MOLECULAR AND CELLULAR BIOLOGY, July 1987, p. 2521–2529
0270-7306/87/072521-09$02.00/0

Vol. 7, No. 7

Evolutionary Changes of Sequences and Factors That Direct Transcription Termination of Human and Mouse Ribosomal Genes

INGRID BARTSCH, CHRISTA SCHONEBERG, AND INGRID GRUMMT*

Institut für Biochemie, 8700 Würzburg, Federal Republic of Germany

Received 4 February 1987/Accepted 2 April 1987

We have analyzed the sequences required for termination of human rDNA transcription. The human ribosomal transcription unit is shown to extend about 350 nucleotides into the 3′-terminal spacer and ends immediately upstream of a region with a distinct sequence heterogeneity. This heterogeneous region contains a cluster of conserved 10-base pair sequence elements which exert a striking homology to the proximal part of the 18-base pair murine rDNA transcription termination signal sequence, termed *SalI* box. Exonuclease III protection assays and in vitro transcription experiments with both homologous and heterologous human-mouse minigene constructs, and extracts from HeLa or Ehrlich ascites cells, reveal a functional analogy of the human sequence to the mouse *SalI* box. It mediates binding of a nuclear protein which functions as a transcription termination factor. The murine signal sequence is recognized by the human factor but not vice versa. The different sequence specificities and electrophoretic properties of the functionally equivalent protein factors suggest that a molecular coevolution has taken place between the termination signal sequences and the genes coding for the termination factors.

The nontranscribed spacers (NTS) separating tandemly repeated human rRNA genes are approximately 30 kilobases in length and exert a limited length heterogeneity within a specific region downstream of the 3′ end of the 28S rRNA coding region. This microheterogeneity in the human 3′-terminal NTS region consists of essentially four fragment length variants that apparently were generated by duplication of two discrete sequence modules of about 700 and 800 base pairs (bp), respectively. Interestingly, this length variation is confined within some hundreds of base pairs of spacer sequences starting at a *SalI* restriction site at position +365 with respect to the 28S rRNA terminus and includes some closely spaced *SalI* sites (14). These clustered *SalI* restriction sites are part of a larger conserved sequence element, $TCC^{C}_{G}N_{0-3}GGGTCGACC^{A}_{G}$, which is present at least five times in the smallest size variant of the human 3′ spacer.

Recently, we have shown that in mouse RNA polymerase I (Pol I) terminates transcription 565 bp downstream of the 3′ end of mature 28S rRNA immediately upstream of a tandemly arranged repeated sequence motif (AGGTCGA-$CCAG^{TA}_{AT}$NTCCG) that contains a *SalI* restriction site (7). This conserved 18-bp sequence (termed *SalI* box) mediates binding of a nuclear factor and functions as a termination signal for transcription by Pol I. For binding to occur, an intact *SalI* box is required. Mutations both within the palindrome and in the distal part of the consensus sequence impaired factor binding and transcription termination (10). The presence of clustered sequence elements at a similar position in the 3′ spacer with an 8-bp homology to the murine termination signal suggests an analogous function in human rDNA transcription. Since, however, the distal part of the murine signal is not found in the human sequence motifs, it seemed unlikely that the same termination factor would recognize the human signal.

To establish a functional role for the human *SalI* boxlike sequences in transcription termination, we have investigated their interaction with a specific protein(s) in extracts derived from human or mouse cells, as well as their ability to direct transcription termination on minigene constructs containing promoter and terminator sequences from human and mouse ribosomal genes, respectively. The results demonstrate that the human factor recognizes both the homologous and the heterologous *SalI* box sequence, whereas the mouse factor does not interact with the human signal sequence.

MATERIALS AND METHODS

Plasmid constructions. Most of the plasmids used have been described before (7, 10). The human subclones were derived from $pAD_{BB}19a$, which was kindly provided by R. D. Schmickel (4). $pUCHrT_1$ contains a fragment of human rDNA extending from +308 to +431 relative to the end of 28S RNA. pUCBH is the analogous mouse clone covering sequences from +335 to +715. In the point mutant pUCBH589/590 the G and T residues at positions +589 and 590 were converted into C and G residues. The insertion mutant $pUCB_7H$ differs from the wild-type pUCBH by the insertion of seven nucleotides, GATCCCC, between nucleotides 12 and 13 of the first *SalI* box (10). These spacer fragments fused to the mouse rDNA promoter fragment (from −169 to +155 with respect to the initiation site) yielded the murine minigene constructs pPTBH, pPT589/590, and $pPTB_7H$, respectively (10). When linked to the human promoter (an *Eco*RI-*Bst*EII fragment from pHrES covering sequences from −513 to +81 relative to the initiation site), the human-mouse constructs pHrBH, pHr589/590, and $pHrB_7H$ were obtained.

The two human minigene constructs pHrPT and $pHrPTT_1$ contain promoter sequences from −411 to +379 and spacer sequences from +47 to +613 (pHrPT) and from +308 to +431 ($pHrPTT_1$). The heterologous construct pMrPHrT contains mouse promoter sequences from −169 to +155 and human 3′ spacer sequences from +308 to +786.

Transcription assays. Extracts for cell-free transcription were prepared from exponentially growing HeLa or Ehrlich ascites cells. S-100 extracts were prepared according to Weil et al. (21) and nuclear extracts were prepared according to Dignam et al. (3). For in vitro transcription, 25 to 200 ng of

* Corresponding author.

template DNA was incubated in a 50-μl assay in the presence of a mixture of nuclear and S-100 extracts. The reaction cocktail contained 12 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.9), 85 mM KCl, 0.12 mM EDTA, 5 mM $MgCl_2$, 10 mM creatine phosphate, 0.6 mM each ATP, CTP, and UTP, 12.5 μM GTP, and 1 to 2 μCi of [α-^{32}P]GTP. The samples were incubated for 60 min at 30°C and processed for gel analysis as described previously (6).

S1 mapping experiments. S1 mapping experiments were performed essentially as described by Berk and Sharp (1). A 30-μg amount of total RNA extracted from HeLa cells was mixed with 1.5×10^5 cpm (specific activity, 10^6 cpm/μg) of a 3' labeled 567-bp *Ban*I fragment (from +47 to +613) from the human 3' spacer. After ethanol precipitation, the nucleic acids were dissolved in 25 μl of hybridization buffer {80% formamide, 0.4 M NaCl, 0.04 M PIPES [piperazine-*N*,*N'*-bis(2-ethanesulfonic acid)], pH 6.4, and 1 mM EDTA}. The reaction mixture was incubated at 80°C for 5 min and then at 65°C for 3 h. After dilution with 250 μl of S1 buffer, the hybrids were treated for 30 min at 37°C with 60 U of S1 nuclease and subjected to electrophoresis on denaturing 6% polyacrylamide gels. If transcripts from the recombinant minigene constructs were to be analyzed, the cell-free transcription assay was scaled up to 125 μl and RNA was synthesized in the presence of 0.5 mM concentrations of the four unlabeled nucleotides. The RNAs transcribed in vitro were incubated with RNase-free DNase and treated the same as the in vivo samples.

Exonuclease III protection experiments. DNA fragments derived from the human or murine 3' spacer region were 5' labeled at the coding or noncoding strand, respectively, and approximately 10,000 to 20,000 cpm (0.5 to 2 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) containing 1 μg of pUC9 cut with *Hpa*II, 10 μg of yeast tRNA, 1 μg of deoxynucleoside triphosphates, and about 15 to 30 μg of extract protein. After the binding reaction, 6 U of exonuclease III was added, and incubation was continued for a further 6 min. The reaction was terminated by addition of 25 μl of 350 mM NH_4-acetate–10 mM EDTA. 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 (CCCGGGATCCTTCG GAGGTCGACCAGTACTCCGGGCGAC) was labeled at the 3' ends and approximately 5,000 cpm (ca. 0.2 pmol of DNA) was incubated in 25 μl of binding buffer (12 mM HEPES, pH 8.0; 120 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM dithiothreitol, 8% glycerol, 1 μg of pUC9 DNA cut with *Hpa*II) in the presence of about 25 to 50 μg of nuclear extract protein. After 15 min at 30°C, samples were loaded onto low-ionic-strength (7 mM Tris hydrochloride, pH 8.0; 4 mM potassium acetate, 1 mM EDTA) 6% polyacrylamide gels and electrophoresed at 10 V/cm for 2 h at 4°C. The gel was then dried and analyzed by autoradiography.

RESULTS

Sequences similar to the murine termination signal are present in the heterogeneity region of the human NTS. Individual human ribosomal transcription units exert a characteristic length variation within a specific region downstream from the 3' end of the transcribed region (4, 12, 14, 17). The nucleotide sequence of this heterogeneous region has been determined for three different size classes and has been shown to be generated by multiple duplications of a fundamental DNA module of 700 to 800 bp (14). Each of these DNA modules contains several *Sal*I restriction sites. Figure 1 shows the nucleotide sequences surrounding the five *Sal*I sites present in the smallest class of cloned spacer fragments as determined by La Volpe et al. (14). Interestingly, 10 nucleotides are highly conserved, eight of which exert a dyad symmetry and are shared by the mouse *Sal*I box consensus sequence. The conserved nucleotides TCCG (nucleotides 15 to 18 of the murine signal), which are indispensable for the interaction with the mouse termination factor, are not found at a similar position in the human sequence. Instead the motif TCC^C_G precedes the conserved palindromic sequence. The similarities in the localization, the repetitious nature, and the primary structure suggest that this sequence element $TCC^C_GN_{0-3}GGGTCGACC^G_A$ (referred to as the human *Sal*I box) may represent a termination signal for human rDNA transcription.

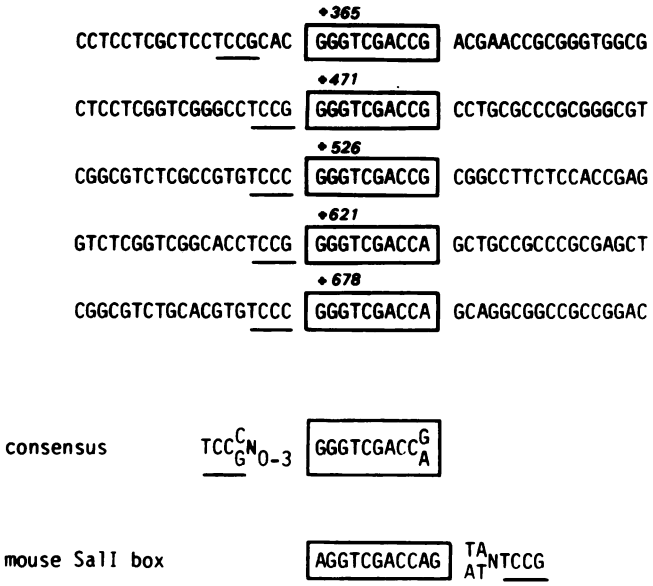

FIG. 1. Alignment of the conserved *Sal*I box sequence motifs present in the human 3' spacer. The sequence is from La Volpe et al. (14). The conserved 10-bp sequence elements containing the *Sal*I restriction sites are boxed. The TCC^C_G motif is underlined. The numbers above the *Sal*I recognition sites mark the distance (in nucleotides) from the 3' end of the 28S RNA coding region. The murine termination signal is shown for comparison.

The 3' end of human pre-rRNA maps immediately upstream of the first *Sal*I box. Recently, we have shown that a part of 3'-terminal spacer sequences are contained in mouse pre-rRNA (9) and that Pol I terminates transcription 565 bp downstream of the 3' end of mature 28S RNA (7). It was not yet known how far the human ribosomal transcription unit extends into the 3' spacer. We therefore mapped the 3' end of human pre-rRNA by S1 nuclease protection experiments. Total RNA from HeLa cells was hybridized to a 3'-end-labeled DNA fragment which encompasses spacer sequences from +47 to +613 relative to the 28S RNA terminus. After treatment with S1 nuclease, one major protected band about 300 nucleotides (nt) long was observed (Fig. 2B, lane 1) which maps the 3' end of human pre-rRNA approximately 345 bp downstream of the 28S RNA coding region, i.e., about 20 bp upstream of the first *Sal*I site at position +365.

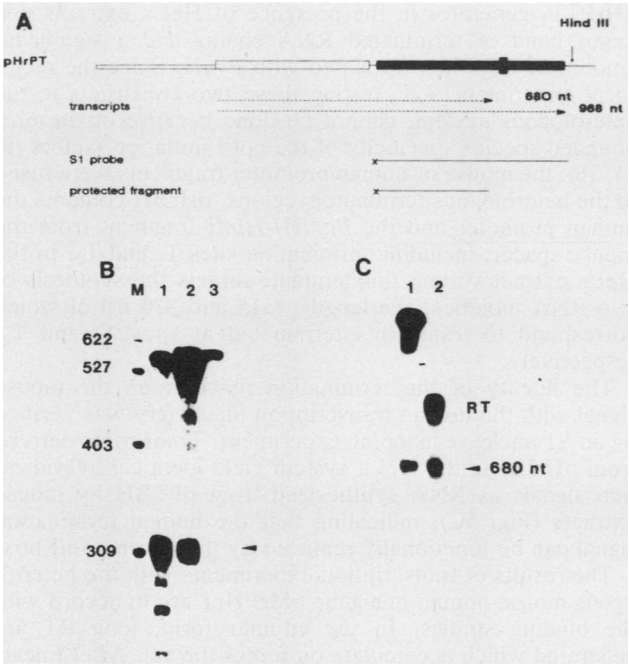

FIG. 2. Termination of human rDNA transcription. (A) Structure of human ribosomal minigene construct pHrPT and the transcripts synthesized in vitro. The thin open bar represents 411 bp of 5′-terminal spacer sequences, the thick open bar represents 379 bp of the transcribed region, and the closed bar represents 3′ spacer sequences. The box marks the *Sal*I box sequence motif at position +365. (B) Nuclease S1 mapping of the 3′ ends of rDNA transcripts synthesized in vivo or in vitro. The 567-bp *Ban*I fragment (from +47 to +613) was 3′ labeled and hybridized to 30 μg of total RNA from HeLa cells (lane 1) or transcripts derived from a 125-μl in vitro transcription assay containing 300 ng of pHrPT as template. As a control, RNA from a nonincubated transcription assay was subjected to hybridization and S1 treatment as well (lane 3). (C) Transcripts synthesized in the HeLa extract system from pHrPT. A 200-ng portion of circular (lane 1) or linear (lane 2) template was used in a 50-μl assay. The 680-nt band represents transcripts terminated in front of the first *Sal*I box; RT are 968 nt.

The 3′ end of transcripts synthesized in vitro from human ribosomal minigene constructs coincides with that of cellular pre-rRNA. To investigate the functional significance of the human *Sal*I box sequence in the process of transcription termination, we fused the human 3′-terminal spacer region to the human rDNA promoter and tested whether these spacer sequences would promote transcription termination on this minigene construct. Figure 2C shows the transcripts generated in the HeLa extract system from clone pHrPT which contains 5′-terminal sequences from −411 to +379 relative to the initiation site and 3′ spacer sequences from +47 to +613 (see also Fig. 2A). In the circular form this template directs the synthesis of one major transcript approximately 680 nt long (Fig. 2C, lane 1). If the template is linearized beyond the 3′ spacer sequences, 968-nt readthrough transcripts (RT) are generated (lane 2).

To check whether the 680-nt transcripts represent faithfully terminated RNA molecules, the 3′ ends of the in vitro products were mapped in parallel to cellular pre-rRNA by an S1 nuclease protection experiment. Transcripts synthesized in the cell-free system (Fig. 2B, lane 2) yield the same protected band as RNA from HeLa cells (lane 1). In addition to correctly terminated RNA chains, a large amount of RT which protect the whole-length fragment is also generated in the extract system. The formation of both the terminated transcripts and the RT is dependent on de novo transcription since reactions that have not been incubated did not contain RNA that hybridized to the labeled probe (lane 3).

HeLa cell extracts contain a protein that binds to the human *Sal*I box sequence. If the human *Sal*I box sequence motif serves a function similar to that of its murine counterpart, then it should interact with a sequence-specific DNA-binding protein. We therefore carried out exonuclease III (exo III) protection experiments, using a 5′-labeled, 154-bp 3′ spacer fragment covering sequences from position +308 to +431. This fragment was incubated with either cytoplasmic or nuclear extracts from HeLa cells and then digested with exo III. In the presence of S-100 extracts that do not contain significant amounts of termination factor (unpublished results), no exo III-resistant fragment caused by protein binding is observed (Fig. 3, lane 1). However, in the presence of nuclear extracts, a new 76-bp fragment that is protected from exo III digestion is generated (lane 2). The size of this fragment reflects the binding of a nuclear protein to the conserved sequence motif and maps the 5′ boundary of the protected region to nucleotide +362 on the coding strand. This position corresponds to the base preceding the human box. To prove the specificity of the observed protein-DNA interaction, the exo III protection experiment was also performed in the presence of specific and unspecific competitor DNA. A 20-fold molar excess of pUCHrT$_1$ DNA completely eliminates the exo III-resistant band (lane 3), whereas the same amount of pUC9 DNA did not affect the binding of the protein to the labeled fragment (lane 4). Thus the signal observed in the exo III protection assay is brought about by the specific interaction of a nuclear factor with the human *Sal*I box sequence.

The human factor recognizes the mouse signal sequence but not vice versa. To study the specificity of sequence recognition of the human and the murine *Sal*I box-binding protein, footprint experiments were carried out with nuclear extracts from both HeLa and mouse Ehrlich ascites cells.

Figure 4 shows the results of exo III assays obtained with the 411-bp *Eco*RI-*Hin*dIII mouse fragment derived from clone pUCBH (7). This fragment extends from positions +335 to +715 in the 3′ spacer of mouse rDNA and includes the two termination signals with the *Sal*I sites at positions +589 and +644. In the presence of nuclear extracts from Ehrlich ascites cells, two main exo III-resistant DNA fragments, 138 and 83 bp long, are observed. The lengths of the protected fragments map the 5′ boundary of the binding region to positions +584 and +639, respectively, i.e., 3 nt upstream of the *Sal*I box sequence. These exo III signals can be specifically competed with plasmid DNA containing an intact *Sal*I box (pUCT$_1$, lane 3), but not with a point mutant (pUCT$_1$589/590) which is unable to interact with the termination factor (lane 4). Substitution of residues GT by CG at positions 589 and 590 (i.e., the third and fourth nucleotide within the first *Sal*I box sequence) abolishes binding of the factor to the signal sequence T$_1$ and leads to an increased interaction with the unmutated distal *Sal*I box T$_2$ (lanes 5 to 8).

When binding to site T$_1$ is impaired, another 110-bp band becomes more prominent (lanes 6 to 8). This band maps a few nucleotides downstream of the first *Sal*I box and cannot be competed by sequences contained in pUCT$_1$. This result suggests that either conformational changes of the DNA or binding of another not yet identified protein to 3′ flanking DNA sequences may account for this exo III-resistant fragment.

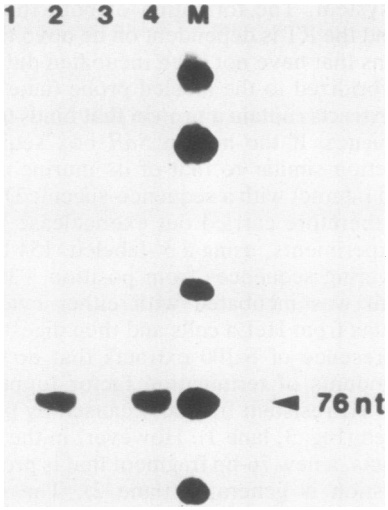

FIG. 3. Binding of a nuclear factor to the human *SalI* box. A human 3′ spacer fragment (from +308 to +431) was 5′ labeled at the coding strand and 20,000 cpm (ca. 1 to 2 ng of DNA) was incubated with about 30 μg of either S-100 (lane 1) or nuclear (lanes 2 to 4) extract proteins. After 15 min, 6 U of exo III was added, and the incubation was continued for another 6 min. The denatured products were separated on 6% sequencing gels along with size marker fragments (pBR322/*HpaII*), shown in lane M. Lanes 1 and 2, No competitor DNA; lane 3, 580 ng of pUCHrT$_1$; lane 4, 560 ng of pUC9. The arrow indicates the position of the 76-nt exo III-resistant fragment.

Essentially the same results are obtained with extracts from HeLa cells (lanes 9 to 16). The lengths, the competition behavior, and relative intensities of the protected fragments are identical to those obtained with the mouse extract, indicating that the human protein recognizes the mouse *SalI* box sequence and does not tolerate base exchanges within the palindromic region.

When an analogous experiment was performed with a labeled fragment from the human 3′ spacer in the presence of mouse extracts, no specific exo III-protected fragment was generated (data not shown). This finding is in accord with previous data which demonstrated that destruction of the integrity of the 18-bp sequence motif abolishes binding of the mouse factor (10) and suggests that the murine termination factor requires a larger specific binding sequence as compared to the human counterpart.

Transcription of heterologous minigene constructs. The exo III footprint studies suggested that the human *SalI* box-binding protein is more promiscuous than the mouse factor. If this binding is a prerequisite for transcription termination, then the murine termination sequence should direct termination in HeLa extracts, whereas the human signal should have no effect in the mouse system.

We therefore tested the transcription of several homologous and heterologous minigene constructs in extracts derived from HeLa or mouse cells. The structure of the four templates used is illustrated in Fig. 5A; the transcripts produced from circular and linear templates are shown in Fig. 5B. Clone pPTBH contains both promoter and terminator sequences from mouse rDNA. In the circular form it directs the synthesis of two classes of transcripts, 385 and 440 nt long, which correspond to RNAs terminated at sites T_1 and T_2, respectively. After truncation with *Eco*RI another defined RNA species is generated which represents RT. Similarly, the analogous human minigene construct pHrPTT$_1$ generates in the presence of HeLa extracts one major band of terminated RNA chains and a significant amount of RT when truncated with *PvuII*. Since the reciprocal experiment, i.e., testing these two constructs in the heterologous system, cannot be done because of the pronounced species specificity of the pol I initiation factors (8, 15, 16), the mouse or human promoter fragments were fused to the heterologous terminator regions. pHrBH contains the human promoter and the *Bam*HI-*Hin*fI fragment from the mouse spacer, including termination sites T_1 and T_2. In the HeLa extract system this template directs the synthesis of two RNA moieties, the lengths (315 and 370 nt) of which correspond to transcripts terminated at sites T_1 and T_2, respectively.

The fidelity of the termination reaction on the mouse signal with the human transcription machinery was verified by an S1 nuclease mapping experiment. Transcripts derived from pHrBH in the HeLa system yield identical hybridization signals as RNA synthesized from pPTBH by mouse extracts (Fig. 5C), indicating that the human termination signal can be functionally replaced by the murine *SalI* box.

The results of transcription experiments with the heterologous mouse-human minigene pMrPHrT are in accord with the binding studies. In the circular form, long RT are generated which accumulate on top of the gel. After linearization with *Eco*RI, 637-nt runoff transcripts are observed. Thus the human *SalI* box elements are not recognized by the mouse termination factor.

The distal part of the murine termination signal is dispensable for interaction with the human factor. The mouse *SalI* box consists of 11 and 4 perfectly conserved bases separated by three more variable nucleotides. Both submotifs are essential for factor binding, and the exact spacing between the two elements has been shown to be functionally important. An insertion of seven nt between the two domains (pPTB$_7$H) distorts the interaction of the murine termination factor(s) with the *SalI* box sequence and abolishes termination (10). Since the human sequence does not contain the TCCG motif downstream of the palindromic sequence, we tested whether the 7-bp insertion between nucleotides 12 and 13 of the first *SalI* box, which eradicates termination in the mouse system, will allow factor binding and termination in HeLa extracts.

Figure 6A shows a schematic view of the 3′-terminal spacer fragment contained in the mouse construct pPTBH or the human-mouse rDNA fusion gene pHrBH (also Fig. 5A) and the corresponding base substitutions or the insertion introduced into the first *SalI* box. The transcripts generated from pPTBH, pPTB$_7$H, and the point mutant pPT589/590 and from the analogous clones containing the human rDNA promoter are shown in Fig. 6B. In the mouse system both the insertion and the point mutations abolished termination at site T_1. The human system, however, responds differently to mutations within the box. Similar to the mouse system, the human factor did not tolerate base substitutions within the palindrome (pHr589/590), resulting in a loss of termination at site T_1 and a preferential usage of site T_2. The insertion mutant pHrB$_7$H, on the other hand, still directs termination at site T_1, although with lower efficiency than the control pHrBH.

Complementary results were obtained in binding experiments with the exo III protection assay. The spacer fragments contained in pUCBH and pUCB$_7$H were 5′ labeled and assayed for their ability to specifically interact with the murine or human *SalI* box-binding proteins. No interaction of the mouse factor with the mutated binding site T_1 present

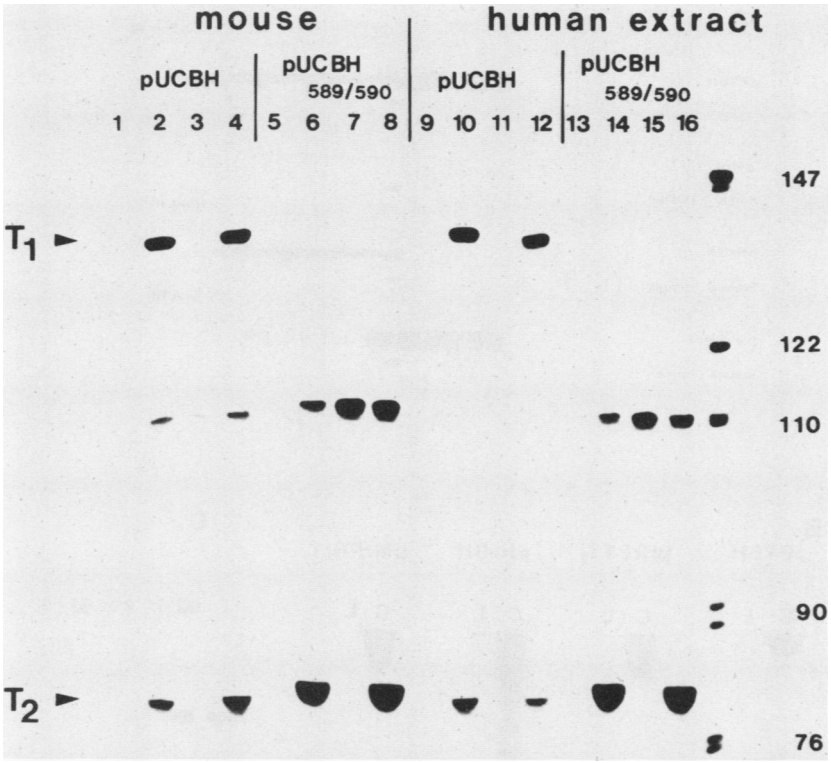

FIG. 4. Interaction of the murine and human factor with mouse termination signal sequences T_1 and T_2. The 411-bp *Eco*RI-*Hin*dIII fragments from plasmid pUCBH and the corresponding point mutant pUCBH589/590 were 5′ labeled at the coding strand. The binding reactions with murine proteins (lanes 1 to 8) contained 10,000 Cerenkov counts (representing 0.5 to 1 ng of DNA) and 15 μg of either S-100 (lanes 1 and 5) or nuclear (lanes 2 to 4 and 6 to 8) extract proteins. The binding reactions with human proteins (lanes 9 to 16) contained 20,000 Cerenkov counts (1 to 2 ng of DNA) and about 30 μg of either S-100 (lanes 9 and 13) or nuclear (lanes 10 to 12 and 14 to 16) extract proteins. The denatured products were separated on 6% sequencing gels along with size markers (lanes M). Lanes 1 to 4 and 9 to 12 show the protected fragments derived from pUCBH; lanes 5 to 8 and 13 to 16 show those from pUCBH589/590. Lanes 1, 2, 5, 6, 9, 10, 13, and 14, No competitor DNA; lanes 3 and 7, 330 ng of $pUCT_1$; lanes 4 and 8, 330 ng of $pUCT_1589/590$; lanes 11 and 15, 660 ng of $pUCT_1$; lanes 12 and 16, 660 ng of $pUCT_1589/590$. The positions of the 138- and 83-nt exo III-resistant fragments which reflect protein binding to sites T_1 and T_2, respectively, are marked.

in $pUCB_7H$ occurs (Fig. 7). Instead, an increased binding to site T_2 is observed (lanes 2 and 3). However, the 7-bp insertion does not affect binding of the human protein. Both the wild-type site T_1 and the insertion mutant are recognized, indicating that the distal part of the mouse signal sequence is not required for binding the human protein.

The human and mouse *Sal*I box-binding proteins are physically different. The different sequence requirements of the mouse and the human factors suggest that Pol I transcription termination in both organisms is brought about by different proteins. As a crude test to check the identity or difference of the mouse and human *Sal*I box-binding protein, a gel mobility shift was used. Nuclear extracts were incubated in the presence of carrier DNA with a double-stranded 39-bp ^{32}P-labeled oligonucleotide (CCCGGGATCCTTCGGAG GTCGACCAGTACTCCGGGCGAC) containing the murine termination signal T_1. When electrophoresed at low ionic strength in a polyacrylamide gel, the protein-DNA complex migrates slower than free DNA (5, 19). Figure 8 shows the complexes formed when the probe was incubated with mouse (lanes 1 to 3) or HeLa (lanes 4 to 6) extracts. Of the two retarded bands, the lower one represents a site-specific DNA-protein complex as demonstrated by competition with $pUCT_1$ and the base substitution mutant $pUCT_1589/590$. The specific complex was not formed in the presence of $pUCT_1$, but complex formation was not affected by the mutant $pUCT_1589/590$. Interestingly, the mobility of the DNA-protein complexes, which mainly depends on the electrophoretic properties of the protein, is different in the two systems. We infer from this result that termination of human and mouse rDNA transcription is brought about by protein factors which, although functionally identical, exhibit different sequence specificity and physicochemical properties.

DISCUSSION

Recently, we have shown that an 18-bp conserved nucleotide sequence which is present eight times in the mouse 3′ NTS (between positions +589 and +1163 relative to the 28S RNA terminus) functions as a transcription terminator for RNA polymerase I. This sequence motif is the primary target for a specific DNA-binding protein, and the interaction of this protein factor with the *Sal*I box sequence has been shown to be a prerequisite for transcription termination (7, 10). A sequence comparison of rDNA transcription units of different eucaryotes reveals conservation of this termination signal sequence in rat DNA (2, 22). A 10-bp repeated sequence with an 8-bp perfect homology to the mouse *Sal*I box is found in the human NTS (14), and no significant homology is observed in lower eucaryotes (11, 13, 18). Thus

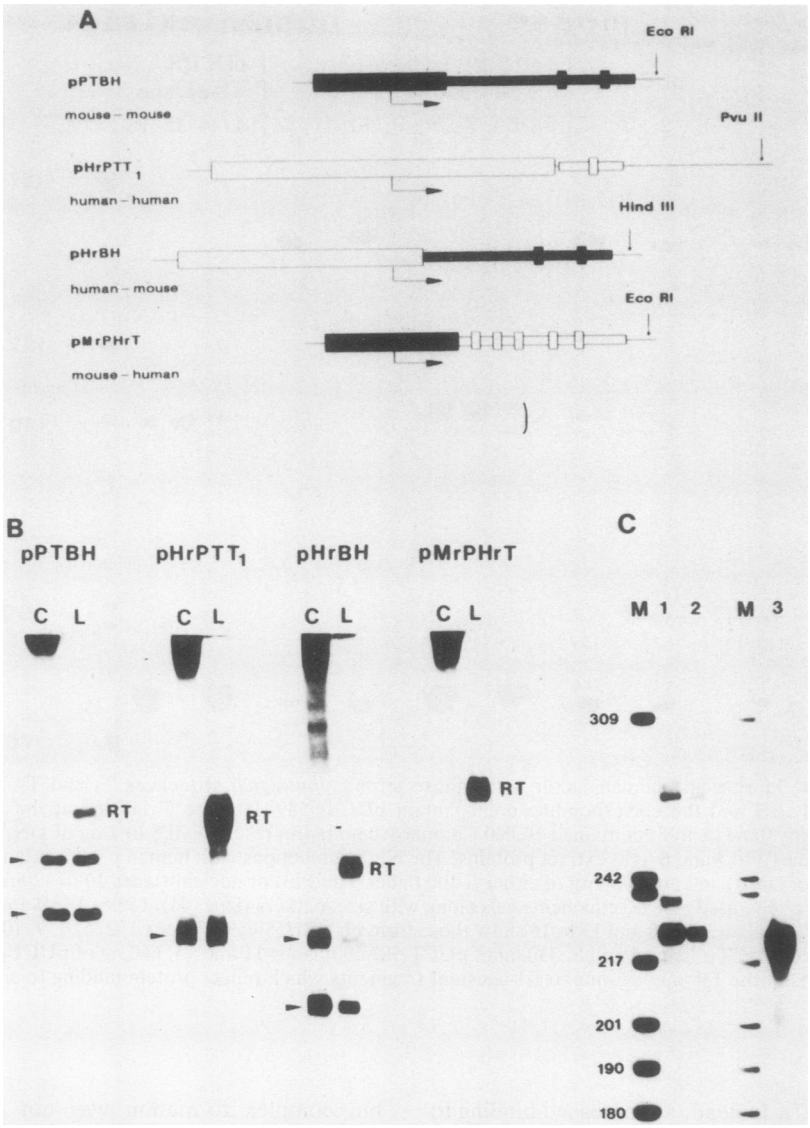

FIG. 5. Transcription of homologous and heterologous human-mouse minigene constructs. (A) Structure of minigene templates of pPTBH, $pHrPTT_1$, pHrBH, and pMrPHrT. The thicker bar represents 5′-terminal rDNA sequences; the thinner bar represents 3′ terminal spacer regions. Mouse rDNA sequences are drawn as solid bars; human sequences are drawn as open bars. The boxes in the 3′ spacer mark the positions of the human and mouse *SalI* box elements. The restriction sites used for truncation of the templates are shown, and the point of transcription initiation is marked by an arrow. The thin line represents pUC9 vector sequences. (B) Transcripts derived from the different minigene constructs in homologous and heterologous extracts. Each template was used both in the circular form (C) or after linearization with the respective restriction enzyme (L). The assays with pPTBH and pMrPHrT contained 100 ng of template DNA and 30 μl of extract from Ehrlich ascites cells. The assays with $pHrPTT_1$ and pHrBH contained 200 ng of DNA and 30 μl of HeLa extract. The arrows point to correctly terminated transcripts. (C) 3′-End mapping of transcripts generated from pPTBH and pHrBH. The *Bam*HI-*Hin*fI fragment from the 3′ spacer of mouse rDNA (from +335 to +715) was 3′-end labeled at the *Bam*HI site and hybridized to RNA transcribed from pPTBH (lane 1) and pHrBH (lane 2) in the mouse and HeLa extract systems, respectively. For control, a reaction with 30 μg of RNA from Ehrlich ascites cells was performed (lane 3). After treatment with S1 nuclease and phenol-chloroform extraction, the ethanol-precipitated samples were denatured and analyzed on a 6% sequencing gel along with size markers (M).

different sequences appear to have evolved to generate correct 3′ ends of pre-rRNA. Moreover, alternative strategies for 3′-end formation of primary rDNA transcripts have been developed. In *Xenopus laevis* and *Drosophila melanogaster*, the formation of pre-rRNA termini has been reported to be brought about by processing rather than by a genuine termination event (13, 20).

To elucidate the molecular specificity of the sequences involved in the formation of pre-rRNA termini, we have investigated whether the sequence element $TCC^{C}_{G}N_{0-3}GG$ $GTCGACC^{A}_{G}$, which is present several times in the heterogeneity region of the human 3′ NTS, serves a similar function as the murine *SalI* box termination signal.

We have shown that the 10-bp sequence element GG $GTCGACC^{A}_{G}$, which is almost identical to the proximal part of the mouse *SalI* box, is recognized by a nuclear factor, which in turn mediates transcription termination. There is a good correlation between factor binding and efficiency of

transcription termination. Substitution of the GT by CG residues abolishes both the interaction with the presumptive termination factor and the termination reaction. By analogy with the mechanism of 3'-end formation of mouse pre-rRNA, we conclude that the human elements are directing true transcription termination and not some type of transcription-linked processing. However, there are no data to unambiguously distinguish between these two possibilities. The human system requires much higher template concentrations to achieve efficient transcription initiation as compared to the mouse system. Thus, at limiting levels of the *SalI* box-binding protein a much higher proportion of RT is generated in the human system (cf. Fig. 5 and 6).

Although the mouse rDNA promoter is not active in the human extract system due to the species specificity of the initiation factor(s) (8, 15, 16), the mouse *SalI* box can functionally replace the human signal; the human sequence, however, is not recognized by the mouse factor. These results are in accord with previous experiments which demonstrated that neither sequence variability nor a change in the spacing between the 11- and 4-bp perfectly conserved sequence motifs is tolerated by the murine protein (10). In humans, a smaller signal sequence mediates factor binding and termination of rDNA transcription. This view is supported by the following observations: (i) insertion of 7 bases between nucleotides 12 and 13 of the mouse *SalI* box eliminates binding of the murine but not the human factor; (ii) point mutations within the conserved bases TCCG (nucleotides 15 to 18 of the murine *SalI* box) strongly reduce the binding and termination activity in the mouse system (10), but have no effect on binding of the human factor (I. Bartsch, unpublished data). Thus, this distal sequence element does not appear to be functionally important for human rDNA transcriptional termination and seems to have been lost during evolution. However, the analogous sequence TCC$^{C}_{G}$ always precedes the human box. At present we do not know whether or not these sequences serve a role in the termination process. As in mice, the human termination signal is repeated several times within a relatively small region of the NTS. Probably most of these sites are functionally equivalent. In the mouse system we could show that all sites function independently and have the same affinity to the termination factor (10). In the human system interaction of

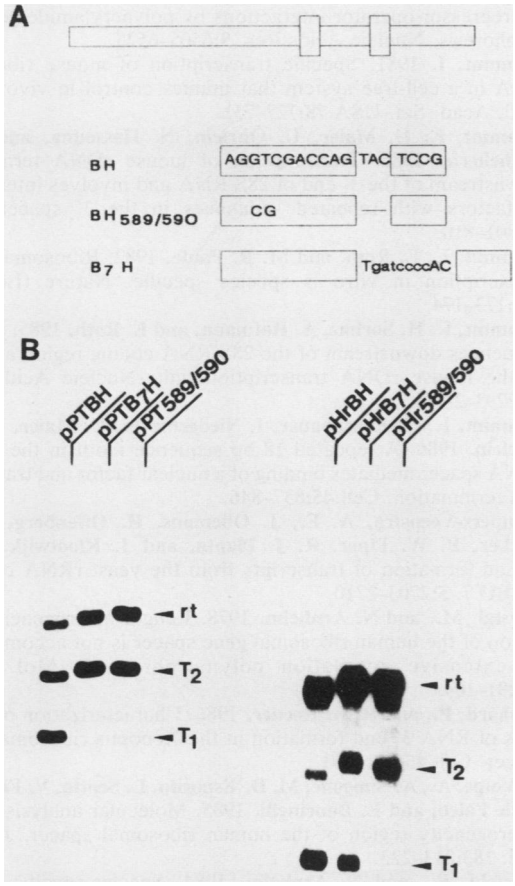

FIG. 6. Effect of mutations within murine termination signal T_1 on transcription termination in mouse and human extract systems. (A) Schematic representation of the 3'-terminal *Bam*HI-*Hin*fI spacer fragment from mouse rDNA which has been fused to either the mouse (pPTBH) or human (pHrBH) promoter. Shown is the sequence of the first *SalI* box, directing transcription termination at site T_1, as well as the base substitutions present in mutant clone BH589/590 and the 7-bp insertion present in the B_7H clones. (B) Transcripts derived from homologous and heterologous minigene constructs containing mutations within the first *SalI* box. The assays with the pPTBH templates contained 100 ng of DNA (truncated with *Eco*RI) and 30 μl of mouse extract. The assays with the pHrBH templates contained 200 ng of linear DNA and 30 μl of HeLa cell extract.

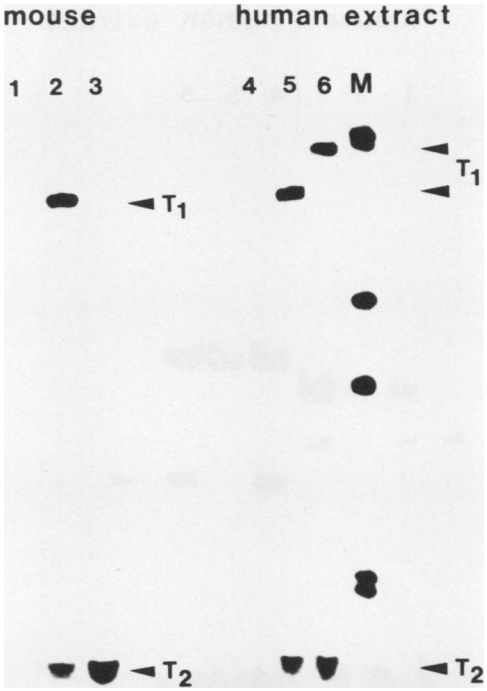

FIG. 7. The human factor binds to insertion mutant $pUCB_7H$. The 266-bp *Ava*I-*Eco*RI fragment from pPTBH and the analogous 273-bp fragment from insertion mutant $pPTB_7H$ containing 3' rDNA spacer sequences from +453 to +715 were 5'-end labeled at the coding strand. The binding reactions with murine proteins (lanes 1 to 3) contained 10,000 Cerenkov counts (0.5 to 1 ng of DNA) and 15 μg of S-100 (lane 1) or nuclear (lanes 2 and 3) extract proteins. The assays with human proteins (lanes 4 to 6) contained 20,000 Cerenkov counts (1 to 2 ng DNA) and about 15 μg of S-100 (lane 4) or nuclear (lanes 5 and 6) extract proteins. Exo III-resistant fragments derived from pUCBH (lanes 1, 2, 4, and 5) and $pUCB_7H$ (lanes 3 and 6) are shown.

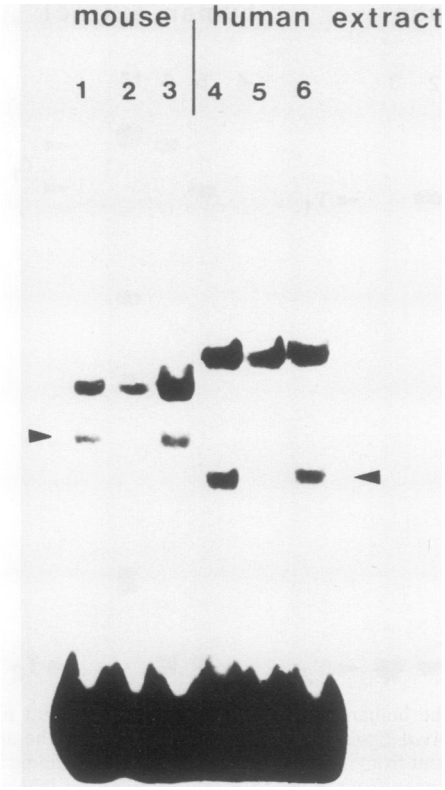

FIG. 8. Electrophoretic mobility shift with human and murine extract proteins. A double-stranded 39-bp synthetic oligonucleotide (CCCGGGATCCTTCGGAGGTCGACCAGTACTCCGGGCGAC) containing the murine termination signal T_1 was labeled at the 3′ ends. For each binding reaction 5,000 Cerenkov counts of labeled DNA (representing about 0.2 pmol of DNA) was incubated at 30°C with 2 μl of extracts containing either 30 μg of mouse (lanes 1 to 3) or 50 μg of human (lanes 4 to 6) nuclear proteins in the presence of 670 ng of competitor DNA (lanes 1 and 4, pUC9; lanes 2 and 5, $pUCT_1$; lanes 3 and 6, $pUCT_1589/590$). After 15 min the samples were electrophoresed on low-ionic-strength gels. The arrows indicate the specific complexes between the labeled oligonucleotide and the human or murine *SalI* box-binding protein.

the factor to the fourth and fifth box is indistinguishable from that of site T_1 (Bartsch, unpublished results). Thus, the tandemly repeated signal sequences may have evolved to ensure termination downstream of the coding region and to prevent RT of the ~30 kilobases of spacer sequences.

A comparison of the nucleotide sequences around the various *SalI* box elements suggests that the repetitive arrangement of termination sites has been generated by a series of duplication events (7, 14). These duplications probably occurred after the termination signal sequences of both species diverged. The changes of the DNA sequences must have been accompanied by a molecular coevolution of the genes coding for Pol I-specific transcription factors. Thus, the divergence of sequences and factors directing transcription termination of rDNA represents another example of evolutionary changes in the structure and function of multigene families. This process, which is probably brought about by unequal crossing over, results in spreading new information through a population and may explain the incompatibility between the rDNA of one species and the transcription machinery of another.

ACKNOWLEDGMENTS

We thank Roy Schmickel for providing the human clone $pAD_{BB}19a$ and E. Zimmer for typing the manuscript.

This work was supported by the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie.

LITERATURE CITED

1. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
2. **Braga, E. A., T. N. Yussifov, and V. V. Nosikov.** 1982. Structural organization of rat ribosomal genes. Restriction endonuclease analysis of genomic and cloned ribosomal DNAs. Gene **20**:145–156.
3. **Dignam, J. D., R. M. Lebowitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **2**:1475–1489.
4. **Erickson, J. M., and R. D. Schmickel.** 1985. A molecular basis for discrete size variation in human ribosomal DNA. Am. J. Hum. Genet. **37**:311–325.
5. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
6. **Grummt, I.** 1981. Specific transcription of mouse ribosomal DNA in a cell-free system that mimics control in vivo. Proc. Natl. Acad. Sci. USA **78**:727–731.
7. **Grummt, I., U. Maier, U. Öhrlein, N. Hassouna, and J.-P. Bachellerie.** 1985. Transcription of mouse rDNA terminates downstream of the 3′ end of 28S RNA and involves interaction of factors with repeated sequences in the 3′ spacer. Cell **43**:801–810.
8. **Grummt, I., E. Roth, and M. R. Paule.** 1982. Ribosomal RNA transcription in vitro is species specific. Nature (London) **296**:173–174.
9. **Grummt, I., H. Sorbaz, A. Hofmann, and E. Roth.** 1985. Spacer sequences downstream of the 28S RNA coding region are part of the mouse rDNA transcription unit. Nucleic Acids Res. **13**:2293–2304.
10. **Grummt, I., H. Rosenbauer, I. Niedermeyer, U. Maier, and A. Öhrlein.** 1986. A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. Cell **45**:837–846.
11. **Kempers-Veenstra, A. E., J. Oliemans, H. Offenberg, A. F. Dekker, P. W. Piper, R. J. Planta, and J. Klootwijk.** 1986. 3′-End formation of transcripts from the yeast rRNA operon. EMBO J. **5**:2703–2710.
12. **Krystal, M., and N. Arnheim.** 1978. Length heterogeneity in a region of the human ribosomal gene spacer is not accompanied by extensive population polymorphism. J. Mol. Biol. **126**:91–104.
13. **Labhard, P., and R. H. Reeder.** 1986. Characterization of three sites of RNA 3′ end formation in the Xenopus ribosomal gene spacer. Cell **45**:431–443.
14. **La Volpe, A., A. Simeone, M. D.'Esposito, L. Scotto, V. Fidanza, A. de Falco, and E. Boncinelli.** 1985. Molecular analysis of the heterogeneity region of the human ribosomal spacer. J. Mol. Biol. **183**:213–223.
15. **Miesfeld, R., and N. Arnheim.** 1984. Species-specific rDNA transcription is due to promoter-specific binding factors. Mol. Cell. Biol. **4**:221–227.
16. **Mishima, Y., J. Finanscek, R. Kominami, and M. Muramatsu.** 1982. Fractionation and reconstitution of factors required for accurate transcription of mammalian ribosomal RNA genes: identification of a species-dependent initiation factor. Nucleic Acids Res. **10**:6659–6670.
17. **Schmickel, R. D., J. R. Waterson, M. Knoller, L. L. Szura, and G. N. Wilson.** 1980. HeLa cell identification by analysis of ribosomal DNA segment patterns generated by endonuclease restriction. Am. J. Hum. Genet. **32**:890–897.
18. **Simeone, A., A. La Volpe, and E. Boncinelli.** 1985. Nucleotide sequence of a complete ribosomal spacer of D. melanogaster. Nucleic Acids Res. **13**:1089–1101.

19. **Strauss, F., and A. Varshavsky.** 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**:889–901.
20. **Tautz, D., and G. A. Dover.** 1986. Transcription of the tandem array of ribosomal DNA in Drosophila melanogaster does not terminate at any fixed point. EMBO J. **5**:1267–1273.
21. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
22. **Yavachev, L. P., O. J. Georgiev, E. A. Braga, T. A. Avdonina, A. E. Bogomolova, V. B. Zhurkin, V. V. Nosikov, and A. A. Hadjiolov.** 1986. Nucleotide sequence analysis of the spacer regions flanking the rat rRNA transcription unit and identification of repetitive elements. Nucleic Acids Res. **14**:2799–2810.