Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 140 (2005) 97–105
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Analysis of the highly efficient pre-mRNA processing region HX1 of *Trypanosoma cruzi*

Claudia P. Ben-Dov[1], Mariano J. Levin, Martín P. Vázquez*

Laboratorio de Biología Molecular de la Enfermedad de Chagas, INGEBI-CONICET, Departamento de Fisiología y Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires, Vuelta de Obligado 2490 2P, 1428 Buenos Aires, Argentina

Received 10 November 2004; received in revised form 31 December 2004; accepted 3 January 2005
Available online 19 January 2005

Abstract

Gene expression in trypanosomes is controlled mainly by post-transcriptional processes. This study was designed to analyse HX1, one of the TcP2β upstream intergenic regions. It is an efficient pre-mRNA processing region that has been widely and successfully used in *Trypanosoma cruzi* transfection vectors. Herein we compared its performance with other regions within the same locus, and we identified the sequence elements responsible for the HX1 efficiency in *trans*-splicing and protein synthesis. Our mutational analysis showed the flexibility of the branch point site selection for HX1 *trans*-splicing process. We demonstrated also that its 12 nt 5′UTR sequence contributes to both *trans*-splicing and translation efficiency.

The natural insertion of the repetitive element short interspersed repetitive element (SIRE) in one of the HX1 polypyrimidine tracts decreases the translated protein level by 40%. In this report, we demonstrated that this reduction is a consequence of a decrease of five-fold in the level of processed mRNA balanced by an increased efficiency of translation due to the inclusion of a 38 nt SIRE specific sequence in the 5′UTR of the mRNA.

Keywords: *trans*-Splicing; Post-transcriptional regulation; Repetitive sequences; Chagas disease

1. Introduction

Genome organization and expression in kinetoplastids is unusual. Transcription of protein-coding genes does not seem to be regulated at the level of transcription initiation [1–3]. In fact, it is presently unclear whether transcription even begins at a discrete location and RNA polymerase II promoters are unknown [3]. Consequently, regulation of gene expression is controlled mainly post-transcriptionally in these organisms [1]. Transcription of protein-coding genes is polycistronic and pre-mRNA maturation is accomplished by two coupled RNA-processing reactions, i.e., *trans*-splicing and 3′ end cleavage–polyadenylation [4–8]. The crucial information to accomplish these processes resides in the intergenic regions. Particularly, *trans*-splicing entails the addition of a 39-nt capped leader sequence (splice leader (SL)), which is derived from the 5′ end of the splice leaderRNA (SL RNA) [2,4], to the 3′ splice site (3′ss) present in the intergenic region. The 3′ss consists of three main *cis*-acting sequences: the branch point site, the AG acceptor site and the pyrimidine (PY) tract. Although branching occurs at one or more A residues upstream of a PY tract [9], no consensus branch point sequence has been determined yet due to the lack of available mapped branch sites [2]. In fact, the ubiquitin fusion gene [10] contains the only mapped branch point in *Trypanosoma cruzi*. The first AG dinucleotide downstream of the branch site is scanned by components of the spliceosome

Abbreviations: CAT, chloramphenicol acetyl transferase gene; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; NEO, neomycin resistant gene; ORF, open reading frame; PY, pyrimidine; pRib, ribosomal promotor; SL, splice leader; SL RNA, splice leader RNA; ss, splice site; UTR, untranslated region

* Corresponding author. Tel.: +54 11 47832871; fax: +54 11 47868578.
E-mail address: mvazquez@dna.uba.ar (C.P. Ben-Dov).
[1] Present address: Programme of Gene Regulation, Centre for Genomic Regulation, 08003 Barcelona, Spain.

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.01.001

and is used as the SL acceptor site [5,10]. The PY tracts are essential features for pre-mRNA maturation, governing both *trans*-splicing and polyadenylation processes [5–7]. In *T. cruzi*, they are U-rich sequences rather than C-rich sequences as described for *Leishmania* sp. intergenic regions [11]. However, it is currently unknown how the efficiency of *trans*-splicing of different genes is determined.

Improvements in transfection techniques and the development of highly efficient expression vectors have paved the way for a better understanding of the genetics of these protozoan parasites. These vectors were conceived taking into account the importance of the post-transcriptional regulation. In *T. cruzi* expression vectors, transcription is driven by the strong constitutive ribosomal Pol I promoter or the inducible T7 phage promoter, but high levels of mature translatable mRNAs are obtained by the presence of a highly efficient processing intergenic region between the promoter and the coding region [12–14].

We have previously described the cloning and organization of four loci coding for the *TcP2β* ribosomal protein genes [15]. The intergenic region upstream of *TcP2β*, named HX1, is a short (176 bp) and efficient region for pre-mRNA processing, and for this reason it has been widely used to drive pre-mRNA processing in transfection vectors [12–14]. One of the *TcP2β* locus, named H1.8, is distinct from the other three, since it is the only locus carrying two *TcP2β* genes arranged in a head to tail tandem array with the repetitive element short interspersed repetitive element (SIRE) present upstream and downstream of each gene of the tandem [15] (Fig. 1). SIRE insertion occurs at a PY tract, disrupting the original *trans*-splicing signal and providing a new 3′ss present in its 3′ end. This determines the generation of a TcP2β-H1.8 mRNA containing, after the SL sequence, 38 nt directly transcribed from SIRE, and a decrease of about 40% in the final protein level [12]. SIRE is repeated more than 1000 times in the *T. cruzi* genome and is inserted mainly in intergenic regions associated with open reading frames [16].

The model of TcP2β intergenic regions provides an excellent starting point to analyse *trans*-splicing and 5′ end pre-mRNA processing in *T. cruzi* since (a) we demonstrate that mutations in HX1 sequence strongly affect efficient processing of mRNA and (b) we show that the natural insertion of SIRE creates an in vivo mutation in which the efficiency of the intergenic region is compromised but not abolished.

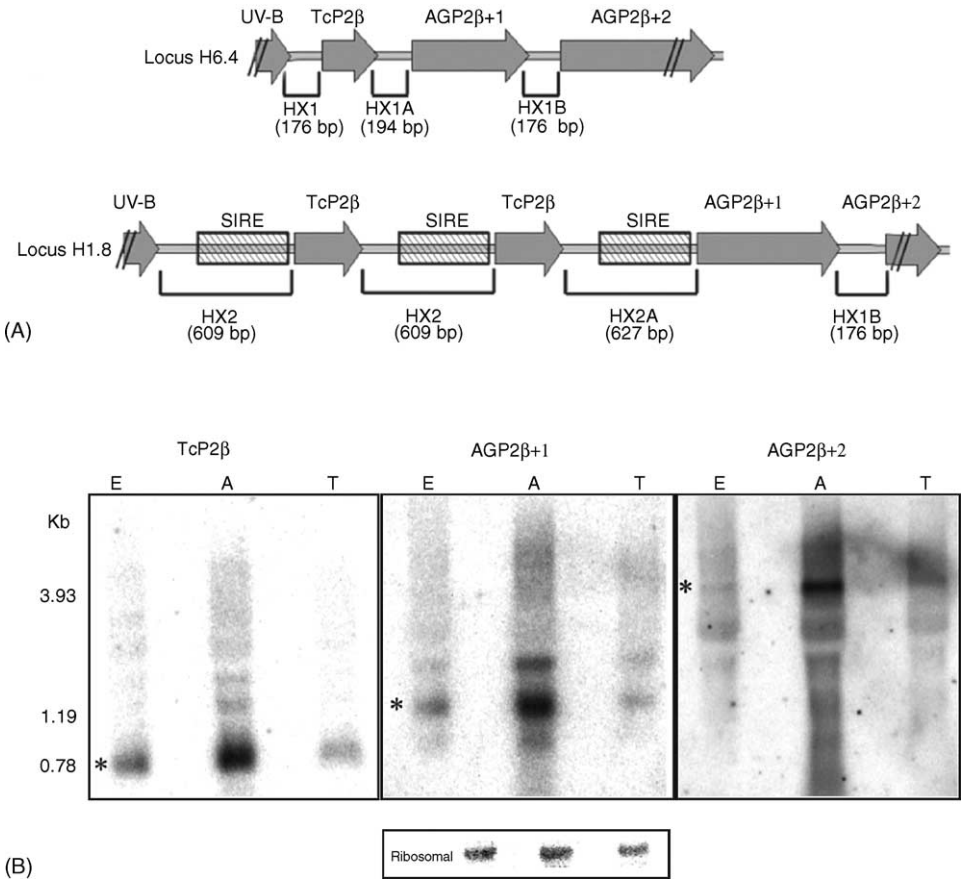

Fig. 1. *TcP2β* loci mRNA expression levels. (A) Schematic comparison between H6.4 and H1.8 loci. Arrows represent ORFs, UV-B, UV DNA damage protein ORF, intergenic regions are represented by a line and its size is indicated in brackets, dashed box symbolizes SIRE element. (B) Northern blot of total RNA from epimastigote (E), amastigote (A) and metacyclic trypomastigote (T) stages. The same membrane was sequentially hybridized with AGP2β+1, AGP2β+2, TcP2β and rRNA probes. The AGP2β+2 image appears overexposed.