

Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene

Appolinaire Djikeng^a, Ludmila Ferreira^b, Maximiliano D'Angelo^c, Pavel Dolezal^d, Tracey Lamb^e, Silvane Murta^f, Veronica Triggs^g, Sebastian Ulbert^h, Alejandro Villarinoⁱ, Sarah Renzi^a, Elisabetta Ullu^{a,j}, Christian Tschudi^{a,*}

^a Department of Internal Medicine, Yale University School of Medicine, PO Box 208022, 333 Cedar Street, New Haven, CT 06520-8022, USA

^b Universidade Federal de Minas Gerais, Departamento de Bioquímica e Imunologia, Belo Horizonte 31270-901, Brazil

^c Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, Vuelta de Obligado 2490, Capital Federal, CA 1428, Argentina

^d Department of Parasitology, Faculty of Science, Charles University, Prague 128 44, Czech Republic

^e Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

^f Centro de Pesquisas "René Rachou", Belo Horizonte 1743, Brazil

^g Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

^h Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066 CX, Netherlands

ⁱ Department of Pathobiology, University of Pennsylvania, 3800 Spruce Street, Philadelphia, PA 19104-6008, USA

^j Department of Cell Biology, Yale University School of Medicine, PO Box 208022, 333 Cedar Street, New Haven, CT 06520-8022, USA

Received 11 September 2000; accepted in revised form 11 December 2000

Abstract

We have previously shown that the poly(A) polymerase (PAP) gene of *Trypanosoma brucei* is interrupted by an intervening sequence. It was postulated that removing this intron by *cis*-splicing requires a yet unidentified U1 small nuclear RNA (snRNA), which in other organisms engages in base-pair interactions across the 5' splice site during early spliceosome assembly. Here we present a characterization of a 75 nucleotide long candidate *T. brucei* U1 snRNA. Immunoprecipitation studies indicate that a trimethylguanosine cap structure is present at the 5' end and that the RNA is bound to core proteins common to spliceosomal ribonucleoprotein particles. The U1 snRNA has the potential for extensive intermolecular base pairing with the PAP 5' splice site. We used block replacement mutagenesis to identify sequences necessary for *in vivo* expression of U1 snRNA. We found that at least two *cis*-acting elements, tRNA-like A and B boxes, located in the 5'-flanking region are necessary for U1 snRNA synthesis; no internal sequences close to the transcription start site are essential, suggesting a promoter architecture distinct from other trypanosome U-snRNA genes. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: U1 small nuclear RNA; Trimethylguanosine cap; *Trypanosoma brucei*; *cis*-splicing; Intron; Promoter architecture; Common core proteins

1. Introduction

It is well established that in trypanosomes *trans*-splicing plays an integral part in the production of mature mRNAs by transferring an identical 39-nucleotide (nt)

spliced leader (SL) sequence from the donor SL RNA to the 5' end of every mRNA. After its discovery in trypanosomes [1–3], *trans*-splicing was subsequently described in nematodes [4], euglenoids [5], and certain trematodes [6]. *Trans*-splicing is mechanistically analogous to *cis*-splicing, the removal of intervening sequences [7,8], and *cis*- and *trans*-splicing share a common set of small nuclear ribonucleoprotein (snRNP) cofactors, namely U2, U4, U5, and U6 snRNPs [9–11]. However, in contrast to *cis*-splicing, *trans*-splicing proceeds without the U1 snRNP [9] and it has been proposed that in *trans*-splicing the SL RNA replaces

Abbreviations: nt – nucleotide(s); PAP – poly(A) polymerase; SL – spliced leader; snRNA – small nuclear RNA; snRNP – small nuclear ribonucleoprotein particle; TMG – trimethylguanosine.

* Corresponding author. Tel.: +1-203-7857332; fax: +1-203-7853864.

E-mail address: christian.tschudi@yale.edu (C. Tschudi).

the function normally carried out by the U1 snRNA in *cis*-splicing [12].

Ever since the discovery of *trans*-splicing in trypanosomes, the belief developed that these organisms lack the machinery to carry out *cis*-splicing. This was based on the apparent absence of intervening sequences and of the U1 snRNA, which in *cis*-splicing engages in base-pair interactions across the 5' splice site during early spliceosome assembly. Since this conviction was not substantiated by experimental data, it was not too surprising that very recently a candidate U1 snRNA was identified in two trypanosomatids, namely *Crithidia fasciculata* and *Leishmania tarentolae* [13], and intervening sequences were discovered in the *Trypanosoma brucei* and *T. cruzi* PAP gene [14]. Examination of the nucleotide sequences of the *C. fasciculata* and *L. tarentolae* U1 snRNA revealed an extensive sequence complementarity with the boundaries of the 5' splice site of the trypanosome introns across 11 nucleotides [14]. Indeed, mutations introduced in this region of the *T. brucei* intron abolished *cis*-splicing in vivo [14], thus supporting a possible interaction of the U1 snRNA with the 5' splice site.

To begin to address the role of U1 snRNA in the removal of intervening sequences in trypanosomes, we report here the identification of the *T. brucei* U1 snRNA, which is 75 nt long and highly similar in sequence to the *C. fasciculata* and *L. tarentolae* U1 snRNA. The *T. brucei* U1 snRNA possesses a trimethylguanosine (TMG) cap structure and is complexed with common core proteins. In vivo expression analysis revealed promoter elements reminiscent of the RNA polymerase III-transcribed U2 snRNA gene of *T. brucei* [15] and of the U4 snRNA gene of *Leptomonas collosoma* [16].

2. Material and methods

2.1. Plasmid constructions

Plasmid pTbU1 represents a subclone of the *T. brucei* U1 snRNA locus of 2049 bp assembled by database mining (accession nos. AZ218384, AQ651810, AQ657735, AQ943660, AQ943659, and AQ9416) and contains the U1 snRNA coding region of 75 and 375 bp of 5' and 450 bp of 3' flanking region in pBluescript II KS (Stratagene). This gene was marked by the insertion of 19 nt with the sequence 5'-TTCCATGGTATG-GCGCCAG-3' at pos. +36 relative to the transcription start site using PCR methodology and two complementary oligonucleotides. The resulting construct (pTbU1tag) was used for subsequent transfection studies. Base substitutions in the U1 snRNA gene and flanking regions were introduced by two sequential PCRs as described previously [15]. The sequences of the A and B boxes (Fig. 1A) were changed to 5'-TAGGATCC-

CAC-3' and 5'-GGATCCTCCA-3', respectively. The –50 and –70 elements (underlined in Fig. 1A) were mutated to 5'-GGATCCGAGG-3' and 5'-GGATCCCT-GCGC-3', respectively. Nucleotides 2–11 of the U1 snRNA coding region were changed to 5'-CACGGA-TCCA-3'. Constructs were verified by DNA sequencing.

2.2. RNA analysis and DNA transfection

Transfections, RNA isolation, primer extension analysis, and Northern blot hybridizations were carried out as described [15,17]. The following oligonucleotides were used in primer extension analysis: U1A, complementary to nt 40–58 of the U1 snRNA; U1tag, complementary to both the U1 and SL RNA tag [18], giving rise to a product of 55 and 72 nt in primer extension analysis for the tagged U1 and SL RNA, respectively.

2.3. Immunoprecipitations

Cell extracts were immunoprecipitated with antibodies against the *T. brucei* common proteins [19] bound to protein A-Sepharose [20] and RNA was processed for Northern blot hybridization. TMG-specific immunoprecipitations on total RNA were carried out essentially as described [21] with a mouse anti-m₃G monoclonal antibody (generously provided by Adrian Krainer, Cold Spring Harbor Laboratory, NY).

3. Results and discussion

3.1. Structure of the *Trypanosoma brucei* U1 snRNA gene

To identify a candidate U1 snRNA gene in *T. brucei*, we used the BLAST algorithm [22] to search the available databases for sequences with similarity to the *C. fasciculata* U1 snRNA gene [13]. This resulted in the identification of a genome survey sequence (accession no. AZ218384) with significant similarity to the first 30 nt of the *C. fasciculata* U1 snRNA: 28 out of 30 nt were identical (Fig. 2). Further database mining allowed us to assemble a contiguous stretch of 2049 bp encompassing the U1 snRNA gene locus, and part of this sequence is shown in Fig. 1A. Whereas there are no recognizable genes in the 1069 bp downstream of the U1 snRNA gene, we identified four tRNA genes in the 905 bp upstream of the transcription start site: tRNA^{Arg}, tRNA^{Thr}, tRNA^{Leu} and tRNA^{Gly}. One characteristic of most Kinetoplastid U-snRNA genes is the presence of a divergently transcribed tRNA gene 95–99 bp upstream of the U-snRNA gene with the tRNA gene internal A and B box elements being required for expression of both the tRNA and the accompanying U-snRNA gene [17]. Whereas the *L. tarentolae* U1

SL sequence. This implies that specific proteins binding to the U1 or the SL RNA direct each of the two RNAs to the appropriate site on the pre-mRNA, namely for the SL RNA the 3' splice site of *trans*-spliced mRNAs and for U1 snRNA the 5' splice site of *cis*-spliced mRNAs. Perhaps in trypanosomes a similar mechanism exists to insure recognition of intron boundaries by the appropriate snRNA.

3.2. Characteristics of the *Trypanosoma brucei* U1 snRNA

Previous experiments in *C. fasciculata* revealed a candidate U1 snRNA with a methylguanosine cap structure. However, the monoclonal antibody used did not differentiate between m⁷G and TMG caps [13]. To examine the structure of the *T. brucei* U1 snRNA 5' end, immunoprecipitations were carried out with anti-TMG antibodies as described in Section 2 (Fig. 3) and Northern blots were probed with a U1 snRNA-specific probe. To control for immunoprecipitations, Northern blots were also probed for the SL RNA, which carries a m⁷G cap structure. Whereas the SL RNA was not precipitated to any detectable level (Fig. 3B, lane 3), most of the U1 snRNA was present in the immunoprecipitate (Fig. 3A, lane 3). Thus, our results support the presence of a TMG cap structure at the 5' end of the *T. brucei* U1 snRNA.

The three Kinetoplastid U1 snRNAs contain a putative binding site for spliceosomal core proteins (underlined for the *T. brucei* sequence in Fig. 2A). It was therefore of interest to test whether the *T. brucei* U1 snRNA assembled with common proteins. Total cell extracts were immunoprecipitated with antibodies raised against common proteins [19] and immunoprecipitates were assayed by Northern blot analysis (Fig.

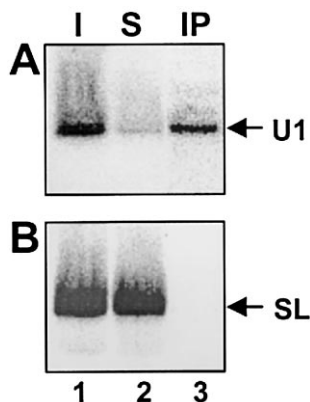


Fig. 3. Northern blot analysis of anti-TMG immunoprecipitates. RNA from input (lane 1, I), supernatant (lane 2, S) and immunoprecipitate (lane 3, IP) was separated on a 8% polyacrylamide -7 M urea gel, blotted onto nylon membrane and hybridized sequentially to probes complementary to the U1 snRNA (U1, panel A) and the SL RNA (SL, panel B).

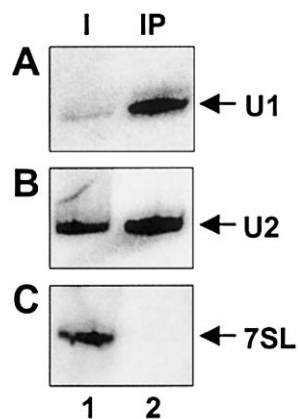


Fig. 4. Northern blot analysis of anti-common protein immunoprecipitates. Total RNA was immunoprecipitated with an antibody specific for *T. brucei* spliceosomal common proteins [19]. RNA from input (lane 1, I) and immunoprecipitates (lane 2, IP) was separated on a 6% polyacrylamide per 7 M urea gel, blotted onto nylon membrane and hybridized sequentially to probes complementary to the U1 snRNA (U1, panel A), U2 snRNA (U2, panel B) and 7SL RNA (7SL, panel C). Only 10% of the input material was loaded on this gel.

4). To control for immunoprecipitations, Northern blots were also probed with a U2 snRNA and 7SL RNA probe. As expected from previous experiments [19], the U2 snRNA was immunoprecipitated with anti-common protein antibodies (Fig. 4B, lane 2). In contrast, the 7SL RNA, which does not associate with common proteins, was not immunoprecipitated to any detectable level (Fig. 4C, lane 2). Probing with a U1 snRNA probe revealed a substantial amount of the U1 snRNA in the immunoprecipitate (Fig. 4A, lane 2), indicating that this RNA forms a core ribonucleoprotein particle.

3.3. Promoter architecture of the *Trypanosoma brucei* U1 snRNA gene

To identify DNA elements required for expression of the *T. brucei* U1 snRNA gene, transcription of different plasmid constructs was analyzed *in vivo* by transfection of insect form trypanosome cells. For this we used templates, which are marked by the insertion of a short sequence (tag) in the coding region. Following transfection, transcripts were detected by primer extension analysis.

A tagged version of the U1 snRNA gene was made by inserting a 19 nt tag at position +36 relative to the transcription start site. The resulting construct containing 375 bp of 5' flanking and 450 bp of 3' flanking sequences (pTbU1tag) was first tested by transient transfection of procyclic trypanosome cells. Total RNA was prepared 4 h post-transfection and primer extended using an oligonucleotide complementary to nt 40–58 of the U1 snRNA. In the control transfection a cDNA of

58 nt was obtained which correspond to the endogenous U1 snRNA (Fig. 5B, lane 2). Samples transfected with the tagged U1 snRNA gene (lane 3) gave two distinct primer extension products. The first of 58 nt in length was derived from endogenous U1 gene expression, whereas the second product of 77 nt had the predicted size for transcripts originating from the tagged U1 snRNA gene. To confirm this result, the primer extension analysis was repeated with an oligonucleotide complementary to the tag. As shown in Fig. 5C, lane 3), this allowed specific detection of transcripts originating from pTbU1tag by the presence of 53–55 nt long primer extension products. The fact that we observed a cluster of three major extension products, rather than a single cDNA species, might indicate that the terminal 3 nt of U1 snRNA are methylated, as described for the *C. fasciculata* U1 snRNA [13].

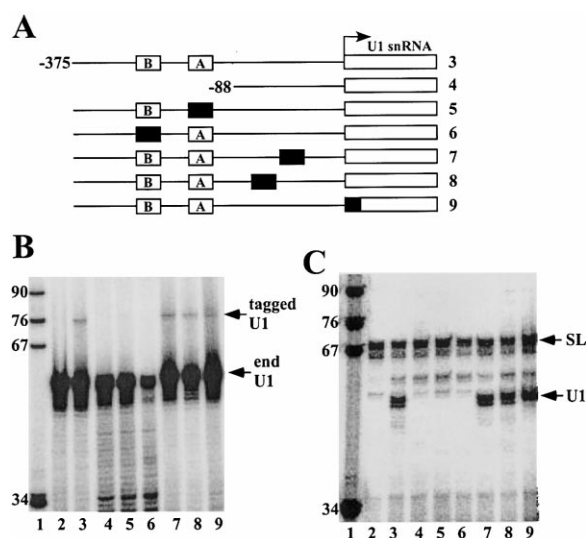


Fig. 5. Primer extension analysis of in vivo transcribed U1 snRNA. (A) Structure of the different constructs (not to scale) used for transfection. Numbers to the left refer to base pairs present upstream of the U1 snRNA transcription start site. Block substitutions are indicated by filled boxes. The number to the right of each construct corresponds to the lane number in the autoradiographs shown below. (B, C) The wild-type tagged U1 snRNA gene or constructs containing mutations as shown in (A) were cotransfected into procyclic *T. brucei* cells with a tagged SL RNA gene [18] to monitor transfection efficiency. Lane 1, MspI-digested pBR322 (sizes are indicated in bp); lane 2, vector DNA; lane 3, wild-type tagged U1 snRNA gene; lane 4, sequences upstream of position -88 were deleted; lanes 5, mutation in the A box; lane 6, mutation in the B box; lanes 7 and 8, sequences around -50 and -70 , respectively, were mutated; lane 9, nt 2–11 of the U1 coding region were substituted. (B) Total RNA preparations were analyzed by primer extension with 5' end labeled oligonucleotide U1A giving rise to two products: one derived from the endogenous U1 snRNA (end U1) and the other from the transfected tagged U1 snRNA (tagged U1). (C) Primer extension analysis was carried out with oligonucleotide U1tag which is complementary to both the U1 and SL RNA tag, giving rise to two cDNA products corresponding to transcripts from the transfected U1 snRNA (U1) and SL RNA gene (SL).

Thus, in the next set of experiments we set out to identify sequence elements essential for U1 snRNA gene expression. Visual inspection of the upstream sequences revealed a number of putative regulatory elements (Fig. 1). Similar to what we found previously for the U2 snRNA gene promoter [15], a degenerate A and B box was located at position -105 and -168 , respectively. On the other hand, an alignment of the upstream sequences with the corresponding sequences of the SL RNA gene revealed elements centered around position -50 and -70 with significant similarity to promoter elements previously localized upstream of the SL RNA transcription start site [18]. As a first test to assay for the importance of the A and B boxes, sequences upstream of -88 were deleted. With this construct the accumulation of U1 snRNA was reduced to undetectable levels (Fig. 5B and 5C, lane 4). In a second set of constructs, both boxes were targeted separately by block substitutions (lanes 5 and 6). Since both mutations did not allow detection of U1 snRNA, this clearly established that U1 snRNA gene expression requires two extragenic elements resembling the A and B box consensus sequences of the internal control regions of tRNA genes.

None of the other sequence substitutions we introduced affected transcription of the U1 snRNA gene at a detectable level. Sequences around position -50 and -70 do not appear to provide promoter function (Fig. 5B and 5C, lanes 7 and 8). Thus, we concluded that the structure of the U1 snRNA promoter most likely resembles that of the U2 snRNA gene with both intragenic and extragenic regulatory elements. However, we were surprised to find that substitution of nt 2–10 of the coding region did not affect U1 snRNA accumulation (Fig. 5B and 5C, lane 9), since the corresponding sequences in the U2 snRNA gene are an integral part of the U2 snRNA gene promoter [15].

Thus taken together, the candidate *T. brucei* U1 snRNA has many of the characteristics of a spliceosomal snRNA and its promoter architecture closely resembles that of the RNA polymerase III-transcribed U2 snRNA gene. However, there is one major distinction in that expression of the U1 snRNA gene does not require intragenic sequences close to the 5' end of the coding region. One important question that remains to be solved is whether the isolated U1 snRNA indeed participates in *cis*-splicing. Future efforts will test whether compensatory mutations in the U1 snRNA 5' end can suppress the observed effect of 5' splice site mutations in the PAP intron [14].

Acknowledgements

We thank Albrecht Bindereif for providing anti-common protein antibodies. L.F., M.D'A., P.D., T.L.,

S.M., V.T., S.U. and A.V. were participants of the 2000 Biology of Parasitism Course at the Marine Biological Laboratory in Woods Hole, MA, where part of this work was carried out. The Biology of Parasitism Course is supported by grants from the Burroughs Wellcome Fund, the Ellison Medical Foundation, and Merck. At Yale University this study received support from National Institutes of Health Grant AI28798 to E.U. E.U. is the recipient of a Burroughs Wellcome Fund Scholar Award in Molecular Parasitology and C.T. is the recipient of a Burroughs Wellcome Fund New Investigator Award in Molecular Parasitology.

References

- [1] Laird PW, Zomerdijk JC, de Korte D, Borst P. In vivo labelling of intermediates in the discontinuous synthesis of mRNAs in *Trypanosoma brucei*. *EMBO J* 1987;6:1055–62.
- [2] Sutton RE, Boothroyd JC. Evidence for trans splicing in trypanosomes. *Cell* 1986;47:527–35.
- [3] Murphy WJ, Watkins KP, Agabian N. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. *Cell* 1986;47:517–25.
- [4] Krause M, Hirsh D. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 1987;49:753–61.
- [5] Tessier LH, Keller M, Chan RL, Fournier R, Weil JH, Imbault P. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in *Euglena*. *EMBO J* 1991;10:2621–5.
- [6] Rajkovic A, Davis RE, Simonsen JN, Rottman FM. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci USA* 1990;87:8879–83.
- [7] Nilsen TW. Trans-splicing of nematode premessenger RNA. *Annu Rev Microbiol* 1993;47:413–40.
- [8] Ullu E, Tschudi C, Günzl A. Trans-splicing in trypanosomatid protozoa. In: Smith DF, Parsons M, editors. *Molecular biology of parasitic protozoa*. Oxford: Oxford University Press, 1996:115–33.
- [9] Hannon GJ, Maroney PA, Nilsen TW. U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. *J Biol Chem* 1991;266:22792–5.
- [10] Maroney PA, Yu YT, Jankowska M, Nilsen TW. Direct analysis of nematode cis- and trans-spliceosomes: a functional role for U5 snRNA in spliced leader addition trans-splicing and the identification of novel Sm snRNPs. *RNA* 1996;2:735–45.
- [11] Tschudi C, Ullu E. Destruction of U2, U4, or U6 small nuclear RNA blocks trans splicing in trypanosome cells. *Cell* 1990;61:459–66.
- [12] Bruzik JP, Steitz JA. Spliced leader RNA sequences can substitute for the essential 5' end of U1 RNA during splicing in a mammalian in vitro system. *Cell* 1990;62:889–99.
- [13] Schnare MN, Gray MW. A candidate U1 small nuclear RNA for trypanosomatid protozoa. *J Biol Chem* 1999;274:23691–4.
- [14] Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, et al. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* 2000;6:163–9.
- [15] Fantoni A, Dare AO, Tschudi C. RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. *Mol Cell Biol* 1994;14:2021–8.
- [16] Li L, Otake LR, Xu Y, Michaeli S. The trans-spliceosomal U4 RNA from the monogenetic trypanosomatid *Leptomonas collosoma*. Cloning and identification of a transcribed tRNA-like element that controls its expression. *J Biol Chem* 2000;275:2259–64.
- [17] Nakaar V, Dare AO, Hong D, Ullu E, Tschudi C. Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. *Mol Cell Biol* 1994;14:6736–42.
- [18] Günzl A, Ullu E, Dorner M, Fragoso SP, Hoffmann KF, Milner JD, et al. Transcription of the *Trypanosoma brucei* spliced leader RNA gene is dependent only on the presence of upstream regulatory elements. *Mol Biochem Parasitol* 1997;85:67–76.
- [19] Palfi Z, Bindereif A. Immunological characterization and intracellular localization of trans-spliceosomal small nuclear ribonucleoproteins in *Trypanosoma brucei*. *J Biol Chem* 1992;267:20159–63.
- [20] Mair G, Ullu E, Tschudi C. Cotranscriptional cap 4 formation on the *Trypanosoma brucei* spliced leader RNA. *J Biol Chem* 2000;275:28994–9.
- [21] Günzl A, Bindereif A, Ullu E, Tschudi C. Determinants for cap trimethylation of the U2 small nuclear RNA are not conserved between *Trypanosoma brucei* and higher eukaryotic organisms. *Nucl Acids Res* 2000;28:3702–9.
- [22] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 1997;25:3389–402.
- [23] Hannon GJ, Maroney PA, Yu YT, Hannon GE, Nilsen TW. Interaction of U6 snRNA with a sequence required for function of the nematode SL RNA in trans-splicing. *Science* 1992;258:1775–80.
- [24] Palfi Z, Lucke S, Lahm HW, Lane WS, Krufft V, Bragado-Nilsson E, et al. The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. *Proc Natl Acad Sci USA* 2000;97:8967–72.