

Analysis of clustered point mutations in the human ribosomal RNA gene promoter by transient expression *in vivo*

(rRNA transcription/RNA polymerase I/species-specific transcription/linker scanning mutants)

MICHELE HALTINER JONES*, R. MARC LEARNED†, AND ROBERT TJIAN

Howard Hughes Medical Institute, and Department of Biochemistry, University of California, Berkeley, CA 94720

Communicated by H. A. Barker, September 21, 1987

ABSTRACT We have mapped the cis regulatory elements required *in vivo* for initiation at the human rRNA promoter by RNA polymerase I. Transient expression in COS-7 cells was used to evaluate the transcription phenotype of clustered base substitution mutations in the human rRNA promoter. The promoter consists of two major elements: a large upstream region, composed of several domains, that lies between nucleotides –234 and –107 relative to the transcription initiation site and affects transcription up to 100-fold and a core element that lies between nucleotides –45 and +20 and affects transcription up to 1000-fold. The upstream region is able to retain partial function when positioned within 100–160 nucleotides of the transcription initiation site, but it cannot stimulate transcription from distances of ≥ 600 nucleotides. In addition, we demonstrate, using mouse-human hybrid rRNA promoters, that the sequences responsible for human species-specific transcription *in vivo* appear to reside in both the core and upstream elements, and sequences from the mouse rRNA promoter cannot be substituted for them.

We previously generated a series of linker scanning mutations in the human rRNA promoter and assayed their activities in cell-free transcription reactions. Two distinct control elements were identified that affect the level of rRNA transcription *in vitro*: a core sequence located between –45 and +20 that has a substantial effect on transcription (>100 -fold) and an upstream element located between –156 and –107 that modulates the level of transcription (3- to 5-fold) (1). The upstream control element (UCE) appears to function only when located within a limited distance from the core element and in a single orientation. DNase “footprinting” experiments on the human rRNA promoter confirm the importance of the core and UCE in directing RNA synthesis. Transcription factors UBF-1 and SL1 purified from human tissue culture cells interact with a region in the human rRNA promoter that extends from –75 to –165 relative to the transcription initiation site and includes the UCE, as well as a region in the core around nucleotide –21 (2).

A unique property of RNA polymerase I transcription demonstrated both *in vivo* (3) and *in vitro* (4, 5) is the requirement for template DNA derived from the same or closely related species. Indeed, although similarly positioned promoter regions have been mapped in human and mouse rRNA genes (for review, see ref. 6), *in vitro* studies with human-mouse hybrid promoters suggest that the mouse upstream region cannot substitute for the human UCE in activating transcription from the human core element (2).

A major goal of the current study was to map the control regions of the human rRNA promoter in detail and to assess the function of the UCE *in vivo*. To this end, we have

evaluated the transcription phenotype of linker scanning and various other promoter mutations in intact primate cells. We have previously described and characterized a transient expression system that involves transfection into mammalian tissue culture cells of a plasmid vector containing the human rRNA promoter fused to the herpes virus thymidine kinase gene (7). Transcripts directed by the hybrid gene were found to be nonpolyadenylated, unstable, and localized predominantly in the nucleus. Analysis of deletion mutants of the rRNA promoter by using this system showed that sequences between –234 and +16 are required for maximal transcription. Here, we have extended these studies to include the analysis of transcription directed by templates containing clustered base substitutions spanning the entire upstream control region and core element of the human rRNA promoter. In addition, we have investigated the sequences responsible for human species-specific transcription *in vivo*.

MATERIALS AND METHODS

Recombinant DNA. The wild-type template contained sequences from –650 to +80 relative to the RNA start site. The majority of the mutants (prHuTKLS), including TKI-UCF, were constructed by inserting the *Bst*EII-*Acc* I (filled-in) fragment of LS plasmids (1) into the *Bst*EII-*Cla* I (filled-in) fragment of prHuTK3. TKCP-120/-107 was constructed by mutagenizing the rRNA promoter in pUC18 by oligonucleotide-directed mutagenesis (–120 ATGATCGTG-CCTGT –107; italics indicate altered nucleotides) and then inserting the *Eco*RI-*Pst* I (filled-in) fragment into the *Cla* I (filled-in)-*Bst*EII (filled-in) fragment of prHuTK3. TKHu-70/Mo-56 was constructed by inserting the *Ban* I (filled-in) fragment of MoHu (2) into the *Cla* I-*Bst*EII (filled-in) fragment of prHuTK3. TKMo-73/Hu-73 was constructed by inserting the *Sal* I (filled-in)-*Bst*EII fragment of HuMo (2) into the *Cla* I (filled-in)-*Bst*EII fragment of prHuTK3. TKLS-54/-73 and TKLS-54/-94 were constructed by inserting the *Eco*RI-*Sac* I fragment of TKLS-54/-45 into TKLS-74/-73 and TK-107/-94, respectively, cut with *Eco*RI and *Sac* I. The series of mutants TKLS-60/0.5kb, TKLS-60/1kb, TKLS-130/0.5kb, and TKLS-130/1kb were constructed by inserting various random fragments of *Drosophila melanogaster* DNA that had been digested with *Sac* I into the *Sac* I site of the appropriate linker scanning mutant (i.e., TKLS-60/-51 or TKLS-130/-120).

DNA Transfections. DNA transfections (8) of COS-7 monkey cells (9) in the presence of DEAE-dextran were performed as described (7). Transfection efficiency was $>10\%$, as measured by indirect immunofluorescence using an anti-

Abbreviation: UCE, upstream control element.

*Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143.

†Present address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142.

A

-250 -200
 ACGGGCCGGCCCTGCGTGTGGCCAGGGCCGGCCGGGAGGGCTCCCCGGCCCGGCGCTGTCCCGCGTGTCTCTGGGTTGACCAAGAGGGCCCGGGCGC

-150 -100
 TCCGTGTGTGGCTGCGATGTTGGCGTTTTTGGGGACAGGTGTCCGTGTCCGCGCTCGCCTGGGCCGGCGCGTGGTTCGGTGACGGACCTCCCGGCCCGC

-50 +1
 GGGAGGTATATCTTTCGCTCCGAGTCCGCATTTTGGGCCCGCGGTTATTGCTGACACGCTGTCTCTGGCGACCTGTCTGCTGGAGAGGTTGGGCCTCCG

B

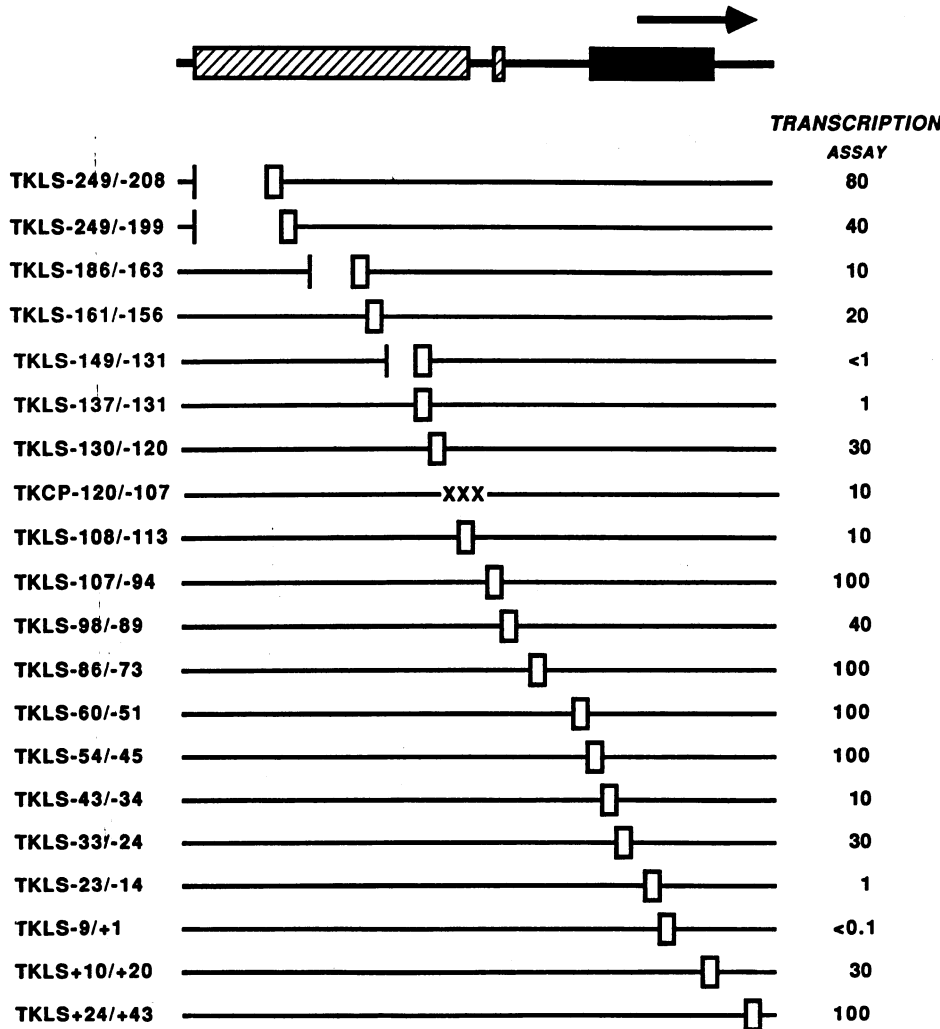
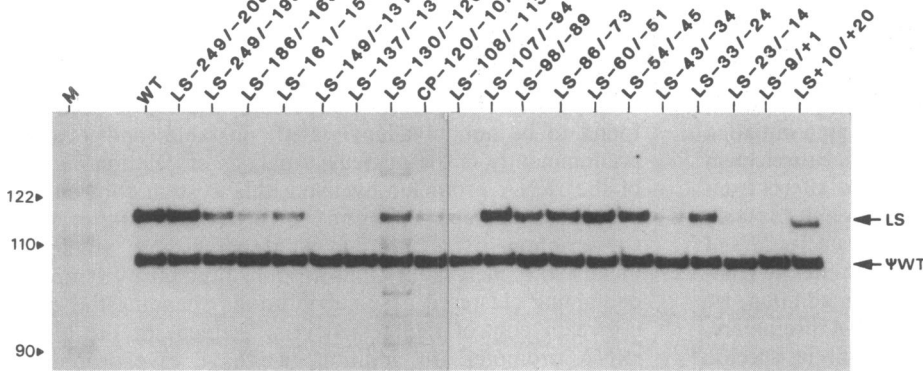


FIG. 1. Effect of linker scanning mutations on rRNA transcription *in vivo*. (A) Nucleotide sequence of wild-type human rRNA promoter. The start site of transcription is designated +1. (B) The upper panel shows a primer extension analysis of RNA prepared from COS-7 cells that were transfected with both a pseudo-wild-type template and templates containing linker scanning mutations. Nucleotide size markers (M) are shown on the left and the arrows on the right indicate the expected position of cDNAs corresponding to transcripts from the pseudo-wild-type control template (Ψ WT) and the mutant templates (LS). LS+10/+20 contains a 1-nucleotide deletion in the transcribed region and thus encodes a 1-nucleotide shorter cDNA. The exact coordinates of the linker insertion, a diagram indicating the relative position of the linker (empty boxes) in the promoter, and the transcription efficiency of each mutant template relative to the pseudo-wild-type control (%) are shown in the lower part of the figure. The coordinates refer to the nucleotides that flank the 8-base-pair (*Sac* I linker (CGAGCTCG). Note that a number of the mutations are not perfect linker substitutions but also contain small deletions (gaps) and insertions. Xs designate several point mutations (see *Materials and Methods*). Also shown is a diagram of the promoter, on the same scale, indicating the position of the core element (solid box) and upstream elements (hatched boxes) relative to the transcription initiation site, which is represented as a heavy black arrow.

body directed against the thymidine kinase protein. The replication efficiency of plasmids containing linker scanning mutants was assumed to be similar because previous studies (7) showed that the parent plasmid (pHuTK3) had the same replication efficiency with and without the entire promoter region.

RNA Preparation and Analysis. COS cells were harvested 40–48 hr after transfection and nuclear RNA was prepared as described (7, 10) and analyzed by primer extension (11) using either a thymidine kinase or an rRNA gene primer (gifts of J. Merryweather, Chiron, Emeryville, CA). Electrophoresis, autoradiography, and scanning densitometry were performed as described previously (12).

RESULTS AND DISCUSSION

Linker Scanning Mutations Define Two Major *in Vivo* Promoter Elements. To determine precisely the sequences involved in human rRNA transcription, we introduced linker scanning mutations into the promoter region (Fig. 1A) previously defined by deletion analysis (7). The vector, p-HuTKLS, contains the rRNA promoter fused to a coding sequence transcribed by RNA polymerase II, the herpes virus thymidine kinase gene. This fusion construct facilitates detection of transcripts from the plasmid without interference from the background of endogenous rRNA. In all experiments, the host cell was cotransfected with plasmids bearing the promoter mutation along with a pseudo-wild-type control plasmid. After 40 hr, RNA was prepared, and the level of transcription directed by the plasmids was measured by primer extension analysis (Fig. 1B, upper panel). The transcription efficiency of each template was evaluated relative to that of the pseudo-wild-type control. The coordinates of each mutation along the promoter and a summary of the results from several different transfections are shown in the lower panel of Fig. 1B.

The effect of deletion (7) and linker scanning mutations on the level of transcription *in vivo* defines two major promoter elements: First, a large upstream control region extends from –234 to –107 relative to the transcription start site and appears to be composed of several smaller domains affecting transcription 2- to >100-fold. For example, note that linker scanning mutations TKLS-149/-131 and TKLS-120/-107 decrease transcription >100-fold and 10-fold, respectively, but are separated by a region (–130 to –120) that has only a 3-fold effect on transcription. Second, a core region extends from –45 to +20 and affects transcription 3- to >1000-fold. In addition, the five nucleotides from –94 to –90, which lie in a large mutation-insensitive region, appear to have a small effect on transcription. In contrast, sequences upstream of –234 (from previous deletion analysis), between –107 and –45 and downstream of +20 have little, if any, effect on transcription. This analysis confirms and extends the *in vivo* deletion analysis to identify two major control regions in the human rRNA gene promoter. Fig. 1B also demonstrates that transcription directed by these mutant templates is initiated at the correct nucleotide.

An 11-bp sequence in the UCE between –174 and –164 [and 3' to the 28S rRNA coding sequences (13)] is nearly identical (10/11 match) to a sequence that appears to function both as a terminator at the 3' end of the mouse rRNA gene and as a promoter element at position –171 relative to the transcription start site (14, 15). Although termination at this site in the UCE was not monitored in these experiments, a mutation that disrupts a region including the 11-bp sequence (TKLS-186/-163) decreases transcription 10-fold.

Whereas mutations in the core element have nearly identical effects on transcription in the transient assay and the *in vitro* system, lesions in the UCE have significantly different effects in the two assays. First, mutations in sequences

between –234 and –156, which decrease *in vivo* transcription 2- to 10-fold, have no measurable effect on transcription *in vitro*. Second, mutations in sequences between –149 and –131 have a greater effect *in vivo* (≥ 100 -fold) than *in vitro* (3- to 5-fold). We believe the reason for these differences involves the function of these two domains in the intact cell, rather than simply a greater sensitivity in the *in vivo* assay, because the effects of the core element and of the region between –130 and –107 were quantitatively similar in the two assays.

DNase footprint analysis of the human rRNA promoter using transcription factors UBF-1 and SL1 purified from HeLa cells (ref. 2; S. Bell, R.M.L., and R.T., unpublished results) revealed an upstream protected region that extends from –75 to –165 and overlaps extensively with sequences in the UCE shown here to affect transcription (Fig. 2). Enhanced cleavage was also observed at nucleotide –21 in the core region. Moreover, there is a strong correlation between the ability of linker scanning mutants to direct transcription and interact with transcription factors. These findings provide strong evidence that the sequence-specific cellular DNA-binding protein UBF-1, in conjunction with SL1, interacts selectively with the UCE and core elements to promote human rRNA transcription. In addition, another factor, which may interact with the region between –234 and –156, was detected *in vivo* by a transcription assay but not by either an *in vitro* transcription or an *in vitro* footprinting assay.

Effect of Distance on the Function of the Upstream Region.

The ability of the upstream region to function from various positions relative to the core element was examined by deleting or duplicating the intervening sequences. We also evaluated the effects of inverting or duplicating a large portion of the upstream region. Note that a complete deletion of the upstream region ($\Delta 5'$ -83) decreases transcription to $\approx 4\%$ of wild type (7). As shown in Fig. 3, it appears that the upstream region is still partially functional when moved 16 nucleotides closer to the core (transcription level of TKLS-98/-73 is reduced 4-fold compared to the parent construct used for these experiments, TKLS-98/-89) but is nonfunctional when moved ≥ 32 nucleotides closer (transcription level is reduced >100-fold for TKLS-98/-57 and TKLS-98/-45). The upstream region also appears to be partially functional when placed as much as 49 nucleotides further upstream from the core (transcription level is 40% for TKLS-54/-94). In contrast, the upstream element appears to be essentially nonfunctional *in vitro* when moved either ≥ 16 nucleotides closer or ≥ 49 nucleotides further from the core element (1). Thus, it appears that the function of the upstream region is somewhat less sensitive to distance alterations in the intact cell than in the *in vitro* system.

LS-107/-156 results in a substitution of the distal portion of the upstream region (at –156) with a second full-length copy that appears to partially compensate for the disruption.

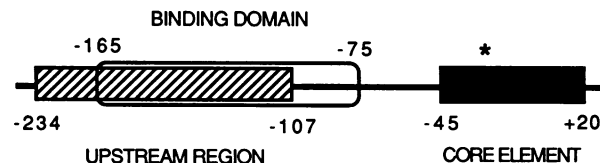


FIG. 2. The upstream control element and the core element overlap binding domains in the rRNA promoter. The boxes represent promoter elements in the rRNA promoter with the numbers below denoting position relative to the RNA start site. A region that is protected and an enhanced cleavage site (at –21) in a DNase footprinting assay using partially purified UBF-1 and highly purified SL1 are shown by an ellipse and an asterisk, respectively, with the coordinates of the binding region shown above the ellipse.

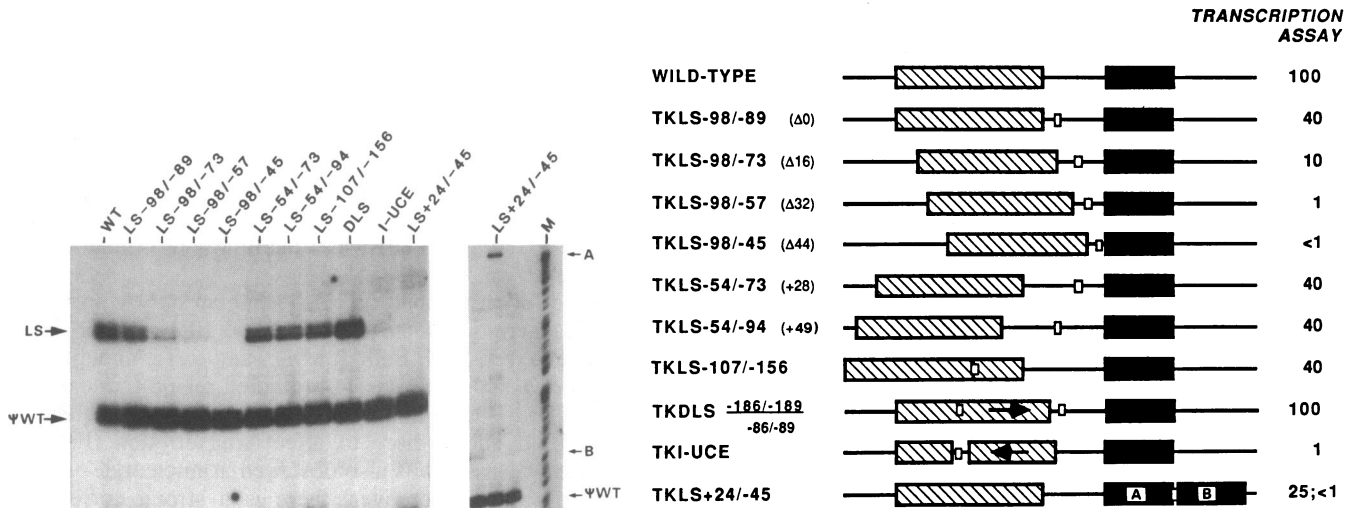


FIG. 3. Effect of altering the spacing, orientation, or number of promoter elements relative to each other. (Upper) Primer extension analysis of RNA prepared from COS-7 cells transfected with both a pseudo-wild-type template (Ψ WT) and templates containing linker scanning mutations (LS). A and B show the expected migration for cDNAs from transcripts starting at the A core element and the B core element, respectively, in TKLS+24/-45. (Lower) Exact coordinates of the linker insertion, a diagram indicating the position and orientation of the upstream element (hatched box) relative to the core element (solid box), and the transcription efficiency of each mutant template relative to the pseudo-wild-type control (%). The numbers in parentheses after some of the mutation names indicate how many nucleotides the upstream element has been moved relative to the core element.

It is probable that only the complete copy is operative, albeit at a reduced level due to the increased spacing from the core (+57 nucleotides). Alternatively, if the upstream region is indeed composed of several subdomains, the proximal portion of the first copy may be functioning as an independent domain. The upstream region cannot function if a major portion of its sequences are inverted, as shown by the mutation I-UCE, which decreases transcription to 1% of wild-type level. It should be noted that this construction (I-UCE) does not invert the entire upstream region but rather the stretch between -86 and -186, which has the predominant effect on the level of transcription. The decrease is not due simply to the presence of the two linkers, since DLS-186/-86, which contains linkers inserted at -86 and -186 but does not invert the intervening sequences, is transcribed at wild-type levels. The inversion could potentially disrupt interactions between UBF/SL1, which have been shown to protect this region, and either the core element or the region upstream of -186.

The effect of the mutation TKLS+24/-45, which contains two tandemly repeated core elements, illustrates two inter-

esting points. First, the presence of the downstream core (B) inhibits transcription from an otherwise competent promoter (A). We obtained a similar result in our *in vitro* assay and believe that the interaction of the transcription machinery with either core region reduces accessibility of the neighboring core region. Second, the B start site is utilized much less frequently (<1% of wild type) than the A start site (25% of wild type), at a level comparable to a template bearing a deletion of the upstream region. It is likely that the reduced activity of the B core is due to interference from the intervening A core element, perhaps even directly by a polymerase initiating at the A core and transverting the B core.

The partial functionality retained by the upstream region when it was positioned 49 nucleotides further upstream from the core prompted us to examine the effect of more drastic spacing mutations on transcription. Therefore, we constructed four additional mutations (Fig. 4), two of which positioned the entire upstream region 500 or 1000 nucleotides further from the core. Because the distal portion of the upstream region appears to have the strongest effect on

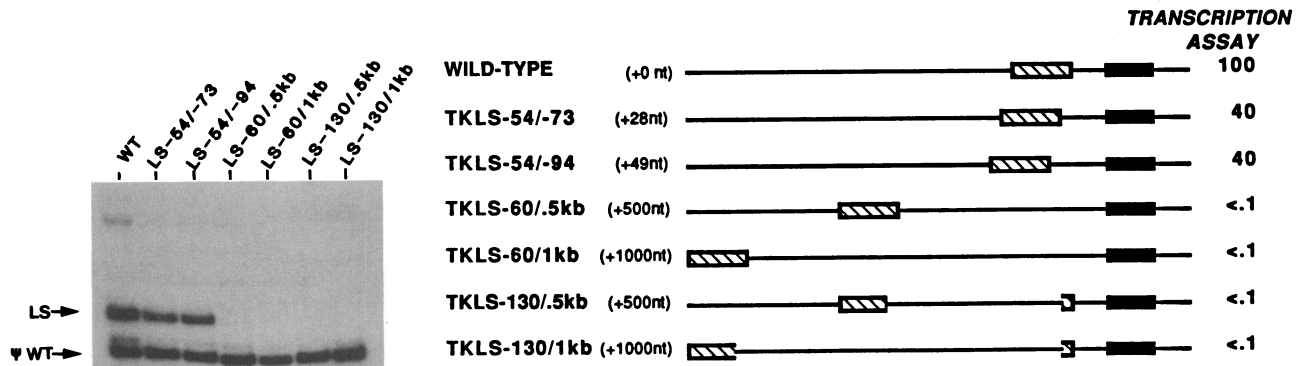


FIG. 4. Effect of increasing the distance between the core and upstream regions. (Left) Primer extension analysis of RNA prepared from COS-7 cells transfected with both a pseudo-wild-type template (Ψ WT) and templates containing linker scanning mutations (LS). (Right) Exact coordinates of the inserted fragment, a diagram indicating the position of the upstream element (hatched box) relative to the core element (solid box), and the transcription efficiency of each mutant template relative to the pseudo-wild-type control (%). The numbers in parentheses indicate how many nucleotides (nt) the upstream element has been moved relative to the core element.

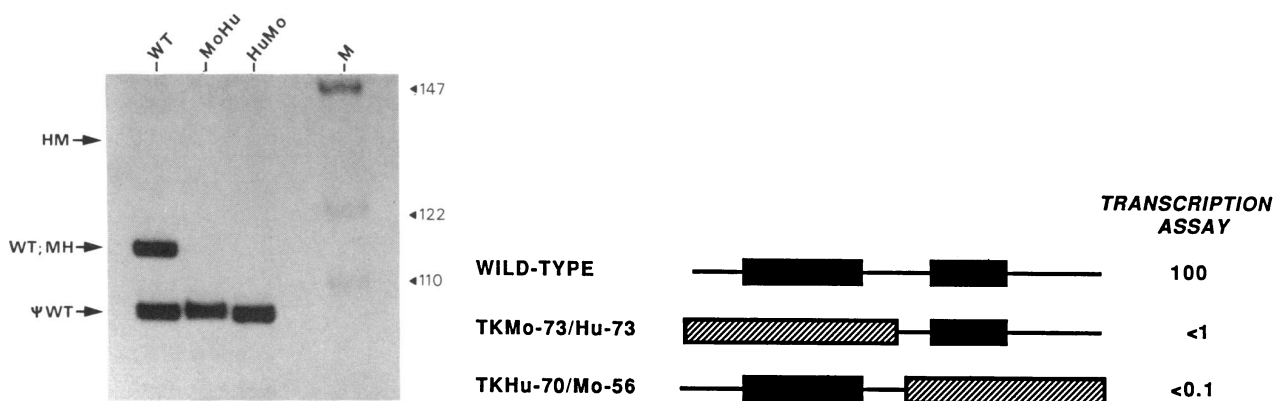


FIG. 5. Effect of mouse promoter elements on human rRNA transcription. (Left) Primer extension analysis of RNA prepared from COS-7 cells transfected with both a pseudo-wild-type template (Ψ WT) and either a wild-type human template (WT), a fusion template with the human core element and mouse upstream sequences (MoHu), or a fusion template with the human upstream element and mouse "core" sequences (HuMo). The arrows on the left indicate the expected position of the corresponding cDNAs from the HuMo template (HM), the MoHu template (MH), the wild-type template (WT), and the pseudo-wild-type template (Ψ WT). Nucleotide size markers (M) are shown on the right. (Right) Coordinates of the fusions, a diagram of the promoters, and their relative transcription efficiencies (%). The solid boxes denote human rRNA promoter elements and the hatched regions denote mouse promoter sequences. Note that because the 3' border of the mouse upstream region has not been mapped, we constructed the fusions by using the human promoter as a prototype.

transcription, we also constructed two mutations that positioned only the distal half of the UCE 500 or 1000 nucleotides further upstream, leaving the proximal domain at its normal distance from the core. Although we reproducibly observe partial functionality of the upstream region when the spacing between the UCE and the core element is increased by 28 or 49 nucleotides (LS-54/-73, LS-54/-94), the activity is abolished when either the entire UCE or the distal domain is positioned ≥ 500 nucleotides further from the core element (LS-60/0.5kb, LS-60/1kb, LS-130/0.5kb, LS-130/1kb).

Mouse Promoter Elements Have No Effect on Human RNA Polymerase I Transcription *in Vivo*. To investigate contributions of the core element and the UCE to the requirement of homologous template for human ribosomal RNA transcription, we evaluated transcription from human-mouse hybrid promoters in our transient expression system (Fig. 5). The template containing the mouse upstream region (14–16) and the human core element (Mo-105/Hu-97) is transcribed at a level (<1% of wild type) equivalent to that of a deletion of the human UCE. Thus, the presence of the mouse upstream region (from -169 to -105) neither stimulates nor inhibits human rRNA transcription. There is no detectable transcription (<0.1% of wild type) from a template containing the human upstream region and the mouse "core" region (17, 18) in human cells (Hu-70/Mo-56), confirming both that the human upstream element alone is not sufficient to direct transcription and that the mouse "core" cannot substitute for the human core element. In addition, a plasmid containing the entire mouse rRNA promoter fused to the thymidine kinase gene of herpes simplex virus promotes transcription in mouse 3T6 cells but not in monkey COS-7 cells (S. Smale and R.T., unpublished data).

Footprint analysis demonstrated extensive protection of the UCE of the Hu-70/Mo-56 fusion template but not of the Mo-105/Hu-97 fusion template or of an intact mouse template, suggesting that the binding of UBF-1 to these sequences is species-specific as well (2). However, these results do not exclude the possibility that some domains of these promoter elements are shared between different species. Presumably, the 11-bp sequence within the UCE in the human promoter that has been characterized as a terminator/promoter signal in the mouse promoter is an example of such a shared domain. The existence of two other regions of

homology, one in the core and one in the upstream region, between the rat, mouse, and human promoters (19) is also consistent with this idea. Further analysis of chimeric promoters, as well as studies on DNA-binding properties and transcriptional activity of purified transcription factors, will be required to identify the common and divergent elements of rRNA transcription.

We thank Keith Yamamoto, Michael Jantzen, and Steve Bell for critical reading of this manuscript, the reviewers for helpful comments, Steve Smale for the plasmid prHuTK3, Ernst Hafen for *Drosophila* DNA fragments, and Karen Ronan for preparing the manuscript. This work was supported in part by National Institutes of Health Grant GM 32856 awarded to R.T.

- Haltiner, M., Smale, S. & Tjian, R. (1986) *Mol. Cell. Biol.* **6**, 227–235.
- Learned, R. M., Learned, T. K., Haltiner, M. M. & Tjian, R. (1986) *Cell* **45**, 847–857.
- Tantravani, R., Miller, D. A., D'Ancona, G., Croce, C. M. & Miller, O. J. (1979) *Exp. Cell Res.* **119**, 387–392.
- Grummt, I., Roth, E. & Paule, M. (1982) *Nature (London)* **296**, 173–174.
- Learned, R. M., Cordes, S. & Tjian, R. T. (1985) *Mol. Cell. Biol.* **5**, 1358–1369.
- Sollner-Webb, B. & Tower, J. (1986) *Annu. Rev. Biochem.* **55**, 801–830.
- Smale, S. & Tjian, R. (1985) *Mol. Cell. Biol.* **5**, 352–362.
- Luthman, H. & Magnusson, G. (1983) *Proc. Natl. Acad. Sci. USA* **77**, 3855–3859.
- Gluzman, Y. (1981) *Cell* **23**, 175–182.
- Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
- McKnight, S. L. & Kingsbury, R. (1982) *Science* **217**, 316–325.
- Rio, D. C. & Tjian, R. (1983) *Cell* **32**, 1227–1240.
- LaVolpe, A., Simeone, A., D'Esposito, M., Scotto, L., Fidanza, V., deFalco, A. & Boncinelli, E. (1985) *J. Mol. Biol.* **183**, 213–223.
- Henderson, S. & Sollner-Webb, B. (1986) *Cell* **47**, 891–900.
- Grummt, I., Kuhn, A., Bartsch, I. & Rosenbauer, H. (1986) *Cell* **47**, 901–911.
- Miller, K. G., Tower, J. & Sollner-Webb, B. (1985) *Mol. Cell. Biol.* **5**, 554–562.
- Grummt, I. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6908–6911.
- Kishimoto, T., Nagamine, M., Sasaki, T., Takakusa, N., Miwa, T., Kominami, R. & Muramatsu, M. (1985) *Nucleic Acids Res.* **10**, 3515–3536.
- Financsek, I., Mizumoto, K., Mishima, Y. & Muramatsu, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3092–3096.