Influence of 5'-flanking sequence on 4.5SI RNA gene transcription by RNA polymerase III

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>gen-2017-0223.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Note</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>12-Dec-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Gogolevskaya, Irina; Institut molekularnoj biologii imeni V A Engel'gardta RAN</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?:</td>
<td>N/A</td>
</tr>
<tr>
<td>Keyword:</td>
<td>non-coding RNA, RNA polymerase III, transcription, promoter, TATA box</td>
</tr>
</tbody>
</table>
Influence of 5'-flanking sequence on 4.5SI RNA gene transcription by RNA polymerase III

Irina K. Gogolevskaya, Danil V. Stasenko, Karina A. Tatosyan, and Dmitri A. Kramerov*

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov st., Moscow, 119991, Russian Federation

*To whom correspondence should be addressed. Tel: +7 499 135 9997; E-mail: kramerov@eimb.ru

Abstract

Short nuclear 4.5SI RNA can be found in three related rodent families. Its function remains unknown. The genes of 4.5SI RNA contain an internal promoter of RNA polymerase III composed of the boxes A and B. Here, the effect of the sequence immediately upstream of the mouse 4.5SI RNA on its transcription was studied. The gene with deletions and substitutions in the 5'-flanking sequence was used to transfect HeLa cells and its transcriptional activity was evaluated from the cellular level of 4.5SI RNA. Single-nucleotide substitutions in the region adjacent to the transcription start site (positions -2 to -8) decreased the expression activity of the gene down to 40-60% of the control. The substitution of the conserved pentanucleotide AGAAT (positions -14 to -18) could either decrease (43-56%) or increase (134%) the gene expression. A TATA-like box (TACATGA) was found at positions -24 to -30 of the 4.5SI RNA gene. Its replacement with a polylinker fragment of the vector did not decrease the transcription level, while its replacement with a GC-rich sequence almost completely (down to 2-5%) suppressed the transcription of the 4.5SI RNA gene. The effect of plasmid sequences bordering the gene on its transcription by RNA polymerase III is discussed.

Keywords: non-coding RNA, RNA polymerase III, transcription, promoter, TATA box, rodents.
Introduction

RNA polymerase III transcribes genes encoding a variety of small non-coding RNAs in eukaryotic cells (Dieci et al. 2007). Such RNAs are critical for transcription of certain protein-coding genes (7SK RNA), splicing (U6 RNA), translation (tRNA, 5S rRNA, 7SL RNA, BC1 RNA, and BC200 RNA), initiation of DNA replication (Y RNA), and macromolecular assembly and/or transport (vault RNA). RNA polymerase III also transcribes Short Interspersed Elements (SINEs), highly numerous copies of which (up to $10^6$) are dispersed throughout the genomes of most multicellular eukaryotes (Kramerov and Vassetzky 2011). SINE transcripts can participate in the formation of new genomic copies of these nonautonomous mobile genetic elements and play a key role in the cell response to stress such as heat shock (Allen et al. 2004; Zovoilis et al. 2016).

Several types of RNA polymerase III promoters are known. The type 1 promoter is typical only for 5S rRNA genes and SINEs originating from this RNA species (Gogolevsky et al. 2009; Orioli et al. 2012; Schramm and Hernandez 2002). It is an internal promoter, i.e., lies within the transcribed DNA sequence. The promoter starts at position +45 and spans over 50 bp; it includes the box A, intermediate element, and box C at short distances from each other. Intermediate element and box C interact with transcriptional factor IIIA (TFIIIA) which recruits another factor, TFIIIC. The majority of tRNA and SINE genes as well as adenovirus genes encoding the small VA$_I$ and VA$_II$ RNAs have the type 2 promoter (Dieci et al. 2007; Kramerov and Vassetzky 2011). This promoter is also internal and includes two 11-nucleotide boxes (A and B). The box A starts at position +10...+12 relative to the transcription start site. The boxes A and B are spaced by 30-40 bp and bind to TFIIIB protein which recruits TFIIIC; the latter in turn recruits RNA polymerase III. The type 3 promoter was described in mammalian genes encoding U6, 7SK, and Y RNAs. It includes three external elements: (i) the distal sequence element (DSE) with SPH and octamer motifs which bind to Staf and Oct1 proteins, respectively, (ii) the proximal sequence element (PSE) at positions -49 to -65 interacting with SNAPc factor, and (iii) the TATA box roughly at position -25 attaching to TATA-binding protein (Dieci et al. 2007; Schramm and Hernandez 2002). Some researchers also recognize the type 4 promoter combining both internal and external elements (Englert et al. 2004). The internal elements are boxes A and B as in the type 2 promoter, while the external ones correspond to PSE and TATA box of the type 3 promoter. This promoter can be found in genes encoding 7SL, BC1, BC200, G22 and vault RNAs as well as in the EBER1 and EBER2 genes of Epstein-Barr virus (Englert et al. 2004; Howe and Shu 1993; Kickhoefer et al. 2003; Ludwig et al. 2005; Martignetti and Brosius 1995).

4.5S RNA (or 4.5S RNA$_I$) synthesized by RNA polymerase III was among the first discovered and sequenced small noncoding RNAs (Ro-Choi et al. 1972); however, its function remains unknown (Gogolevskaya and Kramerov 2002). 4.5S RNA is 98-99 nt long, it is largely localized in the nucleus, and the number of RNA molecules in the cell reaches 1-2 × 10$^6$ (Tatosyan et al. 2017). 4.5S is found only...
among species that belong to three related rodent families: Muridae (mice, rats, and gerbils), Cricetidae (hamsters and voles), and Spalacidae (blind mole-rats, bamboo rats, and zokors) (Gogolevskaya and Kramerov 2002; Gogolevskaya et al. 2010). Presumably, 4.5SI RNA genes emerged in the common ancestor of these rodents about 25 million years ago. The mouse genome contains three 4.5SI RNA genes on chromosome 6 at a distance of 40 kb from each other (Gogolevskaya and Kramerov 2010). 4.5SI RNA genes contain typical boxes A and B of the type 2 RNA polymerase III promoter. Deletion of the sequence upstream of the mouse 4.5SI RNA gene almost completely suppresses its transcription (Gogolevskaya and Kramerov 2010). These data clearly indicate that the 5'-flanking sequences are critical for the transcription of 4.5SI RNA genes. Apparently, the promoter of these genes should be assigned to the type 4.

Here, we studied the effect of the DNA region immediately upstream of one of mouse 4.5SI RNA genes on its transcriptional activity. Nucleotide deletions and substitutions were introduced into the 5'-flanking sequence (positions -1 to -35) and the obtained plasmids were used to transfect HeLa cells where the relative level of 4.5SI RNA was measured. This allowed us to identify a region (positions -24 to -30) crucial for the transcription of the 4.5SI RNA gene. It corresponds by the position but not by the nucleotide sequence to the TATA box of genes with type 4 promoters of RNA polymerase III.

Materials and methods

The constructs containing the 4.5SI RNA gene (Mmu1’) with deletions or substitutions in the 5'-flanking sequence were generated by PCR. The amplified DNA fragments were isolated using electrophoresis in 2% agarose gel and cloned into pGEM-T (Promega). A transcriptional dependence on the orientation of genes transcribed by RNA polymerase III in this vector has been reported (Koval and Kramerov 2009). Hence, plasmids with the same orientation of DNA insertions were selected (the 5′ end of 4.5SI RNA gene faced the site complementary to the M13 Reverse primer in pGEM-T). A NucleoBond PC 100 kit (Macherey-Nagel) was used to isolate plasmids suitable for transfection experiments. The plasmid with the 4.5SI RNA gene (4 µg) was mixed with 1 µg plasmid containing SOR SINE (Borodulina and Kramerov 2001; Gogolevskaya and Kramerov 2010), 500 µl DMEM, and 10 µl TurboFect reagent (Thermo Scientific). Following 20 min incubation, the mixture was added to HeLa cells growing in 60 mm Petri dish and total cellular RNA was isolated in 20 h. RNA (15 µg) was fractionated by electrophoresis in 6% polyacrylamide gel with 6 M urea and transferred onto a Hybond XL membrane (GE Healthcare). The 4.5SI RNA and SOR transcripts were detected by hybridization with 32P-labeled probes obtained by PCR (Gogolevskaya and Kramerov 2002, 2010). Following hybridization, the membrane was washed in 0.1× SSC and 0.1% SDS at 42°C, exposed to an X-ray film, and the radioactivity of RNA bands was measured by a Cyclone phosphorimager. The hybridization signal of the SOR transcript was used to normalize the 4.5SI
RNA level. The relative transcription activity of 4.5SI RNA gene was determined from three independent transfection experiments.

DNA bendability and predicted curvature of 45 bp upstream regions of 4.5SI RNA genes were determined using a model of sequence-dependent DNA bending propensity of trinucleotides, based on the probability of DNase I cutting (Brukner et al. 1995; Vlahovickev et al. 2003).

Results and Discussion

A comparison of the nucleotide sequences immediately upstream of the 4.5SI RNA genes in mouse (Mmu1’, Mmu1, and Mmu2) and rat (Rno1, Rno2, and Rno3) has demonstrated similarities at certain positions (Fig. 1). Conserved nucleotides were found at positions -3 (G in 5 out of 6 genes), -5 (T in 5 out of 6 genes), -7 (C in 6 genes), -14 to -18 (pentanucleotide AGAAT), and -24 to -34 (T(t/a)gcTAC(t/a)T(c/g)A; the nucleotides identical in all 6 genes are shown in uppercase). First, we have generated the constructs of the mouse 4.5SI RNA gene (Mmu1’) with the 19-bp 5’-flanking sequence containing nucleotide substitutions at positions -3, -5, and -7 either individually (constructs 2, 3, and 4) or all together (construct 5). The resulting DNA was introduced to HeLa cells and the level of synthesized 4.5SI RNA was determined. (Human cells contain no 4.5SI RNA gene, which allows monitoring of the gene construct transcription). These nucleotide substitutions decreased the transcription of the 4.5SI RNA gene to 38-69% of that observed for the intact gene (construct 1) (Fig. 1 and Fig. 2A). Unexpectedly, the substitution of non-conserved nucleotides in the same regions (positions -4, -6, and -8; construct 6) also decreased the transcription (56% of control). Apparently, the nucleotide sequence immediately preceding the 4.5SI RNA gene contributes to the transcription efficiency although to a small extent.

Statistical analysis of nucleotide sequences of tRNA genes from the genomes of different species allowed Giuliodori et al. (2003) to find hexanucleotides AGAAAA (mouse) or AAGAAA (human) upstream of the genes at positions -13 to -18. No such hexanucleotides could be identified in the 5’-flanking sequences of tRNA genes in other studied species (Saccharomyces cerevisiae, Schizosaccharomyces pombe, Neurospora crassa, Arabidopsis thaliana, Caenorhabditis elegans, and Drosophila melanogaster). The conserved pentanucleotide AGAAT identified by us at positions -14 to -18 of the 4.5SI RNA genes is clearly similar to the AGAAAA and AAGAAA sequences of the mouse and human tRNA genes. This region of the 5’-flanking sequences of tRNA genes has not been studied experimentally. We have obtained four constructs with replaced conserved pentanucleotide AGAAT at positions -14 to -18. A decreased transcription of the 4.5SI RNA gene down to 43-56% of control was observed for three constructs (7, 8, and 9), while the construct 10 with the pentanucleotide TTTTT instead of AGAAT demonstrated increased (134%) gene expression (Fig. 1 and Fig. 2B). Conceivably, this increase is due to the absence of G and C residues in this motif. Otherwise, this effect can be attributed to the changes in DNA bendability. Giuliodori et al. (2003)
observed alternation of nearby DNA segments with opposite bending propensities in the 5’-flanking regions of yeast tRNA genes. Using Bend.it server we calculated DNA bendability of 45 bp upstream region of mouse and rat 4.5SI RNA genes. It was found that the local minimum of DNA bendability corresponds to the AGAAT pentanucleotide (supplementary data, Fig. S1 A-D). We suggested that this feature favors efficient transcription of the gene. The AGAAT to TTTTT substitution (construct 10) further decreases the DNA bendability (supplementary data, Fig. S1 E), which can increase the gene transcription. Conversely, the AGAAT to TCATG replacement (construct 9) substantially increases the DNA bendability (supplementary data, Fig. S1 F) and decreases the transcription down to 45% of control. Anyway, the nucleotide sequence in this region (positions -14 to -18) exerts a decent effect on the gene transcription by RNA polymerase III.

The subsequent constructs had partially or completely removed murine 19-nucleotide 5’-flanking sequence present in the control construct 1. The removal of 7 bp (construct 11), 15 bp (construct 12), or 19 bp (construct 13) of the 5’-flanking DNA sequence decreased the 4.5SI RNA gene expression to 5%, 26% and 5% of control, respectively (Fig. 1 and Fig. 2C). This result was surprising. The removal of 7 bp including the AGAAT pentanucleotide induced a more significant decrease in the gene expression (9:11 times) than the replacement of the pentanucleotide (constructs 7, 8, and 9) or all these 7 nt (construct 14). It was also conspicuous that the removal of 7 bp in the 5’-flanking sequence decreased the gene expression more than the removal of 15 bp. Apparently, these results cannot be solely attributed to transcriptional significance of the DNA sequence immediately upstream of the gene, i.e., from position -19 to the transcription start site. We assumed that the sequences more distant from the transcription start site were critical for the transcription.

As was mentioned in the Introduction, a TATA box is present roughly at position -25 of genes with type 3 or 4 promoters of RNA polymerase III. In some cases (e.g., 7SL RNA gene (Engler et al. 2004)), the sequence of the box bears little resemblance to the canonical TATA box (TATAAAA); in this case, it is referred to as a TATA:like box. The studied 4.5SI RNA gene has a candidate TATA:like box TACATGA at positions -24 to -30. The construct 1 contains the plasmid sequence CACTAGT, which seems to successfully substitute the murine TATA-like box. The removal of 7 bp from the 19-bp 5’-flanking sequence (construct 11) places the plasmid sequence GCGGCCG lacking A and T residues to the position of the presumable TATA-like box, which we propose to underlie the 20-fold decrease in the gene expression (Fig. 1). The removal of 15 bp (construct 12) places a more AT-rich sequence ACCTGCA to the same position, which can maintain a higher gene expression relative to the construct 11 (Fig. 1 and Fig. 2C). Finally, the construct 13 lacking the entire murine 5’-flanking sequence has a less AT-rich plasmid sequence (GTCGACC), which decreases the gene expression again (Fig. 1 and Fig. 2C).

Thus, the data obtained suggested that the replacement of the TATA-like box with a GC-rich sequence decreases the transcription of the 4.5SI RNA gene. In order to confirm this suggestion, a construct containing 35 bp of the murine 5’-flanking sequence was obtained (construct 15) and the TATA-like box
was replaced in it with a GC-rich sequence GGCGGCC (construct 16). Transfection of these constructs into HeLa cells demonstrated that this replacement decreased the transcription level 47-fold (Fig. 1 and Fig. 2C). For comparison, the constructs containing the AGAAT pentanucleotide (positions -14 to -18) replaced with GCGGC or the TGCCTCCT sequence (positions -5 to -12) replaced with GGCGGCGC were obtained. The former (construct 17) decreased the transcription 1.3-fold; and the latter (construct 18), 1.8-fold (Fig. 1 and Fig. 2C). Thus, although these replacements slightly decreased the transcription of the 4.5SI RNA gene, their effect was only a fraction of that observed after the replacement of the TATA-like box with a GC-rich sequence (construct 16). Apparently, this box is the most important part of the RNA polymerase III promoter apart from the internal boxes A and B.

Overall, the sequence of the presumable TATA-like box can vary widely. We believe that this underlies the observed high transcriptional level of the construct with the 19-bp 5’-flanking sequence (construct 1) introducing a suitable plasmid sequence to the place of the TATA-like box. Further shortening of the 5’-flanking sequence sharply decreased the transcription not due to the deletion per se but rather to the placement of a GC-rich sequence to the position of the TATA-like box. This finding demonstrates that replacement of a functional DNA region with a different sequence not necessarily disturbs the function. In our case, the presence of a GC-rich sequence in the plasmid polylinker made it possible to identify a TATA-like box upstream of the 4.5SI RNA. The data obtained can help in the interpretation of some published data. For instance, the transcriptional assay of the small RNA G22 allowed Ludwig et al. (2005) to demonstrate that the constructs with the 22-bp 5’-flanking sequence lacking the TATA box could be efficiently transcribed by RNA polymerase III, while the removal of the entire 5’-flanking sequence almost completely suppressed the transcription. They concluded that the “upstream sequences, especially those between the transcriptional start site and -22 upstream, were important for basal transcriptional activity.” We believe that the efficient transcription of the construct with the 22-bp 5’-flanking sequence of the G22 gene could be due to a suitable sequence of the plasmid polylinker in the place of the TATA box. The removal of these 22 bp could suppress transcription due to the placement of a GC-rich region of the plasmid polylinker to the TATA box position. Indeed, such GC-rich region is present in the polylinker of the used plasmid pCRII-Topo. Thus we believe that in certain cases, it is important to consider not only what DNA sequences are being removed from the gene constructs, but also which sequences in vector are adjacent to the gene in question.

Data presented here indicate that random sequences placed instead of TATA-like box can effectively promote 4.5SI RNA gene transcription by RNA polymerase III. However, GC-rich sequences in the position of TATA-like box prevent transcription. Hence, it would be an important task to identify nucleotides crucial for TATA-like box function.
Acknowledgments

This work was supported by the Russian Foundation for Basic Research (project nos. 14-04-00616 and 17-04-00686) and the Program of the Russian Academy of Sciences for Molecular and Cell Biology.

References


**Figure captions**

**Fig. 1.** Nucleotide sequences immediately upstream of the 4.5SI RNA gene in the used constructs. The alignment of 5'-flanking sequences of the mouse (Mmu) and rat (Rno) 4.5SI RNA genes as well as the first 14 nt of the transcript is given below. Conserved nucleotides are shown against a black background. The nucleotide sequences upstream of the Mmu1’ 4.5SI RNA gene in 18 constructs are given above. Lowercase characters indicate nucleotides of the pGEM-T polylinker, while uppercase ones correspond to the sequence flanking the mouse 4.5SI RNA gene as well as nucleotide substitutions (bold and underlined) therein. The TATA-like box is framed. Relative transcription of 4.5SI RNA in HeLa cells is given to the right of each construct. The expression of the construct 1 containing the 19-bp 5'-flanking murine sequence was set as 100%. Each mean ± SD is based on three transfection experiments.
**Fig. 2.** Northern hybridization of RNA isolated from HeLa cells transfected by constructs with 4.5SI RNA gene and SOR SINE. Results of three experiments (A, B, and C). Construct numbers are given above the lanes (see Fig. 1 for construct details). Construct 1 was used in all experiments as the standard. Cellular RNA immobilized on the membrane was sequentially hybridized with 4.5SI- and SOR-specific probes. For quantitative analysis (see Fig. 1), the 4.5SI RNA signal was normalized to the SOR transcript signal. The lengths of 4.5SI and SOR RNAs are 98 and 170 nt, respectively.

**Fig. S1.** Curvature propensity and bendability plot of 5’-flanking DNA sequences of 4.5S1 RNA genes. Predicted curvature and bendability are shown in red and green, respectively. Positions of nucleotides with respect to the transcription start site are shown on the X-axis. A and B, mouse genes Mmu1’ and Mmu2, respectively. C and D, rat genes Rno1 and Rno3, respectively. E and F, gene Mmu1’ with the conserved pentanucleotide AGAAT replaced with TTTTT and TCATG, respectively. Note the minimum of bendability between positions -14 and -18; it corresponds to the conserved pentanucleotide AGAAT.

The bend.it server (http://pongor.itk.ppke.hu/dna/bend_it.html#/bendit_intro) was used to predict the curvature propensity and bendability (Vlahovicék, K., Kaján and Pongor, S. (2003): "DNA analysis servers: plot.it, bend.it, model.it and IS" Nucleic Acids Res., 31(13), 3686-7). Window size was 6.
Fig. 1. Nucleotide sequences immediately upstream of the 4.5SI RNA gene in the used constructs. The alignment of 5′-flanking sequences of the mouse (Mmu) and rat (Rno) 4.5SI RNA genes as well as the first 14 nt of the transcript is given below. Conserved nucleotides are shown against a black background. The nucleotide sequences upstream of the Mmu1′ 4.5SI RNA gene in 18 constructs are given above. Lowercase characters indicate nucleotides of the pGEM-T polylinker, while uppercase ones correspond to the sequence flanking the mouse 4.5SI RNA gene as well as nucleotide substitutions (bold and underlined) therein. The TATA-like box is framed. Relative transcription of 4.5SI RNA in HeLa cells is given to the right of each construct. The expression of the construct 1 containing the 19-bp 5′-flanking murine sequence was set as 100%. Each mean ± SD is based on three transfection experiments.
Fig. 2. Northern hybridization of RNA isolated from HeLa cells transfected by constructs with 4.5SI RNA gene and SOR SINE. Results of three experiments (A, B, and C). Construct numbers are given above the lanes (see Fig. 1 for construct details). Construct 1 was used in all experiments as the standard. Cellular RNA immobilized on the membrane was sequentially hybridized with 4.5SI- and SOR-specific probes. For quantitative analysis (see Fig. 1), the 4.5SI RNA signal was normalized to the SOR transcript signal. The lengths of 4.5SI and SOR RNAs are 98 and 170 nt, respectively.