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

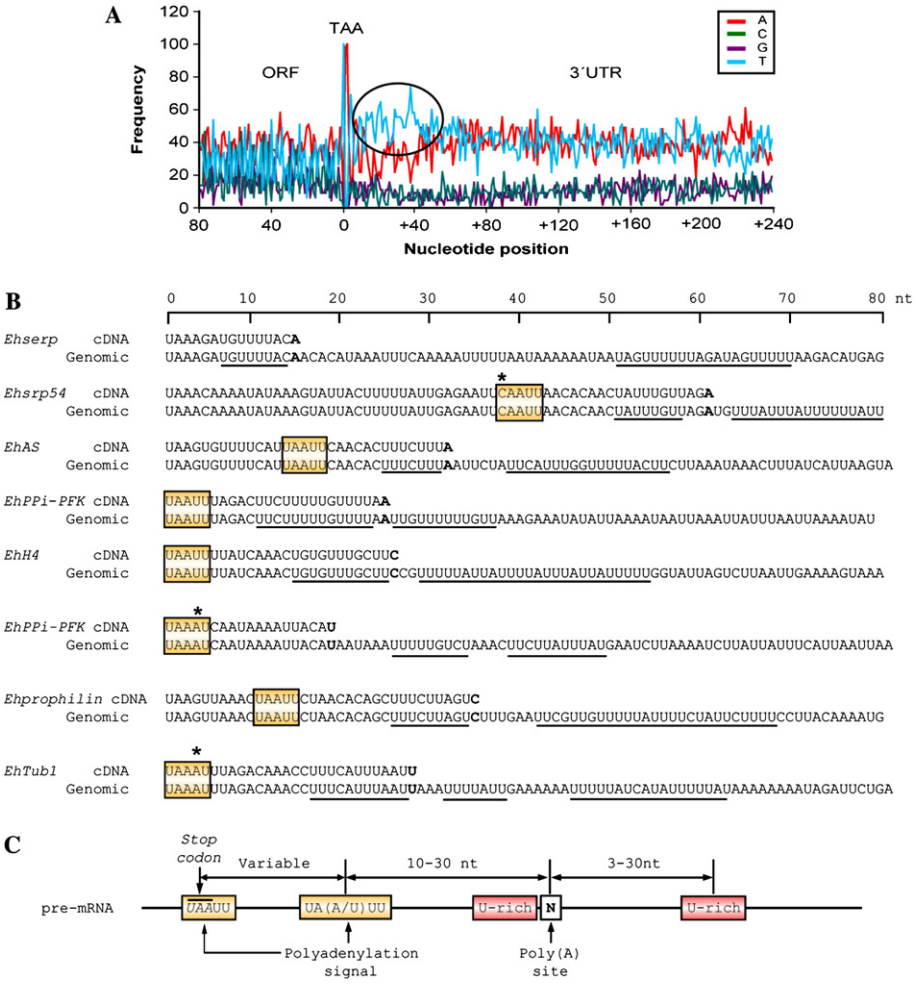

Fig. 1. Nucleotide frequency at the 3′ UTR and putative pre-mRNA 3′ end processing signals in *E. histolytica* genes. (A) Nucleotide (nt) frequencies of 50 different 3′ UTR gene sequences at −80 to +240 base positions with respect to the stop codon site at 0. Ellipse marks the region with a high frequency of T's. Nucleotides are denoted by colored lines (inlay box). (B) Representative multi-alignment of eight cDNA and genomic sequences up to 80 nt from translation stop codon at position 0. The scale at the top indicates the size in nt. Yellow boxes, polyadenylation signal; asterisks, nt that do not match with the consensus polyadenylation motif. The U-rich elements are underlined and the poly(A) site is printed in boldface letter. (C) Schematic representation of pre-mRNA 3′ end processing signals. Positions of the putative RNA processing sequences with respect to the polyadenylation signal, poly(A) site, and TAA stop codon are marked by upper arrows.

upstream the poly(A) site, (ii) the U-rich tract located 1–30 nt upstream the poly(A) site, (iii) the poly(A) site, and (iv) a novel U-rich element located 3–30 nt downstream the poly(A) site. *E. histolytica* 3′ UTR regulatory motifs differ from those described in human and yeast. Thus, this parasite could be a good model to gain insights into the evolutionary changes in pre-mRNA 3′ end processing sequences recognition.

3.2. Putative *E. histolytica* cleavage and polyadenylation factors

Our analyses of the *E. histolytica* genome yield a list of 16 genes that encode putative proteins involved in pre-mRNA 3′ end processing (Table 1). We detected proteins homologous to the four main human CPSF, CstF, CFIm, and CFIIm complexes, as well as additional proteins participating in pre-mRNA 3′ end processing. Interestingly, *E. histolytica* has some proteins homologous to factors present in yeast but not in human, and vice versa, suggesting that it has an intermediate evolutionary position between these organisms. For example, in *E. histolytica*, we identified a homologue of the yeast Psf2, which is absent in human. In contrast, we did not find yeast Pta1p and Hrp1p proteins; neither human CFIm59 and CFIm68 homologues, which are also absent in yeast.

3.2.1. Cleavage/polyadenylation specificity factor

Human CPSF complex consists of four subunits of 160, 100, 73, and 30 kDa, as well as PAP and FIP1 proteins. The yeast homologues are represented by Cft1p/Yhh1p, Cft2p/Ydh1p, Brr5/Ysh1p, Yth1p, Pap1, and Fip1 factors (Proudfoot, 2004). CPSF-160 binds to the

Table 1
Comparisons of *Entamoeba histolytica*, *Homo sapiens*, and *Saccharomyces cerevisiae* pre-mRNA 3′ end processing factors

Entamoeba histolytica			*Homo sapiens*					*Saccharomyces cerevisiae*				
Predicted protein	Locus[a]	mRNA expression	Protein	Accession number[b]	*e* value	H (%)	I (%)	Protein	Accession number[b]	*e* value	H (%)	I(%)
EhCPSF160	448.m00031	−	CPSF160	Q10570	6e−12	39	22	Cftlp/Yhhlp	Q06632	7e−04	39	20
EhCPSFl00	193.m00066	−	CPSF100	Q9P2I0	6e−24	39	22	Cft2p/Ydhlp	Q12102	8e−04	42	17
EhCPSF73	33.m00214	+	CPSF73	Q9UKF6	le−54	64	45	Brr5/Yshlp	Q06224	7e−47	61	40
EhCPSF30	4.m00640	+	CPSF30	095639	2e−28	60	44	Ythlp	Q06102	4e−19	52	38
EhPAP	16.m00324	+	PAP-β	Q9NRJ5	5e−44	49	27	Pap1	P29468	le−35	52	32
EhFIPl	22.m00317	+	FIP1	Q6UN15	4e−07	59	39	Fipl	P45976	5e−06	62	47
—	—		Symplekin	Q92797	—	—	—	Ptalp	Q01329	—	—	—
EhCstF77	19.m00314	nd	CstF77	Q12996	3e−07	53	24	RNA14	P25298	nd	52	18
EhCstF64	2.m00562	nd	CstF64	P33240	9e−07	54	37	RNA15	P25299	0.001	57	26
EhCstF50	24.m00285	nd	CstF50	Q05048	0.003	42	26	—	—	—	—	—
EhCFlm25	40.m002l6	nd	CFIm25	043809	6e−22	55	32	—	—	—	—	—
—	—		CFIm59	Q8N684	—	—	—	—	—	—	—	—
—	—		CFIm68	Q16630	—	—	—	—	—	—	—	—
Unknown			CF Im72	Unknown				Unknown				
EhClpl	55.m00185	nd	C1P1	Q92989	le−13	40	22	Clpl	Q08685	4e−10	45	23
EhPcfll	3.m00613	nd	PCF11	094913	le−04	50	26	Pcfl 1	P39081	0.052	48	24
EhPfs2	77.m00151	nd	—	—	—	—		Pfs2p	P42841	7e−54	51	36
—	—		—	—	—	—	—	Hrpl	Q99383	—	—	—
EhMpel	37.m00231	nd	Rrbp6[c]	Q7Z6E9	3e−10	38	25	Mpel	P35728	0.002	55	38
EhSsu72	l.m00620	nd	PNAS-120[c]	Q9BZS6	7e−27	55	35	Ssu72	P53538	2e−21	52	34
EhPC4/Subl	69.m00181	nd	PC4	P53999	2e−06	65	40	Subl	P54000	5e−06	61	44

nd: not determined.

[a] TIGR *E. histolytica* genome sequence project databases.

[b] Swiss-Prot/TrEMBL databases.

[c] Participation in pre-mRNA 3′ end processing has not been experimentally tested.

polyadenylation signal in the nascent RNA after RNA pol II transcribe the genes. It forms a complex with CPSF-100, CPSF-73, and CPSF-30, and interacts with PAP to initiate the recruitment of other processing factors. PAP adds the poly(A) tails at specific sites of pre-mRNAs. FIP1 interacts with CPSF-160 and PAP, stimulating the polyadenylation reaction (Kaufmann et al., 2004). Thus, CPSF is required for both cleavage and polyadenylation. Yeast processing complex is assembled in a similar way; however, in this organism FIP1 negatively regulates PAP activity (Helmling et al., 2001).

Entamoeba histolytica has the four human CPSF-like proteins, as well as the EhPAP and EhFIP1 factors (Table 1). EhCPSF proteins differ in molecular weight from those described in human and yeast, but they contain the characteristic functional domains. To facilitate their identification, we kept the nomenclature used for human (or yeast proteins in the case of factors absent in human), adding the Eh suffix. EhCPSF-160 (132 kDa) contains a divergent RNA-binding domain (RBD) in the central domain similar to the one described for other CPSF-160 proteins, suggesting that it could bind to RNA. EhCPSF-100 (79 kDa) and EhCPSF-73 (86 kDa) factors contain a metallo-β-lactamase domain (PF00753) in their N-terminus, as described for related human CPSF proteins. EhCPSF-30 (21 kDa) contains four CCHC zinc fingers (PF00642) at the N-terminus, but the C terminal end is less conserved and it does not possess the CCHC knuckle described in other CPSF-30 factors. EhPAP (60 kDa) protein has a conserved N-terminus with three aspartate residues (D_{76}, D_{78}, and D_{130}) involved in nucleotide transfer and the F/YGS motif responsible for ATP binding in mammals PAPs (Garcia-Vivas et al., 2005). EhFIP1 (27 kDa) contains 27 invariant aa residues, a divergent nuclear localization signal (NLS), and an arginine-rich region at the N-terminus, like human factor. Taken together, these data suggest that EhCPSF complex could be organized in a similar way as the one described for human CPSF.

Surprisingly, we do not detect mRNA expression of EhCPSF-160 and EhCPSF-100 under the conditions tested here, whereas EhCPSF-73, EhCPSF-30, EhPAP, and EhFIP1 mRNA were expressed (Table 1), suggesting that transcription of these genes is not coordinated in basal culture conditions.

3.2.2. Cleavage stimulation factor

Human CstF is formed by three proteins of 77, 64, and 50 kDa. CstF RNA binding is mediated by the 64 kDa protein, which recognizes the U/GU-rich motif downstream the poly(A) site. CstF-77 bridges CstF-64 and CstF-50, and interacts with CPSF-160 to stabilize the RNA processing complex. CstF-77 participates in the poly(A) site choice. In yeast, CstF-77 and CstF-64 are represented by RNA 14 and RNA 15 factors, respectively, but no CstF-50 homologue has been identified (Zhao et al., 1999).

In *E. histolytica*, we identified three genes coding for proteins related to human CstF complex (Table 1). EhCstF-64 (35 kDa) contains a RBD (PF00076) and a divergent bipartite NLS (PS00015) at 140–217 and 24–64 aa residues, respectively. It also has a CCH zinc finger domain at the 255–270 aa, and glutamic acid and lysine rich regions at the N-terminus. EhCstF-77 (65 kDa) contains a metallo-β-lactamase motif, whereas EhCstF-50 (55 kDa) has two WD-40 repeats (PDOC00574) as described in human CstF-50 factor.

3.2.3. Cleavage factors I and II

Human CFIm complex consists of four proteins of 25, 59, 68, and 72 kDa, which are required for the initial stage of RNA processing. CFIm-25 factor binds to RNA and interacts with CFIm-68, PAP, and poly(A) binding protein II (Dettwiler et al., 2004). Interestingly, CFIm-68 also interacts with splicing factors. CFIm-59 and CFIm-72 factors remain uncharacterized. Human CFIIm complex is formed by ClP1 and Pcf11 polypeptides. ClP1 interacts with CFIm-25 and CPSF-100 factors (Vries et al., 2000). Intriguingly, the identity of the factor responsible for RNA cleavage remains elusive. Notably, yeast does not contain homologues for CFIm proteins, but it contains the ClP1and Pcf11 factors. Yeast Pcf11 interacts with RNA pol II and it also has a role in efficient transcription termination (Sadowski et al., 2003).

In *E. histolytica*, we found genes encoding proteins related to human CFIm and CFIIm factors (Table 1). Intriguingly, EhCFIm does not seem to be a multiprotein complex since we only identified the 25 kDa subunit encoding gene. EhCFIm-25 presents 55% homology and 32% identity with human CFIm-25. It has a bipartite NLS at 6–22 aa residues, which is not found in the human related CFIm-25 factor. *E. histolytica* CFIIm complex seems to consist of two proteins related to human and yeast Pcf11 and ClP1 factors (Table 1). EhPcf11 (47 kDa) shares sequence similarities with RNA pol II carboxy terminal domain (CTD) binding proteins. The highest similarity region covers 130 aa of the N-terminal known as the CTD interaction domain, which is conserved between yeast, human, and *E. histolytica* proteins. It has a short N-terminus and it did not present the stretch of 20 glutamines described in yeast Pcf11 (Sadowski et al., 2003). EhClP1 (47 kDa) contains the Walker A (A/G-X_4-K-S/T) and B motifs (PF03029), which have been implicated in ATP/GTP binding in yeast and human homologues (Vries et al., 2000).

3.2.4. Additional proteins

In addition to CPSF, CstF, CFIm, and CFIIm, other proteins participate in pre-mRNA 3′ end pro-

cessing. The human transcriptional coactivator PC4 (Sub1 in yeast) acts as an anti-terminator of RNA pol II transcription through interaction with CstF-64 factor (Calvo and Manley, 2001). On the other hand, yeast transcription factor Ssu72 interacts with Pta1p and Ydh1p/Cft2p proteins, which are components of the cleavage and polyadenylation complex, and bridges TFIIB and RNA pol II factors during transcription initiation (Dichtl et al., 2002a,b). Similar interactions between human CPSF and TFIID factors link mRNA 3′ end formation to transcription. In yeast, Psf2 and Mpe1 bridge processing factors to promote the assembly of the pre-mRNA 3′ end processing complex (Ohnacker et al., 2000; Vo et al., 2001). Notably, related Psf2 factor appears to be absent in human.

Entamoeba histolytica contains two genes codifying for proteins related to human and yeast PC4/Sub1 and Ssu72 proteins. EhPC4/Sub1 exhibits high homology (65 and 61%) and identity (40 and 44%) with human PC4 and yeast Sub1, respectively. EhSsu72 (22 kDa) has 52 and 34% homology and identity, respectively, with yeast Ssu72 protein. We also identified the putative EhPfs2 and EhMpe1 factors (Table 1). EhPfs2 sequence presents six WD transducin repeats (PS50294) between 114 and 404 aa residues. In yeast Pfs2, these motifs are essential for cleavage and polyadenylation reaction (Ohnacker et al., 2000). EhMpe1 (43 kDa) contains the CCHC-type zinc finger motif, the so-called A region (from 1 to 27 aa residues), the cysteine-rich B domain (from 202 to 260 aa residues), and a bipartite NLS at the C-terminus, as described in yeast Mpe1 factor (Vo et al., 2001).

3.3. The E. histolytica pre-mRNA 3′ end processing machinery

Based on the knowledge of the pre-mRNA 3′ end processing in human and yeast, and considering that this process seems to be conserved through the evolutionary scale, we propose a hypothetical working model for the pre-mRNA 3′ end processing in *E. histolytica* (Fig. 2). We suggest that EhCPSF complex could bind to the consensus polyadenylation signal UA(A/U)UU, and EhCstF complex could recognize the U-rich motif downstream the poly(A) site to initiate the recruitment of the other processing factors. EhCFIm and EhCFIIm proteins could bind close to the poly(A) site to perform the RNA cleavage (Fig. 2). Finally, EhPAP could synthesize the poly(A) tails, in a reaction stimulated by the interaction with other processing factors. Additional interactions with EhMpe1, EhPC4/Sub1, EhSsu72, and EhPfs2 factors, could contribute to the efficient cleavage and polyadenylation reaction.

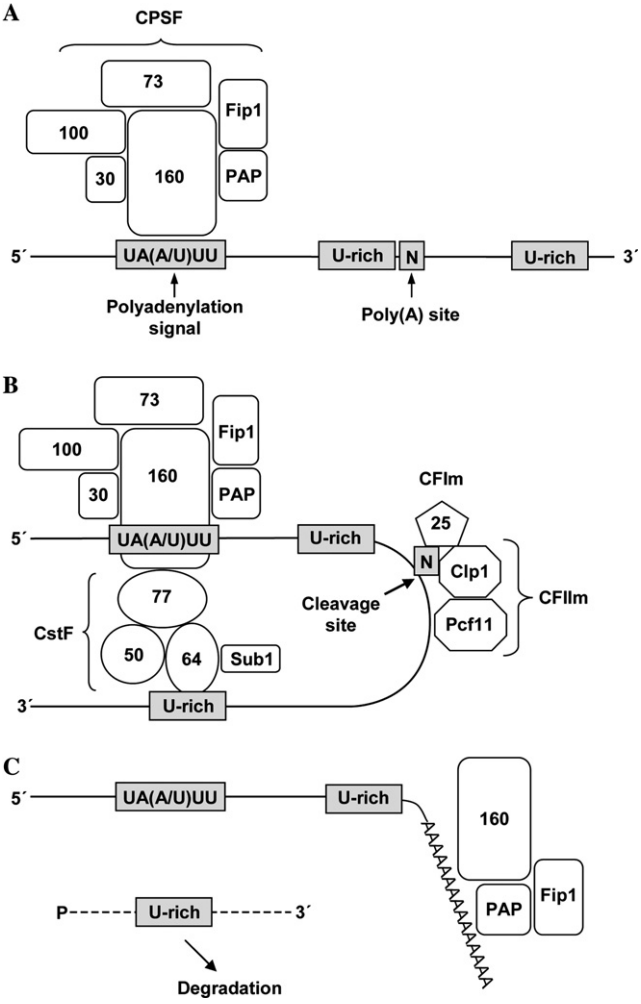

Fig. 2. Working model for the pre-mRNA 3′ end processing machinery of *E. histolytica*. (A) The *E. histolytica* processing complex could assembles through a cooperative binding of EhCPSF, EhFIP1, and EhPAP at the UA(A/U)UU polyadenylation signal. (B) EhCstF could bind the U-rich sequence downstream the poly(A) site. CFIm and CFIIm factors could interact with RNA in the boundaries of the poly(A) site. (C) After RNA cleavage by CFIm and CFIIm factors, PAP bound to CPSF could synthesize the poly(A) tail. Thin line, RNA; gray rectangles, pre-mRNA 3′ end processing signals. Depicted interactions between proteins and RNA are hypothetical.

The molecular function of the predicted *E. histolytica* proteins in pre-mRNA 3′ end formation is currently under experimental investigation.

Acknowledgments

This work was supported by CGPI-IPN (20040056) and CONACyT-México (39888-M) grants. Preliminary gene sequences were obtained from TIGR and Sanger *E. histolytica* genome databases. The TIGR and Sanger sequencing efforts are supported by awards from Institute of Allergy and Infectious Diseases, National Institutes of Health, USA, and from The Wellcome Trust, UK.

References

Bruchhaus, M., Leippe, C., Lioutas, C., Tannich, E., 1993. Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. DNA and Cell Biology 12, 925–933.

Calvo, O., Manley, J.L., 2001. Evolutionary conserved interaction between CstF64 and PC4 links transcription, polyadenylation and termination. Molecular Cell 7, 1013–1023.

Dantonel, J.C., Murthy, K.G., Manley, J.L., Tora, L., 1997. Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature 389, 399–402.

Dettwiler, S., Aringhieri, Ch., Cardinale, S., Keller, W., Barabino, S.M.L., 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein–protein interactions and subcellular localization. The Journal of Biological Chemistry 279, 35788–35797.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S., Keller, W., 2002a. Yhh1p/Ctf1p directly links poly(A) site recognition and RNA polymerase II transcription termination. EMBO Journal 21, 4125–4135.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H., Keller, W., 2002b. A Role for *SSU72* in balancing RNA polymerase II transcription elongation and termination. Molecular Cell 10, 1139–1150.

Garcia-Vivas, J., López-Camarillo, C., Azuara-Liceaga, E., Orozco, E., Marchat, L.A., 2005. *Entamoeba histolytica*: cloning and expression of the poly(A) polymerase EhPAP. Experimental Parasitology (in press).

Helmling, S., Zhelkovsky, A., Moore, C., 2001. Fip1 regulates the activity of poly(A) polymerase through multiple interactions. Molecular and Cellular Biology 21, 2026–2037.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., Keller, W., 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO Journal 23, 616–626.

Keller, W., Bienroth, K., Christofori, G., 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ end processing signal AAUAAA. EMBO Journal 10, 4241–4249.

Loftus, B. et al., 2005. The genome of the protist parasite *Entamoeba histolytica*. Nature 433, 865–868.

López-Camarillo, C., Luna-Arias, J.P., Marchat, L.A., Orozco, E., 2003. *EhPgp5* mRNA stability is a regulatory event in the *Entamoeba histolytica* MDR phenotype. The Journal of Biological Chemistry 278, 11273–11280.

MacDonald, C.C., Wilusz, J., Shenk, T., 1994. The 64-kDa subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Molecular and Cellular Biology 14, 6647–6654.

Ohnacker, M., Barabino, S.M.L., Preker, P.J., Keller, W., 2000. The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3-end-processing complex. EMBO Journal 19, 37–47.

Proudfoot, N., 2004. New perspectives on connecting messenger RNA 3′ end formation to transcription. Current Opinion in Cell Biology 16, 272–278.

Ruegsegger, U., Beyer, K., Keller, W., 1996. Purification and characterization of human cleavage factor Im involved in the 3′ end processing of messenger RNA precursors. The Journal of Biological Chemistry 271, 6107–6113.

Sadowski, M., Dichtl, B., Hubner, W., Keller, W., 2003. Independent functions of yeast Pcf11p in pre-mRNA 3′ end processing and in transcription termination. EMBO Journal 22, 2167–2177.

Vo, A., Minet, M., Schmitter, J.M., Lacroute, F., Wyers, F., 2001. Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for cleavage and polyadenylation of mRNA. Molecular and Cellular Biology 21, 8346–8356.

Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., Keller, W., 2000. Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. EMBO Journal 19, 5895–5904.

Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation and interrelationships with other steps in mRNA synthesis. Microbiology and Molecular Biology Reviews 63, 405–445.