Focus

Transcription of the Kinetoplastid Spliced Leader RNA Gene

D.A. Campbell, N.R. Sturm and M.C. Yu

In recent years, much has been learned about the cis-elements controlling transcription of the kinetoplastid spliced leader (SL) RNA gene. The SL RNA gene contains the first 39 nucleotides that are trans-spliced on to all nuclear-derived mRNAs in these organisms. Transcription initiation is determined by two precisely spaced upstream elements and transcription termination is directed by the downstream poly-T tract, although the RNA polymerase responsible for SL RNA synthesis is still questioned. In this article, David Campbell, Nancy Sturm and Michael Yu review the field of kinetoplastid SL RNA gene transcription, address past proposals in light of current data and discuss some of the differences that appear in the literature.

The spliced leader (SL) RNA gene is of central importance to kinetoplastid mRNA metabolism because its product is the common substrate for the trans-splicing reaction that provides mRNAs with their 5′-cap structures. The cap consists of an inverted 7mG and specific methylation of nucleotides 1–4 and 6 (Refs 1–3). Owing to the organization of the SL RNA genes in large tandem arrays and their transcriptional isolation by non-transcribed spacers, the SL RNA genes were prime candidates for the identification of transcription promoter and termination elements.

In kinetoplastids, transcription has been found to deviate from the mainstream paradigm in several respects: small nuclear RNA (snRNA) genes contain promoters in upstream RNA genes; the arrangement of transcription units yielding monocistronic mRNAs has been abandoned largely in favor of polycistronic organization, in which multiple protein-coding regions are transcribed from a single promoter (see also references on the description and discussion of the complete gene organization of chromosome I in Leishmania major); furthermore, there is a strong body of evidence supporting pol I transcription of some protein-coding genes (e.g. those encoding variant surface glycoprotein (VSG) and procyclin/procyclic acidic repetitive protein) in Trypanosoma brucei. Similar to other eukaryotes, mature mRNAs possess a 5′-cap and a 3′-poly-A tail; however, all kinetoplastid mRNAs possess a common leader sequence of 39–41 nucleotides (nt) that is typically uniform within each species. This common leader, or mini-exon, is transcribed from the SL RNA gene and is added post-transcriptionally via the major trans-splicing pathway. The steady-state SL RNA ranges in length from 96 nt in Leishmania to 141 nt in T. brucei and is found predominantly in non-polyadenylated RNA. Polyadenylated SL RNA has been characterized in short, stumpy forms of T. brucei and amastigotes of Leishmania donovani. Because these are slowly or non-dividing forms of the parasites, this might be a mechanism for downregulating mRNA processing or a transient step in SL RNA turnover.

Which polymerase transcribes the SL RNA gene? Since the discovery of the RNA polymerase more than 16 years ago, the identity of the RNA polymerase (RNA pol) responsible for the transcription of this gene in kinetoplastids has been debated. A recent study concluded that the SL RNA gene is transcribed by pol II, pol III or a polymerase with intermediate character. A definitive identification is increasingly important in light of the dearth of pol II promoters described in kinetoplastids; the only candidate promoters thus far are for the genes encoding actin and heat shock protein 70 (Ref. 18) in T. brucei and the GARP gene20 in T. congolense.

Most studies to determine the polymerase responsible for SL RNA gene transcription have used nuclear run-on assays with disrupted or permeabilized cells (Table 1). Studies using 1,10-phenanthroline, sarkosyl, and tagetin-toxin tend to support pol II transcription of the SL RNA gene25–27 (R.M. Saito, PhD Thesis, 1996, University of California, Los Angeles, USA). An exception in the case of 1,10-phenanthroline indicated pol III transcription of the SL RNA gene22; however, this study also concluded that pol II transcribed the VSG gene, which is now thought to be transcribed by pol I (Ref. 10). Divalent (Mn2+) and monovalent (K+) cation optima have implicated pol III in SL RNA synthesis20,22 (Table 2).

The most common inhibitor used to identify the RNA polymerase transcribing the SL RNA genes is α-amanitin, which has yielded conflicting results (Table 1). α-Amanitin is the standard inhibitor used for most eukaryotes, where pol II is highly sensitive, pol III intermediate and pol I resistant to its effects (Box 1). All researchers find that the kinetoplastid pol II is partially resistant to α-amanitin, as are several other eukaryotic pol II enzymes20–30. In the kinetoplastids, α-amanitin is useful for identifying pol I transcription units, but the minimal and inconsistent differentials in resistance for pol II and pol III essentially invalidate its use in determining the RNA polymerase that synthesizes kinetoplastid SL RNA. Because the promoters for pol II synthesis of snRNA and mRNA are not interchangeable in metazoa30, it is possible that the same will be true in kinetoplastids. If this is the case, the genes encoding actin and tubulin might not be the appropriate pol II control templates for comparison with the SL RNA genes because the mechanisms of transcription initiation are likely to differ between these and mRNA genes; α-amanitin has been reported to inhibit the initiation of RNA synthesis as well as PP, exchange to the same extent as chain elongation28. Cumulative evidence has also been cited in support of either pol II or pol III transcription of the SL RNA gene. The presence of the 5′cap on SL RNA transcripts was used to argue in favor of pol II synthesis29, whereas the identification of putative box A/B elements (Box 1) in the SL RNA transcribed region and the...
presence of a downstream T tract were used to support pol III transcription. Extensive mutagenesis of the SL RNA transcribed region performed in search of transcription and splicing elements eliminated the possibility of box A/B internal promoter elements. Recent studies have confirmed that the T tract is indeed a termination element in *Leishmania tarentolae* and *T. brucei* (G. Laufer and A. Günzl, pers. commun.). A run of more than six Ts is required for termination *in vivo* and *in vitro*; in this GC-rich surrounding, a typical eukaryotic pol III would be predicted to terminate at a run of four Ts. Therefore, the failure of the *T. brucei* SL RNA gene promoter to drive transcription through a luciferase reporter might be explained by the inability of the polymerase to traverse the polyuridine-rich splicing signal at the splice acceptor site that contains a 14-T tract.

Although the identity of the RNA polymerase has not been resolved, the SL RNA gene promoter presents a more unified picture.

The SL RNA gene promoter

The SL RNA gene promoter has been studied *in vivo* in six different kinetoplastid species, and has been assayed by transient transfection using either a chloramphenicol acetyl transferase reporter gene or a geneinternal tag; and by stable transfection of plasmids containing internally tagged SL RNA genes. In these studies, block/linker scan mutation of upstream sequences revealed either two or three promoter elements within 81 bp of the transcription start site in *L. tarentolae*, *Leishmania amazonensis*, *Leptomonas seymouri* and *T. brucei* (summarized in Fig. 1); a single element was identified in *T. cruzi*. Transcription of the SL RNA gene, like that of...
Box 1. General Characteristics of Eukaryotic Nuclear Transcription

Eukaryotes possess three nuclear RNA pols. Each polymerase is recruited to the appropriate gene template by a protein or protein complex, termed a transcription factor, which is bound to specific DNA sequences (termed promoters and initiators) located at precise sites relative to the transcription start point (nucleotide position +1).

RNA pol I transcribes only large ribosomal RNA (rRNA) genes in most eukaryotes. Its activity is resistant to the fungal toxin α-amanitin (see Table I). The promoter for RNA pol I is located between nucleotides –142 and +6. Termination of transcription by RNA pol I is determined by a specific RNA–protein interaction.

| Table I. Typical α-amanitin sensitivities for eukaryotic nuclear RNA polymerases* |
|------------------|---------------|---------------|
| Pol I | Pol II | Pol III |
| Animal cells | >1000 | 0.01–0.15 | 10–25 |
| Insects | >1000 | 0.03–0.06 | >1000 |
| Yeast | 300–600 | ~1 | >1000 |

*aμg ml⁻¹ for 50% inhibition.

RNA pol II is responsible for the synthesis of mRNA and is considered sensitive to α-amanitin. The simplest promoters that direct basal transcription by RNA pol II are the ‘TATA’ box, centered at about position –27 and the initiator element at positions –3 to +6. TBP, first identified as the TATA-binding protein, is a component of transcription factor complexes for all three classes of RNA polymerase. Termination of RNA pol II transcription is complex and might involve poly-T tracts, secondary structure in the nascent transcript and specific protein interactions.

RNA pol III transcribes genes for a variety of small RNA molecules including transfer RNA (tRNA) and 5S rRNA, and has an intermediate sensitivity to α-amanitin in animal cells. The promoters for RNA pol III are located within the transcribed region. tRNA genes possess two elements termed A and B. The A boxes are common elements that interact with the transcription factor TFIIIC. Termination of RNA pol III activity is effected by a tract of at least four T residues in the DNA.

Small nuclear RNA (snRNA) genes can be transcribed by either RNA pol II or RNA pol III. The core promoters for these genes are upstream proximal sequence element (about position –56) and an AT-rich region (about position –27); mutations in the AT-rich region might change the specificity for the RNA polymerase. The promoters that direct RNA pol II transcription of snRNA genes are not interchangeable with those for mRNA synthesis; differential recognition of termination signals in terms of processivity and recognition of termination signals.

genes encoding metazoa snRNAs are genus specific55. A comparative alignment of the wild-type SL RNA gene upstream sequence relative to L. tarentolae is shown in Fig. 2. Point mutation of the two elements in L. tarentolae revealed the core nucleotides to be –66 GACN₉G – 58 and –39 GN₉CCC – 33 (Ref. 47). These core nucleotides are conserved in the essential elements of some other kinetoplastids and are reminiscent of the conserved proximal sequence element (PSE) nucleotides of vertebrate snRNA genes. The core elements of the kinetoplastid promoter determine the transcription start site and have a spacing relative to each other that is necessary for function56. Mutation of the –11 to –2 region in T. brucei32 and a proposed trypanosome initiator element (Inrt) at –10 to –1 in L. seymouri39 yielded transcripts with altered initiation sites.

Key techniques in the understanding of transcription mechanisms are in vitro transcription using crude nuclear extracts and reconstituted transcription using purified fractions. In vitro transcription of SL RNA genes has been developed for T. brucei25, L. tarentolae22 and L. tarentolae46. Although not identical, the nucleotides necessary for transcription in vitro (underlined in Fig. 2) reflect a subset of the nucleotides required in vivo.

Putative transcription factors that bind to DNA upstream of the SL RNA gene have been identified by electrophoretic mobility shift assay (EMSA) in four of the kinetoplastids. In L. tarentolae, point mutations in the –66 to –58 region, which inactivated transcription in vivo, also prevented binding25 (shaded box in Fig. 2). By contrast, both positions –70 to –61 and –40 to –31 bound protein(s) present in L. seymouri nuclear extracts39, and whereas only a single shifting complex was identified in L. tarentolae, two complexes are apparent in Leptomonas. Using DNA footprinting, the Leptomonas EMSA complex I covers positions –74 to –55 and the larger EMSA complex II covers positions –82 to –28 (Ref. 52). Complex II formation was dependent on precise spacing of the two promoter elements25. The L. seymouri PBR-1, which corresponds to complex I, has been identified as a 122-kDa complex, consisting of 57-, 36- and 46-kDa polypeptides, with the 46-kDa polypeptide contacting the DNA directly52.
Transcription initiation signals for the kinetoplastid SL RNA genes are not related to those of the nematode SL RNA gene (Fig. 3), but more closely resemble those of vertebrate snRNA genes. The validity of this analogy will be tested further by comparison of the derived protein sequences for the kinetoplastid SL RNA gene transcription factors and the vertebrate (SNAP and PTF) and invertebrate (Dm(PBF)) counterparts. The presence of a vertebrate-like octamer motif/distal sequence element (DSE, Fig. 3) has not been reported in kinetoplastids; however, two results indicate that additional constraints (eg. an enhancer, nucleosome assembly, chromatin topology or position effects) might influence SL RNA gene transcription: first, in L. tarentolae, the normalized steady-state level of tagged SL RNA is 2–6% of wild-type SL RNA39; second, in Leptomonas collinae, the best expression of tagged SL RNA is obtained with 415 bp (relative to 135 bp) of upstream sequence (J. Goucharov and S. Michaeli, pers. commun.). A conserved sequence element (DSE; Fig. 3) has not been reported in kinetoplastids; however, two results indicate that additional constraints (eg. an enhancer, nucleosome assembly, chromatin topology or position effects) might influence SL RNA gene transcription: first, in L. tarentolae, the normalized steady-state level of tagged SL RNA is 2–6% of wild-type SL RNA39; second, in Leptomonas collinae, the best expression of tagged SL RNA is obtained with 415 bp (relative to 135 bp) of upstream sequence (J. Goucharov and S. Michaeli, pers. commun.).

Future work

It should be possible to determine definitively which polymerase is responsible for SL RNA gene transcription by manipulating pol II53 directly (either by conditional inactivation54 or by altering its drug sensitivity55) or by a combined immunohistochemical26 and immobilized template27 approach. In addition, along with others56, we are purifying the transcription factors that interact with the SL RNA gene promoter elements. This approach will elucidate the protein bridge between the DNA template and the initiating polymerase. It will also provide a perspective into the transcriptional machinery in the early branching euglenoid/kinetoplastid lineage.

Acknowledgements

We thank Vivian Berkett and F. S. (UPDQ), New Jersey Medical School, USA, Arthur Gratz (Tübingen University, Germany), Shula Michaeli (Weizmann Institute of Science, Israel) and their co-workers for critical reading of the manuscript, N.N.S.-G and M.C. were supported by the Microbial Pathogenesis Training Grant (2-T32-AI0723). The studies in our laboratory were supported by a grant from the National Institutes for Health (AI3536).

References


Fig. 2. Multiple alignment of nucleotide sequences upstream of spliced leader RNA genes in Leishmania tarentolae and other kinetoplastids. DNA footprinting assays indicate coverage of positions −74 to −55 by complex I and −82 to −28 by complex II in Leptomonas26, and multiple positions between −49 and −6 in Trypanosoma brucei29. Numbering of the nonRNA proximal sequence element (PSE) corresponds to the human U2 snRNA gene. Key: top line, complete sequence corresponding to positions −78 to −1 of the L. tarentolae gene; n, nucleotides not necessary for transcription (can be mutated without affecting transcription initiation); red underline, complex I DNase I footprint in L. tarentolae; blue underline, complex II DNA footprint in L. tarentolae; black underline (and lower case), amino acids necessary for transcription in vitro; pink shading, nucleotides that eliminate protein binding in the gel shift assay (EMSA) when mutated. −, spaces introduced to maximize the alignment; open boxes, essential nucleotides in vivo, defined by point mutation; nucleotides (upper case), defined by block mutagenesis in the indicated kinetoplastids to be essential in vivo.

Fig. 3. Comparative arrangements of transcriptional promoter and terminator elements for a kinetoplastid (Leishmania) spliced leader (SL) RNA gene, a nematode (Ascaris) SL RNA gene and vertebrate small nuclear RNA (snRNA) genes. Transcribed regions are indicated by the open boxes above the line. Below the line, promoter elements are represented by the crosshatched boxes and termination/T and formation elements by the black boxes; the shaded box indicates a non-functional termination consensus sequence in Ascaris. TBP corresponds to the site of TATA-binding protein interaction. DSE, distal sequence element; PSE, proximal sequence element.
Focus


7 Johnson, P.J. et al. (1987) Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multisite transcription unit adjacent to a VSG gene. Cell 55, 275–281


20 Geondal, E.J. et al. (1989) Characterization of the RNA polymerases of Trypanosoma brucei: trypanosomal mRNAs are composed of transcripts derived from both RNA polymerase II and III. EMBO J. 8, 3385–3399


