Small Nucleolar RNA Clusters in Trypanosomatid

Leptomonas collosoma

GENOME ORGANIZATION, EXPRESSION STUDIES, AND THE POTENTIAL ROLE OF SEQUENCES PRESENT
UPSTREAM FROM THE FIRST REPEATED CLUSTER*  

Received for publication, July 29, 2003
Published, JBC Papers in Press, November 30, 2003, DOI 10.1074/jbc.M308264200

Xue-hai Liang, Avivit Ochaion, Yu-xin Xu‡, Qing Liu, and Shulamit Michaeli§

From the Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel

Trypanosomatid small nucleolar RNA (snoRNA) genes are clustered in the genome. snoRNAs are transcribed polycistronically and processed into mature RNAs. In this study, we characterized four snoRNA clusters in Leptomonas collosoma. All of the clusters analyzed carry both C/D and H/ACA RNAs. The H/ACA RNAs are composed of a single hairpin, a structure typical to trypanosome and archaea guide RNAs. Using deletion and mutational analysis of a tagged C/D snoRNA situated within the snoRNA cluster, we identified 10-nucleotide flanking sequences that are essential for processing snoRNA from its precursor. Chromosome walk was performed on a snoRNA cluster, and a sequence of 700 bp was identified between the first repeat and the upstream open reading frame. Cloning of this sequence in an episome vector enhanced the expression of a tagged snoRNA gene in an orientation-dependent manner. However, continuous transcript spanning of this region was detected in steady-state RNA, suggesting that snoRNA transcription also originates from an upstream-long polycistronic transcriptional unit. The 700-bp fragment may therefore represent an example of many more elements to be discovered that enhance transcription along the chromosome, especially when transcription from the upstream gene is reduced or when enhanced transcription is needed.

In eukaryotes, pre-rRNA transcript undergoes numerous site-specific modifications before cleavage. Two prevalent modifications, 2′-O-methylation and conversion of uridine to pseudouridine, are guided by C/D box and H/ACA small nuclear RNAs (snoRNAs), respectively (1–3). The C/D snoRNAs carry short conserved motives, as do the C box (RUGA) and D box (CUGA), near the 5′ and 3′ ends, respectively, as well as the internal D′ and C′ boxes. Sequences (>10 nt) upstream from the D or D′ box can interact with the target RNA by perfect base pairing. The 5th nt on the target RNA upstream from the D or D′ box in the duplex is methylated, which is known as the “+ 5 rule” (1). The H/ACA snoRNAs that guide pseudouridylation carry two hairpin structures linked by the H box (AnAnnA) in the hinge region and an ACA box 3 nt upstream from the 3′ end. The internal loop interacts with the target RNA to form the pseudouridylation pocket. The uridine on the target RNA to be isomerized is located usually 14–16 nts upstream from the ACA and/or H box (2, 3).

In vertebrates, all guide snoRNAs are intronic (4, 5). Most yeast snoRNAs are encoded by independent genes, some of which are found in clusters. However, seven intronic genes were also described (6). In plants, most snoRNAs are clustered and are independently transcribed, but intronic snoRNA gene clusters also exist (7–9). The intronic snoRNA genes are transcribed from the promoter of host genes by RNA polymerase II (10, 11). The polycistronic snoRNAs in yeast are transcribed by polymerase II from their own promoters carrying the TATA box and the binding site of Rap1p, a transcription factor involved in ribosomal protein expression (6). In plants, promoters expressing polycistronic snoRNA clusters have not yet been identified. In vertebrates and yeast, only a single snoRNA exists per intron. The maturation of intron-encoded snoRNAs involves largely splicing-dependent processing by exonucleases from linearized and debranched lariats (12). In plants, the processing of intronic snoRNA clusters requires endonucleolytic activity (9).

Trypanosomatids are uncellular protozoan parasites that diverged very early from their eukaryotic lineage. These organisms possess unique RNA processing mechanisms such as pre-mRNA trans-splicing and RNA editing (16, 17). A very unusual characteristic of gene expression in these organisms is the presence of polycistronic transcription. Long transcripts generated by RNA polymerase II are processed by both trans-splicing and polyadenylation (18, 19). In Leishmania, transcription along the chromosome is bi-directional. A unidirectional transcript can cover as much as one-third or even half of an entire chromosome (20). So far, there is no evidence of the existence of a polymerase II promoter that transcribes protein-coding genes in any of the trypanosomatid species. The only promoters identified so far are polymerase I promoters, polymerase III promoters that transcribe the U small nuclear RNAs and 7SL RNA genes, and the polymerase II promoter that transcribes the spliced leader RNA (SL RNA) gene (21). Two protein-coding genes, the Trypanosoma brucei variant surface glycoprotein and the EP (procyctin), are transcribed by RNA polymerase I (22). In the absence of transcriptional regulation...
for most of the protein-coding genes, the expression of mRNAs in trypanosomatids is regulated post-transcriptionally by mRNA processing, stability, and translatability (23).

The probable presence of gene clusters carrying both C/D and H/ACA-like RNAs has been reported in studying snoRNA genes in trypanosomatids (24). The first C/D snoRNA identified in *Leptomonas collosoma* was termed snoRNA-2 and was shown to obey the +5 methylation guiding rule (25). Additional C/D snoRNAs were identified in *T. brucei* by immunoprecipitation with anti-fibrillarin antibody (26). C/D snoRNAs were also found in a locus encoding for the spliced leader-associated RNA (SLA1 RNA) gene (26, 27). Transcription analysis in *L. collosoma* and *T. brucei* suggests that snoRNAs are transcribed as polycistronic RNAs by polymerase II (28, 29). The snoRNA genes are organized as reiterated gene clusters that are repeated more than five times in the genome (25). No promoter activity upstream from the snoRNA-2 gene was detected, because expression of the tagged snoRNA-2 gene was dependent on the transcription from the episomal neo gene (28). Although the trypanosomatin C/D snoRNAs exactly fit the prototype C/D snoRNAs in eukaryotes and archaean, the trypanosome H/ACA RNAs possess unique features. The first identified trypanosomatid H/ACA RNA, h1 RNA, is composed of a single hairpin, whereas in other eukaryotes, H/ACA RNAs carry two hairpins connected by the H hinge region (24). h1 has an AGA box but not an ACA box at the 3′ end. Recently, SLA1 was identified as a guide RNA that directs pseudouridylation at position −12 (relative to the 5′ splice site) on the SL RNA. Like h1, SLA1 also carries a single hairpin and possesses an AGA box at the 3′ end (30).

We have demonstrated recently that snoRNA can be silenced in trypanosomatids in *L. collosoma, Leishmania major,* and *T. brucei.* The silencing in *T. brucei* was achieved by conventional RNA using the T7 opposing system. However, in *Leishmania* and *Leptomonas,* the silencing was elicited by antisense RNA. We propose that silencing in *L. major* and *L. collosoma* mostly probably operates by cleaving the mature RNA after its pairing with the antisense transcript by Rnoise III homologue. However, we detected siRNAs specific to the cleaved transcript, and the amount of antisense RNA needed for silencing was very high, suggesting that siRNAs are probably not catalytic and are not used for cleaving the target RNA via an RNA-induced silencing complex. Although the mechanism of snoRNA silencing is currently unknown, silencing of snoRNAs turned out to be efficient for knocking down C/D but less for inactivating H/ACA RNAs. Silencing of snoRNA-2 almost completely eliminated the modification guided by this RNA (31).

In this study, we analyzed the genome organization of snoRNA in mouse genes encoding for both C/D and H/ACA RNAs in *L. collosoma*. Eight novel H/ACA RNAs were identified, and with no exception, all obey the trypanosomatid prototype pseudouridine guidelines, being a single hairpin RNA and carrying an AGA box at the 3′ end. To determine the sequences essential for the expression of a clustered snoRNA gene, we tagged the B2 gene and performed deletion and mutation analysis by changing the sequences flanking the gene. Mutation of the first 10-nt flanking sequences was deleterious to expression. Moreover, expression of the tagged gene was dependent on its orientation in the vector. In fact, very poor expression of the tagged B2 gene was observed when it was present in the opposite orientation with respect to the neo gene in the vector, because of silencing of the transcript by the antisense effect. Chromosome walk was performed to identify potential regulatory sequences that initiate transcription of the repeated clusters. A 700-bp fragment situated upstream from the first snoRNA repeat and downstream from an open reading frame was shown to enhance transcription in an orientation-dependent manner. However, a long transcript from the ORF to the snoRNA cluster was observed, suggesting that snoRNA genes are part of a long polycistronic transcript. This upstream sequence may therefore represent sequences in the genome that can locally promote transcription and hence enhance transcription of genes with special needs.

**Plasmid Construction—**B2 constructs were generated as described previously (28). Briefly, the B2 snoRNA gene was amplified by PCR with the sense primer 26553, 5′-CCGGATCCCGTTGCTATTTATCGCCG-3′, sense, carrying a BamHI site, is specific for the B2 upstream flanking sequence, from position −63 to −82. 35178, 5′-CCGGTACCGTCATTTATCGCCG-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 93 to 108. 35758, 5′-CCGGATACCAAGCGCCAGGCGCA-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 114 to 127. 35759, 5′-CCGGATCCCGGTGGCGTACG-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 155 to 167. 35759, 5′-CCGGATCCCGGTGGCGTACG-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 155 to 167. 41482, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 87 to 120. 42066, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, is complementary to B2 snoRNA, from position 35 to 54. Long B2 antitag, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, carrying an XbaI site, is complementary to the downstream flanking sequence of B2, from position 93 to 108. 3′-35, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, carrying an XbaI site, is complementary to the downstream flanking sequence of B2, from position 114 to 127. 41482, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, carrying a BamHI site, is complementary to the downstream flanking sequence of B2, from position 87 to 120. 42066, 5′-GATCTTCCGAGCGGCACGGGCGGAATTGCTGACGTT-3′, antisense, is complementary to B2 snoRNA, from position 35 to 54. 43362, 5′-CAATCATTGGCAGGATGTTGC-3′, antisense, is complementary to h1 snoRNA, from position 52 to 69. 22182, 5′-AGCTTTCTGCAATCTGACGCGG-3′, sense, is specific to snoRNA-2 snoRNA, from position 18 to 39, carrying a T deletion in the middle. 38815, 5′-GAGGGAGGAAAGTGTTGACG-3′, antisense, is complementary to neo/hygro mRNA, from position 4996 to 5016 on the pX vector. 19208, 5′-CGCGATATGCCAGAGG-3′, antisense, is complementary to h5 snoRNA, from position 21 to 37. 3′-35, 5′-GAGGGAGGAAAGTGTTGACG-3′, antisense, is complementary to neo/hygro mRNA, from position 4996 to 5016 on the pX vector. 19208, 5′-CGCGATATGCCAGAGG-3′, antisense, is complementary to h5 snoRNA, from position 21 to 37. 3′-35, 5′-GACCTTTTGCCCCAGGCAGAGG-3′, antisense, is complementary to B5 snoRNA, from position 59 to 76. 22076, 5′-GCGGATCCTGCCAGAACCGGCGGAGG-3′, sense, is specific to the upstream ORF, from 361 to 378. 3′-35, 5′-GAGGGAGGAAAGTGTTGACG-3′, antisense, is complementary to the intergenic region, from position 653 to 660. 3′-35, 5′-GCGGATCCCTGCCAGGAGG-3′, antisense, is specific to the intergenic region, from position 1050 to 1067.

**Plasmid Construction—**B2 constructs were generated as described previously (28). Briefly, the B2 snoRNA gene was amplified by PCR with the sense primer 26553, specific to the upstream flanking sequence, and antisense oligonucleotides 35178, 35759, and 35759, to generate constructs containing 15-, 35-, and 75-bp 3′-flanking sequences, respectively. The PCR fragments were inserted into the BamHI site of expression vector pX-neo, in two orientations with respect to the neo gene, as depicted in Fig. 3A. The snoRNA-2 constructs were generated as described previously.
(28) and are depicted in Fig. 5. 58-1 carries a tagged snoRNA-2 gene flanked by 290-nT upstream and 104-nT downstream sequences and is inserted into the BamHI site of the pX-neo expression vector in the same orientation as the neo transcript. 58-2 possesses the same construct as 58-1, but it is present in the opposite orientation with respect to the neo gene. 2-3 carries the 700-bp sequence found upstream to the first snoRNA repeat and is inserted into the Xbal site of the 58-2 construct. The repeats from the snoRNA-2 gene with the same orientation as snoRNA-2, but with an orientation opposite that of the neo gene. 2-4 has the same construct as 2-3, except that the orientation of the upstream sequence is opposite that of the snoRNA-2.

Site-directed Mutagenesis—Site-directed mutations were generated in a two-step process by using PCR. The fragments were amplified using sense primer 26553 and antisense primers 44182 and 44185, carrying 5’ and 3’ mutations, respectively. The first-step PCR products were used as a megaprimer to amplify the full-length gene by second-step PCR with antisense primer 26554. The PCR product was then cloned into the BamHI site of the pX-neo vector. The presence of the mutation was confirmed by sequencing, and the orientation was determined. The constructs carrying the 5’ and 3’ mutations were termed M1 and M2, respectively.

Genomic Library Screening—RNPs enriched in the flow-through fraction of a DEAE column served as a source of snoRNAs, as described previously (25). RNA was separated on a 10% polyacrylamide gel. RNA ranging in size from 70 to 90 nts was excised from a preparative gel and was labeled at the 5’ end with [γ-32P]ATP, as described previously (32). Subsequently, the RNA probe was used to screen the λ-phenom genomic library (33). Clean λ DNA was digested with Sau3A1 and subcloned into the BamHI site of the pBluescript KS+ vector. Two clones obtained from different λ phages, termed B6 and B7, were sequenced.

To clone the sequence present upstream from the first 5’ snoRNA repeats, the L. collosoma genomic library was screened with p2ClAI plasmid (24), which contains the entire repeat of the B2 snoRNA cluster. λ DNA from positive plaque was digested with Sall and SacI and subjected to Southern analysis with a B2-specific probe. A 5.5-kb fragment carrying the B2 cluster was subcloned into the pBluescript KS+ vector. By using direct sequencing from the vector, we identified the sequence upstream from the first snoRNA cluster. The plasmid that carries this fragment was termed UPSS.

Cell Growth and Establishing Stable Cell Lines—L. collosoma cells were grown as described previously (34). The pX-neo derived constructs were transfected into L. collosoma wild-type cells, as described previously (34). Stable cell lines were selected by G418 and were obtained from positive clones, and the presence of the constructs carrying the 5’ and 3’ mutations were termed M1 and M2, respectively.

Identification of Two Clusters Encoding for Both C/D and H/ACA snoRNAs—To screen the genomic library for snoRNAs, 70–90-nt RNAs obtained from the flow-through fraction of a DEAE-Sephacel column (32) were separated from the total RNA and recovered. Next, the eluted RNA was dephosphorylated and labeled with [γ-32P]ATP. The RNA probe was used to screen an L. collosoma genomic library, and several positive clones were selected for further analysis. DNA was isolated from positive clones, and the λ DNA was digested by Sma3A1 and subjected to Southern analysis with the RNA probe described above. The hybridizing fragments were isolated from a preparative gel, subcloned, and sequenced. Two novel loci were identified and termed B6 and B7. The structure of the four loci carrying snoRNA genes is depicted in Fig. 1A. Interestingly, all clusters carry both C/D and H/ACA RNAs. An extensive search revealed six novel H/ACA RNAs in previously reported B2 and snoRNA-2 clusters (24, 25, 28). These new H/ACA RNAs are termed h2 to h9, respectively.

The sequences of the putative H/ACA RNAs were folded using the MFOLD program, www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi, and the predicted secondary structure of the eight new H/ACA RNAs is depicted in Fig. 1B. Interestingly, all of them obey a canonical structure consisting of a single hairpin and carrying an AGA box at the 3’ end. The sizes of these H/ACA RNAs are around 70 nts. Fig. 2A illustrates the putative interaction domain of four of the eight H/ACA RNAs. The predicted pseudouridines are positioned 15–17 nt from the AGA box, and the lengths of the flanking duplexes formed between the snoRNA and rRNA are 4–6 bp, suggesting that these RNAs obey the canonical guiding rules for directing pseudouridylation (2, 3). To verify the presence of such RNAs, we subjected total RNA to both primer extension and Northern analysis, and the results, presented in Fig. 2, B and C, indicate the existence of these small RNAs. These results, together with the previously identified H/ACA RNA, such as h1 and SLA1, suggest that all H/ACA identified so far in trypanosomatid possess a single hairpin structure and carry a 3’ AGA box.

Sequences Flanking the snoRNA Genes Govern Its Processing from a Polycistronic Transcript—We reported previously the analysis of a sequence essential for the expression of the snoRNA-2 gene, which is the first snoRNA within its cluster (Fig. 1A). The data suggested that both the 5’ and 3’ sequences are essential for proper expression. Deletion of the entire 5’ flank still afforded expression of the tagged gene yet at a lower level, suggesting that a promoter of the cluster does not lie immediately adjacent to the first gene within a single cluster and may in fact exist upstream from the entire gene cluster (see below). Deletion of the 3’ flanks affected the expression, and a construct carrying only 82 nt of the upstream and 15 nt of the downstream sequence was expressed but at lower levels (28). In this study, we examined the expression of a snoRNA gene embedded within a cluster. Systematic deletion was performed at the 3’ end of the gene. In addition, site-directed mutagenesis was performed to change the sequence adjacent to the gene from either side. Constructs carrying the same upstream sequence, covering the entire upstream intergenic region of 82 nt and the varying 3’ sequence (75, 35 and 15 bp) were generated and cloned into the pX-neo plasmid in two orientations with respect to the neo transcript, as illustrated in Fig. 3A. Stable cell lines were established, and RNA from wild-type and cell lines carrying tagged B2 and different mutants was subjected to primer extension analysis using an antisense primer, which can extend both the wild-type and the tagged snoRNA gene. The results, presented in Fig. 3A, indicate that the 15-bp 3′-flanking sequence is sufficient to efficiently express the tagged B2 gene. Expression of the tagged gene is, however, strictly dependent on its orientation with respect to the neo transcript (Fig. 3A, panel b, lanes 1–3 and 4–6). We have reported recently (31) that the expression of antisense RNA to the snoRNA gene elicits the degradation of the RNA in a mechanism related to RNA. This mechanism also explains

RESULTS

Identification of Two Clusters Encoding for Both C/D and H/ACA Box snoRNAs—To screen the genomic library for snoRNAs, 70–90-nt RNAs obtained from the flow-through fraction of a DEAE-Sephacel column (32) were separated from the total RNA and recovered. Next, the eluted RNA was dephosphorylated and labeled with [γ-32P]ATP. The RNA probe was used to
the poor expression of the tagged B2 gene in cell lines expressing the antisense B2 in the transcriptional direction of the neo gene (Fig. 3A).

To elucidate the flanking sequences that are essential for snoRNA processing, we introduced site-directed mutations to change the first 10 bp upstream and downstream from B2. The results, presented in Fig. 3B, indicate that tagged genes lacking either the first 10 bp upstream or downstream flanking sequences were not expressed, suggesting that these sequences are essential for the expression. To examine whether these mutations affected the processing of the snoRNA but not its transcription, we performed RT-PCR to detect continuous transcripts originating from the episomes that carry the tagged snoRNA transcripts. A sense primer specific for the vector sequence, situated upstream from the tagged B2 gene, and an antisense primer complementary to the tagged B2 coding region were used to amplify cDNA generated using the antisense oligonucleotide. The results (Fig. 3C) indicate that the tagged B2 gene is indeed transcribed into a polycistronic continuous molecule, because the 300-bp fragment was detected in the cell lines carrying the mutations in the flanking sequences (lanes 1 and 2), as well as in the cell line carrying the tagged B2 (lane 3, positive control) but was absent from wild-type cells (lane 4, negative control). The identity of the second molecular weight PCR product is currently unknown. As a control we used the DNA-removed RNA template in a PCR (Fig. 3C, lanes 5–8), and no contamination was detected. These data suggest that the failure to express the tagged snoRNA mutants in the flanking regions stem from the inability to process the polycistronic snoRNA such that it can interact with the binding proteins to generate a stable snoRNP.

Identification of a Sequence Upstream from all B2 Repeats—Our previous study demonstrated that no promoter activity lies immediately upstream from individual snoRNA genes within the cluster. Because the snoRNA genes are reiterated in the genome (24, 28) and carry 5–6 repeats, a promoter may exist upstream from the first repeat. To identify such a regulatory element, we first needed to identify the sequence present upstream from the first cluster. To this end, an L. collosoma genomic library was screened with a B2-specific probe, and DNA isolated from the pure /H9261 clones was subjected to Southern analysis. We detected a clone, which upon digestion with the restriction enzymes SalI and SacI, showed an additional hybridization fragment (Fig. 4B). This clone can carry either the 5’ or the 3’ flank of the gene cluster. We therefore subcloned this fragment and sequenced it. The gene organization is depicted in Fig. 4A. The sequence analysis is presented in Fig. 4C and indicates that this fragment carries the snoRNA cluster as well as the upstream sequence. Bioinformatic analysis identified a part of an ORF with an unknown function as well as a
sequence of 700 bp that is situated downstream from the ORF and upstream from the first repeat. This fragment could potentially carry the promoter of the gene clusters. Interestingly, the 700-bp fragment encodes for two perfect units of Ts analogous to polypyrimidine tracts essential for cis- and trans-splicing.

**Functional Analysis of the 700 Region as a Potential Promoter of the snoRNA Cluster**—The expression of snoRNA in the pX-neo system is orientation-dependent, as we reported previously (28) and also demonstrated in this study (Fig. 3A). We thought to use this observation and enhance the expression of the tagged snoRNA-2 gene by cloning the upstream fragment to the poorly expressed construct. The constructs that were generated are depicted in Fig. 5A. Construct 58-1 carried the tagged snoRNA-2 gene in a transcriptional direction of the neo gene and served as a positive control for maximum expression. Indeed, primer expression indicates strong expression of both the wild-type and the tagged snoRNA-2 gene (Fig. 5B, lane 1). Construct 58-2 carried the snoRNA-2 gene in an opposite orientation with respect to the neo gene, and in this cell line no tagged snoRNA-2 transcript was observed, and the level of the wild-type transcript was also almost completely eliminated (Fig. 5B, lane 2). In addition, the level of h2, which is present in the same construct, was also reduced as a result of silencing (Fig. 5C, lane 2), as we reported previously (31). Interestingly, cloning the 700-bp upstream fragment to the 58-2 construct (construct 2-3) diminished the silencing effect, and the levels of expression of snoRNA-2 and h2 were partially elevated (lane 3 in Fig. 5, B and C). This phenomenon should stem from the increased level of the sense snoRNA-2 transcript generated in this cell line due to the presence of the upstream element. The additional sense snoRNA-2 transcript increases the cellular amount of snoRNA-2, and consequently there was not a sufficient amount of antisense RNA left to pair with all the snoRNA-2 cellular transcripts. If this is indeed the case, we should be able to detect lower levels of antisense snoRNA-2. In addition, because the snoRNA-2 sense transcript originated from the episome may also carry the neo antisense information, silencing of neo transcript should be observed. To this end, cell lines carrying constructs 58-2 and 2-3 were generated and selected on elevated levels of G418. RNA was prepared from the different cell lines and subjected to primer extension, and the results, presented in Fig. 6A, demonstrate that in these cell lines the level of the neo transcript was indeed very different. As expected, the level of the neo transcript increased in cell line 58-2 due to the increase in the copy number of the plasmid. However, in the 2-3 cell line, the level of the neo mRNA first increased, whereas at a very high drug concentrations, the level of neo mRNA decreased, most probably because of the increased levels of antisense RNA to neo that were promoted by the presence of the 700-bp fragment. This phenomenon is clearly observed when comparing the level of neo transcript in cell lines selected under 600 μg/ml G418 (Fig. 6A), demonstrating the increase in the level of neo mRNA in the 2-3 cell line compared with its level in the 58-2 cell line. In all these experiments the level of the h1 RNA was used to control the level of the RNA in the samples, and the RNA used in B is the same as in A. These results strongly suggest that the 700-bp upstream fragment has the ability to promote the expression of down-stream genes. Interestingly, the 700-bp element elicited its effect on snoRNA-2 silencing in an orientation-dependent manner, because cloning of this fragment in the opposite orientation with respect to snoRNA-2 (construct 2-4) did not affect silencing (lane 4 in Fig. 5, B and C).
The snoRNA Repeats Are Transcribed into a Polycistronic Precursor with Upstream ORF—It is now widely accepted that most of the trypanosomatid genes are transcribed as part of a long polycistronic unit. If the 700-bp region functions in vivo as a promoter, it is expected that the region adjacent to this element will be transcriptionally silent. We therefore examined the steady-state level of transcripts around the 700-bp element. RT-PCR was performed with primers specific to different regions, as depicted in Fig. 7A. The results (Fig. 7, B, lane 3, and C, lane 2) demonstrate the presence of a 700-bp transcript covering the region between the open reading frame and the snoRNA. In addition, we detected RT-PCR products that reflect transcription of a continuous transcript covering this region. The absence of DNA contamination (Fig. 7, B, lane 1, and C, lane 3). These results suggest that the 700-bp region is actively transcribed.

FIG. 3. The expression of tagged B2 snoRNA relies on its orientation in the episomal vector. A. panel a, schematic representation of the constructs. The arrows indicate the orientation of tagged B2 gene in the construct with respect to the neo gene in the vector. 1–3 designate the constructs carrying tagged B2 gene in the same orientation as neo, varying at the 3′-flanking sequences. 4–6 designate the constructs carrying the same B2 gene as in 1–3, respectively, but in the opposite orientation. The length of flanking sequences is indicated. The triangle indicates the tag sequence. Panel b, primer extension analysis of B2. The primer extension was performed with primer 20406, which recognizes both wild-type (WT) and the tagged RNA. The identity of the products is marked with arrows and indicated. The arrows above the lanes indicate the orientation of the tagged gene with respect to the neo gene in the vector. Lanes 1–6, total RNA from the cell lines carrying constructs 1–6, respectively, as illustrated in panel a. B. RT-PCR was performed with primers specific to different regions, as depicted in Fig. 7B. The results (Fig. 7, B, lane 3, and C, lane 2) demonstrate the presence of a 700-bp transcript covering the region between the open reading frame and the snoRNA. In addition, we detected RT-PCR products that reflect transcription of a continuous transcript covering this region. The absence of DNA contamination (Fig. 7, B, lane 1, and C, lane 3). These results suggest that the 700-bp region is actively transcribed.

DISCUSSION

In this study we enlarged the repertoire of H/ACA snoRNAs described so far in trypanosomatids and provided evidence of the role of snoRNA flanking regions for processing the long pre-snoRNA precursors. Moreover, we described the identification of a 700-bp fragment situated between the ORF and the first snoRNA cluster within the snoRNA repeats, and we demonstrated the ability of this fragment to enhance the expression of a tagged snoRNA gene in an orientation-dependent manner. However, this element cannot be considered as a classical promoter because a continuous transcript covering this region was detected that spans this region.

All trypanosomatid snoRNA genes identified so far are present in clusters in the genome, which encode for both C/D and H/ACA RNAs, as we described previously (24). In addition, the SLA1 locus also carries both SLA1 and three C/D RNAs (27, 35). All the clusters analyzed so far are transcribed as polycistronic snoRNA precursors by RNA polymerase II.
Recently, we compiled a large number of C/D snoRNAs in both *T. brucei* and *Trypanosoma cruzi* genomes and found that with no exception these snoRNAs are organized in repeated clusters. All clusters that we identified carry both C/D and H/ACA RNAs. This genome organization resembles mostly the organization in plants, which also carry a cluster encoding for mixed C/D and H/ACA RNAs. In yeast, the clusters encode for only one type of snoRNA (6), and in metazoa such as mice, humans, and *Drosophila*, all snoRNAs are intronic, and each intron encodes for only a single snoRNA (9).

The discovery of the eight newly identified H/ACA RNAs in this study raises the possibility that, most, if not all, trypanosome H/ACA-like RNAs are composed of a single hairpin and carry the AGA box. Moreover, the 23 newly identified H/ACA RNAs from the *T. brucei* genome also carry a single hairpin and an AGA box at the 3’ end, suggesting that this is the prototype of the trypanosomatid H/ACA RNA. Another organism possessing single hairpin pseudouridine guides is archaea (36, 37), whereas all eukaryotic H/ACA snoRNAs carry two hairpins connected by the hinge region. Interestingly, archaea were also shown to possess H/ACA-like RNA that carry multiple hairpins. We currently have no evidence of the existence of such RNAs in trypanosomatids. The structure of trypanosomes, archaea, and eukaryotic H/ACA RNA suggests that the single hairpin H/ACA RNA may have been the ancestor of multihairpin H/ACA RNAs that may have originated from the fusion of adjacent single hairpin RNAs. This hypothesis is supported by the fact that each of the two hairpins associates with one set of four core proteins (38).

---

2 X.-h. Liang, Q. Liu, and S. Michaeli, unpublished data.
Based on the fact that in trypanosomatids almost all snoRNA genes are clustered and transcribed into polycistronic precurso-s, the maturation of individual snoRNAs should involve the endonucleolytic cleavage and exonucleolytic trimming, like the
case of yeast polycistronic snoRNAs. In higher eukaryotes, most of the C/D box RNAs carry a short (4–6 bp) stem structure
formed between the 5′/H1 and 3′/H1 coding sequences. This stem
serves as a signal for pre-RNA processing (39, 40). In trypano-
somes, the flanking sequences of most snoRNAs have the po-
tential to form a duplex (24, 28), and it was shown that the
flanking sequences are required for snoRNA accumulation (28).
Indeed, mutational studies performed during this study indi-
cate the importance of the first 10-bp flanking sequence for
snoRNA processing, because in these mutants the polycistronic
transcript was unable to produce a stable tagged B2 transcript.
However, 15 bp of the 3′-flanking sequence was sufficient to
obtain maximal expression of the tagged B2 gene, suggesting
that the processing signals lie very close to the 3′ end of the
snoRNA. The duplex that is potentially formed between the
10-nt flanking sequence may serve as a signal for endonucleo-
lytic cleavage, similar to pre-snoRNA processing by Rnt1 in
yeast (6). Ongoing studies in our laboratory are attempting to
identify the role of exosome components recently identified in
Trypanosomes as described previously and also illustrated in
this study, the expression of antisense RNA to the tagged snoRNA
gene eliminated the production of both the sense-tagged snoRNA trans-
script as well as the wild-type transcript. This silencing was

![Fig. 5. The ability of the 700-bp upstream sequence to enhance transcription. A, schematic representation of the constructs. The orientation of the tagged gene is marked with arrows, and the names of the constructs are given. The sites of restriction enzymes are indicated. pX-neo, the expression vector; h2, the h2 snoRNA; sno2, the snoRNA-2 RNA; up, the 700-bp sequence upstream from B2 repeats. B, primer extension analysis to determine the expression of the tagged snoRNA-2 gene. Primers used in the reaction are 16865 and 43362, complementary to the wild-type and tagged snoRNA-2, and h1 RNA, respectively. The extension products are marked with arrows and are indicated. Lane WT, total RNA from wild type; lanes 1–4, RNA from cell lines carrying the constructs 58-1, 58-2, 2-3, and 2-4, respectively. C, primer extension analysis to determine the level of h2 RNA using primers 22076 and 43388, specific to h2 and B5 snoRNA, respectively. The extension products are marked with arrows and are indicated. Lanes 1–4, RNA from 58-1, 58-2, 2-3, and 2-4 cell lines, respectively. M, marker was as described in Fig. 2. The sizes of the marker in nucleotides (nt) are indicated.](image)

![Fig. 6. The effect of the upstream fragment cloned on the level of neo mRNA and the snoRNA-2 antisense transcript. A, 58-2 and 2-3 cell lines were selected with an elevated G418 level as indicated above the lanes (in µg/ml). Total RNA was prepared from different cell lines and subjected to primer extension using primers 36815 and 43362, complementary to neo mRNA and h1 snoRNA, respectively. The extension products are marked and indicated. B, detection of the level of snoRNA-2 antisense transcript. The same RNA that was used in A was subjected to primer extension using a sense primer 22182 specific to snoRNA-2 that extends the antisense transcript. The product is marked and indicated. The numbers above the lanes indicate the concentration of G418 in which the cell lines were selected. WT, wild type.](image)
accompanying by the production of siRNAs (31). However, we propose that the mechanism of silencing may differ from the conventional RNAi mechanism. The siRNAs we detected may be produced in this cleavage are not catalytic and cannot promote conventional RNAi mechanism. The siRNAs we detected may be produced because it is known that due to steric hindrance transcription cannot take place simultaneously on two strands, the transcription from the other strand encoding for the antisense snoRNA-2 transcript was most probably reduced (construct 2-3). As a result of increasing the snoRNA-2 transcript and decreasing the antisense to snoRNA-2 silencing of this RNA was dimmed. The 700-bp element acts only in an orientation-dependent manner, because attenuation of silencing was detected only when this fragment was present in the transcription direction of the snoRNA but not in the opposite orientation (construct 2-4) (Fig. 5).

Identification of this 700-bp fragment as a potential promoter is not trivial. So far, all other trypanosome small RNA genes were shown to have their own promoters. The small nuclear RNAs and 7SL RNA genes are transcribed by polymerase III by an extragenic tRNA that serves as a promoter (43). The SL RNA gene is transcribed by polymerase II; its promoter as well as the transcription factors and their binding sites were characterized (21). However, proving the existence of RNA polymerase II promoters for protein-coding genes is very elusive. Nevertheless, there are reports to support the existence of such promoters (44, 45). However, in Leishmania and Leptomonas it is possible to obtain expression of a marker from episomes containing only the plasmid backbone; the marker gene is preceded by a splice acceptor and a synthetic poly pyrimidine tract but with no obvious promoter (46, 47). However, research from the laboratory of P. Myler and K. Stuart (Seattle Biomedical Research Institute) supports the existence of promoters along the Leishmania chromosomes. Stably maintained episomes carrying domains from the inflection point that marks the bi-directional transcription chromosome I expressed up to 10-fold more protein from the reporter gene than analogous episomes lacking this region (23). In addition, nuclear run-on experiments on chromosome III also indicated that both promoters and terminators are present in regions where transcription changes direction and where the tRNA gene is found at the end of a transcription unit. Interestingly, as depicted in Fig. 1, we also identified a tRNA-like molecule at the 3’ flank of the snoRNA cluster. Surprisingly, the different coding strand inflection points in Leishmania do not show any sequence similarity. All the data, however, support the existence of a few promoters per chromosome in Leishmania. However, these promoter regions do not share sequence similarity and may not be conventional TATA-containing promoters like promoters in other eukaryotes.

We propose that the 700-bp element that we identified here is one of the several “promoter-like” sequences present along the chromosome, because this sequence is able to enhance the expression of the snoRNA gene in an orientation-dependent manner. Interestingly, this fragment is rich in polypyrimidine tracts that are known to attract the splicing machinery (48, 49). Because transcription and splicing are coupled, and the machinery of polymerase II is equipped with all the splicing and polyadenylation factors, this domain may recruit transcription factors via preferential binding of the splicing and polyadenylation factors. However, the presence of a transcript encoding for the 700-bp fragment does not allow us to define this region as a conventional promoter sequence, because normally promoter regions themselves are not transcribed. However, one can envision a mechanism for transcriptional regulation in these parasites that is assisted by potential elements along the long chromosome. These sequences may operate only in cases where polycistronic transcription originating from the upstream region ceased. These promoter-like sequences may therefore serve asreviving stations to rescue transcription along the chromosome, keeping the chromosome transcriptionally active. However, these elements may also have a special regulatory role when there is an immediate need to enhance
transcription in order to support growth. For instance, snoRNAs are in high demand when the rRNA synthesis level is elevated during accelerated growth. It is too simplistic to assume that transcription of half or more of an entire chromosome depends on initiation only from a single region present at the inflection point that marks the bi-directional transcription of the trypanosomatid chromosomes. It will therefore be of interest to examine other intergenic regions upstream from the first repeated genes for their ability to promote transcription.

Acknowledgments—We thank Elisabetta Ullu and Christian Tschudi for stimulating discussions.

REFERENCES
1. Kiss-László, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., and Kiss, T. (1996) Cell 85, 1077–1088
2. Ni, J., Tian, A. L., and Fournier, M. J. (1997) Cell 89, 565–573
3. Ganz, P., Bortolin, M. L., and Kiss, T. (1997) Cell 89, 799–809
4. Terns, M., and Terns, R. (2002) EMBO J. 21, 1881–1888
5. Liang, X., Liu, L., and Michaeli, S. (2001) J. Biol. Chem. 276, 40313–40318
6. Qu, L. H., Henras, A., Lu, Y. J., Zhou, H., Zhou, W. X., Zhu, Y. Q., Zhao, J., Henry, Y., Caizergues-Ferrer, M., and Bachellerie, J. P. (1999) Mol. Cell. Biol. 19, 1144–1158
7. Brown, J. W., Echeverria, M., and Qu, L. H. (2003) Trends Plant Sci. 8, 42–49
8. Liang, D., Zhou, H., Zhang, P., Chen, Y. Q., Chen, X., Chen, C. L., and Qu, L. H. (2002) Nucleic Acids Res. 30, 2363–2372
9. Brown, J. W., Echeverria, M., and Qu, L. H. (2003) Trends Plant Sci. 8, 42–49
10. Kiss, T. (2001) EMBO J. 20, 3617–3622
11. Kiss, T. (2002) Cell 109, 145–148
12. Filipowicz, W., and Pogacic, V. (2002) Curr. Opin. Cell Biol. 14, 319–327
13. Chanfreau, G., Legrain, P., and Jacquier, A. (1998) J. Mol. Biol. 284, 975–988
14. Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999) EMBO J. 18, 5399–5410
15. van Hoof, A., LENnertz, P., and Parker, R. (2000) Mol. Cell. Biol. 20, 441–452
16. Agabian, N. (1999) Cell 101, 1157–1160
17. Simpson, L., Shiego, S., and Apashevich, R. (2003) RNA (New York) 9, 265–276
18. Huang, J., and van der Ploeg, L. H. (1991) Mol. Cell. Biol. 11, 3180–3190
19. Ullu, E., Matthews, K. R., and Tschudi, C. (1993) Mol. Cell. Biol. 13, 720–725
20. Myler, P. J., Suck, E., McDonagh, P. D., Martinez-Calvillo, S., Schauer, A., Sunkin, S. M., Yan, S., Madhubala, R., Ivens, A., and Stuart, K. (2000) Biochem. Soc. Trans. 28, 527–531
21. Campbell, D. A., Sturm, N. R., and Yu, M. C. (2000) Parasitol. Today 16, 78–82
22. Lee, M. G., and Van der Ploeg, L. H. (1997) Annu. Rev. Microbiol. 51, 463–489
23. Clayton, C. E. (2002) EMBO J. 21, 1881–1888
24. Liang, X., Liu, L., and Michaeli, S. (2001) J. Biol. Chem. 276, 40313–40318
25. Levitan, A., Xu, Y., Ben-Dov, C., Ben-Shlomo, H., Zhang, Y., and Michaeli, S. (1998) Nucleic Acids Res. 26, 1775–1783
26. Dunbar, D. A., Wormsley, S., Lowe, T. M., and Baserga, S. J. (1996) J. Biol. Chem. 271, 14767–14776
27. Roberts, T., Dungan, J., Watkins, K., and Agabian, N. (1996) Mol. Biochem. Parasitol. 83, 163–174
28. Xu, Y., Liu, L., Lopez-Estraño, C., and Michaeli, S. (2001) J. Biol. Chem. 276, 14289–14296
29. Dunbar, D. A., Chen, A. A., Wormsley, S., and Baserga, S. J. (2000) Nucleic Acids Res. 28, 2855–2861
30. Liang, X. H., Xu, X. Y., and Michaeli, S. (2002) RNA (New York) 8, 237–246
31. Liang, X. H., Liu, Q., and Michaeli, S. (2003) Prog. Natl. Acad. Sci. U. S. A. 100, 7521–7526
32. Xu, Y., Ben-Shlomo, H., and Michaeli, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 8473–8478
33. Goldring, A., and Michaeli, S. (1995) Gene (Amst.) 156, 139–144
34. Goldring, A., Zimmer, Y., Ben-Yehuda, E., Gencarová, I., and Michaeli, S. (1996) Exp. Parasitol. 84, 28–41
35. Roberts, T. G., Sturm, N. R., Yee, B. K., Yu, M. C., Hartshorne, T., Agabian, N., and Campbell, D. A. (1998) Mol. Cell. Biol. 18, 4409–4417
36. Rozhdestvensky, T. S., Tang, T. H., Tchirkova, J. V., Brossus, J., Bachellerie, J. P., and Huttenhofer, A. (2003) Nucleic Acids Res. 31, 869–877
37. Bachellerie, J. P., Cavaillé, J., and Huttenhofer, A. (2002) Biochimie (Paris) 84, 775–790
38. Watkins, N. J., Gottschalk, A., Neuhauer, G., Kastner, B., Fabrizio, P., Mann, M., and Lührmann, R. (1998) RNA (New York) 4, 1549–1568
39. Caffarelli, F., Patacchiola, S., De Gregorio, E., Frugapane, P., and Bozzoni, I. (1996) EMBO J. 15, 1121–1131
40. Estvez, A. M., Kempf, T., and Clayton, C. (2001) EMBO J. 20, 3831–3839
41. Robinson, K. A., and Beverley, S. M. (2003) Mol. Biochem. Parasitol. 128, 217–228
42. Nakaar, V., Dare, A. O., Hong, D., Ullu, E., and Tschudi, C. (1994) Mol. Cell. Biol. 14, 6736–6742
43. McKendree, M., Graham, S., Hartmann, C., and Clayton, C. (1998) Exp. Parasitol. 90, 65–76
44. Downey, N., and Donelson, J. E. (1999) Mol. Biochem. Parasitol. 104, 25–38
45. Bellofatto, V., Torres-Munoz, J. E., and Cross, G. A. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 6711–6715
46. Curotto de Lafaille, M. A., Laban, A., and Wirth, D. F. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 7073–7077
47. Wagner, E. J., and Garcia-Blanco, M. A. (2001) Mol. Cell. Biol. 21, 3281–3288
48. Valcarcel, J., and Gebauer, F. (1997) Curr. Biol. 7, R705–R708