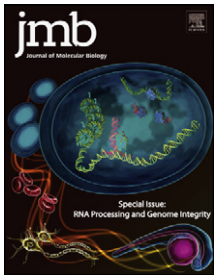

The Canonical Poly (A) Polymerase PAP1 Polyadenylates Non-Coding RNAs and Is Essential for snoRNA Biogenesis in *Trypanosoma brucei*

Vaibhav Chikne[1,†], Sachin Kumar Gupta[1,†], Tirza Doniger[1], K. Shanmugha Rajan[1], Smadar Cohen-Chalamish[1], Hiba Waldman Ben-Asher[1], Liat Kolet[1], Nasreen Hag Yahia[1], Ron Unger[1], Elisabetta Ullu[2], Nikolay G. Kolev[2], Christian Tschudi[3,4] and Shulamit Michaeli[1]

1 - The Mina and Everard Goodman Faculty of Life Sciences and Advanced Materials and Nanotechnology Institute, *Bar-Ilan University, Ramat-Gan 5290002, Israel*
2 - Department of Epidemiology and Microbial Diseases, *Yale School of Public Health, New Haven, CT 06536, USA*
3 - Department of Internal Medicine, *Yale University Medical School, 295 Congress Avenue, New Haven, CT 06536-0812, USA*
4 - Cell Biology, *Yale University Medical School, 295 Congress Avenue, New Haven, CT 06536-0812, USA*

Correspondence to Shulamit Michaeli: *The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 5290002, Israel. shulamit.michaeli@biu.ac.il*
http://dx.doi.org/10.1016/j.jmb.2017.04.015
Edited by Lori A Passmore

Abstract

The parasite *Trypanosoma brucei* is the causative agent of African sleeping sickness and is known for its unique RNA processing mechanisms that are common to all the kinetoplastidea including *Leishmania* and *Trypanosoma cruzi.* Trypanosomes possess two canonical RNA poly (A) polymerases (PAPs) termed PAP1 and PAP2. PAP1 is encoded by one of the only two genes harboring *cis*-spliced introns in this organism, and its function is currently unknown. In trypanosomes, all mRNAs, and non-coding RNAs such as small nucleolar RNAs (snoRNAs) and long non-coding RNAs (lncRNAs), undergo *trans*-splicing and polyadenylation. Here, we show that the function of PAP1, which is located in the nucleus, is to polyadenylate non-coding RNAs, which undergo *trans*-splicing and polyadenylation. Major substrates of PAP1 are the snoRNAs and lncRNAs. Under the silencing of either PAP1 or PAP2, the level of snoRNAs is reduced. The dual polyadenylation of snoRNA intermediates is carried out by both PAP2 and PAP1 and requires the factors essential for the polyadenylation of mRNAs. The dual polyadenylation of the precursor snoRNAs by PAPs may function to recruit the machinery essential for snoRNA processing.

Introduction

In trypanosomes, all mRNAs are *trans*-spliced and polyadenylated. In *trans*-splicing, a small exon, the spliced leader (SL), derived from the small SL RNA, is added to all mRNAs. *Trans*-splicing and polyadenylation are linked, and ~140 nt separate the poly (A) addition site of the upstream gene from the *trans*-splicing addition site of the downstream gene (3′ AG splice site). These coupled processes function to dissect the polycistronic mRNAs to monocistronic mRNAs [1,2]. Mutations in the polypyrimidine tract affect the polyadenylation of the upstream gene [3]. Recent study revealed that the knockdown of the canonical poly (A) polymerase (PAP) affects the 3′ end formation of mRNAs and *trans*-splicing [4]. Using tandem-affinity purification with tagged cleavage and polyadenylation specificity factor (CPSF) 160, a polyadenylation factor [5], and mass spectrometry revealed 10 associated components of the trypanosome polyadenylation machinery. It included homologs to all CPSF subunits, cleavage-stimulating factor (CstF) 50/64, and symplekin present in mammals [5], as well as two hypothetical proteins whose function is currently unknown [4]. RNA interference (RNAi) -mediated knockdown revealed that most of these factors are essential for growth and required for both polyadenylation and *trans*-splicing, strongly supporting

the coupling of these two processes [4]. However, the factors that coordinate the coupling between *trans*-splicing and polyadenylation are unknown. A protein from the splicing machinery U1A but no other U1 small nuclear ribonucleoprotein particle (snRNP)-associated proteins are present in complexes purified with the *Leishmania* polyadenylation factor CPSF73. U1A was shown to affect the polyadenylation of mRNAs, suggesting that this splicing factor is a genuine component of the polyadenylation machinery in these parasites [6].

Trypanosomes possess large repertoire of small nucleolar RNAs (snoRNAs). Our recent studies predict the presence of 83 H/ACA snoRNAs [7], which guide pseudouridylation and 79 C/D snoRNAs, which guide 2′-O-methylations [8]. Studies in *Leishmania* identified 81 H/ACA and 80 C/D snoRNAs [9]. Genome-wide mapping of pseudouridines in the two life stages of *Trypanosoma brucei* indicated that this modification, which is elevated in specific positions in ribosomal RNA (rRNA), may assist in coping with the temperature shift while cycling between the two hosts [7].

In eukaryotes, snoRNAs exhibit very diverse modes of genomic organization. In animals, most snoRNAs are located within introns of host genes that usually encode proteins related to ribosome biogenesis [10]. In yeast, the majority of snoRNAs are encoded by monocistronic genes transcribed from their own polymerase II promoter [11]. In plants, most snoRNAs are found in genomic clusters, which are independent and transcribed from their own promoter or located in introns of protein-coding genes [12]. snoRNA processing is carried out by different mechanisms. In animals, the major pathway for processing intronic snoRNAs depends on intron debranching followed by exonucleolytic trimming of the 5′ and 3′ free ends [13]. Processing of polycistronic pre-snoRNA has been well described in yeast [14]. A key factor in this process is Rnt1p, an RNase III endonuclease that cleaves the pre-snoRNA and liberates the individual snoRNAs with 5′ and 3′ extensions. These extensions are eliminated by Rat1p or Xrn1p 5′ → 3′ exonucleases and the nuclear exosome, which has 3′ → 5′ exonuclease activity [15]. The mature snoRNA ends are protected by the assembly of snoRNP core proteins [16]. In plants, most snoRNAs are processed from polycistronic precursors and this implicates an endonucleolytic cut to release the snoRNAs [17]. Nevertheless, the endonuclease responsible for this cleavage has not been identified. AtRTL2, the closest homolog of Rnt1p in *Arabidopsis*, could be implicated, but no effect on pre-snoRNA accumulation was detected in *atrtl2* mutants [18].

In trypanosomes, the snoRNAs are organized in clusters, which carry interspersed H/ACA and C/D genes [8,9,19,20]. Most of the clusters are transcribed polycistronically by RNA polymerase II, but since conventional pol II promoters do not exist upstream to protein coding genes, no promoter was reported upstream to snoRNA clusters [21]. However, studies in *Leptomonas collosma*, a monogenetic trypanosomatid, demonstrated that 700 bp present upstream to a snoRNA cluster enhanced the expression of a tagged snoRNA gene in an orientation-dependent manner [22]. The trypanosome pre-snoRNA encodes for several snoRNAs ranging from 1 to 9. Among the 50 snoRNA genes, 21 genes encode for a single snoRNA, which we termed solitary snoRNA [8]. All snoRNAs either solitary or clustered are liberated from transcripts that are *trans*-spliced and polyadenylated [2]. The majority of the clusters are also repeated in the genome several times. The *trans*-splicing and polyadenylation stabilize the pre-snoRNA before these are processed to liberate the individual RNAs [8,23]. Similar to plants, the enzyme that liberates the snoRNA from the precursor is currently unknown. It is also unknown how the machinery that polyadenylates mRNA differs from the one that polyadenylates snoRNA transcripts, if at all.

All eukaryotic genomes code for one or several canonical PAPs with similar, catalytic, central and RNA binding domains. Yeast has a single PAP [24], and the genomes of higher eukaryotes have two or three PAP genes. In mammals, two nuclear PAPs, PAP α [25,26] and PAP γ [27–29], were identified. PAP β (TPAP) is involved in the polyadenylation of testis-specific mRNAs [30]. The heterogeneity in PAP functions was recently reviewed [31].The other groups of PAPs are non-canonical PAPs, which include the GLD-2, TRF4/5, and CID1-type PAPs or poly (U) polymerases, and 2′-5′-oligo (A) synthetases. These enzymes share the catalytic domain with canonical PAPs but contain a different nucleotide base-recognition motif. These groups of polymerases are mainly involved in RNA turnover and quality control acting on rRNA, tRNAs, snoRNAs, and cryptic RNA polymerase II transcripts [32–35].

The canonical PAP adds long poly (A) tails to the 3′ end of the mRNAs, and in contrast, the Trfp proteins were shown to add short poly (A) tails to their substrate RNAs, which are assumed to trigger the efficient decay of the RNAs by the recruitment of the nuclear exosome complex [32–34]. Moreover, Trf4p-mediated polyadenylation is involved in the degradation of cryptic unstable transcripts generated by RNA polymerase II [34]. In *Drosophila*, two ncPAPs TRF4–1 and 2 were identified; the first was shown to polyadenylate small nuclear RNAs (snRNAs) [36]. However, more recent studies identified this protein in the cytoplasm and demonstrated its function in mRNA degradation [37]. Oligoadenylation of snoRNA precursors and its role in snoRNA processing were also reported in human cells [38].

The non-canonical PAPs are working in a complex known as the Trf4/Air2/Mtr4p polyadenylation complex (TRAMP) complex, which consists of either one of the two non-canonical PAPs (ncPAPs), Trf4p and Trf5p,