Nucleic Acids Research, Vol. 19, No. 24 ***6699–6704***

Upstream box/TATA box order is the major determinant of the direction of transcription

Licen Xu, Markus Thali and Walter Schaffner*
Institut für Molekularbiologie II der Universität Zürich, Irchel, CH-8057 Zürich, Switzerland

Received October 29, 1991; Revised and Accepted November 26, 1991

ABSTRACT

Mammalian gene promoters for transcription by RNA polymerase II are typically organized in the following order: upstream sequence motif(s)/TATA box/initiation site. Here we report studies in which the order, orientation and DNA sequences of these three elements are varied to determine how these affect polarity of transcription. We have constructed promoters with an 'octamer' upstream sequence ATTTGCAT (or its complement ATGCAAAT) in combination with several different TATA boxes and initiation (cap) sites, and tested these promoters in transfection experiments with cultured cells. TATA boxes derived from the adenovirus major late promoter (TATAAAA), immunoglobulin kappa light chain (TTATATA) and heavy chain (TAAATATA) promoter functioned equally well or even better when inverted. Only the β-globin TATA box (CATAAAA) was poorly active when inverted. In addition, a symmetrical TATA box (TATATATA) derived from a casein gene was very active. Our results suggest that the asymmetry of most TATA boxes (consensus TATAAAA) is not a primary determinant of the polarity of transcription. We also found that the initiation (cap) site, which usually consists of an adenine embedded in a pyrimidine-rich region (PyPyCAPyPyPyPyPy), was permissive towards sequence alterations; even a randomly composed sequence worked well. However, an inverted, hence purine-rich, cap site reduced transcript levels to 1/7th, as did an oligo G sequence. Irrespective of the presence of a cap site, the configuration: 'TATA box/octamer' yielded a strong leftward, rather than rightward transcription. From this, we conclude that the polarity of transcription is primarily determined by the linear order of an upstream sequence relative to a TATA box, rather than by the individual orientations of either of these two elements.

INTRODUCTION

Many promoters for RNA polymerase II transcription contain at least one binding site for a regulatory factor ('upstream sequence'), a TATA box, and an initiation site ('cap site'). Transcription can be further boosted by clusters of upstream binding sites for transcription factors, these can be located nearby or at remote positions (enhancer effect) (reviewed in 1). Upstream sequences such as those binding Sp1 or octamer transcription factors can be experimentally inverted without adverse effect, and in natural promoters these sequences can be found in either orientation (2, 3). However, the TATA box which binds the ubiquitous transcription factor TFIID (4), is usually asymmetrical with the consensus sequence 5′-TATAAAA-3′, although many variations of this sequence have been found. The TATA-binding factor TFIID has recently been cloned from many species (5–12, for review see 13). The factors of yeast and mammals can replace each other for basal level transcription, but not for regulated interaction with upstream factors (14–18).

RNA polymerase II tends to initiate transcription approximately 25 nucleotides downstream of a TATA box. A consensus cap site sequence, 5′-PyPyCAPyPyPyPyPy-3′, has been identified in several mammalian TATA box-containing genes, with transcription beginning at the A position (19). In a few cases, mutations in the cap region (also referred to as initiator box) have been found to result in alternative start sites or a reduction in promoter strength (19–24).

In the present study we have analyzed the contributions of the three elements: upstream sequence, TATA box, and initiation site, to the direction and efficiency of transcription. We have investigated the effect of inverting the TATA box to see whether the asymmetry of this element contributes to the polarity of transcription. In addition, we made permutations of the three elements and determined their effects on the direction of transcription.

We find that the position of the upstream sequence relative to the TATA box is more important in determining the polarity of transcription than the orientation of the TATA sequence (or upstream sequence) itself. The presence of a permissive initiation site can considerably facilitate initiation if it is located at the correct position about 25 bp downstream from the center of the TATA box.

MATERIALS & METHODS

Plasmid constructions

All plasmids were made by standard recombinant DNA techniques (25). Constructions with different TATA boxes: From the reporter gene sβ1 (26), which contains the β-globin coding

* To whom correspondence should be addressed

sequence and the SV40 enhancer downstream of it, the 32 bp SalI-PstI fragment containing the β-globin TATA sequence was deleted and various oligonucleotides were inserted into this site. The TATA sequences contain the 7 TATA nucleotides and 6 flanking nucleotides on either side. The distances between the TATA boxes and the cap site, and between the upstream octamer and the TATA boxes remain the same as in sβ1.

Constructions with altered order of elements: The SacI-PstI fragment with octamer and TATA box was removed from plasmid sβ1 and oligos with either the changed order: TATA/octamer, or only a TATA box were inserted in this region (Figure 2A). In order to circumvent a possible block of transcription by the octamer sequence, we designed another oligonucleotides with the normal cap sequence between the TATA box and the downstream octamer sequence, and added some extra nucleotides upstream of the TATA box to render the distance between the upstream octamer sequence and the TATA box the same as that found between the TATA box and the downstream octamer sequence (Figure 2C). The distance between octamer and TATA box was not changed. The TATA box with 6 flanking nucleotides on either side was exchanged with the octamer and 6 flanking nucleotides. To detect 'leftward' transcription, we constructed a series of plasmids with β-globin coding sequences of different length on both sides of the promoter. With these, the ratio of rightward to leftward transcription could be quantified by using the same RNA probe. For these constructs, the PstI-PstI fragment between two different length β-globin coding sequences was removed from plasmid $E\beta_2\beta_1E$ (26) and oligos with either the natural order of elements, or the changed order: TATA/octamer were inserted in this region. Furthermore, the leftward cap sequence was shifted further left by an inserted random sequence upstream of the changed order promoter (Figure 3). We performed computer searches of all new sequences and junctions to make sure that no other known factor binding motifs were fortuitously created.

Constructions with different cap sequences: SalI-PstI oligonucleotides containing the inverted kappa TATA sequence with either the β-globin initiation sequence (21), the TdT 17mer initiator (27), the Drosophila cuticle protein III initiation sequence (28, 29), one of two β-globin cap mutants, a random cap sequence, or the inverted β-globin cap sequence were inserted into sβ1 (Figure 4). The random cap sequence is an artificial sequence that is totally different from the consensus cap sequence in either orientation. The inverted β-globin cap sequence oligo contains the inverted 6 nucleotide sequences on both sides of the +1 adenine and the original 3 nucleotides sequences flanking the inverted sequence. In these constructions, the distances between the normal cap sequence and the inserted new 'caps', the TATA box and the new 'cap' are 24 nts and about 25 nts, respectively. In Fig 2A, the octamer is also inverted in one construction (No. 4).

Quantitation of transcripts

Transcription was tested by transient transfections of cultured cell lines. 10μg test plasmid and 1μg reference plasmid were transfected by Calcium Phosphate-DNA precipitate procedure. Two days later, cytoplasmic RNA was extracted and analysed. To quantitate RNA transcripts by RNase protection, we prepared radioactive RNA probe from SP6βTS (30) which contains the SP6 promoter and inverted β-globin coding sequence downstream of it. The probe can hybridize with the first 138 nts of the transcript, starting from the initiation site. For other constructions with altered order promoters and different cap sequences, we inserted the appropriate oligos into 'RsaI-SauI T7', a Bluescript plasmid (26) from which we could prepare the probes which detect the specifically initiated transcripts for each construct.

Transcripts were quantitated by excising the radioactive test and reference bands from the gel and measuring by Cerenkov counting. The efficiency of transcription was normalized to the cotransfected 'OVEC ref' plasmid (30). In the experiments with altered cap sites, the transcription efficiency from the β-globin initiation sequence was taken as 100%.

RESULTS

Transcription can be stimulated by inverted TATA boxes or a symmetrical TATA box

Since almost all TATA boxes are asymmetrical, we wanted to test the effect of inverting such TATA boxes. We selected several TATA boxes (Figure 1A) derived from the adenovirus major late promoter (20), immunoglobulin kappa light chain promoter (31), Ig heavy chain promoter (32) and β-globin promoter (21). We used a vector construct derived from recombinant plasmid sβ1 (26) with the β-globin coding sequence and the SV40 enhancer inserted downstream of the reporter gene (Figure 1A), and tested our constructs in HeLa cell and B cell lines (BJA-B and Namalwa). The basic promoter construction contained three elements, namely an upstream sequence, a TATA box, and an initiation (cap) site. The upstream sequence used was an 'octamer' sequence ATTTGCAT (3) from an immunoglobulin kappa gene. To take into account the possible effect of the flanking nucleotides, the TATA boxes together with 6 nucleotides on either side were inverted. The distances between the octamer sequence and the TATA box, and between the TATA box and the cap site were kept constant.

All of these constructions were transcribed in HeLa or B lymphocyte-type cells, although to a variable extent. Somewhat unexpectedly, the asymmetrical consensus TATA box (TATAAAA) from the adenovirus major late promoter worked well in either cell type when inverted (Figure 1, lanes 1 & 2; data in B cells not shown). In fact, in HeLa cells, promoters containing the inverted kappa TATA box and the inverted heavy chain TATA box worked up to 4 fold better than the promoters with TATA boxes in their natural orientation (Figure 1, lanes 3–6). In B cells, the kappa TATA box worked equally well in either orientation, whereas the VH TATA box was twice as active when inverted (not shown). By contrast, the β-globin TATA box in its natural orientation was highly active in both HeLa (Figure 1, lane 9) and B cells, while its inversion reduced activity in both cell types (Figure 1, lane 10). We verified that trancription in vitro from both normal and inverted TATA box promoters was found to be highly sensitive to α-amanitin, hence due to RNA polymerase II activity (data not shown).

We also analyzed a symmetrical TATA box (TATATATA) found in the mouse β casein gene (33), and the adenovirus E4 promoter (34). We modified the flanking sequences so that they were also of inverted symmetry: TACGCGTATATATACGCGTA. This TATA box, which for obvious reasons did not have to be inverted, was highly active and functioned as effectively as the β-globin TATA box in all three cell lines (Figure 1, lane 7). This result is reminiscent of an experiment in yeast, where mutation of TATAAA to TATATA resulted in a promoter which retained considerable activity (35).

When the TATA box was eliminated by a cluster of point mutations, the transcript level was strongly reduced in all three cell lines (Figure 1, lane 11 and data not shown).

Taken together, it appears that most TATA boxes can work in either orientation, and that the asymmetry observed in most TATA sequences is not a major determinant of transcriptional polarity. Therefore, we next addressed the possible contribution of upstream sequence motifs and initiation (cap) sites to transcription polarity.

Upstream sequence/TATA box order is a determinant of transcriptional polarity

We changed the order of elements in our promoter constructs from octamer/TATA/cap to TATA/octamer/cap (Figure 2A).

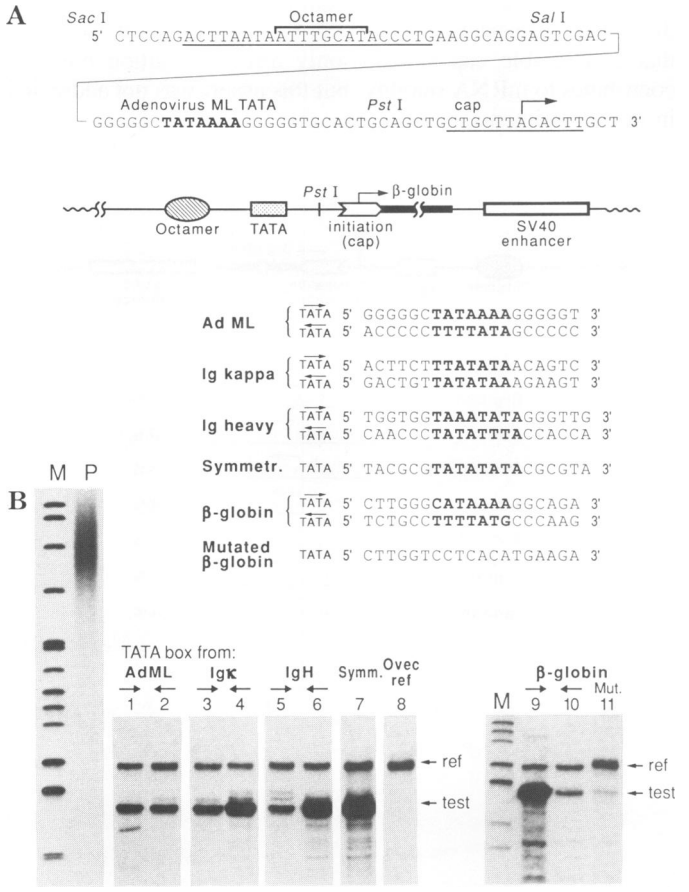

Fig. 1. Transcription from constructions with different TATA boxes. (A) The structure of the reporter gene plasmid and the sequences of different TATA boxes and their inversions are shown. The TATA boxes together with 6 flanking nucleotides on either side were inverted. The distance between the TATA box and cap site of the β-globin coding sequence is the same as in the promoter of β-globin gene. (B) Transcription efficiency of different promoter constructions. M and P indicate marker DNA and untreated RNA probe, respectively. Cytoplasmic RNA samples from Hela cells cotransfected with reporter gene plasmid and reference gene plasmid were analyzed by RNase protection analysis. Transcripts from the reporter gene (test) and from the reference gene OVEC-ref (30) are indicated by arrows in the right side of the gel. The orientations of TATA boxes are shown above each lane. The rightward arrow indicates natural TATA box orientation. The adenovirus TATA box was equally strong in either orientation (lanes 1 and 2), while Igκ and IgH TATA boxes were more active when inverted (lanes 3, 4 and 5, 6). Unlike other inverted TATA boxes, the inverted β-globin TATA box was poorly active (lanes 9 and 10). The promoter with a mutated TATA box was even weaker (lane 11).The Ovec ref (lane 8) served as a control that there was no influence of it on the lower position where test signals were.

This construction failed to give any transcripts detectable by our probe (Figure 2B, lane 2). We wondered whether the failure of the TATA/octamer/cap construction was due to a displacement of the initiation sequence relative to the TATA box. Therefore, an additional cap site was inserted to make the construction: TATA/cap/octamer/cap (Figure 2C, construction 6).

While the previous TATA/octamer/cap yielded no transcripts detectable by our probe, TATA/cap/octamer/cap gave a low but significant level of transcripts from the cap site closest to the TATA box (Figure 2D, lane 6, signal at 'test 2'). This showed that the position of the cap site relative to the TATA box was important. As a control, we also added another upstream sequence to make the construct: octamer/TATA/cap/octamer/cap (Figure 2C, construction 7). This restored the natural configuration octamer/TATA/cap, leaving an additional octamer/cap further downstream. With this construction, transcription was several fold increased from the cap site closest to the TATA box (Figure 2D, lane 7).

To further test the hypothesis that the order upstream box /TATA box was important for transcriptional polarity, we made the constructions shown in Fig 2A, No.3 and 4. The cap site was replaced by a random sequence, and either the octamer motif

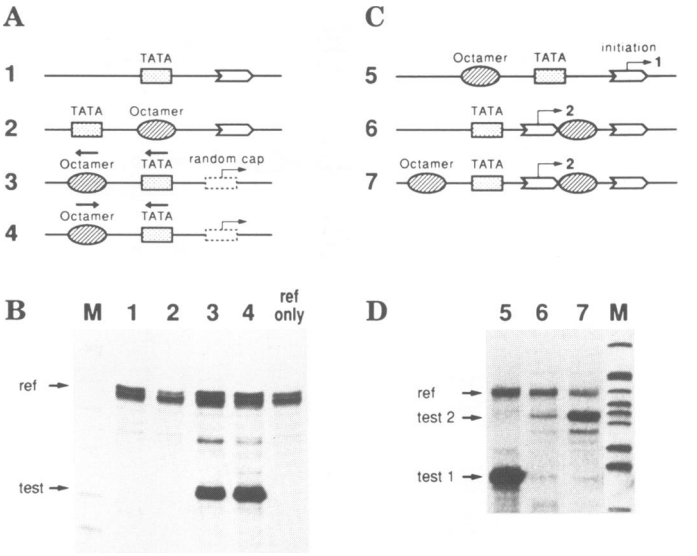

Fig. 2. The order: upstream element/TATA box is a determinant of transcription polarity. All octamer and TATA boxes without any arrow above have natural orientation. The constructions in panel (A) contain the Igκ TATA box. The distance between the 'octamer' sequence and TATA box in construction 1 is the same as in construction 3 and 4. In construction 3 & 4, the cap site was replaced by a random sequence, and the Igκ TATA box (construction 3) or both TATA and upstream 'octamer' motif (construction 4) were inverted. (B) Transcription of constructions 1–4 in HeLa cells. The two configurations 'octamer/TATA (inverted)/random cap' and 'octamer(inverted)/TATA(inverted)/random cap' are recognized as promoters, with transcription starting at the random sequence (lanes 3 and 4). Constructions 1 and 2, by contrast, are not transcribed from the consensus cap site (lanes 1 and 2). Constructions in panel (C) contain a symmetrical TATA box. In constructions 6 & 7 (not drawn to scale), all 'octamer' sequences are at the same distance from the TATA box. (D) Transcription of constructions 5–7 in HeLa cells. When the β-globin initiator sequence was inserted between the symmetrical TATA box and the octamer sequence, transcription was initiated from the newly inserted initiation site (test 2) rather than from the natural position (test 1). Transcription was strongly stimulated by the placement of a second octamer sequence upstream of the TATA box (lane 7). In all cases, RNase protection analysis was done with RNA probes whose 5' end was 10 nucleotides downstream of the TATA box.

and TATA box (construction 3) or only TATA box (construction 4) was inverted. Both of these promoters were transcribed (Figure 2B, lanes 3 and 4). So far, we could not determine with any construction whether or not there was transcription in the 'wrong' or 'leftward' direction. This was rigorously analyzed by constructing two divergent β-globin reporter genes (β1 < β2) under the control of a centrally located single promoter (Figure 3). In this series of constructions, the fractions of rightward to leftward transcription could be determined with the same RNA probe for each transfection experiment. The same three promoters above were used: octamer/TATA box, TATA box/octamer, and random sequence/TATA box/octamer (Figure 3A). Not unexpectedly, transcription followed the order of upstream element→TATA box→initiation site irrespective of the orientation of each element (Figure 3B, lane 1: β_1; lane 2: β_2). Even when the leftward initiation site was replaced by the random sequence (Figure 3A, No.3), transcription ran leftwards and initiated within the random sequence (Figure 3B, lane 3: β_2). This result confirms that the transcription apparatus is sensitive to the relative order of the octamer and TATA elements, rather than their individual orientation, in choosing the polarity of transcription.

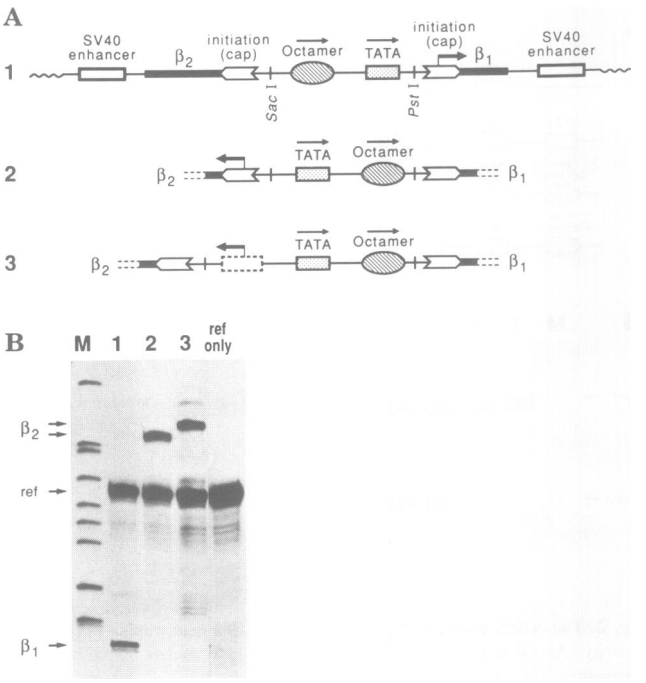

Fig. 3. Leftward transcripts detection confirms that the order: 'upstream element/TATAbox' is a determinant of transcription polarity. (A) The structure of the plasmids used to detect transcription in both directions. Two β-globin coding sequences of different lengths where β1 is much shorter than β2 with natural initiation (cap) sites are on either side of the central promoter region. Further down/upstream of each β-globin sequence is a SV40 enhancer. The order: octamer/TATA box is changed to TATA box/octamer in construction 2. The upstream initiation site is moved further upstream by insertion of a random sequence in construction 3. (B) Transcription followed the direction of octamer→TATA box, without being affected by the individual orientations of the TATA box or octamer elements (β2 in lane 2). Even when the left initiation site was replaced by the random sequence in the changed order promoter, the transcription went exclusively leftward (β2 in lane 3). The rightward and leftward transcriptions are indicated by arrows β1 and β2 on the left side of the gel, respectively. M is DNA molecular weight marker. The signals were normalized by OVEC-ref. 'Ref only' served as a control. RNase protection was the same as described in Figure 2.

The initiation (cap) site, when placed at the correct distance from the TATA box, contributes to transcription efficiency

Finally, we also wanted to investigate the role of the initiation (cap) site. For these experiments we used the construction containing the inverted kappa TATA box since it functions efficiently in HeLa cells. Between the TATA box and the cap site of the β-globin gene, the following sequences were inserted (Figure 4A): the β-globin cap sequence (21), the initiator sequence of the TdT gene (27), a Drosophila cap sequence from cuticle protein III (28, 29), two mutated β-globin cap sequences, the inverted β-globin cap sequence (6 flanking nucleotides on both sides of the start nucleotide), and a random sequence (also used in Figure 2A, No. 3 & 4, and in Figure 3). In all cases, the new initiation sites were dominant over the original β-globin initiation sequence which was rendered non-functional by being shifting to about 50 bp downstream from the TATA box. Since the new cap site extends into the RNA-coding region, it is possible that a favorable cap site not only affects initiation but also contributes to mRNA stability, but this aspect was not addressed in our experiments.

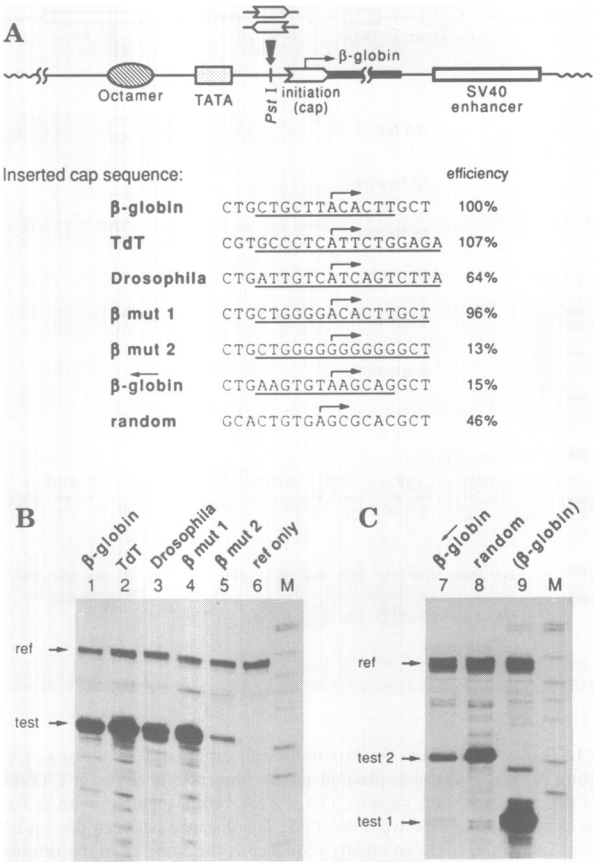

Fig. 4. The sequence of the initiation (cap) site affects the transcript level in HeLa cells. (A) Sequences which were inserted between the inverted kappa TATA box and β-globin cap site. (B) and (C) RNase protection using distinct RNA probes spanning either the inserted sequences or the original β-globin initiation site. Transcription was initiated only from sites upstream of the intact β-globin cap site (about 25 nucleotides downstream from the TATA box). Good initiators such as the β-globin cap, TdT initiation and Drosophila conserved caps gave strong transcription (lanes 1–4 in B, and 8 in C), while the completely mutated (G string) sequence and the inverted cap site gave a low transcript level (about 1/7th of maximum) (lane 5 in B, and 7 in C). The lane designated (β-globin) (lane 9 in C) is the construction without any insertion. M indicates marker DNA. 'Ref only' serves as a control.

The initiator from the terminal deoxynucleotidyltransferase gene (TdT), a paradigmatic mammalian initiation site, was functional and comparable in efficiency to the β-globin initiator region (Figure 4B, lanes 1 and 2). The Drosophila initiation sequence, which is highly conserved in different Drosophila genes, caused initiation to occur at several sites (+3, +1, −3), with an overall efficiency of about 100%, including 64% initiation at the +1 position (Figure 4B, lane 3; weaker exposure not shown). Two mutant β-globin cap sequences were also tested. In the first mutation, three nucleotides before the adenine initiation site were changed to G, yielding the sequence 'CTGGGGACACTTGCT'. Surprisingly, the transcription efficiency remained at 95% (Figure 4B, lane 4). However, when the sequence was further mutated to 'CTGGGGGGGGGGCT', the level of transcripts dropped to 1/7th of that from the wild type cap site. Despite the lack of A residues, accurate initiation occurred at the expected distance from the TATA box (Figure 4B, lane 5). The inverted β-globin cap sequence also reduced transcription severely, most likely because it was purine-rich, similar to the string of G residues. The randomized cap sequence, which was different from the consensus initiation sequence in either orientation, only reduced the level of transcripts by half (lanes 7 and 8 in Figure 4C).

DISCUSSION

Transcriptional polarity is determined by upstream sequence/TATA box order

Many mammalian promoters contain, in addition to one or more binding sites ('upstream sites') for regulatory transcription factors, a TATA box which binds the transcription factor TFIID, and a defined initiation site (cap site). Such promoters can also be subject to remote control by enhancer regions that contain clusters of binding sites for transcription factors (reviewed in 1, 36). Regulated promoters of this kind are found in all eukaryotes, including fungi, while TATA box-independent 'CpG islands' from which transcription of housekeeping genes is initiated (37) are only found in vertebrates and some higher plants.

We asked the question: What is the major determinant of the polarity of the promoter? Could it be an inbuilt polarity of an asymmetrical TATA box sequence (typically TATAAAA), or the order: upstream/TATA, or is the correct order of all three elements required? We were surprised to find that some inverted TATA boxes functioned as well as, or even better than the natural TATA boxes. This could mean that the TFIID factor that recognizes the TATA box can bind to many AT-rich sequences of some 7−8 bp length. Binding can be facilitated by the presence of a tetranucleotide TATA (or CATA) sequence, even though not all known TATA boxes contain it. Such a TATA motif is also present in all our inverted TATA boxes except β-globin, which works poorly and has TATG instead. When the TATG was mutated to TATA in β-globin inverted TATA box, the transcriptional level increased 2 to 3 fold, but was still only 1/3 of the level of natural β-globin TATA box (data not shown).

Our studies of changed order promoters clearly show that it is primarily the relative order: upstream motif/TATA box that determines the polarity of transcription, while the nucleotide sequence of the cap site region affects the precision and efficiency of initiation. This simple type of polarity determination may be overridden by special promoter configurations. For example, in some cases 'upstream' sites are found downstream of the TATA box (38), and in one case a TATA-like sequence contributes to transcriptional efficiency from a position immediately downstream of a viral cap site (39). The elucidation of these examples may have to await more detailed factor binding/mutagenesis experiments. Such cases should however not be confused with those where transcription from a 'standard' promoter is boosted by an enhancer located within an intron a few hundred base pairs further downstream (40). Our findings may also help to explain why in mammals there always seems to be at least one upstream site close to the TATA box (4): A TATA/cap configuration by itself may not suffice to ensure unidirectional transcription (In our experiments, we did not address the question whether or not the inefficient 'no upstream' promoter (TATA-cap) was affected by the orientation of the TATA box). This situation may not hold for lower eukaryotes, where the initiation site is more important (see below).

The cap site tolerates sequence alterations but requires proper spacing from TATA box

Recently it was shown that the initiator (cap) site from the terminal deoxynucleotidyltransferase (TdT) gene that fits to a loose consensus PyPyCAPyPyPyPyPy is sufficient to serve as a precise, albeit not very efficient initiation site even in a promoter that lacks a TATA box (27). The significance of the cap site region in mammals was overlooked for a long time, because early studies showed that the TATA box is the dominant element in the selection of the initiation point. For example, when the cap site is eliminated, other sequences about 25 bp downstream of the TATA box were recruited as new initiation points (42, reviewed in 43). Previous studies in our institute have addressed more thoroughly the effect of spacing on cap site usage. It was found that even in the absence of a TATA box, residual transcription can be quite efficient if a strong upstream site is appropriately spaced relative to a cap site (44). Our experiments show that a good cap site is no longer used when it is moved 24 bp downstream of its normal location relative to the TATA box (Figure 4, lane 9 compared to lanes 1−8). In addition, we find that although diverse sequences can serve as good cap sites, initiation is inhibited by certain unfavorable sequences, such as a string of G residues. In contrast to our findings with initiation sites in mammalian cells, the initiation regions in Drosophila and yeast appear to be less permissive. For example, in many Drosophila genes, the initiation region is highly conserved with a consensus PyATCA$^{G}/_{T}$TPy (initiator A underlined)(28). In yeast, the initiation sequence seems particularly important: The location of the TATA box can vary considerably without affecting the site of initiation (45, 46).

The concept of strict spacing and loose consensus sequence which can be derived from our data with TATA boxes and cap sites is not generally true for all components of a mammalian promoter. We have also tested a set of octamer/TATA/cap promoters where the distance between octamer site and globin TATA box was reduced stepwise. In 10 constructions where the spacing varied from 24 to 5 nucleotides, transcriptional efficiency remained unchanged. With a spacing of only 3 or 1 nucleotides, or an overlap of sites by 3 nucleotides, transcription was strongly reduced, presumably as a result of steric hinderance between octamer binding factor and TFIID (P. Matthias, P. Künzler and W. Schaffner, unpublished observations; see also 47). From this result, it appears that the spacing between the octamer upstream motif and the TATA box is less constrained than the TATA/initiation site spacing. However, an element of strigency is ensured by the octamer sequence itself which tolerates little sequence alteration without considerable loss of function (47).

Why the multitude of different gene-specific TATA boxes?

In view of our findings that a variety of different TATA sequences serve as efficient TATA boxes, it seems paradoxical that gene specific TATA box sequences are conserved in evolution.

From our data, one might deduce that all it takes to make a reasonable TATA box is a cluster of some 7 As and Ts flanked by more GC-rich sequence, and that the actual sequence would not be conserved. However, quite the opposite is true. For example, instead of TATAAAA, the β-globins of all mammals have a deviant CATAAAA box (48). The many mammalian α-interferon genes all have a characteristic TATTTAA box (49, 50), while β interferon genes have a TATAA$^{A}/_{G}$T box (51). a and β casein genes have TTTAAATA and TATATATA boxes, respectively (33). Why have these sequences been conserved? It has been speculated that there are variants of TFIID for specific interaction with different TATA boxes, but so far there is only limited evidence for this (52). Another possibility would be that TFIID assumes a characteristic conformation upon binding to a given TATA box, such that there is an optimal interplay with factors bound to upstream sites. Some TATA boxes may also interact with sequence-specific factors other than TFIID. These factors might synergize with TFIID to activate transcription, or, they might repress transcription in nonpermissive cell types by blocking the TATA box. Such special functions may not be detected in our transient expression assays that measure basic transcription functions. Whatever the reason for gene-specific TATA box sequences, the fact that they have been conserved between species throughout 80 million years of mammalian evolution strongly suggests that they are required for optimal gene function.

ACKNOWLEDGMENTS

We are grateful to Dr. Hans-Peter Müller for many helpful discussions; Drs. Pamela J. Mitchell, Charles Weissmann and David Arnosti for critical reading of the manuscript. We thank Dr. David Arnosti for a kind gift of HeLa nuclear extract which was used in the RNA polymerase II test, and Mr. Hanspeter Gerber for helpful advice of α-amanitin test. We thank Fritz Ochsenbein for expert graphic work. This work was supported by the Kanton of Zürich and the Swiss National Science Foundation.

REFERENCES

1. Müller, H-P. and Schaffner, W. (1990) TIG, 6, 300–304.
2. Gidoni, D.,Kadonaga, J. T., Barrera-Saldana, H., Takahashi, K., Chambon, P., Tjian, R. (1985) Science, 230, 511–517.
3. Parslow, T. G., Blair, D. L., Murphy, W. J., and Granner, D. K. (1984) Proc. Natl. Acad. Sci., USA, 81, 2650–2654.
4. Sawadogo, M. and Roeder, R. G. (1985) Proc. Natl. Acad. Sci., USA 82, 4394–4398.
5. Cavallini, B., Faus, I., Matthes, H., Chipoulet, J. M., Winsor, B., Egly, J.-M., and Chambon, P. (1989) Proc. Natl. Acad. Sci. USA 86, 9803–9807.
6. Eisenmann, D. M., Dollard, C., and Winson, F. (1989) Cell, 58, 1183–1191.
7. Hahn, S., Buratowski, S., Sharp, P. A., and Guarente, L. (1989) Cell, 58, 1173–1181.
8. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F., and Tjian, R. (1990) Cell, 61, 1179–1186.
9. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M., and Roeder, R. G. (1990) Nature, 346, 387–390.
10. Horikoshi, M., Wang, C. K., Fuji, H., Cromlish, J. A., Weil, P. A., and Roeder R. G. (1989) Nature, 341, 299–303.
11. Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R., and Berk, A.J. (1990) Science, 248, 1646–1650.
12. Peterson, M. G., Tanese, N., Pugh, B. F., and Tjian, R. (1990) Science, 248, 1625–1630.
13. Lewin, B. (1990) Cell, 61, 1161–1164.
14. Buratowski, S., Hahn, S., Sharp, P. A., and Guarente, L. (1988) Nature, 334, 37–42.
15. Cormack, B. P., Strubin, M., Ponticelli, A. S., and Struhl, K. (1991). Cell, 65, 341–348.
16. Gill, G., and Tjian, R. (1991) Cell, 65, 333–340.
17. Pugh, B. F. and Tjian, R. (1990) Cell, 61, 1187–1197.
18. Smale, S. T., Schmidt, M. C., Berk, A. J., and Baltimore, D. (1990) Proc. Natl. Acad. Sci., USA 87, 4509–4513.
19. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) Science, 209, 1405–1414.
20. Concino, M. F., Lee, R. F., Merryweather, J. P., and Weinmann, R. (1984) Nucleic Acids Res., 12, 7423–7433.
21. Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) Cell, 32, 695–706.
22. Jones, K. A., Luciw, P. A., and Duchange, N. (1988) Genes Dev., 2, 1101–1114.
23. Talkington, C. A., and Leder, P. (1982) Nature, 298, 192–195.
24. Tokunagan K., Hirose, S., and Suzuki, Y. (1984) Nucleic Acids Res., 12, 1543–1558.
25. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
26. Heuchel, R., Matthias, P., and Schaffner, W. (1989) Nucleic Acids Res., 17, 8931–8947.
27. Smale, S. T. and Baltimore D. (1989) Cell, 57, 103–113.
28. Hultmark, D. R., Klemenz, R., and.Gehring, J. W (1986) Cell, 44, 429–438.
29. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J., and Davidson, N. (1982) Cell, 29, 1027–1040.
30. Westin, G., Gerster, T., Müller, M. M., Schaffner, G., and Schaffner, W. (1987) Nucleic Acids Res., 15, 6787–6798.
31. Seidman, J. G., Max, E. E., and Leder P. (1979) Nature, 280, 370–375.
32. Ballard, D.W., and Bothwell, A. (1986) Proc. Natl. Acad. Sci., USA 83, 9626–9630.
33. Yoshimura, M., and Oka, T. (1989) Gene, 78, 267–275.
34. Pei, R., and Berk, A. J. (1989) J. Virology, 63, 3499–3506.
35. Chen, W., and Struhl, K. (1988) Proc. Natl. Acad. Sci., USA, 85, 2691–2695.
36. Müller, M. M., Gerster, T., and Schaffner, W. (1988) Eur. J. Biochem., 176, 485–495.
37. Sehgal, A., Patil, N., and Chao, M. (1988) Mol. Cell. Biol., 8, 3160–3167.
38. Stuve, L. L. and Myers, R. M. (1990) Mol. Cell. Biol., 10, 972–981.
39. Carcamo, J., Maldonado, E., Cortes, P., Ahn, M. H., Kasai, Y., Ilho-ha, Flint, J., and Reinberg, D. (1990) Genes Dev, 4, 1611–1622.
40. Banerji, J., Olson, L. and Schaffner, W. (1983) Cell, 33, 729–740.
41. Schatt, M. D., Rusconi, S., and Schaffner, W. (1990) EMBO J., 9, 481–487.
42. Grosschedl, R. and Birnstiel M. L. (1980) Proc. Natl. Acad. Sci., USA 77, 1432–1436.
43. Breathnach, R., and Chambon, P. (1981) Ann. Rev. Biochem., 50, 349–383.
44. Wieland, S., Schatt, M. D., and Rusconi, S. (1990) Nucleic Acids Res., 18, 5113–5118.
45. Chen, W. and Struhl, K. (1985) EMBO J., 4, 3273–3280.
46. Struhl, K. (1987) Cell, 49, 295–297.
47. Wirth, T., Staudt, L., and Baltimore, D. (1987) Nature, 329, 174–178.
48. Efstratiadis, A., Posakony, J. W., Maniatis, T., et al. (1980) Cell, 21, 653–668.
49. Henco, K., Brosius, J., Fujisawa, A., Fujisawa, J.-I., Haynes, J. R., Hochstadt, J., Kovacic, T., Pasek, M.,. Schamböck, A, Schmid, J., Todokoro, K., Wälchli, M., Nagata, S., and Weissmann, C. (1985) J. Mol. Biol., 185, 227–260.
50. Kelly, K. A. and Pitha, P. M. (1985) Nucleic Acids Res., 13, 825–839.
51. Weissmann, C., and Weber, H. (1986) The interferon genes. Progress in Nucleic Acid Research and Molecular Biology, 33, 251–300.
52. Tamura, T., Sumita, K., Hirose, S., and Mikoshiba, K. (1990) EMBO J., 9, 3101–3108.