The trans-Spliceosomal U4 RNA from the Monogenetic Trypanosomatid Leptomonas collosoma

CLONING AND IDENTIFICATION OF A TRANSCRIBED tRNA-LIKE ELEMENT THAT CONTROLS ITS EXPRESSION*

(Received for publication, September 13, 1999)

Liu Li‡, Leo R. Otake§, Yu-xin Xu¶, and Shulamit Michaeli‡

From the ‡Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel the §Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06536-0812, and the ¶Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

U4 small nuclear RNA is essential for trans-splicing. Here we report the cloning of U4 snRNA gene from Leptomonas collosoma and analysis of elements controlling its expression. The trypanosome U4 RNA is the smallest known, it carries an Sm-like site, and has the potential for extensive intramolecular base pairing with the U6 RNA. Sequence analysis of the U4 locus indicates the presence of a tRNA-like element 86 base pairs upstream of the gene that is divergently transcribed to yield a stable small tRNA-like RNA. Two additional tRNA genes, tRNAPro and tRNAGly, were found upstream of this element. By stable expression of a tagged U4 RNA, we demonstrate that the tRNA-like gene, but not the upstream tRNA genes, is essential for U4 expression and that the B box but not the A Box of the tRNA-like gene is crucial for expression in vivo. Mapping the 2′-O-methyl groups on U4 and U6 small nuclear RNAs suggests the presence of modifications in canonical positions. However, the number of modified nucleotides is fewer than in mammalian homologues. The U4 genomic organization including both tRNA-like and tRNA genes may represent a relic whereby trypanosomatids "hired" tRNA genes to provide extragenic promoter elements. The close proximity of tRNA genes to the tRNA-like molecule in the U4 locus further suggests that the tRNA-like gene may have evolved from a tRNA member of this cluster.

In trypanosomes all pre-mRNAs are produced by trans-splicing. In this process a common short spliced leader (SL)3 derived from a small RNA, the SL RNA, is added to each mRNA. trans-splicing is mechanistically related to cis-splicing (1). As in cis-splicing, U snRNAs are required. Trypanosome counterparts to U2, U4, and U6 have been characterized and shown to function in trans-splicing (2, 3). Trypanosome snRNAs are generally smaller than their cis-splicing counterparts. Only recently was the trypanosome U5 homologue identified and its gene cloned and sequenced (4, 5). Evidence supports the presence of a trypanosome tri-snRNP complex carrying the U4/U6/U5 snRNAs (5). However, the trypanosome U5 snRNA may have a unique role in trans-splicing based on its potential to interact with the SL RNA intron region by base pairing, as supported by in vivo cross-linking experiments (4) and phylogenetic conservation (5).

The majority of trypanosome snRNAs carry a divergent Sm site. The Sm site was shown in other eukaryotes to bind core proteins that are common to all snRNPs and are recognized by sera from autoimmunne patients (6). The trypanosome snRNPs are the only ones so far that are not recognized by anti-Sm sera that recognize Sm proteins from yeast to man. Surprisingly, however, trypanosomes do possess Sm proteins, because recently an SmE homologue was identified among the core proteins that bind the SL RNA (7).

In mammals and yeast, U snRNA genes, except U6, are transcribed by RNA polymerase II (8). The promoters of these genes include distal and proximal elements and TATA boxes. In trypanosomes, in contrast, all small RNAs including U2, U3, U4, U6, and 7SL RNA are transcribed by polymerase III (9). In the last three cases, A and B box elements located upstream of the gene were shown to be essential for transcription in vivo. These control elements are part of the internal control regions of divergently oriented tRNA genes or in the case of the U2 snRNA, of a tRNA-like gene (10).

To study further the assembly of trypanosome U snRNAs in vivo, we have cloned and sequenced the U4 gene from the monogenetic trypanosomatid Leptomonas collosoma. This completes the set of spliceosomal RNAs from this trypanosomatid species. The U4 gene locus represents a new mode of genomic organization, harboring both tRNA-like and tRNA genes upstream of the coding region. Like U2 snRNA, but unlike U6 snRNA, the B box sequence of the tRNA-like molecule plays a major role in the expression of the gene.

EXPERIMENTAL PROCEDURES

Oligonucleotides—Oligonucleotide 424, 5′-GCACATATGAACAT-CACTGAGG-3′; 425, 5′-ATACTAAATAACCCAGTACTCC-3′; 25005, 5′-CCCCAGCCTGTTGCGAGTACA-3′ antisense to positions 94–113 of the U4 RNA; 25002, 5′-GGGAATCGAACCCGGGCCTC-3′ antisense complementary to positions 1–20 of U4; 26264, 5′-GGGCGTCTGCGCCGCTGCGC-3′ antisense complementary to positions 120 to 140 in the tRNA-like RNA; 24772, 5′-CAGCTAC-TCTATACGGAGGAAAGAT-3′ carrying insertion (underlined) to NOI site in position 82; 24422, 5′-AGGGCTGCCTGCC AGCTCCAAAGGCGGG-3′ sense complementary to positions 144 to 171 carrying substitutions (underlined) in the B box positions —153 to —162; 24427, 5′-TTCCTCTGGGCTCAGACACCGCGCAGGGCCG-3′ sense complementary to positions —102 to —129 carrying substitutions.

This paper is available on line at http://www.jbc.org
Trypanosomatid U4 snRNA

polyacylamide–7 m urea gel and electroblotted onto a Nytran membrane (Hybond, Amersham Pharmacia Biotech). Hybridization with labeled oligonucleotides was performed at 42 °C in 5× SSC, 0.1% SDS, 5× Denhardt’s solution, and 100 µg/ml salmon sperm DNA. Primer extension was performed using end-labeled oligonucleotide (100,000 cpm/pmol). Reactions were performed as described previously (12) were analyzed on a 6% polyacrylamide–7 m urea gel next to DNA sequencing reaction performed with the same primer. Primer extension reaction for mapping of 2′-O-methylations was performed at 0.02 and 5 mM concentrations of dNTPs using 5′-end-labeled oligonucleotides complementary to the 3′-end of U4, U5, and U6 snRNAs (13, 14).

RESULTS AND DISCUSSION

The U4 Gene Is Linked to tRNA-like and Two tRNA Genes—By inverse PCR using primers derived from the Trypanosomatid brucei U4 snRNA sequence (2), a Sau3AI 600-bp genomic fragment was cloned and sequenced. An RNA probe synthesized from this clone was used to screen a genomic library. A 1.2-kilobase AluI fragment carrying the gene was subcloned and sequenced. The sequence contains the U4 coding region 432 bp upstream and 253 bp downstream of sequences (Fig. 1). Sequence analysis indicated the presence of two tRNA genes, tRNAPro and tRNAArg, located at positions –223 and –372, respectively. The tRNAPro is 90% identical to the tRNAPro of humans, mice, and Caenorhabditis elegans. Positions 4–53 of tRNA Gly are identical to the tRNAGly of humans, mice. The tRNAPro is transcribed in the same direction as the U4 gene, whereas the tRNAArg is divergently transcribed. The A boxes of tRNAPro and tRNAArg (5′-TGGTCTAGTGG-3′ and 5′-TGGTCTAGTGG-3′, respectively) conform to the consensus box A sequence 5′-TTRYNAGTGG-3′ (the most conserved positions are underlined) and 5′-TTRYNAGTGG-3′, respectively) are consistent with the box B consensus 5′-GTTCCRANNC-3′. Both tRNAs can fold into the canonical tRNA cloverleaf structure.

The genomic arrangement of the L. collosoma U4 locus does not resemble the loci of T. brucei U3 (15), L. collosoma U5 (5), and L. collosoma U6 (11, 16, 17); each of which has a divergently transcribed tRNA gene located 95–96 bp upstream of the U snRNA sequences (Fig. 2). We therefore examined an tRNA-like element upstream of the U4 gene, analogous to that described for U2 snRNA (10). Indeed, bona fide A and B boxes were found (5′-TCCGGAGTGGG-3′ and 5′-GTTCTAGTGGC-3′, respectively), both of which agree with the consensus sequences (underlined). In comparing the consensus sequences with the corresponding boxes in the U2 locus (10), the A box of U2 is located at position –104 whereas the U4 A box is located at position –107. The U2 A box sequence (5′-TGCC-
CCGGGTGT-3') deviates more from the consensus sequence than the U4 A box. The U4 box A contains all the highly conserved positions (underlined), whereas the U2 A box contains two deviations at positions 8 and 12 with G and T instead of A and G. The B box in the U2 locus (5'-GGTTCGAGCCT-3') is located at position 2155 and that of U4 is located at position 2153; both the U2 and U4 B boxes agree well with the consensus. Thus, the tRNA-like boxes of U2 and U4 are positioned similarly with respect to the snRNA coding regions.

The U4 gene organization is unique when compared with other U snRNAs (Fig. 2), because it not only contains a tRNA-like element like the T. brucei U2 locus, but also carries two additional upstream tRNA genes, similar to U3, 7SL RNA, U6, and U5 snRNA gene loci. However, the companion tRNA located 95–98 bp upstream of the snRNAs is always divergently transcribed, whereas the tRNA gene adjacent to the tRNA-like sequence is transcribed from the same strand as the U4 gene. The presence of tRNA-like and tRNA sequences suggests that the former may have originated from a bona fide tRNA gene that had accumulated major changes and lost its identity. This tRNA may have been part of a tRNA cluster because most tRNA genes in trypanosomatids are clustered. In the case of 7SL, U5, U6, and U3, the identity of tRNA gene was kept, whereas in the case of U2 and U4, tRNA gene mutations accumulated but the identity of the A and B boxes was preserved including their proper spacing (Fig. 3A).

Because the tRNAs upstream of small RNA genes are transcribed (16, 18), it was of interest to examine whether the tRNA-like gene is also transcribed. Primer extension analysis was performed using an oligonucleotide (26264), and a discrete extension product was obtained, suggesting the presence of tRNA-like RNA transcript (Fig. 3B). Comparison of the tRNA-like structure to the tRNAPro secondary structure is presented in Fig. 3A. Deviations of the tRNA-like structure from the canonical are mainly in the D stem and Tc stem-loops. The acceptor stem is extended at the 5'-end and the anti-codon loop is composed of 5 instead of 7 nt. However, the RNA can still be folded into the cloverleaf structure, and the A and B boxes are located in their conserved positions.

To examine the genomic organization of the U4 gene, L. collosoma DNA was subjected to Southern blot analysis. A single hybridization band was observed after digesting the DNA using restriction enzymes with 4-bp and 6-bp recognition sites, suggesting that the U4 gene, like all other U snRNA genes in trypanosomatids (2), is a single copy gene (Fig. 4A).
Northern blot analysis indicates a single transcript of 114 nt (Fig. 4B).

The U4 RNA and the Relatedness to Its Homologues—The U4 coding sequence is highly conserved among trypanosomatid species: 87% identity with T. brucei (2) and 89% with Leishmania ternaolae (GenBank EBI Data Bank Accession no. X97621) U4 RNA (Fig. 5). The RNA is shorter (114 nt) compared with metazoans (142–145 nt) (20), higher plants (150–154 nt) (21), and budding yeast (160 nt) (22). The predicted secondary structure of U4 snRNA contains four stem-loop structures (I–IV) similar to higher eukaryotic U4 snRNAs. Unlike other unicellular organisms such as Physarum polycephalum, Saccharomyces cerevisiae, and Schizosaccharomyces pombe, the trypanosomatid U4 snRNAs possess a full stem-loop (IV) at the 3′-end (Fig. 5) (21). The trypanosomatid U4 snRNAs possess a full stem-loop (IV) at the 3′-end (Fig. 5) (21). The potential for base pair interactions between L. collosoma U4 and U6 is presented in Fig. 6. The first interacting domain (Stem II) is composed of 17 perfect base pairs as in T. brucei (2). This stem is more stable than in yeast, where 11 perfect base pairs are disrupted by a bulged C and followed by a perfect 5 base-paired duplex (23). Stem I is composed of a 9-bp duplex disrupted by two bulged nt in both L. collosoma and T. brucei.

The position of the bulged A in U4 snRNA is conserved in both trypanosomatid species, whereas the position of the bulged nt in the U6 duplex is different (the bulged A in T. brucei is located between positions 4 and 5 of the duplex, whereas the bulged U in L. collosoma is located between positions 6 and 7, Fig. 6). In yeast, stem I is composed of 8 perfect base-paired nt (23). Interestingly, the single sequence difference (U at position 57) between T. brucei and L. collosoma U4 RNAs is compensated for by the A at position 51, which is bulged in the T. brucei U4-U6 duplex.

The majority of changes between the trypanosomatid U4 RNAs are in the central domain and central loop (Fig. 6). The central domain and 3′-stem-loop in the yeast U4 RNA is quite tolerant to mutation (24). Interestingly, however, when the T. brucei stem-loop structure, which is homologous to the 3′-stem-loop of the yeast RNA was used to replace this homologous domain in yeast, the chimeric RNA could not complement a null allele of the yeast U4 snRNA (25). This chimeric U4 RNA did not associate well with the U6 snRNA, suggesting that in addition to base pair interactions in stem I and II, other interactions are required to promote U4/U6 complex formation. It was found that only a single mutation in the 3′-stem-loop in yeast resulted in a cold-sensitive phenotype (24). This suggests that the overall structure of the 3′-stem-loop, rather than its particular sequence, may be important for the U4 function.
some Sm site can be derived: AAAN4G (where in 71% cases N is a U). Based on the strong deviation of these sites from the consensus Sm-like sequence and the finding that trypanosome proteins are not recognized by anti-Sm sera, it was proposed that trypanosome core proteins and their corresponding binding site may differ significantly from those in mammals. Affinity selection using antisense biotinylated oligonucleotides to \( T. brucei \) snRNP and SL RNP RNAs suggests that core proteins are shared among all these particles (31). Moreover, antibodies raised against these proteins immunoprecipitate all the spliceosomal RNAs including U6 and SL RNA (32). The recent finding of an SmE homologue in trypanosomes, which carries both Sm motifs 1 and 2, suggests that Sm proteins do exist in trypanosomes (7). Yet, the lack of recognition of these proteins by anti-Sm sera is intriguing. Changes in both the Sm-like site and the core proteins may have co-evolved. These deviations may stem from the need of these snRNAs to interact with different capping enzymes, because SL RNA has a unique cap structure and U5 unlike U2 and U4 does not possess a trimethylguanosine cap (5), whereas the U6 most probably possesses an inverted cap, like in all other eukaryotes (2).

**Extragenic Sequence Elements Controlling the Expression of the U4 Gene**—To investigate the elements that control expression of the U4 gene, the coding region was tagged in the internal loop III by inserting 6 bp between position 82 and 83 (Figs. 5 and 7). The marked gene carrying 210 bp of upstream sequence carrying the tRNA-like coding region and 91 bp of downstream sequence was cloned into the pBS expression vector (Fig. 7A). RNA prepared from transfected cell lines was analyzed by primer extension using oligonucleotide (250000) complementary to the 3'-end of the U4 snRNA. The results indicate that the tagged gene was expressed efficiently (Fig. 7B, lane 2), suggesting that the tRNA-like element, but not the tRNA\(^{Pro}\) and tRNA\(^{Glu}\), is important for the U4 expression. To further explore elements essential for the expression, we mutated the A and B boxes separately and established cell lines. Analysis of U4 expression (Fig. 7B) demonstrated that mutation of the A box had no effect, whereas mutation of the B box completely inactivated the gene. To control for gene dosage (plasmid copy number), the level of the neo mRNA encoded by the pBS vector was examined by primer extension and found to be identical in all cell lines (Fig. 7B, bottom). This pattern of dependence on the tRNA-like extragenic A and B box elements is identical to that obtained for the U2 gene, except that mutation in the \( T. brucei \) box A also had a slight effect on expression (10). The \( T. brucei \) U2 B box was identified as a binding site for a nuclear protein (33). The protein might be involved in chromatin organization. Indeed, a B box located downstream to the yeast U6 coding region was shown to affect chromatin organization (34).

**Mapping 2'-O-Methyl Modifications on the L. collosoma U4, U5, and U6 snRNAs**—An important characteristic of U snRNAs is their specific modification in regions highly conserved in evolution. In the vertebrate U4 and U6, modifications are clustered in the 5'-terminal region of U4 and the central region of U6 (35). Modifications are less numerous in yeast (36). We therefore examined the modifications of snRNAs from trypanosomes, which diverged much earlier than yeast from the eu- karyotic lineage. The position of the 2'-O-methyl groups was determined by primer extension in the presence of low dNTP concentration (14), which causes the reverse transcriptase to pause one nt before the methylated site (13). The presence of specific stops in low (0.02 mM) but not high (5 mM) dNTP concentrations (Fig. 8) suggests 2'-O-methylation at positions A63 and A72 of U4. A63 position is conserved in vertebrates and is the nearest nt to the potential stem I of the U4/U6 pairing site.
duplex. A72 is in the vicinity of U4/U6 stem III (see Fig. 6), but this position is not modified in vertebrates (35). In U6 snRNA 2′-O-methyls can be assigned to positions G42, G47, U50, C53, and U56, located in stem I and stem II of the U4/U6 duplex (Figs. 6 and 8). The modified nt of trypanosome U6 RNA agrees well with their positions in vertebrate U6 (35), where they appear in regions involved in intramolecular base pairing and are expected to be at or near the spliceosome catalytic center. Interestingly, U5 is modified (Fig. 8) in position A19 of the invariant loop, which is the one position that deviates from the canonical U5 invariant loop sequence (5). We have suggested that this position interacts with the SL RNA +2 position by base pairing (4, 5). Another U5 modification appears in position A28, which seems to be specific to trypanosome, because no modified nt were found in other eukaryotes downstream of the invariant loop.

Recent study has shown that 2′-O-methylation of vertebrate U6 snRNA is guided by box C+D snoRNAs (37), suggesting that modification may take place in nucleolar or coiled bodies that have a close relationship with the nucleoli. Indeed the modification of U6 position U47 was shown to be guided by a guide RNA that also guides the methylation of 28 S rRNA. So far, no snoRNAs that have potential to guide modification of trypanosome U4 or U6 snRNAs have been discovered, although modification may take place in nucleolar or coiled bodies (Fig. 8) of RNA (Grosjean, H., and Benne, R., eds) pp. 201–227, ASM Press, Wash-

Acknowledgments—We thank Joan A. Steitz for her interest in the study and critical comments on the manuscript. We acknowledge Maile Ray for performing inverse PCR to initially identify and characterize L. collosoma U4 snRNA gene and Albrecht Bindereif for communicating unpublished results.

REFERENCES

*Downloaded from http://www.jbc.org/ by guest on December 25, 2017*
The trans-Spliceosomal U4 RNA from the Monogenetic Trypanosomatid Leptomonas collosoma: CLONING AND IDENTIFICATION OF A TRANSCRIBED tRNA-LIKE ELEMENT THAT CONTROLS ITS EXPRESSION

Liu Li, Leo R. Otake, Yu-xin Xu and Shulamit Michaeli

doi: 10.1074/jbc.275.4.2259

Access the most updated version of this article at http://www.jbc.org/content/275/4/2259

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 35 references, 16 of which can be accessed free at http://www.jbc.org/content/275/4/2259.full.html#ref-list-1