription initiation site, respectively. The fact that the spacing between the TATA box and the initiation site is well conserved among *Drosophila*, human, and *Xenopus* genes might be associated with the conserved size of TFIIBs among the three species. If this is the case, *S. cerevisiae*, whose genes contain variable spacing (40 to 120 bp) between the TATA box and initiation site (38), may be expected to have a somewhat distinct TFIIB or possibly to employ another factor for initiation site selection.

Structural Motifs in TFIIB. Both TFIIB (22, 23) and TFIID τ (ref. 39; reviewed in ref. 17) contain imperfect direct repeats, although the amino acid sequences in the repeats are distinct between the two molecules. A detailed mutational analysis of yeast TFIID τ showed that the residues comprising the direct repeats are essential for DNA binding (35, 40, 41). TFIIB interacts with the DNA–TFIID complex and affects the DNase I footprinting region (5, 6). These results suggest that TFIIB interacts with DNA, albeit in a sequence-independent manner,

FIG. 3. Analysis of genomic DNA and mRNA encoding TFIIB. (A) mRNA analysis. *Drosophila* RNA was hybridized with a random-primered probe containing either total cDNA sequences (lanes 1 and 2) or only 3' sequences (lane 3). The filter was washed under either low (L; lanes 1 and 3) or high (H; lane 2) stringency conditions. (B) Genomic DNA analysis. *Bam*HI (lane 1) and *Eco*RI (lane 2)-digested *Drosophila* genomic DNA was hybridized with a nick-translated probe containing “middle region” dTFIIB sequences.

FIG. 4. Structural analysis of TFIIB. (A) Comparison of deduced amino acid sequences of dTFIIB, hTFIIB, and xTFIIB. Only nonidentical amino acids are shown. Similar amino acids are shaded according to the following groupings (E, D), (K, R, H), (N, Q), (A, I, L, M, V), (F, W, Y), (S, T), (C), (G), and (P). hTFIIB and xTFIIB each contain a single amino acid insertion between residues 9 and 10 of the *Drosophila* sequence. The direct repeat is indicated by arrows. Basic amino acids in the "basic repeat" regions are marked "+" above the sequence. (B) TFIIB direct repeat. The repeated regions of dTFIIB, hTFIIB, and xTFIIB are aligned. The direct repeat can be subdivided into regions 1 and 2 (see text). Identical and similar amino acids are shaded. Altered positions of conserved amino acids between the species are indicated by vertical lines. (C) Helical wheel depiction of the first (Left) and second (Middle) basic repeats in dTFIIB and the dTFIID basic repeat (Right). (D) σ sequence similarities. σ subregions 2.1, 2.2, and 2.4 are aligned with TFIIB. In region 2.2, two alternative alignments are shown. Identical and similar amino acids between TFIIB and the σ factor are shaded. Note that the xTFIIB sequence is identical to the hTFIIB sequence in the compared regions except that serine (S) at position 205 is replaced with asparagine (N) in the xTFIIB sequence.

and that the direct repeats (like those in TFIID γ) may be important for these interactions. Among TFIIBs from the three species, less-conserved regions are found in the N-terminal regions of each repeat. Interestingly, the amino acid changes alter the positions of the conserved amino acids found between the direct repeats (Fig. 4B). These findings suggest that the direct repeats can be subdivided into structural subdomains 1 and 2 (Fig. 4B) and that the divergent amino acid sequence in subdomain 1 could reflect subtle differences in the function of TFIIB, or in associated protein-protein interactions, in the

different species. It is noteworthy that amino acid sequences of two potential hydrophobic helical regions in the second repeat (residues 226–246 and 251–266 in hTFIIB; ref. 22) are well conserved among the three species.

As previously reported, hTFIIB contains two basic repeat regions near the N-terminal and C-terminal regions of the first direct repeat (22), and these regions are predicted to form an α -helical structure (Fig. 4C). In these regions, basic amino acids are found at regular intervals (Fig. 4A); some of these residues have been postulated to be possible targets for acidic

activators. It is worthwhile to note that all the basic amino acids in these regions are perfectly conserved among different species, although amino acid changes are observed in the intervening residues. The helical wheel depictions of the basic repeat regions of dTFIIB and dTFIID τ are shown in Fig. 4C. Interestingly, the positions of the basic amino acids are somewhat more dispersed in dTFIID τ , whereas they are concentrated in dTFIIB. This suggests that the basic regions of TFIIB and TFIID τ might be recognized by distinct domains of an acidic activator like GAL4 or GAL4-VP16 and/or that interactions of the same activator domain with TFIIB and TFIID τ might vary in strength.

Human transcription factors TFIIE α/β (24, 26) and RAP 30 (the small subunit of TFIIF) (20), which interact with RNA polymerase II, contain regions with sequence similarity to σ subregions 2.1/2.2 (reviewed in ref. 34). This region is important for σ binding to core RNA polymerase, and recent studies indicate that the related region in RAP 30 is important for binding to RNA polymerase II (42). TFIIB likely interacts with RNA polymerase II when RNA polymerase II binds to the DNA-TFIID-TFIIB complex. The sequence of the σ 2.1/2.2-related region is well conserved among TFIIB from the three species (Fig. 4D), consistent with a possible role in RNA polymerase interactions. Like the TATA-binding factor TFIID τ (43), TFIIB also has a region with sequence similarity to the σ subregion 2.4 (23), a region important for DNA binding. Although this sequence similarity was originally noted for the major σ factors (23), this region of TFIIB also has sequence similarity, albeit reduced, to various minor σ factors (Fig. 4D). Along with the observation that TFIIB affects the DNase I footprinting region of TFIID (5, 6), this suggests that the σ 2.4-related region may be involved (with the direct repeat) in DNA binding. We have not discussed the previously noted sequence similarity between TFIIB and the RNA polymerase β subunit (23) because the sequence in question is poorly conserved even in RNA polymerase from the different species and might not be significant. Because the sequence similarities between σ factors and the eukaryotic initiation factors are low, it is important to determine if the relevant regions of all the eukaryotic factors are functional and truly homologous to the bacterial counterparts (as shown for RAP 30; ref. 42).

In conclusion, we have isolated the cDNA encoding *Drosophila* TFIIB and demonstrated structural similarities and dissimilarities between TFIIB from the various species and between TFIIB and TFIID τ . The sequence information for TFIIB from the various species should be useful in studies of structure-function relationships.

We are grateful to the members of C. Wu's laboratory for considerable advice on fly techniques, A. Hoffmann for assistance on Ni-agarose affinity chromatography, R. Bernstein for plasmid G6HIV and GAL4-VP16 protein, T. Kornberg for the *Drosophila* cDNA library, K. Doi for oligonucleotides, R. Wenthold for his support in carrying out this study, M. Brenner for a critical reading of the manuscript, and D. Schoenberg for editing the manuscript. S.Y. is supported by Nippon Suisan Kaisha, Ltd. M.H. is an Alexandrine and Alexander L. Sinsheimer Scholar. S.M. is supported by Fellowship GM13244 from National Institutes of Health. A part of this study was supported by National Institutes of Health Grants CA42567 and AI27397 (to R.G.R.) and GM45258 (to M.H.), by funds from Sankyo Co., Ltd. (to M.H.), and by general support from the Pew Trusts to The Rockefeller University.

1. Scott, M. P. & Carroll, S. B. (1987) *Cell* **51**, 689-698.
2. Roeder, R. G. (1991) *Trends Biochem. Sci.* **16**, 402-408.
3. Ptashne, M. & Gann, A. A. (1990) *Nature (London)* **346**, 329-331.
4. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. (1988) *Science* **241**, 1335-1338.
5. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. (1989) *Cell* **56**, 549-561.

6. Maldonado, E., Ha, I., Cortes, P., Weis, L. & Reinberg, D. (1990) *Mol. Cell. Biol.* **10**, 6335-6347.
7. Inostroza, J., Flores, O. & Reinberg, D. (1991) *J. Biol. Chem.* **266**, 9304-9308.
8. Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. & Roeder, R. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9158-9162.
9. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. (1988) *Cell* **54**, 665-669.
10. Horikoshi, M., Hai, T., Lin, Y. S., Green, M. R. & Roeder, R. G. (1988) *Cell* **54**, 1033-1042.
11. Workman, J. L., Abmayr, S. M., Cromlish, W. A. & Roeder, R. G. (1988) *Cell* **55**, 211-219.
12. Stringer, K. F., Ingles, C. J. & Greenblatt, J. (1990) *Nature (London)* **345**, 783-786.
13. Lin, Y. S. & Green, M. R. (1991) *Cell* **64**, 971-981.
14. Sharp, P. A. (1991) *Nature (London)* **351**, 16-18.
15. Parker, C. S. & Topol, J. (1984) *Cell* **36**, 357-369.
16. Wampler, S. L., Tyree, C. M. & Kadonaga, J. T. (1990) *J. Biol. Chem.* **265**, 21223-21231.
17. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. & Roeder, R. G. (1990) *Nature (London)* **346**, 387-390.
18. Peterson, M. G., Tanese, N., Pugh, B. F. & Tjian, R. (1990) *Science* **248**, 1625-1630.
19. Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R. & Berk, A. J. (1990) *Science* **248**, 1646-1650.
20. Sipta, M., Burton, Z. F. & Greenblatt, J. (1989) *Nature (London)* **341**, 410-414.
21. Horikoshi, M., Fujita, H., Wang, J., Takada, R. & Roeder, R. G. (1991) *Nucleic Acids Res.* **19**, 5436.
22. Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M. & Roeder, R. G. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9553-9557.
23. Ha, I., Lane, W. S. & Reinberg, D. (1991) *Nature (London)* **352**, 689-695.
24. Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M. & Roeder, R. G. (1991) *Nature (London)* **354**, 398-401.
25. Peterson, M. G., Inostroza, J., Maxon, M. E., Flores, O., Admon, A., Reinberg, D. & Tjian, R. (1991) *Nature (London)* **354**, 369-373.
26. Sumimoto, H., Ohkuma, Y., Sinn, E., Kato, H., Shimasaki, S., Horikoshi, M. & Roeder, R. G. (1991) *Nature (London)* **354**, 401-404.
27. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. & Tjian, R. (1990) *Cell* **61**, 1179-1186.
28. Muhich, M. L., Iida, C. T., Horikoshi, M., Roeder, R. G. & Parker, C. S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9148-9152.
29. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985) *Cell* **40**, 37-43.
30. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367-382.
31. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60-89.
32. Sawadogo, M. & Roeder, R. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4394-4398.
33. Hisatake, K., Malik, S., Roeder, R. G. & Horikoshi, M. (1991) *Nucleic Acids Res.* **19**, 6639.
34. Helmann, J. D. & Chamberlin, M. J. (1988) *Annu. Rev. Biochem.* **57**, 839-872.
35. Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P. A. & Roeder, R. G. (1990) *Cell* **61**, 1171-1178.
36. Reinberg, D. & Roeder, R. G. (1987) *J. Biol. Chem.* **262**, 3310-3321.
37. Storz, G., Tartaglia, L. A. & Ames, B. N. (1990) *Science* **248**, 189-194.
38. Struhl, K. (1989) *Annu. Rev. Biochem.* **58**, 1051-1077.
39. Cavallini, B., Faus, I., Matthes, H., Chipoulet, J. M., Winsor, B., Egly, J. M. & Chambon, P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9803-9807.
40. Reddy, P. & Hahn, S. (1991) *Cell* **65**, 349-357.
41. Yamamoto, T., Horikoshi, M., Wang, J., Hasegawa, S., Weil, P. A. & Roeder, R. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2844-2848.
42. McCracken, S. & Greenblatt, J. (1991) *Science* **253**, 900-902.
43. Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil, P. A. & Roeder, R. G. (1989) *Nature (London)* **341**, 299-303.
44. Chasman, D. I., Leatherwood, J., Carey, M., Ptashne, M. & Kornberg, R. D. (1989) *Mol. Cell. Biol.* **9**, 4746-4749.