An undecamer DNA sequence directs termination of human ribosomal gene transcription

Christa Pfleiderer+, Amke Smid, Ingrid Bartsch and Ingrid Grummt*
Institut für Biochemie, Röntgenring 11, 8700 Würzburg, FRG

Received June 5, 1990; Revised and Accepted July 19, 1990

ABSTRACT

Previously we have shown that a repetitive 18 bp sequence motif, the Sal box (AGGTCGACC-AGA/TT/ANTCCG), present in the 3' terminal spacer of mouse rDNA constitutes a termination signal for RNA polymerase I (pol I). Similar sequence elements which are functionally analogous to the murine terminator are present in the spacer of human rDNA. However, the human termination signal is shorter encompassing only 11 bp (GGGTCGACCAG) which correspond to the proximal part of the mouse sequence. Two out of the five human Sal box elements are functionally inactive due to natural point mutations which damage factor binding. A similar sequence motif with a 10 of 11 base identity with the downstream terminators is located upstream of the human transcription initiation site. The upstream element interacts with the same factor(s) as the downstream terminators and is also capable to stop elongating human RNA polymerase I. Despite the human and mouse factors exert different electrophoretic mobilities in gel retardation assays, UV-crosslinking and proteolytic clipping experiments indicate that both the sizes and the tertiary structure of the Sal box binding proteins of both species are very similar. When bound to DNA, both the human and the mouse factor terminate transcription of pol I from the heterologous species. The results implicate that changes in signal sequences necessary for termination have been accompanied by compensatory changes in the DNA binding domain of the protein(s) interacting with the termination signal. In contrast, the protein-protein interactions between the termination factor and the transcribing RNA polymerase I appear to have been conserved during evolution.

INTRODUCTION

During the last few years some insights into the molecular mechanisms which mediate transcription termination of eukaryotic RNA polymerases (pol) have emerged. It has been shown that two major processes, i.e. transcription termination, 3½ terminal processing or a combination of both events, generate correct 3' ends of transcripts synthesized by class I, II and III RNA polymerases, respectively. In each case specific protein-nucleic acid interactions appear to play an important role (1). We have studied for a while the mechanism of transcription termination at the ribosomal RNA genes from mouse. Both the DNA sequence signaling termination has been identified and the *trans*-acting protein factor interacting with the terminator sequence has been purified and functionally characterized. It has been shown that the murine rRNA genes possess multiple terminators in the 3' terminal external spacer which consist of an 18 bp sequence motif AGGTCGACCAGA/TT/ANTCCG (the Sal box). Pol I terminates transcription 11 bases upstream of the first termination signal (2, 3). Subsequently, the 3' ends of the primary transcripts are trimmed by 10 nucleotides in a sequence-dependent processing reaction (4). Termination depends on the binding of a specific nuclear factor (TTF I) to the Sal box sequence (5). Deletions, insertions, or point mutations within this sequence element decrease or eliminate factor binding and similarly inhibit or abolish transcription termination (3). Interestingly, the termination signal is orientation-dependent and specific for pol I. Purified factor TTF I bound to the Sal box sequence element directs transcription termination with pol I from species as divergent as mouse and yeast but fails to affect elongation both by prokaryotic and eukaryotic class II and III RNA polymerases, suggesting that its function is more subtle than simply blocking the elongating enzyme (6).

Transcription termination at the human rDNA cluster appears to be mechanistically similar to mouse. We have shown that the 3' end of human pre-rRNA maps upstream of a conserved sequence element TCCC/GN$_{0-3}$GGGTCGACCAG which has a 10 bases perfect match with the murine Sal box (7). In functional analogy to the mouse terminator, the human Sal box is recognized by a nuclear protein which appears to be distinct from its murine counterpart. Since the mouse factor requires a longer (18 bp) sequence motif for binding, it does not interact with the human terminator. The human factor, however, efficiently binds to the murine Sal box. Whether or not the human factor can functionally replace the murine TTF I, i.e. is capable to stop mouse pol I transcription, was not yet known.

In addition to the Sal box elements downstream of the 28S rRNA coding region, a nearly identical sequence is found

* To whom correspondence should be addressed

+ Present address: Children's Cancer Research Institute, St Anna Children's Hospital, Kinderspitalgasse 6, 1090 Vienna, Austria

upstream of the transcription start site (8, 9, 10). Such promoter-proximal terminators appear to have a dual function. They positively affect transcription initiation and prevent transcriptional interference by readthrough of polymerases from the spacer that separates each rDNA repeat (11, 12). Whether or not termination and initiation are functionally linked is an open question which has to await further study of the molecular mechanisms governing the initiation and termination process.

In this communication we report on experiments which determine the minimal nucleotide sequence that constitutes the human rDNA termination signal. By comparison with the mouse system we demonstrate that both the terminator sequences and their recognizing proteins have changed during evolution although their mode of action has been conserved. Thus the phenomenon of concerted evolution of sequences and factors mediating the read-out of multigene families (13, 14) is not restricted to protein-DNA interactions at the rDNA promoter. The results suggest that similar molecular mechanisms are involved in the species-specificity of both rDNA transcription initiation and termination.

MATERIALS AND METHODS

Plasmid constructions

The plasmids $pHrT_1$, $pHrT_3T_4$ and $pHrT_5$ contain fragments from the human 3' terminal rDNA spacer covering nucleotides 308–431 ($pHrT_1$), 521–678 ($pHrT_3T_4$) and 673–786 ($pHrT_5$) downstream of the end of 28S rRNA (see Fig. 1). In pHrSB1A, pHrSB1B, pHrSB1C and pHrSB* synthetic oligonucleotides the sequence of which is shown in Figure 5A are inserted into pUC9. The construct pHrSB1A was cleaved with Hind III and a human rDNA promoter fragment (from −411 to +379) was inserted yielding the artificial minigene pHrP-SB1A. The plasmids $pHrPT_1$, $pHrPT_2$, $pHrPT_5$ are similar in structure as pHrP-SB1A, except that they contain human 3' terminal spacer sequences covering elements T_1 (nucleotides 308–431), T_2 (nucleotides 428–523), or T_5 (nucleotides 630–786), respectively. The human-mouse minigene $pHrPMrT_2$ contains the murine terminator element T_2 (from +604 to +685) fused to the human promoter. The homologous mouse-mouse construct

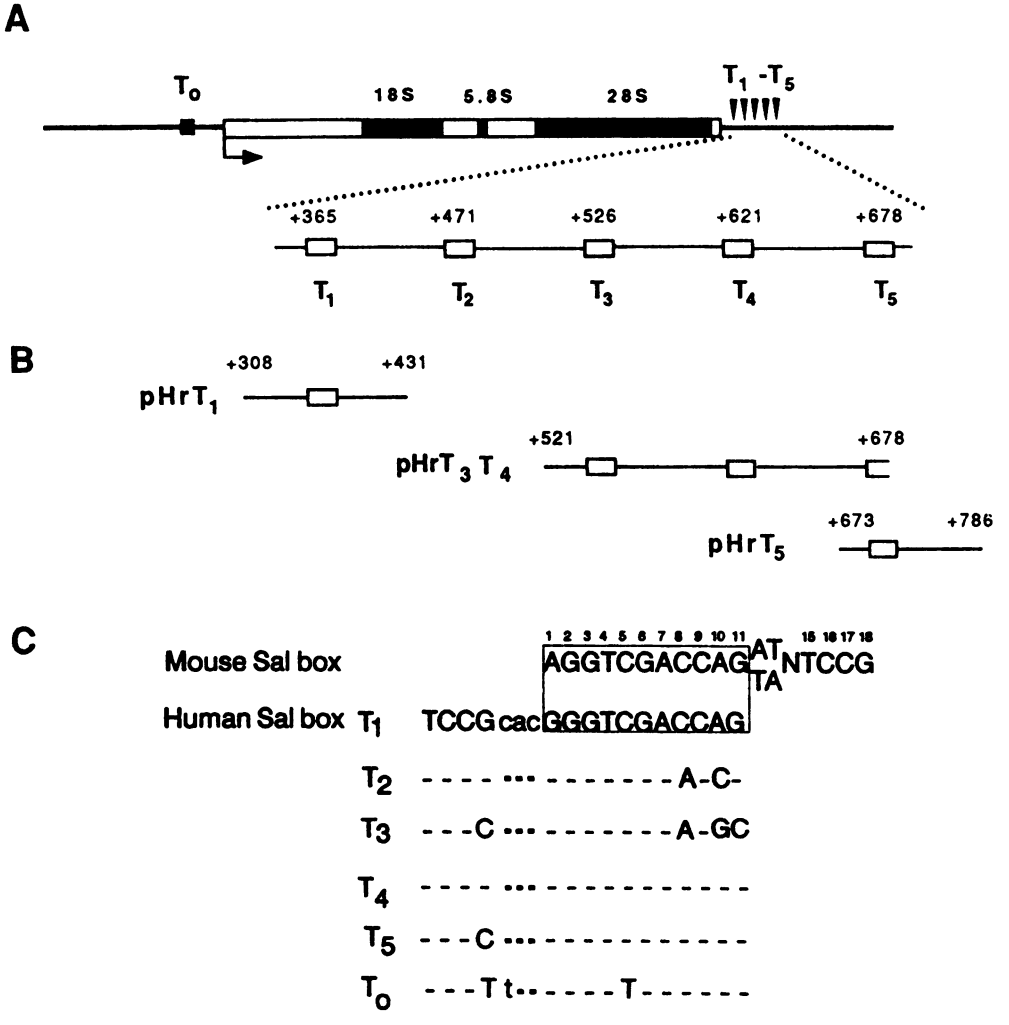

Fig. 1. A diagram showing the structure of the human ribosomal gene transcription unit and the rDNA constructs containing the tandemly repeated Sal box elements T_1-T_5. A) Structure of the human rDNA transcription unit. The diagram shows one ribosomal repeat unit which encodes 45S pre-rRNA. The hatched bars represent the 18S, 5.8S and the 28S RNA coding sequences, the open bars transcribed spacer regions, and the thin line intergenic spacer sequences. The arrows mark the positions of the clustered termination signals in the 3' spacer region (T_1-T_5), the box depicts the promoter-proximal terminator To. The blow up shows the 3' terminal spacer region containing the Sal box elements T_{1-5}. The numbers above the boxes refer to the distance (in nucleotides) of the individual elements relative to the 3' end of 28S RNA. B) rDNA fragments used in the exonuclease III protection assay. C) Sequence of the individual Sal box elements T_{1-5} present in the 3' terminal human rDNA spacer as compared to the upstream terminator T_0 and to the consensus sequence of the murine Sal box.

pMrPMrT$_2$ contains the mouse (from −169 to +155) instead of the human promoter fragment. The construct pHrP2x contains two identical human promoter fragments (−411 to +379) in head to tail orientation. pUCBH contains 3′ terminal mouse rDNA spacer sequences extending from nucleotides +334 to +712 with respect to the end of 28S rRNA (3). pHrT$_0$ contains a synthetic 32 bp oligonucleotide 5′CTGTCCTTGGGTTGACCAGAGGGACCCGGATC cloned into the Sma I site of pUC9.

In vitro transcription assay

The soluble cell-free transcription system and the analysis of the RNA synthesized have been described before (15, 16). Usually 50−100 ng of template DNA were incubated with 15 μl of a mixture of S-100 and nuclear extracts in a total volume of 25 μl containing 12 mM HEPES pH 7.9, 0.1 mM EDTA, 0.5 mM DTE, 5 mM $MgCl_2$, 75 mM KCl, 10 mM creatine phosphate, 12% glycerol (v/v), 0.66 mM each of ATP, CTP, and UTP, 0.01 mM GTP and 1.5 μCi of (α-^{32}P)GTP. After incubation for 60 min at 30°C the nucleic acids were extracted, precipitated and the labelled RNA was analyzed on nondenaturing 4% polyacrylamide gels.

Chromatographic procedures

All binding experiments were performed with partially purified protein fractions from Ehrlich ascites or HeLa cells, which were obtained by chromatography of nuclear extracts on DEAE-Sepharose and Heparin Ultrogel as described by Bartsch et al. (5).

Exonuclease III protection experiments

DNA fragments derived from the 3′ terminal spacer region were 5′-end labelled and approximately 20,000 cpm (about 1 ng of DNA) were incubated for 15 min at 30°C in 25 μl of binding buffer (12 mM HEPES, pH 8.0, 75 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM dithiothreitol, 4 mM NaF, 12% glycerol) in the presence of 1 μg of pUC9 cut with Hpa II, 10 μg of yeast tRNA, and about 10 μg of partially purified Sal box binding protein. After binding, 6 U of exonuclease III were added and incubation was continued for another 6−12 min depending on the individual fragments used. The DNA was purified by phenol-chloroform extraction and electrophoresed on 6% sequencing gels.

Electrophoretic mobility shift assay

A double stranded 39 bp synthetic oligonucleotide (CCCGGGATCCTTCGGAGGTCGACCAGTACTCCGGGCGAC) was 3′ end labelled, and 20,000 cpm (2−3 fmoles) were incubated with 15 μg nuclear protein fractions in a total volume of 25 μl of binding buffer (12 mM Tris-HCl, pH 8.0; 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM dithiothreitol, 2 μg poly(dI-dC), 10 μg tRNA, 8% glycerol) in the presence of different amounts of competitor DNA. After incubation for 20 min at room temperature, the samples were loaded onto low ionic strength (30 mM Tris-borate, pH 8.3, 0.8 mM EDTA) 8% polyacrylamide gels and electrophoresed at 13 V/cm for 2.5 h

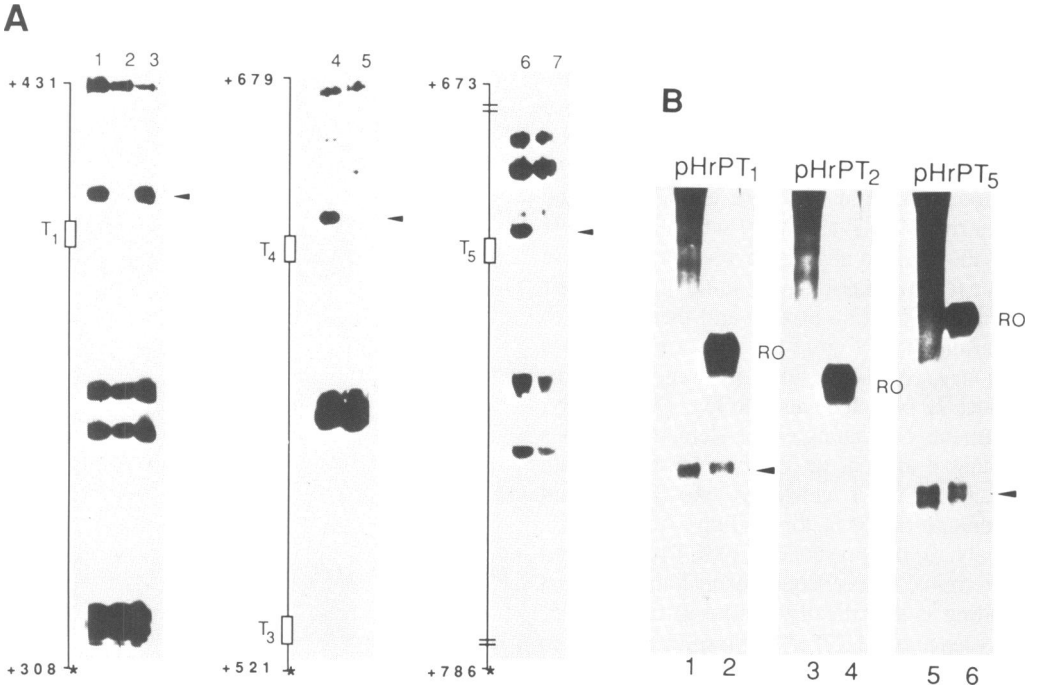

Fig. 2. Natural point mutations within the human terminator elements T$_2$ and T$_3$ eliminate factor binding and transcription termination. A) Exonuclease III protection assay. The 154 bp Eco RI-Hind III fragment from pHrT$_1$ (lanes 1−3), the 186 bp Eco RI-Hind III fragment from pHrT$_3$T$_4$ (lanes 4 and 5) and the 321 bp Eco RI-Pvu II fragment from pHrT$_5$ (lanes 6 and 7) were incubated with 5 μl of nuclear extract proteins. After 15 min, 6 units of exonuclease III were added and the incubation was continued for another 6 min (lanes 1−5) and 12 min (lanes 6 and 7), respectively. The denatured products were separated on 6% sequencing gels along with size markers (pBR322/Hpa II). Lane 1: no competitor DNA; lanes 2, 5 and 7: 500 ng of pUCT$_1$; lanes 3,4, and 6: 500 ng of pUCT$_1$* which is similar to pUCT$_1$ except two base mutations in the Sal box which abolish factor binding (3). The arrows indicate the positions of the 115 bp, the 125 bp and the 120 bp fragments, which correspond to factor binding to the first, the fourth and the fifth human Sal box. The position of the third Sal box element which is not recognized by the factor is shown as well. B) In vitro transcription assay. 100 ng of the template DNAs pHrPT$_1$, pHrPT$_2$, and pHrPT$_5$ were transcribed in a standard 25 μl reaction in the presence of 15 μl HeLa extract either in circular form (lanes 1, 3, and 5) or after digestion with Pvu II (lanes 2, 4, and 6). The arrows mark the terminated transcripts, the run off transcripts are labelled RO.

at room temperature. The gel was then dried and exposed for autoradiography.

UV-crosslinking

A 5-bromo-2'-deoxyuridine substituted labelled DNA probe was prepared by annealing a 14 nt primer (5'GATCAAAAAAACCA) with a synthetic oligonucleotide encompassing the murine Sal box sequence (5'GATCCTTCGGAGGTCGACCAGTACTCCGG-TGGTTTTTTTGATC) and synthesizing the second strand by Klenow DNA polymerase in the presence of 75 μM dGTP, 150 μM 5-bromo-2'-deoxyuridine triphosphate and 25 μCi each of (α-^{32}P)dATP and (α-^{32}P)dCTP (3,000 Ci/mmole). After removal of the unincorporated nucleotides by chromatography on Sephadex G-50 and phenol-chloroform extraction, about 3×10^5 cpm of labelled DNA were incubated for 20 min at room temperature with approximately 30 μg of partially purified mouse or human Sal box binding protein in a 25 μl reaction containing the same binding conditions as those used for the mobility shift assay. The mixture was irradiated for 10 min at 302 nm. After addition of $CaCl_2$ to a final concentration of 3.5 mM, DNA was digested for 30 min at 30°C with 2 μg of DNase I and 20 units of micrococcal nuclease and the proteins were analyzed by electrophoresis on SDS polyacrylamide gels. Proteolytic digestion of the bound proteins was carried out for 5 min at 37°C in the presence of 0.1 and 1 μg of proteinase K, respectively.

RESULTS

Natural point mutations within the human Sal box elements inactivate individual terminators

We have previously identified transcription termination signals in the spacer of human rDNA which exhibit structural similarities to the murine terminator, the Sal box. The 3' end of 45S pre-rRNA maps in front of the first element, T1 (7). The sequence of this terminator element T1 (see Fig. 1C) contains an 8 bp perfect palindrome which is identical to the proximal part of the murine Sal box. There are similar sequence elements located further downstream (termed T_2-T_5) some of which have a few base exchanges within the box. To investigate whether all of these elements are functional terminators, we studied both their interaction with the Sal box binding protein and their ability to direct transcription termination. Figure 2A shows exonuclease III (exo III) protection experiments with different fragments encompassing individual Sal box elements (see Fig. 1B). Clearly, incubation of the fragments containing either element T1 (lanes 1–3), T3 and T4 (lanes 4 and 5), or T5 (lanes 6 and 7) with partially purified factor TTF I generates exo III-resistant fragments which are diagnostic for factor binding to element T1, T4 and T5, respectively. In the presence of pUCT1, a plasmid containing murine 3' terminal spacer sequences including the first Sal box (3), this binding is specifically competed (lanes 2, 5, 7), whereas in the presence of pUCT1*, a plasmid identical to pUCT1 but with two base exchanges in the palindromic part of the Sal box which abolish factor binding (3), no competition occurs (lane 3). Interestingly, this specific interaction between the factor and the human Sal box sequence is not observed at elements T_2 (data not shown) and at T_3 (lanes 4 and 5). As in these two elements the cytosine at position 8 and the adenine at position 10 are converted into an adenine and guanine or cytosine, respectively (see Fig. 1C), these base exchanges are probably responsible for the inability of the factor to interact with the DNA.

This result suggests that the natural base exchanges within the T2 and T3 box will probably functionally inactivate these elements. To test the termination activity of the different human Sal box elements different minigenes were constructed in which the human rDNA promoter was fused to 3' terminal spacer fragments containing either element T_1 ($pHrPT_1$), T_2 ($pHrPT_2$) or T_5 ($pHrPT_5$), respectively. These three minigene constructs were used in the cell-free transcription system either in circular form or after cleavage with Pvu II (Fig. 2B). In the circular form long read-through transcripts were generated from all templates which accumulate on top of the gel (lanes 1, 3, and 5). After digestion with Pvu II defined run-off transcripts (RO) were generated from all templates (lanes 2, 4, and 6). Besides these read-through transcripts the minigenes $pHrPT_1$ and $pHrPT_5$ direct the synthesis of transcripts with a defined length which correspond to RNA molecules terminated at site T_1 and T_5, respectively. No terminated transcripts were detectable when $pHrPT_2$ was used as template (lanes 3 and 4). This result demonstrates the strict correlation between factor binding and termination activity, and suggests that some of the repeated terminator elements have been inactivated by mutation.

The promoter-proximal human Sal box is a functional terminator

Previously it has been shown that a terminator sequence is present upstream of the rDNA promoter of mouse and *Xenopus* (8, 9,

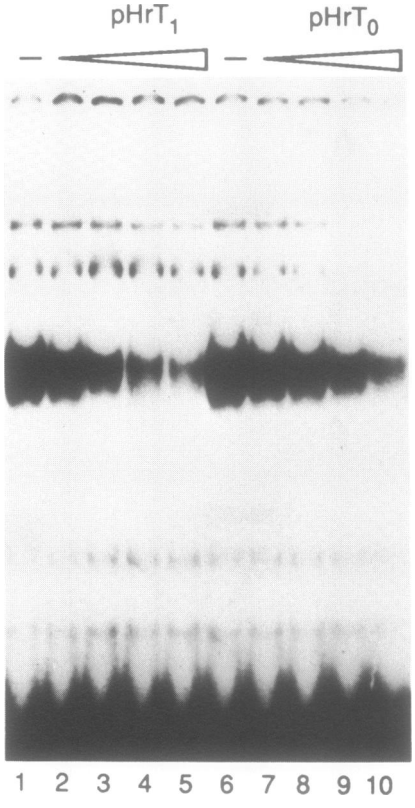

Fig. 3. Competitive binding experiments demonstrating the different affinities of the human Sal box binding protein to the termination site T1 and to the upstream terminator T_0. 2–3 fmoles (20,000 cpm) of a labelled Sal box-specific oligonucleotide were incubated with 15 μg of fractionated HeLa nuclear extract proteins in the presence of increasing amounts of competitor DNA. The competitor DNAs used were $pHrT_1$ (lanes 2–5) and $pHrT_0$ (lanes 7–10) at 25 fmoles (lanes 2 and 7), 50 fmoles (lanes 3 and 8), 100 fmoles (lanes 4 and 9) or 200 fmoles (lanes 5 and 10) per reaction. The total amount of DNA in the binding assay was kept at 200 fmoles by filling up with pUC9 vector DNA. Lanes 1 and 6 show the DNA-protein complexes formed in the presence of 200 fmoles of pUC9.

17, 18). Sequence analysis of the region preceding the human rDNA transcription initiation site reveals a Sal box motif extending from position −185 to −175 (10, 19). This sequence element GGGTTGACCAG has one base exchange (C to T) at the fifth nucleotide as compared to the functional downstream elements T_1, T_4 and T_5. To investigate whether the Sal box binding factor interacts with this upstream element containing this natural point mutation, the binding activities of the upstream and downstream elements were compared by competitive mobility shift analysis. For this, fractionated human extract proteins were incubated with a labelled Sal box oligonucleotide in the presence of increasing amounts of plasmid DNA containing either the human downstream terminator T_1 ($pHrT_1$) or the T_o sequence ($pHrT_o$). As shown in Figure 3, both DNAs efficiently compete for factor binding as demonstrated by the decrease in signal intensity in the presence of an excess of competitor DNA. However, the efficiency of pHrTo to compete for factor binding is about 50% lower than that of pHrT1. Thus the mutation within the upstream box lowers the affinity but does not eliminate factor binding.

Next we studied whether the upstream box would constitute a genuine transcription terminator. For this, a construct was used

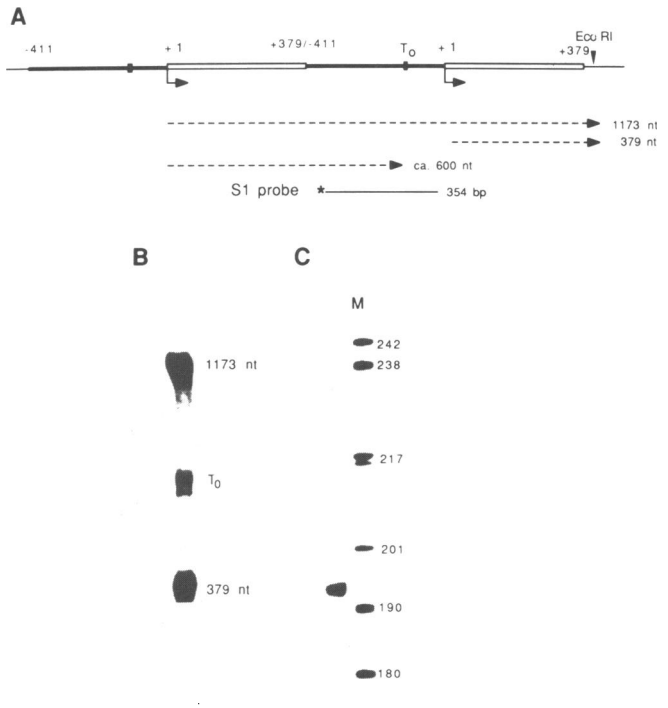

Fig. 4. Identification of a termination signal upstream of the human ribosomal gene promoter. A) Schematic representation of the human rDNA construct pHrP2x and of the transcripts synthesized *in vitro*. The thin bar represents 5′ terminal spacer sequences, the thicker bar 379 bp of the transcribed region. The arrow indicates the transcription start site, the boxes mark the promoter-proximal upstream terminator T_o. B) Transcripts generated from pHrP2x. pHrP2x was linearized with Eco RI and 100 ng were used as template in a 25 µl transcription assay containing 15 µl nuclear extract from HeLa cells. The 1173 and 379 nt RNA molecules represent run-off transcripts initiated from the first and second rDNA promoter, respectively. Transcripts marked T_o represent RNA molecules initiated at the first gene promoter and terminated at the promoter-proximal upstream terminator of the second rDNA fragment. C) Nuclease S1 mapping of the 3′ ends of transcripts terminated at site T_o. A 354 bp Hinf I fragment from the spacer region was 3′ end labelled and hybridized to transcripts derived from a 125 µl transcription assay containing 500 ng of pHrP2x digested with Eco RI. The S1 nuclease-resistant DNA fragment was electrophoresed on a 6% sequencing gel along with a pBR322/Hpa II size marker (M).

which contains two tandem rDNA promoter fragments encompassing two identical sequences from −411 to +379 with respect to the transcription start site (pHrP2x). This construct was truncated with Eco RI and used as template in the cell-free transcription system (Fig. 4A). The length of run-off transcripts started from the first promoter should be 1173 nt, from the second promoter 379 nt. As shown in Figure 4B these two classes of RNA molecules are synthesized in HeLa extracts. However, in addition to these run-off transcripts a third species of RNA molecules of approximately 600 nt is synthesized. This length corresponds roughly to the distance between the first promoter and the upstream terminator located at the second gene fragment. To prove that this RNA in fact represents transcripts terminated at To, a nuclease S1 mapping analysis of the 3′ ends of these molecules was performed. When hybridized to an end-labelled Hinf I fragment derived from the spacer preceding the transcribed region, a 193 bp S1 nuclease-resistant fragment was generated (Fig. 4C) indicating that the 3′ ends of the terminated transcripts map 10 bp upstream of the Sal box. Thus, similar to mouse, Xenopus and Drosophila, the human rDNA transcription unit also contains a promoter-proximal active terminator.

The conserved TCCC/G motif preceding the human Sal box is not required for factor binding and transcription termination

The 18 bp consensus sequence of the mouse Sal box has a bipartite structure consisting of the proximal 11 bp motif which is almost identical to the human Sal box and a highly conserved distal TCCG motif (nucleotides 15−18) which is separated from the proximal part by three more variable nucleotides (see Fig. 1C). Interestingly, all human Sal box elements are also flanked by such a TCCC/G sequence, but in contrast to mouse this motif is located upstream of the box. In the following experiments we studied whether natural sequences flanking the box at either site contribute to factor binding. In particular, the effect of the TCCC/G motif on the strength of factor-DNA interaction was to be addressed. For this, mobility shift assays were performed in the presence of different competitor DNAs. The competitors used were pHrT1, a plasmid containing human 3′ terminal spacer sequences extending from nucleotides +308 to +431 (20), or cloned synthetic oligonucleotides (see Fig. 5A) which encompass either 11 bp of the first human Sal box (pHrSB1A) or contain additional upstream nucleotides including the TCCG motif (pHrSB1B) or downstream flanking sequences (pHrSB1C). The complexes formed in the presence of these individual plasmid DNAs are shown in Figure 5B, the quantitation of the binding-competition experiments is depicted in Figure 5C. Clearly, the presence of flanking sequences does not affect the interaction between the Sal box and its cognate binding factor. The competition by pHrSB1A (lanes 1−5), pHrSB1B (lanes 6−10), pHrSB1C (lanes 11−15) was as efficient as by $pHrT_1$ (lanes 16−20) indicating that the adjacent gene sequences do not influence binding of TTF I. No competition for factor binding was observed in the presence of pHrSB*, a mutant clone with two base exchanges at positions 10 and 11 of the Sal box sequence (lanes 21−25). This sequence GGGTCGACCGA has originally been published for the first human Sal box by La Volpe et al. (20). In our studies we used a different human rDNA clone which was isolated and kindly provided to us by J. Sylvester and R. Schmickel. The sequence of box T_1 of this functionally active clone contains a perfect Sal box consensus sequence GGGTCGACCAG. At present it is not clear whether the clone

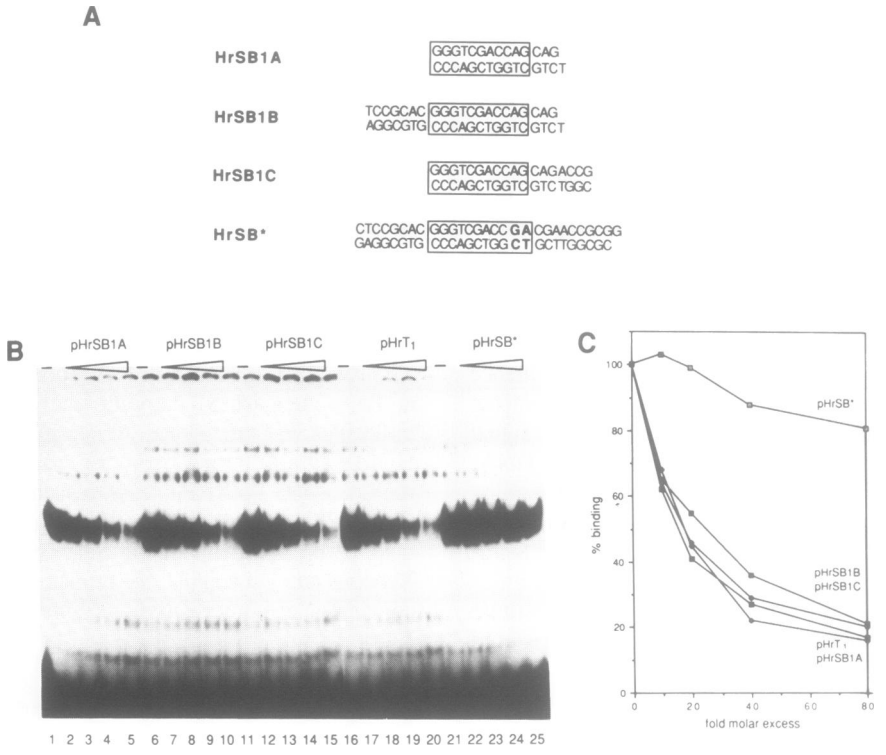

Fig. 5. Effect of DNA sequences flanking the human Sal box on factor binding. A) Sequence of oligonucleotides used for competition. B) Competitive mobility shift assay. Quantitative comparison of binding of HeLa nuclear extract proteins to the labelled murine Sal box probe in the presence of increasing amounts of competitor DNA. The assay conditions and the amounts of competitor used (25, 50 100 and 200 fmoles) were identical to those of legend to Fig. 3. The competitor DNAs were pHrSB1A (lanes 2–5), pHrSB1B (lanes 7–10), pHrSB1C (lanes 12–15), $pHrT_1$ (lanes 17–20) and pHrSB* (lanes 22–25). The total amount of DNA in the assay was kept at 0.2 pmoles by filling up with pUC9 vector DNA. Lanes 1, 6, 11,16 and 21 show the DNA-protein complexes formed in the presence of 200 fmoles of pUC9. The complexes were separated by gel electrophoresis and quantitated by autoradiography. C) Quantitation of specific DNA-protein complexes formed in the presence of increasing amounts of competitor DNA. The relative amounts of Sal box-factor complexes, as determined by densitometric scanning of appropriately exposed autoradiograms, were plotted against the molar excess of specific competitor DNA present in the assay. In the absence of specific competitor DNA (B, lanes 1, 6, 11, 16 and 21), the fraction of bound to free oligonucleotide probe represents 100% total binding.

La Volpe et al. (20) described represents a natural mutant which inactivates the first terminator or whether the AG to GA transition represents a sequencing error.

To investigate whether the 11 bp Sal box sequence on its own is sufficient to support human rDNA transcription termination, and to determine whether flanking gene sequences affect the efficiency of the termination process, the 11 bp oligonucleotide pHrSB1A was fused to the human rDNA promoter fragment yielding the minigene construct pHrP-SB1A. The termination activity was tested *in vitro* in parallel to pHrPT1, the construct containing natural gene sequences. In contrast to the murine terminator (4, 21), the TCCG/C motif or other gene sequences flanking the human Sal box did not affect either the efficiency or the accuracy of the termination process (Fig. 6). The synthetic 11 bp Sal box oligonucleotide exhibits the same capacity to stop human pol I progression as does the natural site T1 with its neighboring gene sequences.

UV-crosslinking of the murine and human Sal box binding protein

In an attempt to compare the physicochemical properties of the mouse and human Sal box binding proteins we analyzed the electrophoretic mobilities of specific protein-DNA complexes in gel retardation assays and determined the size of the individual proteins by UV-crosslinking. As shown in Fig. 7A, nuclear extract proteins partially purified from Ehrlich ascites and HeLa cells, respectively, yield different Sal box-specific DNA-protein complexes. With mouse proteins (lanes 1 and 2) four different complexes (designated A–D) were formed all of which are specific as demonstrated by competition with an excess of wild-type oligonucleotide (SB, lane 2). A mutant oligonucleotide (SB*) with a G to C and a T to G transversion of the third and fourth nucleotide within the 18 bp murine Sal box sequence did not affect the formation of the individual DNA-protein complexes (lane 1). These heterogenous complexes either represent defined proteolytic intermediates of the extremely protease-sensitive murine termination factor TTF I or different members of a gene family of Sal box binding proteins (Smid et al., manuscript in preparation). In the human system, however, there is only one specific complex formed whose mobility is significantly higher than the murine complexes (lanes 3 and 4). The differences in the electrophoretic mobility of the human and mouse Sal box-protein complexes suggest that the proteins which bind to this sequence element may differ either in size or charge. Since a direct comparison of the physical properties of the proteins recognizing the termination signal of the murine and human rDNA transcription unit has to await the purification of the respective proteins, we performed UV-crosslinking experiments which allow the identification of a specific DNA binding protein in crude protein mixtures. The DNA probe used for crosslinking was prepared by annealing a short primer to a single stranded Sal box oligonucleotide. Second strand synthesis was carried out in the presence of ^{32}P-dATP, ^{32}P-dCTP and the thymidine analogue bromo-deoxyuridine (Br-dUTP). ^{32}P-labelled DNA

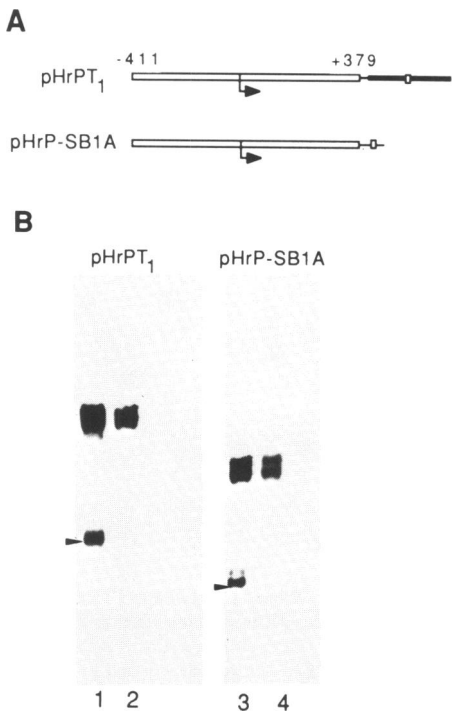

Fig. 6. The 11 bp Sal box is sufficient for human pol I transcription termination. A) Schematic representation of the rDNA minigene constructs used. B) In vitro transcription assay. A 25 μl transcription reaction contained 100 ng of template DNA pHrPT$_1$ or pHrP-SB1A, both digested with Pvu II, and 15 μl of HeLa nuclear extracts. To unambiguously identify the terminated transcripts, the in vitro transcription reaction was also performed in the presence of 400 ng of pUCBH (2), a plasmid that contains a 381 bp bp 3′ terminal fragment from the mouse rDNA spacer including two Sal box elements and, therefore, efficiently competes for factor binding (lanes 2 and 4). The arrows mark the terminated transcripts.

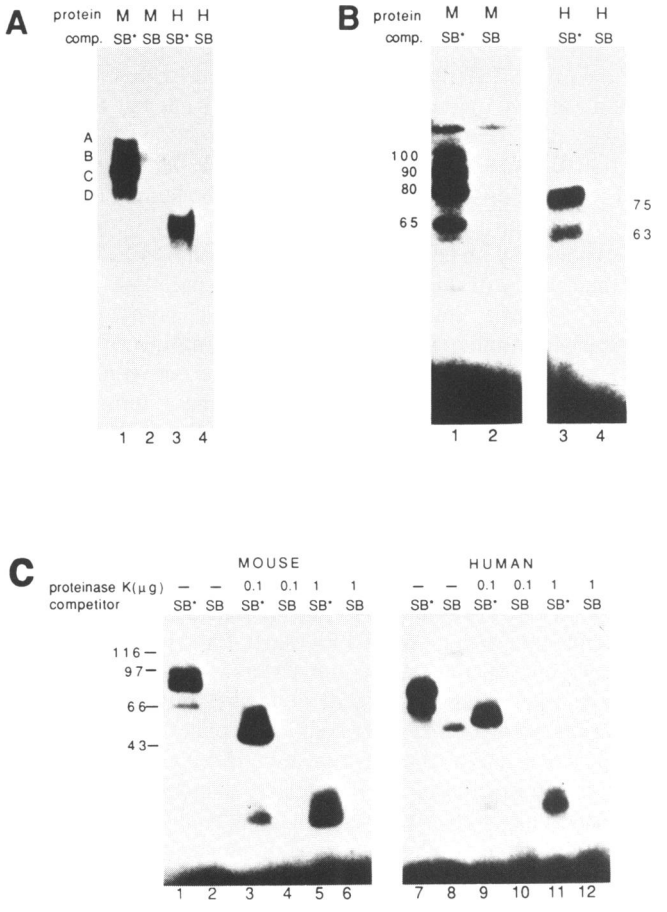

Fig. 7. Identification of murine and human Sal box binding factors by UV-crosslinking. A) Gel retardation assay showing the electrophoretic mobilities of the murine (M) and human (H) protein-DNA complexes. Binding was performed in the presence of an excess of mutant (SB*) oligonucleotide (lanes 1 and 3) or wild-type Sal box (SB) oligonucleotide (lanes 2 and 4) and the complexes were analyzed by electrophoresis on an 8% native polyacrylamide gel. The individual murine DNA-protein complexes are designated A, B, C, and D. B) UV-crosslinking of the murine (lanes 1 and 2) and human (lanes 3 and 4) Sal box binding proteins. The non-coding strand of a Sal box oligonucleotide probe was ^{32}P-labelled and the thymidine residues were substituted by Br-dUTP. The probe was incubated with partially purified Sal box binding protein in the presence of a 400-fold molar excess of unlabelled mutant (lanes 1 and 3) or wild-type Sal box oligonucleotide (lanes 2 and 4). The reactions were UV irradiated and processed as described in Materials and methods. The molecular weights (in kd) of the crosslinked proteins are indicated. C) Identification of the DNA-binding domain of mouse and human TTF I by UV-crosslinking. Partially purified murine (lanes 1−6) and human (lanes 7−12) Sal box binding proteins were bound to the Br-dUTP-substituted labelled DNA probe and then treated with the indicated amounts of proteinase K for 5 min at 37°C. The protease was inhibited by addition of 2 mM PMSF, and UV-crosslinking and processing of the probes were performed as described in Materials and methods. The crosslinked proteins were analyzed by electrophoresis on a 10% SDS polyacrylamide gel. The numbers represent the sizes of protein molecular weight markers.

molecules substituted with Br-dUTP were incubated with partially purified Sal box binding proteins from both species in the presence of wild-type or mutant oligonucleotide, and the protein-DNA complexes formed were crosslinked by exposure to UV-light. After digestion with DNase I and micrococcal nuclease, the labelled proteins were analyzed by electrophoresis on SDS gels. As shown in Figure 7B, UV-irradiation of the binding reactions labels four murine proteins with molecular weights of 65, 80, 90 and 100 kd (lane 1), respectively. None of these proteins was labelled in the presence of specific competitor DNA (lane 2). When probe DNA alone was irradiated or when UV-irradiation was omitted, no nuclease-resistant labelled protein was generated (data not shown). In fractionated human nuclear extracts two proteins with apparent molecular weights of 63 and 75 kd are labelled (lane 3) which are competed by the Sal box oligonucleotide (lane 4). Thus, despite the remarkable differences in the electrophoretic mobility of the mouse and human DNA-protein complexes, the molecular weight of the human proteins closely corresponds to the 65 and 80 kd mouse proteins. This finding indicates that the different mobilities of the DNA-protein complexes of both organisms is due to a different charge rather than different sizes of the individual Sal box binding proteins.

Further support for the structural relatedness of the proteins interacting with the murine or human termination signal derived from proteolytic cleavage of the protein-DNA complexes and UV-crosslinking of the cleavage products (Fig. 7C). Limited proteolytic digestion of the Sal box-TTF I complexes with proteinase K results in a faster migrating complex whose mobility is very similar in mouse and man (5). Analysis of the crosslinked proteins present within these protease-treated complexes reveals that the size of both the proteolytic intermediates (Fig. 7C, lanes 3 and 9) and the end products (lanes 5 and 11) is practically identical in both species (47−50 and 19 kd), demonstrating that the DNA-binding domains of human and mouse TTF I are similar in structure.

The human Sal box binding protein terminates mouse rDNA transcription

In view of the different binding specificities of the murine and human Sal box binding proteins, it was of interest to find out whether the human and mouse termination factors can functionally substitute each other and block the progression of pol I from the different species. For this, the human and mouse Sal box binding proteins were partially purified by chromatography on heparin-Ultrogel and MonoQ-FPLC columns and the activity of the individual factor preparations was tested in the cell-free transcription system at homologous and heterologous minigene templates. The minigene constructs $pMrPMrT_2$ and $pHrPMrT_2$ represent fusions between the mouse or human promoter, respectively, and a mouse 3′ terminal spacer fragment including the second Sal box T_2. Because of the marked species-specificity of rDNA transcription initiation (21) those constructs are transcribed only in the presence of their homologous extracts derived from cultured Ehrlich ascites or HeLa cells, respectively (Fig. 8). In this experiment we used mainly cytoplasmic extracts which exhibit very low termination activity on their own (lanes 1, 3, 5, 7). When these extracts were complemented with mouse TTF I (lanes 2 and 6) or human TTF I (lanes 4 and 8), most of the transcripts were terminated at the Sal box. Interestingly, the shift of read-through to terminated transcripts was observed after addition of either mouse or human TTF I, irrespective whether mouse or HeLa cell extracts were used for transcription. This result demonstrates that the termination factors of both species interact with the transcriptional machinery from both human and mouse. The reciprocal experiment, i.e. transcription termination by mouse TTF I at a minigene construct containing the human Sal box, cannot be performed since the mouse factor does not recognize the human terminator sequence (7). However, transcription of a template containing the mouse promoter fused to the first human Sal box was efficiently terminated by human TTF (data not shown) indicating that the human and mouse Sal box binding proteins can functionally substitute each other despite the differences in their target sequences.

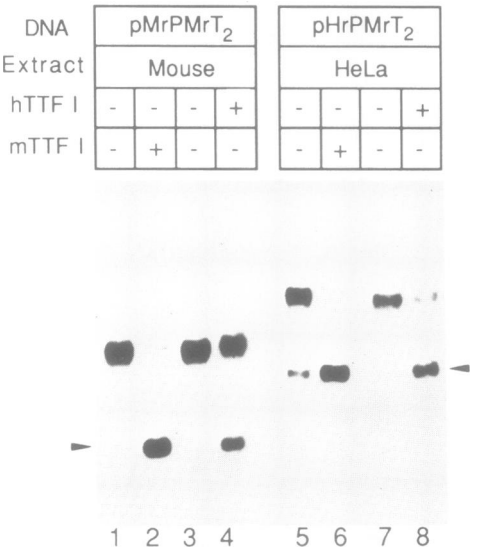

Fig. 8: Functional substitution of human and mouse TTF I. 50 ng of $pMrPMrT_2$ or $pHrPMrT_2$ cleaved with Eco RI were transcribed in the presence of 5 μl of mouse extract (lanes 1–4) or 7.5 μl HeLa extract (lanes 5–8) in the absence and presence of partially purified human and mouse TTF I. The mouse extract was complemented with 4 μl mTTF I (lane 2) or 7.5 μl hTTF I (lane 4), the human extract with 6 μl mTTF I (lane 6) or 7.5 μl hTTF I (lane 8).

DISCUSSION

Termination of transcription by eukaryotic RNA polymerase I, although factor dependent, bears little resemblance to factor-dependent termination in bacteria. In *E. coli*, termination signals precede the actual sites of termination and exhibit considerable sequence flexibility which makes them difficult to recognize. A protein factor, usually r, interacts with the RNA to assist termination by the polymerase. On the contrary, in pol I transcription terminators it is the DNA to which the factor binds and the signal follows, rather than precedes, the site of termination. This is true for mouse and human, and probably for all eukaryotic class I genes. Although in mouse both the *cis*-acting sequences and the *trans*-acting factors governing rDNA transcription termination have been characterized (3, 5), we are still ignorant of the molecular mechanisms which direct individual steps of the termination process, i.e. the arrest of polymerase progression, the formation of correct 3′ ends and the release of nascent RNA chains. Certainly the stop of the elongation reaction is not brought about by a simple protein blockage mechanism, since we have shown that the factor bound to the Sal box termination signal stops only pol I but not heterologous polymerases, implying that very specific interactions between the termination factor and the transcribing enzyme are involved in the termination reaction (6). The results presented in this paper showing that the human factor interacts with the mouse transcription machinery as well as the finding that the mouse factor bound to its target sequence also blocks pol I from species as distant as yeast (6), suggest that despite different DNA recognition specificities, eukaryotic pol I termination factors have maintained one common protein domain which functionally interacts with pol I. Probably the overall structure of pol I termination factors has been conserved in evolution although the DNA binding properties have diverged. This implies that changes in the sequences necessary for termination have been accompanied by compensatory changes in the part of the protein that interacts with the respective termination signal.

The human 3′ terminal spacer fragment used in this study contains five Sal box elements (T_1–T_5), two of which (T_2 and T_3) contain two or three base exchanges (see Fig. 1). Interestingly, these natural point mutations (C to A and A to C/G transversion at positions 8 and 10, respectively) erase factor binding and transcription termination indicating that these nucleotides play a vital role in the DNA-protein interaction. Indeed, the guanine at position 8 on the coding strand of the Sal box sequence is contacted by both the mouse and human termination factor as revealed by methylation interference experiments (M. Finsterer, unpublished data). In this context it may be worth recalling that also in rat rDNA there are repeated Sal box elements which function as transcription terminators. Most of the eight boxes contain one or more point mutations at different positions, but in contrast to human, not all of them impair factor binding. Thus certain base exchanges within the Sal box consensus sequence are tolerated without affecting DNA-protein interaction and termination activity (23).

There is one intriguing difference in mouse and rat as compared to the human terminator sequence. The sequence of nucleotides 15–18 of the rodent Sal box motif is TCCG. This distal part is an integral part of the rodent termination signal since both point

mutations and insertions between the proximal and distal part of the murine Sal box motif reduce or abolish factor binding (3). Surprisingly, the TCCG/C motif which in human rDNA—contrary to the mouse terminators—always precedes the palindromic part of the human Sal box, does not seem to play an essential role in factor binding and transcription termination. A cloned oligonucleotide which contains only 11 nucleotides of the Sal box is as efficient in factor binding as longer sequences containing the TCCG sequence or additional downstream sequences. Also the efficiency and accuracy of transcription termination as determined in the cell-free system were not affected by flanking gene sequences. Therefore, if there is a functional role of the conserved TCCG motif in 3′ end formation of pre-rRNA, it is not detected in the binding and transcription assay used in this study.

Interestingly, ribosomal genes from various mammals contain a Sal box element in the vicinity of the transcription initiation site whose sequence and position relative to the transcription start site is highly conserved. This upstream terminator stops pol I molecules which have started in the spacer separating individual transcription units and thus prevents transcriptional interference, i.e. disruption of preinitiation complexes at the rDNA promoter (11, 12). Furthermore, it seems to function as an upstream promoter element which exerts a positive effect on the efficiency of initiation (8). Although the present study did not address the question of the role of this upstream terminator in rDNA transcription, we have demonstrated that in human rDNA the 5′ terminal Sal box which exerts a 10-of-11 base identity with the termination signals at the end of the gene represents an active terminator which stops elongating pol I. As shown by competitive gel retardation experiments this human upstream Sal box element exhibits a weaker affinity to its recognizing protein. The binding strength is reduced by a factor of 2. A similar decrease in binding and/or termination efficiency of the upstream terminator has also been observed in mouse suggesting that lowering the binding strength of the promoter-proximal terminator may be functionally relevant.

On the other hand, it is also possible that the upstream and downstream terminators are recognized by different proteins. Both in unfractionated mouse extracts and throughout the purification of TTF I, we routinely observe heterogenous Sal box-protein complexes in gel retardation assays. The four complexes with different electrophoretic mobilities contain 65, 80, 90 and 100 kd proteins which have been identified by UV-crosslinking. None of these proteins represents the 105 kd protein which we have recently purified by sequence-specific affinity chromatography (5). Peptide sequence analysis of this homogenous protein demonstrated that this polypeptide which copurifies with both the specific binding and termination activity is Poly(ADP-ribose)synthetase (Smid et al., manuscript in preparation). Apparently this enzyme co-elutes at four different chromatographic resins together with the Sal box binding proteins which are present in the cell in so minute amounts that they are detected only by activity but not at the protein level. The distinct proteins found in the heterogenous DNA-protein complexes could represent different members of a family of Sal box binding proteins. Alternatively, this heterogeneity could be due to post-translational modification of the Sal box binding protein(s), or to limited proteolysis of an originally larger factor molecule. In HeLa extracts only one complex is detected after electrophoresis in 8% native polyacrylamide gels regardless whether it was formed on the murine or human Sal box oligonucleotide. The human Sal box-TTF I complex has a higher electrophoretic mobility than the fastest migrating mouse complex. Surprisingly, this complex contains two proteins with molecular weights of 63 and 75 kd which closely match the sizes of the two smallest mouse proteins. Thus, the marked differences in the electrophoretic mobilities of the human and mouse Sal box binding proteins are not caused by size differences of the respective proteins but rather appear to reside in the charge of the factor proteins. On the other hand, proteolytic digestion of the DNA-protein complex with proteinase K yields proteolytic intermediates and a protease-resistant core with similar migration properties in both species indicating that the DNA binding domain of human and mouse TTF I has a similar structure. This finding together with the observation that after proteinase K treatment similar size proteins are labelled by UV-crosslinking suggest that the pol I transcription termination factors of both species have diverged both in their target sequence specificity and in some protein domain(s) involved in DNA binding. The part of the protein, however, which is thought to interact with the polymerase and thus mediates the termination event appears to be functionally conserved as demonstrated by the ability of the human factor to block elongation of mouse pol I. Although we do not yet know how the factor interacts with the transcription machinery and thus exerts its biological function we favour the idea that not only at the rDNA promoter but also at the rDNA terminator a coevolutionary adjustment in the pol I cofactor genes to mutations within their respective binding sites has taken place. This rapid evolution of signal sequences within multigene families is likely to provide a major driving force for appropriate compensatory changes in the DNA-interacting proteins (13, 14). Cloning of the genes coding for TTF I of both organisms, defining their functional domains and comparing their primary structure will help to elucidate the molecular mechanisms by which the individual steps of the termination reaction are brought about and will contribute to our understanding how evolutionary changes are spread within multigene families.

ACKNOWLEDGEMENTS

We thank W. Hädelt for help with extract preparations and C. Fieger for participating in part of the experiments. This work was supported by the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie.

REFERENCES

1. Manley, J., Proudfoot, N. J., and Platt, T. (1990). Genes Dev. **3**, 2218–2222.
2. Grummt, I., Maier, U., Öhrlein, A., Hassouna, N., and Bachellerie, J.-P. (1985). Cell **43**, 801–810.
3. Grummt, I., Rosenbauer, H., Niedermeyer, I., Maier, U., and Öhrlein, A. (1986). Cell **45**, 837–846.
4. Kuhn, A. and Grummt, I. (1989). Genes Dev. **3**, 224–231.
5. Bartsch, I., Schoneberg, C., and Grummt, I. (1988). Mol. Cell. Biol. **8**, 3891–3897.
6. Kuhn A., Bartsch I., and Grummt, I. (1990). Nature **344**, 559–562.
7. Bartsch, I., Schoneberg, C., and Grummt, I. (1987). Mol. Cell. Biol. **7**, 2521–2529.
8. Grummt, I., Kuhn, A., Bartsch, I., and Rosenbauer, H. (1986). Cell **47**, 901–911.
9. Henderson, S. and Sollner-Webb, B. (1986). Cell **47**, 891–900.
10. Parker, K. A. and Bond, U. (1989). Mol. Cell. Biol. **9**, 2500–2512.
11. Bateman, E. and Paule, M. R. (1988). Cell **54**, 985–992.
12. Henderson, S. L., Ryan, K., and Sollner-Webb, B. (1989). Genes Dev. **3**, 212–223.

13. Dover, G. A. (1982). Nature **299**, 111–117.
14. Dover, G. A. and Flavell, R. B. (1984). Cell **38**, 622–623.
15. Grummt, I. (1981). Proc. Natl. Acad. Sci. USA **78**, 727–731.
16. Clos, J., Buttgereit, D., and Grummt, I. (1986). Proc. Natl. Acad. Sci. USA **83**, 604–608.
17. Labhart, P. and Reeder, R. H. (1986). Cell **45**, 431–443.
18. McStay, B. and Reeder, R. H. (1986). Cell **47**, 913–920.
19. Financsek, I., Mizumoto, K., Mishima, Y., and Muramatsu, M. (1982). Proc. Natl. Acad. Sci. USA **79**, 3092–3096.
20. La Volpe, A., Simeone, A., D'Esposito, M. D., Scotto, L., Fidanza, V., de Falco, A., Bonicelli, E. (1985). J. Mol. Biol. **183**, 213–223.
21. Kuhn, A., Normann, A., Bartsch, I., and Grummt, I. (1988). EMBO J. **7**, 1497–1502.
22. Grummt, I., Roth, E., and Paule, M. (1982). Nature **296**, 173–174.
23. Kermekchiev, M. B. and Grummt, I. (1987). Nucleic Acids Res. **15**, 4131–4143.