DNA binding site selection by RNA polymerase II TAF$_{II}$250–TAF$_{II}$150 complex recognizes the Initiator

Gillian E.Chalkley and C.Peter Verrijzer[1]

Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

Present address: MGC Department of Molecular and Cell Biology and Centre for Biomedical Genetics, Leiden University Medical Centre, PO Box 9503, 2300 RA Leiden, The Netherlands

[1]Corresponding author
e-mail: verrijzer@mail.medfac.leidenuniv.nl

Basal transcription factor TFIID comprises the TATA-box-binding protein, TBP, and associated factors, the TAF$_{II}$s. Previous studies have implicated TAF$_{II}$250 and TAF$_{II}$150 in core promoter selectivity of RNA polymerase II. Here, we have used a random DNA binding site selection procedure to identify target sequences for these TAFs. Individually, neither TAF$_{II}$250 nor TAF$_{II}$150 singles out a clearly constrained DNA sequence. However, a TAF$_{II}$250–TAF$_{II}$150 complex selects sequences that match the Initiator (Inr) consensus. When in a trimeric complex with TBP, these TAFs select Inr sequences at the appropriate distance from the TATA-box. Point mutations that inhibit binding of the TAF$_{II}$250–TAF$_{II}$150 complex also impair Inr function in reconstituted basal transcription reactions, underscoring the functional relevance of Inr recognition by TAFs. Surprisingly, the precise DNA sequence at the start site of transcription influences transcriptional regulation by the upstream activator Sp1. Finally, we found that TAF$_{II}$150 specifically binds to four-way junction DNA, suggesting that promoter binding by TFIID may involve recognition of DNA structure as well as primary sequence. Taken together, our results establish that TAF$_{II}$250 and TAF$_{II}$150 bind the Inr directly and that Inr recognition can determine the responsiveness of a promoter to an activator
Keywords: core promoter/initiator/TAFs/TFIID/transcription

Introduction

The transcription control regions of eukaryotic structural genes can be classified into two categories: (i) a core promoter, comprising the transcription start site and flanking sequences that interact with the general transcription machinery; and (ii) binding sites for gene-specific regulators that can be localized proximally or distally to the start site of transcription. The core, or basal, promoter nucleates the assembly of a pre-initiation complex (PIC), containing RNA polymerase II (RNA pol II) and the general transcription factors (GTFs), TFIIA, B, D, E, F and H (reviewed in Orphanides *et al.*, 1996; Roeder, 1996; Hampsey, 1998).

PIC assembly is initiated by binding of TFIID to the core promoter followed by recruitment of the basal machinery, either in a stepwise manner or as a pre-assembly of RNA pol II and GTFs (Orphanides *et al.*, 1996; Roeder, 1996; Hampsey, 1998). TFIID is an evolutionarily conserved multiprotein complex comprising the TATA-box-binding protein, TBP, and a set of tightly associated factors, the RNA pol II TAFs (Burley and Roeder, 1996; Verrijzer and Tjian, 1996; Hoffmann *et al.*, 1997; Hahn, 1998; Lee and Young, 1998). TBP not only functions in RNA pol II transcription, but is essential for transcription by all three eukaryotic RNA polymerases (I, II and III) (Hernandez, 1993; Lee and Young, 1998). Genes transcribed by the different RNA polymerases are each characterized by a unique core promoter structure. Class-specific promoter recognition is achieved by the association of TBP with at least four different sets of TAFs, each dedicated to a distinct class of genes (Hernandez, 1993; Lee and Young, 1998).

What makes an RNA pol II core promoter? Although core promoters are far from uniform, a number of general motifs have been recognized (Smale, 1997). The TATA-box (consensus: TATAAA) is located ~25–30 nucleotides upstream of the transcription start site of many genes and can direct accurate initiation of transcription (Smale, 1997). The initiator element (Inr) is a sequence that encompasses the start site of transcription and can direct initiation of transcription in the absence of a TATA-box (Smale and Baltimore, 1989). Functional studies suggested PyPyAN(T/A)PyPy as the optimal initiator sequence (Smale, 1997), whereas sequence comparisons of *Drosophila* genes indicated a TCA(T/G)TPyPy consensus (Arkhipova, 1995). A third motif, the downstream promoter element [DPE; consensus: PuG(A/T)CGTG], was identified in *Drosophila* TATA-less promoters and is located ~30 nucleotides downstream of the transcription start site (Burke and Kadonaga, 1996, 1997). Recently, a G-rich element adjacent to the TATA-box and bound by TFIIB has been identified (Lagrange *et al.*, 1998; Qureshi and Jackson, 1998). It is pertinent to note that natural promoters frequently contain divergent core elements or lack one or more of them altogether. Moreover, sequences other than the motifs discussed above can also modulate basal promoter strength.

What proteins mediate core promoter function? Several factors have been identified that bind to core promoters (reviewed in Smale, 1997). RNA pol II itself recognizes features of the Inr which might assist the correct positioning of the polymerase on the promoter (Carcamo *et al.*, 1991; Weis and Reinberg, 1997). Core promoter structure can modulate the differential requirements for TFIIE, TFIIF and TFIIH *in vitro* and several GTFs have been shown to contact promoter DNA (reviewed in Orphanides *et al.*, 1996; Roeder, 1996; Hampsey, 1998; Robert *et al.*,

1998). Although the contacts between the GTFs and the basal promoter appear to be largely sequence-independent, specificity might result from accumulated weak preferences of individual factors within the PIC. Finally, a number of transcriptional regulators such as TFII-I, E2F, YY1 and USF can stimulate transcription, not only via binding to enhancer elements, but also through interaction with binding sites that coincide with core promoter sequences of specific genes (reviewed in Smale, 1997). Although the proteins listed above are likely to play a role in the selection of at least a number of promoters, among the various promoter binding factors, TFIID has emerged as the prime general core promoter recognizing factor.

TFIID is a sequence-specific DNA-binding GTF, which is involved in transcription of all, or almost all, structural genes (Hahn, 1998). Several of the TFIID subunits have been implicated in core promoter selectivity (Verrijzer and Tjian, 1996; Hampsey and Reinberg, 1997; Smale, 1997; Hahn, 1998). In addition to TATA-box binding by TBP, accumulated evidence suggests that the TAFs play a key role in the functioning of the other core promoter elements (Verrijzer and Tjian, 1996; Hoffmann et al., 1997; Hahn, 1998). Early DNase I footprinting studies already noted that, on certain promoters, the TFIID footprint is extended compared with TBP alone, suggesting that one or more of the TAFs contact DNA (Sawadogo and Roeder, 1985; Nakatani et al., 1990; Zhou et al., 1992; Emanuel and Gilmour, 1993; Wang and Van Dyke, 1993; Kaufmann and Smale, 1994; Purnell et al., 1994; Verrijzer et al., 1994). Indeed, using recombinant TAFs, it has been shown that some TAFs bind promoter DNA directly and can mediate core promoter specificity in reconstituted basal transcription reactions (Verrijzer et al., 1994, 1995). Functional studies with highly purified TFIID also showed that the TAFs contribute to basal activities of non-TATA core elements in the context of TATA-less as well as TATA-containing promoters (Kaufmann and Smale, 1994; Martinez et al., 1994; Verrijzer et al., 1994, 1995; Burke and Kadonaga, 1996, 1997). It should be noted, however, that TAF-independent TATA-less transcription has also been described (Aso et al., 1994; Weis and Reinberg, 1997). Finally, TFIID containing a TBP subunit defective in DNA binding can no longer function on TATA-only promoters, but still supports transcription from Inr-containing promoters (Martinez et al., 1995). Thus, specific TBP–DNA contacts might be dispensable for Inr-mediated transcription and instead, TAFs may target TFIID to TATA-less promoters.

Which TAFs are responsible for recognition of basal promoter elements? In vitro transcription and DNA binding experiments using recombinant partial TBP–TAF complexes, revealed that together, TAF$_{II}$250 and TAF$_{II}$150 can mediate core promoter discrimination (Verrijzer et al., 1994, 1995). DNA cross-linking experiments using highly purified TFIID also revealed that these two TAFs are in intimate contact with the core promoter DNA (Sypes and Gilmour, 1994; Verrijzer et al., 1994, 1995). Furthermore, within purified TFIID, TAF$_{II}$60 can be specifically cross-linked to the DPE (Burke and Kadonaga, 1997). The TAFs however, act not only by promoting TFIID–promoter binding. Depending on the core promoter sequence, TAFs, in particular TAF$_{II}$250, can also inhibit TFIID–promoter

interactions (Kokubo et al., 1993; Verrijzer et al., 1995; Burley and Roeder, 1998; Liu et al., 1998). Transcription reactions using TFIID depleted for TAF$_{II}$150 revealed that this TAF is required for Inr function (Hansen and Tjian, 1995; Kaufmann et al., 1996, 1998). Additionally, TFIIA, which shares many characteristics with the TAFs, has also been implicated in core promoter discrimination (Hansen and Tjian, 1995; Emami et al., 1997; Martinez et al., 1998). Finally, studies in mammalian and yeast cells demonstrated that TAF$_{II}$250 or its yeast homologue TAF$_{II}$145, function as core promoter selectivity factors in vivo (Shen and Green, 1997; Wang and Tjian, 1997).

Here, we used an unbiased DNA binding site selection procedure to investigate potential sequence determinants for promoter recognition by TAF$_{II}$250 and TAF$_{II}$150. The functional relevance of TAF–DNA interactions was assessed in DNA binding and reconstituted transcription assays. The role of core promoter sequence in determining the responsiveness to a transcriptional activator was also addressed. Finally, we have tested the ability of TAF$_{II}$250 and TAF$_{II}$150 to recognize structured DNA. Our results provide evidence for direct recognition of the Inr by a dimeric TAF$_{II}$250–TAF$_{II}$150 complex that contributes to core promoter selectivity of TFIID.

Results

Sequence requirements for DNA binding by TAFs

The two largest TFIID subunits, TAF$_{II}$250 and TAF$_{II}$150, can form a stable complex with TBP and with each other (Verrijzer et al., 1994). As discussed above, these two TAFs are involved in DNA binding and core promoter selectivity of RNA pol II. Previously, a random DNA binding site selection procedure demonstrated that a Drosophila TFIID fraction preferentially binds the Inr sequence (Purnell et al., 1994). In order to identify the preferred binding sequences for TAF$_{II}$250 and TAF$_{II}$150 we performed binding site selection experiments using a pool of oligonucleotides that were random at 34 positions flanked by primer and cloning sequences. The length of the random DNA sequence was motivated by previous footprinting and cross-linking experiments which indicated that TAF$_{II}$250 and TAF$_{II}$150 may contact promoter DNA over >30 nucleotides (Sypes and Gilmour, 1994; Verrijzer et al., 1994, 1995; Oelgeschläger et al., 1996). Recombinant human TAF$_{II}$250 is expressed at higher levels and is less susceptible to proteolytic degradation than Drosophila TAF$_{II}$250. Since in a previous study we did not detect any functional differences between human and Drosophila TAF$_{II}$250 (Chen et al., 1994), we decided to use the human protein in our experiments. Recombinant TAFs were immunopurified from extracts prepared from Sf9 cells infected with baculoviruses expressing HA-tagged human TAF$_{II}$250 or Flag-tagged Drosophila TAF$_{II}$150 (Figure 1A).
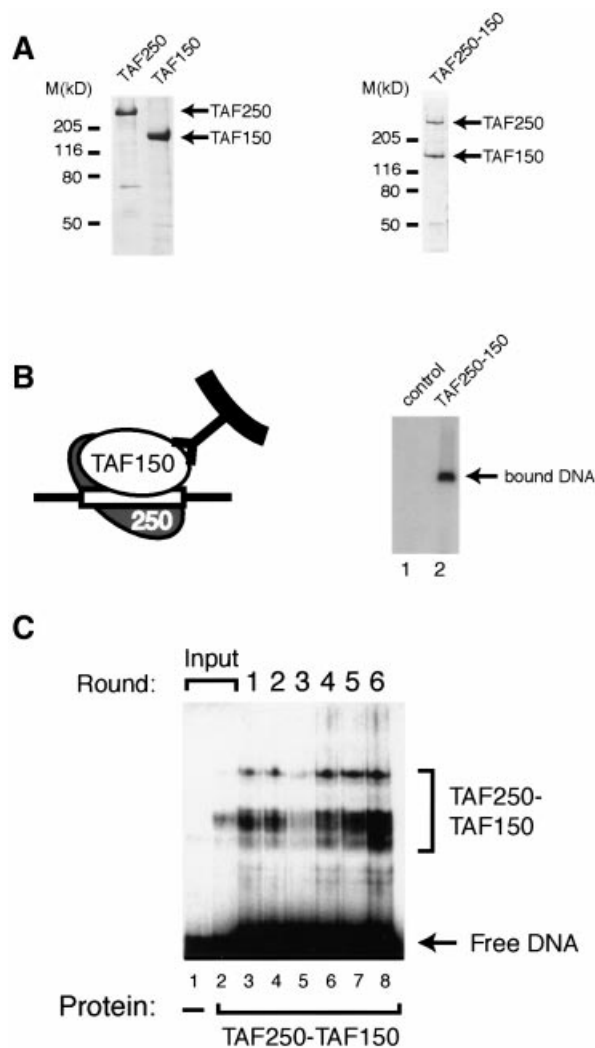
Either the individual TAFs or an in vitro assembled dimeric TAF$_{II}$150–TAF$_{II}$250 complex was immobilized on immunoaffinity beads and incubated with the random oligonucleotide pool in the presence of an excess amount of non-specific competitor DNA. Next, proteins were immunoprecipitated and extensively washed prior to the isolation of any bound DNA (Figure 1B). After PCR amplification the recovered DNA was used in a subsequent

round of selection. These steps were repeated for a total of six rounds of DNA binding selection. When either $TAF_{II}150$ or $TAF_{II}250$ was used alone, subsequent rounds of selection did not yield significant increases in the proportion of DNA bound. In contrast, when the dimeric $TAF_{II}250$–$TAF_{II}150$ complex was used, we observed a strong enrichment in the proportion of DNA bound in sequential rounds of selections (data not shown).

The selection process with the $TAF_{II}250$–$TAF_{II}150$ complex was monitored directly by electrophoretic mobility shift assays (EMSAs). The DNA recovered after each round of binding was labelled at comparable specific activities and tested for binding to the $TAF_{II}250$–$TAF_{II}150$ complex (Figure 1C). The band-shift assay revealed a

clear enrichment in bound DNA after subsequent rounds of selection, indicating that the $TAF_{II}250$–$TAF_{II}150$ complex had selected preferred DNA binding sequences from the randomized pool.

To identify potential sequence motifs selected by the TAFs, oligonucleotides bound in the final round of selection were recovered, subcloned and their sequences determined. All the clones analysed contained unique DNA sequences, demonstrating that they originated from independently selected oligonucleotides. Of the 62 oligonucleotide sequences selected by the $TAF_{II}250$–$TAF_{II}150$ complex, 56 could be aligned. After tabulation, a consensus sequence of Y(C/t)AN(T/a)YY was derived that is closely related to the Inr consensus sequence (Figure 1D; Y, pyrimidine, capital letters indicate the preferred base, lower case letters indicate the next preferred base). In addition to the tabulated sequences (those that most closely fit the Inr consensus), the selected oligonucleotides were enriched for partial Inr sequences. No clear consensus sequence could be found after sequencing of the DNA pools obtained after six rounds of binding by either $TAF_{II}250$ or $TAF_{II}150$ alone (data not shown). Furthermore, within these sequences there was no enrichment for the Inr motif. These results suggest that a $TAF_{II}250$–$TAF_{II}150$ dimer, but not the isolated TAFs, preferentially binds the Inr element.

Binding site selection by a TBP–$TAF_{II}250$–$TAF_{II}150$ complex

We next asked if the TAFs would also select an Inr consensus sequence if their position on the DNA is restricted by association with TBP. For these experiments, we assembled a TBP–$TAF_{II}250$–$TAF_{II}150$ complex *in vitro* and synthesized a pool of oligonucleotides that are random at 20 positions flanked by a spacer sequence, an optimal TATA-box, primer and cloning sequences (Figure 2A). We reasoned that the binding of TBP to the TATA-box

Fig. 1. Selection of DNA sequences from a pool of quasi-random oligonucleotides with a $TAF_{II}250$–$TAF_{II}150$ complex.
(**A**) Recombinant FLAG-tagged $TAF_{II}150$ or HA-tagged $TAF_{II}250$ were expressed in Sf9 cells and, after extract preparation, immunopurified using an anti-Flag or anti-HA column, respectively. The purified TAFs were either used individually or assembled into a dimeric $TAF_{II}250$–$TAF_{II}150$ complex. The purified TAFs and $TAF_{II}250$–$TAF_{II}150$ complex were analysed by SDS–PAGE followed by Coomassie Blue staining. The positions of the TAFs and molecular weights (kDa) of protein standards are indicated. (**B**) The purified $TAF_{II}250$–$TAF_{II}150$ complex was tested in DNA binding assays using a pool of radiolabelled oligonucleotides that were random at 34 positions (open box), flanked by primer and cloning sequences (solid bar, see Materials and methods for details). Binding reactions were performed with control beads (lane 1) or an immobilized $TAF_{II}250$–$TAF_{II}150$ complex (lane 2). Bound DNA was isolated and analysed by PAGE. (**C**) Enrichment of $TAF_{II}250$–$TAF_{II}150$ binding sequences during subsequent rounds of selection (rounds 1–6). DNA recovered after each round of binding was labelled at comparable specific activities and tested for binding by the $TAF_{II}250$–$TAF_{II}150$ complex. After incubation in the absence (lane 1) or presence (lanes 2–8) of $TAF_{II}250$–$TAF_{II}150$ complex, the resulting protein–complexes were analysed by EMSA. The positions of free DNA and the $TAF_{II}250$–$TAF_{II}150$–DNA complex are indicated. (**D**) The oligonucleotide sequences recovered after six rounds of selection were aligned and the frequencies of each base, as well as those of purines (R) and pyrimidines (Y), at each position are indicated. This quantification was used to obtain the consensus sequence shown below the table.
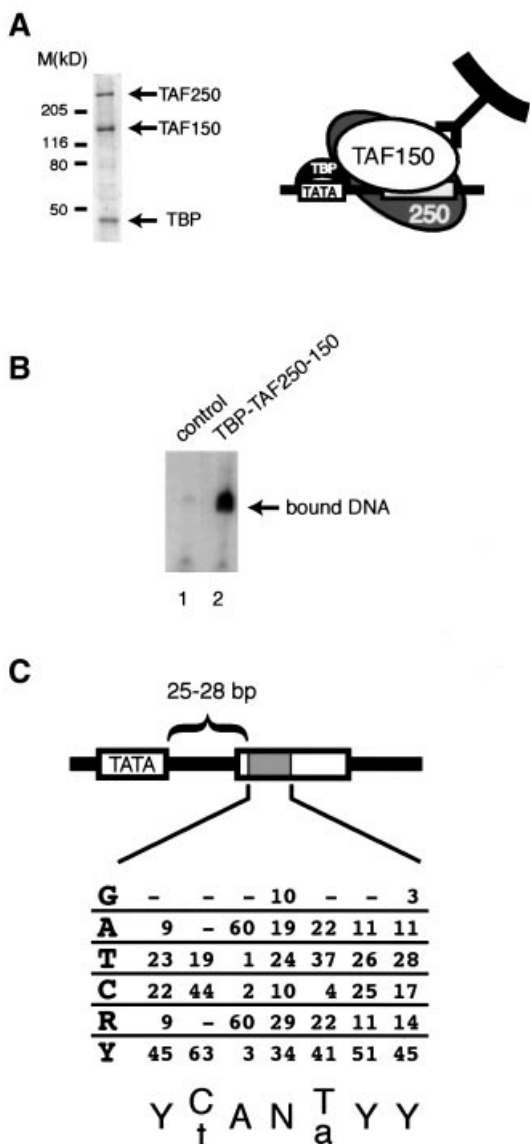
G.E.Chalkley and **C.P.Verrijzer**

A

B

C

Fig. 2. Selection of DNA sequences from a pool of quasi-random oligonucleotides with a TBP–TAF$_{II}$250–TAF$_{II}$150 complex. (**A**) Recombinant, purified TBP, TAF$_{II}$150 and TAF$_{II}$250 were assembled into a trimeric TBP–TAF$_{II}$250–TAF$_{II}$150 complex. The resulting complex was analysed by SDS–PAGE followed by Coomassie Blue staining. The positions of TBP and TAFs and molecular weights (kDa) of protein standards are indicated. (**B**) The purified TBP–TAF$_{II}$250–TAF$_{II}$150 complex was used in DNA binding assays using a pool of radiolabelled oligonucleotides that contained a TATA-box 23 bp upstream of a stretch of 20 random basepairs (open box). The TATA-box and random sequences are flanked by primer and cloning sequences (solid bar, see Materials and methods for details). Binding reactions were performed with control beads (lane 1) or an immobilized TBP–TAF$_{II}$250–TAF$_{II}$150 complex (lane 2). Bound DNA was isolated and analysed by PAGE. (**C**) Sequences of the oligonucleotides recovered after six rounds of selection were aligned and the frequencies of each base, as well as those of purines (R) and pyrimidines (Y), at each position are indicated. This quantification was used to obtain a consensus sequence. A TATA–Inr spacing of 25, 26, 27 and 28 bp was present in 12, 18, 25 and eight of the selected sequences, respectively.

would constrain the freedom of the TAFs to interact with the remainder of a bound DNA molecule. Consequently, the location of sequences critical for TAF binding should also be restricted.

The TATA-box-containing oligonucleotides were used in a site selection procedure similar to that described above for the TBP–TAF$_{II}$250–TAF$_{II}$150 complex (Figure 2B). During these binding experiments we again noted a clear enrichment in the proportion of DNA bound in subsequent rounds of selection (data not shown). Oligonucleotides bound in the final round of selection were recovered, subcloned and 70 unique clones were sequenced.

Sequence analysis of the selected oligonucleotides revealed that 63 out of 70 contained a good match to the Inr consensus 25–28 nucleotides downstream of the TATA-box (Figure 2C). Tabulation of these sequences resulted in a Y(C/t)AN(T/a)YY consensus. Outside the Inr area we failed to detect any obvious sequence constraints. A TATA–Inr spacing of 25, 26, 27 and 28 bp was present in 12, 18, 25 and eight of the selected sequences, respectively. The location of the selected Inr element is in good agreement with the spacing found within natural promoters where the TATA-box and Inr element are typically separated by ~25–30 nucleotides. The remaining oligonucleotides that were sequenced (seven out of 70) contained stretches of A/T-rich sequences, indicating that they were selected via TBP binding, instead of recognition by the TAFs. The selection of Inr sequences at a restricted position by a trimeric TBP–TAF$_{II}$250–TAF$_{II}$150 complex strongly supports the notion that the TAF$_{II}$250–TAF$_{II}$150 complex recognizes the Inr.

Inr mutations inhibit binding of a TAF$_{II}$250–TAF$_{II}$150 complex

Do mutations within the Inr affect TAF binding? To test this directly, we performed band-shift assays with canonical and various mutant Inr elements. Figure 3 shows that a recombinant, purified TAF$_{II}$250–TAF$_{II}$150 complex can bind efficiently to an oligonucleotide containing a consensus Inr element (lanes 1 and 13). The integrity of the complex was confirmed with antibody inhibition and supershift experiments (data not shown). Next, a set of distinct oligonucleotides each containing distinct point mutations in the Inr sequence were used in binding competition experiments (Figure 3). Increasing amounts of unlabelled wild-type Inr DNA or various mutant Inr elements were added to binding reactions containing the TAF$_{II}$250–TAF$_{II}$150 complex and radiolabelled wild-type Inr. As expected, oligos A and F, containing a consensus Inr, competed efficiently for binding to the TAF$_{II}$250–TAF$_{II}$150 complex (compare lanes 1 and 13 with lanes 3, 4 and lanes 15, 16, respectively). In contrast, equal molar amounts of oligonucleotides containing distinct mutant Inr sequences, failed to compete efficiently for TAF$_{II}$250–TAF$_{II}$150 binding to the wild-type Inr (B–E, lanes 5–12, G and H, lanes 15–20). Thus, specific point mutations within the Inr sequence inhibit recognition by the TAF$_{II}$250–TAF$_{II}$150 complex.

Inr mutations that inhibit TAF binding also impair basal transcription

Do the mutations that inhibit TAF binding also impair Inr function during transcription? In order to investigate this possibility, we generated eight different transcription templates by cloning the various Inr oligonucleotides used in the DNA binding experiment, into a parental vector. The resulting plasmids (A–H) all contain three Sp1 sites
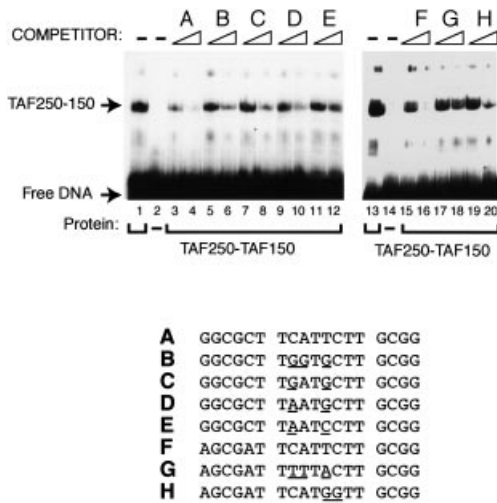
A GGCGCT TCATTCTT GCGG
B GGCGCT TGGTGCTT GCGG
C GGCGCT TGATGCTT GCGG
D GGCGCT TAATGCTT GCGG
E GGCGCT TAATCCTT GCGG
F AGCGAT TCATTCTT GCGG
G AGCGAT TTTTACTT GCGG
H AGCGAT TCATGGTT GCGG

Fig. 3. Mutations within the Inr sequence inhibit binding of a TAF$_{II}$250–TAF$_{II}$150 complex. DNA binding of a TAF$_{II}$250–TAF$_{II}$150 complex to double-stranded radiolabelled oligonucleotides containing a consensus Inr (A, lanes 1–12; F, lanes 13–20) was competed with unlabelled oligonucleotides containing distinct mutant Inr elements and analysed by EMSA. Lanes 2 and 14 contain no protein. Binding was tested in the absence of a specific competitor DNA (lane 1, probe A; lane 13, probe F) or in the presence of a 5- to 50-fold excess of cold competitor DNA corresponding to either the wild-type Inr (A, lanes 3 and 4; F, lanes 15 and 16) or various mutant Inrs (B–E, lanes 5–12; G and H, lanes 17–20). The position of free DNA and the TAF$_{II}$250–TAF$_{II}$150–DNA complex is indicated. The nucleotide sequences of the Inr region of the various oligonucleotides used in the EMSA experiment are listed and mutant nucleotides within the Inr element are underlined.

and a TATA-box upstream of the wild-type or mutant Inr sequences. Apart from the Inr sequence, all plasmids are identical.

The effects on basal promoter strength of the various Inr mutations were tested in reconstituted transcription reactions using a partially purified *Drosophila* embryo extract that provided the general transcription machinery, including endogenous TFIID. RNA transcripts were visualized by primer extension analysis. Figure 4A shows that all templates containing a mutant Inr (B–E, G and H, lanes 1–5, 7 and 8) were significantly weakened in their ability to direct basal transcription, compared with those containing a consensus Inr (A and F, lanes 1 and 6). The transcription levels of several independent experiments were quantified by PhosphorImager analysis and tabulated in Figure 4C. The effects of the distinct Inr mutations ranged from an ~3- (template E) to 25-fold (template G) reduction of basal promoter strength. In most cases, mutations in the Inr sequence also lead to a change in the start site of transcription as determined by primer extension analysis (indicated by an arrowhead in Figure 4C). We conclude that Inr mutations that affect TAF binding also impair basal promoter strength.

The Inr sequence can determine the response to an activator

Next, we investigated the effects of the Inr mutations on transcriptional activation. Transcription reactions were performed either in the absence or presence of increasing amounts of the activator Sp1 (Figure 4B). The transcription levels of several independent experiments were quantified by PhosphorImager analysis and tabulated in Figure 4C.
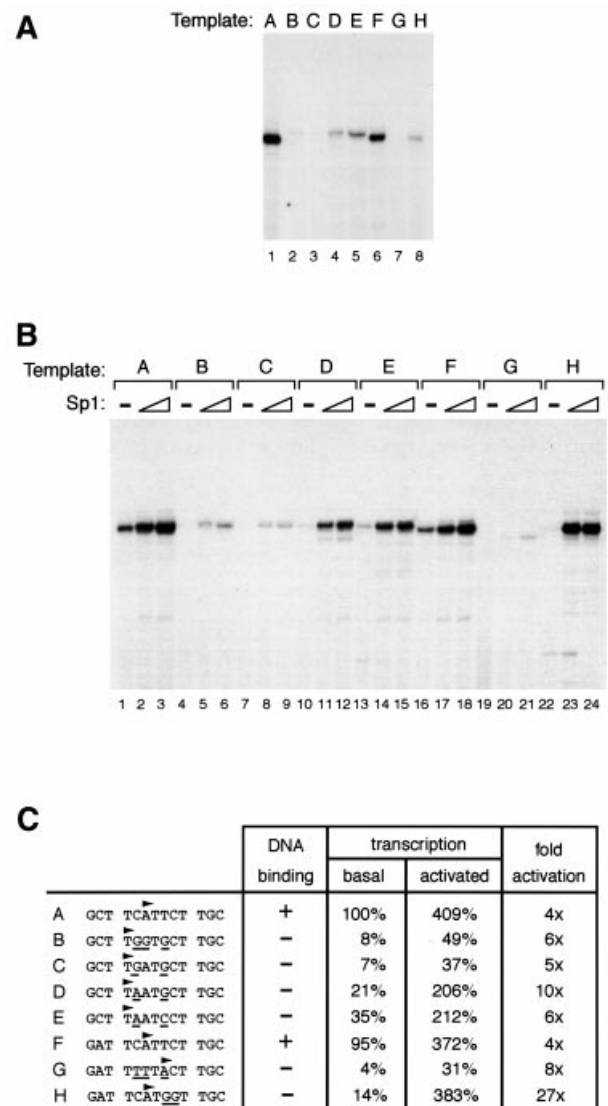

		DNA binding	transcription		fold activation
			basal	activated	
A	GCT TCATTCT TGC	+	100%	409%	4x
B	GCT TGGTGCT TGC	−	8%	49%	6x
C	GCT TGATGCT TGC	−	7%	37%	5x
D	GCT TAATGCT TGC	−	21%	206%	10x
E	GCT TAATCCT TGC	−	35%	212%	6x
F	GAT TCATTCT TGC	+	95%	372%	4x
G	GAT TTTTACT TGC	−	4%	31%	8x
H	GAT TCATGGT TGC	−	14%	383%	27x

Fig. 4. Effects of distinct Inr mutations on basal and activated transcription. (**A**) Basal transcription reactions with templates containing mutant Inrs. Distinct transcription templates were generated containing the various Inrs (indicated by the letters above the lanes, A–H) previously tested for binding to TAF$_{II}$250–TAF$_{II}$150 by EMSA (Figure 3). The basal transcription levels supported by these templates were compared in a *Drosophila* transcription system in the absence of an activator. Transcription products were visualized by primer extension. (**B**) Templates A–H were also tested for their ability to mediate Sp1 activation. As indicated, transcription reactions were performed in either the absence or presence of increasing amounts of recombinant, purified Sp1. Transcription products were visualized by primer extension. (**C**) Summary of the results from the DNA binding and transcription experiments. Mutant nucleotides within the Inr element are underlined. The transcription start sites on the distinct templates were determined by running sequence reactions, performed with the primer also used for primer extensions, in parallel with the transcription reactions. The major transcription start site is indicated with an arrowhead above the initiating nucleotide. The relative binding affinities determined by the band-shift competition experiments (Figure 3) are indicated qualitatively (+, wild-type affinity; −, decreased affinity). The transcription levels of five independent experiments were quantified by PhosphorImager analysis. The amount of basal and activated transcription directed by the distinct templates is expressed as a percentage of the amount of basal transcription directed by the wild-type template A (set at 100%). The variation between independent experiments was <10%. The fold-activation in the presence of Sp1 is also indicated.

Strikingly, we found that the effect of Inr mutations on activated transcription is not always proportional to that on basal transcription. Instead, distinct Inr mutations that reduce basal transcription to a similar extent can have very different consequences on the amount of activation by Sp1.

On the wild-type templates, Sp1 induced a 4-fold stimulation of transcription (Figure 4C; A, lanes 1–3; F, lanes 16–18). On mutant templates B, C and E, Sp1 gave a comparable, only marginally stronger, 5- to 6-fold activation. Thus, on these templates the Inr mutations lead to a similar reduction of basal and activated transcripion. In contrast, template H (Figure 4C; lanes 22–24), in spite of a much lower level of basal transcription, still supported a high level of activated transcription that is similar to that of the wild-type promoters (templates A and F). As a consequence of the reduced basal level, the activation by Sp1 was much stronger on this template than on the wild-type templates (27- versus 4-fold). Similarly, Sp1 activation on promoters D (Figure 4C; lanes 10–12) and G (lanes 19–21) was also significantly stronger than on the wild-type promoter (10- and 8-fold, respectively, compared with 4-fold). The mutant templates therefore fall into two classes. On templates B, C and E the levels of basal and activated transcripiton were reduced to a similar extent. In contrast, on templates D, G and H, basal transcription was much more impaired than activated transcription, resulting in an increased activation by Sp1. These experiments demonstrate that the DNA sequence at the start site of transcription can be a critical determinant of the responsiveness of a promoter to a transcriptional activator.

The results of the DNA binding and transcription experiments with the various Inr mutants are summarized in Figure 4C. Inr mutations that inhibit binding of the TAF$_{II}$250–TAF$_{II}$150 complex also impair basal promoter strength. Surprisingly, the effects of the activator Sp1 are determined, in part, by the structure of the basal promoter.

TAF$_{II}$150 recognizes structured DNA

Previous experiments established that TAF$_{II}$150 can bind DNA by itself and protects downstream promoter sequences, including the Inr, of the AdML and hsp70 promoters against DNase I digestion (Verrijzer *et al.*, 1994, 1995). Nevertheless, we failed to detect any obvious sequence motifs common to DNA areas protected by TAF$_{II}$150 (data not shown) nor did we detect any selected sequence preferences for this TAF in the site selection assays. These results indicate that the sequence requirements for TAF$_{II}$150 binding are not stringent enough to allow the determination of a clear consensus. Additionally, recognition of DNA structure, rather than specific sequences, may contribute to TAF$_{II}$150 binding.

To test this idea directly we performed band-shift assays with four-way junction DNA (4WJ DNA) and purified recombinant TAF$_{II}$150, TAF$_{II}$250 or a TAF$_{II}$250–TAF$_{II}$150 dimeric complex. As shown in Figure 5, TAF$_{II}$150 bound efficiently to synthetic 4WJ DNA (probe C, lane 7) but not to the corresponding duplex 'arms' (probes A and B, lanes 5 and 6) or the 'Y-form' (probe D, lane 8). These results indicate that, rather than a particular sequence, TAF$_{II}$150 recognizes specific features of the DNA structure. Competition experiments indicated that TAF$_{II}$150
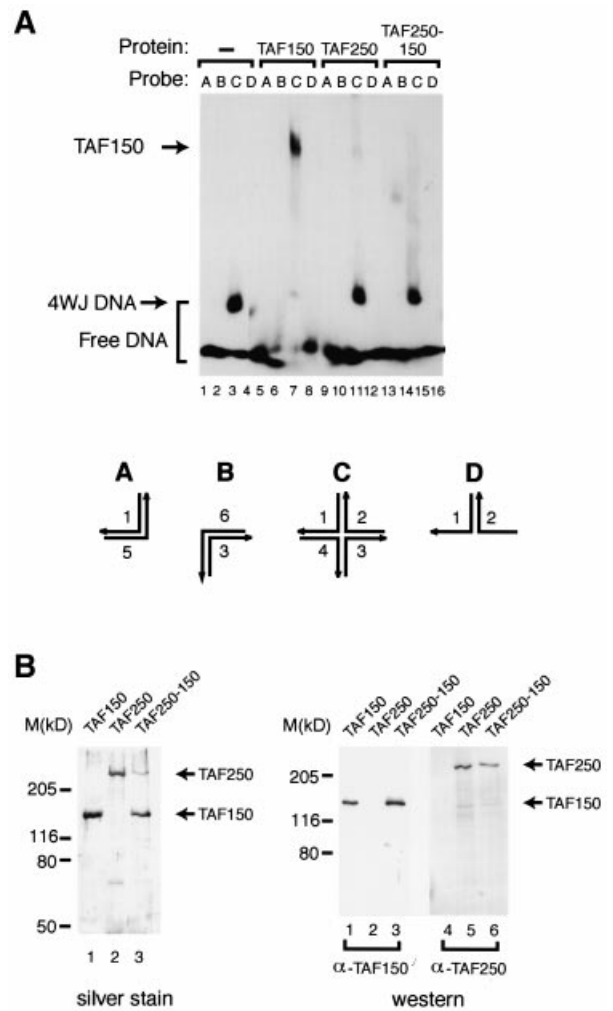

Fig. 5. Recognition of structured DNA by TAF$_{II}$150. (**A**) DNA binding of TAF$_{II}$150, TAF$_{II}$250 and TAF$_{II}$250–TAF$_{II}$150 complex to distinct structured DNA probes was compared. The labelled probes comprise two distinct duplex arms (probe A, lanes 1, 5, 9 and 13; probe B, lanes 2, 6, 10 and 14), the corresponding synthetic cruciform 4WJ DNA (probe C, lanes 3, 7, 11 and 15) and 'Y-form' (probe D, lanes 4, 8, 12 and 16) were incubated with either no protein (lanes 1–4), TAF$_{II}$150 (lanes 5–8), TAF$_{II}$250 (lanes 9–12) or TAF$_{II}$250–TAF$_{II}$150 complex (lanes 13–16) and analysed by EMSA. The positions of the free DNA, 4WJ DNA and the TAF$_{II}$150–4WJ DNA complex are indicated. The structures of probes A to D are illustrated schematically. (**B**) Silver-stained SDS–PAGE gel of the purified recombinant TAF$_{II}$150 (lane 1), TAF$_{II}$250 (lane 2) and TAF$_{II}$250–TAF$_{II}$150 complex (lane 3) used in the EMSA shown in (A). The right-hand panel shows a Western blot analysis of the same complexes with antibodies directed against either TAF$_{II}$150 (lanes 1–3) or TAF$_{II}$250 (lanes 4–6). The positions of TAFs and molecular weights (kDa) of protein standards are indicated.

binds 4WJ DNA with an affinity that is more than one order of magnitude greater than that for the corresponding duplex arms (data not shown).

It has been noted previously that TAF$_{II}$250 contains a region of similarity to AT-hook DNA binding domains (Aravind and Landsman, 1998). Typical AT-hook domains have been implicated in the recognition of structured DNA. However, we did not observe binding of TAF$_{II}$250 to 4WJ DNA (Figure 5; lane 11). Surprisingly, a dimeric TAF$_{II}$250–TAF$_{II}$150 complex did not bind cruciform DNA either (Figure 5; compare lanes 15 and 7). This result indicates that the association with TAF$_{II}$250 impairs the

binding of $TAF_{II}150$ to 4WJ DNA. Silver-staining and Western blot analysis showed that approximately equal amounts of $TAF_{II}150$ were present in the binding reactions containing $TAF_{II}150$ or the $TAF_{II}250–TAF_{II}150$ complex (Figure 5B). The negative effect of $TAF_{II}250$ on $TAF_{II}150$ binding to 4WJ DNA is reminiscent of its inhibition of TBP binding to the TATA-box. It should be noted that in this and in previous studies (Chen *et al.* 1994; Verrijzer *et al.*, 1994, 1995) we observed efficient DNA binding of recombinant $TAF_{II}250–TAF_{II}150$, $TBP–TAF_{II}250$ and $TBP–TAF_{II}250–TAF_{II}150$ complexes. It is most likely that $TAF_{II}250$ has to bind to a correctly spaced Inr element in order to neutralize its inhibition of DNA binding by associated TBP and $TAF_{II}150$.

These experiments suggest that DNA binding by $TAF_{II}150$ is mediated by recognition of structural features of the DNA rather than by a strictly defined primary sequence.

Discussion

TFIID is the first basal factor to bind to a core promoter where it nucleates the recruitment of RNA pol II and the basal machinery. The results presented here provide evidence that the Inr sequence is bound specifically by a complex of the TFIID subunits $TAF_{II}250$ and $TAF_{II}150$. Binding of these TAFs to the Inr correlates with core promoter strength since mutations that inhibit TAF binding also impair basal transcription. Furthermore, we observed that the precise DNA sequence at the start site of transcription can be an important determinant of the level of activated transcription directed by an upstream activator. Finally, our results revealed that $TAF_{II}150$ specifically recognizes DNA structure, suggesting that promoter binding by TFIID may rely, in part, on 'indirect readout'.

Promoter recognition by TFIID
From the data presented here and in other studies, a picture of TFIID emerges in which its subunit architecture reflects the organization of basal promoters. In other words, distinct core promoter elements can be considered to form an array of binding sites for distinct TFIID subunits. Thus, the TATA-box is bound by TBP (Burley and Roeder, 1996), the Inr by a $TAF_{II}250–TAF_{II}150$ dimer (this study) and the DPE by a $TAF_{II}60–TAF_{II}40$ heterotetramer (Burke and Kadonaga, 1997). Such an arrangement of promoter-bound DNA is illustrated in Figure 6. At present, a role in promoter recognition for some of the other TAFs cannot be excluded either (see e.g. Oelgeschläger *et al.*, 1996). Likewise, sequences other than the core motifs uncovered so far, can help determine basal promoter strength.

An interesting feature of the core promoter motifs is their relatively flexible sequence requirements: many A/T-rich sequences can impart TATA activity, the Inr consensus is rather loose and partial DPEs have been described. However, the need for multiple, correctly juxtaposed elements greatly increases the specificity of TFIID binding. Such combinatorial requirements for binding are not limited to TFIID but also involve TFIIB, RNA pol II and possibly other basal transcription factors. Thus, a multitude of individually relatively weak protein–DNA and protein–protein interactions together, make PIC forma-
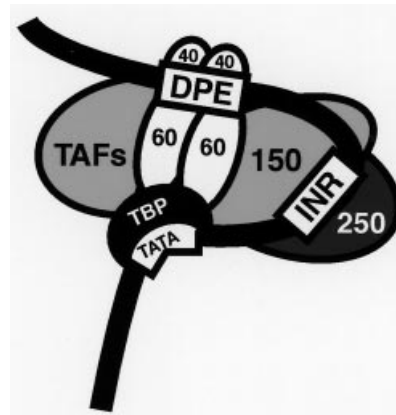

Fig. 6. A tentative model for TFIID–promoter binding. This cartoon illustrates that distinct TFIID subunits target different core promoter elements. TFIID is formed by a multitude of TAF–TAF and TBP–TAF interactions, of which only a few are indicated here. It should be noted that apart from interactions between particular pairs of TFIID subunits, the disposition of TAFs within TFIID is unknown. Within the TFIID–promoter complex, TBP binds the TATA-box, $TAF_{II}250$–$TAF_{II}150$ complex binds the Inr and $TAF_{II}60$–$TAF_{II}40$ complex binds the DPE. Depicted here is an idealized promoter that contains all three core elements recognized by TFIID. Within natural promoters, one or more of the core elements is typically lacking or highly degenerated. Moreover, sequences other than the three core elements shown here can also influence promoter strength. The promoter DNA is depicted wrapped around TFIID. DNA binding might induce isomerization of TFIID that could influence the communication with activators. See text for discussion of the model and references.

tion and initiation of transcription a highly specific process that does not occur randomly on the genome.

The main conclusion of the present study, that a $TAF_{II}250–TAF_{II}150$ complex targets the Inr, agrees well with results from other studies. First, we previously demonstrated, by reconstitution of TFIID with recombinant subunits, that both these TAFs are required for discrimination between Inr-containing and Inr-less promoters (Verrijzer *et al.*, 1995). Promoter selectivity results not only from stabilization of TFIID–DNA binding, but also from an inhibition of TBP–TATA-box interactions in the absence of a docking site for $TAF_{II}250$ (Kokubo *et al.*, 1993; Verrijzer *et al.*, 1995; Burley and Roeder, 1998; Liu *et al.*, 1998; see also Chen *et al.*, 1994; Verrijzer *et al.*, 1994). Secondly, DNA cross-linking and other DNA binding studies suggested that both these TAFs are in intimate contact with the promoter DNA, including the Inr region (Gilmour *et al.*, 1990; Sypes and Gilmour, 1994; Verrijzer *et al.*, 1994, 1995; Oelgeschläger *et al.*, 1996). It should be noted that the 135 kDa subunit of human TFIID, which can be cross-linked to the AdML promoter (Oelgeschläger *et al.*, 1996) is now considered likely to be the human $TAF_{II}150$ (Martinez *et al.*, 1998). Thirdly, $TAF_{II}150$ was purified from *Drosophila* embryos and from human cells as an essential cofactor for Inr-dependent transcription reconstituted with $TAF_{II}150$ stripped TFIID (Hansen and Tjian, 1995; Kaufman *et al.*, 1996, 1998). Finally, studies in mammalian and yeast cells demonstrated that $TAF_{II}250$ and its yeast homologue $TAF_{II}145$, function as core promoter selectivity factors *in vivo* (Shen and Green, 1997; Wang *et al.*, 1997).

The dual requirement for $TAF_{II}150$ and $TAF_{II}250$ could be the result of both proteins directly contacting the Inr. Alternatively, one protein may specifically recognize the

Inr sequence, whereas the other protein stabilizes the complex by making sequence-independent DNA contacts. Since $TAF_{II}150$ can bind DNA by itself, this protein would be a good candidate for the latter function with $TAF_{II}250$ providing specific Inr recognition. Finally, it is possible that the binding of $TAF_{II}150$ and $TAF_{II}250$ to each other induces a conformational change that exposes the DNA binding domain.

DNA binding by $TAF_{II}150$ alone does not depend on the precise Inr sequence nor can $TAF_{II}150$, in the absence of $TAF_{II}250$, mediate Inr function (Verrijzer *et al.*, 1994, 1995; Kaufman *et al.*, 1998; this study). Since the DNase I footprint of $TAF_{II}150$ on the AdML or the *Drosophila* hsp70 promoter extends significantly beyond the Inr element, additional DNA sequences appear to contribute to $TAF_{II}150$ binding (Verrijzer *et al.*, 1994, 1995). However, we failed to identify a critical consensus sequence for $TAF_{II}150$ DNA binding. Instead, our results indicate that DNA secondary structure can be an important determinant for $TAF_{II}150$ binding. We also observed that association with $TAF_{II}250$ inhibits binding of $TAF_{II}150$ to 4WJ DNA. Interestingly, $TAF_{II}250$ has a similar inhibitory effect on DNA binding by TBP (Kokubo *et al.*, 1993; Verrijzer *et al.*, 1995; Burley and Roeder, 1998; Liu *et al.*, 1998). It is probable that docking of $TAF_{II}250$ on a correctly positioned Inr element neutralizes its inhibition of DNA binding by associated proteins TBP and $TAF_{II}150$. Indeed, here and in previous studies we observed efficient DNA binding of recombinant $TAF_{II}250$–$TAF_{II}150$, TBP–$TAF_{II}250$ and TBP–$TAF_{II}250$–$TAF_{II}150$ complexes to Inr-containing promoters (Chen *et al.* 1994; Verrijzer *et al.*, 1994, 1995).

The functional significance of the recognition of structured DNA by $TAF_{II}150$ is unclear at this moment. It is possible that the Py-rich Inr region adopts a secondary structure that deviates from regular B-form DNA. It has also been proposed that the promoter DNA wraps around the TFIID complex and forms a nucleosome-like structure (Oelgeschläger *et al.*, 1996; Hoffmann *et al.*, 1997). It is therefore of interest that 4WJ DNA is believed to resemble a DNA crossover structure such as the point of DNA entry and exit of a nucleosome. Likewise, DNA wrapping around TFIID may create a DNA crossover point that is recognized by $TAF_{II}150$, which may stabilize a stereo-specific nucleoprotein structure.

The core promoter and transcriptional activation

Gene-specific activators are the main regulators of gene expression in eukaryotic cells. This has led to the perception that transcription is controlled strictly via enhancers and that core promoters are merely passive docking sites for the basal machinery. However, a number of recent reports, including this study, emphasize that the basal promoter structure can be a major determinant of the effects elicited by a transcriptional activator. When fused to the GAL4 DNA binding domain, the activation domain of VP16 or one of the SP1 activation domains, shows different activation profiles on distinct core promoters (Emami *et al.*, 1995). In *Drosophila* embryos it has been elegantly demonstrated that the core promoter structure can determine selectivity for particular natural activators *in vivo* (Ohtsuki *et al.*, 1998). Finally, the results presented here show that the precise sequence at the start site

of transcription can significantly influence the level of activation achieved by the transcriptional activator Sp1.

Taken together, these studies demonstrate that recognition of the basal promoter can play a prominent role during transcriptional activation by upstream binding regulators. Thus, the great diversity among natural core promoters might allow different genes to respond differently to a particular activator.

How can the core promoter sequence influence the responsiveness to activating signals? One attractive possibility is that TFIID adopts distinct conformations on different core promoters that are either more or less receptive to activating signals from particular activators. Such a mechanism is not unprecedented since a number of examples of DNA-induced allosteric effects during transcriptional regulation have been described (reviewed in Lefstin and Yamamoto, 1998; see also Chi and Carey, 1996; Emami *et al.*, 1997). Reversibly, activators can induce extended TFIID–DNA contacts, probably by changes in the TFIID conformation (Horikoshi *et al.*, 1988). It will be important to obtain direct proof of isomerization of TFIID by techniques such as atomic force microscopy or protease sensitivity mapping.

An alternative explanation for the influence of core promoter structure on transcriptional activation could be provided by the binding of distinct TFIID-related complexes. However, depletion of the transcription extract with antibodies against either TBP, $TAF_{II}150$ or $TAF_{II}250$ effectively abolished transcription (data not shown). Therefore, we consider this an unlikely model for the Inr effects described here. Nevertheless, it is important to stress that a number of TFIID-related complexes have been described (reviewed in Lee and Young, 1998). Transcription of a subset of snRNA genes by RNA pol II is mediated by SNAPc, a distinct TBP–TAF complex that binds its target promoters via the proximal sequence element and directs selective activation by Oct-1 (Hernandez, 1993; Das *et al.*, 1995; Lee and Young, 1998). MOT-1 is a negative regulator of RNA pol II transcription that associates with TBP (Auble *et al.*, 1994; van der Knaap *et al.*, 1997). Human $TAF_{II}30$ is present in a separate TFIID complex and is required for oestrogen receptor function (Jacq *et al.*, 1994). Surprisingly, a TFTC, TBP-lacking TAF complex, has been identified that can mediate transcription in the absence of TFIID (Wieczorek *et al.*, 1998). Human $TAF_{II}105$ is a tissue-specific, substoichiometric subunit of TFIID that is involved in expression of anti-apoptotic genes (Yamit-Hezi and Dikstein, 1998). Finally, a *Drosophila* tissue-specific TBP-related factor, TRF, has been identified as part of a multi-protein complex that is distinct from TFIID (Hansen *et al.*, 1997). Thus, it is tempting to speculate that transcription initiation on certain RNA pol II promoters may involve distinct TFIID-related complexes.

In summary, our findings reveal that the Inr is specifically recognized by a $TAF_{II}250$–$TAF_{II}150$ complex. Surprisingly, the Inr sequence not only determines the basal promoter strength but also influences the responsiveness of a promoter to activating signals. These results indicate that recognition of the core promoter may be more intimately tied to the regulation of transcription by activators than previously anticipated.

Materials and methods

Expression and purification of recombinant proteins

Constructs used to express recombinant HA-tagged hTAF$_{II}$250, FLAG-tagged dTAF$_{II}$150, Flag-tagged Sp1 and 6× His-tagged hTBP have been described previously (Chen *et al.*, 1994; Verrijzer *et al.*, 1995; Yokomori *et al.*, 1998; Ryu *et al.*, 1999). The TAFs and Sp1 were expressed in Sf9 cells using the baculovirus expression system and purified essentially as described (Verrijzer *et al.*, 1995; Chen and Tjian, 1996). Briefly, recombinant baculoviruses were plaque purified and amplified. For protein expression, Sf9 cells were infected at a m.o.i. of ~5 and harvested 48 h post infection. All protein procedures were carried out at 4°C or on ice using HEMG buffer (25 mM HEPES–KOH pH 7.6, 0.1 mM EDTA, 12.5 mM MgCl$_2$ and 10% glycerol) containing 1 mM DTT, 0.2 mM AEBSF [-(2-aminoethyl)benzenesulfonyl fluoride], 1 µM pepstatin, 0.01% NP-40 and varying amounts of KCl. Whole-cell extracts were prepared by sonication in 0.4 M KCl–HEMG containing 0.1% NP-40. After centrifugation at 100 000 *g*, the tagged proteins were immuno-purified from the supernatant using protein A–Sepharose beads (Pharmacia) covalently conjugated (Harlow and Lane, 1988) with either anti-HA (12CA5, Zhou *et al.*, 1992) or anti-FLAG (Kodak) monoclonal antibodies. The TAF$_{II}$250–TAF$_{II}$150 and TBP–TAF$_{II}$250–TAF$_{II}$150 complexes were assembled *in vitro* as described (Chen and Tjian, 1996). For band-shift experiments, proteins were eluted under native conditions using peptides corresponding to the appropriate epitope (HA: YPYDVPDYA; FLAG: DYKDDDDK) in HEMG buffer containing 100 mM KCl. Recombinant His-tagged human TBP was expressed in *Escherichia coli* BL21 and purified as described (Yokomori *et al.*, 1998). Briefly, after induction, extracts were prepared by sonication in HNGN (25 mM HEPES–KOH pH 7.6, 0.5 M NaCl, 10% glycerol, 0.1% NP-40) containing 0.5 mg/ml lysozyme, 0.2 mM PMSF, 0.1 mM sodium metabisulfite and 0.2 mM AEBSF. TBP was purified by NTA-Ni (Qiagen) chromatography and eluted with HNGN, 0.01% NP-40 containing 500 mM imidazole (pH 7.5). The eluate was diluted to 200 mM NaCl and loaded onto a 2 ml SP Sepharose fast flow column (Pharmacia), equilibrated with HEMG–200 mM KCl. After extensive washes, the column was developed with a linear gradient of 200–800 mM KCl in HEMG and TBP with a high specific activity eluted from the column at ~500 mM KCl. All protein fractions were aliquotted and stored at –80°C. Proteins and complexes immobilized on beads were either used directly or stored at 4°C.

Binding site selection

All DNA binding reactions were performed in DB buffer (12.5 mM HEPES–KOH pH 7.6, 0.05 mM EDTA, 6.25 mM MgCl$_2$, 5% glycerol, 0.01% NP-40, 50 mM NaCl, 0.2 mM AEBSF, 0.1 mM leupeptin and 0.1 mM pepstatin). The *in vitro* binding site selection procedure was derived from Pollock and Treisman (1990) with modifications described below. For site selection with the TAFs we used a double-stranded oligonucleotide probe comprising a 34 bp random DNA sequence flanked by primer and cloning sequences (*Bam*HI, *Kpn*I), PV64: ACGGAT-CGGTCAGCGGATCCGGTTC(N)$_{34}$GAGGCGGTACCAGTGCAAGC-TCAGC; reverse primer PV62: GCTGAGCTTGCACTGGTACC-GCCTC and forward primer PV63: ACGGATCGGTCAGCGGAT-CCGGTTC. For the experiments with the TBP–TAF$_{II}$250–TAF$_{II}$150 complex we used a probe, loosely based on the *Drosophila* hsp70 promoter, that contains a TATA-box (underlined) 23 bp upstream of 20 bp random sequence flanked by primer and cloning sequences (*Bam*HI, *Kpn*I), PV121: CGGGATCC*TATAAATA*GCGGCGCTTCGT-CTACGGAGCGA(N)$_{20}$ GAGGCGGTACCAGTGCAAGCTCAGC, PV62 as reverse primer and PV121 as forward primer: CGGGAT-CCTATAAATAGCGGCGCTT.

TAF$_{II}$150, TAF$_{II}$250–TAF$_{II}$150 or TBP–TAF$_{II}$250–TAF$_{II}$150 were immobilized via the TAF$_{II}$150 Flag epitope on M2 affinity gel (Kodak). TAF$_{II}$250 was immobilized on anti-HA coated protein A beads. About 15 µl of the appropiate TAF beads were pre-incubated in 25 µl DB buffer containing 0.1 mg/ml BSA and 10 µg poly(dGdC)·(dGdC). Next, TAF beads were added to 50 µl DB containing: 0.1 mg/ml BSA, 1 µg poly(dGdC)·(dGdC) and 100 pmol of PV64 or PV121 double-stranded probes end-labelled with polynucleotide kinase (PNK) using standard procedures (Sambrook *et al.*, 1989). After incubation on a rotating wheel for 1.5 h at room temperature, unbound probe was removed by extensive washes with DB buffer. Bound DNA was eluted from the beads at 45°C in 200 µl elution buffer (5 mM EDTA, 0.5% SDS, 100 mM NaOAc, 50 mM Tris pH 8.0), phenol extracted and recovered by EtOH precipitation. The DNA was resuspended in TE (10 mM Tris pH 8.0, 0.1 mM EDTA) and amplified by nine PCR cycles in the presence of [α-^{32}P]dCTP

to body-label the oligonucleotide pools. The amplified DNA was purified on a 6% polyacrylamide gel (35.5:1, acrylamide:bis) containing 0.5× TBE buffer, eluted according to standard procedures (Sambrook *et al.*, 1989) and used in further rounds of selection. Recovered DNA from the final round of selection was subcloned into pBluescript and individual clones were sequenced on both strands.

DNA binding assays

In order to monitor the enrichment in TAF binding sites, samples of recovered DNA were taken after each round of selection and stored. These DNA samples were amplified and body-labelled by six PCR cycles in the presence of [α-^{32}P]dCTP and purified on a 6% polyacrylamide gel using standard procedures (Sambrook *et al.*, 1989). Probes were quantified by Cerenkov counting and ethidium bromide staining of agarose gels. Approximately equal amounts of DNA were tested for binding of TAF$_{II}$250–TAF$_{II}$150 in band-shift experiments essentially as described previously (Verrijzer *et al.*, 1995). Briefly, binding reactions were carried out for 30 min at 28°C in 20 µl DB buffer containing 80 ng poly(dGdC)·(dGdC), 50 µg/ml BSA, 0.05% NP-40 and 4 µl of the eluted TAF$_{II}$150–TAF$_{II}$250 complex. Samples were analysed on a 5% polyacrylamide gel (35.5:1 acrylamide:bis), containing 0.5× Tris–glycine buffer, 0.01% NP-40. For competition experiments a double-stranded oligonucleotide, probe A: GGCGCTTCATTCTTGCGG, containing a consensus Inr (underlined), was prepared by end-labelling with T4 PNK and purified using standard procedures (Sambrook *et al.*, 1989). Binding reactions were for 30 min at 28°C carried out in 20 µl DB buffer containing 200 ng poly(dGdC)·(dGdC), 50 µg/ml BSA, 1 mM DTT, ~10 fmol probe A in the absence or presence of a 5- to 50-fold excess of unlabelled specific competitor oligonucleotide (B–H). Apart from the mutations indicated in Figure 3, the other oligonucleotides are similar to A. The oligonucleotides used to assemble 4WJ DNA binding experiments have been described before (oligonucleotides 1–6; Bianchi, 1988). For the cruciform binding assays, oligonucleotides 1 and 3 were end-labelled using T4 PNK. The appropriate combinations of oligonucleotides were annealed by incubation at 95°C for 2 min, followed by 10 min at 65°C, 10 min at 37°C and 10 min at room temperature. Annealed cruciforms were purified on a 10% non-denaturing acrylamide gel run at 4°C. About 1000 c.p.m. of probe was used in each binding reaction. EMSA was performed essentially as described (Bianchi, 1988). Binding reactions were carried out in DBF buffer (8% Ficoll, 16 mM HEPES–KOH pH 7.6, 13 mM MgCl$_2$, 5 mM KCl, 20 mM NaCl, 1 mM EDTA, 1 mM spermidine, 0.5 M DTT, 200 µg/ml BSA) in the presence of 5 ng poly(dGdC)·(dGdC). Samples were analysed as described above on a 4% polyacrylamide gel run at 4°C.

In vitro transcription reactions

Synthetic oligonucleotides were cloned into pBluescript to create a transcription template (pSTI) that contains three consensus Sp1 binding sites, a TATA-box and a consensus Inr flanked by *Xho*I and *Nhe*I sites. The distinct oligonucleotides used in the DNA binding expriments (A–H) were cloned into the parental vector using the *Xho*I and *Nhe*I sites. The relevant part of the sequence of the core promoter containing the wild-type Inr (template A), with the TATA-box and Inr underlined, is: CCGGAGTATAAATAGAGGCGCTTCCTCGAGACGATTCATTCT-TGCGGCTAG. All other templates were similar with the exception of the Inr mutations indicated in Figure 3. Transcription reactions and primer extension analysis were carried out essentially as described (Kadonaga, 1990; Verrijzer *et al.*, 1995). Transcription reactions were performed in a volume of 25 µl and contained 100 ng of template. The transcription machinery was provided by addition of 1 µl heparin 0.4 M fraction (Austin and Biggin, 1996). The transcription start sites were determined by running sequence reactions, performed with the same primer as used for the primer extensions, in parallel with the transcription reactions. Quantification of the transcription gels was by PhosphorImager analysis (Molecular Dynamics)

References

Aravind,L. and Landsman,D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.*, **26**, 4413–4421.

Arkhipova,I.R. (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics*, **139**, 1359–1369.

Aso,T., Conaway,J.W. and Conaway,R.C. (1994) Role of core promoter structure in assembly of the RNA polymerase II preinitiation complex. A common pathway for formation of preinitiation intermediates at many TATA and TATA-less promoters. *J. Biol. Chem.*, **269**, 26575–26583.

Auble,D.T., Hansen,K.E., Mueller,C.G., Lane,W.S., Thorner,J. and Hahn,S. (1994) Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.*, **8**, 1920–1934.

Austin,R.J. and Biggin,M.D. (1996) Purification of the *Drosophila* RNA polymerase II general transcription factors. *Proc. Natl Acad. Sci. USA*, **12**, 5788–5792.

Bianchi,M.E. (1988) Interaction of a protein from rat liver nuclei with cruciform DNA. *EMBO J.*, **7**, 843–849.

Burke,T.W. and Kadanoga,J.T. (1996) *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.*, **10**, 711–724.

Burke,T.W. and Kadonaga,J.T. (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAF$_{II}$60 of *Drosophila*. *Genes Dev.*, **11**, 3020–3031.

Burley,S.K. and Roeder,R.G. (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.*, **65**, 769–799.

Burley,S.K. and Roeder,R.G. (1998) TATA box mimicry by TFIID: autoinhibition of Pol II transcription. *Cell*, **94**, 551–553.

Carcamo,J., Buckbinder,L. and Reinberg D. (1991) The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc. Natl Acad. Sci. USA*, **88**, 8052–8056.

Chen,J.-L. and Tjian,R. (1996) Reconstitution of TATA-binding protein associated factor/TATA-binding protein complexes for *in vitro* transcription. *Methods Enzymol.*, **273**, 208–217.

Chen,J.-L., Attardi,L., Verrijzer,C.P., Yokomori,K. and Tjian,R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93–105.

Chi,T. and Carey,M. (1996) Assembly of the isomerized TFIIA–TFIID–TATA ternary complex is necessary and sufficient for gene activation. *Genes Dev.*, **10**, 2540–2550.

Das,G., Hinkley,C.S. and Herr,W. (1995) Basal promoter elements as selective determinants of transcriptional activator function. *Nature*, **374**, 657–660.

Emami,K.H., Navarre,W. and Smale,S.T. (1995) Core promoter specificities of the Sp1 and VP16 transcription activation domains. *Mol. Cell. Biol.*, **15**, 5906–5916.

Emami,K.H., Jain,A. and Smale,S.T. (1997) Mechanism of synergy between TATA and initiator: synergistic binding of TFIID following a putative TFIIA-induced isomerization. *Genes Dev.*, **11**, 3007–3019.

Emanuel,P.A. and Gilmour,D.S. (1993) Transcription factor TFIID recognizes DNA sequences downstream of the TATA element in the *hsp*70 heat shock gene. *Proc. Natl Acad. Sci. USA*, **90**, 8449–8453.

Gilmour,D.S., Dietz,T.J. and Elgin,S.C. (1990) UV cross-linking identifies four polypeptides that require the TATA box to bind to the *Drosophila* hsp 70 promoter. *Mol. Cell. Biol*., **10**, 4233–4238.

Hahn,S. (1998) The role of TAFs in RNA polymerase II transcription. *Cell*, **95**, 579–582.

Hampsey,M. (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Rev.*, **62**, 465–503.

Hampsey,M. and Reinberg,D. (1997) Transcription: why are TAFs essential? *Curr. Biol.*, **7**, R44–R46.

Hansen,S. and Tjian,R. (1995) TAFs and TFIIA mediate differential utilization of the tandem Adh promoters. *Cell*, **82**, 565–575.

Hansen,S.K., Takada,S., Jacobson,R.H., Lis,J.T. and Tjian,R. (1997) Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell*, **91**, 71–83

Harlow,E. and Lane,D. (1988) *Antibodies*: *A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hernandez,N. (1993) TBP, a universal eukaryotic transcription factor? *Genes Dev.*, **7**, 1291–1308.

Hoffmann,A., Oelgeschläger,T. and Roeder,R.G. (1997) Considerations of transcriptional control mechanisms: Do TFIID–core promoter complexes recapitulate nucleosome-like functions? *Proc. Natl Acad. Sci. USA*, **94**, 8928–8935.

Horikoshi,M., Hai,T., Lin,Y.S., Green,M.R. and Roeder,R.G. (1988)
Transcription factor ATF interacts with the TATA factors to facilitate establishment of a pre-initiation complex. *Cell*, **54**, 1033–1042.

Jacq,X., Brou,C., Lutz,Y., Davidson,I., Chambon,P. and Tora,L. (1994) Human TAF$_{II}$30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell*, **79**, 107–117.

Kadonaga,J.T. (1990) Assembly and disassembly of the *Drosophilia* RNA polymerase II complex during transcription. *J. Biol. Chem.*, **265**, 2624–2631.

Kaufmann,J. and Smale,S. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.*, **8**, 821–829.

Kaufmann,J., Verrijzer,C.P., Shao,J. and Smale,S.T. (1996) CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev.*, **10**, 873–886.

Kaufmann,J., Ahrens,K., Koop,R., Smale,S.T. and Müller, R. (1998) CIF150, a human cofactor for TFIID-dependent initiator function. *Mol. Cell. Biol.*, **18**, 233–239.

Kokubo,T., Gong,D.-W., Yasmashita,S., Horikoshi,M., Roeder,R.G. and Nakatani,Y. (1993) Drosophila 230-kD TFIID subunit, a functional homolog of the human cell cycle product, negatively regulates DNA binding of the TATA box-binding subunit of TFIID. *Genes Dev.*, **7**, 1033–1046.

Lagrange,T., Kapanidis,A.N., Tang,H., Reinberg,D. and Ebright,R.H. (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.*, **12**, 34–44.

Lee,T.I. and Young,R.A. (1998) Regulation of gene expression by TBP-associated proteins. *Genes Dev.*, **12**, 1398–1408.

Lefstin,J.A. and Yamamoto,K.R. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature*, **392**, 585–588.

Liu,D., Ishima,R., Tong,K.I., Bagby,S., Kobuko,T., Muhandiram,D.R., Kay,L.E., Nakatani,Y. and Ikura,M. (1998) Solution structure of a TBP-TAF$_{II}$230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell*, **94**, 573–583.

Martinez,E., Chiang,C.-M., Ge,H. and Roeder,R.G. (1994) TAFs in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.*, **13**, 3115–3126.

Martinez,E., Zhou,Q., L'Etoille,N.D., Oelgeschläger,T., Berk,A.J. and Roeder,R.G. (1995) Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. *Proc. Natl Acad. Sci. USA*, **92**, 11864–11868.

Martinez,E., Ge,H., Tao,Y., Yuan,C.-X., Palhan,V. and Roeder,R.G. (1998) Novel factors and TFIIA mediate functional core promoter selectively by the human TAF$_{II}$150-containing TFIID complex. *Mol. Cell. Biol.*, **18**, 6571–6583.

Nakatani,Y., Horikoshi,M., Brenner,M., Yamamoto,T., Bresnard,F., Roeder,R.G. and Freese,E. (1990) A downstream initiation element required for efficient TATA box-binding and *in vitro* function of TFIID. *Nature*, **348**, 86–88.

Oelgeschläger,T., Chiang,C.-M. and Roeder,R.G. (1996) Topology and reorganization of a human TFIID–promoter complex. *Nature*, **382**, 735–738.

Ohtsuki,S., Levine,M. and Cai,H.N. (1998) Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.*, **12**, 547–556.

Orphanides,G., Lagrange,T. and Reinberg,D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev.*, **10**, 2657–2683.

Pollock,R. and Treisman,R. (1990) A sensitive method for the determination of protein–DNA binding specificities. *Nucleic Acids Res.*, **18**, 6197–6204.

Purnell,B.A., Emanuel,P.A. and Gilmour,D.S. (1994) TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.*, **8**, 830–842.

Qureshi,S.A. and Jackson,S.P. (1998) Sequence-specific DNA binding by the *S.shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell*, **1**, 389–400.

Robert,F., Douziech,M., Forget,D., Egly,J.M., Grenblatt,J., Burton,Z.F. and Coulombe,B. (1998) Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol. Cell*, **2**, 341–351.

Roeder,R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.*, **21**, 327–335.

Ryu,S., Zhou,S., Ladurner,A.G. and Tjian,R. (1999) The transcriptional co-factor complex CRSP is required for activity of the enhancer binding protein Sp1. *Nature*, **397**, 446–450.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sawadogo,M. and Roeder,R.G. (1985) Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell*, **43**, 165–175.

Shen,W.-C. and Green,M.R. (1997) Yeast TAFμ145 functions as a core promoter selectivity factor, not a general coactivator. *Cell*, **90**, 615–624.

Smale,S.T. (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta*, **1351**, 73–88.

Smale,S.T. and Baltimore,D. (1989) The 'initiator' as a transcription control element. *Cell*, **57**, 103–113.

Sypes,M.A. and Gilmour,D.S. (1994) Protein–DNA crosslinking of a TFIID complex reveals novel interactions downstream of the transcription start. *Nucleic Acids Res.*, **22**, 807–814.

van der Knaap,J.A., Borst,J.W., van der Vliet,P.C., Gentz,R. and Timmers,H.T.M. (1997) Cloning of the cDNA for the TATA-binding protein associated factor(II)70 subunit of transcription factor B-TFIID reveals homology to global transcription regulators in yeast and *Drosophila. Proc. Natl Acad. Sci. USA*, **94**, 11827–11837.

Verrijzer,C.P. and Tjian,R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.*, **21**, 338–342.

Verrijzer,C.P., Yokomori,K., Chen, J.-L. and Tjian,R. (1994) *Drosophila* TAF$_{II}$150: Similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science*, **264**, 933–941.

Verrijzer,C.P., Chen,J.-L., Yokomori,K. and Tjian,R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115–1125.

Wang,J.C. and Van Dyke,M.W. (1993) Initiator sequences direct downstream promoter binding by human transcription factor IID. *Biochim. Biophys. Acta*, **1216**, 73–80.

Wang,E.H., Zou,S. and Tjian,R. (1997) TAF$_{II}$250-dependent transcription of cyclin A is directed by ATF activator proteins. *Genes Dev.*, **11**, 2658–2669.

Weis,L. and Reinberg,D. (1997) Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding protein associated factors and initiator-binding proteins. *Mol. Cell. Biol.*, **17**, 2973–2984.

Wieczorek,E., Brand,M., Jacq,X. and Tora,L. (1998) Function of TAF$_{II}$-containing complex without TBP in transcription by RNA polymerase II. *Nature*, **393**, 187–191.

Yamit-Hezi,A. and Dikstein,R. (1998) TAF$_{II}$105 mediates activation of anti-apoptotic genes by NF-κB. *EMBO J.*, **17**, 5161–5169.

Yokomori,K., Verrijzer,C.P. and Tjian,R. (1998) An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. *Proc. Natl Acad. Sci. USA*, **95**, 6722–6727.

Zhou,Q., Liebermann,P.M., Boyer,T.G. and Berk,A.J. (1992) Holo-TFIID supports transcriptional stimulation by diverse activators and from TATA-less promoters. *Genes Dev.*, **6**, 1964–1974.