MOLECULAR AND CELLULAR BIOLOGY, Aug. 1990, p. 3859–3867
0270-7306/90/083859-09$02.00/0

Vol. 10, No. 8

Yeast and Human TATA-Binding Proteins Have Nearly Identical DNA Sequence Requirements for Transcription In Vitro

C. RICHARD WOBBE AND KEVIN STRUHL*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received 8 March 1990/Accepted 25 April 1990

We have analyzed the DNA sequence requirements for TATA element function by assaying the transcriptional activities of 25 promoters, including those representing each of the 18 single-point mutants of the consensus sequence TATAAA, in a reconstituted in vitro system that depends on the TATA element-binding factor TFIID. Interestingly, yeast TFIID and HeLa cell TFIID were virtually identical in terms of their relative activities on this set of promoters. Of the mutated elements, only two had undetectable activity; the rest had activities ranging from 2 to 75% of the activity of the consensus element, which was the most active. In addition, mutations of the nucleotide following the TATAAA core strongly influenced transcriptional activity, although with somewhat different effects on yeast and HeLa TFIID. The activities of all these promoters depended upon TFIID, and the level of TFIID-dependent transcription in vitro correlated strongly with their activities in yeast cells. This suggests that the in vivo activities of these elements reflect their ability to functionally interact with a single TATA-binding factor. However, some elements with similar activities in vitro supported very different levels of transcriptional activation by GAL4 protein in vivo. These results extend the degree of evolutionary conservation between yeast and mammalian TFIID and are useful for predicting the level of TATA element function from the primary sequence.

Most eucaryotic RNA polymerase II promoters contain a TATA element, consensus sequence TATAAA, that is required for transcription in vivo and in vitro. The DNA sequence requirements for TATA element function have been most extensively analyzed by comparing native promoter sequences (1, 2), by saturation mutagenesis of a specific TATA element (7, 16), or by random selection of functional TATA elements (36a). However, unlike the situation with upstream activation sequence or enhancer elements, it has not been possible until recently to obtain a TATA element-binding protein pure enough for detailed in vitro analyses. As a result, ambiguity has arisen over the number of distinct types of TATA elements and the number of TATA element-binding proteins. From a variety of observations in yeast and mammalian cells, it has been suggested that there may be multiple TATA element-binding proteins, distinguishable not only by their specific sequence requirements but also by their response to transcriptional activator proteins (13, 16, 21, 36, 38). In addition, other experiments (19, 24) indicate that there may be factors that have a TATA element as part of their recognition sequence but that do not activate and may repress transcription by RNA polymerase II. Clearly, resolution of these issues will require more detailed knowledge of the sequence requirements for TATA element function and the number and properties of the protein(s) interacting with these sequences.

Biochemical fractionation of transcriptionally active extracts from a variety of eucaryotic sources has shown that multiple factors in addition to RNA polymerase II are required to reconstitute accurate basal transcription from a number of genes in vitro (25, 32). One of these, TFIID, binds the TATA element, and this binding is required for its activity (10, 27, 28). Owing to the apparent instability of this activity during purification, it has not been possible to obtain preparations of mammalian TFIID of adequate purity for extensive biochemical characterization. However, the yeast *Saccharomyces cerevisiae* contains a 26-kilodalton (kDa) protein that can substitute for mammalian TFIID in DNA binding and in transcription reactions reconstituted with the other HeLa factors (4, 6). With this reconstitution assay, yeast TFIID has been purified to near homogeneity, and the gene encoding this protein has been cloned (5, 11, 14, 22, 34). The availability of pure yeast TFIID should facilitate detailed biochemical studies of its role in transcriptional initiation.

Saturation mutagenesis of the yeast *his3* T_R TATA element indicated that 17 out of the 18 possible mutations of the sequence TATAAA are defective for GAL4-activated transcription in vivo (7). Surprisingly, however, three of these mutant TATA elements function moderately well in combination with GCN4 activator protein, and a double mutant (TATTTA) functioned nearly as well as the wild-type element in combination with GAL4 (16). These results could be explained either by invoking multiple TATA factors or by proposing that the relative activity of the TATA element depends on other factors besides the primary sequence and its affinity for a single TATA-binding factor.

In this study, we have used this reconstituted in vitro transcription system to assay the activity of wild-type and a large number of mutant TATA sequences that have been characterized for their ability to function in vivo (7, 16). Our results indicate that HeLa and yeast TFIID factors have very similar sequence specificities and that the level of TFIID-dependent transcription in vitro correlates closely with promoter activity in yeast cells. However, some elements with similar activities in vitro behave very differently in terms of their ability to activate transcription in combination with GAL4 protein. The evolutionary and mechanistic implications of these results are discussed.

* Corresponding author.

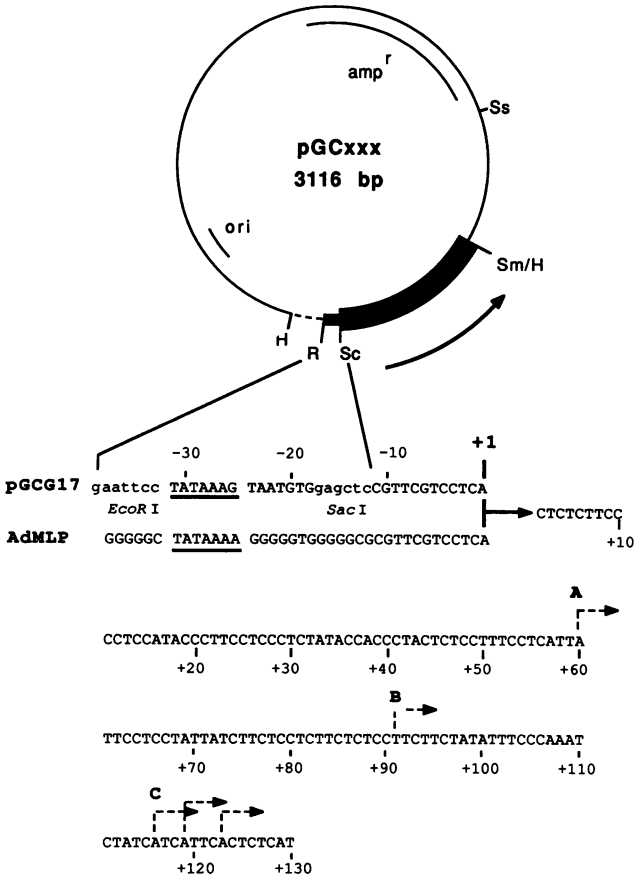

FIG. 1. Structure of transcription assay substrate plasmid. (Top) Structure of plasmid indicating relative positions of the G-less cassette (thick black line, with arrow below indicating direction of transcription), the *his3* TATA oligonucleotide (medium black line), the pUC19 polylinker with a deleted *Sac*I site (dotted thin line), and pUC19 sequences (thin solid line) containing the origin and ampicillin resistance gene. Restriction site abbreviations: H, *Hin*dIII; R, *Eco*RI; Sc, *Sac*I; Ss, *Ssp*I; Sm/H, junction of the *Sma*I and blunted *Hin*dIII sites of pUC18 created during the course of construction. (Bottom) Sequence of the promoter region, from nucleotide −37, and the first 130 nucleotides of the coding strand of constructs containing either the wild-type adenovirus major late promoter (AdMLP) or the wild-type *his3* TATA oligonucleotide (pGCG17). *Eco*RI and *Sac*I sites are indicated in lowercase letters, and the 7-nucleotide TATA sequence is underlined. Sequences are identical downstream from nucleotide −11. The +1 initiation nucleotide is indicated by the bold arrow. Internally initiated transcripts A, B, and C (see Fig. 2) are labeled in boldface, and the initiating nucleotides are indicated by dotted arrows.

MATERIALS AND METHODS

Construction of transcription assay templates with mutated TATA sequences. The transcription assay substrate used was a derivative of plasmid pML(C_2AT)$_{19}$Δ-50 (33) (Fig. 1). The G-less cassette was excised from pML(C_2AT)$_{19}$Δ-50 as a *Bst*UI fragment and blunt-end ligated to a double-stranded oligonucleotide with the sequence indicated in Fig. 1. The ligation product was digested with *Eco*RI and *Sma*I, isolated by agarose gel electrophoresis, and ligated into *Eco*RI-*Sma*I-digested pUC18. This plasmid was digested with *Sma*I and *Hin*dIII, and the *Hin*dIII end was blunted with Klenow DNA polymerase. Following religation, the resultant molecule was cleaved with *Eco*RI and *Ssp*I, and the G-less cassette-containing fragment was ligated into *Eco*RI-*Ssp*I-digested pUC19 in which the normal *Sac*I site had been destroyed by using T4 DNA polymerase. To insert mutant TATA oligonucleotides, this plasmid was digested with *Eco*RI and *Sac*I, gel purified, and ligated to the *Eco*RI-*Sac*I TATA oligonucleotides from plasmids described previously (7), generating the pGC plasmids; the last three digits in the pGC plasmid designations refer to the TATA allele numbers described elsewhere (7, 16). For a clone lacking the TATA sequence, the G-less cassette was excised from pML(C_2AT)$_{19}$Δ-50 with *Bst*UI and *Sma*I and ligated into *Sma*I-digested pUC18.

Preparation of HeLa transcription factors. Nuclear extracts were prepared from HeLa cells as described previously (35), except that the extract was dialyzed against buffer C (20 mM Tris chloride [pH 7.9, 4°C], (0.1 mM EDTA, 20% glycerol, 2 mM dithiothreitol [DTT], 0.2 mM phenylmethylsulfonyl fluoride) containing 100 mM KCl. The crude extract was chromatographed on phosphocellulose as described by Reinberg and Roeder (31). The flowthrough (containing TFIIA), 0.5 M KCl (containing TFIIB, TFIIE, and TFIIF), and 1.0 M KCl (containing TFIID) fractions were further purified by DEAE Sephacel chromatography as described elsewhere (30, 31). The TFIIA DEAE fraction was concentrated by precipitation with 60% saturated ammonium sulfate and dialyzed against buffer C containing 0.1 M KCl. The TFIIB DEAE flowthrough fraction was applied to a 4.0-ml phosphocellulose column equilibrated with buffer C plus 0.1 M KCl and eluted with buffer C plus 0.5 M KCl. Protein concentrations of the fractions were as follows: TFIIA, 11.5 mg/ml; TFIIB, 0.88 mg/ml; TFIID, 1.15 mg/ml; and TFIIE/F, 0.85 mg/ml.

Purification of yeast TFIID. *S.cerevisiae* TFIID was purified by a modification of a procedure described previously (4). BJ926 cells (19 liters) were grown in YPD medium to an OD_{600} of 5.0 with vigorous aeration. Cells (240 g) were harvested and washed once with 600 ml of ice-cold H_2O, suspended with 240 ml of Z buffer (50 mM Tris chloride [pH 7.9, 4°C], 10 mM $MgCl_2$, 1 M sorbitol) containing 30 mM DTT and incubated at 23°C for 15 min. The cells were repelleted with and resuspended in 750 ml of Z buffer containing 1 mM DTT and 75,000 U of Zymolyase. Following incubation at 30°C for 40 min with gentle agitation, the spheroplasts were pelleted at 2,000 × *g* for 5 min and washed three times with ice-cold Z buffer containing 1 mM DTT. The final pellet was washed once with 250 ml of disruption buffer (50 mM Tris chloride [pH 7.9, 4°C], 10 mM $MgSO_4$, 1 mM EDTA, 10 mM potassium acetate, 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride, 0.4 μg of pepstatin A per ml) and suspended in 250 ml of the same buffer. The spheroplasts were disrupted with 15 strokes of a tight-fitting pestle in a Dounce homogenizer. An equal volume of disruption buffer containing 20% glycerol and 0.8 M $(NH_4)_2SO_4$ was added, and the lysate was gently rocked for 30 min at 4°C, followed by centrifugation at 100,000 × *g* for 90 min at 4°C. Proteins in the supernatant were precipitated with ammonium sulfate, and TFIID activity was purified by chromatography over heparin-agarose, DEAE-cellulose, and Mono S resins as described previously (4). The Mono S fraction was concentrated by precipitation with ammonium sulfate at 60% saturation, and the pellet was suspended with 0.2 ml of buffer C containing 0.25 M KCl. This was applied directly to a Sephacryl S-200 HR fast protein liquid chromatography column (1 by 30 cm) and eluted at 0.1 ml/min with buffer C plus 0.25 M KCl. Fractions (0.25 ml) were collected and assayed for TFIID activity. The peak of TFIID activity (10

μg/ml; purified approximately 1,000-fold from the heparin-agarose fraction) eluted at a position expected for a globular protein of approximately 26 kDa and was of comparable purity to the Superose 12 fraction of Hahn et al. (15) when examined by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis.

In vitro transcription assays. Transcription assays (20 μl) were carried out as described previously (33) except that 25 μM [α-^{32}P]UTP was used as the label. Reaction mixes were incubated at 30°C for 60 min and contained 2.3 μg of TFIIA, 0.18 μg of TFIIB, 1.7 μg of TFIIE/F, 5 U of calf thymus RNA polymerase II (purified as described before [20]), and the amounts of HeLa or yeast TFIID indicated. The amount of HeLa or yeast TFIID added was saturating under these reaction conditions. Reactions were terminated by addition of 200 μl of stop mixture containing 20 mM Tris chloride (pH 7.9), 1 mM EDTA, 200 mM NaCl, 100 μg of yeast tRNA per ml, 0.2% SDS, and 100 μg of proteinase K per ml, incubated for 30 min at 37°C, extracted with phenol, and precipitated with ethanol. The products were electrophoretically separated in a 5% denaturing polyacrylamide gel and visualized by autoradiography. Nucleotide incorporation into specific products was quantitated with a Betascope 603 blot analyzer (Betagen), and the results shown are the average of duplicate analyses.

RESULTS

Replacement of the adenovirus major late TATA with a *his3* TATA oligonucleotide. With the scheme outlined in Fig. 1, the TATA element of the adenovirus major late promoter in the plasmid pML(C_2AT)$_{19}$Δ-50 (33) was replaced with an *Eco*RI-*Sac*I oligonucleotide containing the TATA site and some surrounding sequences from the yeast *his3* gene used previously for in vivo studies (7). The two principal advantages of this plasmid are that it allows direct comparison of the activity of different TATA sequences in an otherwise identical sequence background and that the G-less cassette simplifies the transcriptional assay.

In transcription reaction mixes containing partially purified factors from HeLa cells, the promoter containing the "wild-type" *his3* TATA element retained transcriptional activity, albeit at a level 30 to 40% of that of the TATA element in the adenovirus major late promoter (Fig. 2A and B). Using an in vitro complementation assay (4, 6), we also purified a TFIID activity from *S. cerevisiae* that functions in transcription assays containing either the adenovirus or *his3* TATA sequence. In all cases, transcription was dependent upon a TATA sequence and upon TFIID, and it originated principally from the correct +1 initiation site of the adenovirus promoter. In these assays and in those described below, saturating amounts of HeLa or yeast TFIID were used, and the levels of transcription with the two factors were comparable.

In addition to the major transcription product originating from the +1 site, several other faster mobility bands were routinely observed, three of which were designated A, B, and C (Fig. 2B). The appearance of these products was also dependent on TFIID and inhibited by α-amanitin, but was independent of the presence of a TATA sequence. Primer extension analysis (not shown) confirmed that these minor transcripts were initiated at sites within the G-less cassette that were located about 25 to 30 base pairs (bp) downstream from sequences that weakly resembled a consensus TATA element (see Fig. 1). Interestingly, these internal starts responded differently to HeLa and yeast TFIID, possibly

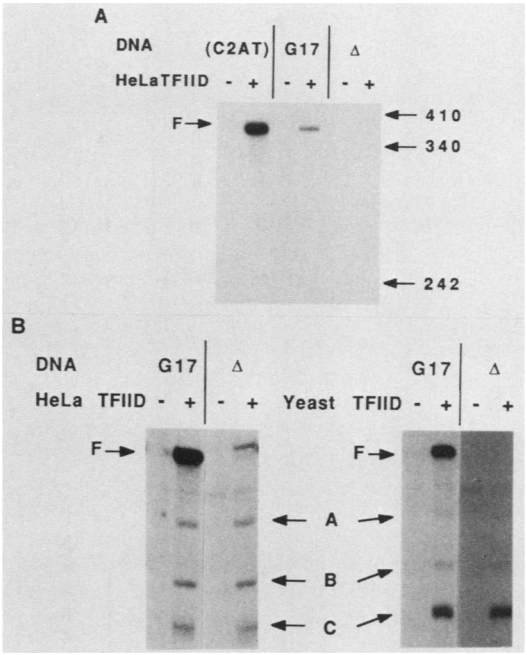

FIG. 2. Analysis of transcription from promoters containing the adenovirus major late or yeast *his3* TATA sequence. Reactions were carried out as described in Materials and Methods, and mixes were supplemented with 2.3 μg of HeLa TFIID or 30 ng of yeast TFIID (S-200 fraction) where indicated. (A) Transcription from substrates containing the adenovirus TATA (C_2AT), the *his3* TATA (G17), or no TATA (Δ) in the absence or presence of HeLa TFIID. Numbers at right indicate the length (in nucleotides) of molecular size markers. (B) Transcription from substrates containing the *his3* TATA or no TATA with HeLa TFIID (left) or yeast TFIID (right). A, B, and C designate internally initiated transcripts discussed in the text. A small amount of a product of readthrough from upstream initiations, which is six nucleotides larger than the +1 transcript, is seen in the lane from the reaction containing HeLa TFIID and no TATA. Arrow F, Migration expected for full-length, +1-initiated transcription product.

reflecting slight differences in the sequence preferences of these two factors (see below). Because these internally initiated transcripts occurred equally in the presence and absence of the normal TATA element, they were used to normalize transcriptional levels from the +1 site of the TATA mutants described below.

Transcriptional activity of promoters with TATA elements mutated at positions 1 to 6. Figure 3 shows the results of transcription assays of derivatives in which the original TATA oligonucleotide was replaced with mutants of the sequence TATAAA in the absence of TFIID or in the presence of yeast or HeLa TFIID. A wide range of transcriptional activities were observed, but regardless of the TATA sequence, TFIID-dependent transcription always initiated at the normal +1 site. Strikingly, the amount of transcript observed in reactions containing HeLa or yeast TFIID was virtually identical for any given mutant; this will be discussed in detail below. The same relationship was observed even at subsaturating levels of HeLa or yeast TFIID and over a range of template and salt concentrations (data not shown).

The results of quantitation of the radioactivity present in the full-length transcript are shown in Table 1, and several general observations are worthy of note. As suggested by Fig. 3, the transcriptional activities of any given mutant with

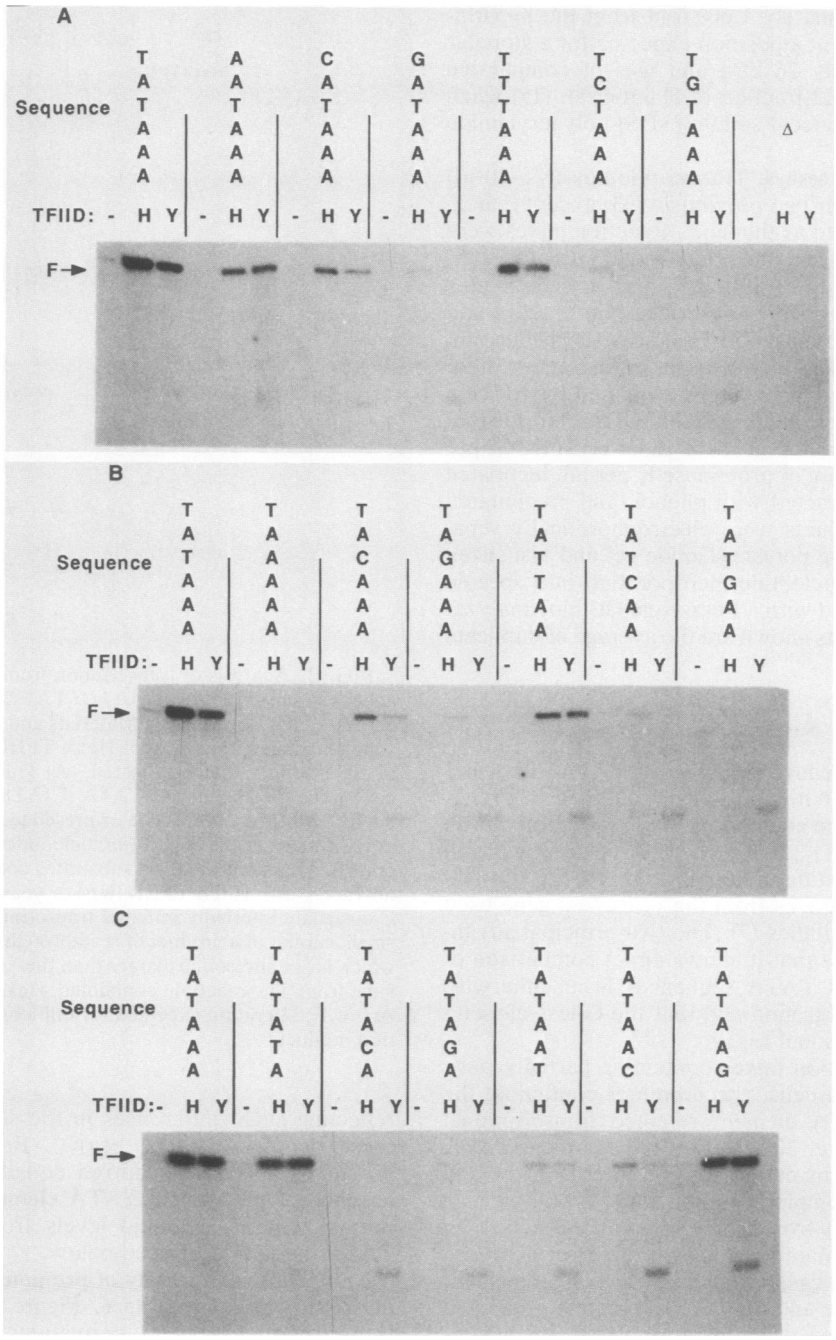

FIG. 3. Analysis of transcription from mutations in positions 1 to 6 of the *his3* TATA sequence in the absence (—) of TFIID or in the presence of HeLa (H) or yeast (Y) TFIID. (A) Positions 1 and 2. (B) Positions 3 and 4. (C) Positions 5 and 6. Reactions were carried out as described in the legend to Fig. 2 and Materials and Methods. Sequences are indicated above corresponding lanes. Δ, TATA deletion.

HeLa and yeast TFIID were indistinguishable. The sequence TATAAA, generally considered to be the TATA consensus, yielded the highest transcriptional activity. In comparison, any mutation of this sequence reduced activity, but only two (TAAAAA and TATGAA) reduced transcription to undetectable levels. In general, mutations in the first and sixth positions were the least detrimental (any mutant retained at least 10% of wild-type activity), while mutations in positions 3 and 4 were the most detrimental. Finally, maintenance of A+T content did not appear to be a major determinant of activity. For example, the most active mutations at positions 3 and 6 were T to C and A to G changes, respectively, and a T to C change at position 1 retained 30% of wild-type activity.

Also indicated in Table 1 are the results of previous analyses of these TATA sequences in vivo in a variety of promoter contexts (7, 16). In general, there was a direct relationship between the relative activity of a given TATA sequence in vivo and in vitro with purified TFIID. With only two exceptions, AATAAA and TATAAT, sequences >15% as active in vitro as TATAAA scored as having detectable basal or GCN4-induced expression, with derivatives having

TABLE 1. Relative transcription of TATA mutants

Wild-type sequence		Mutant nucleotide	Relative in vitro activity[a]		In vivo activity[b]		
Positions(s)	Nucleotide(s)		HeLa TFIID	Yeast TFIID	Basal[c]	GCN4[c]	GAL4[c,d]
1	T	A	30	33	−	−	−
		C	26	30	(+)	+/−	−
		G	8	5	−	−	−
		T	100	100	(++)	++	++
2	A	A	100	100	(++)	++	++
		C	7	4	−	−	−
		G	3	1	−	−	−
		T	54	62	(+)	+	−
3	T	A	<1	<1	−	−	−
		C	16	20	(+)	+/−	−
		G	2	2	−	−	−
		T	100	100	(++)	++	++
4	A	A	100	100	(++)	++	++
		C	2	1	−	−	−
		G	<1	<1	−	−	−
		T	19	25	(+)	+/−	−
5	A	A	100	100	(++)	++	++
		C	5	3	−	−	−
		G	5	5	−	−	−
		T	72	77	(++)	++	+
6	A	A	100	100	(++)	++	++
		C	8	5	−	−	−
		G	54	58	(+)	+	−
		T	21	10	−	−	−
7	G[e]	A	172	100	ND	ND	+++
		C	38	35	ND	+	−
		G	72	77	ND	ND	+
		T	170	107	ND	ND	+++
4 and 5	AA	AT	72	77	(++)	++	+
		CT	4	5	ND	ND	−
		GT	3	3	ND	ND	−
		TT	39	41	ND	ND	++

[a] Relative to transcription of TATAAA template (0.2 to 0.4 pmol of nucleotide full-length product in 60 min), which was set at 100.
[b] Symbols defined in reference 16.
[c] Data from Harbury and Struhl (16).
[d] Data from Chen and Struhl (7).
[e] I.e., TATATAG.

30 to 40% activity in vitro being more active in vivo. On the other hand, only sequences with activity greater than ~70% of that of the wild type in vitro displayed GAL4-activated expression in vivo, with one exception, discussed below.

Transcriptional activity of TATA elements mutated at position 7. Previous observations indicated that the first nucleotide downstream of the sequence TATAAA, a G in the *his3* promoter, had a marked influence on transcriptional activity (16). This effect was also observed when the first six nucleotides were TATATA. Mutants at this position were also tested in the in vitro system, and the results are shown in Fig. 4 and Table 1. In agreement with the in vivo experiments, substitution with an A or T increased activity, whereas substitution with a C reduced activity. However, the response of the HeLa and yeast TFIID-containing reaction mixes to these changes differed: while a G to C change reduced activity by 70% in both cases, G to A or G to T mutations increased activity almost 2.5-fold with HeLa TFIID, but only 1.3-fold with yeast TFIID. Thus, while the identity of the seventh nucleotide appears to be nearly as important as the identity of the first six nucleotides, it has more influence with HeLa TFIID than with yeast TFIID, presumably reflecting differences in protein-DNA contacts made by the two factors.

Transcriptional activity of double mutations at positions 4 and 5. Although almost all single mutations of TATAAA significantly reduce transcription in vivo (7) and in vitro (Table 1), the double mutant TATTTA displayed nearly wild-type activity when tested for its ability to support transcriptional activation by GAL4 protein in vivo (16). However, the related double mutants TATCTA and TAT GTA as well as all other single mutations of TATATA were

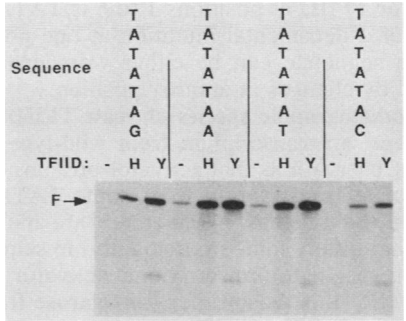

FIG. 4. Analysis of transcription from mutations at position 7 in the absence of TFIID (—) or in the presence of HeLa (H) or yeast (Y) TFIID. Sequences are indicated above corresponding lanes. See Fig. 2 legend for explanation of arrow F.

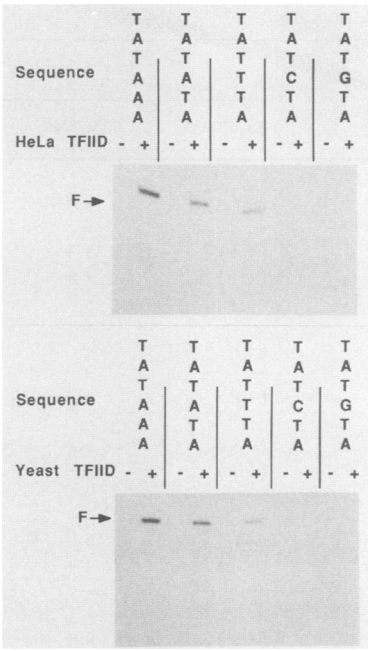

FIG. 5. Analysis of transcription from double mutations at positions 4 and 5 in the absence of TFIID (—) or in the presence of HeLa (top) or yeast (bottom) TFIID. Sequences are indicated above lanes. See Fig. 2 legend for explanation of arrow F.

functionally defective (16). In vitro analysis of these TAT NTA double mutants indicated that any mutation of the A at position 4 was detrimental, ranging from a 2-fold reduction by an A to T change to a nearly 30-fold reduction accompanying an A to G mutation (Fig. 5, Table 1). The observation that TATTTA was less efficient than TATATA in TFIID-dependent transcription in vitro contrasts with its apparently higher activity during GAL4-dependent activation in vivo. The relatively lower activity of the sequence TATTTA in vitro was also seen in transcription reactions with a crude yeast extract (29).

A more subtle effect was noted when the activities of elements with the sequence TATNTA were compared with those of elements with the sequence TATNAA, where N is any nucleotide. When the fourth nucleotide was a C, G, or T, changing the fifth nucleotide from an A to a T increased activity twofold, but when the fourth nucleotide was an A, this change at nucleotide 5 reduced transcription about 30%. These observations suggest that (i) the optimal recognition sequence for TFIID at positions 1 to 6 is TATAAA and (ii) the effect of a detrimental mutation at one position in the recognition sequence can be either worsened or partially suppressed by changes at another position.

One chromatographic species of yeast TFIID is necessary and sufficient for transcription from wild-type and mutant templates. Observations that activator proteins can differentially stimulate transcription from some TATA sequences have led to the suggestion that there are multiple TFIID activities that differ, sometimes slightly, in sequence recognition and response to transcriptional activator proteins (16, 21, 36, 38). For this reason, heparin-agarose fractions from the purification of yeast TFIID were assayed with three templates (TATTTA, TATAAG, and CATAAA) whose properties in vivo were most suggestive of multiple TFIID activities (16). As shown in Fig. 6, the activity required for transcription from each of the mutants precisely and entirely copurified with the activity required for transcription from the wild-type sequence, which itself behaved like a single species. In addition, the same activity reconstituted transcription of the internally initiated transcripts A, B, and C, whose putative TATA sequences deviate significantly from the consensus.

DISCUSSION

To analyze the DNA sequence requirements for TFIID-dependent transcription in vitro, we have constructed a relatively simple assay substrate consisting of two fundamental promoter elements: the initiator of the adenovirus major late promoter (sequences from −11 to +10) fused to an easily assayable transcription unit (33) and a TATA element on a synthetic oligonucleotide which can easily be deleted or replaced with mutants. These substrates were assayed for transcriptional activity in vitro in a system consisting of the partially purified HeLa transcription factors TFIIA, TFIIB, TFIIE/F, and polymerase II supplemented with either partially purified HeLa TFIID or highly purified yeast TFIID. The major conclusion from this work are that (i) HeLa and yeast TFIID are extremely similar in terms of their DNA sequence preferences, (ii) the level of TFIID-dependent transcription in vitro correlates strongly, but not absolutely, with the level of transcription in yeast cells, and (iii) a single TFIID can account for the transcriptional activity of a variety of TATA sequences.

DNA sequence requirements for human and yeast TFIID are similar. The fact that yeast TFIID can functionally replace mammalian TFIID for accurate transcription in vitro indicates that these proteins are evolutionarily conserved (4, 6). The striking similarity in DNA sequence requirements significantly extends the degree of evolutionary conservation between yeast and mammalian TFIID. When the relative activity of each mutant with HeLa TFIID was plotted against the activity of that mutant with yeast TFIID, almost all the points fell very close to a line with a slope of 1, which indicates identical specificity (Fig. 7). Thus, the details of a crucial protein-DNA interaction that determines both the level of transcription and choice of initiation site of most (if not all) genes have been maintained between two phylogenetically distant species. As argued previously (39), such strong conservation might reflect difficulties in changing the ancestral code of promoter recognition that existed in primitive organisms that contained multiple transcription units.

In this regard, it may be significant that eucaryotic TATA elements are similar in sequence to the consensus −10 element of a large number of bacterial promoters (17). If similarity in the target DNA sequences reflects similarity in the protein sequences, then a key DNA-binding domain may have been preserved from bacterial RNA polymerase holoenzymes to yeast and mammalian TFIIDs. Viewed in this light, eucaryotic TFIID might be thought of as an easily dissociable subunit of RNA polymerase II, much as σ^{70} is a dissociable subunit of *Escherichia coli* RNA polymerase that confers sequence specificity to the majority of bacterial promoters (18).

HeLa and yeast TFIID can be distinguished by the internally initiated transcripts and the activities of two mutations at position 7. This almost certainly reflects differences between the proteins, either within the DNA-binding domains or in regions found in one protein that are not present in the other. One obvious difference is that the apparent molecular weight of the crude HeLa TFIID is about 120,000 (30),

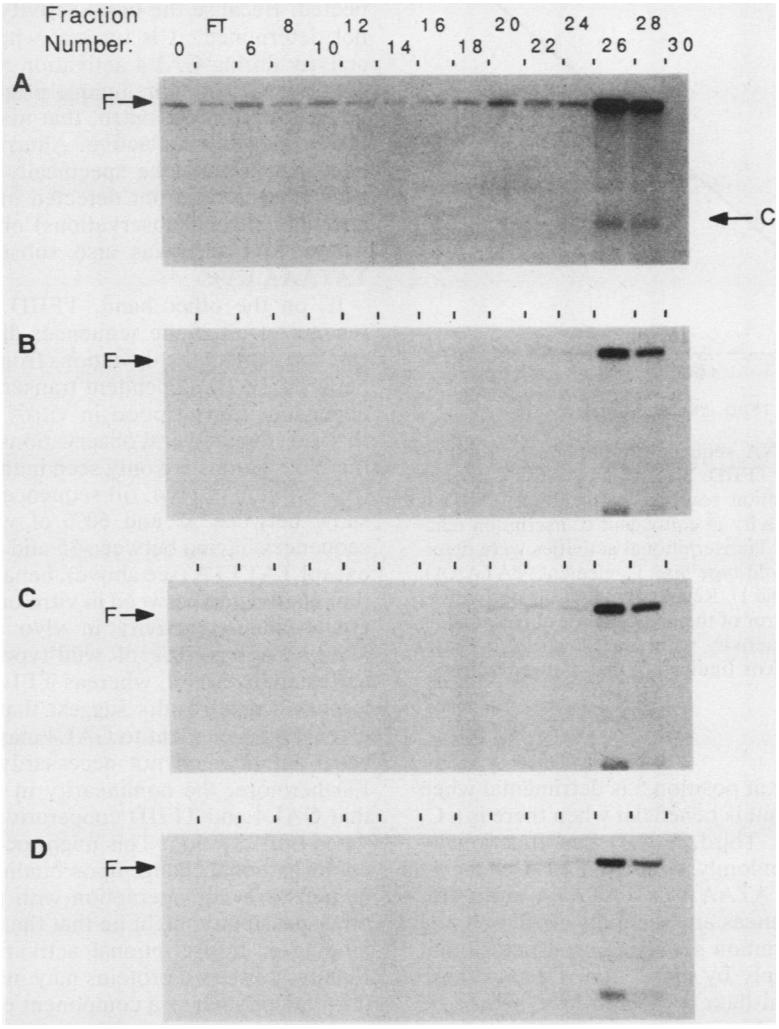

FIG. 6. Assay of heparin-agarose chromatography fractions for TFIID activity on substrates containing (A) *his3* wild-type TATA, (B) TATAAG, (C) TATTTA, or (D) CATAAA. As described in Materials and Methods, yeast crude extract (950 mg of protein) was prepared and applied to a heparin-agarose column (2.5 by 10.2 cm), which was eluted with a 250-ml linear gradient of 0.04 to 0.6 M KCl. Fractions of 8.5 ml were collected, and 3 μl of those indicated (top) were assayed in reactions containing HeLa TFIIA, TFIIB, TFIIE/F, and polymerase II as described in the text. Arrow C, Internally initiated transcript C. Lane 0, No fraction added to reaction mix. FT, Flowthrough. See Fig. 2 legend for explanation of arrow F.

whereas it is 26,000 for highly purified yeast TFIID (3). In perspective, however, the few cases of moderate divergence in sequence utilization are overshadowed by the many similarities, suggesting that the DNA-binding domains of these two proteins are very similar.

Nature of the TFIID interaction site. Our previous analysis of the sequence requirements of a yeast TATA element were carried out in vivo on intact promoters and hence were complicated by the possibility of multiple proteins having distinct specificities and by differential effects of upstream activator proteins (7, 16). The experiments here directly address the inherent specificity of a TFIID interaction site because they were done with highly purified yeast TFIID as well as simple promoters in which the TATA elements are necessary and sufficient for transcription. Our results indicate that TATAAA is a highly active TFIID recognition sequence and that any single point mutation of this sequence reduces activity. Although the pattern of nucleotide substitutions is not simple, mutations at positions 1 and 6 are generally tolerated better than mutations at positions 3 and 4, maintenance of the overall A+T content does not seem to be particularly important, and transversions to G or C residues appear to be most deleterious. These observations are consistent with and expand upon mutational analyses (9, 12, 23, 26) and promoter sequence comparisons of genes from a wide variety of eucaryotic organisms (1, 2).

For several reasons, the consensus for a transcriptionally active TFIID-binding site must be regarded as rather incomplete. First, it is clear that so-called flanking sequences are important. Position 7, the first nucleotide downstream of TATAAA, has a marked influence on TFIID activity in vitro, suggesting that it should be considered part of the element. Moreover, the previous in vivo analysis suggested that positions 8 and 9 might play a role and did not address the importance of positions upstream of TATAAA. Thus, depending on the flanking sequences, some TATAAA elements might be poor substrates, whereas some single-mutant elements might be good substrates. Second, within the TATAAA sequence, the effect of a given nucleotide change can be influenced by the identity of the other residues. For

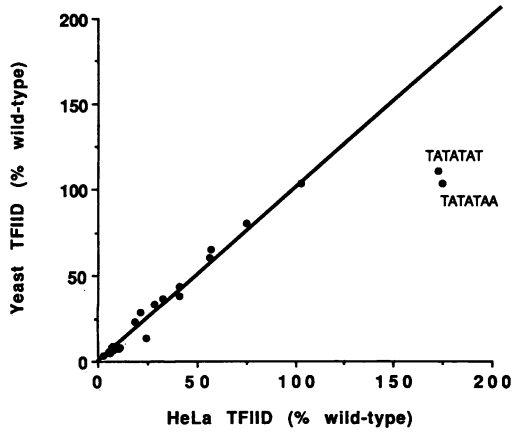

FIG. 7. Comparison of DNA sequence requirements for transcription by yeast and HeLa TFIID. The relative activity of each TATA derivative in transcription reactions with HeLa TFIID is plotted against its relative activity in equivalent transcription reactions containing yeast TFIID. Transcriptional activities were determined relative to that of the wild-type *his3* T_R element (TATAAA), which was defined as 100 (Table 1). Results for all of the derivatives are within the experimental error of the diagonal line corresponding to equivalent transcriptional activity with yeast and HeLa TFIID except for two TATA sequences (indicated) that showed a preference for the HeLa factor.

example, an A to T change at position 5 is detrimental when there is an A at position 4 but is beneficial when there is a C, G, or T at that position. Third, TFIID can functionally interact with natural or randomly selected TATA elements whose sequences lack a TATAAA or TATATA motif (15, 36a). These different sequences are generally A+T rich and may derive from some common sequence or structure that cannot be ascertained simply by inspection. Despite these uncertainties, the empirical data presented here should be useful for predicting the relative activity of a TFIID site from its primary sequence.

Implications for the mechanism of transcriptional initiation in vivo. The results presented here indicate that the activity of a given TATA element in supporting basal or GCN4-activated transcription in vivo is directly related to its TFIID-dependent activity in vitro on a simple promoter. This strongly suggests that, for the sequences examined, one TATA-binding factor, TFIID, is sufficient to account for activity in vivo. Furthermore, all the sequences proposed as candidates for binding alternative factors are moderately good substrates for TFIID-dependent transcription in vitro. Our inability to detect other TFIID-like activities that act through some of these sequences is also consistent with a single protein factor. Thus, the contribution of the TATA element to the level of transcription from many intact promoters can be generally explained by the inherent activity of TFIID on the TATA sequence.

Despite the general validity of the above conclusion, there are some interesting deviations from the correlation between TFIID-dependent transcription in vitro and TATA-dependent transcription in vivo. The clearest example is that the activities of the sequences TATTTA, TATAAG, TTTAAA, and TATATAC are similar in vitro (about 35 to 60% as active as TATAAA) but are drastically different in vivo when coupled to the enhancer recognized by GAL4 protein (Table 1) (16). In this context, TATTTA is nearly as efficient as TATAAA and hence appears to be more active than expected. Because the basal activity of TATTTA in vivo was not determined, it is unclear whether its surprisingly high activity during GAL4 activation represents an effect of the *gal* enhancer or some unique property of TATTTA in vivo, not reconstituted in vitro, that makes promoters containing this sequence more active. Alternatively, a distinct TFIID-like protein could be specifically active on TATTTA, but such a factor was not detected in our fractionation (Fig. 6 and unpublished observations) or in yeast crude extracts, where TATTTA was also substantially less active than TATAAA (29).

If, on the other hand, TFIID is the only activity that recognizes all of the sequences discussed above, how then can one explain the deviations from the excellent correlation between TATA-dependent transcription in vivo and TFIID-dependent transcription in vitro? In attempting to answer this question, several observations should be considered: (i) these deviations are only seen in the case of GAL4-activated transcription in vivo, (ii) sequences displaying this behavior show between 35 and 60% of wild-type activity, (iii) all sequences having between 35 and 60% of wild-type activity, except TATTTA (see above), behave in this manner, and (iv) the relationship between in vitro or in vivo basal activity and GAL4-induced activity in vivo appears to be nonlinear (TATATA has 70% of wild-type activity in vitro and is activated by GAL4, whereas TTTAAA has 60% activity but is not). These results suggest that the responsiveness of a given TATA element to GAL4 may be related to its inherent basal activity and not necessarily to its sequence per se. Furthermore, the nonlinearity in this relationship suggests that GAL4 and TFIID cooperatively interact to direct activated transcription. This might occur if TFIID underwent a conformational change upon binding DNA and some conformations prevent interaction with particular activators. Another possibility might be that the activator alters the DNA-binding or transcriptional activation properties of TFIID. Finally, these two proteins may interact allosterically with a third factor such as a component of the basic transcriptional machinery. These models implicate a functional, but not necessarily direct, interaction between upstream activator proteins and TFIID, and our results suggest that the properties of this interaction may differ between GAL4 and GCN4.

The observations here do not exclude the possibility of factors other than TFIID that bind to promoter-proximal elements and stimulate transcription. A number of promoters lack a discernible TATA box, and while some may still have binding sites for TFIID (15, 36a), others probably do not (8, 36a, 37). A curious case in point is the human U2 small nuclear RNA gene, a TATA-less gene transcribed by RNA polymerase II, in which conversion of a critical element to TATATAT inhibits RNA polymerase II-dependent transcription (24). Moreover, a fragment containing this sequence in the U6 gene is necessary for transcription by RNA polymerase III, and mutations of this sequence stimulate transcription by RNA polymerase II. Clarification of these issues should be facilitated by the availability of the cloned gene for yeast TFIID (14, 22, 34) as well as TFIID mutants with altered transcriptional properties (11).

ACKNOWLEDGMENTS

We thank Ulla Hansen for supplying some of the HeLa cells, Rebecca Sundseth for a valuable collaboration in preparing the HeLa transcription factors, and Rebecca Sundseth, Vicki Singer, and Fred Ponticelli for fruitful discussions.

This work was supported by a Merck postdoctoral fellowship to

C.R.W. and by a Public Health Service grant from the National Institutes of Health (GM 30186) and a grant from the Lucille Markey trust to K.S.

LITERATURE CITED

1. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
2. **Bucher, P., and E. N. Trifonov.** 1986. Compilation and analysis of eukaryotic Pol II promoter sequences. Nucleic Acids Res. **14**:10009–10026.
3. **Buratowski, S., S. Hahn, L. Guarente, and P. A. Sharp.** 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell **56**:549–561.
4. **Buratowski, S., S. Hahn, P. A. Sharp, and L. Guarente.** 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature (London) **334**:37–42.
5. **Cavallini, B., I. Faus, H. Matthes, J. M. Chipoulet, B. Winsor, J.-M. Egly, and P. Chambon.** 1989. Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. Proc. Natl. Acad. Sci. USA **89**: 9803–9807.
6. **Cavallini, B., J. Huet, J. L. Plassat, A. Sentenac, J.-M. Egly, and P. Chambon.** 1988. A yeast activity can substitute for the HeLa TATA box factor. Nature (London) **334**:77–80.
7. **Chen, W., and K. Struhl.** 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.
8. **Chen, W., and K. Struhl.** 1989. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. EMBO J. **8**:261–268.
9. **Concino, M. F., R. F. Lee, J. P. Merryweather, and R. Weinmann.** 1984. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription *in vitro*. Nucleic Acids Res. **12**:7423–7433.
10. **Davison, B. L., J.-M. Egly, E. R. Mulvihill, and P. Chambon.** 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature (London) **301**:680–686.
11. **Eisenmann, D. M., C. Dollard, and F. Winston.** 1989. *SPT15*, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. Cell **58**: 1183–1191.
12. **Giangrande, A., C. Mettling, M. Martin, C. Ruiz, and G. Richards.** 1989. *sgs3* TATA: effects of point mutations on expression *in vivo* and protein binding *in vitro* with staged nuclear extracts. EMBO J. **8**:3459–3466.
13. **Gill, G., and M. Ptashne.** 1988. Negative effect of the transcriptional activator GAL4. Nature (London) **334**:721–723.
14. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the *SPT15* suppressor of Ty element insertions. Cell **58**:1173–1181.
15. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA **86**:5718–5722.
16. **Harbury, P. A. B., and K. Struhl.** 1989. Functional distinctions between yeast TATA elements. Mol. Cell. Biol. **9**:5298–5304.
17. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
18. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57**: 839–872.
19. **Hobson, G. M., M. T. Mitchell, G. R. Molloy, M. L. Pearson, and P. A. Benfield.** 1988. Identification of a novel TA-rich DNA-binding protein that recognizes a TATA sequence within the brain creatine kinase promoter. Nucleic Acids Res. **16**: 8925–8944.
20. **Hodo, H. G., and S. P. Blatti.** 1977. Purification using polyethylenimine precipitation and low molecular weight subunit analysis of calf thymus and wheat germ DNA-dependent RNA polymerase II. Biochemistry **16**:2334–2343.
21. **Homa, F. L., J. C. Glorioso, and M. Levine.** 1988. A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. Genes Dev. **2**:40–53.
22. **Horikoshi, M., C. K. Wang, H. Fujii, J. A. Cromlish, P. A. Weil, and R. G. Roeder.** 1989. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature (London) **341**:299–303.
23. **Huang, D. H., M. Horikoshi, and R. G. Roeder.** 1988. Activation of the adenovirus EIIa late promoter by a single point mutation which enhances binding of transcription factor IID. J. Biol. Chem. **263**:12596–12601.
24. **Lobo, S. M., and N. Hernandez.** 1989. A 7-bp mutation converts a human RNA polymerase II RNA promoter into a RNA polymerase III promoter. Cell **58**:55–67.
25. **Matsui, T., J. Segall, P. A. Weil, and R. G. Roeder.** 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255**:11992–11996.
26. **Myers, R. M., K. Tilly, and T. Maniatis.** 1986. Fine-structure genetic analysis of a β-globin promoter. Science **232**:613–618.
27. **Nakajima, N., M. Horikoshi, and R. G. Roeder.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol. Cell. Biol. **8**:4028–4040.
28. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region specific DNA-binding activity. Cell **36**:357–369.
29. **Ponticelli, A. S., and K. Struhl.** 1990. Analysis of yeast *his3* transcription in vitro: biochemical support for multiple mechanisms of transcription. Mol. Cell. Biol. **10**:2832–2839.
30. **Reinberg, D., M. Horikoshi, and R. G. Roeder.** 1987. Factors involved in specific transcription by mammalian RNA polymerase II: functional analysis of initiation factors IIA and IID, and identification of a new factor operating at sequences downstream of the initiation site. J. Biol. Chem. **262**:3322–3330.
31. **Reinberg, D., and R. G. Roeder.** 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.
32. **Samuels, M., A. Fire, and P. A. Sharp.** 1982. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. **257**:14419–14427.
33. **Sawadogo, M., and R. G. Roeder.** 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc. Natl. Acad. Sci. USA **82**:4394–4398.
34. **Schmidt, M. C., C. C. Kao, R. Pei, and A. J. Berk.** 1989. Yeast TATA-box transcription factor gene. Proc. Natl. Acad. Sci. USA **86**:7785–7789.
35. **Shapiro, D. J., P. A. Sharp, W. W. Wahli, and M. J. Keller.** 1988. A high-efficiency HeLa cell nuclear transcription extract. DNA **7**:47–55.
36. **Simon, M. C., T. M. Fisch, B. J. Benecke, J. R. Nevins, and N. Heintz.** 1988. Definition of multiple, functionally distinct TATA elements, one of which is a target in the *hsp70* promoter for E1A regulation. Cell **52**:723–729.

36a. **Singer, V. L., C. R. Wobbe, and K. Struhl.** 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev. **4**:636–645.

37. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57**:103–113.
38. **Struhl, K.** 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. Mol. Cel. Biol. **6**:3847–3853.
39. **Struhl, K.** 1986. Yeast *HIS3* expression in *Escherichia coli* depends upon fortuitous homology between eukaryotic and prokaryotic promoter elements. J. Mol. Biol. **191**:221–229.