MINIREVIEW

Site-Specific Initiation of Transcription by RNA Polymerase II (43583)

RICHARD KOLLMAR AND PEGGY J. FARNHAM[1]

Program in Cell and Molecular Biology and McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Abstract. **RNA polymerase II initiates transcription at specific DNA sequences. Studies using sequence analysis and molecular genetics suggest a simple and universal model of start-site selection by RNA polymerase II. Two consensus sequences occur at fixed positions in promoters from higher eukaryotes and their viruses: the TATA box around −30 and the initiator at the start site of transcription. Both consensus sequences function as positioning elements that control site-specific initiation. As a first step during initiation, the basal transcription factor TFIID binds to the TATA box; regulatory transcription factors can tether TFIID to promoters without a consensus TATA box. TFIID then directs the assembly of other basal transcription factors and RNA polymerase II into a preinitiation complex. Finally, RNA polymerase II searches for the best match to the initiator consensus about 30 base pairs downstream of the TATA box to select the exact start site. The transcriptional activity of a start-site sequence generally correlates with its similarity to the initiator consensus, suggesting that there is only one type of initiator.** [P.S.E.B.M. 1993, Vol 203]

All RNA polymerases in all organisms initiate transcription at specific sites (1–11). A major goal of molecular biologists is to understand how the cell achieves such selectivity and accuracy in gene expression. Here, we will review how RNA polymerase II (DNA-directed RNA polymerase, EC 2.7.7.6) initiates transcription at specific sites in mammals. This enzyme transcribes the genes that code for proteins in eukaryotes. When DNA is added to a crude cell extract, RNA polymerase II starts transcription at the same sites as in the cell (12). This demonstrates that start sites are selected through specific interactions among proteins and DNA. We will first describe the proteins that compose the transcriptional machinery, RNA polymerase II and its transcription factors. Next, we will catalog the DNA sequences that control start-site selection. Such positioning elements have been identified and characterized by sequence analysis and by molecular genetics; we will examine how closely the structure of these elements correlates with their function. We will then present a simple model of site-specific initiation by RNA polymerase II. Finally, we will discuss three aspects of start-site selection that are not yet fully understood: (i) the universal role of the transcription factor TFIID, (ii) the ability of regulatory transcription factors to direct site-specific initiation, and (iii) the function of the initiator, a positioning element that includes the start site of transcription.

The Transcriptional Machinery

Three types of proteins take part in the initiation of messenger RNA synthesis: RNA polymerase II, basal transcription factors, and regulatory transcription factors. The properties of these proteins have been reviewed (8, 13–18) and are summarized here. RNA polymerase II transcribes the genes that code for proteins and most small nuclear RNAs in eukaryotes. The purified enzyme binds to DNA and initiates preferen-

[1] To whom requests for reprints should be addressed at McArdle Laboratory for Cancer Research, University of Wisconsin, 1400 University Avenue, Madison, WI 53706.

0037-9727/93/2032-0127$3.00/0

tially at single-stranded nicks, gaps, and ends. On closed circular templates, the purified enzyme starts at specific sites, but not always at the same sites that it uses in the cell (19). Thus, RNA polymerase II cannot initiate transcription selectively and accurately by itself, perhaps because a strong preference for specific nucleotide sequences would reduce its fidelity and processivity. Site-specific initiation requires basal transcription factors in addition to RNA polymerase II. To date, seven such proteins have been described: TFIIA, -B, -D, -E, -F, -H, and -J. The basal factors have been purified from yeast, fruit fly, rat, and human; the genes for TFIIA, -B, -D, -E, and -F have been cloned. Only TFIID binds to a specific DNA sequence, the TATA box (see below). Regulatory transcription factors control the rate of initiation at specific promoters in different tissues or at different stages of cell growth and development. Many of these factors bind to a specific DNA sequence. They activate or repress transcription by changing the chromatin structure (20) or by interacting with RNA polymerase II and the basal factors.

Before RNA synthesis begins, RNA polymerase II and basal factors bind to a promoter in an ordered series of steps and form a preinitiation complex at the start site (17, 21). The major late promoter of human adenovirus 2 has served as a prototype to study this process. First, TFIID binds to the TATA box upstream of the start site. Next, TFIIA and -B associate with TFIID at the promoter. RNA polymerase and TFIIF together bind to this initial complex. Binding of TFIIE, -H, and -J completes the assembly of the preinitiation complex. Only then can RNA polymerase II begin to synthesize RNA. Initiation at other promoters follows the same principles: Site-specific initiation by RNA polymerase II requires a common set of basal transcription factors; some of these factors bind to the promoter before RNA polymerase II and serve as adapters between the DNA and the enzyme.

Positioning Elements

During assembly of the preinitiation complex, components of the transcriptional machinery bind to specific DNA sequences. Some of these sequences control the position of the preinitiation complex and thus the start site of transcription; we will call them positioning elements. (Positions in promoters are given as distances in base pairs from the start site at +1. A positive sign denotes a downstream position, in the direction of transcription; a negative sign denotes an upstream position. There is no Position 0.) In the next three sections, we will describe how two complementary approaches, sequence analysis and molecular genetics, have been used to identify and characterize three positioning elements in mammalian promoters: the TATA box, the proximal sequence element, and the initiator.

Sequence Analysis

Sequence analysis is used to identify consensus sequences that occur in many promoters at similar positions relative to the start site and that could, therefore, be positioning elements. The positioning function of these sequences is then confirmed by molecular genetics (see below).

Sequence analysis requires a data base of promoters and an algorithm to examine their sequences. The Eukaryotic Promoter Database (22) now has 502 entries whose start sites have been mapped precisely by RNA sequencing, nuclease protection, or primer extension. This data base contains only sequences that are transcribed by RNA polymerase II in a higher eukaryote, which excludes phycophyta, fungi, myxomycetes, and protozoa. The basic algorithm to identify positioning elements aligns promoter sequences at their start sites and generates a frequency table that shows how often each base occurs at each position; a group of adjacent, conserved nucleotides indicates a consensus (23–25). This algorithm has two limitations: If a start site has not been mapped accurately, the promoter will not be aligned accurately. If an element is conserved, but its position varies among promoters, the element may not be detected. An improved form of the basic algorithm tolerates mapping errors and varying positions (26). The basic algorithm has also been modified for refining the consensus of a known positioning element (27, 28). Promoter sequences are aligned to the given consensus, and a new consensus is derived from the best matches; these steps are repeated until the consensus stabilizes. Once a consensus has been found, the similarity of other sequences to this consensus can be calculated from its frequency table (28). Unlike a nucleotide-by-nucleotide comparison with the explicit consensus, this calculation provides a quantitative measure of similarity among sequences.

So far, three positioning elements have been identified by sequence analysis (Table I). First, Goldberg and Hogness (23) found an AT-rich consensus when they compared only 16 promoters, mostly of histone genes from the fruit fly and sea urchin. The so-called TATA box is accentuated by a higher GC content in promoters than in the genome (29). To refine the TATA box consensus, Bucher (28) has analyzed the sequences in the Eukaryotic Promoter Database; sequences with a significant similarity (greater than 79%) to the consensus occur around position −30 in about 78% of the entries. The median similarity to the TATA box consensus of all possible 15-nucleotide sequences is about 44%, and only 1 in 309 sequences has a similarity of 79% or more (data not shown). Second, a consensus called the proximal sequence element occurs between about −50 and −60 in the promoters of 39 vertebrate small nuclear RNA genes (30). Third, transcription

Table I. Consensus Sequences and Examples of Positioning Elements in Eukaryotic Promoters

Promoter[a] (Ref.)	Sequence[b]	Similarity to the consensus[c] (%)
TATA box	STATAWAWRSSSSSS[d]	
Ad2 major late (32, 33)	−32 CTATAAAAGGGGGTG... −12	98
SV40 early (34–37)	...TTATTTATGCAGAGG...	85
Human β-globin (38)	...GCATAAAAGTCAGGG...	91
Proximal sequence element	STSDCCNTRNS[e]	
Toad U2 snRNA (39)	−61 CTCTCCCCATG −51	
Human U1 snRNA (40, 41)	−80 ...GTGACCGTGTGTG −50	
Initiator	KCABHYBY[f]	
Ad2 major late (43)	−6 GTCCTCACTCTCTTCCG +11	97
Mouse TdT (43, 44)	−3 CTCATTCT +5	95
Human PBGD (45)	−2 TCAGTGTCCTGG +10	95
Ad2 IVa2 (46–48)	−6 CGTCTCAGAGTGGTCCG +11	89
AAV2 P5 (49)	−7 G GTCTCCATTTTGAAGCG +11	88
Mouse DHFR (50, 51)	−11 ATTTCGCGCCAAACTTGACG +9	84

[a] Abbreviations: Ad2, human adenovirus 2; SV40, simian virus 40; snRNA, small nuclear RNA; TdT, terminal deoxynucleotidyltransferase; PBGD, porphobilinogen deaminase; AAV2, adeno-associated virus type 2; DHFR, dihydrofolate reductase.
[b] Sequence on the nontemplate strand from 5′ to 3′. Promoter elements are aligned at their best match to the consensus. B, C or G or T; D, A or G or T; H, A or C or T; K, G or T; N, any base; R, A or G; S, C or G; W, A or T; Y, C or T; the numbers refer to the boundaries of the promoter elements as defined by molecular genetics; ellipsis (...) indicates that the boundary is not known or not shown.
[c] Similarity to the consensus was calculated according to Ref. 28; a similarity greater than 79% (TATA box) or 81% (initiator) is considered significant.
[d] Consensus (Ref. 28): the distance from the underlined T in the consensus to the start site is 30 ± 2 base pairs in 60% of the entries in the Eukaryotic Promoter Database. Transcription starts 30 base pairs downstream of the underlined bases.
[e] Consensus (Ref. 30): the 3′-end lies between −53 and −50.
[f] Consensus (Ref. 28): start sites are underlined.

frequently starts at an A within several pyrimidine nucleotides (24). This consensus, which includes the start site, is called the initiator (31). Sequences with a significant similarity (greater than 81%) to a refined initiator consensus occur at the start site in about 60% of the entries in the Eukaryotic Promoter Database (28). The median similarity to the initiator consensus of all possible eight-nucleotide sequences is about 41%, and only 1 in 27 sequences has a similarity of 81% or more (data not shown).

The results of sequence analysis thus suggest that most, but not all, promoters of higher eukaryotes contain one or two common sequence elements at similar positions: a TATA box around −30 and an initiator at the start site. Less common elements like the proximal sequence element may be found by analyzing other select groups of promoters.

Molecular Genetics

Molecular genetics is used to identify positioning elements in individual promoters and to test the positioning function of consensus elements that have been identified by sequence analysis. Molecular genetics examines the characteristic effects of several types of mutations on site-specific initiation: Placing a positioning element into a different sequence context should result in initiation at a new site, but at the same distance from the element. Changing the nucleotide sequence of a positioning element should change the start-site pattern. Inserting or deleting sequences downstream of a positioning element should lengthen or shorten the transcript by the same number of nucleotides. Inserting or deleting sequences upstream of a positioning element should not change the length of the transcript.

The TATA box was the first positioning element whose function was confirmed by molecular genetics (Table I). The major late promoter of adenovirus 2 contains a sequence around −30 that is very similar to the TATA box consensus. When sequences of this promoter between −32 and −12 are inserted into a plasmid, transcription starts in the plasmid sequence 30 base pairs downstream of the TATA box (32). Changing the A at −28 to a C reduces transcription and causes initiation at additional start sites around +1 (33); this point mutation reduces the similarity to the TATA box consensus from 98% to 84%. The early promoter of simian virus 40 also contains a sequence similar to the TATA consensus that directs initiation about 30 base pairs downstream (34). When sequences immediately downstream of this TATA box are deleted, transcription starts at new sites about 30 base pairs downstream of the TATA box (35–37); a deletion immediately upstream of the TATA box does not change the start site (37). A single point mutation in a TATA box

can lead to a disease: In a patient with β-thalassemia and low levels of β-globin messenger RNA, the TATA box in both copies of the β-globin gene contained a C instead of an A at −28 (38); this substitution reduces the similarity to the TATA box consensus from 91% to 79%. These and many other studies show that the TATA box is a positioning element that directs site-specific initiation about 30 base pairs downstream.

The promoters of vertebrate small nuclear RNA genes that are transcribed by RNA polymerase II lack a TATA box but contain a proximal sequence element (Table I). The proximal sequence element in the promoter of the toad U2 small nuclear RNA gene (39) and the region between −80 and −50 in the promoter of the human U1 small nuclear RNA gene, which includes a proximal sequence element (40, 41), are necessary for site-specific initiation. These results suggest that the proximal sequence element substitutes for the missing TATA box in promoters of small nuclear RNA genes and directs site-specific initiation about 50 base pairs downstream. Surprisingly, promoters that contain both a proximal sequence element and a TATA box are transcribed by RNA polymerase III; it seems that the TATA box and the proximal sequence element play a role in selecting both the correct RNA polymerase and the correct start site (42).

The Initiator(s)

In several promoters, sequences that direct site-specific initiation have been found at the start site (Table I): between −6 and +11 in the major late promoter of adenovirus 2 (43), between −3 and +5 in the promoter of the mouse terminal deoxynucleotidyltransferase (DNA nucleotidylexotransferase) gene (43, 44), between −2 and +10 in the erythroid-specific promoter of the human porphobilinogen deaminase gene (45), between −6 and +11 in the IVa2 promoter of adenovirus 2 (46–48), between −7 and +11 in the P5 promoter of adeno-associated virus type 2 (49), and between −11 and +9 in the promoter of the mouse dihydrofolate reductase gene (50, 51). Using the similarity to the initiator consensus as a quantitative measure, we found that the start sites of all six promoters closely match the initiator consensus. This finding suggested that all six promoters contain an initiator and that the initiator functions as a positioning element.

To examine the relationship among these start-site sequences and the initiator consensus further, we compiled for each sequence a list of point mutations whose effects on transcriptional activity and start-site pattern have been published (Table II). We calculated the similarity to the initiator consensus for each mutant and compared our results with the experimental data. Mutations that do not change the similarity to the initiator consensus do not affect the activity or the start-site pattern. Mutations that reduce the similarity to the consensus also reduce the activity. Mutations that introduce better matches to the consensus at positions other than +1 change the start-site pattern. For example, in mutant GCDI of the dihydrofolate reductase promoter, plasmid sequences next to −11 fortuitously create a sequence with higher similarity than the wild-type start site; the major start site in this mutant is −11 (50). We also examined a start site that bears no apparent similarity to the initiator consensus: At the promoter of the small ribosomal protein 16 gene from the mouse, transcription starts within a stretch of 12 pyrimidines (52) that matches the initiator consensus poorly (Table II). To date, there is no evidence that this stretch of pyrimidines functions as a positioning element by itself. Mutations that reduce the similarity to the initiator consensus (Table II, CM.5 to 7) reduce the promoter activity *in vivo* (53); mutations that increase the similarity (CM.2 to 4) increase the promoter activity *in vitro* (54). Similar results have been obtained for the start site of the major late promoter of simian virus 40 (J. Mertz, personal communication, October 1992), which is only 65% similar to the initiator consensus. Finally, the sequence around the start site of some, but not all, promoters functions as an initiator when placed downstream of binding sites for the transcription factor Sp1 (47). We found a close correlation between the transcriptional activity of these sequences and the similarity of their start site to the initiator box consensus (Fig. 1A). These findings suggest that the function of any sequence as a start site depends on its similarity to the initiator consensus.

We noticed two types of discrepancies between predicted and observed initiator functions. First, we could predict new start sites downstream of +1 in mutants of the promoter of the small ribosomal protein 16 gene (Table II), but not to the exact nucleotide. The *in vivo* start sites reported for these mutants (53) also differ slightly from the *in vitro* start sites (54). Similarly, the reported start sites of the entries in the Eukaryotic Promoter Database fall anywhere between positions +1 and +6 of the consensus (28). Such small discrepancies may arise because the methods used to map start sites are inherently inaccurate (28). Second, we predicted full activity for mutants 2A2, 2A3, and 2A4 of the porphobilinogen deaminase promoter (Table II), but they are inactive (45). They contain mutations only downstream of the start site, between +5 and +10, that abolish binding of a human nuclear protein (45). These mutants are probably inactive because the binding site for an essential activator protein has been altered. Similarly, transcription from the promoter of human immunodeficiency virus 1 starts at a poor match to the initiator consensus at +1 (48% similarity) rather than at a better match at +5 (81% similarity) (47, 55). A cellular transcription factor, LBP-1, binds to this promoter between −16 and +27 and contacts the DNA

Table II. Similarity to the Consensus and Transcriptional Activity of Mutant Initiators

Mutant[a] (Ref.)	Sequence[b]	Similarity to the consensus[c] (%)			Transcriptional activity[d]
Ad2 major late	GTCCTCACTCTCTTCCG		97		
AT+1 (33)	GTCCTCtCTCTCTTCCG	36	84	36	+
Mutant (46)	tTCgagtCTCcCTTCCG		51	82	+
Mouse TdT	CCCTCATTCTGG		95		
C4 (44)	CCCTCAgTCTGG		97		++
−5 (43)	tCCTCATTCTGG		95		++
−4 (43)	CtCTCATTCTGG		95		++
+7 (43)	CCCTCATTCTGt		95		++
F2 (44)	ataTCATTCTGG		95		++
−3, −2 (43)	CCgaCATTCTGG		90		+
−2 (43)	CCCaCATTCTGG	86	90		++
F1 (44)	ataTCATcgaGG		84		++
F3 (44)	CCCTCATcgaGG		84		++
C2 (44)	CCCTgATTCTGG	75	69		+
C3 (44)	CCCTCgTTCTGG		70		+
C1 (44)	CCCTgggTCTGG		[45]		−
Human PBGD	TCAGTGTCCTGG		95		
2A4 (45)	TCAGTGTCCTtc		[95]		−
2B5 (45)	TCAcTGTCCTGG		93		++
2B2 (45)	gCAGTGTCCTGG		91		++
2A2 (45)	TCAGTGgagTGG		[90]		−
2A3 (45)	TCAGTGgaCTGG		[90]		−
2B4 (45)	TCtGTGTCCTGG		[81]		−
2B3 (45)	TgAGTGTCCTGG		[69]		−
Ad2 IVa2	GTCTCAGAGTGGTCCG		89		
Mutant (46)	tTgagtGAGcGGTCCG		[43]		−
+2/−8 (48)	cTCgaccAGTGGTCCG		[32]		−
AAV2 P5	GGTCTCCATTTTGAAGCG		88		
pP5+1 mt (49)	GGTCTaaAgTTgGAAGCG		[62]		−
Mouse DHFR	CGATTTCGCGCCAAACTTGACGGCA		84		
GCDI (50)	tcATTTCGCGCCAAACTTGACGGCA	92	84		+
−1, −2 (51)	CGATTTCGCGCagAACTTGACGGCA	89	[32]		++
A (51)	CGgccctatattcAAaTccAgtaat		[34]	85	++
−5, −9, −15 (51)	CGATgTCGtGCCAAACTTGACGGCA		84		++
−5, −9, −12 (51)	CtATaTCGtGCCAAACTTGACGGCA		84		++
+5, +6 (51)	CGATTTCGCGCCAAACTgtACGGCA		80	74	++
−2, −6 (51)	CGATTTCtCGCaAAACTTGACGGCA	82	[58]		++
−2 (51)	CGATTTCGCGCgAAACTTGACGGCA		[58]		−
Mouse rpS16 (52)	...CTTCCCTTTTCC...		71		
CM.3 (53, 54)	...CTTCaCTTTTCC...		[96]	75	++/+++[e]
CM.2 (53, 54)	...CTTCCCaTTTCC...		[67]	89	++/+++
CM.4 (53, 54)	...CTgCCaTTgTCC...		[60]	88	++/+++
CM.1 (53, 54)	...CTaCCCTTTTCC...		[65]	75	++/++
CM.6 (53, 54)	...gTgCgagTgTaC...		[37]	60	+/++
CM.5 (53, 54)	...CTTCgagTgaCC...		[37]	59	+/++
CM.7 (53, 54)	...gaagggaaaagg...		[28]	52	+/++
Trout TPG-3 (108)	...ACATTTTATCCATCAAT...	87	66	83	

[a] See Footnote *a* to Table I for abbreviations; additional abbreviations: rpS16, small ribosomal protein 16; TPG-3, trout protamine gene 3. Mutant designations are taken from the references.

[b] Sequences on the nontemplate strand from 5′ to 3′ aligned at the start site of the wildtype. Wildtype sequences are from Table I. Lowercase letters indicate point mutations; underlined letters are the start sites; an ellipsis (...) shows that the boundary ot the initiator is not known. If the position of a start site cannot be inferred precisely from the reference, it is estimated from the similarity to the initiator consensus.

[c] Similarity to the consensus was calculated according to Ref. 28; a similarity greater than 81% is considered significant. If there is no initiation at the wild type start site, the value for that position is set between brackets. If there is initiation upstream or downstream of the wildtype start site, the value for the new site is indicated to the left or to the right, respectively.

[d] From the references: ++, as active as the wildtype; +++, more active; +, less active; −, no transcription.

[e] Activities *in vivo*/in vitro.

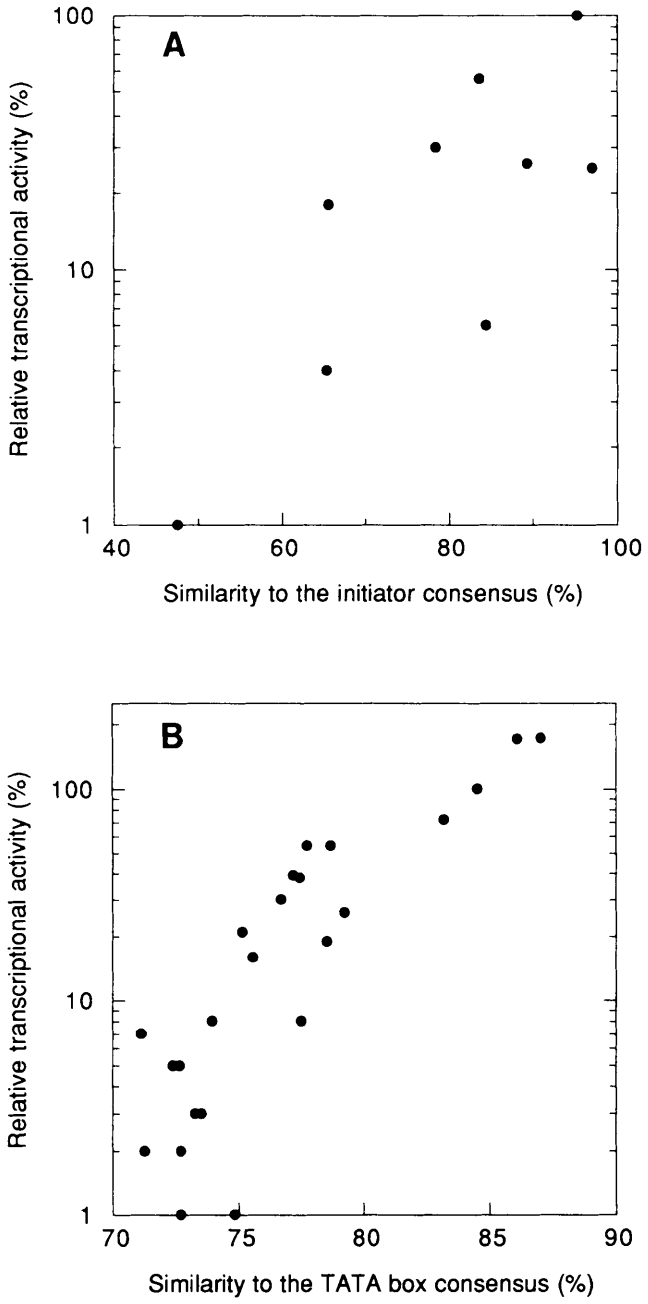

Figure 1. Correlation between transcriptional activity and similarity to the initiator or the TATA box consensus. (A) O'Shea-Greenfield and Smale (47) have tested whether sequences between −6 and +11 from several cellular and viral promoters function as initiators when placed downstream of binding sites for the transcription factor Sp1; (B) Wobbe and Struhl (84) have systematically changed 1 or 2 base pairs throughout the T_R TATA box from the yeast *his3* gene and have determined the transcriptional activities *in vitro* for the wildtype and 24 mutant sequences. We calculated for each sequence the similarity to the initiator or the TATA box consensus (28) and plotted the result against the relative activity.

directly between −3 and −1, +8 and +11, and +16 and +18 (56). LBP-1 may thus block initiation at +5, but not at +1. These observations show that predictions of initiator activity from the similarity to the initiator consensus can be wrong if another promoter element overlaps with the start site.

The six start-site sequences described at the beginning of this section (Table I) were originally assigned to different families because different transcription factors bind at or next to the initiator elements (57). However, these factors seem to function independently of the initiators. The human CAP site-binding factor requires sequences of the major late promoter of adenovirus 2 between +8 and +16, downstream of the initiator, for binding and activation of transcription (58, 59). The transcription factor TFII-I substitutes for the basal factor TFIIA and binds around the start sites of the major late promoter of adenovirus 2, of the terminal deoxynucleotidyltransferase promoter, and of the promoter of human immunodeficiency virus 1 (60). These regions contain a common sequence, YAYTCYYY, which resembles the initiator consensus. However, TFII-I also binds independently to a dissimilar sequence, CACGTG, the binding site for the activator protein USF (60). The transcription factor YY1 binds to the P5 promoter of adeno-associated virus type 2 between −10 and +13 (61) and is required for transcription from the isolated P5 initiator (49). At several other promoters, however, YY1 binds upstream or downstream of the start site and activates or represses transcription (49, 61–64); only the 3′-end of the consensus of these YY1 binding sites, CSKCCATYTT, resembles the initiator consensus. A nuclear protein binds to the porphobilinogen deaminase promoter immediately downstream of the initiator (45). The 60-kDa transcription factor E2F and the 180-kDa protein HIP1 bind to the dihydrofolate reductase promoter between −9 and −1 (50, 65, 66). E2F plays a role in the control of cell growth and has been cloned recently (67); HIP1 has not been characterized further, and its relationship to E2F is unknown. We proposed previously that binding of HIP1 controls the start site of transcription in the dihydrofolate reductase promoter (50). However, mutations between −9 and −1 that abolish protein binding and growth regulation do not affect transcription *in vitro*; conversely, a mutation at −2 that reduces the similarity to the initiator consensus and abolishes transcription *in vitro* does not affect protein binding (51, 66, Table II). Because of these more recent results, we now believe that E2F binds next to the initiator and mediates growth regulation and that the initiator controls the start site independently of E2F.

To summarize, these six promoters contain binding sites for different transcription factors that overlap with, but do not coincide with, the initiators. Their initiators, however, belong to the same family. Are there other families of initiators? It is possible that different initiator elements will be found when site-specific initiation is studied at more promoters. However, we examined all start sites whose positioning function has been studied rigorously. The available data can be explained simply and consistently with the similarity of the start sites to

the initiator consensus. Therefore, we propose that there is only one type of initiator that functions as a positioning element at the start site.

A Simple Model of Start-Site Selection

In 1980, Grosschedl and Birnstiel (31, 68) presented a simple model of start-site selection that distinguished between two classes of positioning elements. They noticed that the TATA box consensus was better defined than the initiator consensus; the latter seemed too ambiguous to direct site-specific initiation on its own. They proposed that a dominant "selector" element, such as the TATA box, guides RNA polymerase to a correct window of initiation. The initiator then determines the exact start site within that window. The central feature of this model is the hierarchical relationship between a selector and an initiator. What has been learned about site-specific initiation since 1980 is consistent with this model. Many more promoters have been analyzed; most of them contain a TATA box and an initiator as positioning elements. The components of the transcriptional machinery have been characterized; a common set of basal factors supports initiation by RNA polymerase II at all the various promoters. The assembly of the preinitiation complex has been investigated; TFIID binds to the TATA box, the selector, before RNA polymerase II and the other basal factors bind to the promoter. In the next sections, we will address three unresolved questions: (i) Is TFIID required for initiation at promoters without a TATA box? (ii) Can regulatory factors participate in start-site selection? and (iii) How does the initiator function?

Is TFIID Required for Initiation at Promoters without a TATA Box?

During assembly of the preinitiation complex, TFIID is the first basal transcription factor to bind to the promoter. Some promoters, however, do not contain a consensus TATA box, the binding site for TFIID. In this section, we will first describe the structure of TFIID; we will then examine whether TFIID is necessary for initiation at promoters without a TATA box and how TFIID could bind to such promoters.

TFIID can be obtained in different forms and from different sources (17, 69). TFIID in fruit fly and mammals is a large complex that comprises a TATA-binding protein (TBP) and several tightly bound TBP-associated factors; TFIID in yeast is a single peptide, the TATA-binding protein. TBP also plays a role in transcription by RNA polymerases I and III (70, 71). Native TFIID has been partially purified from yeast, fruit fly, rat, and human. TFIID in humans exists in at last two forms that contain different TBP-associated factors (72, 73). The TBP genes of yeast, fruit fly, mouse, human, and the plant arabidopsis have been cloned, and recombinant TBP has been expressed in bacterial and mammalian cells. About 180 amino acids at the carboxy terminus of TBP, which contain the DNA-binding domain, are conserved, whereas the amino-terminal domain varies in length and composition among species. TBP-associated factors have been purified, but their genes have not been cloned yet. Recombinant TBP and purified TBP-associated factors do not form a complete TFIID complex *in vitro*. However, recombinant TBP expressed in human cells binds the TBP-associated factors *in vivo*; the resulting native TFIID complex can be purified to homogeneity by immunoaffinity chromatography, if the recombinant TBP contains an "epitope tag" (74).

The requirement for TFIID in the absence of a consensus TATA box has been examined only for a few promoters: the IVa2 promoter of adenovirus 2 (75, 76), the major late promoter of adenovirus 2 with a mutant TATA box (46), and artificial promoters comprising an initiator and binding sites for the transcription factor Sp1 (44, 77). The transcriptional activities of these promoters were measured in nuclear extract, after heat-inactivation of TFIID or removal of TFIID with an anti-TBP antibody, or with purified RNA polymerase II and basal transcription factors; in all cases, addition of TFIID was necessary for transcription. In contrast, it has been reported that heat treatment of nuclear extract abolishes transcription from the uroporphyrinogen decarboxylase promoter, which contains a TATA box, but not from the porphobilinogen deaminase promoter, which lacks a TATA box (45). It is possible, however, that the heat treatment in these experiments inactivated another transcription factor instead of TFIID. For instance, the uroporphyrinogen decarboxylase promoter (78), but not the porphobilinogen deaminase promoter (79), contains a binding site for transcription factor Sp1, which is more heat sensitive than TFIID in nuclear extract (80, 81). These few examples suggest that TFIID is necessary for transcription by RNA polymerase II from any promoter, whether it contains a TATA box or not.

How does TFIID bind to promoters without a TATA box? One possibility is that TFIID can bind to other sequences than the TATA box. We used the similarity to the consensus as a quantitative measure to analyze the available data. Recombinant human TBP binds to sequences from several promoters that occur around −30, but do not look like TATA boxes (82). The more similar these sequences are to the TATA box consensus, the better they bind TBP (Table III); each sequence is the best or one of the best matches to the consensus within 100 base pairs around the start site of its promoter (data not shown). In the promoter of the yeast *his3* gene, various sequences with similarities to the consensus between 61% and 97% can substitute for a TATA box *in vivo* (83). However, most mutations of one or two base pairs in this TATA box reduce tran-

Table III. Similarity to the TATA Box Consensus and Affinity for TBP of Sequences in Eukaryotic Promoters around −30

Promoter[a]	Similarity to TATA box consensus[b] (%)	Percentile similarity[c]	Relative affinity for recombinant human TBP[d]
Ad2 major late	98	100.0	1
pSVsKC2	84	99.9	0.5
Human DHFR	83	99.9	0.3
Mouse rpL32	62	91.3	0.25
SV40 major late	68	96.6	0.15
Mouse TdT	62	91.3	0.15
Mouse IRF-1	70	97.6	0.15
pSVsKA1	52	72.5	0.05

[a] See Table I for abbreviations; additional abbreviations: rpL32, large ribosomal protein 32; IRF-1, interferon regulatory factor-1; pSVsKC2 and pSVsKA1, mutants of the SV40 major late promoter (Ref. 114).
[b] Calculated according to Bucher (28; S. Wiley, personal communication, October 1992).
[c] Percentage of all possible 15-nucleotide sequences that are less similar to the TATA box consensus (S. Wiley, personal communication, October 1992).
[d] From Ref. 82.

scription with purified mammalian RNA polymerase II and basal factors (84). We found a close correlation between the transcriptional activity of these mutants and their similarity to the TATA box consensus (Fig. 1B). Together, these findings show that the ability of TFIID to bind to a sequence and to direct basal transcription depends on the similarity of that sequence to the TATA box consensus. TFIID may bind *in vitro* to isolated promoter sequences that do not match the TATA box consensus; however, TFIID probably cannot distinguish such promoter sequences from random sequences in the genome. Thus, binding of TFIID to DNA alone cannot account for site-specific initiation at promoters that lack a TATA box.

A second possibility is that other components of the transcriptional machinery tether TFIID to promoters without a TATA box. Many regulatory transcription factors bind to specific DNA sequences (see above) and interact directly with TFIID. For example, binding of purified TFIID extends farther downstream when the activator proteins USF, GAL4, or ATF bind next to the TATA box (85–87). Both native TFIID and recombinant TBP support a basal level of transcription from a promoter with a TATA box; only native TFIID, however, supports a high level of transcription in the presence of activator proteins (69), indicating that the TBP-associated factors are necessary for the interaction between TFIID and activator proteins. Native TFIID also supports transcription from a promoter without a consensus TATA box, but with an activator binding site (44). This result shows that activator proteins can tether TFIID to promoters that lack a TATA box. Does TFIID have to bind to DNA at all to direct complex assembly, or is tethering to other components of the transcriptional machinery sufficient? Some point mutations in the carboxy terminus of TBP that abolish binding to the TATA box do not affect transcription from the major late promoter of adenovirus 2 in the presence of RNA polymerase II and the other basal factors (88). This result suggests that interactions other than between TFIID and the TATA box, perhaps between RNA polymerase II and the initiator (see below), can position the preinitiation complex. To conduct similar experiments with native TFIID and activator proteins, one could express epitope-tagged TBP mutants in human cells and obtain TFIID complexes with altered DNA-binding abilities.

In conclusion, TFIID appears to be required for transcription by RNA polymerase II from all promoters. TFIID is a complex transcription factor that binds not only to the TATA box, the basal factors, and RNA polymerase II, but also, through the TBP-associated factors, to regulatory transcription factors. At promoters that lack a consensus TATA box, TFIID may bind only weakly to the sequence around −30, but is stabilized by activator proteins.

Can Regulatory Factors Participate in Start-Site Selection?

If regulatory factors can tether TFIID or other components of the preinitiation complex to the promoter, they should be able to influence the start site of transcription. Here, we will examine whether regulatory transcription factors and their binding sites have positioning function.

A binding site for the activator protein GCN4 can replace a TATA box in yeast without affecting the start site; site-specific initiation requires both the DNA-binding domain and the acidic activation domain of GCN4 (89). A cluster of four binding sites for the glucocorticoid receptor, a transcription factor, can direct site-specific initiation in the absence of a TATA box; transcription starts about 45–55 base pairs downstream within sequences that are similar to the initiator consensus (90, 91). The human activator protein ETF stimulates transcription from promoters with ETF binding sites, but only if they lack a TATA box; ETF may have no effect if TFIID can use a TATA box to bind directly to a promoter (92). The erythroid transcription factor eUSF binds to the promoter of the duck histone H5 gene downstream of the start site (93). eUSF stabilizes binding of TFIID in the presence or absence of a TATA box. Both TFIID and eUSF are necessary for site-specific initiation in the absence of the TATA box. At the start sites of the major late promoter of adenovirus 2, the terminal deoxynucleotidyltransferase promoter, the porphobilinogen deaminase promoter,

the P5 promoter of adeno-associated virus type 2, and the dihydrofolate reductase promoter, a binding site for a regulatory transcription factor overlaps with an initiator; these sequences form compact promoters that can direct site-specific initiation by themselves (see above). These regulatory factors probably interact with TFIID or other basal factors and stabilize the preinitiation complex at the start site.

Together, these results indicate that regulatory transcription factors play a role both in regulation and in start-site selection. The binding site for such a regulatory factor can function as a selector element and compensate for the lack of a TATA box if the factor tethers TFIID to the promoter. The distinction between regulatory and positioning function may be merely theoretical: An activator protein that stabilizes preinitiation complexes nearby also increases the probability that RNA polymerase II initiates nearby. This model does not conflict with the finding that activator proteins act at various distances (18). First, only some activators may have both functions. Second, an activator by itself may not be sufficient to direct initiation; only specific combinations and arrangements of promoter elements support initiation of transcription (94–96).

How Does the Initiator Function?

The initiator is a common positioning element that determines the exact start site of transcription (see above). Therefore, we expect that a common component of the transcriptional machinery interacts with the initiator during the assembly of the preinitiation complex. Here, we will discuss whether one of the basal transcription factors or RNA polymerase II interacts with the initiator and whether selectors, such as the TATA box, and the initiator act at different steps during initiation.

No basal transcription factor, except TFIID, has been shown to interact with a specific DNA sequence (15, 17). A comparison of the transcriptional machineries of yeast and mammals suggests that TFIID also does not interact with the initiator. Start-site selection in yeast is different from that in higher eukaryotes, as transcription does not start at a fixed distance from the TATA box, but between 40 and 120 base pairs downstream (3). Within that window, various sequences may act as initiators: TCRA, RRYRR (97), YAWR (98), YAAR (99), ACAGATC immediately upstream of +1 (100), or a T-rich sequence between −30 and −10 followed by an A-rich sequence between −8 and +15 (101). In yeast nuclear extract, transcription from the major late promoter of adenovirus 2 starts between 63 and 69 base pairs, instead of 30 base pairs, downstream of the TATA box (102). However, when mammalian RNA polymerase II and basal factors are combined with either human or yeast TFIID, transcription starts about 30 base pairs downstream of the TATA box (103). Mutations in the *SHI* gene, which does not code for TFIID, affect the spacing between TATA box and start sites in yeast (104). These results show that factors other than TFIID measure the distance between the TATA box and the start site.

RNA polymerase II itself could select the start site, once it has joined the preinitiation complex. In support of this hypothesis, mutations in the two large subunits of yeast RNA polymerase II alter the start-site pattern and reduce the accuracy of initiation (105). Purified human and bovine RNA polymerase II initiates transcription at several specific sites in the major late promoter and the IVa2 promoter of adenovirus 2 (46, 106). Two initiation sites lie within AT-rich sequences, the TATA box at −30 in the major late promoter and the inverted TATA box at +19 in the IVa2 promoter. The other initiation sites, however, lie within sequences that are between 80% and 89% similar to the initiator consensus and include the authentic start sites in both promoters. Mutations that reduce the similarities of the initiators around +1 to the consensus from 97% to 82% (major late) or from 89% to 65% (IVa2) specifically abolish initiation at +1 by purified RNA polymerase II. These results show that RNA polymerase II does interact with specific sequences, including the initiator. Transcription of closed circular DNA by purified RNA polymerase II alone is at least as efficient as by the complete transcriptional machinery in crude cell extracts; the transcriptional machinery may therefore not be activating an idle enzyme, but bringing a highly active enzyme to the biologically important start sites (107).

Several lines of evidence suggest that the initiator acts later during initiation than a selector. First, a TATA box dominates the selection of the start site in higher eukaryotes within a narrow window at a fixed distance; only within that window does the initiator determine the exact start site. Transcription from the promoter of the trout protamine gene 3 initiates 31 base pairs downstream of a TATA box at a poor match to the initiator consensus, despite better matches at −5 and +5 (Table II). Inserting 1, 2, or 3 base pairs at −25 gradually reduces initiation at +1 and increases initiation at −5; deleting 1 or 3 base pairs at −25 gradually increases initiation at +5. There is no initiation between the start sites at −5, +1, and +5 (108). Second, an initiator never dominates start-site selection in promoters that contain Sp1 binding sites, a TATA box, and an initiator in various combinations, positions, and orientations (47). Third, TFIID and TFIIA together can bind to a TATA box without other factors; formation of a stable complex at an isolated initiator also requires TFIIB, RNA polymerase II, and TFIIF, which enter the preinitiation complex after TFIID and TFIIA (46). However, TBP, TFIIB, RNA polymerase II, and only the RAP30 subunit of TFIIF cannot bind stably to an

isolated initiator (109). Fourth, the preinitiation complex assembles first at the TATA box and then extends to the initiator. In the major late promoter of adenovirus 2, a restriction site between TATA box and initiator is protected against cleavage by the restriction enzyme only after TFIIE and δ, a basal factor from rat liver, join the initial complex of TFIID, -B, -F, and RNA polymerase II (110, 111). Fifth, the start site is not yet completely fixed when the preinitiation complex has formed. Initiation can be primed *in vitro* with dinucleotides. Addition of dinucleotides that are complementary to sequences between −4 and +6, but not farther away from +1, shifts the start site to the corresponding position in the initiators of the major late promoter of adenovirus 2, of the EIV promoter of human adenovirus 5 (112), and of the terminal deoxynucleotidyltransferase promoter (113). These results indicate that a common factor, probably RNA polymerase II itself, recognizes the initiator at a later step during assembly of the preinitiation complex.

In conclusion, selectors, such as the TATA box or binding sites for regulatory proteins, and the initiator may play different roles in site-specific initiation. Selectors bind or tether TFIID to the promoter in the first step of initiation and control the position of the preinitiation complex. The initiator interacts later with a common component of the complex, probably with RNA polymerase II, and controls the exact start site. This conclusion agrees with the simple hierarchical model of start-site selection presented above. We do not exclude the possibility that a factor like TFII-I (see above) can also bind to an initiator-like sequence and tether TFIID to the promoter.

Summary

RNA polymerase II initiates transcription at specific DNA sequences in an ordered series of steps with the help of a common set of transcription factors. We have examined the structural and functional relationships among positioning elements, the DNA sequences that control start-site selection. On the basis of this analysis, we propose a simple and universal model of initiation by RNA polymerase II in higher eukaryotes: Two types of positioning elements, the selectors and the initiator, reflect the hierarchy of steps in the initiation process. Selectors, such as the TATA box and binding sites for activator proteins, act first and direct the basal transcription factor TFIID to the correct promoter region. TFIID can bind directly to promoters that contain a TATA box or can be tethered by activator proteins to promoters that lack a consensus TATA box. TFIID then assembles the remaining basal factors and RNA polymerase II into the preinitiation complex. The initiator interacts with RNA polymerase II at a later step and determines the exact start site. There is only one family of initiator elements; the sequences that are most similar to the initiator consensus serve as start sites within the promoter region specified by the selectors. This model should be a useful guideline in future studies of promoters and the initiation process.

Note added in proof. Smale and colleagues have confirmed that the start-site sequence of the core promoter of human immunodeficiency virus 1 is not functionally analogous to the initiators in the major late promoter of adenovirus 2 and the promoter of the mouse terminal deoxynucleotidyltransferase gene (B. Zenzie-Gregory, P. Sheridan, K. A. Jones, S. T. Smale, submitted). One of the TBP-associated factors has now been cloned and identified as the cell-cycle regulatory gene CCG1 (115, 116).

Work in this laboratory was supported by United States Public Health Service Grants CA07175, CA23076, and CA45240 to P. J. F. We thank Arnold Berk and Danny Reinberg for sending us their manuscripts before publication, Steve Wiley and Janet Mertz for communicating unpublished results, and Andrew Buermeyer, Dick Burgess, Karen Gould, Carol Gross, Judith Kimble, Janet Mertz, Lynda Schilling, Jill Slansky, Bill Sugden, Nancy Thompson, and Steve Wiley for discussions and comments on the manuscript. We apologize to those colleagues whose work we did not cite because of space limitations.

1. McClure WR. Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem **54**:171–204, 1985.
2. Reiter W-D, Palm P, Zillig W. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eucaryotic pol II promoters. Nucleic Acids Res **16**:1–19, 1988.
3. Struhl K. Molecular mechanisms of transcriptional regulation in yeast. Annu Rev Biochem **58**:1051–1077, 1989.
4. Costanzo MC, Fox TD. Control of mitochondrial gene expression in *Saccharomyces cerevisiae*. Annu Rev Genet **24**:91–113, 1990.
5. Clayton DA. Replication and transcription of vertebrate mitochondrial DNA. Annu Rev Cell Biol **7**:453–478, 1991.
6. Joshi CP. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res **15**:6643–6653, 1987.
7. Jarrell KA, Meselson M. *Drosophila* retrotransposon promoter includes an essential sequence at the initiation site and requires a downstream sequence for full activity. Proc Natl Acad Sci USA **88**:102–104, 1991.
8. Sawadogo M, Sentenac A. RNA polymerase B (II) and general transcription factors. Annu Rev Biochem **59**:711–754, 1990.
9. Geiduschek EP, Tocchini-Valentini GP. Transcription by RNA polymerase III. Annu Rev Biochem **57**:873–914, 1988.
10. Kunkel GR. RNA polymerase III transcription of genes that lack internal control regions. Biochim Biophys Acta **1088**:1–9, 1991.
11. Reeder RH. Regulatory elements of the generic ribosomal gene. Curr Opin Cell Biol **1**:466–474, 1989.
12. Weil PA, Luse DS, Segall J, Roeder RG. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484, 1979.
13. Young RA. RNA polymerase II. Annu Rev Biochem **60**:689–715, 1991.

14. Heintz N (Ed). Transcription—A special issue. Trends Biochem Sci **16**:393–447, 1991.
15. Weinmann R. The basic RNA polymerase II transcriptional machinery. Gene Expression **2**:81–91, 1992.
16. Zawel L, Reinberg D. Advances in RNA polymerase II transcription. Curr Opin Cell Biol **4**:488–495, 1992.
17. Zawel L, Reinberg D. Initiation of transcription by RNA polymerase II: A multi-step process. Prog Nucleic Acid Res Mol Biol **44**:67–108, 1993.
18. Ptashne M. A Genetic Switch: Phage λ and Higher Organisms (2nd ed). Cambridge, MA: Cell Press and Blackwell Scientific Publications, 1992.
19. Lewis MK, Burgess RR. Eucaryotic RNA polymerases. In: Boyer PD, Ed. The Enzymes. Vol 15, Nucleic Acids, Part B (3rd ed). New York: Academic Press, pp109–153, 1982.
20. Felsenfeld G. Chromatin as an essential part of the transcriptional mechanism. Nature **355**:219–224, 1992.
21. Conaway JW, Conaway RC. Initiation of eukaryotic messenger RNA synthesis. J Biol Chem **266**:17721–17724, 1991.
22. Bucher P. The Eucaryotic Promoter Database EPD. Release 30. Postfach 10.2209, D-6900 Heidelberg: EMBL Nucleotide Sequence Data Library, 1992.
23. Goldberg ML. Sequence analysis of *Drosophila* histone genes [Dissertation]. Stanford, CA: Stanford University, 1979.
24. Busslinger M, Portmann R, Irminger JC, Birnstiel ML. Ubiquitous and gene-specific regulatory 5′ sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res **8**:957–977, 1980.
25. Breathnach R, Chambon P. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem **50**:349–383, 1981.
26. Bucher P, Trifonov EN. Compilation and analysis of eucaryotic POL II promoter sequences. Nucleic Acids Res **14**:10009–10026, 1986.
27. Penotti FE. Human DNA TATA boxes and transcription initiation sites: A statistical study. J Mol Biol **213**:37–52, 1990.
28. Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol **212**:563–578, 1990.
29. Nussinov R. Sequence signals in eucaryotic upstream regions. Crit Rev Biochem Mol Biol **25**:185–224, 1990.
30. Dahlberg JE, Lund E. The genes and transcription of the major small nuclear RNAs. In: Birnstiel ML, Ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Berlin: Springer-Verlag, pp38–70, 1988.
31. Grosschedl R, Birnstiel ML. Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription *in vivo*. Proc Natl Acad Sci USA **77**:7102–7106, 1980.
32. Sassone-Corsi P, Corden J, Kédinger C, Chambon P. Promotion of specific in vitro transcription by excised "TATA" box sequences inserted in a foreign nucleotide environment. Nucleic Acids Res **9**:3941–3958, 1981.
33. Concino MF, Lee RF, Merryweather JP, Weinmann R. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription *in vitro*. Nucleic Acids Res **12**:7423–7433, 1984.
34. Wasylyk B. Transcription elements and factors of RNA polymerase B promoters of higher eukaryotes. CRC Crit Rev Biochem **23**:77–120, 1988.
35. Benoist C, Chambon P. *In vivo* sequence requirements of the SV40 early promoter region. Nature **290**:304–310, 1981.
36. Mathis DJ, Chambon P. The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. Nature **290**:310–315, 1981.
37. Ghosh PK, Lebowitz P, Frisque RJ, Gluzman Y. Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc Natl Acad Sci USA **78**:100–104, 1981.
38. Orkin SH, Sexton JP, Cheng T-C, Goff SC, Giardina PJV, Lee JI, Kazazian HH Jr. ATA box transcription mutation in β-thalassemia. Nucleic Acids Res **11**:4727–4734, 1983.
39. Parry HD, Tebb G, Mattaj IW. The *Xenopus* U2 gene PSE is a single, compact, element required for transcription initiation and 3′ end formation. Nucleic Acids Res **17**:3633–3644, 1989.
40. Skuzeski JM, Lund E, Murphy JT, Steinberg TH, Burgess RR, Dahlberg JE. Synthesis of human U1 RNA: II. Identification of two regions of the promoter essential for transcription initiation at position +1. J Biol Chem **259**:8345–8352, 1984.
41. Murphy JT, Skuzeski JT, Lund E, Steinberg TH, Burgess RR, Dahlberg JE. Functional elements of the human U1 RNA promoter: Identification of five separate regions required for efficient transcription and template competition. J Biol Chem **262**:1795–1803, 1987.
42. Dahlberg JE, Lund E. How does III × II make U6? Science **254**:1462–1463, 1991.
43. Smale ST, Baltimore D. The "initiator" as a transcription control element. Cell **57**:103–113, 1989.
44. Smale ST, Schmidt MC, Berk AJ, Baltimore D. 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, 1990.
45. Beaupain D, Eléouët JF, Roméo PH. 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, 1990.
46. Carcamo J, Buckbinder L, Reinberg D. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc Natl Acad Sci USA **88**:8052–8056, 1991.
47. O'Shea-Greenfield A, Smale ST. 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, 1992.
48. Kasai Y, Chen H, Flint SJ. Anatomy of an unusual RNA polymerase II promoter containing a downstream TATA element. Mol Cell Biol **12**:2884–2897, 1992.
49. Seto E, Shi Y, Shenk T. YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. Nature **354**:241–245, 1991.
50. Means AL, Farnham PJ. Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol Cell Biol **10**:653–661, 1990.
51. Means AL. Identification, purification and characterization of HIP1: A protein that positions transcription initiation of the dihydrofolate reductase gene [Dissertation]. Madison, WI: University of Wisconsin, Madison, 1991.
52. Wagner M, Perry RP. Characterization of the multigene family encoding the mouse S16 ribosomal protein: Strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. Mol Cell Biol **5**:3560–3576, 1985.
53. Hariharan N, Perry RP. 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, 1990.
54. Chung S, Perry RP. Cell-free transcription of a mouse ribosomal-protein-encoding gene: The effects of promoter mutations. Gene **100**:173–180, 1991.
55. Jones KA, Luciw PA, Duchange N. Structural arrangements of transcription control domains within the 5′-untranslated leader regions of the HIV-1 and HIV-2 promoters. Genes Dev **2**:1101–1114, 1988.
56. Kato H, Horikoshi M, Roeder RG. Repression of HIV-1 transcription by a cellular protein. Science **251**:1476–1479, 1991.

57. Weis L, Reinberg D. Transcription by RNA polymerase II: Initiator-directed formation of transcription competent complexes. FASEB J **6**:3300–3309, 1992.
58. Garfinkel S, Thompson JA, Jacob WF, Cohen R, Safer B. Identification and characterization of an adenovirus 2 major late promoter CAP sequence DNA-binding protein. J Biol Chem **265**:10309–10319, 1990.
59. Safer B, Reinberg D, Jacob WF, Maldonado E, Carcamo J, Garfinkel S, Cohen R. Interaction of CAP sequence site binding factor and transcription factor IID preceding and following binding to the adenovirus 2 major late promoter. J Biol Chem **266**:10989–10994, 1991.
60. Roy AL, Meisterernst M, Pognonec P, Roeder RG. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature **354**:245–248, 1991.
61. Shi Y, Seto E, Chang L-S, Shenk T. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell **67**:377–388, 1991.
62. Park K, Atchison ML. Isolation of a candidate repressor/activator, NF-E1 (YY-1, δ), that binds to the immunoglobulin κ 3′ enhancer and the immunoglobulin heavy-chain μE1 site. Proc Natl Acad Sci USA **88**:9804–9808, 1991.
63. Hariharan N, Kelley DE, Perry RP. δ, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. Proc Natl Acad Sci USA **88**:9799–9803, 1991.
64. Flanagan JR, Becker KG, Ennist DL, Gleason SL, Driggers PH, Levi B-Z, Appella E, Ozato K. Cloning of a negative transcription factor that binds to the upstream conserved region of moloney murine leukemia virus. Mol Cell Biol **12**:38–44, 1992.
65. Blake MC, Jambou RC, Swick AG, Kahn JW, Azizkhan JC. Transcriptional initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. Mol Cell Biol **10**:6632–6641, 1990.
66. Means AL, Slansky JE, McMahon SL, Knuth MW, Farnham PJ. The HIP1 binding site is required for growth regulation of the dihydrofolate reductase gene promoter. Mol Cell Biol **12**:1054–1063, 1992.
67. Nevins JR. E2F: A link between the Rb tumor suppressor protein and viral oncoproteins. Science **258**:424–429, 1992.
68. Grosschedl R, Birnstiel ML. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc Natl Acad Sci USA **77**:1432–1436, 1980.
69. Pugh BF, Tjian R. Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. J Biol Chem **267**:679–682, 1992.
70. Sharp PA. TATA-binding protein is a classless factor. Cell **68**:819–821, 1992.
71. White RJ, Jackson SP. The TATA-binding protein: A central role in transcription by RNA polymerases I, II and III. Trends Genet **8**:284–288, 1992.
72. Timmers HTM, Sharp PA. The mammalian TFIID protein is present in two functionally distinct complexes. Genes Dev **5**:1946–1956, 1991.
73. Timmers HTM, Meyers RE, Sharp PA. Composition of transcription factor B-TFIID. Proc Natl Acad Sci USA **89**:8140–8144, 1992.
74. Zhou Q, Lieberman PM, Boyer TG, Berk AJ. Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes Dev **6**:1964–1974, 1992.
75. Carcamo J, Lobos S, Merino A, Buckbinder L, Weinmann R, Natarajan V, Reinberg D. Factors involved in specific transcription by mammalian RNA polymerase II: Role of factors IID and MLTF in transcription from the adenovirus major late and IVa2 promoters. J Biol Chem **264**:7704–7714, 1989.
76. Carcamo J, Maldonado E, Cortes P, Ahn MH, Ha I, Kasai Y, Flint J, Reinberg D. A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. Genes Dev **4**:1611–1622, 1990.
77. Pugh BF, Tjian R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev **5**:1935–1945, 1991.
78. Romana M, Dubart A, Beaupain D, Chabret C, Goossens M, Romeo P-H. Structure of the gene for human uroporphyrinogen decarboxylase. Nucleic Acids Res **15**:7343–7356, 1987.
79. Chretien S, Dubart A, Beaupain D, Raich N, Grandchamp B, Rosa J, Goossens M, Romeo P-H. Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. Proc Natl Acad Sci USA **85**:6–10, 1988.
80. Farnham PJ, Cornwell MM. Sp1 activation of RNA polymerase II transcription complexes involves a heat-labile DNA binding component. Gene Expression **1**:137–148, 1991.
81. Kollmar R, Lindstrom MJ, Farnham PJ. Heat sensitivity and Sp1 activation of complex formation at the syrian hamster carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase promoter *in vitro*. J Biol Chem **267**:385–391, 1992.
82. Wiley SR, Kraus RJ, Mertz JE. 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, 1992.
83. Singer VL, Wobbe CR, Struhl K. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes Dev **4**:636–645, 1990.
84. Wobbe CR, Struhl K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. Mol Cell Biol **10**:3859–3867, 1990.
85. Sawadogo M, Roeder RG. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175, 1985.
86. Horikoshi M, Carey MF, Kakidani H, Roeder RG. Mechanism of action of a yeast activator: Direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. Cell **54**:665–669, 1988.
87. Horikoshi M, Hai T, Lin Y-S, Green MR, Roeder RG. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. Cell **54**:1033–1042, 1988.
88. Yamamoto T, Horikoshi M, Wang J, Hasegawa S, Weil PA, Roeder RG. A bipartite DNA binding domain composed of direct repeats in the TATA box binding factor TFIID. Proc Natl Acad Sci USA **89**:2844–2848, 1992.
89. Chen W, Struhl K. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. EMBO J **8**:261–268, 1989.
90. Schatt MD, Rusconi S, Schaffner W. A single DNA-binding transcription factor is sufficient for activation from a distant enhancer and/or from a promoter position. EMBO J **9**:481–487, 1990.
91. Wieland S, Schatt MD, Rusconi S. Role of TATA-element in transcription from glucocorticoid receptor-responsive model promoters. Nucleic Acids Res **18**:5113–5118, 1990.
92. Kageyama R, Merlino GT, Pastan I. Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box. J Biol Chem **264**:15508–15514, 1989.
93. Bungert J, Kober I, Düring F, Seifart KH. Transcription factor eUSF is an essential component of isolated transcription complexes on the duck histone H5 gene and it mediates the interaction of TFIID with a TATA-deficient promoter. J Biol Chem **223**:885–898, 1992.

94. Taylor ICA, Kingston RE. Factor substitution in a human HSP70 gene promoter: TATA-dependent and TATA-independent interactions. Mol Cell Biol **10**:165–175, 1990.
95. Xu LC, Thali M, Schaffner W. Upstream box/TATA box order is the major determinant of the direction of transcription. Nucleic Acids Res **19**:6699–6704, 1991.
96. Wang W, Gralla JD. Differential ability of proximal and remote element pairs to cooperate in activating RNA polymerase II transcription. Mol Cell Biol **11**:4561–4571, 1991.
97. Hahn S, Hoar ET, Guarente L. Each of three "TATA elements" specifies a subset of the transcription initiation sites at the *CYC-1* promoter of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA **82**:8562–8566, 1985.
98. Furter-Graves EM, Hall BD. DNA sequence elements required for transcription initiation of the *Schizosaccharomyces pombe* ADH gene in *Saccharomyces cerevisiae*. MGG **223**:407–416, 1990.
99. Healy AM, Zitomer RS. A sequence that directs transcriptional initiation in yeast. Curr Genet **18**:105–109, 1990.
100. Rathjen J, Mellor J. Characterisation of sequences required for RNA initiation from the *PGK* promoter of *Saccharomyces cerevisiae*. Nucleic Acids Res **18**:3219–3225, 1990.
101. Maicas E, Friesen JD. A sequence pattern that occurs at the transcription initiation region of yeast RNA polymerase II promoters. Nucleic Acids Res **18**:3387–3393, 1990.
102. Lue NF, Flanagan PM, Sugimoto K, Kornberg RD. Initiation by yeast RNA polymerase II at the adenoviral major late promoter in vitro. Science **246**:661–664, 1989.
103. Buratowski S, Hahn S, Sharp PA, Guarente L. Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature **334**:37–42, 1988.
104. Furter-Graves EM, Furter R, Hall BD. *SHI*, a new yeast gene affecting the spacing between TATA and transcription initiation sites. Mol Cell Biol **11**:4121–4127, 1991.
105. Hekmatpanah DS, Young RA. Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. Mol Cell Biol **11**:5781–5791, 1991.
106. Mishoe H, Brady JN, Lancz G, Salzmann NP. *In vitro* transcription initiation by purified RNA polymerase II within the adenovirus 2 major late promoter region. J Biol Chem **259**:2236–2242, 1984.
107. Dynan WS, Burgess RR. *In vitro* transcription by wheat germ RNA polymerase II: Initiation of RNA synthesis on relaxed, closed circular template. J Biol Chem **256**:5866–5873, 1981.
108. Kovacs BJ, Butterworth PHW. The effect of changing the distance between the TATA-box and cap site by up to three base pairs on the selection of the transcriptional start site of a cloned eukaryotic gene *in vitro* and *in vivo*. Nucleic Acids Res **14**:2429–2442, 1986.
109. Killeen M, Coulombe B, Greenblatt J. Recombinant TBP, transcription factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II. J Biol Chem **267**:9463–9466, 1992.
110. Conaway RC, Garrett KP, Hanley JP, Conaway JW. Mechanism of promoter selection by RNA polymerase II: Mammalian transcription factors α and $\beta\gamma$ promote entry of polymerase into the preinitiation complex. Proc Natl Acad Sci USA **88**:6205–6209, 1991.
111. Conaway JW, Bradsher JN, Conaway RC. Mechanism of assembly of the RNA polymerase II preinitiation complex: Transcription factors δ and ϵ promote stable binding of the transcription apparatus to the initiator element. J Biol Chem **267**:10142–10148, 1992.
112. Samuels M, Fire A, Sharp PA. Dinucleotide priming of transcription mediated by RNA polymerase II. J Biol Chem **259**:2517–2525, 1984.
113. Zenzie-Gregory B, O'Shea-Greenfield A, Smale ST. Similar mechanisms for transcription initiation mediated through a TATA box or an initiator element. J Biol Chem **267**:2823–2830, 1992.
114. Nandi A, Das G, Salzman NP. Characterization of a surrogate TATA box promoter that regulates in vitro transcription of the simian virus 40 major late gene. Mol Cell Biol **5**:591–594, 1985.
115. Ruppert S, Wang EH, Tjian R. Cloning and expression of human $TAF_{II}250$: A TBP-associated factor implicated in cell-cycle regulation. Nature **362**:175–179, 1993.
116. Hisatake K, Hasegawa S, Takada R, Nakatani Y, Horikoshi M, Roeder RG. The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. Nature **362**:179–181, 1993.