Available online at www.sciencedirect.com
SCIENCE DIRECT®
Experimental Parasitology 110 (2005) 184–190
Experimental Parasitology
www.elsevier.com/locate/yexpr

Entamoeba histolytica: Comparative genomics of the pre-mRNA 3′ end processing machinery

César López-Camarillo [a,*], Esther Orozco [b], Laurence A. Marchat [c]

[a] *Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, San Lorenzo # 290, Col. del Valle, CP 03110, MexicoD.F., Mexico*
[b] *Departamento de Patología Experimental, CINVESTAV IPN, Ave. Instituto Politécnico Nacional # 2508, A.P. 14-740, CP 07300, Mexico D.F., Mexico*
[c] *Programa Institucional de Biomedicina Molecular, Escuela Nacional de Medicina y Homeopatía del IPN, Guillermo Massieu Helguera #239 Fracc. La Escalera, Ticoman, CP 07300, Mexico D.F., Mexico*

Received 17 January 2005; received in revised form 25 February 2005; accepted 28 February 2005
Available online 27 April 2005

Abstract

We report here the pre-mRNA 3′ end processing machinery in *Entamoeba histolytica*. Comparative analysis of the putative sequences participating in the pre-mRNA 3′ end processing of *E. histolytica* genes shows similitude and differences to those described for yeast and human transcripts. By a genomic survey, we identified 16 putative genes encoding for cleavage/polyadenylation factors in this parasite. *E. histolytica* pre-mRNA 3′ end processing machinery does not seem to contain homologous genes coding for human Symplekin, CFIm59, and CFIm68 proteins, neither sequences related to yeast Pta1p and Hrp1p. Protein sequence comparisons among *E. histolytica*, yeast, and human showed little variation in their functional domains through evolutive scale. *E. histolytica* pre-mRNA 3′ end processing machinery appears to be in an intermediate evolutionary position between mammals and yeast. From these analyses, we propose a hypothetical working model for the pre-mRNA 3′ end processing machinery in *E. histolytica*.

Index Descriptors and Abbreviations: *Entamoeba histolytica*; pre-mRNA 3′ end processing; polyadenylation; comparative genomics; aa, amino acid; CF, cleavage factor; CPSF, cleavage/polyadenylation specificity factor; CstF, cleavage stimulation factor; CTD, C-terminal domain; mRNA, messenger ribonucleic acid; NLS, nuclear localization signal; nt, nucleotides; PAP, poly(A) polymerase; pre-mRNA, precursor messenger ribonucleic acid; RBD, RNA binding domain; RNA pol II, RNA polymerase II; UTR, untranslated region

1. Introduction

The pre-mRNA 3′ end processing occurs in a two-step coupled reaction, denoted as cleavage and polyadenylation (Zhao et al., 1999). Both processes depend on *trans*-acting factors interacting in a coordinated way with *cis*-sequence motifs. First, the primary transcript is cleaved at the poly(A) site. Then, several adenosine residues are added to the 3′ end of the RNA fragment to form a poly(A) tail, producing mature mRNAs that can be translated. Poly(A) tails can be degraded or they may be readenylated in the cytoplasm. The poly(A) tail controls mRNA nuclear export, stability, and translation (López-Camarillo et al., 2003).

In mammals, three sequence elements are required for mRNA 3′ end formation: (i) the polyadenylation signal represented by the canonical AAUAAA hexanucleotide or related sequences, which are found 10–30 nt upstream the poly(A) site, (ii) the U/GU-rich element located downstream the poly(A) site, and (iii) the poly(A) site denoted generally by the CA dinucleotide (Zhao et al., 1999). The tetrameric cleavage/polyadenylation specificity factor (CPSF) recognizes the polyadenylation signal

* Corresponding author.
E-mail address: genomicas@yahoo.com.mx (C. López-Camarillo).

0014-4894/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.exppara.2005.02.024

(Keller et al., 1991). CPSF and FIP1 proteins interact with poly(A) polymerase (PAP) and tether it to RNA substrate to form the initial pre-mRNA processing complex (Kaufmann et al., 2004). Then, the trimeric cleavage stimulation factor (CstF) binds to U/GU-rich motif and contacts CPSF enhancing its RNA binding capacity (MacDonald et al., 1994). Finally, the cleavage factors (CFIm and CFIIm) are recruited to the poly(A) site to perform the RNA cleavage (Ruegsegger et al., 1996). Interestingly, CPSF and CstF also interact with RNA polymerase II (RNA pol II) revealing a functional link between transcription termination and pre-mRNA 3′ end processing (Dantonel et al., 1997).

Most RNA 3′ end processing factors have been characterized in mammals and in *Saccharomyces cerevisiae*. Notably, yeast RNA 3′ end processing machinery presents some additional factors, although it lacks some of the mammalian proteins (Proudfoot, 2004).

In *Entamoeba histolytica*, the protozoan parasite responsible for human amoebiasis, little is known about pre-mRNA 3′ end processing. The genomic raw information obtained from the *E. histolytica* genome project (Loftus et al., 2005) offers new experimental approaches to study gene expression and facilitates the determination of the complete repertoire of genes involved in RNA metabolism. To understand the posttranscriptional gene regulation in this parasite, we performed a genomic survey and in silico analysis of the pre-mRNA 3′ end processing machinery. We also analyzed mRNA expression of some these factors by RT-PCR assays. Our results showed that *E. histolytica* pre-mRNA 3′ end processing signals differ from those described in human and yeast. In contrast, pre-mRNA 3′ end processing factors are well conserved, suggesting a high conservation of these mechanisms through evolution.

2. Materials and methods

2.1. Genomic and cDNA sequences analyses

Entamoeba histolytica cDNA and genomic sequences were obtained from GenBank and Sanger databases. Nucleotide frequencies and multi-alignments of 3′ UTRs were performed using Bioedit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequence alignments of eukaryotic cleavage/polyadenylation factors were carried out using ClustalW with gap penalties of 10 (http://www.ch.embnet.org/software/ClustalW.html). Conserved sequences for each factor were used to screen the *E. histolytica* databases on TIGR (http://www.tigr.org/tdb/e2k1/eha1/) and Sanger servers (http://www.sanger.ac.uk/Projects/E_histolytica/) using BLASTP. Homologous protein sequences were defined by the following criteria: (i) identity and homology greater than 20 and 35%, respectively, with the query sequence; (ii) *e* value lower than 0.02; (iii) absence of stop codons in the coding sequence; and (iv) presence of conserved functional domains reported for the homologous eukaryotic proteins. To determine significant *e* values and identity/homology percentages, aa sequences were compared with human and yeast related proteins using BLAST. Functional domains were predicted by ScanProsite and Pfam programs (http://us.expasy.org/tools/scanprosite/) (http://www.sanger.ac.uk/Software/Pfam/).

2.2. RT-PCR experiments

Total RNA was obtained from clone A trophozoites (HM1-IMSS strain) and submitted to RT-PCR assays using specific internal primers for EhCPSF processing factors. Amplified products were analyzed by 6% PAGE and ethidium bromide staining.

3. Results

3.1. Putative pre-mRNA 3′ end processing signals in E. histolytica genes

To identify *E. histolytica* sequences participating in pre-mRNA 3′ end processing, we aligned 320 nt of 50 different genomic sequences (80 nt corresponding to coding region and 240 nt to 3′ UTR). The analysis of the single-nucleotide frequencies at each position showed that the AT content was higher in 3′ UTRs (80%) than in coding regions (68%). In addition, we found a high frequency of T's in the first 10–50 nt of 3′ UTRs (Fig. 1A). We also analyzed 15 genes whose cDNA and 3′ UTR genomic sequences were available, as well as the genomic sequences of the 50 genes described in Fig. 1A (data not shown). Fig. 1B shows a representative summary of these analyses. Multiple alignments of the cDNA and first 80 nt of 3′ UTR genomic sequences evidenced two U-rich regions, one located from 1 to 30 nt upstream the poly(A) site (Bruchhaus et al., 1993), and the other one from 3 to 30 nt downstream the poly(A) site. Similar U-rich motifs participate in the initial step of pre-mRNA 3′ end processing in mammals and yeast (Zhao et al., 1999). Notably, the poly(A) site does not seem to have a consensus sequence. The consensus polyadenylation signal UA(A/U)UU (Bruchhaus et al., 1993) was only found in 65% of the genes and 8% of them include the stop codon (Fig. 1B). These findings suggest that 35% of *E. histolytica* genes could use different regulatory sequences for pre-mRNA 3′ end processing.

In summary, *E. histolytica* genes contain four 3′ UTR *cis*-acting motifs, which could be required for mRNA 3′ end formation (Fig. 1C): (i) the consensus polyadenylation signal or variants of this sequence located 10–30 nt