Cell, Vol. 58, 1173–1181, September 22, 1989, Copyright © 1989 by Cell Press

Isolation of the Gene Encoding the Yeast TATA Binding Protein TFIID: A Gene Identical to the *SPT15* Suppressor of Ty Element Insertions

Steven Hahn,* Stephen Buratowski,†‡ Phillip A. Sharp,‡ and Leonard Guarente‡
*Fred Hutchinson Cancer Research Center
1124 Columbia Street
Seattle, Washington 98104
†Center for Cancer Research and
‡Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Summary

We report the cloning of the gene that encodes the yeast TATA binding protein TFIID. TFIID contains 240 amino acids and has no obvious sequence similarity to other known proteins. TFIID was synthesized in vitro and in two separate assays behaved identically to the protein purified from yeast. TFIID bound to TATA elements from the adenovirus major late promoter (TATAAAA) and the yeast *LEU2* promoter (TATTTAA) and formed protein–DNA complexes stable to electrophoresis only in the presence of TFIIA. In vitro–synthesized yeast TFIID also complemented a mammalian in vitro transcription system that lacked TFIID. Comparison of the yeast TFIID gene with the yeast *SPT15* gene (suppressor of Ty element insertions) showed that the two genes are identical. This finding indicates that the yeast TFIID activity defined in vitro is responsible for specific transcription in vivo.

Introduction

The TATA element is an essential component of many promoters transcribed by RNA polymerase II. The binding of the general transcription factor TFIID to the TATA element is the first step in the pathway of transcription initiation by RNA polymerase II (Davison et al., 1983, Fire et al., 1984; Van Dyke et al., 1988; Buratowski et al., 1989). This TFIID–DNA complex nucleates the assembly of the other general transcription factors TFIIA, TFIIB, and TFIIE, and RNA polymerase II into an active transcription complex (Van Dyke et al., 1988; Buratowski et al., 1989). It is likely that one or more steps in this assembly pathway serve as targets for regulation by gene-specific transcriptional activators and repressors.

Recent work has demonstrated the striking conservation among eukaryotes in the function of TFIID (Buratowski et al., 1988; Cavallini et al., 1988). Yeast TFIID can substitute for its mammalian counterpart in an in vitro transcription reaction, suggesting the conservation of functional interactions between TFIID and the rest of the transcription machinery. At least one other general transcription factor, TFIIA, is also functionally conserved between yeast and mammals (Hahn et al., 1989b). In addition to this conservation in the general transcription machinery, mechanisms governing its stimulation by transcriptional activators are also conserved in eukaryotes. The yeast GAL4 will activate transcription in higher cells (Kakidani and Ptashne, 1988; Webster et al., 1988), and the c-*jun* oncogene product and the human glucocorticoid and estrogen receptors will function in yeast (Struhl, 1988; Metzger et al., 1988; Schena and Yamamoto, 1988). Purification of general transcription factors from yeast and the cloning of their genes will assist in the understanding of the mechanisms of transcriptional control.

The yeast TFIID protein has been purified and shown to be a monomer of about 27 kd (Buratowski et al., 1988; Hahn et al., 1989). The protein binds with moderate affinity and high specificity to yeast and mammalian TATA elements, which match the consensus TATA sequence (TATAAA and TATATA), and to at least several TATAs that differ from the consensus (Hahn et al., 1989a). Further, TFIID will promote transcription from these various TATA elements in vitro.

The yeast transposable element Ty1 can generate mutations by inserting itself into the promoter regions of yeast genes (Roeder et al., 1980; Errede et al., 1980). Recombination via the 350 base direct repeats at the ends of Ty1 leaves the repeat, a delta element, inserted at the promoter. Several of these delta element insertions have been isolated in the *HIS4* promoter that cause a His^- phenotype. This phenotype results because transcription initiates within the delta element instead of the normal *HIS4* site, generating a nontranslatable *HIS4* mRNA. Revertants that generate a His^+ phenotype define a set of genes termed *SPT* (suppressor of Ty insertions) that restores initiation of at least some fraction of the transcripts at the normal *HIS4* site (Winston et al., 1984; Winston et al., 1987; Fassler et al., 1988). Thus, the *SPT* genes are candidates for loci that encode components of the transcriptional machinery.

Here we describe the cloning and initial characterization of the gene encoding yeast TFIID. In this work and in the accompanying paper (Eisenmann et al., 1989), it is shown that the TFIID gene is identical to a gene isolated as a suppressor of Ty insertions, *SPT15*.

Results

Isolation of the TFIID Gene

About 2 μg of the 27 kd yeast TFIID polypeptide was purified and subjected to total amino acid composition and N-terminal sequence analysis (see Experimental Procedures). Fifteen N-terminal residues were determined with high confidence: A-EERLKEFKEANKI-F (where – indicates an undetermined amino acid). The sequence of the central 13 consecutive amino acids was then used to clone the TFIID gene. As described in Experimental Procedures, two degenerate 14 base oligonucleotides were used as primers in the polymerase chain reaction (PCR) to amplify a 39 bp yeast genomic DNA fragment. This fragment was sequenced and shown to encode the 13 TFIID amino acids. The sequence of this fragment was used to

Cell
1174

```
                .         .         .         .         .         .
   1  GAATTCGTTCAAGTGGTCCGTAATATTCCCGTCTTTACAAAGCTGGATTACCATCTCTAA   60
                .         .         .         .         .         .
  61  TGCCAACTTCCATGCATATAGCTCAGGCCCCACCGTGTGCAGCTCCGTGCTTCGCAGCTC   120
                .         .         .         .         .         .
 121  CTGCAGAGCATCCTCGGGGATTGGGAACCTCTCATTTAGCAAGTAATTCACATAACACAG   180
                .         .         .         .         .         .
 181  ATTTAGAAACCATTTCCATTGTGACTTTTCCCGACATTGCGAGAGTAGCCCATGAAAACT   240
                .         .         .         .         .         .
 241  CGTCTTCACCCTGCGGTGCTGTTTCAGCTTAATGCAAAGCATCACGCCGACATACTGGAA   300
                .         .         .         .         .         .
 301  TACGGATGCCCAATTTTGATACAACTCATCCTGCAAATTTACCATGTACTGGACTAATTC   360
                .         .         .         .         .         .
 361  ATTGCAATTTCTTAGTGCAATCTTATAGTGGAACTTACTGTCTCTCATAAGTGGCAAGTC   420
                .         .         .         .         .         .
 421  ATGTAACAGCAGAAACTCGCAACGCATGATCTCTTCTACCAAATCTGTGTCGCTCTGGTG   480
                .         .         .         .         .         .
 481  CGTTTGTAACCGTTCTTTCAAACTGGAAATGTAAAGCTCTGCTAGGTCAAAATTATACGT   540
                .         .         .         .         .         .
 541  CTCCTGTATCAATAACTCCACCATCTCAAACGTGACCTTACTATCCTCCAGAACTGAAAG   600
                .         .         .         .         .         .
 601  CGTACATTTCGTTTTCAATAGCTGAAACATCTGGATAGACATGTTCATGAGGCCATAATA   660
                .         .         .         .         .         .
 661  CTGCTTCAACCCTTCCTCAGAACCGATTTTATTCGCAATTGATATGCATGGTCTCTGTAT   720
                .         .         .         .         .         .
 721  TCCTGTGCTAAGTGGTATACTTGTGAAATACTAAGTTTGTCGCCAAGATTTTCCATGAAT_  780
                .         .         .         .         .         .
 781  TTGTACTTCTTTCGAAATCGTTCAATTTCTACCAATACTGATTCCCCTCTGATAGCTGAG   840
                .         .         .         .         .         .
 841  ATGTCGGGATTCCCTTTGCTGATAGATCTAACTCATCTCTTTACGTATTTTAATTGTGAA   900
                .         .         .         .         .         .
 901  GCCGTAAATAGTTATCTTCCAAGTTTCTCTTACGCGAGCTTTTTGGGAAAAGAAAAAAAT   960
                .         .         .         .         .         .
 961  TTGAAGATCTACATATAAAACATGGCTTCAAAGGATTACTAATGACTTTTTTTACCTTGA   1020
                .         .         .         .         .         .
1021  TAGGTATTCTTGATGGTAAGAGTAAACAAGGGACGTGAAAATTACAGTAGTTACTGTTTT   1080
                .         .         .         .         .         .
1081  TTTTGGACTATAAGATCGGGGGAAAGATAACACATAAGAAATAAAACGACTACTAGTTAG   1140
                .         .         .         .         .         .
1141  ACTGCTCTGCGGAAGAAGCAAGGAAGTAAAGGCTGCATTTTATTTTTCTTTTCTAGTCCA   1200
                .         .         .         .         .         .
1201  ACATAAACAGGTGTATCAAGAGAAACTTTTTTAATTATGGCCGATGAGGAACGTTTAAAG   1260
                                              M   A   D   E   E   R   L   K
                .         .         .         .         .         .
1261  GAGTTTAAAGAGGCAAACAAGATAGTGTTTGATCCAAATACCAGACAAGTATGGGAAAAC   1320
      E   F   K   E   A   N   K   I   V   F   D   P   N   T   R   Q   V   W   E   N
                .         .         .         .         .         .
1321  CAGAATCGAGATGGTACAAAACCAGCAACTACTTTCCAGAGTGAAGAGGACATAAAAAGA   1380
      Q   N   R   D   G   T   K   P   A   T   T   F   Q   S   E   E   D   I   K   R
                .         .         .         .         .         .
1381  GCTGCCCCAGAATCTGAAAAAGACACCTCCGCCACATCAGGTATTGTTCCAACACTACAA   1440
      A   A   P   E   S   E   K   D   T   S   A   T   S   G   I   V   P   T   L   Q
                .         .         .         .         .         .
1441  AACATTGTGGCAACTGTGACTTTGGGGTGCAGGTTAGATCTGAAAACAGTTGCGCTACAT   1500
      N   I   V   A   T   V   T   L   G   C   R   L   D   L   K   T   V   A   L   H
                .         .         .         .         .         .
1501  GCCCGTAATGCAGAATATAACCCCAAGCGTTTTGCTGCTGTCATCATGCGTATTAGAGAG   1560
      A   R   N   A   E   Y   N   P   K   R   F   A   A   V   I   M   R   I   R   E
```

```
1561  CCAAAAACTACAGCTTTAATTTTTGCCTCAGGGAAAATGGTTGTTACCGGTGCAAAAAGT  1620
      P  K  T  T  A  L  I  F  A  S  G  K  M  V  V  T  G  A  K  S

1621  GAGGATGACTCAAAGCTGGCCAGTAGAAAATATGCAAGAATTATCCAAAAAATCGGGTTT  1680
      E  D  D  S  K  L  A  S  R  K  Y  A  R  I  I  Q  K  I  G  F

1681  GCTGCTAAATTCACAGACTTCAAAATACAAAATATTGTCGGTTCGTGTGACGTTAAATTC  1740
      A  A  K  F  T  D  F  K  I  Q  N  I  V  G  S  C  D  V  K  F

1741  CCTATACGTCTAGAAGGGTTAGCATTCAGTCATGGTACTTTCTCCTCCTATGAGCCAGAA  1800
      P  I  R  L  E  G  L  A  F  S  H  G  T  F  S  S  Y  E  P  E

1801  TTGTTTCCTGGTTTGATCTATAGAATGGTGAAGCCGAAAATTGTGTTGTTAATTTTTGTT  1860
      L  F  P  G  L  I  Y  R  M  V  K  P  K  I  V  L  L  I  F  V

1861  TCAGGAAAGATTGTTCTTACTGGTGCAAAGCAAAGGGAAGAAATTTACCAAGCTTTTGAA  1920
      S  G  K  I  V  L  T  G  A  K  Q  R  E  E  I  Y  Q  A  F  E

1921  GCTATATACCCTGTGCTAAGTGAATTTAGAAAAATGTGATGGGGAAGGAGTAGACGAAAA  1980
      A  I  Y  P  V  L  S  E  F  R  K  M

1981  GAAAAAAAGGTTTTCTATTTGTTCCATTTTCTCAATTATTAATGGTCCTCAAAGAAATAA  2040

2041  AAGAAAAGGAAGAAGAAGTAATTGTAATATCAAACGGTTTTTTATAGTATATTCTTCTTA  2100

2101  TTCTATATTTATATATCAATGTTTTATAATAAGATGTTTATTCATAGCATATCTGGTGGA  2160

2161  TCGTCTCTATTAAGCGCCAGCGAGGTGTTTGCCTCTGCATTTTTCAGCAAAGCAAGCTCC  2220

2221  CTTTCCAGCTTGAATCTATGTTCACGCTCATCCGACAATTCTTTTTCATACTTTCTTTGT  2280

2281  GTACTCGTAAGCACTTTTTTTAAACTCACTTGTCATTATTGAAAGTGAACGTGATCCAGAA  2340

2341  CCGCTTGTGGGGCTTCCTACAGAGGAAGGTGAACTTGGATCCCAAGTCACTGGCGAACTC  2400

2401  GCTGGTGATGACATGCCGAAATTATGTCTGCTTGAATTC  2439
```

Figure 1. Sequence of the Yeast TFIID Gene

Shown is the sequence of a 2.44 kb EcoRI genomic DNA fragment with the translated 240 amino acid TFIID open reading frame.

design a unique probe for the TFIID coding sequence. As measured in a yeast genomic Southern blot, this 29 base probe hybridized to a single chromosomal locus at high stringency (data not shown). The probe was used to isolate λ phage from a yeast genomic library, all of which contained a common 2.4 kb EcoRI fragment. The sequence of this fragment is shown in Figure 1.

The EcoRI fragment encodes an open reading frame of 240 amino acids with a predicted molecular weight of 27,000. Residues 2–18 of this open reading frame correspond exactly with the amino acid sequence of TFIID. Experiments described in the next section provide proof that this polypeptide specifies the previously defined TFIID activity. At neutral pH, TFIID has a net basic charge with a predicted isoelectric point of 9.4. Table 1 lists the amino acid composition of the protein.

From the known pathway of transcription initiation, TFIID must interact not only with DNA, but also with TFIIA and possibly with TFIIB, RNA polymerase II, and perhaps even other transcription factors (Buratowski et al., 1989; Van Dyke et al., 1988). Therefore, we expect that a large part of the TFIID protein surface will be involved in protein–protein and protein–DNA interactions. Figure 2 shows a map of the acidic and basic residues in TFIID as well as a hydropathy plot of the protein. The amino-terminal quarter of the protein is hydrophilic and contains a mixture of both positively and negatively charged residues. A stretch of 24 uncharged amino acids follows with a net hydrophobic character. This region is succeeded by a stretch of about 150 amino acids with a net basic charge. Several regions within this basic stretch have hydrophobic character. Since many DNA binding do-

Table 1. Amino Acid Composition of TF11D

	n	n (%)
Alanine	23	9.5
Cysteine	2	0.8
Aspartic acid	10	4.1
Glutamic acid	19	7.9
Phenylalanine	15	6.2
Glycine	12	5.0
Histidine	2	0.8
Isoleucine	19	7.9
Lysine	22	9.1
Leucine	16	6.6
Methionine	5	2.1
Asparagine	8	3.3
Proline	11	4.6
Glutamine	8	3.3
Arginine	15	6.2
Serine	14	5.8
Threonine	16	6.6
Valine	16	6.6
Tryptophan	1	0.4
Tyrosine	6	2.5

mains from proteins such as yeast GCN4 and rat C/EBP are highly basic (Hope and Struhl, 1986; Landschultz et al., 1989), we speculate that some of this basic region in TFIID may be involved in DNA binding. The protein contains two cysteine residues at positions 76 and 162. Upon reduction with β-mercaptoethanol or dithiothreiotol, TFIID isolated from yeast decreases in mobility by about 2 kd on SDS–PAGE, which suggests that the two cysteines can form a disulfide bridge in the native protein. This disulfide bond would link together most of the basic region of the polypeptide.

TFIID contains neither zinc finger–type structures nor obvious homology to the helix-turn-helix DNA binding motif. Comparison of the TFIID amino acid sequence with the

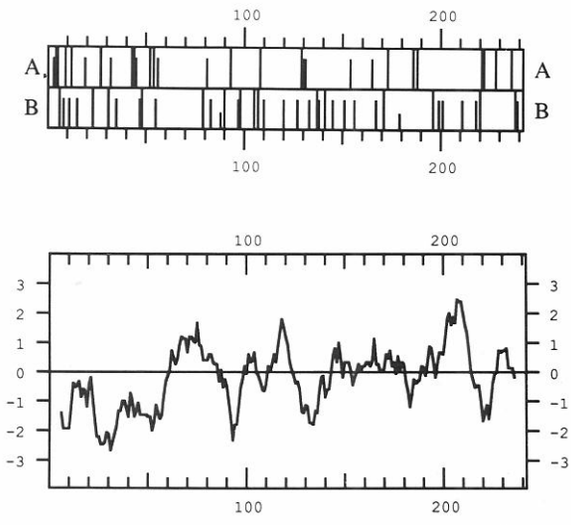

Figure 2. The TFIID Polypeptide

(Top) Plot of acidic (A) and basic (B) residues in TFIID. Charged residues are indicated by a vertical line. (Bottom) Hydropathy plot (Kyte and Doolittle, 1982) of TFIID. Hydrophobic residues are plotted above the central line.

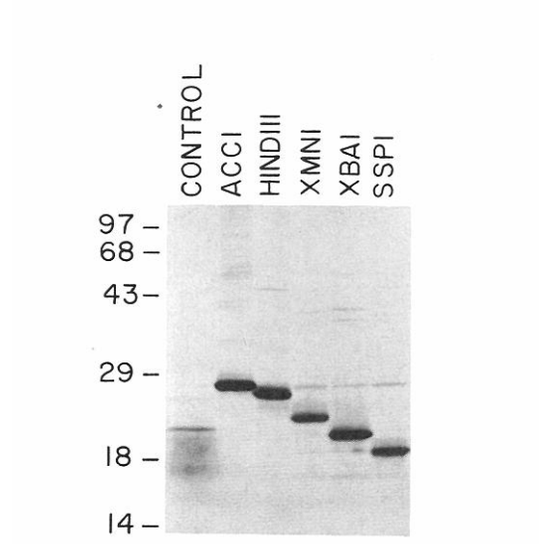

Figure 3. In Vitro Translation of TFIID

TFIID was transcribed and translated in vitro with [^{35}S]methionine as described in Experimental Procedures. Products were run on a 15% SDS–PAGE gel and visualized by autoradiography. Control translation contained no TFIID mRNA. DNA templates for in vitro transcription were cut with AccI (full-length TFIID), HindIII (14 amino acid C-terminal deletion), XmnI (50 amino acid deletion), XbaI (68 amino acid deletion), and SSP1 (81 amino acid deletion).

available protein and nucleic acid data bases translated in all six reading frames showed no obvious similarity between TFIID and other known sequences. Since TFIID and the sigma 70 subunit of bacterial RNA polymerase perform somewhat analogous functions, i.e., recognition of promoter sequences that include a TATA motif and polymerase complex formation, the amino acid sequences of these two proteins were directly compared. The region of greatest identities spanned 18 amino acids, of which 7 were identical from positions 97–114 of TFIID and positions 580–597 of sigma 70. Interestingly, this region of sigma 70 corresponds to sequences conserved among many bacterial sigma subunits. This conserved region 4 is thought to be involved in promoter recognition and may contain a helix-turn-helix DNA binding motif (see Hellmann and Chamberlin, 1988, for review). The significance of this similarity is doubtful, however, for several reasons. First, the sequence similarity between the two proteins does not extend beyond this short tract. About one-half of all proteins in the NBRF data base show as much overall similarity with TFIID as does sigma 70 when compared with the FASTP algorithm (Lipman and Pearson, 1985). Second, the amino acids conserved between the two proteins do not correspond well with the amino acids most highly conserved among sigma subunits. Finally, the conserved region 4 of sigma 70 is thought to interact with the −35 region of bacterial promoters and not the −10 TATA-like sequence.

The 240 Amino Acid Open Reading Frame Encodes TFIID Function

To test whether or not the predicted polypeptide specifies TFIID, the 240 amino acid open reading frame in the 2.4 kb EcoRI fragment was first cloned downstream of the

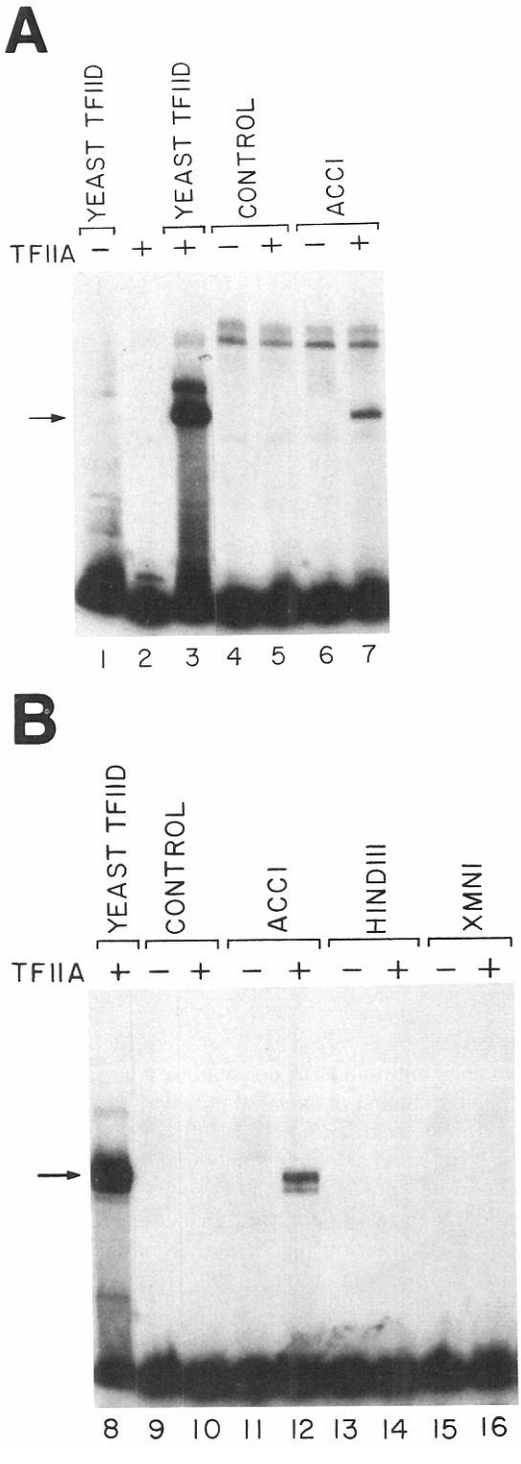

Figure 4. Gel Retardation Assay for TFIID Binding

Protein and DNA were mixed as described in Experimental Procedures and loaded to Tris–glycine acrylamide gels.

(A) The 87 bp *LEU2* TATA (TATTTAA) DNA substrate from pMS1 (see Experimental Procedures).

(B) The 125 bp adenovirus major late promoter TATA (TATAAAA) DNA substrate from pRW (Chodosh et al., 1986). Reactions contained either no additions (lane 2), yeast TFIID (lanes 1, 3, and 8), 1 μl of control in vitro translation (lanes 4, 5, 9, and 10), 1 μl of full-length in vitro–translated TFIID (AccI; lanes 6, 7, 11, and 12), or 1 μl of C-terminal deleted TFIID polypeptides (lanes 13–16). Yeast TFIIA was added where indicated.

SP6 RNA polymerase promoter. SP6 polymerase was used to synthesize mRNA from template cut with AccI, which was then translated in vitro using a rabbit reticulocyte system (Figure 3). Full-length mRNA was translated to generate a 27 kd polypeptide with the same mobility as the purified yeast TFIID. Several C-terminal deletion polypeptides were also generated in vitro by cutting the template with restriction enzymes internal to the TFIID coding sequence (Figure 3).

To test for DNA binding activity, the gel retardation assay system described by Buratowski et al. (1989) was used. In this system, purified yeast TFIID or TFIIA alone does not yield a protein–DNA complex stable to gel electrophoresis. However, addition of both TFIIA and TFIID allows formation of stable TFIIA–TFIID–DNA complexes (Figure 4A, lanes 1–3). Thus, the TFIID synthesized in vitro would be expected to form stable complexes resistant to gel electrophoresis only in the presence of TFIIA.

To test for binding, unlabeled protein synthesized in vitro was incubated with ^{32}P-labeled DNA and yeast TFIIA and loaded on a native polyacrylamide gel. The in vitro–synthesized protein yielded TFIID–TFIIA–DNA complexes with the same mobility as that obtained with TFIID purified from yeast. Figure 4 shows that the in vitro–constructed protein binds to both an 87 bp fragment containing the yeast *LEU2* TATA, TATTTAA (Hahn et al., 1989a) (Figure 4A, lanes 6 and 7) and to a 125 bp fragment containing the adenovirus major late promoter TATA, TATAAA (Figure 4B, lanes 11 and 12). The control reticulocyte lysates, which received no TFIID mRNA, did not form TFIIA-dependent complexes with DNA (lanes 4, 5, 9, and 10). A contaminating DNA binding activity unrelated to TFIID was present in the reticulocyte lysate that bound to the *LEU2* TATA fragment (lanes 4–7) but not to the major late promoter TATA fragment. The binding of both TFIID purified from yeast and in vitro–synthesized TFIID to the *LEU2* TATA fragment was competed effectively by 100 ng of p(dI-dC)·(dI-dC), which has a high affinity for TFIID (Hahn et al., 1989). All the C-terminal truncated TFIID polypeptides failed to bind in this assay (Figure 4B and data not shown). The shortest truncation tested (HindIII-cut template) lacked only 14 amino acids from the C-terminal end of TFIID.

As final proof of TFIID function, the in vitro–translated protein was added to a mammalian in vitro transcription system lacking TFIID (Buratowski et al., 1988) (Figure 5). Using the adenovirus major late promoter template, no transcription was detected in the absence of TFIID (lane 1). The addition of purified yeast TFIID restored transcription in vitro (lane 2). The addition of 1 μl of the control translation reaction, which did not recieve TFIID mRNA, produced nearly undetectable levels of transcription (lane 3). However, the addition of 1 μl of in vitro–translated TFIID produced at least a 10-fold increase in transcription over the control lysate (lane 4). From the DNA binding and transcription complementation activities of the in vitro–synthesized protein, we conclude that the 240 amino acid open reading frame encodes the yeast TFIID activity.

Identity of the TFIID Clone and *SPT15*

Because of the possibility that at least some of the genes

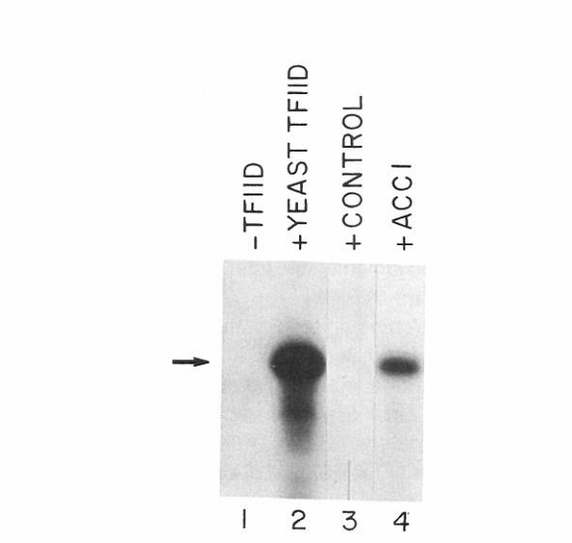

Figure 5. In Vitro–Synthesized TFIID Complements In Vitro Transcription

In vitro transcription from the adenovirus major late promoter −G cassette fusion template. All reactions contained HeLA TFIIA, TFIIB, TFIIC, and TFIIE, and calf thymus RNA polymerase. Purified yeast TFIID or 1 μl of in vitro translation reactions were added as indicated.

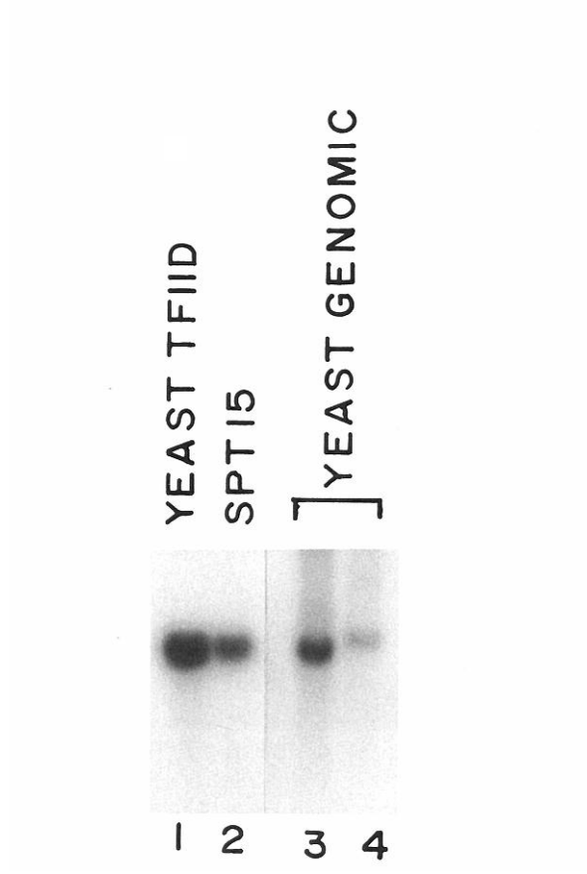

Figure 6. TFIID Is Identical to the *SPT15* Gene

Southern blot analysis of the TFIID and *SPT15* clones probed with an AccI to BamHI fragment of TFIID (positions 735–2377). Lanes 1–3 are PstI- plus BamHI-digested DNA, lane 4 is EcoRI-digested DNA. Lane 1, TFIID clone; lane 2, *SPT15* clone (pDE52-2; Eisenmann et al., 1989); and lanes 3 and 4, yeast genomic DNA. The differences in the hybridization signals are due to differing amounts of DNA loaded in each lane.

identified in the *SPT* mutant screen encode general transcription factors, and given the size of the *SPT15* functional subclone (Eisenmann et al., 1989), the restriction map of TFIID was compared with those of *SPT* clones isolated in the Winston laboratory (this work and Eisenmann et al., 1989). This comparison revealed that the TFIID and *SPT15* clones had identical restriction maps. Parallel restriction digests of the TFIID clone and pDE25-2 (a plasmid containing the *SPT15* locus; Eisenmann et al., 1989) with four different pairs of enzymes gave identical digestion products (data not shown). Further, the *SPT15* DNA hybridizes in a Southern blot when probed with TFIID at high stringency (Figure 6). From this and other evidence discussed in Eisenmann et al., we conclude that *SPT15* is identical to the TFIID gene.

Discussion

Here we report the isolation and preliminary analysis of a yeast gene encoding TFIID. This is an important step in the molecular analysis of the RNA polymerase II transcription machinery. Aside from subunits of RNA polymerase II, TFIID is the first of the general transcription factors to be characterized at the gene and sequence level. The isolation of this gene opens many avenues of investigation for the molecular, genetic, and biochemical analysis of TFIID's interaction with the rest of the transcription machinery.

The amino acid sequence of TFIID displays no obvious similarity with other known proteins. It was suggested that TFIID function may be analogous to prokaryotic sigma factors (Chen and Struhl, 1988; Ptashne, 1988). This proposal is reasonable as both TFIID and sigma are thought to position their respective polymerases at promoters, both recognize specific DNA sequences, and because of the homology between RNA polymerase II large subunit and the large subunit of bacterial polymerase. Comparison of the amino acid sequence of TFIID and sigma 70 reveals a short region of identity; however, for reasons discussed previously, this similarity is probably fortuitous. Unlike sigma factor, TFIID may not directly contact RNA polymerase II in directing initiation. After formation of the TFIID–DNA complex in mammalian systems, TFIID promotes the binding of TFIIA and TFIIB, which is essential for the subsequent binding of RNA polymerase II (Buratowski et al., 1989). Furthermore, in yeast, there is no fixed distance between the position of the TATA element and the site of transcription initiation (Hahn et al., 1985; Chen and Struhl, 1985; Nagawa and Fink, 1985; McNeil and Smith, 1986). This finding may indicate that the protein that recognizes the TATA element TFIID may not directly contact RNA polymerase. Since TFIIB has been reported to bind both DNA (Buratowski et al., 1989) and RNA polymerase (Reinberg and Roeder, 1987), it is interesting to speculate that if any of the general transcription factors is evolutionarily related to sigma factor, it would be TFIIB.

Previous work indicates that TFIID interacts both with DNA and with TFIIA, and possibly with other components of the transcription machinery. From the amino acid se-

quence of TFIID, the domain of the protein involved in DNA binding is not yet clear. Biochemical, genetic, and molecular methods are currently being used to define this domain as well as those domains involved in protein–protein interactions with other components of the transcriptional machinery. Initial experiments suggest that TFIID must have an intact C-terminal end for formation of TFIIA–TFIID–DNA complexes that are detectable by gel electrophoresis. However, it is not yet clear if the C-terminus of TFIID is required for interactions with DNA and/or with TFIIA, since TFIIA is necessary for the formation of a TFIID–DNA complex that withstands the gel-shift assay under our conditions.

The fact that the gene encoding TFIID is identical to the previously described *SPT15* gene, which alters the specificity of transcription initiation in vivo (Eisenmann et al., 1989), shows that TFIID is an important component of the transcription apparatus. Until now, TFIID has been defined solely by in vitro assays. The *SPT15* suppressor allele isolated by Eisenmann et al. apparently creates some subtle change in the specificity of TFIID interaction with either the promoter or the rest of the transcription machinery to shift the site of the initiation complex. Several models for the action of the suppressor are discussed in the accompanying paper. These models await testing by direct biochemical assays.

It is striking that the two alternative approaches of biochemistry and genetics converged on the same general transcription factor. The experiments described here and by Eisenmann et al. indicate that TFIID is critical for transcription in vivo. While the TFIID gene is essential for viability (Eisenmann et al., 1989) and is present as a single copy gene (this work and Eisenmann et al., 1989), these experiments do not rule out the existence of other TFIID-like proteins. Such proteins, if they exist, would not recognize the same spectrum of TATA elements as that recognized by TFIID. We showed previously that purified yeast TFIID protein binds with high affinity to several TATA elements that vary from the consensus sequence. Thus, it is possible, but certainly not proven, that all TATA elements in yeast function by the binding of the protein encoded by the TFIID gene.

Experimental Procedures

Proteins

TFIID was purified from S. cerevisiae through the superose 12 column as described (Buratowski et al., 1988). Starting with 440 g of yeast, about 4 μg of TFIID was obtained. TFIIA from yeast was purified on heparin and DEAE (Hahn et al., 1989b) and was approximately 50-fold purified from whole cell extracts.

Amino Acid Sequence Determination

About 2 μg of TFIID was run on a 12% SDS–PAGE gel, transferred to Immobilon membrane (Millipore), and stained with Coomassie blue as described (Matsudaira, 1987). The filter-bound TFIID was excised and used for amino acid analysis. N-terminal sequence determination and total amino acid composition analysis were performed by William Lane and David Andrews at the Harvard Microchemistry Facility.

Cloning of the TFIID Gene

A genomic DNA fragment encoding 13 N-terminal amino acid residues was amplified using PCR (Gil et al., 1988). Two degenerate oligonucleotides of sequence

```
   A   C   G   C   T
5' GATTTTATTGGCCT 3'
   T           A
               T
```

(oligonucleotide 1)

```
    A   AA  TT  A
5' GAGGAGCGGCTTAA 3'
           C   C
           A   G
```

(oligonucleotide 2)

were used in the PCR reaction. The reaction contained 0.1 μg of [γ-^{32}P]labeled oligonucleotide 1 (about 10^{10} cpm), 0.4 μg of unlabeled oligonucleotide 2, 0.5 μg of yeast genomic DNA, 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 200 μM each dATP, dGTP, dCTP, and dTTP, 0.01% gelatin, 2.5 U of AmpliTaq polymerase (United States Biochemical) in a total volume of 25 μl. After an initial denaturation at 94°C for 4 min, the reaction was carried out for 30 cycles at 94°C for 1 min, 48°C for 2 min, and 70°C for 1 min. The reaction products were precipitated with ethanol and fractionated on an 18% acrylamide sequencing gel. The expected 39 bp product (about 300 cpm) was excised an used as a template in a second PCR reaction. This amplification process was repeated a second time (final yield of 28,000 cpm per product) and the 39 bp fragment was sequenced by modified Maxam–Gilbert chemistry (Bencini et al., 1984). As the identity of several bases in the 39 bp product was ambiguous, the deduced sequence was used to design a 29 base oligonucleotide probe lacking the region of ambiguity. The ^{32}P-labeled oligonucleotide was used to probe a yeast genomic Southern blot. Hybridization was for 12 hr at 37°C in 5% SDS, 1 mM EDTA, 0.17 M sodium phosphate (pH 7), 0.6 M NaCl. Filters were washed in tetramethylammonium chloride as described (Wood et al., 1985). When washed from 55°C to 65°C, the probe hybridized to a single genomic DNA fragment. The 29 base probe was then used to screen a yeast genomic λ EMBL library (from strain S288C) containing 20–24 kb genomic DNA inserts. Filters were hybridized and washed at 61°C as described (Wood et al., 1985). Eleven positive phage were purified and found by Southern analysis to contain an identical 200 bp Sau3A fragment that hybridized to the 29 base probe. Restriction analysis showed that this segment was contained on a 2.4 kb EcoRI genomic fragment. The EcoRI fragment was subcloned into M13 for sequence analysis. The DNA sequence was determined using Sequenase enzyme (United States Biochemical) by subcloning DNA fragments to M13 and by using synthetic oligonucleotide primers. The sequence from positions 855–2132 (Figure 1) was determined on both DNA strands. Sequences were aligned and analyzed using DNA Strider software provided by C. Marck.

Sequence Analysis

Similarity between the TFIID amino acid sequence and other known sequences was determined using GenePro software. The TFIID sequence was compared both with the PIR protein data base and with the GenBank and EMBL nucleic acid data bases translated in all six reading frames. Helix-turn-helix motif homology was determined using a modified version of the method described by Dodd and Egan (1987).

In Vitro Synthesis of TFIID

TFIID expression was placed under the control of the SP6 polymerase promoter by cloning a fragment of the gene in pSP65 (Promega). First, in vitro mutagenesis was used to place a BamHI site 10 bp upstream of the TFIID translation start site. This DNA was cut with BamHI at positions 1227 and 2376 (Figure 1) and cloned into the BamHI site of pSP65 to create pSH227. For in vitro transcription of full-length TFIID RNA, this plasmid was cut with AccI (position 1970). In vitro transcription reactions contained 10 mM Tris (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 100 U of RNasin, 0.5 mM each ATP, CTP, GTP, and UTP, 2.5 μg of DNA template, 50 U of SP6 polymerase in a total volume of 100 μl and was incubated for 1 hr at 40°C.

In vitro translation was carried out in nuclease treated rabbit reticulocyte lysates (Promega). For synthesis of ^{35}S-labeled protein, reactions contained 35 μl of lysate, 40 U of RNasin, 1 mM amino acids minus methionine (Promega), 1 μg of RNA, 40 μCi of [^{35}S]methionine

(1100 Ci/mM) in a total volume of 50 μl. Reactions were incubated for 60 min at 30°C and frozen at −70°C. For synthesis of unlabeled protein, 1 mM unlabeled methionine was substituted for the labeled amino acid. These in vitro translation reactions were used directly in DNA binding and in vitro transcription reactions.

DNA Binding

DNA binding of TFIID was assayed using the gel retardation system described by Buratowski et al. (1989). TFIID was incubated for 20 min at room temperature with about 0.5 ng of labeled DNA in 60 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 2 μg of poly(dG-C)·(dG-C), 4 mM Tris (pH 8), 4% glycerol, 0.1% Brij 58, 50 μg/ml BSA in a total volume of 20 μl. Where indicated, 0.5 μg of DEAE-purified yeast TFIIA (Hahn et al., 1989b) was added. Reactions were loaded directly to Tris–glycine native acrylamide gels (Buratowski et al., 1989) and run at 4°C.

The 87 bp *LEU2* TATA fragment used for the gel retardation assay contained *LEU2* sequences from −135 to −94 (Andreadis et al., 1982) cloned into the BamHI site of pSP64 (Promega). The plasmid containing this segment, pMS1, was cut with HindIII and EcoRI in the polylinker and labeled at the HindIII site on the coding strand. For assaying binding to the adenovirus major late promoter, a 125 bp fragment from plasmid pRW (Chodosh et al., 1986) containing major late promoter sequences from −53 to +33 was used. pRW was cut with HindIII and EcoRI and labeled at the HindIII site on the coding strand.

In Vitro Transcription

In vitro transcription using fractionated HeLa cell extracts was exactly as described (Buratowski et al., 1988). For transcription of the adenovirus major late promoter, the −G cassette template pΔ-50ML(C_2AT) (Sawadogo and Roeder, 1985) was used. All reactions lacked HeLa TFIID and contained either purified yeast TFIID or reticulocyte lysate containing in vitro–synthesized unlabeled TFIID. Reactions were incubated for 45 min at 30°C. One hundred units of RNAse T1 was added and incubated for 15 min at 30°C. Samples were processed as described (Buratowski et al., 1988) and loaded on 6% urea acrylamide gels.

Acknowledgments

We thank Fred Winston and Dave Eisenmann for sharing unpublished results and for plasmid pDE25-2, and William Lane and David Andrews for amino acid sequence analysis. S. H. thanks Ed Winter, John McGrath, Mike Gallatin, and Tom St. John for advice on cloning, Ernie Tolentino for oligonucleotide synthesis, Steve Henikoff for help with homology searches, and Mike Schrag for technical assistance. This work was supported by funds from the Fred Hutchinson Cancer Research Center, grant 1RO1GM42551 from the National Institutes of Health to S. H., grant 5RO1GM30454 from NIH to L. G., grant PO1CA42063 and Cancer Center Support (Core) grant P30CA14051 from NIH, and from Cooperative Agreement CDR-8803014 from the National Science Foundation to P. A. S.

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

Received July 20, 1989; revised July 24, 1989.

References

Andreadis, H., Usu, Y.-P., Kohlaw, G. B., and Schimmel, P. (1982). Nucleotide sequence of yeast *LEU2* 5′-noncoding-region has sequences cognate for leucine. Cell *31*, 319–325.

Bencini, D. A., O'Donovan, G. A., and Wild, J. R. (1984). Rapid chemical degredation sequencing. Biotechniques *2*, 4–5.

Buratowski, S., Hahn, S., Sharp, P. A., and Guarente, L. (1988). Function of a yeast TATA element binding protein in a mammalian transcription system. Nature *334*, 37–42.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549–561.

Cavallini, B., Huet, J., Plassat, J., Sentenac, A., Egly, J., and Chambon, P. (1988). A yeast activity can substitute for the HeLa cell TATA box factor. Nature *334*, 77–80.

Chen, W., and Struhl, K. (1985). Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. EMBO J. *4*, 3273–3280.

Chen, W., and Struhl, K. (1988). Saturation mutagenesis of a yeast *HIS3* TATA element: genetic evidence for a specific TATA-binding protein. Proc. Natl. Acad. Sci. USA *85*, 2691–2695.

Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986). A single polypeptide possesses the binding and transcription activities of the adenovirus major late promoter factor. Mol. Cell. Biol. *6*, 4723–4733.

Davison, B. L., Egly, J., Mulvihill, E. R., and Chambon, P. (1983). Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature *301*, 680–686.

Dodd, I. B., and Egan, J. B. (1987). Systematic method for the detection of potential λ cro-like DNA-binding regions in proteins. J. Mol. Biol. *194*, 557–564.

Eisenmann, D. M., Dollard, C., and Winston, F. (1989). *SPT15*, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. Cell *58*, this issue.

Errede, B., Cardillo, T. S., Sherman, F., Dubois, E., Deschamps, J., and Wiame, J.-M. (1980). Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. Cell *22*, 427–436.

Fassler, J. S., and Winston, F. (1988). Isolation and analysis of a novel class of suppressor of Ty insertion mutations in *S. cerevisiae*. Genetics *118*, 203–212.

Fire, A., Samuels, M., and Sharp, P. A. (1984). Interactions between RNA polymerase II, factors and template leading to accurate transcription. J. Biol. Chem. *259*, 2509–2516.

Gil, G., Smith J. R., Goldstein, J. L., Slaughter, C. A., Orth, K., Brown, M. S., and Osborne, T. F. (1988). Multiple genes encode nuclear factor 1–like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proc. Natl. Acad. Sci. USA *85*, 8963–8967.

Hahn, S., Hoar, E., and Guarente, L. (1985). Each of three TATA elements specifies a subset of the transcription initiation sites at the CYC1 promoter of *S. cerevisiae*. Proc. Natl. Acad. Sci. USA *82*, 8562–8566.

Hahn, S., Buratowski, S., Sharp, P. A., and Guarente, L. (1989a). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA *86*, 5718–5722.

Hahn, S., Buratowski, S., Sharp, P. A., and Guarante, L. (1989b). Identification of a yeast protein homologous in function to the mammalian general transcription factor TFIIA. EMBO J., in press.

Helmann, J. D., and Chamberlin, M. J. (1988). Structure and function of bacterial sigma factors. Annu. Rev. Biochem. *57*, 839–872.

Hope, I. A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*, 885–894.

Kakidani, H., and Ptashne, M. (1988). GAL4 activates expression in mammalian cells. Cell *52*, 161–167.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681–1688.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. Science *227*, 1435–1441.

Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. *262*, 10035–10038.

McNeil, B. J., and Smith, M. (1986). Transcription initiation of the *S. cerevisiae* iso-1-cytochrome c gene. Multiple independent TATA sequences. J. Mol. Biol. *187*, 363–378.

Metzger, D., White, J., and Chambon, P. (1988). The human estrogen receptor functions in yeast. Nature *334*, 31–36.

Nagawa, F., and Fink, G. R. (1985). The relationship between the TATA sequence and transcription initiation sites at the *HIS4* gene of *S. cerevisiae*. Proc. Natl. Acad. Sci. USA *82*, 8557–8561.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.

Reinberg, D., and Roeder, R. G. (1987). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. J. Biol. Chem. *262*, 3310–3321.

Roeder, S., Farabaugh, P., Chaleff, D., and Fink, G. R. (1980). The origins of gene instability in yeast. Science *209*, 1375–1380.

Sawadogo, M., and Roeder, R. G. (1985). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantatative in vitro assay. Proc. Natl. Acad. Sci. USA *82*, 4394–4389.

Schena, M., and Yamamoto, K. (1988). Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. Science *241*, 965–968.

Struhl, K. (1988). The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. Nature *332*, 649–651.

Van Dyke, M. W., Roeder, R. G., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science *241*, 1335–1338.

Webster, N., Jin, J. R., Green, S., Hollis, M., and Chambon, P. (1988). The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 *trans*-activator. Cell *52*, 169–178.

Winston, F., Chaleff, D., Valent, B., and Fink, G. (1984). Mutations affecting Ty-mediated expression of the *HIS4* gene in *S. cerevisiae*. Genetics *107*, 179–197.

Winston, F., Dollard, C., Malone, E., Clare, J., Kapakos, P., Farabough, P., and Minehart, P. (1987). Three genes are required for *trans* activation of Ty transcription in yeast. Genetics *115*, 649–656.

Wood, W. I., Gitschier, J., Lasky, L., and Lawn, R. M. (1985). Base composition-independent hybridization-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA *82*, 1585–1588.

GenBank Accession Number

The accession number for the sequence reported in this paper is M27135.