The EMBO Journal vol.9 no.6 pp.1915 - 1921, 1990

Transcription of a nematode *trans*-spliced leader RNA requires internal elements for both initiation and 3′ end-formation

Gregory J.Hannon, Patricia A.Maroney, Deborah G.Ayers, Joseph D.Shambaugh and Timothy W.Nilsen[1]

Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, Cleveland, OH 44106, USA

[1]To whom all correspondence should be addressed

Communicated by P.A.Sharp

We have used block substitution mutagenesis and *in vitro* transcription to define sequence elements important for efficient initiation and 3′ end-formation of the *trans*-spliced leader RNA (SL RNA) of the parasitic nematode *Ascaris lumbricoides*. These experiments indicate that the SL RNA has an unusual promoter structure containing elements which include the 22 nt *trans*-spliced leader exon itself. Efficient transcription is correlated with the binding of a factor to the 22 nt (SL) sequence; mutations within the SL which abolish transcription lead to a loss in binding of this factor. In addition to internal sequences, synthesis of SL RNA *in vitro* requires an element centered 50 bases upstream of the cap site. Mutations within this element dramatically affect the level of SL RNA synthesis but do not affect accuracy of initiation. Finally, all of the information required for accurate 3′ end-formation of SL RNA lies within the transcribed region. Thus, the arrangement of sequences necessary for the synthesis of SL RNAs does not resemble that of sequences important for the synthesis of vertebrate U snRNAs despite the similarities between SL RNAs and U snRNAs.

Key words: Ascaris lumbricoides/RNA polymerase II/spliced leader RNA/transcription

Introduction

A fraction of messenger RNAs in nematodes contains a common 22 nt spliced leader (SL) sequence at their 5′ ends. The SL is acquired from a small non-polyadenylated spliced leader RNA (SL RNA) *via* a *trans*-splicing reaction (for review see Blumenthal and Thomas, 1988; Nilsen, 1989). SL RNAs and the genes encoding them have been characterized in some detail in the free-living nematode *Caenorhabditis elegans* and in the parasitic nematodes *Brugia malayi* and *Ascaris lumbricoides* (Krause and Hirsh, 1987; Takacs *et al.*, 1988; Nilsen *et al.*, 1989). This analysis has revealed that nematode SL RNAs bear several striking similarities to the vertebrate U snRNAs essential for *cis* splicing. In particular, SL RNAs are synthesized by RNA polymerase II, possess trimethylguanosine cap structures and contain functional binding sites for proteins with Sm antigenic determinants (Van Doren and Hirsh, 1988; Bruzik *et al.*, 1988; Thomas *et al.*, 1988; Nilsen *et al.*, 1989). Furthermore, based upon sequence comparison, it has been suggested that SL RNA genes have sequence elements in their flanking regions resembling those known to be essential for transcriptional initiation and 3′ end-formation of vertebrate snRNAs (Bruzik *et al.*, 1988; Nilsen *et al.*, 1989).

We recently described a cell-free system derived from *A.lumbricoides* embryos which catalyzed accurate and efficient synthesis of the *A.lumbricoides* SL RNA (Maroney *et al.*, 1990). We have used this cell free system and SL RNA genes altered by site-directed mutagenesis to characterize SL RNA transcriptional control elements. We find that the SL RNA gene has a highly unusual promoter structure which contains elements both upstream of and within the transcribed region. Mutations within the 22 nt SL sequence itself abolish *in vitro* transcription. DNase I footprint analysis indicates that a potential transcription factor(s) binds to the SL sequence; mutations within the SL which reduce transcription result in a loss of footprinting activity.

We also show by mutational analysis that elements directing accurate 3′ end-formation lie within the transcribed region. We conclude that SL RNA synthesis is mediated by elements quite distinct from those used in the synthesis of vertebrate U snRNAs.

Results

Ascaris lumbricoides SL RNA genes are reiterated in tandem within the 5S rRNA gene cluster such that each tandem repeat of 1 kb contains a 5S rRNA gene and an SL RNA gene, with both genes expressed in the same transcriptional orientation (Nilsen *et al.*, 1989). We have previously shown that a 336 bp subclone containing an SL gene, 165 bases of 5′ flanking sequence and 63 bases of 3′ flanking sequence was transcribed *in vitro* at the same efficiency as an SL RNA gene present on an intact 1 kb repeat fragment (Maroney *et al.*, 1990). Neither the efficiency nor the accuracy of *in vitro* transcription was affected when this construct was further shortened to include only 99 bp of 5′ flanking sequence or 32 bases of 3′ flanking sequence (data not shown). To define potential transcriptional control elements, we used directed mutagenesis techniques to introduce block replacements throughout this region of 5′ flanking sequence, the SL RNA coding sequence and the first 32 bases of 3′ flanking sequence. Replacements were designed such that each nucleotide in the altered region was changed. Mutagenesis was carried out on plasmids containing a functional 5S rRNA gene since this made it possible to control for any variation in amount or quality of plasmid DNA by normalizing the level of SL RNA synthesis to 5S rRNA synthesis.

Mutational analysis of the 5′ flanking region

As shown in Figure 1, most block substitutions within the 5′ flanking region of the SL RNA, including the alteration of the 10 bp immediately adjacent to the cap site, had little or no effect on the efficiency of *in vitro* transcription. Other block substitutions, including regions −24 to −34 and −35

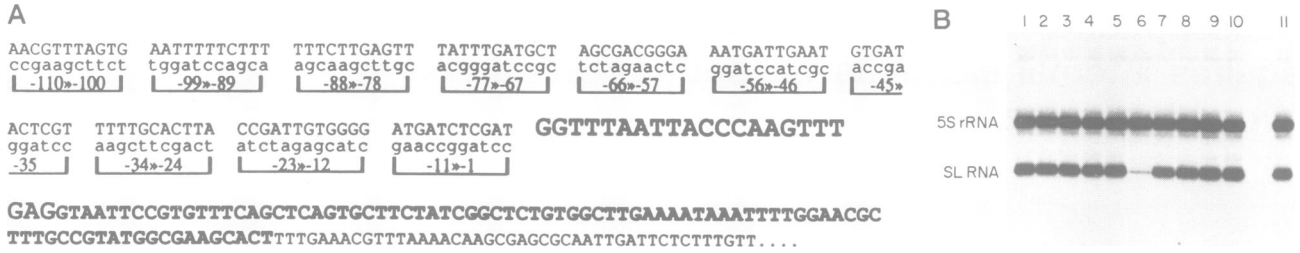

Fig. 1. Efficient synthesis of SL RNA requires a sequence element centered 50 nt upstream from the cap site. SL RNA synthesis directed by a wild-type template (**lane 11**) or templates altered by site-directed block mutagenesis (bases −110 to −100, **lane 1**; −99 to −89, **lane 2**; −88 to −78, **lane 3**; −77 to −67, **lane 4**; −66 to −57, **lane 5**; −56 to −46, **lane 6**; −45 to −35, **lane 7**; −34 to −24, **lane 8**; −23 to −12, **lane 9**; −11 to −1, **lane 10**) was assayed in embryo extract (**B**). The nucleotide sequence of a portion of the wild-type template is shown in (**A**) in upper case letters. The nucleotide sequence of individual mutants is shown in lower case letters. The SL RNA coding sequence is designated by bold letters with the 22 nt SL sequence in larger type. Transcription reactions were assembled and analyzed as described in Materials and methods.

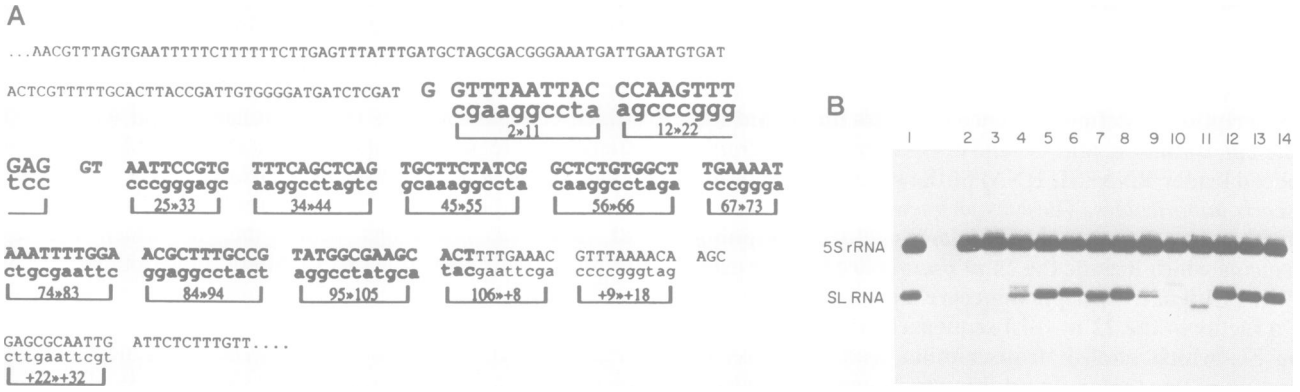

Fig. 2. The 22 nt SL sequence as well as additional internal elements are required for efficient and accurate SL RNA synthesis. SL RNA synthesis directed by a wild-type template (**lane 1**) or templates altered by site-directed block mutagenesis (bases 2−11, **lane 2**; 12−22, **lane 3**; 25−33, **lane 4**; 34−44, **lane 5**; 45−55, **lane 6**; 56−66, **lane 7**; 67−73, **lane 8**; 74−83, **lane 9**; 84−94, **lane 10**; 95−105, **lane 11**; 106 to +8, **lane 12**; +9 to +18, **lane 13**; +22 to +32, **lane 14**) was assayed in embryo extract (**B**). The partial nucleotide sequence of wild-type and mutant templates is shown in (**A**). Transcription reactions were assembled and analyzed as described in Materials and methods.

to −45, resulted in only a slight decrease in transcriptional efficiency. The most dramatic effect was observed when the region from −46 to −56 was altered (Figure 1, lane 6). In this case a >10-fold reduction in efficiency of transcription was observed. No alterations within the 5′ flanking region appeared to effect the accuracy of transcription initiation.

Mutational analysis within the SL RNA coding region

The first block substitution analyzed within the coding region encompassed the first 11 bases of the 22 nt SL sequence. Almost no SL RNA was synthesized from this altered template (data not shown). Because it seemed possible that this effect might be due to alteration of the nucleotide normally used for initiation, an additional substitution encompassing bases 2−11 of the SL sequence was tested. A marked reduction in accumulation of SL RNA was also observed with this template and with a template in which bases 12−22 were altered (Figure 2, lanes 2 and 3). When templates containing sequential 7−11 base alterations throughout the rest of the SL RNA coding sequence were tested, it was apparent that internal elements in addition to the SL sequence itself were required for accurate and efficient SL RNA synthesis (Figure 2). Alterations of bases 25−33 or 74−83 led to reduced accumulation of SL RNA and the appearance of products of aberrant size (Figure 2, lanes 4 and 9). Similar but less dramatic effects were seen upon alteration of bases 56−66 (Figure 2, lane 7). Substitutions encompassing bases 84−94 and 95−105 led to greatly reduced accumulation of SL RNA transcripts, and no transcripts of correct size were observed with these mutant templates (Figure 2, lanes 10 and 11). Substitution of bases 34−44, 45−55 or 67−73 had little or no effect on SL RNA synthesis (Figure 2, lanes 5, 6 and 8). Similarly, alteration of the three 3′ terminal nucleotides and the first eight bases of 3′ flanking sequence did not effect either accuracy or efficiency of SL RNA synthesis (Figure 2, lane 12).

Since all of the mutations lying within the transcribed region would result in the synthesis of altered transcripts, we considered the possibility that the differences we observed in accumulation of *in vitro* transcripts could result from differential stability of mutant SL RNAs in the extract. To control for this possibility, a pulse-chase experiment was performed. Transcription reactions identical to those shown in Figure 2 were incubated for one hour at which point α-amanitin was added to 1 μg/ml along with a 10-fold excess of unlabeled UTP. Half of each reaction was analyzed immediately while the other half was incubated for an additional hour. By comparing the levels of each SL RNA at the two time points, it was possible to monitor any decay of reaction products synthesized during the first hour of incubation. This analysis indicated that neither the wild type SL RNA nor any of the mutant SL RNAs significantly decayed during the second hour of incubation (data not shown).

Both Hernandez and Weiner (1986) and Neuman de Vegvar *et al.* (1986) have previously shown that 3′ end-

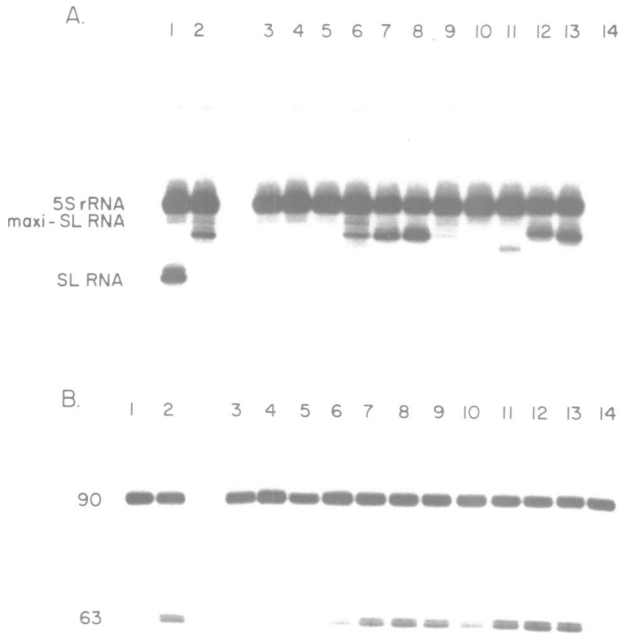

Fig. 3. Primer extension analysis of wild-type and mutant SL RNAs synthesized in embryo extract. SL RNA synthesis from wild-type templates without (**lanes 1**) or with a net 12 base insertion between bases 39 and 43 (maxi-SL) (**lanes 2**) and maxi-templates containing block substitutions (−56 to −46, **lanes 3**; 2−11, **lanes 4**; 12−22, **lanes 5**; 25−33, **lanes 6**; 56−66, **lanes 7**; 67−73, **lanes 8**; 74−83, **lanes 9**; 84−94, **lanes 10**; 95−105, **lanes 11**; 106 to +8, **lanes 12**; +9 to +18, **lanes 13**) was assayed in embryo extract; **lanes 14** did not contain added template. Aliquots of reaction products were either analyzed directly on denaturing gels (**A**) or subjected to primer extension analysis (**B**) as described in Materials and methods. The size of primer extension products was determined with reference to a sequencing ladder electrophoresed in parallel lanes. The nucleotide sequence of substitutions is shown in Figures 1 and 2. Transcription reactions were assembled and processed as described in Materials and methods.

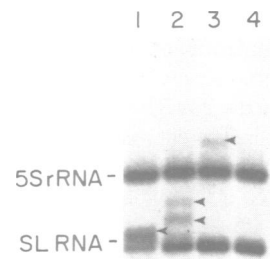

Fig. 4. SL RNA 3′ end-formation does not require 3′ flanking sequence elements. SL RNA synthesis directed by linear transcription templates as described in the text containing 1 (**lane 1**), 12 (**lane 2**), 26 (**lane 3**) or 100 (**lane 4**) bases of 3′ flanking sequence was assayed in embryo extract. The template used in lane 1 was prepared by digesting the 106 to +8 mutant template with *Eco*RI (see Figure 2); the template used in lane 2 was prepared by digestion of the +9 to +18 mutant template with *Eco*RI and *Sma*I (see Figure 2); the template used in lane 3 was prepared by digestion of the +22 to +32 mutant template with *Eco*RI (see Figure 2); and the template used in lane 4 was prepared by digestion of a wild-type template with *Eco*RI and *Hin*dIII. In each case, appropriate restriction fragments were gel-isolated prior to use as transcription templates. In lanes 1−3, transcripts larger than mature SL RNA are evident and are indicated by arrows. The size of these RNAs suggests that they arose by polymerase running off the template. Transcription reactions were assembled and analyzed as described in Materials and methods.

formation of U snRNAs is functionally coupled to initiation from an snRNA promoter. Thus, the failure to observe accumulation of transcripts with any particular mutant could have been due to a defect in initiation and/or 3′ end-formation. To distinguish between these possibilities and to provide information relevant to the nature of the aberrantly sized transcripts observed with templates containing substitutions at 25−33, 74−83, 84−94 and 95−105, we synthesized a marked SL RNA gene (see Materials and methods) which would be suitable for primer extension analysis. The marked gene contained a net 12 base insertion between bases 39 and 43 of the SL RNA coding sequence. This position of insertion was chosen since mutations in this region did not appear to affect the efficiency or accuracy of SL RNA synthesis (see Figure 2, lane 5). After determining that the marked gene was transcribed *in vitro* with comparable efficiency to the wild-type gene (Figure 3A, lanes 1 and 2), the same block substitutions as those described above were introduced into the marked template. These substitutions had the same quantitative and qualitative effects on SL RNA synthesis from the marked gene as they did on SL RNA synthesis from the wild-type template (Figure 3A). In addition to direct analysis, aliquots of transcription reactions containing marked templates were analyzed by primer extension using an oligodeoxynucleotide primer complementary to bases 44−63 of the marked genes. Accurately initiated SL RNA transcripts synthesized from marked templates would thus yield a primer extension product of 63 nucleotides. When the marked gene was transcribed in extract and reaction products subjected to primer extension analysis, the expected 63 nucleotide extension product was observed (Figure 3B, lane 2) as well as an extension product of 90 nucleotides. The 90 nucleotide extension product was observed even in the absence of added marked template (Figure 3B, lanes 1 and 14) and results from the fortuitous hybridization of the primer to an unidentified RNA present in these extracts. Figure 3(B) shows primer extension analysis of transcription products produced from marked templates containing block substitutions. Only three block substitutions (−46 to −56, 2−11 and 12−22) caused a substantial reduction in initiation of SL RNA transcripts (Figure 3B, lanes 3−5). Reactions containing templates with substitutions at positions 25−33 and 84−94 yielded only slightly reduced levels of primer extension product (Figure 3B, lanes 6 and 10), even though accumulation of transcripts synthesized from the 84−94 template appeared to be greatly reduced as judged by the direct analysis of transcription products (see Figure 3A, lane 10). Surprisingly, wild-type levels of primer extension product were observed when transcripts synthesized from the 74−83 and 95−105 templates were analyzed (Figure 3B, lanes 9 and 11). These templates yielded essentially no SL RNA transcripts of the correct size (Figure 3A, lanes 9 and 11). In no case did we obtain evidence for altered sites of initiation of SL RNA transcription. These results indicated that the element located at −46 to −56 and the SL sequence itself were both required

for efficient initiation of SL RNA synthesis. Furthermore, the lowered accumulation of aberrantly sized transcripts synthesized from templates with substitutions encompassing much of the rest of the SL RNA coding sequence was not due to inaccurate or inefficient initiation but, rather, to altered specificity and efficiency of 3′ end-formation. In this regard, the apparent lowered accumulation of SL RNA transcripts observed with the 25−33, 74−83, 84−94 and 95−105 templates (Figure 3A, lanes 6 and 9−11) probably results from the dissociation of RNA polymerase from the template at nonspecific sites throughout the 3′ flanking region. To explore the determinants of 3′ end-formation further, we analyzed the effects of mutations within the 3′ flanking region.

Mutational analysis within the 3′ flanking region

As described above, accurate 3′ end-formation of SL RNAs synthesized *in vitro* was observed with only 32 nt of 3′ flanking sequence. We previously noted that the sequence GTTTAAAACAAGC lies nine nucleotides downstream of the SL RNA 3′ terminus (Nilsen *et al.*, 1989). This sequence contains a 12/13 match to the vertebrate U snRNA 3′ end-formation box ($GTTTN_{0-3}AAAG/ANNAGA$) known to be essential for accurate 3′ end-formation of U snRNAs (Hernandez, 1985; Neuman de Vegvar *et al.*, 1986; You *et al.*, 1985). It seemed quite likely that this sequence element would contribute to 3′ end-formation of the SL RNA. To test this possibility, block substitutions were introduced which obliterated the putative 3′ end-formation signal. Surprisingly, neither this mutation nor any other mutation within the 3′ flanking region altered either the position or efficiency of 3′ end-formation (Figure 2, lanes 13 and 14). To exclude the unlikely possibility that a vector sequence could in some way substitute for an element normally present in the 3′ flanking region, we used isolated restriction fragments containing wild-type or mutant SL RNA genes as transcription templates. These fragments, which contained a functional 5S rRNA gene and the SL RNA coding sequence followed by varying amounts of normal or altered 3′ flanking sequence, were transcribed efficiently in the extract (Figure 4). Accurate 3′ end-formation was observed with as little as one base of mutant 3′ flanking sequence (Figure 4, lane 1), although efficiency of 3′ end-formation with this template was ~50% of that observed for SL RNAs synthesized from a template containing 12 bases of 3′ flanking sequence (Figure 4, lane 2). Further increases in efficiency of 3′ end-formation were observed with templates containing 26 or 100 bases of additional 3′ flanking sequence (Figure 4, lanes 3 and 4). This experiment indicated that all of the sequence information necessary to specify accurate 3′ end-formation of SL RNA lay within the SL RNA coding sequence itself and that more than 12 bases of 3′ flanking sequence were required for efficient 3′ end-formation (Figure 2, lanes 13 and 14 and data not shown). This requirement does not appear to involve specific sequence recognition, since every base in this region can be changed without affecting efficiency or accuracy of 3′ end-formation and probably reflects a minimal size necessary to accommodate 3′ end-formation components.

DNase I footprint analysis

The foregoing experiments indicated that an element located between −46 and −56, the 22 nt SL sequence and other regions within the SL RNA coding sequence were important transcriptional control elements for SL RNA synthesis. To determine if any of these sequences were potential binding sites for *trans*-acting factors we performed DNase I footprint analysis of end-labeled SL RNA genes. When these genes were incubated with the whole cell extract, a region of protection which comprised the 22 nt SL sequence itself and 4−5 bases both 5′ and 3′ of this sequence was observed (Figure 5). Equivalent protection was observed when either the coding or non-coding strands were analyzed (Figure 5A and B). To explore the specificity of this interaction, we also analyzed the footprints produced on genes containing substitutions within the SL region. These substitutions (2−11 and 12−22) resulted in the loss of a detectable footprint (Figure 5 and data not shown). In addition, a number of competition experiments were performed. Competition for the SL binding factor was observed with genes containing an intact 22 nt sequence and with double-stranded oligodeoxynucleotides corresponding to the 22 nt sequence, but no competition was observed when genes with altered 22 nt sequences or oligodeoxynucleotides of unrelated sequence were used (data not shown). Furthermore, the ability to generate a footprint was not affected by block mutations lying outside of the 22 nt region (Figure 5). In similar competition experiments, we determined that a plasmid containing several copies of the SL sequence effectively inhibited SL RNA synthesis at concentrations which prevented footprinting, whereas neither SL RNA synthesis nor footprinting was affected by inclusion of a plasmid lacking these sequences (Figure 6).

To date, we have not obtained convincing evidence for factors which might interact with the −46 to −56 region or the other regions within the coding sequence, although we have not attempted to extensively vary footprinting conditions, nor have we tried potentially more sensitive techniques, such as gel shift assays, which might reveal the association of factors with these sites.

To explore the possibility that the factor or factors associated with the SL sequence might have both DNA and RNA binding activities analogous to *Xenopus* TFIIIA, we attempted to compete for the SL footprinting activity with synthetic RNAs exactly corresponding to the SL RNA. Even at levels 10-fold above those necessary for effective competition with DNA fragments or double-stranded oligodeoxynucleotides, we observed no competition with RNA. We concluded from these experiments that a potential transcription factor specifically recognizes the 22 nt SL sequence. Association of this factor is correlated with the transcription of SL RNA genes *in vitro*.

Discussion

We have shown that the SL RNA of *A.lumbricoides* has a complex promoter structure containing internal elements. All of the sequences needed for efficient and accurate transcription and 3′ end-formation *in vitro* are found within the first 56 bases of 5′ flanking sequence and the SL RNA coding sequence; there is no requirement for 3′ flanking sequence. It should be noted that the elements we have identified are necessary for *in vitro* transcription and may not reflect all of the sequences required for maximally efficient *in vivo* expression. Since the effects of enhancers or enhancer-like sequences are often not observed in *in vitro* systems, a

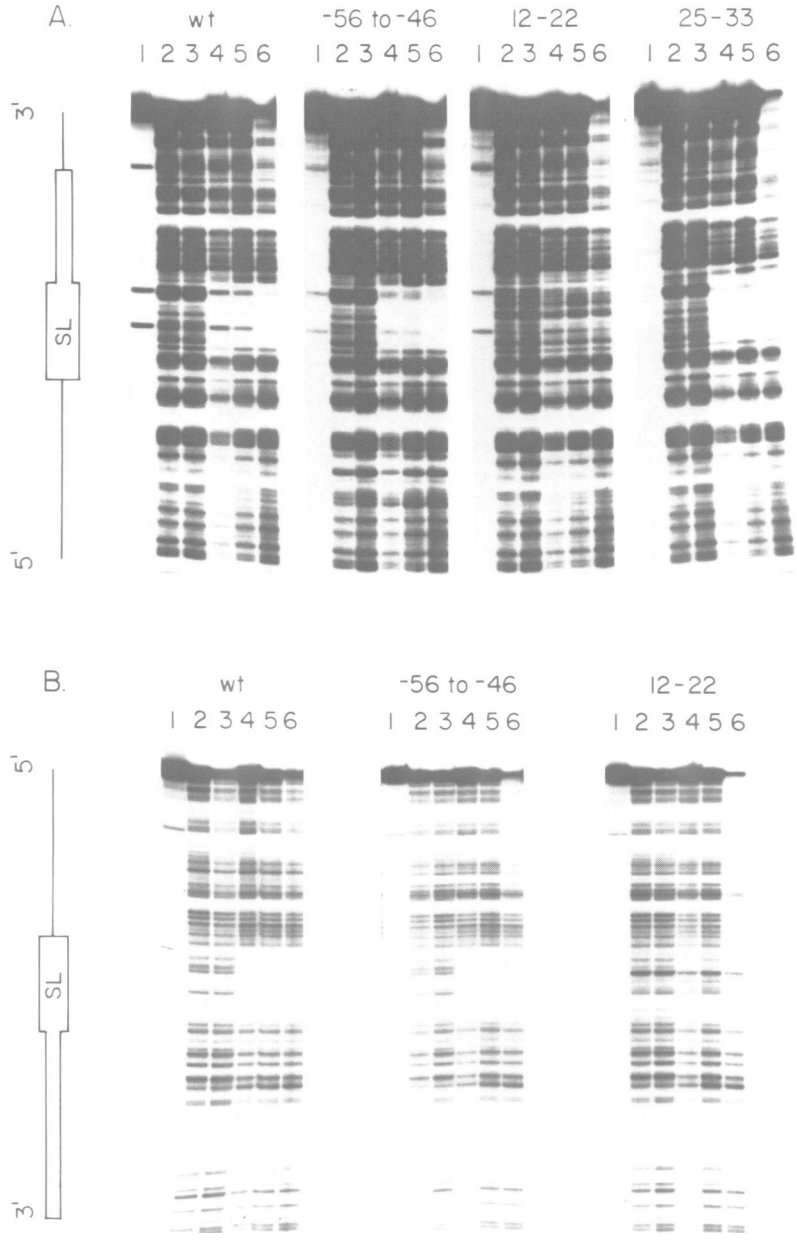

Fig. 5. A potential transcription factor specifically binds to the 22 nt SL sequence. Restriction fragments containing wild type or mutant SL RNA genes (as indicated) were labeled on the coding (**A**) or non-coding (**B**) strands and subjected to DNase I footprint analysis as described in Materials and methods. A schematic representation of the SL RNA gene (boxed) and flanking regions (lines) is shown to the left of each panel of footprints. Positioning of the schematic representation relative to the autoradiogram was determined with reference to a sequencing ladder electrophoresed in parallel lanes. The nucleotide sequence of substituted regions is shown in Figures 1 and 2. Labeled fragments were incubated without (**lanes 1–3**) or with (**lanes 4–6**) 3 μl of whole cell extract. After incubation, reactions were digested with DNase I at 0 (**lanes 1**), 0.3 (**lanes 2**), 0.6 (**lanes 3**), 5 (**lanes 4**), 10 (**lanes 5**) or 15 μg/ml (**lanes 6**) before being analyzed by gel electrophoresis as described in Materials and methods. The labeled bands which appear in the protected regions in footprints designated wt and −56 to −46 in A do not reflect cleavage by DNase I but result from nicks in the labeled fragment used for these analyses; compare lanes 1 with lanes 5 and 6.

complete understanding of SL RNA gene transcriptional control signals awaits the development of appropriate *in vivo* assay systems.

The element centered at −50 and the 22 nt SL element together appear to be necessary and sufficient for efficient initiation of SL RNA synthesis; alteration of either sequence leads to a dramatic reduction in accumulation of SL RNA transcripts (see Figures 1, 2 and 3). While the 22 nt sequence has been perfectly conserved among widely divergent nematodes (reviewed in Nilsen, 1989), similar conservation is not observed in the element centered at −50. In particular, there is no sequence identity in this region between the *A.lumbricoides* SL RNA gene and the *C.elegans* SL RNA gene. This may explain why the *C.elegans* gene is transcribed at a very low efficiency in the *Ascaris* extract (Maroney *et al.*, 1990). While the position of this upstream element is similar to that of the proximal sequence elements (PSE) of U snRNA promoters, it may not serve a similar role. Several lines of evidence suggest that PSE elements are functionally analogous to the TATA elements of protein coding genes in specifying the sites of initiation of U snRNA synthesis (see Dahlberg and Lund, 1988; Parry *et al.*, 1989

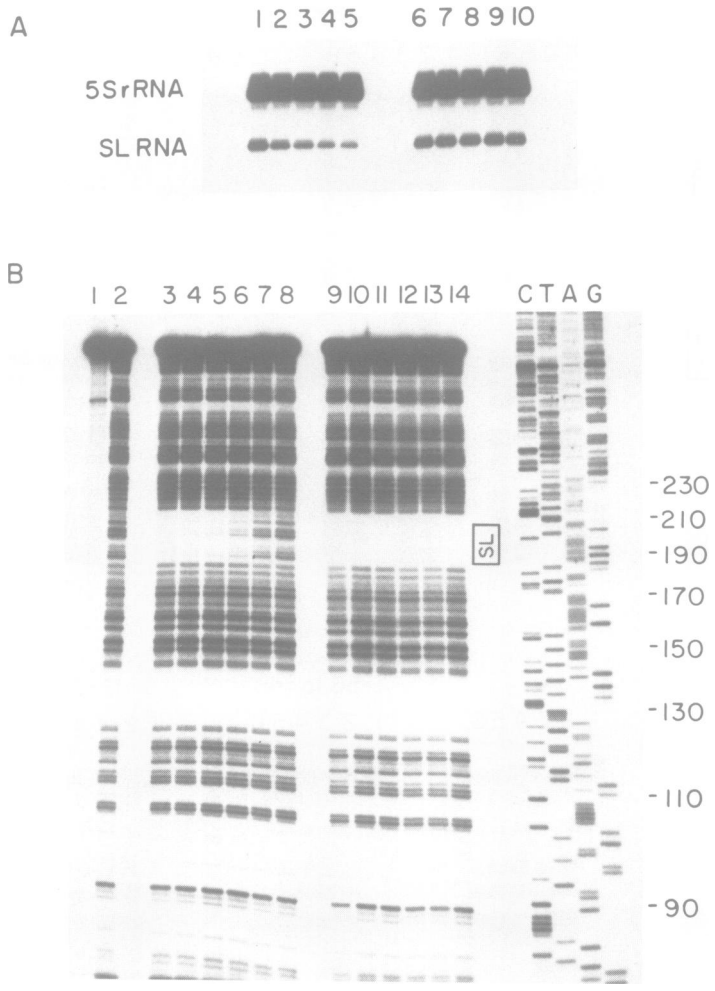

Fig. 6. Synthesis of SL RNA *in vitro* is correlated with the binding of a potential transcription factor to the 22 nt SL sequence. Aliquots of *A.lumbricoides* embryo extract were preincubated for 15 min on ice with varying amounts of a plasmid containing four tandem copies of the 22 nt SL sequence or a control plasmid lacking these sequences, before being used for transcription (**panel A**) or footprinting (**panel B**) reactions. Panel A shows transcription reactions containing wild-type template (see Figure 1, lane 11) and either specific (**lanes 1–5**) or nonspecific (**lanes 6–10**) competitor. Each reaction contained 7.5 μl of embryo extract which had been preincubated with 0 (lanes 1 and 6), 6.6 (lanes 2 and 7), 13.3 (lanes 3 and 8), 33.3 (lanes 4 and 9) or 66.6 ng (lanes 5 and 10) of competitor DNA per μl of extract. Transcription reactions were assembled and analyzed as described in Materials and methods. Panel B shows footprinting reactions using the same labeled restriction fragment used in Figure 5 (panel B, wt). Footprinting reactions contained 3 μl of embryo extract which had been preincubated with 0 (**lanes 3** and **9**), 3.3 (**lanes 4** and **10**), 6.6 (**lanes 5** and **11**), 13.3 (**lanes 6** and **12**), 33.3 (**lanes 7** and **13**) or 66.6 ng (**lanes 8** and **14**) per μl of extract of specific (lanes 3–8) or nonspecific (lanes 9–14) competitor DNA. **Lane 1** shows a reaction where the fragment was incubated with extract but not digested with DNase while in **lane 2** the fragment was digested with 0.5 μg/ml DNase I in the absence of extract. For size reference, a dideoxy sequencing ladder of a fragment of known sequence electrophoresed in parallel is shown. The position of the 22 nt SL sequence is schematically depicted adjacent to the protected region. Footprinting reactions were assembled and analyzed as described in Materials and methods.

for review). Alterations in the −50 element of the SL RNA gene do not lead to the use of aberrant initiation sites (see Figures 1 and 3). We currently do not know which element(s) direct accurate initiation of SL RNA synthesis, since no mutation we have tested affected the position of transcription initiation (see Figures 1–3). It seems possible that either the upstream element or the 22 nt sequence can independently fix the site of initiation.

The mechanism of 3′ end-formation of SL RNA transcripts is also quite different from that used to generate the 3′ ends of U snRNAs. For U snRNAs, sequences designated 3′ end-formation signals located ~15 bases downstream of the mature 3′ terminus are necessary for 3′ end-formation (Hernandez, 1985; You *et al.*, 1985; Ciliberto *et al.*, 1986) and, with the exception of the *Xenopus laevis* U1B snRNA (Ciliberto *et al.*, 1986), sequences within snRNA genes appear not to be required (reviewed in Dahlberg and Lund, 1988). By contrast, several regions within the SL RNA coding sequence appear to be required for accurate 3′ end-formation and no flanking sequence requirement is seen.

We do not at present know whether the promoter and 3′ end-formation elements involved in SL RNA synthesis are peculiar to this gene or reflect an unusual mode of snRNA synthesis in nematodes. We have recently isolated genomic fragments containing *A.lumbricoides* U1 and U2 snRNAs and both genes are efficiently transcribed in the embryo extract (unpublished observations). It will be of interest to determine whether their promoter structures resemble those of vertebrate snRNAs or are more similar to the SL RNA promoter described here.

The presence of internal elements in an RNA polymerase II promoter is not unique. Transcription of both the *Drosophila* transposable element 'Jockey' and a yeast Ty element involves the utilization of internal control elements

(Mizrokhi *et al.*, 1988; Yu and Elder, 1989). Furthermore, transcription of certain mouse ribosomal protein genes requires a gene internal sequence (Moura-Neto *et al.*, 1989; Narayanan *et al.*, 1989), and the minimal promoter or 'initiator' region of a terminal transferase gene contains the first few bases of the mature transcript (Smale and Baltimore, 1989). All of these genes are apparently transcribed by the 'normal' RNA polymerase II which transcribes protein coding genes. In the case of genes transcribed by 'modified' pol II (snRNAs), Mattaj and colleagues have shown that a short gene internal sequence appears to be necessary for stable transcription complex formation on the *Xenopus* U2 snRNA gene *in vivo*. However, with the exception of the cap site nucleotide itself, this sequence does not affect either accuracy or efficiency of transcription (Tebb and Mattaj, 1988).

Finally, as noted above, the 22 nt SL sequence has been perfectly conserved over wide evolutionary distance in nematodes. This conservation has been interpreted as suggesting that the SL sequence itself, present on *trans*-spliced mRNAs, must have some important post-transcriptional function. The observations reported here suggest the interesting possibility that the extreme sequence conservation of nematode SL sequences could instead result from constraints imposed by the binding specificity of a transcription factor(s).

Materials and methods

Extracts and in vitro transcription

Whole cell extracts were prepared from 32 cell *A.lumbricoides* embryos as described (Maroney *et al.*, 1990). *In vitro* transcription reactions in 25 μl contained 15 μl of extract, 3.0 mM $MgCl_2$, 60 mM KCl, 2 mM DTT, 0.5 mM ATP, CTP and GTP, 50 μM UTP, 10 μCi [α-^{32}P]UTP, 20 mM creatine phosphate, 12 mM Tris, pH 7.9, and 500 ng template as indicated in the figure legends. Following a 2 h incubation at 30°C, reactions were diluted to 250 μl in 250 mM NaOAc, 20 mM Tris, pH 7.5, 1 mM EDTA and 0.25% SDS, digested with 100 μg/ml proteinase K for 30 min at 42°C and extracted with phenol/chloroform. Reaction products were analyzed on 8 M urea, 6% polyacrylamide gels. Following electrophoresis, labeled transcripts were visualized by autoradiography. In reactions where marked templates were used, transcription products were assayed by primer extension using the primer oligonucleotide specified in the text essentially as described (Takacs *et al.*, 1988) except that extensions were carried out at 42°C.

Transcription templates and site directed mutagenesis

An ~1 kb *ClaI* restriction fragment containing a 5S rRNA gene and an SL RNA gene (Nilsen *et al.*, 1989) was subcloned into pBSm13+ (Stratagene) and used to transform *dut*$^-$*ung*$^-$ *E.coli* (CJ 236) (Kunkel *et al.*, 1987). The plus strand of the plasmid was rescued and prepared as described (Vieira and Messing, 1987) and subjected to site-directed mutagenesis exactly as described (Kunkel *et al.*, 1987). Oligodeoxynucleotides used for mutagenesis contained 10 bases of complementary sequence on either side of the 7–12 bases altered in each construct. Appropriate mutations were identified by nucleotide sequence analysis of mini-preps of individual colonies. The nucleotide sequence of each block mutation is shown in Figures 1 and 2.

The marked *A.lumbricoides* SL RNA template was also created by site-directed mutagenesis using the oligodeoxynucleotide (5′ CGTGTTTCAGACTCTAGAGGATCCCCGTGCTTCTAT 3′) which contained two ten base blocks of complementarity flanking a 16 base block of non-complementary sequence. Since the two complementary blocks were separated by four nucleotides of SL RNA sequence, mutagenesis with this oligonucleotide created a net 12 base insertion within the SL RNA coding sequence. This marked construct was then subjected to additional site-directed mutagenesis as detailed for the unmarked gene. A phage T7 transcription template used for the synthesis of transcripts for use in competition experiments was constructed using appropriate oligonucleotides and PCR exactly as described (Maroney *et al.*, 1990).

DNase I footprint analysis

Footprinting reactions were carried out essentially as described (Carbon *et al.*, 1987). 395 base restriction fragments containing either wild type or mutant SL RNA genes as indicated in the legend to Figure 5 were end-labeled on either strand (10 fmol, 30 000 c.p.m.) and incubated with or without extract for 15 min on ice in a 25 μl reaction containing 12% glycerol, 3 mM $MgCl_2$, 60 mM KCl, 1 mM DTT and 1 μg, poly(dI·dC) (Pharmacia). DNase I was then added in the indicated amounts to each reaction. After 90 s, digestions were stopped by the addition of 25 μl of 0.6 M NaCl, 0.5% SDS and 10 mM EDTA. Following phenol/chloroform extraction and ethanol precipitation, digestion products were analyzed on 6% denaturing polyacrylamide gels. Regions of protection were determined with reference to a sequencing ladder electrophoresed in parallel.

Acknowledgements

We thank Drs D.Setzer, F.Rottman and J.A.Steitz for critically reading the manuscript. This research was supported by grants from the John D. and Catherine T.MacArthur Foundation, the public health service GM-31528 and a PYI award of the NSF (TWN).

References

Blumenthal,T. and Thomas,J. (1988) *Trends Genet.*, **4**, 305–308.
Bruzik,J.P., van Doren,K., Hirsh,D. and Steitz,J.A. (1988) *Nature*, **335**, 559–562.
Carbon,P., Murgo,S., Ebel,J., Krol,A., Tebb,G. and Mattaj,I.W. (1987) *Cell*, **51**, 71–79.
Ciliberto,G., Dathan,N., Rainer,F., Philipson,L. and Mattaj,I.W. (1986) *EMBO J.*, **5**, 2931–2937.
Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer Verlag, Berlin, pp. 38–71.
Hernandez,N. (1985) *EMBO J.*, **4**, 1827–1837.
Hernandez,N. and Weiner,A.M. (1986) *Cell*, **47**, 249–259.
Krause,M. and Hirsh,D. (1987) *Cell*, **49**, 753–761.
Kunkel,T.A., Roberts,J.D. and Zakovr,R.A. (1987) *Methods Enzymol.*, **154**, 367–383.
Maroney,P.A., Hannon,G.J. and Nilsen,T.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 709–713.
Mizrokhi,L.J., Georgieva,S.G. and Ilyin,Y.V. (1988) *Cell*, **54**, 685–691.
Moura-Neto,R., Dudov,K.P. and Perry,R.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3997–4001.
Narayanan,H., Kelley,D.E. and Perry,R.P. (1989) *Genes Dev.*, **3**, 1789–1800.
Neuman de Vegvar,H.E., Lund,E. and Dahlberg,J.E. (1986) *Cell*, **47**, 259–266.
Nilsen,T.W. (1989) *Exp. Parasitol.*, **69**, 413–416.
Nilsen,T.W., Shambaugh,J., Denker,J.A., Chubb,G., Faser,C., Putnam,L. and Bennett,K. (1989) *Mol. Cell. Biol.*, **9**, 3543–3547.
Parry,H.D., Scherly,D. and Mattaj,I.W. (1989) *Trends Biochem. Sci.*, **14**, 15–19.
Smale,S. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
Takacs,A.M., Denker,J.A., Perrine,K.G., Maroney,P.A. and Nilsen,T.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7932–7936.
Tebb,G. and Mattaj,I.W. (1988) *EMBO J.*, **7**, 3785–3792.
Thomas,J.D., Conrad,R.C. and Blumenthal,T. (1988) *Cell*, **54**, 533–539.
Van Doren,K.V. and Hirsh,D. (1988) *Nature*, **335**, 556–559.
Vieira,J. and Messing,J. (1987) *Methods Enzymol.*, **153**, 3–34.
You,C.-Y., Ares,M.,Jr and Weiner,A.M. (1985) *Cell*, **42**, 193–202.
Yu,K. and Elder,R.T. (1989) *Mol. Cell. Biol.*, **9**, 3667–3678.

Received on January 16, 1990; revised on March 6, 1990