Volume 17 Number 18 1989 Nucleic Acids Research

Transcription of a human U6 small nuclear RNA gene *in vivo* withstands deletion of intragenic sequences but not of an upstream TATATA box

Gary R.Kunkel*+ and Thoru Pederson

Cell Biology Group, Worcester Foundation for Experimental Biology, 222 Maple Avenue, Shrewsbury, MA 01545, USA

Received April 17, 1989; Revised May 31, 1989; Accepted August 16, 1989

ABSTRACT

Most eukaryotic genes transcribed by RNA polymerase III contain internal control regions. U6 small nuclear RNA genes are transcribed by RNA polymerase III but are unusual in that, at least *in vitro*, their expression does not require intragenic sequences. Here we show that this is true as well *in vivo*. A human U6 gene devoid of all but the first 6 and last 10 base-pairs was expressed efficiently after transfection into human 293 cells. We also report data extending the previous identification of 5′ flanking sequences important for human U6 gene transcription. Deletion-substitution of a 10 base-pair upstream sequence encompassing the TATATA element (−29 to −24) abolished U6 transcription. A double point mutation in the middle of this element (TATATA → TAGCTA) reduced U6 transcription but not to the extent brought about by TATATA deletion-substitution. These results establish that, *in vivo*, transcription of human U6 small nuclear RNA is independent of intragenic sequences between nucleotides 6 and 98, and requires the upstream TATATA box.

INTRODUCTION

Unlike the genes for the other abundant small nuclear RNAs (U1−U5), U6 RNA is apparently transcribed by RNA polymerase III (1−3). Irrespective of this obvious difference in the transcription machinery used to synthesize U6 versus U1−U5 RNAs, U6 genes from a variety of organisms contain 5′ flanking transcription control elements similar or identical to ones adjacent to polymerase II-transcribed RNA genes (3−8). In the human U6 gene these elements are located within the first 67 5′ flanking nucleotides and upstream of nucleotide −149 (6). An additional 5′ flanking sequence element including a TATA sequence is present in all U6 genes sequenced so far (TATATA between −29 and −24, inclusive, in the human U6 gene), and has been shown to be important for efficient polymerase III-directed transcription of a *Xenopus borealis* U6 gene (8) as well as a human 7SK RNA gene (9).

The presence of pol II-like transcription control elements upstream from the U6 and 7SK genes identifies these transcription units as novel members of the class III gene family (10,11). Like other genes transcribed by pol III, the human U6 gene has an internal element similar to the intragenic control region A block (1). The A block-like internal region of a *Xenopus* U6 gene can be deleted with only minor effect on transcription in injected oocytes (3). In addition, deletion of all but the initiating G nucleotide from the mouse U6 gene has little effect on *in vitro* transcription (7). However these experiments leave open the possibility that a heretofore undiscovered internal element controls transcription of the U6 gene *in vivo*. Here we demonstrate that extensive internal deletion of the human U6 gene, leaving intact only the cap site region and the normal U6 terminator, results in continued

efficient expression after transfection of human cells. We also show that an upstream element encompassing the TATATA sequence is critically important for *in vivo* expression.

MATERIALS AND METHODS

Cells

Human 293 cells were grown in monolayers in Dulbecco-modified Eagle's medium with 10% fetal calf serum.

DNA constructions

The parent plasmid DNA used for construction of internal deletions was pGEM/U6 (1). Sequences were excised using exonuclease III digestion and recombined at added *Xho*I linkers as described previously (6). DNA fragments restricted from plasmid pGEM1 (Promega Biotec, Madison, WI) by *Hae*III digestion were blunt-end-ligated at the *Xho*I site of the deleted U6 constructions after opening with *Xho*I and trimming of 5′ overhangs with S1 nuclease.

U6 templates with mutations or deletion/substitution in the TATATA upstream element were generated by oligonucleotide-directed mutagenesis of the U6 maxigene (6) inserted into a M13mp18 vector grown in a dut$^-$ ung$^-$ host (12,13). The oligonucleotides used to construct these mutants were:

Double point mutant: TGGCTTTAGCTATCTTGT

Deletion/substitution: ATTTCGATTTCTTTGGGAGGTACCGACTTGTGGAAAGGACG

Mutant U6 DNA inserts from phage RF DNA were excised and recloned into the pGEM1 vector for transfection experiments.

All plasmid constructions were verified by restriction enzyme digestion and dideoxy sequencing.

In vitro transcription, transfection, RNA isolation and primer extension

Transcription of plasmid templates in a HeLa cell S100 extract, transfection of 293 cells, isolation of RNA, extraction of transfected plasmid DNA by Hirt lysis and primer extension analysis of U6 gene transcripts were carried out as described previously (6). Oligodeoxynucleotides used for primer extension were:

'U6 primer': TATGGAACGCTTCAC

'348 primer': TATGGCCCCTGCCACCAT.

RESULTS

U6 genes with extensive internal deletions are expressed efficiently in vivo.

We constructed several human U6 gene templates with extensive internal deletions and compared the transient expression from such templates to that of the previously characterized U6 maxigene (6). The first two templates constructed lacked U6 gene sequences from nucleotides +24 to +86 (Figure 1A). The deleted internal sequence block was substituted with restriction fragments from pGEM1 vector DNA to create new U6 maxigenes whose transcripts would be detected as longer primer extension products clearly separable from any premature primer extension stops occurring on highly abundant endogenous U6 RNA (see Figure 1A). Primer extension analysis of RNAs produced from these internal deletion templates (dl(+24,+86)) using the 'U6 primer' (see Materials and Methods) showed that transient expression is very similar to the level from the U6 maxigene containing the entire U6 sequence (arrowheads, Figure 1B). This shows that *in vivo* transcription of the human U6 gene does not depend on the A block consensus sequence (+48 to +59) or on any other element residing between +24 and +86.

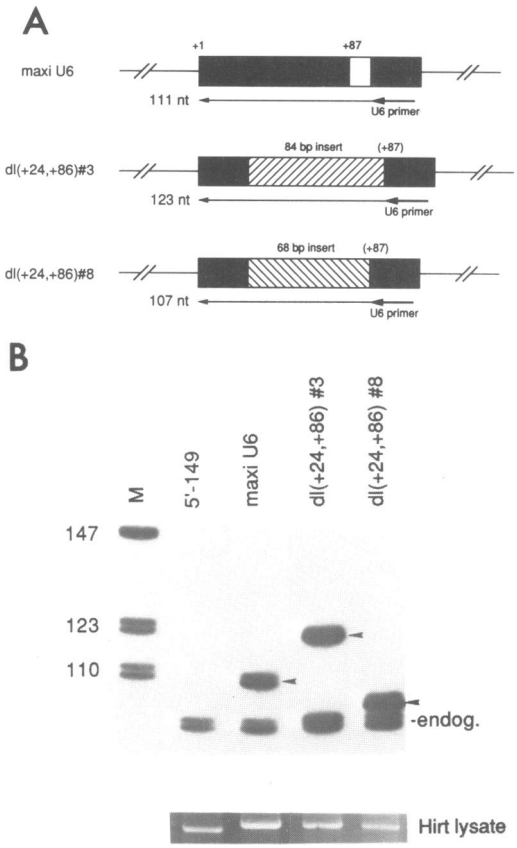
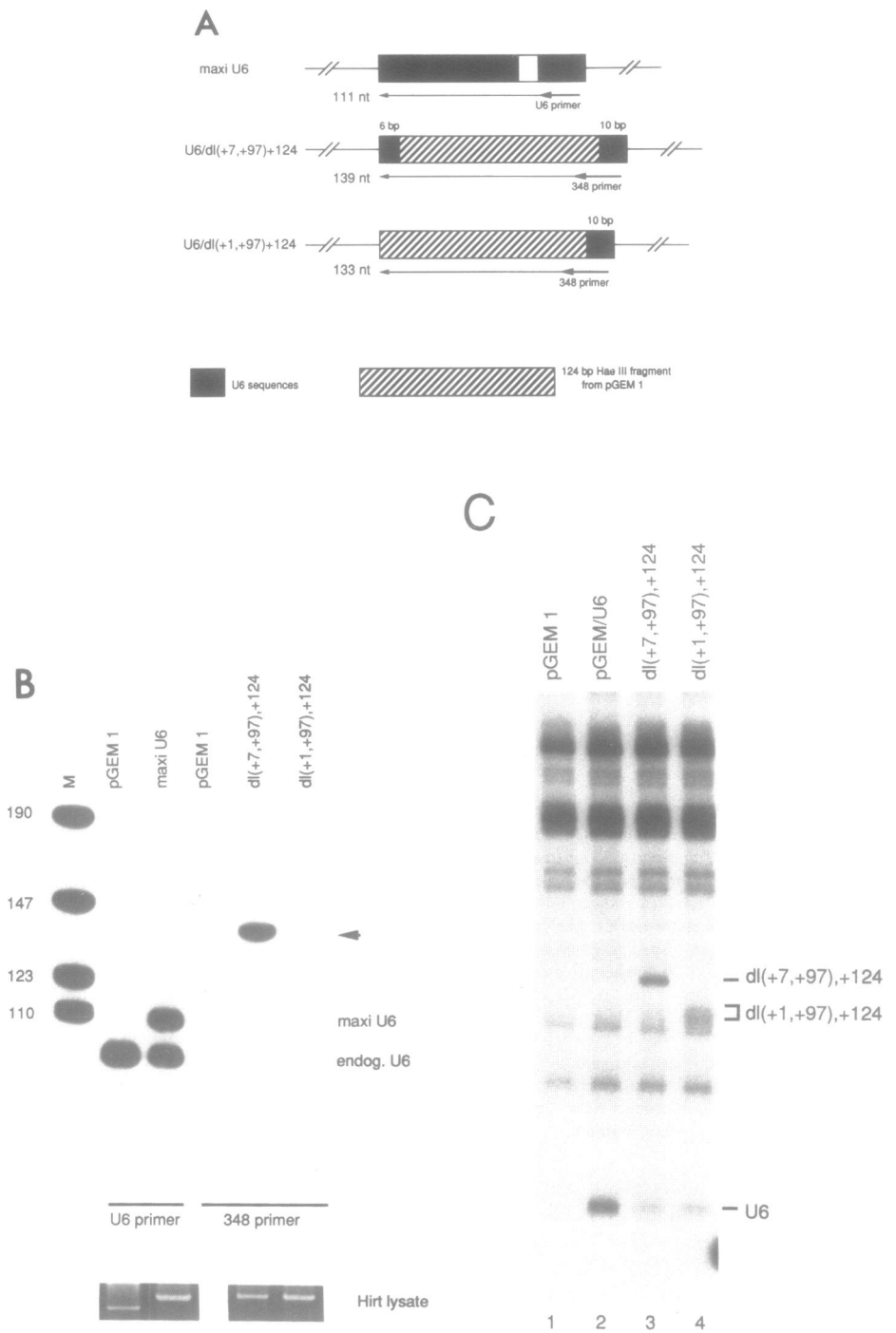


Figure 1. Human U6 genes containing a deletion of intragenic sequences between +24 and +86 are efficiently expressed *in vivo*. *A*. Templates used for transfection studies. Portions filled in black denote U6 coding sequences, the unfilled region (+87) represents the maxigene insert (see ref. 6) and the stippled regions correspond to pGEM1 inserts. Each deletion template, containing a different pGEM1/*Hae*III restriction fragment, was constructed as described in Materials and Methods. Under each diagram are indicated the position of U6 primer hybridization and the expected length of primer extension product. These three templates all contain the same U6 gene 5′ and 3′ flanking sequences. *B*. Primer extension analysis of RNAs transcribed from U6 templates and electrophoresis of maxigene plasmid DNAs from transfected 293 cells. (Top panel) U6 primer-extension products fractionated on a 10% polyacrylamide/8.3 M urea gel. Lane M: ^{32}P-labeled *Msp*I restriction fragments of pGEM1 DNA run as markers. Lane 5′−149: primer extension of RNAs produced from cells transfected with a U6 maxigene lacking 5′ flanking sequences upstream of −149 (see ref. 6). Endog.: primer extension product from endogenous U6 RNA. (Lower panel) U6 plasmid DNAs isolated by Hirt lysis from transfected 293 cells at the time of harvest for RNA isolation. Nucleic acids were digested with RNase A and *Eco*RI, electrophoresed on a 1% agarose gel and stained with ethidium bromide.

To extend this analysis we constructed templates with yet further deletion of internal U6 gene sequence (Figure 2A). These plasmids contained only a U6 decanucleotide sequence 3′-ward of position +97 (CCATATTTTT) and either the first 6 nucleotides of the U6 gene (dl(+7,+97)+124) or no 5′ end U6 nucleotides (dl(+1,+97)+124) (Figure 2A). A 124 bp *Hae*III restriction fragment derived from pGEM1 was inserted to replace U6 transcribed sequence. The sequence of this fragment contained no apparent similarity to


A
maxi U6
111 nt
U6 primer
6 bp
10 bp
U6/dl(+7,+97)+124
139 nt
348 primer
10 bp
U6/dl(+1,+97)+124
133 nt
348 primer
U6 sequences
124 bp Hae III fragment from pGEM 1
B
M
pGEM 1
maxi U6
pGEM 1
dl(+7,+97),+124
dl(+1,+97),+124
190
147
123
110
maxi U6
endog. U6
U6 primer
348 primer
Hirt lysate
C
pGEM 1
pGEM/U6
dl(+7,+97),+124
dl(+1,+97),+124
dl(+7,+97),+124
dl(+1,+97),+124
U6
1
2
3
4

a pol III A block element. Both constructions included the entire 5′ and 3′ flanking regions of the efficiently-expressed U6 maxigene (6). Primer extension using the 'U6 primer' verified a high level of expression from the 'wild-type' U6 maxigene (Figure 2B). In order to detect RNAs transcribed from the U6 deletion templates shown in Figure 2A a different oligodeoxynucleotide was used as a primer (designated '348 primer'; see Materials and Methods). It can be seen in Figure 2B that expression from dl(+7,+97)+124 is quite high. Quantitation of the relative levels of expression of the maxiU6 and dl(+7,+97)+124 templates is not straightforward since, of necessity, different primers were used. However the primers were kinased to the same specific radioactivity (results not shown), identical amounts of RNA were used for primer extension and the transfection efficiency of each plasmid was similar (Figure 2B, bottom). In contrast, the level of expression from dl(+1,+97)+124, a template containing the U6 5′ flanking region directly abutted to an abnormal start site sequence (CCGGCATC), is much lower and these transcripts apparently have heterogeneous 5′ ends. The low level of expression observed from dl(+1,+97)+124 may be a consequence of the instability of transcripts with altered start sites, or the start site sequence may itself directly affect the efficiency of U6 transcription (8).

We also compared the relative levels of *in vitro* transcription from these U6 deletion templates to that from a wild-type U6 template in a HeLa cell S100 extract (Figure 2C). Transcription from dl(+7,+97)+124 was comparable, albeit somewhat reduced, to the level observed with pGEM/U6 (compare band labeled 'U6' in lane 2 of Figure 2C with that labeled 'dl(+7,+97)+124' in lane 3). As in the transfection experiments, dl(+1,+97)+124 produced a heterogeneous set of transcripts (Figure 2C, lane 4). However these readily detectable *in vitro* transcripts from dl(+1,+97)+124 were synthesized at relatively higher levels than seen in the transfection experiments. Therefore instability of the dl(+1,+97)+124 transcripts is not apparent *in vitro*, at least not to an extent that would account for the drastically reduced level of expression observed *in vivo*. Since we have been unable to differentiate the relative effects of transcription and RNA stability in transfected cells, the exact role of the immediate start site sequence (+1 to +6) is still undetermined. However our results demonstrate that human U6 transcription is influenced little, if at all, by the internal sequence from +7 to +97.

We were surprised by the presence of a band produced from the U6 deletion templates migrating at the exact position of wild-type U6 RNA (see lanes 3 and 4, Figure 2C). The relative level of this band is extract-dependent, and it was not seen previously after *in vitro* transcription of U6 maxigenes with a normal 5′ flanking sequence. We found that transcription of the U6-sized band from 5′ flanking region deletion mutants of plasmids containing the U6 maxigene in the same S100 extract used for the experiment in Figure

Figure 2. Transient *in vivo* expression and *in vitro* transcription of U6 templates with extensive deletions of internal sequences. *A*. Templates used for transfection studies. Portions filled in black denote U6 coding sequences, the lengths of which are indicated above each box in the deletion constructions. Under each diagram are indicated the position of primer hybridization and the expected length of primer extension product if initiation occurs at the first nucleotide after −1. All templates contain the same extent of 5′ and 3′ flanking sequences. *B*. Primer extension analysis of RNAs produced from U6 templates (top panel) and direct analysis of plasmid DNAs from transfected cells (lower panel). The particular construction tested is noted at the top of each lane, and the primer used is indicated at the bottom. Lane pGEM1: a mock transfection with plasmid lacking an insert of human DNA. Endog. U6: primer extension product from endogenous U6 RNA. Maxi U6: primer extension product from the U6 maxigene. The arrow indicates the primer extension product from the U6/dl(+7,+97)+124 template. *C*. Transcripts synthesized from U6 templates in a HeLa S100 extract. 500 ng of plasmid template were used for each reaction and the total transcripts from each reaction were loaded on a 10% polyacrylamide/8.3 M urea gel. The mobilities of normal U6 RNA and the U6 internal deletion/substitution transcripts are indicated.

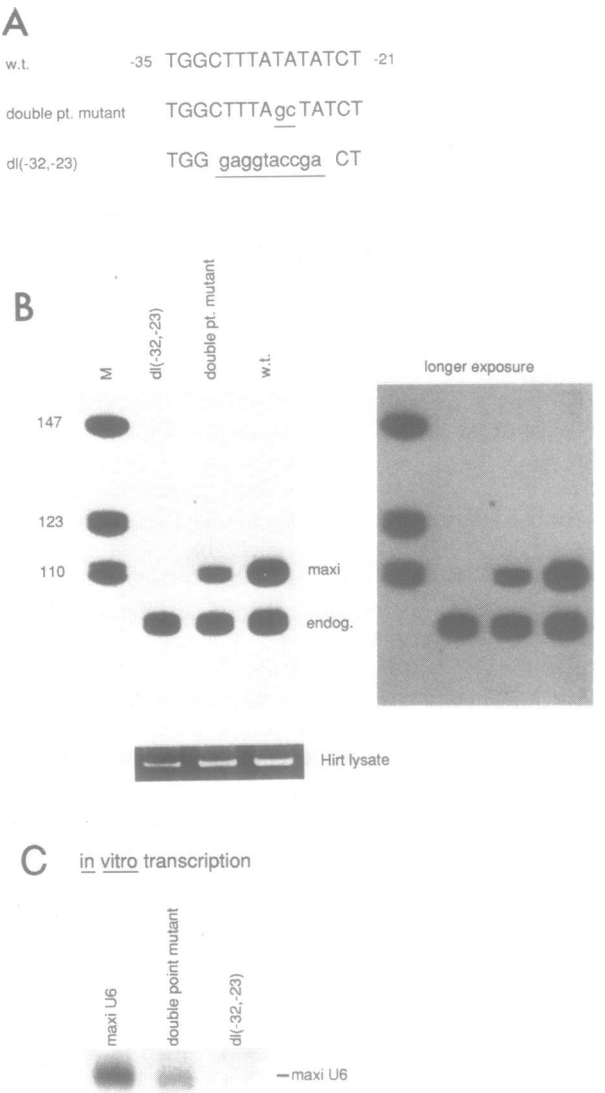


Figure 3. An upstream region encompassing the TATATA sequence is important for U6 expression. *A*. Sequence changes in the mutant templates. Oligonucleotide-directed mutagenesis was used to effect the indicated sequence changes in the TATATA region of the human U6 maxigene. *B*. Primer extension analysis of U6 RNAs produced from cells transfected with the mutant templates (top panels) and electrophoresis of plasmid DNAs isolated from transfected cells (lower panel). The particular construction tested is noted at the top of each lane. Lane M: ^{32}P-labeled *Msp*I restriction fragments of pGEM1 DNA used as markers. (w.t.) refers to cells transfected with the U6 maxigene (Figure 1A, top construction) containing a normal 5′ flanking sequence. *C*. *In vitro* transcription of U6 maxigene RNAs. 500 ng of plasmid template were used for each transcription reaction in a HeLa cell S100 extract.

2C was dependent on 5′ flanking sequence upstream of −332 (results not shown). This far upstream region is present in the U6 deletion templates, thus explaining the transcription of the U6-sized band in some preparations of S100 extract.

An element encompassing the TATATA sequence is important for in vivo expression of the human U6 gene.

We have shown previously that at least two elements in the 5′ flanking region significantly affect transient expression of transfected U6 maxigenes (6). Deletion of a distal region upstream of −149 reduces expression to a low (basal) level, whereas further deletion to −43 eliminates U6 maxigene transcription. Another possible control element includes the TATA sequence found in the proximal region of U6 genes so far isolated from many organisms (1,2,4,14−16). Our previous result demonstrating a drastic decrease of transcription *in vitro* for a U6 template in which nucleotides −28 to −13 were deleted is consistent with a role for the TATA sequence (6). RNA polymerase III-directed transcription of the *Xenopus* U6 gene is dependent on a region including the TATA element (8). In order to investigate the effect of the TATATA sequence located proximal to position −43 (−29 to −24) we constructed two mutant U6 maxigene templates (Figure 3A). As shown in Figure 3B, a U6 maxigene containing a double point mutation disrupting two nucleotides in each overlapping TATA sequence (TATATA → TAGCTA) is expressed at a decreased level. However complete removal of a 10 base-pair element including the TATATA sequence and substitution with a segment of 10 unrelated base pairs to preserve spacing of upstream elements (dl(−32,−23) in Figure 3A) almost completely eliminates expression (Figure 3B). A longer exposure of the autoradiograph shown in Figure 3B revealed a very faint band corresponding to the maxiU6 transcript, and there was no evidence for aberrantly initiated U6 transcripts with either mutant template. We also compared the S100 *in vitro* transcription of these TATATA mutants with the wild-type 5′ flanking U6 maxigene. As shown in Figure 3C, transcription of the (TATATA → TAGCTA) double point mutation was decreased but still clearly visible. However U6 maxigene synthesis from the dl(−32,−23) template was not detected.

DISCUSSION

The present experiments demonstrate that virtually all sequences within a human U6 gene are dispensable for high-level expression following transfection of deleted genes into human 293 cells. It has previously been shown that *in vitro* transcription of a mouse U6 RNA gene is not dependent on internal sequences, except possibly the initiating G nucleotide (7). We cannot address the transcriptional importance of the first 6 base-pairs in the human U6 gene because those deletions were not recovered in our subcloning procedure. Expression from a transfected gene containing the entire U6 5′ flanking region abbutted to an abnormal start site sequence (dl(+1,+97)+124) is greatly reduced and there is evidence that these transcripts have heterogeneous 5′ ends (Figure 2B,C). We cannot rule out the possibility that transcription initiation occurs in this deleted gene at a high level *in vivo*, but that these transcripts are unstable. The unusual 5′ end structure on U6 RNA (17) may have a role in metabolic stability, and we do not know whether any of these 'U6 transcripts' contain this modification. A recent study has shown that single and double point mutations in the start site region of a *Xenopus* U6 gene have a dramatic effect on the level of U6 RNA produced in injected oocytes (8). However this investigation also did not examine the relative stabilities of transcripts produced from the mutant templates.

Our most extensive deletion constructions still contained the (somewhat extended) normal

U6 terminator sequence. The primer extension assay does not address the question of faithful termination, but Northern gel blot analysis of RNA following transfection of the dl(+24,+86)#3 plasmid (Figure 1A) revealed a band of the expected size for correct termination (results not shown). *In vitro* transcription of a mouse U6 gene lacking its normal terminator stops after the first run of 4 T's encountered in the plasmid vector sequence (7).

Our results also show that a 10 bp block encompassing the TATATA sequence in the human U6 gene is of critical importance for efficient expression *in vivo*. A similar dependence of the TATA region for RNA polymerase III transcription of a *Xenopus* U6 RNA gene has recently been demonstrated (8). In addition, after this paper was submitted a report appeared demonstrating a role of the TATA region for polymerase III transcription of the human U6 gene (18). In contrast to these recent results (8,18), we did not detect an appreciable level of U6 transcripts generated independent of the TATA element. Similar to the *Xenopus* case (3), we find that the efficiency of human U6 RNA expression is only moderately affected by a 2 bp mutation in the middle of the TATATA sequence (Figure 3B). In contrast, *in vitro* transcription by polymerase III of a human 7SK RNA gene is drastically reduced by mutation within its TATATA element (9). The apparent plasticity of the U6 gene's core TATA sequence to mutation is reminiscent of the degeneracy of octamer motif sequences recognized by OBP100 *in vitro* (19). It is to be noted that there are numerous precedents for polymerase III-transcribed genes requiring upstream sequences, including a TATA element (*e.g.* 20–22). The unique feature of the vertebrate U6 (3,7; this report) and 7SK (9,23) RNA genes studied so far is the complete lack of a requirement for most, and perhaps all, of the intragenic sequences.

ACKNOWLEDGEMENTS

This research was supported by NIH grant GM-21595-14 to T.P. We thank John Goodchild for synthesizing oligonucleotides in the Worcester Foundation's central facility, supported by NIH Cancer Center core grant P30-12708, and Molly Rhoadhouse for expert assistance in gene constructions and plasmid isolations. We are grateful to Ann Kleinschmidt, John Maddalena, Sandra Mayrand, Jeffrey Patton and Jin Wang for advice and criticism throughout the course of this work.

*To whom correspondence should be addressed

+Present address: Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843, USA

REFERENCES

1. Kunkel, G.R., Maser R.L., Calvet J.P., and Pederson T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8575–8579.
2. Reddy, R., Henning D., Das G., Harless M., and Wright D. (1987) *J. Biol. Chem.* **262**, 75–81.
3. Carbon, P., Murgo S., Ebel J.-P., Krol A., Tebb G., and Mattaj I.W. (1987) *Cell* **51**, 71–79.
4. Krol, A., Carbon P., Ebel J.-P., and Appel B. (1987) *Nucleic Acids Res.* **15**, 2463–2478.
5. Bark, C., Weller P., Zabielski J., Janson L., and Pettersson U. (1987) *Nature* **328**, 356–359.
6. Kunkel, G.R. and Pederson T. (1988) *Genes Dev.* **2**, 196–204.
7. Das, G., Henning D., Wright D., and Reddy R. (1988) *EMBO J.* **7**, 503–512.
8. Mattaj, I.W., Dathan N.A., Parry H.D., Carbon P., and Krol A. (1988) *Cell* **55**, 435–442.
9. Murphy, S., DiLiegro C., and Melli M. (1987) *Cell* **51**, 81–87.
10. Sollner-Webb, B. (1988) *Cell* **52**, 153–154.
11. Folk, W.R. (1988) *Genes Dev.* **2**, 373–375.
12. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488–492.
13. Kunkel, T.A., Roberts J.D., Zakour R.A. (1987) *Methods Enzymol.* **154**, 367–382.

14. Das, G., Henning D., and Reddy R. (1987) *J. Biol. Chem.* **262**, 1187–1193.
15. Brow, D.A. and Guthrie C. (1988) *Nature* **334**, 213–218.
16. Tani, T. and Ohshima, Y. (1989) *Nature* **337**, 87–90.
17. Epstein, P., Reddy R., Henning D., and Busch H. (1980) *J. Biol. Chem.* **255**, 8901–8906.
18. Lobo, S.M. and Hernandez, N. (1989) *Cell* **58**, 55–67.
19. Baumruker, T., Sturm R., and Herr W. (1988) *Genes Dev.* **2**, 1400–1413.
20. Larson, D., Bradford-Wilcox, J., Young, L.S., and Sprague, K.U. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3416–3420.
21. Tyler, B.M. (1987) *J. Mol. Biol.* **196**, 801–811.
22. Ullu, E. and Weiner, A.M. (1985) *Nature* **318**, 371–374.
23. Kleinert, H. and Benecke, B.-J. (1988) *Nucleic Acids Res.* **16**, 1319–1331.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.