MOLECULAR AND CELLULAR BIOLOGY, July 1993, p. 3841–3849
0270-7306/93/073841-09$02.00/0

Vol. 13, No. 7

Mechanism of Initiator-Mediated Transcription: Evidence for a Functional Interaction between the TATA-Binding Protein and DNA in the Absence of a Specific Recognition Sequence

BEATRICE ZENZIE-GREGORY, ANITA KHACHI, ISLA P. GARRAWAY, AND STEPHEN T. SMALE*

Howard Hughes Medical Institute, Molecular Biology Institute, and Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, California 90024-1662

Received 28 January 1993/Returned for modification 25 March 1993/Accepted 9 April 1993

Promoters containing Sp1 binding sites and an initiator element but lacking a TATA box direct high levels of accurate transcription initiation by using a mechanism that requires the TATA-binding protein (TBP). We have begun to address the role of TBP during transcription from Sp1-initiator promoters by varying the nucleotide sequence between −14 and −33 relative to the start site. With each of several promoters containing different upstream sequences, we detected accurate transcription both in vitro and in vivo, but the promoter strengths varied widely, particularly with the in vitro assay. The variable promoter activities correlated with, but were not proportional to, the abilities of the upstream sequences to function as TATA boxes, as assessed by multiple criteria. These results confirm that accurate transcription can proceed in the presence of an initiator, regardless of the sequence present in the −30 region. However, the results reveal a role for this upstream region, most consistent with a model in which initiator-mediated transcription requires binding of TBP to the upstream DNA in the absence of a specific recognition sequence. Moreover, in vivo it appears that the promoter strength is modulated less severely by altering the −30 sequence, consistent with a previous suggestion that TBP is not rate limiting in vivo for TATA-less promoters. Taken together, these results suggest that variations in the structure of a core promoter might alter the rate-limiting step for transcription initiation and thereby alter the potential modes of transcriptional regulation, without severely changing the pathway used to assemble a functional preinitiation complex.

A critical DNA sequence element in many genes transcribed by mammalian RNA polymerase II is a TATA box, which is located 25 to 30 nucleotides from the site of transcription initiation (reviewed in reference 17). By interacting with a protein called TATA-binding protein (TBP), the TATA box determines the location of the transcription start site and directs the formation of a functional preinitiation complex. It has been well documented that several genes do not contain obvious TATA boxes and that some of these TATA-less genes contain initiator (Inr) elements that overlap and specify the precise transcription start site (1, 2, 9, 13, 19). In several laboratories, proteins that recognize Inr elements have been identified (3, 13, 16, 18), and it appears that, both in vitro and in vivo, transcription of genes lacking TATA boxes proceeds through a rate-limiting step different from that used during transcription of genes containing TATA boxes (4, 24).

Transcription from TATA-less promoters is absolutely dependent on TBP (4, 6, 15, 20), but the role of TBP remains unknown. Possibly, TBP or a TBP-associated protein mediates communication between the upstream activator proteins and the general factors positioned at the start site, solely through protein-protein interactions and without contacting the promoter DNA. Alternatively, a TBP-associated protein might be important for Inr-mediated transcription, leading to only an indirect requirement for TBP. Another alternative is that TBP could be directed by proteins interacting with the Inr to bind to the −30 region in the complete absence of a functional TATA box. Finally, Wiley et al. (23) recently proposed that many TATA-less promoters contain functional, nonconsensus TBP binding sites, "with direct binding of hTFIIDτ [TBP] to the −30 regions of many TATA-less promoters probably also contributing to determining the initiation site."

Previously, we showed that synthetic promoters containing the terminal deoxynucleotidyl transferase (TdT) Inr direct accurate transcription in the absence of a TATA box (19). The TATA independence of this transcription was demonstrated by placing different sequences at the −30 region, most of which did not function as TATA boxes by themselves (14, 19, 20). In the study reported here, we have extended our previous results by varying more extensively and in a controlled manner the sequences between −14 and −33 in a promoter containing Sp1 sites and the TdT Inr. In these experiments, we found a high degree of variability in the promoter strengths, especially when using an in vitro transcription assay. Each promoter directed detectable transcription from the correct start site, but some promoters were very weak. One of the weaker promoters contained the −30 region from the authentic murine TdT gene (12). Our results suggest that during transcription initiation from Inr-containing promoters, TBP is directed to bind to the −30 region regardless of the sequence present in that region.

MATERIALS AND METHODS

Plasmid DNAs. DNAs used for these studies were derived from a plasmid described previously (plasmid III) (20), containing the simian virus 40 (SV40) 21-bp repeats inserted into the *Bgl*II site of the pSP72 vector (Promega). To generate the series III plasmids, oligonucleotides containing the eight sequences designated a through h shown in Fig. 1 and flanked with *Bgl*II (upstream) and *Sac*I (downstream)

* Corresponding author.

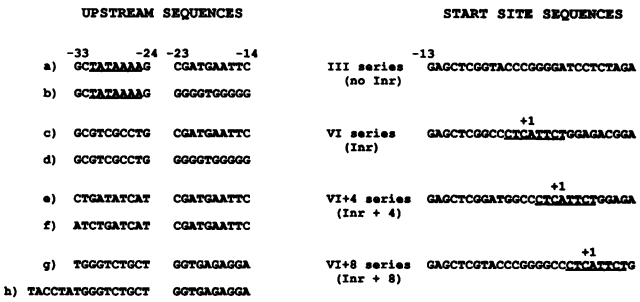

FIG. 1. Four series of plasmids containing synthetic promoters with variable sequences between −14 and −33. On the left, sequences a through h are the variable sequences that were placed between −14 and −33 in the four different series of plasmids. Sequences a and b contain consensus TATA boxes (underlined), c and d are highly G+C rich, e and f contain sequences with moderate A+T content, and g and h contain the upstream sequence found in the authentic murine TdT promoter. Sequence h is 6 bp longer than the other sequences so that the effect of the TAT found in the TdT promoter could be tested. On the right, the sequences which were placed downstream of sequences a through h are shown. The III series lacks a functional Inr element. The VI series contains the TdT Inr (underlined), with the start site located 26 bp downstream of the TATAAA in sequences a and b. The VI+4 and VI+8 series also contain the TdT Inr (underlined), but the Inr is shifted 4 and 8 bp, respectively, farther downstream relative to the TATAAA.

were inserted into plasmid III digested with *Bgl*II and *Sac*I. To generate the series VI plasmids, an oligonucleotide containing the TdT Inr, flanked by *Sac*I and *Bam*HI sites (Fig. 1) (19), was inserted into the series III plasmids digested with *Sac*I and *Bam*HI. For the series VI+4 and VI+8 plasmids, the oligonucleotides shown in Fig. 1 were inserted into the series III plasmids digested with *Sac*I and *Bam*HI (series VI+4) or *Sac*I and *Hin*dIII (series VI+8). All plasmids were purified with Qiagen columns. At least three independent preparations of each plasmid in series III and series VI were made and tested.

Plasmids for in vivo transfection experiments were prepared by cleaving the pSP72-based plasmids described above with *Hpa*I and *Bam*HI and inserting the promoter fragments into the *Cla*I (filled in with Klenow fragment) and *Bgl*II sites of plasmid pSVPyTK (19). This vector contains the herpes simplex virus thymidine kinase (HSV-TK) coding region as a reporter gene, an SV40 origin of replication, and a polyomavirus origin of replication and large-T-antigen gene.

In vitro transcription experiments. Nuclear extracts were prepared and in vitro transcription reactions were performed as described previously (14). Primer extension reactions used the SP6 promoter primer (Promega). Experiments yielding significant results were repeated with at least two different extract preparations.

In vivo transfection experiments. COS7 cells were grown in Dulbecco's modified Eagle's medium supplemented with 5% calf serum and were transfected with 2 μg of plasmid DNA by a DEAE-dextran method, as described previously (21). RNA was isolated by a Nonidet P-40 lysis method (21) and was analyzed by primer extension with the following oligonucleotide complementary to HSV-TK sequences: 5′-GGG GTACGAAGCCATACGCG-3′.

DNA-binding assays. Gel mobility shift assays were performed as described by Kao et al. (10) by using 3 μg of poly(dG-dC) nonspecific competitor DNA and a 4% native polyacrylamide gel in 0.5× Tris-borate-EDTA. The human TBP expressed in bacteria was partially purified by chromatography on DEAE-dextran and heparin Sepharose.

RESULTS

In vitro transcription of synthetic promoters containing variable sequences downstream of Sp1 sites, in the absence of an Inr. To carefully test the influence of the upstream region on transcription from an Sp1-Inr promoter, we prepared and analyzed four series of plasmids containing synthetic promoters. The first series of eight plasmids (III series) (Fig. 1) contains multiple Sp1 binding sites upstream of eight distinct nucleotide sequences of approximately 20 bp (Fig. 1), all in the absence of an Inr. Two of these plasmids (III-a and III-b) contain consensus TATA boxes (TATAAAA) beginning 12 bp downstream from the most proximal Sp1 site (−24 to −33 in Fig. 1). These two plasmids differ from each other only at the 10 nucleotides immediately downstream from the TATA box (−14 to −23 in Fig. 1), with plasmid III-a being A+T rich and plasmid III-b being G+C rich within this region. The next two plasmids (III-c and III-d) contain a highly G+C-rich sequence in place of the TATA box described above, 12 bp downstream from the Sp1 sites. Like the TATA-containing plasmids, these two plasmids differ from each other only at the 10 nucleotides downstream of this G+C-rich sequence. Next, plasmids III-e and III-f possess a sequence with a 30% G+C content in place of the TATA box. Plasmid III-e contains the sequence CTGATATCAT, which is somewhat related to a TATA box, but plasmid III-f contains the sequence ATCTGATCAT, which appears to be unrelated to any known functional TATA element. The final two plasmids, III-g and III-h, contain in place of the TATA box the actual sequence found in the −30 region of the authentic murine TdT gene (12). Plasmid III-h is slightly larger than the first seven plasmids, containing TdT sequences extending to −39 in the TdT promoter (rather than to −33 as in plasmid III-g). These extra nucleotides allowed us to include the sequence TAT, which is present in the TdT promoter between −33 and −35. Our previous studies (14) suggested that this location is too far from the start site for the TAT to function as a TATA box, but tests with plasmid III-h will allow us to confirm this hypothesis. Thus, with the exception of plasmid III-h, the eight plasmids described above are nearly identical, varying only at the 20 bp downstream from the Sp1 sites.

In vitro transcription analysis of plasmids III-a through III-h, all lacking Inr elements, revealed that the abilities of the different 20-bp sequences to function as TATA boxes were highly variable (Fig. 2A). As with all of the experiments described here, we relied on reproducibility of the results (see Materials and Methods and figure legends) rather than including an internal control, which prevents accurate interpretation of the results by greatly increasing the number of background bands in each experiment. (We have also observed competition between plasmids when using an internal control.) Plasmids III-a and III-b (Fig. 2A, lanes 1 and 2), containing consensus TATA elements, directed efficient initiation from a cluster of sites located about 25 to 30 bp from the TATA. The promoter in plasmid III-a was slightly stronger than that in plasmid III-b, suggesting that the sequence between −14 and −23 relative to the start site may have some influence on promoter strength. The remaining sequences revealed little or no ability to function as TATA boxes (Fig. 2A, lanes 3 to 8). No signals above the background level were detected with plasmid III-c, III-d, III-f, III-g, or III-h, and only a very weak signal was detected with

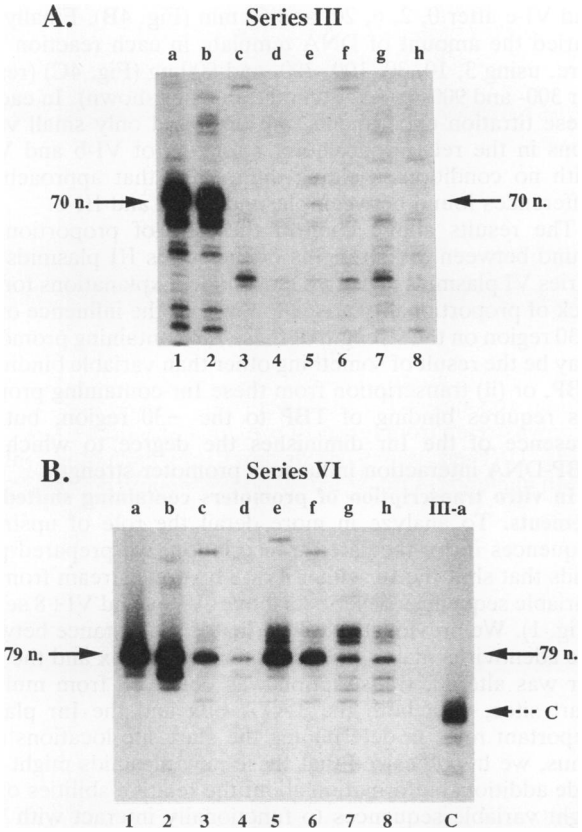

FIG. 2. In vitro transcription of plasmids containing variable upstream sequences, with or without an Inr element. In vitro transcription reactions were carried out with 300 ng of template DNA and 100 μg of HeLa cell nuclear extract. RNA transcripts were analyzed by primer extension with an SP6 promoter primer. Similar results were found with at least three different preparations of DNA and with at least two different extract preparations. (A) Series III plasmids. The arrows point to the location of the 70-nucleotide (n.) band expected if the sequence between −26 and −31 contains a functional TATA box. (B) Series VI plasmids. The control (lane C) was plasmid III-a, which reveals that this panel was derived from a shorter autoradiographic exposure than panel A. The arrows correspond to 79 nucleotides, the expected size of cDNA products resulting from transcripts that initiate at the correct start site within the TdT Inr. For both panels, lowercase alphabetical lane designations correspond to the upstream sequences shown in Fig. 1.

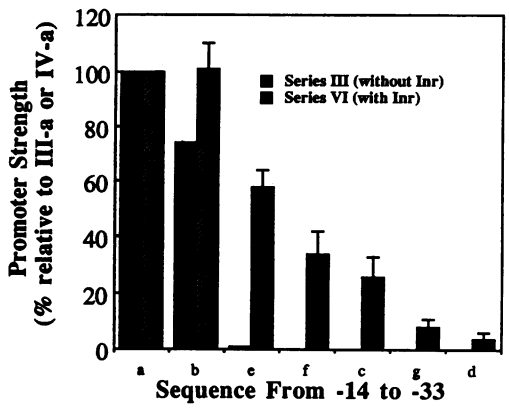

FIG. 3. Lack of proportional transcription between series III and series VI promoter strengths. The bands obtained from four independent experiments similar to the experiments described for Fig. 2 were quantitated by densitometry analysis, and the mean for each promoter was plotted as a percentage of the strength of the signal obtained with either plasmid III-a or plasmid VI-a. Error bars represent the standard deviation for each signal. It should be noted, however, that the purpose of the percentages is to reveal only that the strengths of the series III plasmids are not directly proportional to the strengths of the series VI plasmids. The accuracy of the numbers is limited by the accuracy of the densitometry analysis and by the fact that the autoradiographic exposures were performed in the presence of an intensifying screen. These limitations do not influence the results or the conclusions drawn from the data.

plasmid III-e (Fig. 2A, lane 5). This weak signal had an intensity similar to that of the widely heterogeneous background bands observed in every lane. However, because plasmid III-e contains the sequence at −30 with the most homology to an active TATA box, it is likely that this weak signal actually reflects a low level of specific transcription mediated by the ATAT functioning as a TATA box. This specific transcription with plasmid III-e was reduced by approximately 100-fold (as determined by densitometry analysis) relative to the transcription directed by plasmid III-a. Thus, these results reveal the relative abilities of the variable sequences to function as TATA boxes in the presence of Sp1 sites and in the absence of an Inr element.

In vitro transcription of synthetic Sp1-Inr promoters containing variable sequences between −14 and −33. Next, we inserted a fragment containing the TdT Inr into each of the eight plasmids described above, creating the plasmids referred to as series VI (Fig. 1). The insertion was designed so that in plasmids VI-a and VI-b (Fig. 1), the start site directed by the Inr was located at the same distance relative to the TATA box as it is in the adenovirus major late promoter. In vitro transcription with these promoters again revealed highly variable promoter strengths (Fig. 2B). The TATA-containing promoters (VI-a and VI-b) (Fig. 2B, lanes 1 and 2) directed the highest levels of transcription, with no significant differences between them. The strengths of the promoters in plasmids VI-c, VI-d, VI-e, and VI-f were reduced to varying degrees (Fig. 3). In addition, the strengths of the VI-g and VI-h promoters were similar to each other and reduced in strength by approximately 20-fold relative to VI-a. (With some plasmids, additional bands were observed in the vicinity of the correct start site. We do not know whether these bands represent specific transcription start sites or whether they are artifacts of the primer extension reaction.)

These results reveal the strong degree to which the sequence between −14 and −33 can influence Inr-mediated transcription, even if the sequence does not contain a functional TATA box. In general, with increasing G+C content of the upstream DNA, the promoter strengths decreased. The simplest explanation for these results is that TBP must bind to the upstream DNA during Inr-mediated transcription and that the variable promoter activities reflect the variable affinities of the TBP-DNA interaction. Given this possible explanation, we might have expected to find that the activity of an Inr-containing promoter would be completely dependent on the strength of the TATA box and that the Inr would simply enhance the strength of the promoter in direct proportion to the strength of the TATA box. However, this is clearly not observed, as the graph in Fig. 3 reveals that the signals obtained with the series III plasmids were not directly proportional to those obtained with the series VI plasmids. For example, the ratio of the strength of III-a to that of III-e is 100:1, but the ratio of the

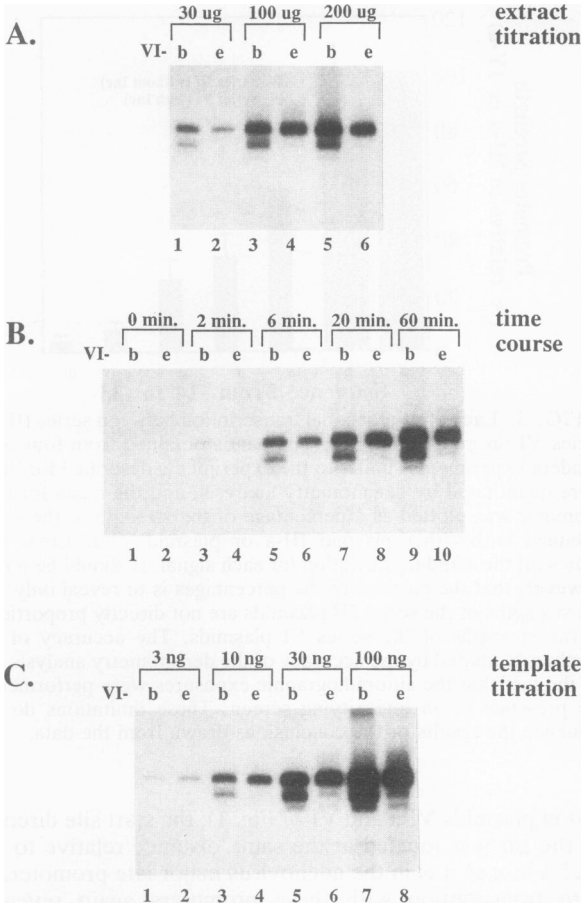

FIG. 4. Titration experiments with plasmids VI-b (lanes b) and VI-e (lanes e). (A) An extract titration experiment was performed with 300 ng of plasmids VI-b and VI-e with 30, 100, or 200 μg of HeLa nuclear extract. (B) In vitro transcription reactions were performed with 100 μg of nuclear extract and 300 ng of plasmid VI-b or plasmid VI-e. Reactions were terminated after 0, 2, 6, 20, or 60 min. (C) In vitro transcription reactions were performed with plasmids VI-b and VI-e with 3, 10, 30, or 100 ng of template DNA.

strength of VI-a to that of VI-e is only about 2:1. Thus, because the promoter strength found in the presence of an Inr is not completely dependent on the strength of the TATA box, we cannot yet reach a conclusion as to whether the variable promoter strengths result from the variable abilities of TBP to bind to the upstream DNA.

Titrations of in vitro transcription reactions with plasmids VI-b and VI-e. One possible explanation for the lack of proportionality between the promoter strengths of the III series and VI series plasmids is that the reactions may have been performed under nonlinear conditions. For example, it is possible that use of appropriately linear conditions would reveal a large difference in promoter strengths between plasmids VI-b and VI-e, similar to the large difference observed between plasmids III-b and III-e. To address this possibility, we measured the relative signals obtained with plasmids VI-b and VI-e by using three separate titration experiments.

First, we varied the amount of nuclear extract in the reaction mixtures, using 30, 100, or 200 μg of protein (Fig. 4A). Next, we analyzed the time course of the transcription reaction, measuring the amount of RNA obtained with VI-b and VI-e after 0, 2, 6, 20, and 60 min (Fig. 4B). Finally, we varied the amount of DNA template in each reaction mixture, using 3, 10, 30, 100, 300, and 900 ng (Fig. 4C) (results for 300- and 900-ng reaction mixtures not shown). In each of these titration experiments, we observed only small variations in the relative promoter strengths of VI-b and VI-e, with no conditions yielding differences that approach the differences found between plasmids III-b and III-e.

The results above confirm the lack of proportionality found between the strengths of the series III plasmids and series VI plasmids. The two most likely explanations for this lack of proportionality are as follows: (i) the influence of the −30 region on the strength of these Inr-containing promoters may be the result of something other than variable binding of TBP, or (ii) transcription from these Inr-containing promoters requires binding of TBP to the −30 region, but the presence of the Inr diminishes the degree to which the TBP-DNA interaction influences promoter strength.

In vitro transcription of promoters containing shifted Inr elements. To analyze in more detail the role of upstream sequences in Inr-mediated transcription, we prepared plasmids that shift the Inr either 4 or 8 bp downstream from the variable sequences described above (VI+4 and VI+8 series) (Fig. 1). We previously showed that if the distance between the adenovirus major late promoter TATA box and the TdT Inr was altered, transcription was observed from multiple start sites, with both the TATA box and the Inr playing important roles in determining the start site locations (14). Thus, we hypothesized that these new plasmids might provide additional information about the relative abilities of the eight variable sequences to functionally interact with TBP and to act as TATA boxes.

The in vitro transcription results shown in Fig. 5 reveal a complex array of efficiencies and locations of transcription start sites. In Fig. 5A, showing results obtained with the VI+8 series (note that in this series, the sequences between −22 and −41 relative to the Inr start site are variable), the plasmids containing consensus TATA elements (Fig. 5A, lanes 1 and 2) directed efficient transcription from heterogeneous start sites that extend from the location dictated by the TATA box to the location dictated by the Inr. In both plasmids, about 30% of the transcription began at the TATA-dictated site (+1 in Fig. 5A), about 5% began at the Inr-dictated site (+8 in Fig. 5A), and the remaining 65% began at locations between those dictated by the TATA and the Inr. With the remaining six plasmids shown in Fig. 5A, the bulk of the transcription start sites were dictated by the Inr. With plasmid VI+8-e, a very weak band, similar to that observed with plasmid III-e, was observed 25 bp downstream of the ATAT, but much higher levels of transcription initiated within the Inr element. As found with the VI series plasmids, the efficiencies of initiation within the Inr were somewhat related to the G+C content found approximately 25 to 30 bp upstream from the start site. For example, plasmid VI+8-d directs the lowest level of specific initiation and contains the sequence GGGGTGGGG in the −30 region, rather than the sequence ATGAATT found in plasmids VI+8-a, -c, -e, and -f (Fig. 1). Among the plasmids VI+8-c through VI+8-h, various degrees of heterogeneity were observed in the selection of transcription start sites, most likely reflecting more minor differences in the upstream sequences (see Discussion). The primary result is that, consistent with the results obtained with the series III plasmids, the upstream sequences in plasmids VI+8-c through -h do not contain specific elements that strongly contribute to determining the transcription start site.

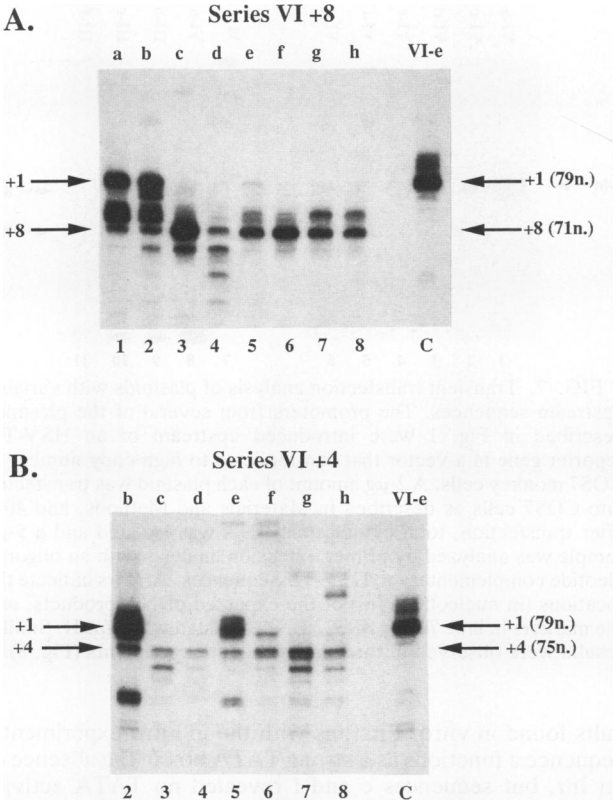

FIG. 5. In vitro transcription of plasmids containing Inr elements shifted 8 or 4 bp downstream from the variable sequences. The series VI+4 and VI+8 plasmids were created by inserting the appropriate start site oligonucleotide containing a shifted Inr (Fig. 1) downstream from the variable sequences a through h. In vitro transcription reactions were then performed with 100 μg of HeLa nuclear extract and 300 ng of plasmid DNA. The results of the primer extension analyses are shown. (A) Series VI+8 plasmids. Arrows point to the start site found with the series VI plasmids (+1, 79 nucleotides [n.]) or to the start site expected after the Inr was shifted 8 bp further downstream (+8, 71 n.). (B) Series VI+4 plasmids. Arrows point to the start site found with the series VI plasmids (+1, 79 n.) or to the start site expected after the Inr was shifted 4 bp further downstream (+4, 75 n.). We do not understand the significance of the upper band that is observed specifically with the VI+4-h plasmid but not with the VI+4-g plasmid. The location of this band suggests that it is activated by the TATGG sequence at −35, but it is not clear why it was not observed with plasmid VI-h or VI+8-h. For both panels, lane C shows a control reaction with plasmid VI-e, and lowercase lane letters correspond to the upstream sequences shown in Fig. 1.

In Fig. 5B, results from transcription experiments with the VI+4 series plasmids are shown. These results are similar to those found with the VI+8 series, in that in most plasmids lacking a strong TATA box, the bulk of transcription initiation sites shift downstream along with the Inr, in this case 4 bp downstream. However, one obvious exception is plasmid VI+4-e, which exhibits high levels of transcription, most of which begins 25 bp downstream of the ATAT rather than within the Inr. Taken together, the results found with plasmids III-e, VI-e, VI+4-e, and VI+8-e suggest that the ATAT sequence contains a significant affinity for TBP. The affinity is too weak to allow the ATAT sequence to function efficiently in plasmid III-e and to allow it to manipulate the start site in plasmid VI+8-e. However, in plasmid VI+4-e, in which the Inr is shifted downstream from the ATAT by only 4 bp, the close spacing appears to allow the ATAT to cooperate with the Inr.

A final important result revealed by the VI+4 series is that found with plasmid VI+4-f. With this plasmid, a small but significant amount of transcription initiated at the nucleotide dictated by the −30 sequence rather than at the nucleotide dictated by the shifted Inr. This suggests that the −30 sequence in plasmid VI+4-f (sequence f) possesses an ability to function as a TATA box that is greater than that of sequences c, d, g, and h but less than that of sequences a, b, and e.

Thus, the experiments described above appear to provide a very sensitive assay for assessing the ability of a particular sequence to function as a TATA box, and they allow us to order the TATA activities of the −30 sequences as follows: a and b > e > f > c, d, g, and h. Because this order closely matches the order of promoter strengths found with the series VI plasmids, it provides strong support for the idea that TBP contacts the upstream sequences during Inr-mediated transcription (see Discussion).

Relative affinities of the variable sequences for TBP. To further evaluate the abilities of the variable sequences to function as TATA boxes, we tested the relative affinities of these sequences for partially purified human TBP expressed in *Escherichia coli*. A gel mobility shift experiment with TBP and a TATA box oligonucleotide derived from the adenovirus E1B promoter revealed a specific protein-DNA complex (Fig. 6A, lanes 1 and 8). Competition experiments were then performed by using oligonucleotides containing six of the variable sequences. Very little competition was found with 0.2 or 2 ng of cold competitor derived from sequence a (Fig. 6A, lanes 9 and 10), but 20 or 200 ng strongly inhibited formation of the gel shift complex (Fig. 6A, lanes 2, 11, and 12). In contrast, 200 ng of competitors c, d, e, f, and g had no effect on the gel shift complex (Fig. 6A, lanes 3 to 7), revealing that the affinity of TBP for these sequences is significantly lower than its affinity for sequence a.

Further titrations of unlabeled competitors revealed that the affinities of TBP for these sequences were variable (Fig. 6B). Quantitation of duplicate titration experiments (with an Ambis phosphoimager) suggested that the TBP affinity for sequences e, f, c, g, and d is reduced by approximately 40-, 75-, 80-, 100-, and 400-fold, respectively, relative to the affinity of TBP for sequence a.

The results of the experiments described above are consistent with the relative abilities of the upstream sequences to function as TATA boxes, determined as shown in Fig. 3 and 5. For example, the results in Fig. 3 suggest that the affinity of TBP for sequence e is 100-fold lower than the affinity of TBP for sequence a, which is in agreement with the 40-fold difference found in this experiment. The degree to which these numbers differ from each other can easily be attributed to inaccuracies in the densitometry analysis used to quantitate the data in Fig. 2A and to the inaccuracies associated with measuring affinities by the gel shift technique. In addition, it should be noted that the gel shift experiments are measuring the combined affinities of TBP for the different A+T-rich regions found at both −30 and −20 in the variable sequences. The important result is that both the gel shift and the functional assays agree that the relative affinities of TBP for sequences a and e are very different from the small twofold difference in transcription signals detected between plasmid VI-a and plasmid VI-e. Our affinity measurements also agree with previous reports of the relative affinities of different sequences for TBP (8).

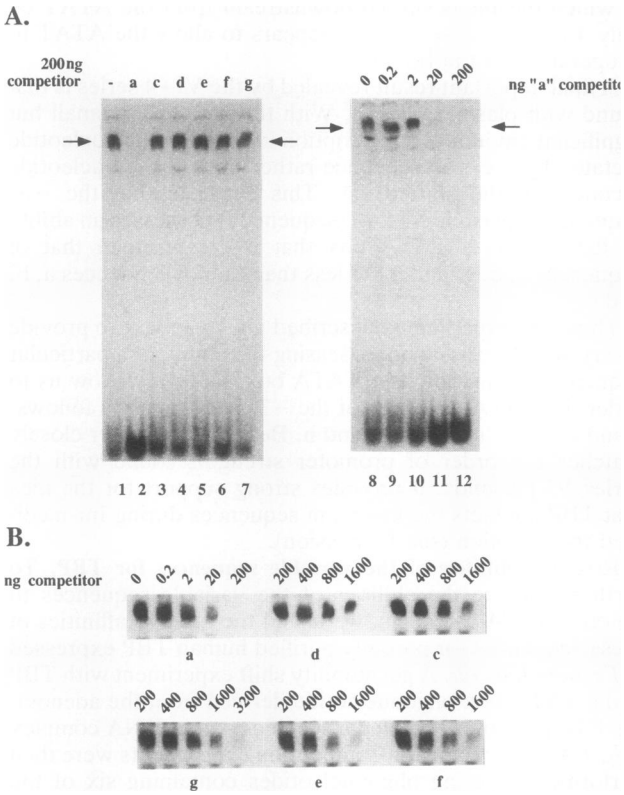

FIG. 6. Gel mobility shift analysis of the interaction of TBP with the variable upstream sequences. (A) Partially purified human TBP expressed in *E. coli* was used in a gel mobility shift assay with conditions as described by Kao et al. (10). Lanes 1 and 8 show the complex observed with a ^{32}P-labeled oligonucleotide derived from the adenovirus E1B TATA box (10). Lanes 2 to 7 show the results of competition experiments with 200 ng of oligonucleotides a, c, d, e, f, and g, respectively. Lanes 9 to 12 show the results of competition experiments with 0.2, 2, 20, and 200 ng, respectively, of competitor a. The arrows point to the specific protein-DNA complexes observed. (B) Gel shift experiments were performed with titrations of competitors a, c, d, e, f, and g. This experiment is representative of two independent experiments, which were quantitated with a phosphoimager (Ambis).

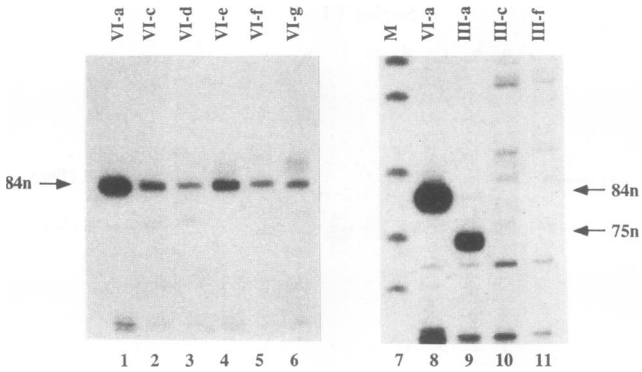

FIG. 7. Transient transfection analysis of plasmids with variable upstream sequences. The promoters from several of the plasmids described in Fig. 1 were introduced upstream of an HSV-TK reporter gene in a vector that can replicate to high copy number in COS7 monkey cells. A 2-μg amount of each plasmid was transfected into COS7 cells as described in Materials and Methods, and 40 h after transfection, total cytoplasmic RNA was isolated and a 5-μg sample was analyzed by primer extension analysis with an oligonucleotide complementary to HSV-TK sequences. Arrows indicate the locations (in nucleotides [n]) of the expected cDNA products, and the markers in lane 7 are pBR322 DNA digested with *Hpa*II. Similar results were observed in three independent experiments (Fig. 8).

Although a previous report (23) suggested that the affinity of TBP for the TdT upstream region was greater than that determined here, that study used a fragment from the TdT promoter that extended further upstream and included the A+T-rich sequence that we showed in Fig. 2 and 5 to have no influence on promoter activity.

Promoter activities detected with a transient transfection assay. To extend the in vitro analysis described above, we tested the relative promoter strengths by using a transient transfection assay. Several of the promoters from series III and series VI were introduced upstream of an HSV-TK reporter gene in a vector containing an SV40 origin of replication (see Materials and Methods). This vector replicates to high copy number when transfected into COS7 monkey cells, allowing us to analyze the RNA by primer extension with a primer complementary to HSV-TK sequences.

Figure 7 shows typical results from the in vivo analysis, and Fig. 8 depicts the relative promoter strengths derived from three independent experiments. These results reveal both similarities to and important differences from the results found in vitro. First, as with the in vitro experiments, sequence a functions as a strong TATA box in the absence of an Inr, but sequences c and f revealed no TATA activity (Fig. 7, lanes 9 to 11). Moreover, as shown with the in vitro experiments diagrammed in Fig. 3, the promoter strengths produced by these sequences in the absence of an Inr were not proportional to the promoter strengths produced in the presence of an Inr (Fig. 7, lanes 1 to 6 and Fig. 8; compare the relative strengths of III-a and VI-a with the relative strengths of III-f and VI-f).

Also consistent with the in vitro results were some of the relative promoter strengths found with the series VI plasmids. The VI-a promoter (Fig. 7, lane 1) was stronger than the VI-e promoter (Fig. 7, lane 4), which was stronger than the VI-c, -d, -f, and -g promoters (Fig. 7, lanes 2, 3, 5, and 6). An important difference between the in vitro and in vivo results, however, was that in vivo, the strengths of these last

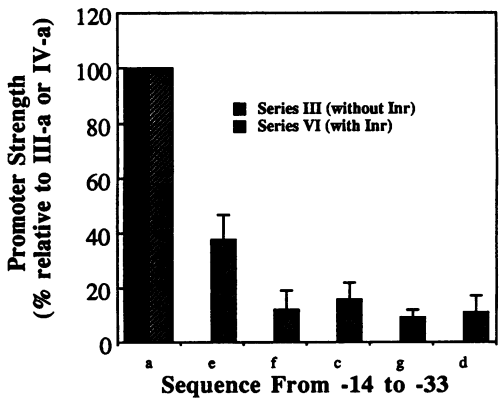

FIG. 8. The upstream sequence has little influence on Inr-mediated transcription in vivo. The results from three separate experiments similar to those shown in Fig. 7 were quantitated by densitometry analysis. Error bars refer to standard deviations. Plasmids III-d, III-e, and III-g were not tested.

four promoters were similar. Although slight differences are visible in Fig. 7, these differences were not reproducible or statistically significant. Thus, in the absence of a strong TATA box, the upstream sequence had much less influence on Inr-containing promoters in vivo than it did in vitro. This result is consistent with a recent report by Colgan and Manley (4), who showed that TBP (and presumably the TBP-DNA interaction) was not involved in determining the rate of transcription from Sp1-Inr promoters. Therefore, the in vitro analysis described above suggests that TBP contacts the −30 region during transcription initiation but that the Inr diminishes the degree to which the affinity of this interaction influences promoter strength. Figures 7 and 8 suggest that in vivo, with promoters that lack a strong TATA box, the Inr nearly eliminates the influence of the TBP-DNA affinity on promoter strength.

DISCUSSION

Our results address four issues related to transcription from promoters containing an Inr element. First, the data suggest that Inr-mediated transcription can proceed regardless of the sequence present in the −30 region. Second, the data provide strong indirect evidence supporting a model in which, during Inr-mediated transcription, TBP must contact the upstream DNA irrespective of the DNA sequence. Third, in vivo, the presence of an Inr strongly diminishes the degree to which promoter strength depends on the affinity of TBP for the upstream DNA. And finally, the −30 region of the TdT promoter cannot act as a TATA box in the absence of the Inr and plays no role in determining the location of the transcription start site in the presence of an Inr.

Inr-mediated transcription proceeds regardless of the upstream sequence. Taking together the data from our previous reports (14, 19, 20) and from the series VI, VI+4, and VI+8 plasmids, we have shown that accurate Inr-mediated transcription can proceed with well over a dozen different sequences placed in the −25 to −30 region, some of which have a very high G+C content. Although this is not an exhaustive analysis, it strongly suggests that any sequence will be compatible with Inr-mediated transcription. It is apparent, however, that depending on the upstream sequence, variable degrees of heterogeneity in the selection of start sites can be found. The most dramatic examples of this heterogeneity are found with plasmids like VI+8-a and VI+4-e, in which an offset strong or weak TATA element alters the locations of the transcription start sites, even though most of the transcription observed is clearly dependent on the presence of the Inr. In these instances, it appears that the preinitiation complex forming over the core promoter must make a choice between forming an optimal alignment relative to the Inr or an optimal alignment relative to the upstream sequences. With plasmid VI+4-e, for example, an optimal alignment with the Inr (leading to initiation from the A within the Inr) forces TBP to interact with an unfavorable upstream sequence. Apparently, with this plasmid, TBP prefers to interact with the GATATC upstream sequence, forcing the protein(s) at the Inr to interact with the DNA in a manner that is less than optimal (leading to initiation from the 3 nucleotides at the 5′ end of the Inr). Under these circumstances, however, the Inr still stimulates transcription, even though most of the transcription initiates slightly upstream from the Inr's preferred location. From this example, it follows that the more subtle heterogeneity found with other plasmids most likely results from similar influences of the upstream sequences.

TBP contacts the promoter during Inr-mediated transcription in a non-sequence-specific manner. An examination of our results strongly suggests that the variable promoter strengths observed in vitro as well as in vivo result largely from the variable affinities of TBP for the upstream sequences. On the basis of the gel shift data and the transcription data from the series III, VI+4, and VI+8 plasmids, the relative abilities of the upstream sequences to function as TATA boxes are ordered as follows: a and b > e > f > c, d, g, and h. This precisely follows the order of promoter strengths found in vitro with the series VI plasmids, implicating a TBP-DNA interaction in transcription from these plasmids. The most reasonable explanation is that TBP is directed by proteins interacting at the Inr to interact with the upstream DNA and that it must interact with the upstream DNA for the Inr-containing promoters to function. Because every upstream sequence tested is compatible with Inr-mediated transcription, it follows that TBP can be directed to bind with some affinity to virtually any sequence. It might not be surprising that TBP can interact with most nucleotide sequences, as it has been shown to interact with the minor groove of DNA (11, 22). Furthermore, as described above, there may be a few nucleotides of flexibility in the actual positioning of the TBP-DNA complex relative to the Inr, allowing TBP to choose the sequence for which it possesses the highest affinity within an approximately 5-bp window.

It remains formally possible that although TBP contacts this region, the TBP-DNA interaction is not absolutely essential for promoter function. In addition, it is possible that TBP does not contact the DNA during transcription from plasmid VI-d, which contains the weakest promoter in this series. However, these two possibilities seem unlikely in light of the considerable influences on promoter strength of the upstream sequences, which, with the exceptions of sequences a and b, appear to have extremely low affinities for TBP.

A final alternative to the model described above is that G+C-rich sequences reduce promoter strength by preventing bending or unwinding of the promoter DNA during initiation. However, sequences a, e, and f have identical G+C contents, yet they yield promoters with different strengths, which appear to reflect their relative affinities for TBP. In addition, it should be noted that these two models are not mutually exclusive. In fact, although our data strongly implicate TBP binding in Inr function, a secondary effect of G+C content could result in the variations observed between plasmids VI-c and VI-d, which vary only downstream of the TATA region between nucleotides −14 and −23. Indeed, TBP is not thought to interact this close to the transcription start site, suggesting that this region is important either for bending or melting or for interacting with another component of the general transcription machinery. An important role for the region between the TATA box and the Inr has been described previously by Conaway et al. (5), who suggested that the region was important for interacting with a component of the transcription factor IID (TFIID) complex, containing TBP and associated proteins (7, 25).

The affinity of the TBP-DNA interaction is relatively unimportant for promoter strength in vivo. Although the in vitro experiments revealed a strong difference in promoter strength depending on the upstream sequence, the influence of this sequence on transcription in the transient transfection assay was much less dramatic. This result is consistent with those obtained in the studies of Colgan and Manley (4), who demonstrated that overexpression of TBP had little influence in vivo on the strength of TATA-less promoters but strongly

increased the strength of TATA-containing promoters. Both our study and that of Colgan and Manley suggest that the TBP-DNA interaction can play a role in determining the initiation frequency from core promoters that depend primarily on a TATA box but not from core promoters that rely on an Inr element.

Colgan and Manley (4) proposed the intriguing possibility that these differences between TATA-containing and TATA-lacking promoters may be important for generating promoters that respond to different types of activators. Our results extend this proposal by suggesting that these changes could be carried out without severely altering the pathway used to assemble a functional preinitiation complex. After TBP is directed to bind to the upstream DNA in a TATA-less promoter, it would be poised to recruit the remaining general transcription factors into a functional preinitiation complex in a manner that closely resembles that used during preinitiation complex formation from a TATA-containing promoter (3, 17). Indeed, we previously showed that some of the later steps during transcription initiation, as measured by behavior in the presence of Sarkosyl or dinucleotide triphosphates, are very similar for promoters that contain and that lack TATA boxes (24).

The −30 region of the TdT promoter is not important for core promoter function. Our results are consistent with those obtained by Wiley et al. (23) in a previous study of the SV40 major late promoter, but the two sets of results lead to different conclusions. The previous study focused on mutating a critical A+T-rich sequence element in the −30 region of the SV40 major late promoter. The results clearly demonstrated that the SV40 major late promoter actually contains a weak TATA box. From these results, Wiley et al. suggested that, in general, TATA-less promoters contain sequence elements at −30 that contain a specific affinity for TBP and that contribute to determining the location of the transcription start site. In contrast, however, we have demonstrated that the G+C-rich −30 sequence of the TdT promoter plays no role in promoter strength or in determining the location of the transcription start site and that any, or almost any, −30 sequence is compatible with accurate Inr-mediated transcription. These conclusions are consistent with our previous mutational analyses of the intact TdT promoter, which failed to reveal a strong influence of the −30 region on promoter strength (12, 19). Therefore, our results suggest a model that is very different from that proposed by Wiley et al. (23) in that we found that it is not necessary for the −30 region of TATA-less promoters to contain a specific binding site for TBP. The reason for this discrepancy is that the SV40 major late promoter studied by Wiley et al. (23) actually contains a weak TATA box, whereas the TdT promoter does not.

One important question that remains is whether the natural TdT −30 sequence is important to the overall structure of the TdT promoter. Possibly, the unusually low affinity of the TdT −30 sequence for TBP is needed for appropriate temporal and tissue-specific regulation of TdT expression. In one scenario, the promoter may rely on some unknown mechanism to overcome the low affinity during appropriate stages of lymphoid differentiation. Alternatively, appropriate regulation may rely on the rate-limiting step found during transcription from TATA-less promoters, which is different from that found with TATA-containing promoters (4, 24). Finally, we previously reported that a sequence element at the −30 region of the TdT promoter is highly homologous to a sequence element at the −30 region of the promoter for a gene called λ5, whose expression is also restricted to early lymphocytes (12). Because it appears that TBP must bind to this region during transcription initiation, this conserved sequence element may be an ideal target for regulatory interactions with a repressor protein that blocks TBP binding or with an activator protein that stabilizes TBP binding.

ACKNOWLEDGMENTS

We thank Arnold Berk, Qiang Zhou, and Paul Lieberman for helpful discussions and for critically reading the manuscript and Anne O'Shea-Greenfield and Joyce Soong for plasmid DNAs.

This work was supported by NIH grant DK43726 and by the Howard Hughes Medical Institute.

REFERENCES

1. **Ayer, D. E., and W. S. Dynan.** 1988. Simian virus 40 major late promoter: a novel tripartite structure that includes intragenic sequences. Mol. Cell. Biol. **8**:2021–2033.
2. **Beaupain, D., J. F. Eleouet, and P. H. Romeo.** 1990. Initiation of transcription of the erythroid promoter of the porphobilinogen deaminase gene is regulated by a cis-acting sequence around the cap site. Nucleic Acids Res. **18**:6509–6515.
3. **Carcamo, J., L. Buckbinder, and D. Reinberg.** 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc. Natl. Acad. Sci. USA **88**:8052–8056.
4. **Colgan, J., and J. L. Manley.** 1992. TFIID can be rate limiting in vivo for TATA-containing, but not TATA-lacking, RNA polymerase II promoters. Genes Dev. **6**:304–315.
5. **Conaway, J. W., E. Travis, and R. C. Conaway.** 1990. Transcription initiated by RNA polymerase II and purified transcription factors from liver: a complex set of promoter sequences governs formation of the initial complex. J. Biol. Chem. **265**:7564–7569.
6. **Cormack, B. P., and K. Struhl.** 1992. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell **69**:685–696.
7. **Dynlacht, B. D., T. Hoey, and R. Tjian.** 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell **66**:563–576.
8. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA **86**:5718–5722.
9. **Hariharan, N., and R. P. Perry.** 1990. Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. Proc. Natl. Acad. Sci. USA **87**:1526–1530.
10. **Kao, C. C., P. M. Lieberman, M. C. Schmidt, Q. Zhou, R. Pei, and A. J. Berk.** 1990. Cloning of a transcriptionally active human TATA binding factor. Science **248**:1646–1650.
11. **Lee, D. K., M. Horikoshi, and R. G. Roeder.** 1991. Interaction of TFIID in the minor groove of the TATA element. Cell **67**:1241–1250.
12. **Lo, K., N. R. Landau, and S. T. Smale.** 1991. LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. Mol. Cell. Biol. **11**:5229–5243.
13. **Means, A. L., and P. J. Farnham.** 1990. Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell. Biol. **10**:653–661.
14. **O'Shea-Greenfield, A., and S. T. Smale.** 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J. Biol. Chem. **267**:1391–1402.
15. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.
16. **Roy, A. L., M. Meisterernst, P. Pognonec, and R. G. Roeder.** 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature (London) **354**:245–248.
17. **Sawadogo, M., and A. Sentenac.** 1990. RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. **59**:711–754.
18. **Seto, E., Y. Shi, and T. Shenk.** 1991. YY1 is an initiator

sequence-binding protein that directs and activates transcription in vitro. Nature (London) **354:**241–245.

19. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57:**103–113.
20. **Smale, S. T., M. C. Schmidt, A. J. Berk, and D. Baltimore.** 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA **87:**4509–4513.
21. **Smale, S. T., and R. Tjian.** 1985. Transcription of herpes simplex virus *tk* sequences under the control of wild-type and mutant human RNA polymerase I promoters. Mol. Cell. Biol. **5:**352–362.
22. **Starr, D. B., and D. K. Hawley.** 1991. TFIID binds in the minor groove of the TATA box. Cell **67:**1231–1240.
23. **Wiley, S. R., R. J. Kraus, and J. E. Mertz.** 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the −30 region of TATA-less promoters. Proc. Natl. Acad. Sci. USA **89:**5814–5818.
24. **Zenzie-Gregory, B., A. O'Shea-Greenfield, and S. T. Smale.** 1992. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. J. Biol. Chem. **267:**2823–2830.
25. **Zhou, Q., P. M. Lieberman, T. G. Boyer, and A. J. Berk.** 1992. Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes Dev. **6:**1964–1974.