Minimal Components of the RNA Polymerase II Transcription Apparatus Determine the Consensus TATA Box

Gudrun Bjornsdottir
Dartmouth College

Lawrence C. Myers
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the Life Sciences Commons

Recommended Citation
Bjornsdottir, Gudrun and Myers, Lawrence C., "Minimal Components of the RNA Polymerase II Transcription Apparatus Determine the Consensus TATA Box" (2008). Open Dartmouth: Faculty Open Access Articles. 3822.
https://digitalcommons.dartmouth.edu/facoa/3822
Minimal components of the RNA polymerase II transcription apparatus determine the consensus TATA box

Gudrun Bjornsdottir and Lawrence C. Myers*

Department of Biochemistry, Dartmouth Medical School, Hanover, NH 03755, USA

Received January 16, 2008; Revised March 6, 2008; Accepted March 10, 2008

ABSTRACT

In *Saccharomyces cerevisiae*, multiple approaches have arrived at a consensus TATA box sequence of TATA(T/A)A(A/T)(A/G). TATA-binding protein (TBP) affinity alone does not determine TATA box function. To discover how a minimal set of factors required for basal and activated transcription contributed to the sequence requirements for a functional TATA box, we performed transcription reactions using highly purified proteins and *CYC1* promoter TATA box mutants. The TATA box consensus sequence is a good predictor of promoter activity. However, several nonconsensus sequences are almost fully functional, indicating that mechanistic requirements are not the only selective pressure on the TATA box. We also found that the effect of a mutation at a certain position is often dependent on other bases within a particular TATA box. Although activators and coactivators strongly influence TBP recruitment and stability at promoters, neither Mediator, the activator Gal4-V16, nor TFIID specifically compensate for the low transcription levels of the weak TATA boxes. The addition of Mediator to purified transcription reactions did, however, increase the functional selectivity for certain consensus TATA sequences. Transcription in whole-cell extracts or in vivo with these TATA box mutants indicated that factors, other than those in our purified system, may help initiate transcription from weak TATA boxes.

INTRODUCTION

Efficient and accurate transcription initiation of eukaryotic genes requires specific sequences in the core promoter. The core promoter of a gene includes the transcription start site as well as the region immediately surrounding this site. Various functional DNA motifs, known as core promoter elements, assist in the recruitment, assembly and initiation of the RNA polymerase II (RNA Pol II) transcription machinery (1). Originally assumed to be largely invariant in strength and sequence, core promoters and their elements have been revealed to be highly variable (2). Hence, the core promoter can be a point of regulation for gene-specific transcription. Little is known, however, about what mechanisms and transcription factors are required to read and execute the information encoded by core promoter elements (1). Although core promoter elements are currently better defined in metazoans, the core promoter elements in the yeast *Saccharomyces cerevisiae* appear to share many of the same features (2). The TATA box is perhaps the best-known core promoter element, but the exact requirement for, and role of, TATA-like sequences in a core promoter is still not well understood.

Early studies led to the view that most mRNA synthesis requires a TATA box in order to recruit the TATA-binding protein (TBP) to a promoter. Over the past decade, however, many studies have demonstrated large classes of promoters that do not follow this rule (3). Among the most striking were studies that revealed the TATA box is not a general component of all Pol II core promoters, despite the TBP requirement for all Pol II transcription. Detailed analyses have shown that only ~20% of yeast core promoters (4) and ~10% of mammalian core promoters (2) have a readily identifiable TATA box. Why the TATA box is required for the transcription of some genes, but not others, and whether the mechanism of initiation differs according to the presence of a TATA box are open questions.

Biochemical, genetic and bioinformatic approaches have been taken to determine the sequence requirements for the TATA box. Initial genetic studies on the TR TATA element of the *S. cerevisiae HIS3* promoter indicated very stringent sequence requirements for the TR element in vivo (5,6) and in an extract-based transcription system.
in vitro (7). A similar, but not identical, pattern was observed for the HIS3 T_R TATA box in a reconstituted system using a partially fractionated basal transcription system (8). Although replacing single bases within the T_R element led to severe impairment of HIS3 transcription in vivo (5), the fact that the replacement of the entire T_R element by a variety of DNA sequences could lead to high levels of HIS3 transcription (9) suggested a greater variability in functional TATA boxes.

Initial efforts to identify functional TATA boxes and define the TATA box consensus sequence focused on single-gene analyses. A recent study (4), however, reevaluated the S. cerevisiae TATA box consensus sequence using a genome-wide bioinformatic approach. This statistical method, based on the combination of genome-wide sequence conservation and experimental data, identified the extent of TATA box usage and the most highly utilized TATA box sequences. The consensus sequence TATA(T/A)A(T/A)(A/G) determined in this study revealed a previously unidentified preference for (A/G) at the eighth position. It is currently unknown what dictates the selection of this consensus sequence. Indirect factors, such as the counter-selection of sequences bound by other specific DNA-bound proteins may influence the consensus sequence. Direct factors, on the other hand, shape the consensus based on their direct mechanistic requirements for transcription.

TBP binding to the TATA box (10–12) is a direct determinant of the consensus sequence. TBP binds to the minor groove of the DNA helix and has many contact points on the TATA box. Even though TBP affinity to TATA boxes plays a role in dictating the consensus (13), experiments have failed to show any direct correlation between TBP affinity and transcriptional activity (14,15). TATA box DNA bound by TBP exhibits a dramatic 80° bend (10,11,16). Although the structure of the TBP–DNA complex and the bending angle of the DNA provide important clues to how the sequence of the TATA box influences transcription, it is still unclear exactly how differences in transcriptional activity correspond to differences in the TBP–DNA structure (15,17–20). Although TBP affinity for TATA boxes and the DNA bending induced by TBP play an important role in dictating the consensus TATA box, TBP may not be the only factor influencing the sequence.

Activators, coactivators and general transcription factors (GTFs) other than TBP may also dictate the bases of the TATA box. Activators interact physically with both components of the basal transcription machinery (21) and with coactivator complexes (22–24). The activator fusion protein Gal4-VP16 can stabilize TBP on TATA boxes, and this effect is more pronounced on weak TATA boxes (25). It has also been shown that TATA box utilization in a partially defined basal transcription system differs from activated transcription in vivo (8). The role played by activators in TATA box utilization also may include coactivators, such as Mediator and TFIID.

Mediator enhances basal transcription, and is required for activated transcription in vivo and in vitro (26). It has physical and functional interactions with both activators and the core transcription machinery. Biochemical and structural studies have shown that Mediator is likely to interact with many components of the core RNA Pol II transcription machinery (27), including TBP (28,29). TBP recruitment to certain promoters in vivo is impaired in Mediator mutants (30–33), and the recruitment of GTFs to a preinitiation complex assembled from extracts is impaired by Mediator mutants (34). The functional interplay between TBP and Mediator suggests Mediator may influence TATA box utilization.

Another coactivator complex that is likely to be important for TATA box usage is TFIID. TFIID consists of TBP and TBP-associated factors (TAFs) and is important for the expression of, and TBP recruitment to, a subset of genes in vivo (35–37). Genome-wide studies have revealed both a physical and a functional interaction between TFIID and TATA-less promoters in vivo (4,38). Other studies have shown that mutations, which weaken the TATA box of the yeast CYC1 promoter, convert this promoter from a TFIID independent to a TFIID-dependent promoter (39).

To address the origin of the TATA box consensus sequence and determine what components of the transcription machinery influence its usage, we have functionally evaluated the yeast TATA box consensus sequence using a biochemical approach. This is the first systematic analysis of transcriptional activities of different TATA box sequences in vitro using a highly purified S. cerevisiae transcription system capable of activated transcription. In addition, we have compared the transcription levels in the purified system to levels seen in a whole-cell extract-based system, and to the transcription levels of the same TATA box variants in a reporter assay in vivo. Our studies show that the ability of the purified apparatus to initiate transcription is a strong, but not exclusive, determinant of a consensus TATA box. There are likely mechanistic subtleties to TATA box utilization that involve both factors tested in our current study as well as factors yet to be characterized.

MATERIALS AND METHODS

Transcription templates

All transcription templates that were used in this study are derivatives of the DNA plasmid pJJ470 (40). This template contains a truncated CYC1 core promoter (−139 to −36, relative to the translational start site) that retains only one of the two primary functional TATA boxes identified in studies of the native CYC1 promoter (41). Mutations in the wild-type CYC1 TATA box in our template (TATATAAA) were introduced by amplifying an XhoI–AflIII fragment from pJJ470 using PCR primer pairs consisting of the primer pUCaFlR2 (5′-GCTGGCCCTTTTCGCATCAG-3′) and one of the primers listed in Supplementary Table S1, depending on the desired mutation(s). The PCR products were digested with XhoI and AflIII and ligated into pJJ470 that had been digested with the same enzymes, creating pJJ470(M) plasmids. Two minor TATA-like sequences have also been identified in the CYC1 core
promoter sequence (41). Mutations in these TATA-like sequences were introduced by amplifying the pJJ470(M) plasmids using the primers pJJ470F1 (5'-CGCGCAG GTCCCTTTGTAGCATAAAATTACTA-3') and pJJ470R1 (5'-CTCAAAAGGACCTAAGTATAAGAAGAACTCTTACGAGAAG-3'). The product was digested with PpuMI and re-ligated, creating pJJ470(MA) plasmids. All constructs were confirmed by DNA sequencing.

Protein purification and extract preparation

TFIID was purified from a yeast strain SHY761 (a generous gift from Dr Steven Hahn) in which both Taf1p and Taf7p are Flag-tagged (24). The cells were grown and lysed in a blender as previously described (42). The extract (2 g of protein) was adjusted to 175 mM KOAc via dialysis and fractionated on a 400-ml BioRex 70 column as described (43) with the following modifications: the loading buffer was 50 mM Hepes–KOH (pH 7.6), 1 mM EDTA, 10% glycerol, 175 mM KOAc, 5 mM β-mercaptopetoethanol and protease inhibitors. The column was washed once with the loading buffer and twice in the same buffer with 300 mM KOAc. The elution buffer was 20 mM HEPES–KOH (pH 7.6), 0.01% NP-40, 10% glycerol, 150 mM KOAc, 0.5 mM DTT, protease inhibitors. Peak fractions were pooled and dialyzed against dialysis buffer [20 mM HEPES–KOH (pH 7.6), 0.01% NP-40, 10% glycerol, 150 mM KOAc, 0.5 mM DTT, protease inhibitors] to 300 mM KOAc. FLAG purification was performed as described (44) except that the elution buffer had 0.001% NP-40. The peak fractions from four 200 μl FLAG columns were pooled and fractionated on a Mono S 5/5 column as described (43). The TFIID fraction was concentrated by dialysis in a membrane with a molecular weight cutoff of 12–14 000 kDa against 20 mM HEPES–KOH pH 7.6, 10% glycerol, 0.001% NP-40, 150 mM KOAc, 30% (w/v) PEG 20 000, 1 mM DTT and protease inhibitors.

The transcription factors RNA Pol II, rTBP, TFIIB, rTFIIE, TFIIF, TFIH, Mediator and Gal4-VP16 were purified as previously described (45). The pLGSD5 reporter plasmids. Functional evaluation of the yeast consensus TATA sequence TATA(T/A)(T/A)(A/G) using a purified basal transcription system

To functionally evaluate the TATA(T/A)(T/A)(A/G) consensus sequence identified by Pugh and colleagues (4), we tested transcriptional activities of 23 variations (Figure 1B) of this sequence using a G-less cassette assay with the minimal set of purified factors capable of basal transcription (51,52). The G-less cassette assay template pJJ470 (40) has a Gal4-binding site upstream of a CYC1 core promoter fragment, which has the consensus TATA box TATATAATTTAGTGTGTGTATTTGTGTTTGCGTGT using a purified basal transcription system

Basal transcription activities of the CYC1 TATA box mutants in vitro were measured in the presence of the general transcription factors TBP, TFIIB, TFIIE, TFIIF, TFIH and RNA Pol II. In our assay, most consensus TATA boxes showed high transcription activities, whereas most nonconsensus TATA boxes showed low transcription activities (Figure 2). A second TATA-like, but nonconsensus, sequence (TATAAAC) in the CYC1 promoter is offset from the wild-type TATA box
Figure 2. Transcription activities of CYC1 TATA box mutants in a purified system in vitro. Mutations that represent another consensus base (4) at that position are underlined. Mutations that are sensitive to mutations in the DNA-binding surface of TBP (4) are marked with an asterisk. Transcription levels were measured in a G-less cassette assay. Basal transcription was measured in the presence of the highly purified general transcription factors RNA Pol II, TBP, TFIIB, TFIIE, TFIIF and TFIIH. Enhanced transcription was measured in the presence of purified GTFs and Mediator. Activated transcription was measured in the presence of purified GTFs, Mediator and Gal4-VP16. Shown is the quantification of transcription signals, expressed as a percentage of the signal from a wild-type TATA box template. Representative primary data are shown in Figures 3–5. The CYC1 core promoter/G-less cassette has two major start sites for transcription. We chose only the upstream start site (longer transcript) for quantification. Most columns represent the mean of at least three independent experiments. Reactions with signals that were almost undetectable were only measured twice. Error bars indicate the SEM. RNA recovery controls were run for most reactions and showed a consistent recovery over all samples checked.
by +2 bp. This sequence makes a small, if any, contribution to CYC1 transcription. The mutation of T to A in the first position of the wild-type sequence, which does not affect the alternate sequence, almost eliminates all transcription in our purified reactions (Figure 2). As observed in previous studies (8), the difference in transcription levels cannot be attributed only to differences in TBP affinity since the reactions were performed in the presence of excess TBP. This conclusion is also supported by our data showing that increasing the amount of TBP did not have a differential effect on the transcription of a wild-type template as compared to one of the weaker mutant TATA boxes (Supplementary Figure S1).

All TATA consensus and nonconsensus sequences tested had transcription signals lower than the wild-type, indicating that the TATATAAA in the CYC1 promoter is optimized for its transcriptional activity. Mutation to a nonconsensus base in the central positions (T3–A6) of the TATA box generally had the most detrimental effects on transcription (Figure 2). Very little selectivity was observed for consensus bases in the eighth position (Figure 2) in reactions reconstituted from the purified basal transcription machinery. A nonconsensus C or T at the eighth position gave a transcription signal similar to a consensus G and wild-type A. The only other nonconsensus single base mutation that gave >70% of the signal observed with the wild type was the mutation of the fourth position to a T (Figure 3). One of the criteria used to define consensus TATA boxes via the statistical approach (4) was TBP sensitivity. Promoters classified as 'TBP sensitive' showed decreased transcription in vivo in yeast strains with mutations in the DNA-binding surface of TBP (53). The three nonconsensus TATA boxes that exhibited high levels of transcription in our assay (TATTTAAA, TATATAAC and TATATAAT) (Figures 2 and 3), as well as the consensus sequences, all were scored as being TBP sensitive (4). Among the criteria used in the study, TBP sensitivity may be the most accurate in predicting a TATA box that will function with the basal transcription machinery. Generally, however, the yeast consensus TATA sequence [TATA(T/A)A(T/A) (A/G)] defined through bioinformatics approaches (4), is a good predictor of transcriptional activity of the CYC1 TATA box in vitro. One aspect of TATA box mechanism apparent in our studies that was not clearly apparent from the bioinformatic approach was the functional codependency of changes in different positions in the TATA box.

codependency of bases within the CYC1 TATA box for transcription

Two results in the initial mutation analysis, where single base changes were made to the CYC1 TATA box sequence, helped reveal a functional coupling between bases in the CYC1 TATA box. In basal reactions with purified factors, the consensus TATA box TATATAAAA gave a surprisingly low transcription signal (<40% of wild type) and the nonconsensus TATA box TATTTTTAAA gave a surprisingly high transcription signal (>75% of wild type) (Figure 2). As described above, the nonconsensus TATA box TATTTTTAAA is TBP sensitive in vivo, which correlated with its ability to facilitate transcription in vitro. However, if a second change is made in this TATA box such that the T in the fifth position is changed to another consensus base (A), the resulting TATTTAAA sequence drops transcription below 20% of wild type (Figure 2). In accordance with our observations above, TATTTAAA was not classified as TBP sensitive (4). Of the two possible consensus bases in the fifth position, only T allows a TATA box with a T in the fourth position to facilitate basal transcription with purified factors. Changing the consensus T to a consensus A in the fifth position is also coupled to the eighth position.

Although many yeast promoters contain and conserve the consensus TATATAAAA TATA box (4), using this sequence in our basal reactions led to low levels of relative transcription (Figures 2 and 4, lanes 1–3 versus 10–12). An A in the fifth position, however, does not exclusively lead to low levels of transcription in vitro. The Adenovirus Major Late (AdML) promoter has an A in the fifth position and has been shown to give strong transcription signals in vitro in both yeast and metazoan transcription systems (54, 55). We noticed that the AdML TATA box (TATAAAG) had a consensus G in the eighth position as opposed to the A in the eighth position of the CYC1 TATA box. We therefore decided to test if substituting the A in the eighth position of TATAAAG for a G, creating TATAAAAG, would work better in the context of the CYC1 promoter. The double mutant gave significantly higher relative transcription levels than the single mutant (Figures 2 and 4). Thus, our results show that in the context of the CYC1 promoter, the ability to use an A in the fifth position of the TATA box relies on having a G in the eighth position. Although the substitution of a G for an A in the eighth position led to the largest suppression of the in vitro defect of the A in the fifth position, both a nonconsensus C and T in the eighth position also reproducibly led to small increases in transcription. The suppressing effect of G in the eighth position was specific for an A in the fifth position as neither a G nor any other alternative base suppressed the
and 5). In both of these cases, the substitution of the T in the presence of Mediator than in its absence (Figures 2 and 5). Substitution of a T in the eighth position of the TATA box, but is functional with either A, C or G. Shown are basal, enhanced basal and activated transcription activities of CYC1 TATA box mutants and wild type in a purified system in vitro. Lanes 1–3 show the transcription decrease with a T5A mutation in the TATA box. Placing a G in the eighth position compensates for the transcription defect in the TATAAAAA sequence (lanes 4–6). Placing a G in the eighth position of the wild-type TATA box, however, leads to no increase in transcription (lanes 7–9). G-less cassette assays were performed in the presence of highly purified RNA Pol II, TBP, TFIIB, TFIIE, TFIIF and TFIIH, in addition to the factors indicated. Salt concentration was 165 mM KOAc and reaction time was 60 min for all reactions. Mutated bases are shown in bold.

Mediator increases the selectivity for consensus bases at the sixth and eighth position of the TATA box, but neither Mediator nor Gal4-VP16 specifically compensate for low transcription levels of weak TATA boxes

The minimal purified basal transcription system is not capable of responding to DNA-bound transcriptional activators (40). The addition of purified Mediator to the minimal basal transcription system enhances basal transcription and enables this machinery to respond to activators (40 and Figure 5, lanes 1–3). We repeated the transcription assays on the wild-type and mutant CYC1 TATA boxes with Mediator and activators to determine whether the step(s) of transcription affected by these factors functionally interacted with the TATA box sequence.

First, we tested how Mediator-enhanced basal transcription was affected by the TATA sequence. The pattern of TATA box-dependent transcription observed using the purified general transcription factors was largely unchanged when Mediator was added to the system (Figure 2). Mediator does not specifically compensate for low transcription levels of weak TATA boxes. Mediator’s stabilizing physical and/or functional interactions with TBP and, perhaps, the DNA itself do not specifically facilitate the step of transcription initiation governed by the TATA box sequence. The addition of Mediator to the reactions, however, did enhance the selectivity of the transcription apparatus for consensus bases at the sixth and eighth position. The replacement of a consensus A at either the sixth or eighth position with a nonconsensus T led to a much greater decrease in relative transcription in the presence of Mediator than in its absence (Figures 2 and 5). In both of these cases, the substitution of the T for the A prevented Mediator from enhancing basal transcription (Figures 2 and 5, lanes 1, 2 versus 10, 11). The ability of a nonconsensus base at these positions to prevent the enhancement of basal transcription by Mediator was specific to T. Neither the replacement of A with G at the sixth position nor the replacement of A with C at the eighth position had a similar effect as replacement with T (Figures 2 and 5).

Second, we looked at the effect of the hybrid activator Gal4-VP16 on TATA box utilization. We measured transcriptional activities of the CYC1 TATA box mutants in a system containing Mediator and Gal4-VP16 in addition to the factors used to measure basal transcription. In the pJJ470 vector, Gal4-VP16 binds to a Gal4 site upstream of the CYC1 core promoter and stimulates transcription in the presence of Mediator. The addition of Gal4-VP16 to the Mediator based reactions did not change the pattern observed in the absence of the activator (Figure 2). Interestingly, even though Mediator had little to no enhancement of basal transcription with the TATATAAT and TATATATAA TATA boxes, Mediator was still able to facilitate a similar fold activation of transcription as with the wild-type TATA box (Figures 2 and 5, lanes 2, 3 versus 11,12). The addition of the activator Gal4-VP16 leads to no further TATA box selectivity beyond that generated by Mediator alone.

TFIID does not specifically compensate for low transcription levels of weak TATA boxes

TATA-less promoters (4,38) as well as some TATA containing promoters with a weak mutated TATA box (39) have been shown to be TFIID dependent, while most promoters with strong consensus TATA box are not (4,38). Hence, we decided to test if TFIID can promote high levels of transcription from weak TATA boxes. If TFIID facilitates transcription initiation at promoters without a strong consensus TATA box in vitro, then we would expect to see a larger effect of purified TFIID on
weak TATA box mutants than on the wild type or strong mutants. We chose a subset of mutants and tested their transcription activities in a purified system with or without TFIIID. TFIIID was purified from a strain that had a Flag tag on the Taf1p and Taf7p subunits of TFIIID (24). We supplemented all TFIIID transcription reactions with enough recombinant TBP to establish an ~1:1 stoichiometry among TBP and the TAFs as well as to assure identical TBP concentrations in reactions with and without TFIIID (48). In addition to TFIIID and/or TBP, the transcription reactions were performed in the presence of RNA Pol II, TFIIB, TFIIH, TFIIF, TFIIIA, and Mediator as described (48). TFIIID dependency was tested both in the presence and absence of Gal4-VP16. As shown previously, TFIIID stimulated transcription with the consensus CYC1 TATA box (48). TFIIID showed an equal fold stimulation of transcription on a variety of mutant TATA boxes of varying strength (Figure 6). These results show that purified TFIIID alone does not specifically compensate for low transcription levels of weak TATA boxes.

Components of a whole-cell extract are able to enhance transcription from weak TATA boxes

Our results have shown that neither Mediator, Gal4-VP16, nor TFIIID compensate specifically for low transcription levels from weak TATA boxes in a purified system. Additional factors, independently or in association with these minimal components of the transcription machinery, may compensate specifically for low relative transcription levels from mutant TATA boxes. In order to test whether weak mutant TATA boxes had the potential to direct high levels of transcription, we performed transcription assays with a subset of our templates using a transcription competent whole-cell extract (46). The whole-cell extract was supplemented with Gal4-VP16 in the transcription reactions, while all other factors were supplied by the extract. The results from the transcription assays performed with the whole-cell extract show that transcription levels normalized to the wild-type TATA box were overall higher than in the purified system (Figure 7). Specifically, we observed an increase in the relative transcription levels of some of the weak TATA box mutants (Figure 7). The overall pattern of TATA box dependence remained similar to the purified system as the very weakest TATA boxes in the purified system (TAGAAAA and TATAGAAA) also gave the weakest transcription in the extract system (Figure 7). The very strongest mutants in the purified system (relative transcription to wild type >70%) exhibited only a slight increase in the extract transcription system (Figure 7). Only one mutant TATA box (TATATAAG) showed a relative transcription signal slightly >100%. These results show that the wild-type TATA box is still optimal for transcription in the extract system. The ability of the extract system to utilize some nonconsensus TATA boxes more efficiently than the purified system indicates that there likely are additional factors in the whole-cell extract participating in the assembly of active preinitiation complexes on weak TATA box promoters. Also of note is that the codependence between TATA box bases in the fifth and the eighth positions (TATAAAA versus TATAAAG) and the discrimination between bases in the eighth position (TATATAAT versus TATATAAC or TATATAAG) observed in the complete purified system (see above) do not occur in the extract-based system (Figure 7).

Transcription activities of TATA box mutants in vivo correlate with their activities in an extract-based system

To determine whether the results of either the extract transcription system or the purified transcription system accurately reflected the discrimination of TATA boxes in vivo, we performed a reporter assay with the CYC1 promoter TATA box and mutant variants. We introduced our mutated TATA boxes into a lacZ reporter plasmid in which a GAL1, 10 upstream activation sequence (UAS) in front of the CYC1 promoter drives the expression of the reporter gene (49). This system requires an activator to obtain a detectable signal. We induced transcription by expression of Gal4-VP16(F456A). Gal4-VP16(F456A), a point mutant with slightly attenuated activation potential, is used since expression of wild-type Gal4-VP16 is toxic to yeast (49). We found that the results of the extract-based system in vitro are highly representative of signal measured using a β-galactosidase assay, with respect to the relative transcription activities of the TATA box.
mutants (Figure 7). The only notable difference was that the reporter assay showed a much greater decrease in relative transcription for the very weakest TATA boxes compared to the extract system (Figure 7). Generally, these results indicate that the extract-based system in vitro is a good indicator of how different TATA boxes direct transcription in vivo in the presence of chromatin.

**DISCUSSION**

We have performed a systematic analysis to functionally evaluate the recently derived yeast TATA box consensus sequence (4) using a highly purified transcription system capable of basal, Mediator-enhanced, TFIID-enhanced and activated transcription. We have compared these results to transcription using an identical promoter in extracts and in vivo. Considerable effort has gone into studying the intricate mechanisms and structural details of TBP binding to the TATA box (56). Comprehensive structural studies have noted the disparity that cocrystal structures of TBP with multiple mutant TATA boxes are largely the same despite 20-fold differences in transcription among these mutants in vivo (15). These studies have concluded that factors beyond the static structure of the complex must be responsible for variations in transcriptional efficiency. Ultimately the critical ‘output’ of a TATA box is the ability to facilitate accurate and efficient transcription initiation. Our studies represent the first effort to evaluate the functional properties of a yeast TATA box using a highly purified transcription system capable of activated transcription. Our study demonstrates that the disparity between structural compatibility and transcription activity extends to a minimal purified system. We have made several important observations.

In general, the functional TATA box sequences best utilized by the purified transcription machinery correspond closely to the consensus sequence TATA (T/A)(T/A)(A/G) derived via a genome-wide statistical approach (4). It is likely that selective pressure on TATA boxes in gene expression originates from the minimal set of components required to initiate transcription from a core promoter. These results are complimentary to those obtained for the HIS3 TATA box using a partially defined transcription system (8). The high levels of relative transcription for the nonconsensus sequences TATTTAAA and TATATAAC observed in our purified system as well as the genome-wide statistical study ranking these sequences as ‘TBP-sensitive’ demonstrate that these two sequences are mechanistically consensus TATA boxes. Why these two sequences did not score highly in the criteria of conservation and location in the genome-wide statistical study is unclear, although it is almost certainly the result of indirect rather than direct selection. The inability of the consensus TATA box TATTTAAAAA to produce high levels of transcription in our purified system is in opposition to the results of the genome-wide statistical approach. This TATA box scored high in both the location analysis and TBP sensitivity criteria used in this study (4). One possibility is that this consensus sequence produces less efficient transcription by design.
and is used as part of a broad-based regulatory scheme. A second possibility is that this difference may be a result of context dependence. A TATA box that works well in one promoter context might not work well in another context. TBP binding to TATA boxes has recently been shown to be dependent on the flanking sequences (57). It is likely that promoter context, either flanking sequence or the presence of another core promoter element such as an initiator (2), may influence the ability for certain TATA box sequences to function. Our studies have already shown that other bases within the TATA box influence the effect of an A in the fifth position on transcription. The suppression of the transcription defect in the consensus TATAAAAA sequence by the replacement of an A in the eighth position by a consensus G provides initial evidence that the A5 is locally context dependent.

Despite extensive contacts between Mediator and the core transcription machinery, Mediator does not specifically compensate for low levels of weak TATA boxes. Previous studies from our lab have shown that Mediator can enhance basal transcription in purified reactions by compensating for limiting amounts of a general transcription factor, such as TBP (58). Since our data and others (8) show that increasing TBP concentration does not increase relative transcription from a mutant TATA box, it is perhaps not surprising that Mediator does not compensate for weak TATA boxes. Mediator does, however, seem to interact functionally with the sixth and the eighth position of the TATA box in the purified system. Substitution of a T in the sixth or the eighth position of the CYC1 TATA box prevents Mediator from enhancing basal transcription. Structural models of Mediator within a preinitiation complex (27) suggest it makes contacts with several components of the basal transcription apparatus. It is currently unclear how Mediator makes formation of an active preinitiation complex more efficient. The result that a T in the sixth or eighth position inhibits this action, however, suggests that Mediator may allosterically modulate interactions between general transcription factors that occur near the 3' end of the TATA box, such as the TBP-TFIIB interaction. In contrast to these results, the fold activation by Gal4-VP16, facilitated by Mediator in the purified transcription system, does not vary with the sequence of the TATA box. It is unlikely that the step of transcription initiation stimulated by Gal4-VP16 in the purified reactions is the same step as regulated by the TATA box. It is unlikely that the step of transcription initiation stimulated by Gal4-VP16 in the purified system is the same step as regulated by the TATA box. It is unlikely that the step of transcription initiation stimulated by Gal4-VP16 in the purified system is the same step as regulated by the TATA box. It is unlikely that the step of transcription initiation stimulated by Gal4-VP16 in the purified system is the same step as regulated by the TATA box. It is unlikely that the step of transcription initiation stimulated by Gal4-VP16 in the purified system is the same step as regulated by the TATA box.

TFIID, in vivo, is specifically required for the transcription of most TATA-less promoters (4,38) as well as the CYC1 promoter with a weak mutated TATA box (39). Generally, TFIID has less widespread effects on TATA-containing promoters (38). Hence, it was somewhat surprising that holo-TFIID did not specifically compensate for weak relative transcription from nonconsensus TATA boxes in the purified system. Several explanations for this result are possible. First, TFIID might require factors in addition to those in the purified system to help initiate transcription from TATA-less promoters. Second, TFIID may have some DNA sequence requirements for its action that are not present in the fragment of the CYC1 promoter fused to the G-less cassette transcription template. Lastly, it is possible that TFIID exerts its influence on TATA-less promoters in vivo through interactions with nucleosomes. TFIID is known to have interactions with acetylated (59) and methylated (60) histones. Although neither Mediator nor TFIID specifically compensated for weak TATA boxes in the CYC1 promoter, some factor in yeast extracts does.

The results from the transcription assays performed with the whole-cell extract show that transcription levels normalized to the wild-type TATA box were overall higher than in the purified system. This pattern was largely replicated using the same promoter in a reporter assay system in vivo. There was, however, a notable difference in the weak TATA boxes TAGAAAAA, TATAGAAA, TATATGAA and TATATTAA, which showed considerably lower relative signals in the reporter assays in vivo compared to the signals in the extract transcription system in vitro. The source of this difference is unknown, but one explanation could be that there is a factor in vivo which has a negative effect on initiation of certain weak TATA boxes. Perhaps this factor is not present in its active form in the extract that was used in our experiments. Given the extensive coupling between mRNA transcription and processing (61), it is also possible that weak TATA boxes could lead to less efficiently processed, and hence less stable, transcripts in vivo. Lastly, chromatin-based mechanisms of repression have been shown to suppress transcription initiation from weak promoter like elements (62,63). These processes would, of course, not be reflected in our nonchromatin transcription system in vitro.

The relative lack of sensitivity to nonconsensus TATA mutants in our extract and transcription assays in vivo is somewhat surprising given the greater sensitivity to TATA box mutations observed in some other yeast studies. Detailed analyses of the HIS3 TR TATA box in extracts (7) and in vivo (5,64) has revealed a much greater sensitivity to nonconsensus mutations in this promoter. However, this could be due to the fact that the native HIS3 TR TATA box (TATAAAGT) already has two nonconsensus bases in the seventh and eighth positions that may sensitize it to further nonconsensus mutations in the first six positions. The ability of the extract to facilitate higher levels of relative transcription than the purified system for nonconsensus TATA boxes suggests that there are additional factors that assist in transcription of weak TATA containing or TATA-less promoters. Fractionating these extracts and adding them back to our purified system is a promising approach for future attempts to identify these factors.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank Dr S. Hahn for the Flag-tagged TFIID strain and Dr P.A. Weil for TFIID antibodies. We also thank Benjamin Guidi for assistance in the preparation of
purified general transcription factors. Funding for this work was provided by an institutional research grant from the Hitchcock Foundation and by the NIH (GM62483). G.B. was supported in part by a Rosaline Borrisson Graduate Fellowship and the Fulbright Foundation.

Conflict of interest statement. None declared.

REFERENCES

interaction with the C-terminal repeat domain of RNA polymerase II. Cell, 77, 599–608.