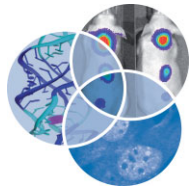
Advanced Review



Deciphering molecular mechanisms of mRNA metabolism in the deep-branching eukaryote *Entamoeba histolytica*

César López-Camarillo,[1]* Itzel López-Rosas,[1]
Juan David Ospina-Villa[2] and Laurence A. Marchat[2,3]

Although extraordinary rapid advance has been made in the knowledge of mechanisms regulating messenger RNA (mRNA) metabolism in mammals and yeast, little information is known in deep-branching eukaryotes. The complete genome sequence of *Entamoeba histolytica*, the protozoan parasite responsible for human amoebiasis, provided a lot of information for the identification and comparison of regulatory sequences and proteins potentially involved in mRNA synthesis, processing, and degradation. Here, we review the current knowledge of mRNA metabolism in this human pathogen. Several DNA motifs in promoter and nuclear factors involved in transcription, as well as conserved polyadenylation sequences in mRNA 3′-untranslated region and possible cleavage and polyadenylation factors, are described. In addition, we present recent data about proteins involved in mRNA decay with a special focus on the recently reported P-bodies in amoeba. Models for mechanisms of decapping and deadenylation-dependent pathways are discussed. We also review RNA-based gene silencing mechanisms and describe the DEAD/DExH box RNA helicases that are molecular players in all mRNA metabolism reactions. The functional characterization of selected proteins allows us to define a general framework to describe how mRNA synthesis, processing, and decay may occur in *E. histolytica*. Taken altogether, studies of mRNA metabolism in this single-celled eukaryotic model suggest the conservation of specific gene expression regulatory events through evolution.

How to cite this article:
WIREs RNA 2014, 5:247–262. doi: 10.1002/wrna.1205

INTRODUCTION

Messenger RNA (mRNA) metabolism includes cellular processes that are critical for cell survival because they contribute to turning genes on/off and regulating mRNA turnover in response to environmental signals and cell–cell communication. In eukaryotic cells, the main steps of mRNA metabolism include: (1) transcription, which is the synthesis of a mRNA precursor (pre-mRNA) from a particular DNA segment; (2) cotranscriptional modifications of the pre-mRNA molecule that include the addition of a 7-methylguanosine at the 5′-end of the nascent molecule (capping), intron removal (splicing), and pre-mRNA

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: genomicas@yahoo.com.mx

[1]Genomics Sciences Program, Autonomous University of Mexico City, Mexico City, Mexico

[2]Institutional Program of Molecular Biomedicine, National School of Medicine and Homeopathy of the National Polytechnic Institute, Mexico City, Mexico

[3]Biotechnology Program, National School of Medicine and Homeopathy of the National Polytechnic Institute, Mexico City, Mexico

Conflict of interest: The authors have declared no conflicts of interest for this article.

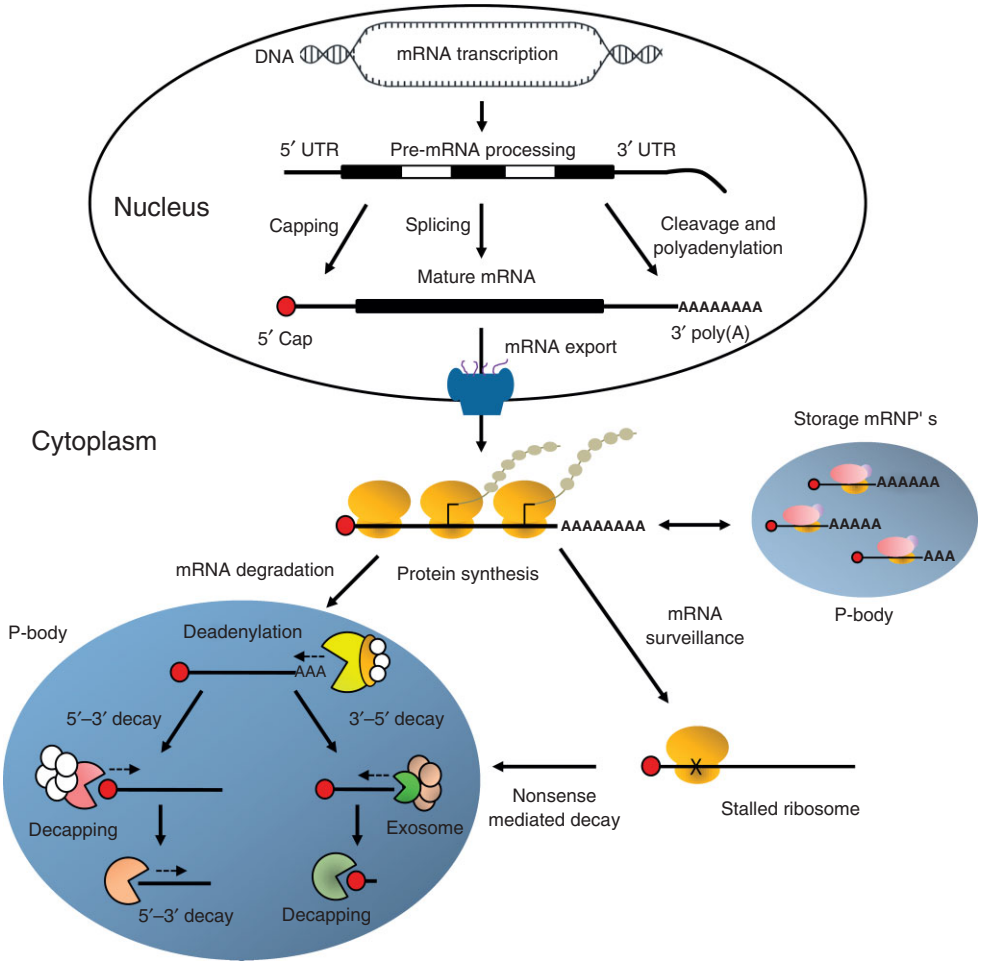


FIGURE 1 | Messenger RNA (mRNA) life cycle. Transcript synthesis, to translation and degradation occur in the form of dynamic protein-bound messenger ribonucleoprotein complexes (mRNPs). Within an mRNP complex, a transcript is bound by a changing set of proteins that mediate the cotranscriptional and post-transcriptional events that make up an mRNA life cycle. Precursor mRNAs are synthesized in nucleus by RNA polymerase II. Then, transcripts undergo capping, splicing, and 3′-end cleavage and polyadenylation in a cotranscriptional coupled reaction. Mature mRNAs are exported to the cytoplasm, translated to proteins in ribosomes, and then degraded. Remarkably, mRNA degradation and storage occur in cytoplasmic structures denoted as P-body. The mRNA decay starts with the deadenylation reaction followed by mRNA body degradation in 5′ → 3′ direction (decapping and 5′ → 3′decay) or alternatively deadenylated transcripts are reduced in 3′ → 5′ direction by the exosome complex and decapping enzymes. The mRNAs are subject to surveillance mechanisms, which check for the presence of premature stop codons. Upon detection of nonsense codons, aberrant mRNAs are degraded through nonsense-mediated decay.

3′-end cleavage followed by poly(A) tail addition; and (3) degradation of mRNA in the cytoplasm (Figure 1). Transcription and pre-mRNA processing events have generally been studied separately; however, these processes are intimately coupled *in vivo*.[1]

To date, little is known about mRNA metabolism in *Entamoeba histolytica*, the protozoan responsible for human amoebiasis. This parasite causes intestinal dysentery and hepatic abscesses that result in 70,000–100,000 deaths a year, making it a leading cause of parasitic death in humans.[2] However, the high incidence of asymptomatic infections suggests that differential expression of virulence genes may explain the variable amoebiasis outcomes. The publication of the *E. histolytica* genome sequence a few years ago[3] provides an extraordinary tool to identify and functionally characterize novel genes that are important for pathogen survival and host infection. It is also a very important step forward in studying and understanding the molecular mechanisms regulating gene expression, particularly the proteins and sequence elements that are relevant for mRNA metabolism in this primitive eukaryote. In this review, we focus on the current knowledge about the molecular basis for transcription, pre-mRNA 3′-end processing, mRNA degradation, and RNA-based

gene silencing mechanisms in *E. histolytica*. We also describe the DEAD/DExH box RNA helicases family of *E. histolytica* because these ubiquitous enzymes are central players in mRNA metabolism.

DNA SEQUENCES AND NUCLEAR FACTORS INVOLVED IN TRANSCRIPTION REGULATION

RNA polymerase II (RNA pol II) is the main enzyme responsible for transcription of protein-encoding genes in eukaryotic cells. It associates with six general transcription factors to form the preinitiation complex (PIC), which is positioned around the transcription initiation site through the interaction of TATA-box-binding protein (TBP) with promoter.[4] PIC also contains several pre-mRNA processing factors, such as the transcriptional coactivator/polyadenylation factor PC4, the cleavage stimulating factor 64-kDa subunit (CstF-64), and the cleavage and polyadenylation specificity factor 160-kDa subunit (CPSF-160).[5,6]

Current Knowledge of Transcription in *E. histolytica*

Although a large number of articles have characterized the transcriptome of *E. histolytica* in different conditions, *cis*-elements and *trans*-acting factors participating in mRNA synthesis remain poorly understood in this parasite. Only a few DNA motifs have been identified in promoters and several genes encoding transcription factors have been cloned and characterized (Table S1, Supporting Information).

E. histolytica genome contains one annotated RNA pol II largest subunit (RPB1) gene.[7] EhRNA Pol II has several unique features including a highly divergent α-amanitin-binding region, which explains why trophozoites are resistant to this drug.[8] Strikingly, the canonical heptapeptide repeats (YSPTSPS) found in the carboxy terminal domain (CTD) of other eukaryotic homologs are absent in *E. histolytica* enzyme. However, *in silico* analyses showed that the noncanonical CTD of EhRNA Pol II retains the proline/serine-rich motifs, which represent potential phosphorylation sites. In spite of its sequence divergence, potential modifications of the CTD by kinases and phosphatases could modulate protein–protein interactions, opening the possibility of CTD functions in transcription and pre-mRNA processing in amoeba. The existence of RNA pol II with a noncanonical CTD has been reported in other protozoan parasites, including *Plasmodium*, *Giardia*, *Trypanosoma*, *Leishmania*, and *Trichomonas*.[9] Notably, the pseudo-CTD is essential for RNA synthesis by RNA Pol II in *Trypanosoma brucei*.[10] In *E. histolytica*, neither the recombinant RNA pol II protein nor functions of potential modifications have been characterized yet. Therefore, further studies are required to confirm the potential role of the noncanonical CTD in the coordination of transcription and pre-mRNA processing events in this human pathogen.

The core regulatory elements in *E. histolytica* promoters include an atypical TATA box, the initiator (Inr) element, as well as a GAAC element (with a variable position between the TATA and Inr sequences) that is able to determine the transcription start site independently of these elements.[11] Interestingly, TATA box is recognized *in vitro* by two functional TBPs, namely EhTBP and EhTBP-related factor 1 (EhTRF1).[12,13] Functional characterization of a few promoters allowed the identification of several DNA sequences and transcription factors that control transcription in amoeba. Purdy et al. reported five upstream regulatory elements (URE1 to URE5) in the galactose and galactose *N*-acetyl-galactosamine inhibitable lectin (Gal/GalNAc lectin) gene (*hgl5*) promoter.[14] Notably, URE3 and URE4 sequences in DNA were recognized by URE3-binding protein (URE3-BP) and enhancer-binding proteins 1 and 2 (EhEBP1 and EhEBP2), respectively. EhEBP1 and EhEBP2 contain RNA recognition motifs but no known DNA-binding domain, whereas URE3-BP contains EF hand motifs, suggesting a role for Ca^{2+}.[15,16] Indeed, in the presence of a high calcium concentration, URE3-BP was unable to bind DNA and was tethered to the amoebic plasma membrane by EhCA2 phospholipid-binding protein.[17,18] Transcriptome profiling of dominant-positive URE3-BP trophozoites showed that URE3-BP regulates the expression of genes that may play a role in trophozoite motility.[19] Amoebae expressing a dominant positive URE3-BP mutant presented a more elongated morphology, were more efficient to invade the intestinal epithelium, and produced larger liver abscesses.[20] Taken altogether, these data indicate that URE3-BP is likely a key transcription factor for the regulation of virulence genes in *E. histolytica*.

Transcriptional analysis of multidrug resistance *EhPgp1* and *EhPgp5* genes revealed the relevance of distinct DNA sequences in promoter regions. Notably, we found that two C/EBP-like motifs act as *cis*-activating elements for *EhPgp1* gene expression, through interaction with a nuclear C/EBP-like protein.[21] In another study based on genomic surveys of the parasite genome, 34 putative proteins that contain the conserved MYB DNA-binding domains with three imperfect conserved

repeats of 52 residues (R1, R2, and R3) were identified. Family I members, known as EhMybR2R3, contain only R2 and R3 repeats, and were differentially expressed in trophozoites grown in basal culture conditions. Notably, purified recombinant EhMyb10 protein was able to bind a consensus Myb recognition element found in gene promoters related to signal transduction, vesicular transport, heat shock response, and virulence. This suggests that the EhMyb10 transcription factor is involved in the regulation of genes participating in several different pathways in *E. histolytica*.[22] Proteins of family III, called EhMybSHAQKYF, only have a single repeat. The EhMyb-dr protein, which is the first developmentally regulated transcription factor identified in *E. histolytica*, binds a CCCCCC promoter motif present within 200 bp of the translation start site and regulates expression of a subset of amoebic stage-specific genes.[23]

Using a bioinformatic approach, Pearson et al. demonstrated that the AAACCTCAATGAAGA sequence is a conserved H_2O_2-responsive motif (HRM) that was specifically recognized by the nuclear HRM-binding protein (HRM-BP) transcription factor to upregulate gene expression following H_2O_2 exposure in *E. histolytica*.[24] On the other hand, *E. histolytica* has a nuclear high-mobility-group box chromosomal protein 1 (EhHMGB1) that binds and bends DNA. Its overexpression led to the modulation of 20 transcripts that were previously reported as regulated in an animal model of amoebiasis.[25] Notably, transcripts corresponding to the light subunit of the Gal/GalNAc lectin, an important virulence factor in *E. histolytica*, were upregulated, confirming the relevance of EhHMGB1 in modulating the virulence properties of trophozoites.[25]

PRE-mRNA 3′-END PROCESSING

In higher eukaryotes, pre-mRNA 3′-end processing is an essential step in mRNA biogenesis.[26] Pre-mRNA molecules synthesized in nucleus by RNA pol II undergo cotranscriptional processing before they can be exported and translated to proteins in cytoplasm. The first modification of nascent eukaryotic mRNA is the addition of a 7-methylguanosine at the 5′-end (capping) to protect mRNA from degradation by exonucleases; the cap structure also promotes 3′-end polyadenylation, splicing, nuclear export, and translation.[27] During the mRNA elongation step, introns (noncoding sequences) are removed from nascent pre-mRNA by splicing, which takes place on spliceosome.[28] Finally, when RNA pol II transcribes the 3′-untranslated region (3′-UTR), the recognition of specific sequence elements promotes the dissociation of RNA pol II from the DNA template to end transcription and induce the synthesis of a poly(A) tail at the pre-mRNA 3′-end.[5,6] These cotranscriptional modifications are intimately coupled because individual molecular machineries act in a coordinated network orchestrated by the RNA pol II-CTD.[29,30] Notably, the CTD heptapeptide repeats play a pivotal role to link mRNA synthesis and processing through their interactions with several factor of each molecular machineries, stimulating capping, splicing, and cleavage and polyadenylation reactions.[31] This functional integration between molecular machineries permits an efficient transcription termination and processing of nascent pre-mRNA.

Cis and *Trans*-Active Elements That Direct Pre-mRNA 3′-End Cleavage and Polyadenylation in Eukaryotic Cells

Polyadenylation greatly influences gene expression because poly(A) tail controls mRNA stability, export, localization, and translation.[32] Cleavage and polyadenylation of pre-mRNA depend on protein complexes interacting in a coordinated way with *cis*-sequences located within the mRNA 3′-UTR. These motifs are known as: (1) the cleavage and polyadenylation site [poly(A) site] represented by the CA dinucleotide in human; (2) the polyadenylation signal represented by the conserved AAUAAA hexanucleotide or variants, which is located 10–30 nt upstream the poly(A) site; and (3) the low conserved U/GU-rich element (UUUU, UGUA, or UAUA) typically found 10–30 nt downstream the poly(A) site.[33] The protein–protein interaction network of human pre-mRNA 3′-end processing complex has been recently reviewed.[26] The core machinery is composed of four complexes: the CPSF, the CstF, and cleavage factors (CFIm and CFIIm). Additional proteins are also required, including the poly(A) polymerase (PAP), symplekin, FIP1, and poly(A)-binding protein (PABP).[34] The pre-mRNA 3′-end processing occurs in a two-step process. In a simplified model, CPSF complex binds to the polyadenylation signal or variants through its 160 subunit (CPSF-160). CPSF-73 acts as the endonuclease responsible for the cleavage step. The CstF-64 binds to the U/GU-rich motif and is required for cleavage, but apparently not for polyadenylation. CPSF-160 and FIP1 proteins recruit PAP and tether it to RNA substrate to form the initial pre-mRNA processing complex; then, CFIm and CFIIm are recruited to the poly(A) site vicinity and RNA cleavage can occur. Notably, CFIm bound

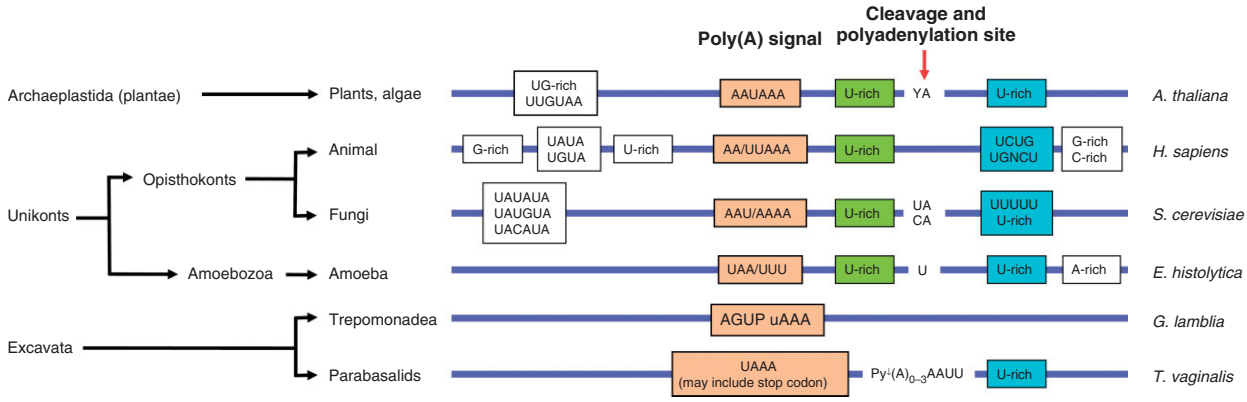


FIGURE 2 | Comparison of polyadenylation sequences among eukaryotic organisms. Left, a simplified version of the eukaryotic evolutionary tree. Right, pre-messenger RNA (mRNA) 3′-untranslated region (UTR) with polyadenylation sequences reported for a representative species of each arm.

to the UGUAA sequence participates in the selection of the poly(A) site.[35] Finally, PAP interacting with CPSF, FIP1, CFIm25, and PABP initiates the poly(A) tract synthesis.

Putative *Cis*-Elements for Pre-mRNA 3′-End Cleavage and Polyadenylation in *E. histolytica*

To identify polyadenylation sequences in *E. histolytica*, we performed a computational analysis of a large expressed sequence tag collection containing 2348 raw cDNA sequences obtained from the virulent *E. histolytica* HM1:IMSS strain.[36] Our results showed that pre-mRNA 3′-end processing motifs are represented by an AU-rich domain corresponding to the consensus UA(A/U)UU polyadenylation signal or variants, the poly(A) site denoted by the U residue and flanked by two U-rich tracts, and a novel A-rich element (Figure 2). Remarkably, *E. histolytica* polyadenylation signal and the poly(A) site are quite different to those described in other *Opisthokonts*. The amoeba UAAUU or UAUUU variants were present in 27 and 11% sequences, respectively. In addition, 13% sequences contain various UA(U/A)UU motifs, whereas 49% sequences do not contain UAAUU nor UAUUU motifs. Data obtained from genomic sequences confirmed that the AU-rich motif is located downstream the stop codon and upstream the poly(A) site. These findings were corroborated by a complementary analysis of mRNA 3′-end regions from *E. histolytica* genes with mapped poly(A) site. Interestingly, the polyadenylation signal was included into the translation stop codon UAA in several genes, which supports the hypothesis that stop codon may represent the polyadenylation signal ancestor.[37] In agreement with this view, it was recently reported that the UAAA tetranucleotide, which includes the UAA stop codon, corresponds to the polyadenylation signal in the related protozoan *Trichomonas vaginalis*.[38] The general organization of conserved motifs surrounding poly(A) site in *E. histolytica* was confirmed by Hon et al. by exploring a larger set of mRNA 3′-UTR ($n = 5018$).[39] Details about 3′-end processing signals in other protists have been recently reviewed by Clayton and Michaeli.[40] Notably, the molecular organization of polyadenylation sequences appears to be poorly conserved through evolutionary scale. Although very little is known in protozoan parasites, data obtained from *E. histolytica*, *Giardia*, *T. vaginalis*, *Plasmodium*, and *Toxoplasma* suggest that the polyadenylation signal seems to be species-specific in these organisms[36,40] (Figure 2).

E. histolytica Contains Conserved Polyadenylation Factors

By search in the *E. histolytica* genome, we identified 16 components of the putative cleavage and polyadenylation machinery[41,42] (Table 1). These include: the four subunits of CPSF (160, 100, 73, and 30 kDa), FIP1 and PAP homologs, the three subunits of CstF (77, 64, and 50 kDa), the two subunits of CFIIm (ClP1 and PCF11), and only the 25-kDa subunit of CFIm, as well as additional factors such as Mpe1, Psf2, Ssu72, and PC4. However, genes coding for symplekin, CFIm59, and CFIm68 proteins were not found. Intriguingly, protein sequences of *E. histolytica* pre-mRNA processing factors are well conserved, although the putative polyadenylation sequences that they may recognize are quite different to those described in human and yeast. *E. histolytica* seems to be the protozoan parasite with the largest polyadenylation apparatus identified so far. Indeed, *Giardia* has PAP, CPSF-73, CPSF-30, and PABP,

TABLE 1 | Summary of the Factors Involved in mRNA Polyadenylation in *Entamoeba histolytica*

Protein	Accession Number[1]	ORF Name[2]	Reference
Cleavage and polyadenylation			
CPSF-160	C4M386	EHI_106110	López-Camarillo et al.[41]
CPSF-100	C4M6Y0	EHI_033130	
CPSFF-73	C4M297	EHI_136170	
CPSF-30	C4M297	EHI_067580	
PAP	Q51D88	EHI_012040	
FIP1	C4M765	EHI_083400	
CstF-77	B1N2W3	EHI_098370	
CstF-64	C4LSN8	EHI_151990	
CstF-50	C4LSW0	EHI_152770	
CFIm-25	C4M2T1	EHI_077110	
CIP1	C4LYE5	EHI_008100	
PCF11	C4LTC6	EHI_045130	
Psf2	C4M1D0	EHI_170080	
Mpe1	C4LXP1	EHI_014000	
Ssu72	C4M1T3	EHI_027340	
PC4/Sub1	C4M1H2	EHI_192520	
PABP1	C4LWS1	EHI_198750	
PABP2	C4M6Y2	EHI_033250	
Cytoplasmic polyadenylation			
GLD-2a	C4LV10	EHI_092480	López-Camarillo et al.[43]
GLD-2b	C4LVL9	EHI_103770	
GLD-2c	C4M7J7	EHI_170350	
TRF4-a	C4M9B4	EHI_003910	
TRF4-b	C4LXS6	EHI_087330	
TRF4-c	C4MBH4	EHI_073170	

[1]UniProt Knowledgebase.
[2]Pathema-Entamoeba.

whereas *Trypanosoma* has the four subunits of CPSF, FIP1, two PAPs, CstF-50, CFIm25, CIP1, and two PABPs.[40] Several *E. histolytica* proteins have been cloned and characterized in an effort to understand the 3′-end processing in amoebozoa.

PAP Family in *E. histolytica*

The EhPAP, which is thought to synthesize the poly(A) tail at pre-mRNA 3′-end in *E. histolytica*, is a 522-amino acid polypeptide, which shares 25–32% identity and 43–52% similarity with nuclear PAPs from other organisms.[44] The conserved PAP central catalytic domain contains three invariant aspartate (D) residues involved in nucleotide transfer and the F/YGS motif responsible for ATP binding in eukaryotic PAPs, which indicates that EhPAP belongs to the polymerases-like superfamily of nucleotidyl transferases. In the three-dimensional (3D) model, both functional domains fold as a U-shaped structure that likely orients the incoming ATP and RNA molecules for poly(A) tail extension.[43] Although it has a divergent RNA-binding domain (RBD), the purified recombinant EhPAP was able to form complexes with the *EhPgp5* gene 3′-UTR. *EhPap* mRNA expression was modulated through cell cycle progression, with a major expression in G1 and S phases. EhPAP was immunodetected in nuclear and cytoplasmic extracts of trophozoites. Notably, inmunofluorescence and microscopy laser confocal assays revealed that EhPAP is located in punctuate nuclear *foci* and in cytoplasmic dots, as it has been described for several factors involved in mRNA polyadenylation in other organisms.[44]

In addition to the canonical EhPAP, *E. histolytica* also has six unconventional PAPs (EhGLD-2a, EhGLD-2b, EhGLD-2c, EhTRF4-a, EhTRF4-b, and EhTRF4-c), which do not share any structural and evolutionary relationship with the nuclear PAP[43] (Table 1). EhTRF4-a is highly related to EhTRF4-b, EhTRF4-c, EhGLD-2a, and EhGLD-2c, but it presents less similarity with EhGLD-2b. Proteins from *E. histolytica*, other eukaryotic cells, and bacteria are all related to the minimal nucleotidyl transferase (MNT) domain that is characteristics of the polymerase-like superfamily of nucleotidyl transferases. The evolution of this superfamily may have included bursts of rapid divergence linked with the emergence of new functions as well as a number of horizontal gene transfer events.[48] Our preliminary assays evidenced that the *EhTRF4*-a gene is highly transcribed in basal culture conditions in comparison with the other genes. The functional characterization of unconventional PAPs currently in progress will help us to understand their relevance in *E. histolytica* gene expression regulation.

Functional Aspects of EhCFIm25 Cleavage Factor

The *EhCFIm25* gene is an intron-less 768-bp sequence that encodes a 256-amino acid polypeptide, which shares 27–35% identity with homologous proteins from plants to human, including other protozoan parasites. As the human homolog, the predicted *E. histolytica* CFIm25 protein is an unconventional Nudix protein.[49] Notably, three of the four conserved E residues of the Nudix consensus sequence are replaced by K residues, and the last G residue is substituted by the hydrophilic S residue. However,

TABLE 2 | Summary of the Factors Involved in mRNA Decay in *Entamoeba histolytica*

Protein	Accession Number[1]	ORF Name[2]	Reference
Decapping			
DCP2	C4M5G6	EHI_058810	López-Rosas et al.[45]
Lsm1	B1N3A8	EHI_188020	
Lsm2	C4LU49	EHI_068580	
Lsm3	C4LSH4	EHI_151310	
Lsm4	C4LUD9	EHI_049370	
Lsm6	C4M187	EHI_188130	
Lsm7	C4M939	EHI_025840	
edc3	C4LWU0	EHI_198940	
dhh1	C4LYI1	EHI_093900	
XRN2	C4MB40	EHI_133330	
Deadenylation			
NOT1	C4M3Y6	EHI_178720	López-Rosas et al., unpublished data
NOT2	C4M790	EHI_044180	
NOT3	C4M7Y3	EHI_119550	
NOT4	C4M9N9	EHI_080710	
CAF1	Q56AY2	EHI_048150	López-Rosas et al.[45]
CAF1-like	C4LZS1	EHI_039000	López-Rosas et al., unpublished data
Exosome			
RRP4-	C4M3T2	EHI_163510	López-Rosas et al., unpublished data
RRP6	C4M054	EHI_021400	
RRP40-	C4M073	EHI_004770	
RRP41-	C4M2G1	EHI_040320	
RRP42-	C4LZA9	EHI_000580	
RRP43-	C4M182	EHI_188080	
RRP46-	C4M8Y9	EHI_086520	
MTR3-	C4LW17	EHI_126330	
DIS3-	C4MAJ9	EHI_160720	
RNA interference			
AGO2-1	C4LVQ2	EHI_186850	Zhang et al.[46]
AGO2-2	C4LVV2	EHI_125650	
AGO2-3	C4LY31	EHI_177170	
RdRP	C4M6W7/	EHI_139420	
Dicer-like	C4LU64	EHI_068740	Abed and Ankri[47]

[1]UniProt Knowledgebase.
[2]Pathema-Entamoeba.

these changes affect neither the classical $\alpha/\beta/\alpha$ sandwich Nudix fold nor the RNA 3′-UTR-binding capacity of the recombinant protein. This suggests that, as the human homolog, EhCFIm25 is not able to cleave RNA.[50] In agreement with its possible role in nuclear processes, the endogenous EhCFIm25 was immunodetected in protein nuclear extracts, although it was also found in cytoplasmic extracts. Intriguingly, its apparent molecular weight was higher than expected, which suggests that EhCFIm25 interacts with itself or other nuclear proteins to form a stable dimer through interwoven structures and strong residues association. Although it does not have a classical RBD, EhCFIm25 was able to form specific RNA–protein complexes with the *EhPgp5* mRNA 3′-UTR used as probe, in the absence of any other mRNA 3′-end processing factors. This interaction could be through the extended loop structure at the N-terminal region, as it has been described for the human factor.[51,52] Multiple alignment of CFIm25 proteins from diverse organisms suggested that the two amino acid residues that bind RNA in the human CFIm25 protein are conserved in all sequences. Interestingly, L_{135} and Y_{217} amino acid residues in EhCFIm25 protein are in close proximity with RNA in the predicted 3D model of EhCFIm25 bound to RNA (Figure 3) (Ospina-Villa et al., our unpublished data), which is in agreement with their possible role in RNA binding. Interestingly, Far-Western experiments and pull-down assays followed by electrospray ionization (ESI)-mass spectrometry (MS)/MS tandem mass spectrometry evidenced that EhCFIm25 is able to interact with EhPAP, another member of the pre-mRNA 3′-end processing machinery in *E. histolytica*. In human, PAP is recruited to the RNA molecule at the beginning of transcription.[1] Then, it is possible that EhCFIm25 may be introduced into the processing complex in the early steps of the cleavage/polyadenylation reaction through its interaction with EhPAP.

Stochastic Noise of Polyadenylation in *E. histolytica*

Alternative splicing and polyadenylation, which generate multiple isoforms from a single pre-mRNA, are the major mechanisms for expanding the diversity of transcriptome and proteome in response to physiological conditions.[53,54] However, these alternative isoforms might also result from aberrant rather than regulated RNA processing events. In order to quantify the extent of stochastic noise in splicing and polyadenylation in *E. histolytica*, Hon et al. sequenced the poly(A)$^+$ transcriptome and compared it to the genome sequence.[39] Results showed that most identified poly(A) sites are genuine and fall within 100 nt downstream the stop codon. They also observed the presence of alternative sites located within ±5 nt to the corresponding constitutive site, which is correlated with the occurrence of

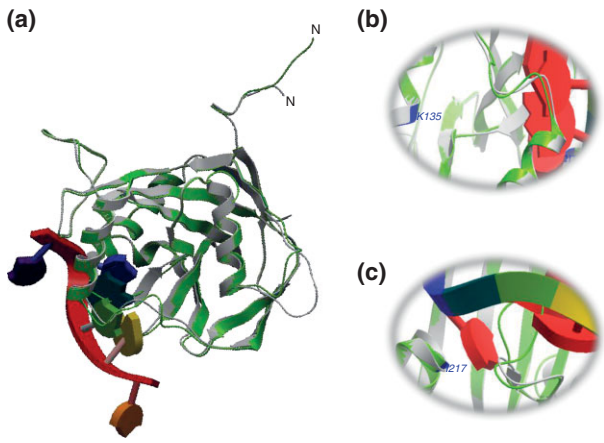


FIGURE 3 | Prediction of amino acid residues of EhCFIm25 that could interact with RNA. (a) Overlapping of predicted three-dimensional model of EhCFIm25 (green) with crystallographic structure of human CFIm25 (gray). (b and c) Magnification showing the proximity of L135 (b) and Y217 (c) amino acid residues in EhCFIm25 with RNA. RNA template sequence is denoted in red.

constitutive cleavage events. On the basis of these observations, the authors concluded that most of such microheterogeneity is likely to be related to stochastic noise of constitutive cleavage events. In summary, it seems that the functional impact of alternative polyadenylation events is limited to a small proportion of genes in *E. histolytica*.[39]

PATHWAYS OF mRNA DEGRADATION

In mammals, transcripts synthesized by RNA pol II are distinguished from those made by RNA pol I and III by the presence of a 5′-end cap structure and a 3′-end poly(A) tail that act as mRNA stability determinants. mRNA molecules interact with cytoplasmic eIF4E and PABP to protect transcripts from exonuclease attack and stimulate translation. After several rounds of translation in the cytoplasm, most mRNAs undergo degradation by the deadenylation-dependent pathway,[55,56] which is initiated by poly(A) tail shortening by the deadenylases CAF1 and CCR4/NOT1–5 complex or by PARN. Transcripts can follow one of two alternative decay pathways: (1) 5′-end decapping by DCP1–DCP2 complex, which is stimulated by Lsm1–7 proteins, followed by $5' \rightarrow 3'$ digestion by exonuclease XRN1 or (2) $3' \rightarrow 5'$ degradation by exosome complex. In the latter case, the scavenger-decapping DCPs enzyme hydrolyzes the remaining cap structure. Alterations of any regulatory elements participating in these reactions can modify both mRNA stability and deadenylation rate.[57]

Decapping and Deadenylation Proteins in *E. histolytica*

Nearly all genes involved in mRNA decapping and deadenylation have been identified in *E. histolytica* genome, indicating that mRNA degradation machineries are well conserved in this parasite[45] (Figure 4). *Ehdcp2*, *Ehxrn2*, *Ehlsm1–6* (with the exception of *Ehlsm5* gene), *Ehedc3*, and *Ehdhh1* decapping-associated genes are present in amoeba (Table 2). Intriguingly, homologs for major DCP1 decapping and XRN1 exoribonuclease are missing, which indicates that mRNA degradation mechanisms may occur with a minimal set of proteins in lower eukaryotic cells. Interestingly, both EhXRN2 and EhDCP2 are localized diffusely in cytoplasmic foci, and they colocalized in discrete cytoplasmic structures resembling P-bodies. Gene expression studies showed that *Ehdcp2* and *Ehxrn2* gene expression was differentially modulated in trophozoites under stress conditions. After heat shock, *Ehxrn2* gene expression was 1.5- to 12-fold upregulated, whereas the *Ehdcp2* was two-fold repressed.[45] *Ehdcp2* expression was also significantly repressed after nitrosative stress and DNA damage induced by UV irradiation. This suggests that DNA/RNA damage events might induce the repression of pre-mRNA processing mechanisms to avoid the generation of damaged capped transcripts. Similar findings evidencing a transient repression of 3′-end processing after DNA/RNA damage have been reported in other organisms.[58]

On the other hand, *in silico* screening of *E. histolytica* genome databases revealed the presence of two intron-less genes encoding CAF1 and CAF1-like proteins, which share 83% identity, but intriguingly no homologs for CCR4, PARN, PAN2, or PAN3 deadenylases were identified (Lopez-Rosas et al., our unpublished data). In addition, *E. histolytica* genome contains only four *NOT* genes, suggesting the existence of a divergent CAF1/NOT complex lacking CCR4 and NOT5 subunit (Figure 4 and Table 2). These data indicate that *E. histolytica* has a reduced number of genes involved in mRNA deadenylation. Another possibility is that orthologs are hard to find using bioinformatics tools.

Functional Insights of EhCAF1 Deadenylase

We have recently performed the functional characterization of EhCAF1, which seems to be the major deadenylase in *E. histolytica*. Using biochemical analysis, we showed that the recombinant EhCAF1 protein exhibits RNA binding and deadenylase activities *in vitro* (Lopez-Rosas et al., our unpublished data). Interestingly, immunofluorescence assays and confocal

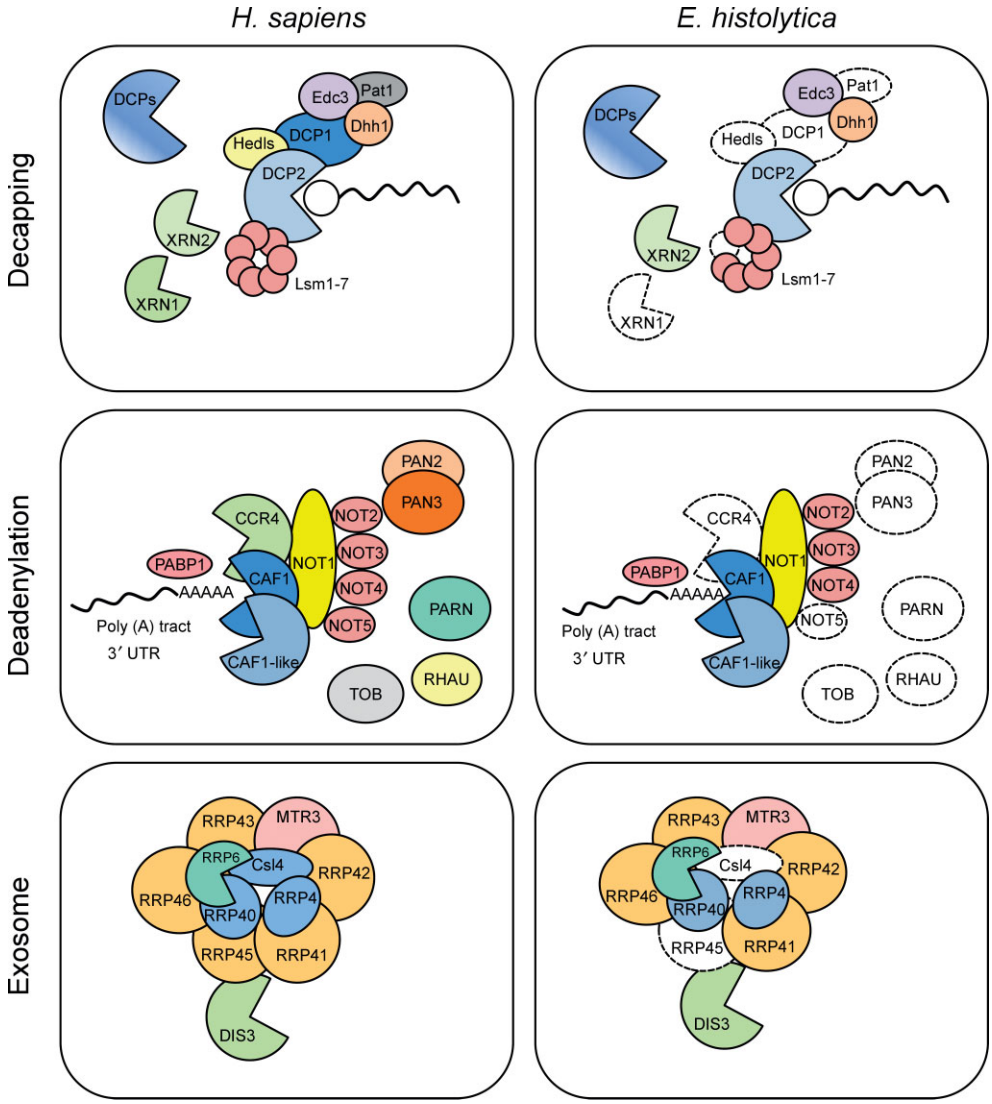


FIGURE 4 | Comparison of the major messenger RNA (mRNA) degradation pathways in *Entamoeba histolytica* and *Homo sapiens*. Upper panel: Decapping (5′ → 3′ decay) proteins. *H. sapiens* has the decapping enzymes DCP1, DCP2, DCP-scavenger (DCPs), and EDC1–3, DHH1, PAT1, Hedls, and Lsm1–7-associated proteins. Of these, *E. histolytica* lacks DCP1, XRN1, Pat1, Hedls, and Lsm5 protein-encoding genes. Middle panel: Deadenylation (3′ → 5′ decay). *E. histolytica* contains an incomplete CCR4–CAF1–NOT complex. Amoeba has genes for CAF1, CAF1-like, NOT1–4, and PABP proteins, whereas CCR4, NOT5, PARN, and PAN2-PAN3 deadenylases are lacking, as well as the accessory TOB and RHAU proteins. Lower panel: Exosome (3′ → 5′ decay): *E. histolytica* contains almost all genes described for human exosome, but it lacks the RRP45 and Csl4 protein-encoding genes. Geometric shapes with continuous outline correspond to proteins present in both *H. sapiens* and *E. histolytica*. Geometric shapes with dotted outline indicate proteins that are absent in *E. histolytica*.

microscopy evidenced the coimmunolocalization of EhCAF1 with EhXRN2 and EhDCP2 decapping enzymes in P-bodies-like structures.[45] In addition, using pull-down experiments and ESI-MS/MS tandem mass spectrometry, 10 EhCAF1-interacting proteins were identified. Notably, one protein corresponds to the endoribonuclease L-PSP of bacterial origin. Using specific antibodies, L-PSP protein was colocalized with the exosome EhRrp41 ribonuclease (see below), linking deadenylation and exosome machineries in this parasite. These data argue for the existence of a complex containing the EhCAF1 deadenylase, a novel L-PSP endoribonuclease, and the exosome EhRrp41 exoribonuclease in *E. histolytica* (López-Rosas et al., our unpublished data).

Exosome Proteins in *E. histolytica*

The exosome complex, which is responsible for the 3′ → 5′ degradation of cytoplasmic deadenylated

transcripts, is a key player in mRNA turnover. In the nucleus, exosome proteins are also involved in several RNA processing reactions, including 3′-end trimming of rRNA, snRNA, and snoRNA, as well as mRNA surveillance and degradation of cryptic unstable transcripts.[59] Both the nuclear and cytoplasmic exosome complexes consist of a nine-subunit ring-shaped structure, which associates with Dis3 (Rrp44), the catalytic subunit responsible for both $3' \rightarrow 5'$ exonucleolytic and endonucleolytic activities in yeast and human.[60] A genomic survey revealed that *E. histolytica* contains seven genes encoding Rrp6, Rrp41, Rrp43, Rrp46, Mtr3-Rrp42, and the catalytic subunit Dis3 proteins, as well as accessory stabilizing Rrp4 and Rrp40 proteins, but it lacks *Rrp45* and *Csl4* genes (Figure 4 and Table 2). As described above, the colocalization of EhRrp41 with EhCAF1 and EhL-PSP in trophozoites (Lopez-Rosas et al., our unpublished data) suggests the existence of cooperative interactions between mRNA decay machineries in *E. histolytica*. Functional assays currently in progress will confirm the activity of these enzymes.

P-Bodies as Centers of mRNA Degradation in *E. histolytica*

In higher eukaryotes, mRNA degradation and RNA-based gene silencing occur in specialized cytoplasmic *foci* variously referred to as mRNP granules, mRNA-decay *foci*, GW182-bodies, and mRNA processing bodies (P-bodies). These structures are enriched in RNA substrates and mRNA turnover proteins, and represent mRNA processing centers in which nontranslating transcripts are sorted and either silenced or degraded.[61] However, data about the presence of P-bodies and their role in mRNA decay are scarce in protozoan parasites. Recently, we reported experimental evidence for P-bodies in *E. histolytica*.[45] EhXRN2 exoribonuclease and EhDCP2 decapping protein, which represent P-bodies markers in eukaryotic organisms, were localized in cytoplasmic *foci* in a pattern resembling P-body organization. Additional factors linked to mRNA decay, including the EhCAF1 deadenylase and the EhAGO2-2 protein involved in RNA interference (RNAi), were also localized to cytoplasmic *foci*. As these proteins are involved in different mRNA decay events, it was proposed that these cytoplasmic *foci* represent sites of mRNA decay in *E. histolytica*. Interestingly, we found that EhCAF1 coimmunoprecipitated with EhXRN2 but not with EhDCP2 or EhAGO2-2, which provides a novel link between deadenylation and $5' \rightarrow 3'$ mRNA degradation. In addition, several features of P-bodies were found in *E. histolytica*. For instance, the number of EhCAF1-containing *foci* significantly decreased after inhibition of transcription and translation. It was also demonstrated that EhCAF1 protein colocalized with poly(A)$^+$ RNA, and during silencing of the *Ehpc4* gene by RNAi, EhAGO2-2 colocalized with small interfering RNAs (siRNAs) in cytoplasmic *foci*. The presence of decapping, deadenylation, and RNAi proteins within P-body-like structures in the low eukaryote *E. histolytica* suggests that the organization of mRNA decay proteins in well-defined cytoplasmic structures has been conserved from the early evolution of eukaryotic lineage. Cytoplasmic *foci* resembling P-bodies have also been described in the protozoan parasite *Trypanosoma*; however, their potential roles in mRNA degradation events have not been comprehensively addressed.[62,63] No data are available about the existence of other RNA-containing cytoplasmic *foci*, such as stress granules and temporal asymmetric bodies, in *E. histolytica*.

RNA DEGRADATION IN RNA-BASED GENE SILENCING

mRNA degradation also represents a powerful gene silencing mechanisms in eukaryotic cells. RNAi was first described in *Caenorhabditis elegans* as a potent and specific gene interference phenomena caused by double-stranded RNA (dsRNA) formation.[64] The general mechanism initiates with the introduction of a dsRNA whose sequence matches a specific gene target; then dsRNA is cleaved by the cytoplasmic Dicer endonuclease into siRNAs of 21–23 nucleotides.[65] siRNAs are then recognized by Argonaute (AGO) protein into RNA-induced silencing complex (RISC). The siRNA acts as a guide to direct the multiprotein complex to its transcript target, which is cleaved by AGO resulting in an effective gene knockdown. RNAi machinery has been found to be widely conserved in eukaryotic cells. In addition to classical primary siRNAs generated by Dicer-mediated endonucleolytic cleavage, there are also small RNAs produced by a Dicer-independent mechanism, such as Piwi-interacting RNA (piRNA), secondary siRNA [synthesized by an RNA-directed RNA polymerase (RdRP)], and primal small RNA (priRNA).[66,67]

Another cellular mechanism that uses RISC proteins to achieve RNA-based gene silencing is represented by the microRNA (miRNA) pathway. miRNAs are a class of evolutionary conserved noncoding small RNAs that act as negative regulators of gene expression.[68] They are transcribed by RNA Pol II as hairpin-shaped primary-miRNAs (pri-miRNA), which are processed by the microprocessor complex

that includes the RNase III-type endonuclease Drosha and the dsRNA-binding protein DGCR8. The resulting precursor miRNAs (pre-miRNA) are exported to cytoplasm by exportin 5. Then they are cleaved by cytoplasmic Dicer into mature 22-nucleotides RNAs and recruited by RISC to function as guide molecules to target mRNA. Imperfect binding of mature miRNAs to 3′-UTR of transcripts induces translation repression or mRNA degradation.

Current Knowledge of RNA-Based Gene Silencing in *E. histolytica*

In *E. histolytica*, it has been shown that dsRNA, siRNA, and short-hairpin RNAs are effective RNA tools for RNA-based gene silencing, suggesting that the machinery for small RNA-mediated silencing is functional in this protozoan.[69,70] Indeed, several genes of the RNAi pathway have been identified in the parasite genome, including three *Ago2*-related genes (*Ago2-1*, *Ago2-2*, and *Ago 2-3*) that encode proteins with both PAZ and Piwi domains, as well as one *RdRP* gene and an additional gene that encodes a protein with an incomplete RdRP domain.[46] Notably, *E. histolytica* lacks a canonical *Dicer* gene, but contains a *Dicer*-like gene that encodes a protein containing a single RNase III domain[47] (Table 2).

On the other hand, Zhang et al. described a population of small 27-nt RNAs, which have an unusual 5′-polyphosphate structure similar to those generated by the Dicer-independent mechanism in *C. elegans*.[66] In addition, most of these 27-nt small RNAs associate with an *E. histolytica* Piwi-related protein and map antisense to genes. Most small RNAs mapped antisense to genes (36%), others map to intergenic regions (25%), as mixed hits (10%), sense to genes (18%), to repetitive regions (5%), and only a small number (0.32%) map to retrotransposons elements. Interestingly, experimental data from microarray analysis, and Northern blot assays strongly suggested that these small RNAs are able to regulate the expression of their cognate genes in *E. histolytica*.[71] These data are indicative of an amplified silencing mechanism similar to that previously described in *C. elegans*.

Intriguingly, neither Drosha nor DGC8 homologs have been found in *E. histolytica* genome, suggesting the absence of the miRNA pathway. However, two exportin-like predicted proteins were identified in this parasite.[46] In addition, miRNAs have been computationally identified in *E. histolytica*.[72] However, their biological significance remains to be demonstrated.

DEAD/DEXH-BOX RNA HELICASES AS ESSENTIAL FACTORS FOR mRNA METABOLISM

mRNA metabolism frequently requires the participation of RNA helicases to unwind dsRNA molecules. Most RNA helicases belong to DEAD-box and DExH-box families of the SFII superfamily, which is characterized by the presence of eight conserved functional domains that are essential for ATP hydrolysis, RNA-binding, and unwinding activities.[73,74]

The DEAD/DExH-Box RNA Helicase Family in *E. histolytica*

Because of the relevance of RNA helicases for the proper realization of molecular events in mRNA metabolism, we initiated the study of these proteins in *E. histolytica*.[75] Using bioinformatics approaches, we reported that *E. histolytica* genome contains 20 *EhDead* and 13 *EhDexh-box* genes, in agreement with the large RNA helicase families previously described in *Leishmania major* and *Plasmodium falciparum* parasites (Table S2). The helicase domain of all predicted EhDEAD-box proteins contains the conserved DEAD (Asp-Glu-Ala-Asp) motif II and the seven additional functional motifs (I, Ia, Ib, III, IV, V, and VI) that are required for ATP hydrolysis, RNA-binding, and unwinding activities.[46] Most EhDEAD-box proteins also possess the conserved aromatic F residue upstream of motif I and the Q motif with six residues including the invariant Q, which forms a loop–helix–loop structure involved in the interaction with motif I and ATP.[76] Similarly, most EhDExH-box helicases have the conserved helicase domain with the DExH (Asp-Glu-x-His) motif II. Several EhDEAD/EhDExH proteins share high sequence similarity with previously characterized helicases, which can help to hypothesize as to their function in different events related to RNA in *E. histolytica*. For example, EhDEAD1 is the *E. histolytica* homolog of human DDX3X and DDX3Y, yeast DED1, and *Xenopus laevis* AN3 proteins that are involved in mRNA translation, mRNA export, and cell cycle regulation, whereas EhDExH1 and EhDExH10 are homologs of the U5snRNP-200-kDa factor that participates in pre-mRNA splicing. Based on phylogenetic analysis, the absence of specific motifs, and the presence of introns, EhDEAD and EhDExH-box helicases can be clustered into distinct subfamilies, in which gene duplication, mutations, and noncoding DNA sequence addition events could have generated the different proteins. In addition, gene fusion events have led to the formation of

multidomains in EhDExH1 and EhDExH10 proteins, which probably increase their catalytic activity. Interestingly, BLAST analysis suggested that gene fusion is an evolutionary conserved event in the eukaryotic lineage of EhDExH1 orthologs, probably as a result of horizontal gene transfer between bacteria and eukaryotes.[75]

Modulation of RNA Helicases Expression During Amoeba Infection

Each RNA helicase may be involved in a specific cellular process.[75] Indeed, several works have shown that RNA helicase expression is related to virulence and infection mechanisms. Thus, transcripts corresponding to *EhDead5*, *EhDead12*, *EhDead13*, *EhDead14*, *EhDead16*, *EhDead18*, *EhDead19*, and *EhDead20* genes were identified in a cDNA library obtained from the highly virulent *E. histolytica* HM1:IMSS strain.[77] Transcription of *EhDead5* and *EhDead16* genes was downregulated during invasion of intestinal cecum in mouse.[78] Moreover, proteomics studies revealed that EhDEAD16 was associated to phagosome fractions.[79] Therefore, the expression of specific RNA helicases might allow trophozoites to regulate mRNA metabolism during infection process.

Functional Characterization of EhDEAD1

To date, EhDEAD1 is the only *E. histolytica* RNA helicase that has been functionally characterized.[80] EhDEAD1 is a 48-kDa protein that exhibits RNA-dependent ATPase activity *in vitro*, even when it has an incomplete Q motif. ATPase activity is lower in the presence of *E. histolytica* poly(A)$^+$ RNA, which suggests that EhDEAD1 function could be more related to tRNA or rRNA than to mRNA molecules. It is also possible that additional RNA-binding proteins could provide RNA specificity to EhDEAD1. As expected for an RNA helicase, EhDEAD1 is able to bind and unwind heteroduplex RNA molecules with 5′ overhangs in an ATPase-dependent manner. As several homologs, it also exhibits a potential single-stranded RNA reannealing activity, probably through the characteristic clusters of arginine and glycine residues at the C-terminus. *EhDead1* gene is mainly transcribed in cell cycle S phase, suggesting that EhDEAD1 might be involved in S phase and/or posterior cell cycle steps or boundaries. This hypothesis is in agreement with the observation that S to G2/M transition seems to be facilitated when *EhDead1* gene expression was inhibited by antisense RNA. Intriguingly, the ectopic expression of EhDEAD1 is unable to rescue a DED1 defective yeast, indicating that these homologous proteins are not functionally conserved in spite of their high sequence homology.[80]

CONCLUSION

Over the past decade, the knowledge of mRNA synthesis, 3′-end processing, and decay mechanisms, mainly in human and yeast, has increased significantly and these events are now considered as key players in gene expression regulation. However, very little is known about these molecular events in protozoan parasites. Initial efforts in the study of mRNA life cycle in *E. histolytica* have revealed that these canonical mechanisms are generally conserved in this primitive eukaryote, although subtle differences were evidenced. Notably, the transcription apparatus of *E. histolytica* shares several features with superior organisms, but the existence of a noncanonical CTD in EhRNA Pol II suggests that protein–protein interactions regulating the coordination of the different events could be distinct in this human pathogen. Genome analysis showed the presence of most polyadenylation factors, whereas several proteins involved in mRNA degradation are missing. Preliminary data about RNA-based gene silencing indicate that RNAi and miRNA pathways might be partially functional. We now have a general idea of how mRNA synthesis, 3′-end processing, and degradation may occur in *E. histolytica*, but the functional characterization of the main molecular players remains to be addressed. Future works focused on the biochemical dissection of mRNA processing and decay may help in the understanding of post-transcriptional gene regulation in deep-branching eukaryote.

REFERENCES

1. Mandel C, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. *Cell Mol Life Sci* 2008, 65:1099–1122. doi: 10.1007/s00018-007-7474-3.
2. Jackson TF. Epidemiology. In: Ravdin JI, ed. *Amebiasis*. London: Imperial College Press; 2000, 47–63.
3. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 2005, 433:865–868. doi: 10.1038/nature03291.

4. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. *Annu Rev Biochem* 2003, 72:449–479. doi: 10.1146/annurev.biochem.72.121801.161520.
5. Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* 2003, 17:1321–1327. doi: 10.1101/gad.1093603.
6. He X, Khan AU, Cheng H, Pappas DL Jr, Hampsey M, Moore CL. Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. *Genes Dev* 2003, 17:1030–1042. doi: 10.1101/gad.1075203.
7. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res* 2011, 39:D612–D619. doi: 10.1093/nar/gkq1006.
8. Lioutas C, Tannich E. Transcription of protein-coding genes in *Entamoeba histolytica* is insensitive to high concentrations of α-amanitin. *Mol Biochem Parasitol* 1995, 73:259–261.
9. Guo Z, Stiller JW. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics* 2004, 45:69. doi: 10.1186/1471-2164-5-69.
10. Das A, Bellofatto V. The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in *Trypanosoma brucei*. *PLoS One* 2009, 4:e6959. doi: 10.1371/journal.pone.0006959.
11. Singh U, Rogers JB. The novel core promoter element GAAC in the hgl5 gene of *Entamoeba histolytica* is able to direct a transcription start site independent of TATA or initiator regions. *J Biol Chem* 1998, 273:21663–21668. doi: 10.1074/jbc.273.34.21663.
12. de Dios-Bravo G, Luna-Arias JP, Riveron AM, Olivares-Trejo JJ, Lopez-Camarillo C, Orozco E. *Entamoeba histolytica* TATA-box binding protein binds to different TATA variants in vitro. *FEBS J* 2005, 272:1354–1366. doi: 10.1111/j.1742-4658.2005.04566.x.
13. Castanon-Sanchez CA, Luna-Arias JP, de Dios-Bravo MG, Herrera-Aguirre ME, Olivares-Trejo JJ, Orozco E, Hernandez JM. *Entamoeba histolytica*: a unicellular organism containing two active genes encoding for members of the TBP family. *Protein Expr Purif* 2010, 70:48–59. doi: 10.1016/j.pep.2009.12.007.
14. Purdy JE, Pho LT, Mann BJ, Petri WA Jr. Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol Biochem Parasitol* 1996, 78:91–103.
15. Gilchrist CA, Holm CF, Hughes MA, Schaenman JM, Mann BJ, Petri WA Jr. Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J Biol Chem* 2001, 276:11838–11843. doi: 10.1074/jbc.M007375200.
16. Schaenman JM, Gilchrist CA, Mann BJ, Petri WA Jr. Identification of two *Entamoeba histolytica* sequence-specific URE4 enhancer-binding proteins with homology to the RNA-binding motif RRM. *J Biol Chem* 2001, 276:1602–1609. doi: 10.1074/jbc.M006866200.
17. Gilchrist CA, Leo M, Line CG, Mann BJ, Petri WA Jr. Calcium modulates promoter occupancy by the *Entamoeba histolytica* Ca^{2+}-binding transcription factor URE3-BP. *J Biol Chem* 2003, 278:4646–4653. doi: 10.1074/jbc.M211271200.
18. Moreno H, Linford AS, Gilchrist CA, Petri WA Jr. Phospholipid-binding protein EhC2A mediates calcium-dependent translocation of transcription factor URE3-BP to the plasma membrane of *Entamoeba histolytica*. *Eukaryot Cell* 2010, 9:695–704. doi: 10.1128/EC.00346-09.
19. Gilchrist CA, Baba DJ, Zhang Y, Crasta O, Evans C, Caler E, Sobral BW, Bousquet CB, Leo M, Hochreiter A, et al. Targets of the *Entamoeba histolytica* transcription factor URE3-BP. *PLoS Negl Trop Dis* 2008, 2:e282. doi: 10.1371/journal.pntd.0000282.
20. Gilchrist CA, Moore ES, Zhang Y, Bousquet CB, Lannigan JA, Mann BJ, Petri WA Jr. Regulation of virulence of *Entamoeba histolytica* by the URE3-BP transcription factor. *MBio* 2010, 1:e00057–10. doi: 10.1128/mBio.00057-10.
21. Marchat LA, Gomez C, Perez DG, Paz F, Mendoza L, Orozco E. Two CCAAT/enhancer binding protein sites are cis-activator elements of the *Entamoeba histolytica EhPgp1* (*mdr*-like) gene expression. *Cell Microbiol* 2002, 4:725–737. doi: 10.1046/j.1462-5822.2002.00220.x.
22. Meneses E, Cárdenas H, Zárate S, Brieba LG, Orozco E, López-Camarillo C, Azuara-Liceaga E. The R2R3 Myb protein family in *Entamoeba histolytica*. *Gene* 2010, 455:32–42. doi: 10.1016/j.gene.2010.02.004.
23. Ehrenkaufer GM, Hackney JA, Singh U. A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in *Entamoeba histolytica*. *Cell Microbiol* 2009, 11:898–910. doi: 10.1111/j.1462-5822.2009.01300.x.
24. Pearson RJ, Morf L, Singh U. Regulation of H_2O_2 stress-responsive genes through a novel transcription factor in the protozoan pathogen *Entamoeba histolytica*. *J Biol Chem* 2013, 288:4462–4474. doi: 10.1074/jbc.M112.423467.
25. Abhyankar MM, Hochreiter AE, Hershey J, Evans C, Zhang Y, Crasta O, Sobral BW, Mann BJ, Petri WA Jr, Gilchrist CA. Characterization of an *Entamoeba histolytica* high-mobility-group box protein induced during intestinal infection. *Eukaryot Cell* 2008, 7:1565–1572. doi: 10.1128/EC.00123-08.
26. Chan S, Choi EA, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* 2011, 2:321–335. doi: 10.1002/wrna.54.
27. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and

biological targets. *Cell* 2009, 136:731–745. doi: 10.1016/j.cell.2009.01.042.

28. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009, 136:701–718. doi: 10.1016/j.cell.2009.02.009.
29. Proudfoot N. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 2004, 16:272–278.
30. Rigo F, Martinson HG. Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. *Mol Cell Biol* 2008, 28:849–862. doi: 10.1128/MCB.01410-07.
31. Jasnovidova O, Stefl R. The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip Rev RNA* 2013, 4:1–16. doi: 10.1002/wrna.1138.
32. Eckmann CR, Rammelt C, Wahle E. Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* 2011, 2:348–361. doi: 10.1002/wrna.56.
33. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 2005, 33:201–212. doi: 10.1093/nar/gki158.
34. Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 2012, 3:385–396. doi: 10.1002/wrna.116.
35. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* 2005, 19:1315–1327. doi: 10.1101/gad.1298605.
36. Zamorano A, López-Camarillo C, Orozco E, Weber C, Guillen N, Marchat LA. In silico analysis of EST and genomic sequences allowed the prediction of cis-regulatory elements for *Entamoeba histolytica* mRNA polyadenylation. *Comput Biol Chem* 2008, 32:256–263. doi: 10.1016/j.compbiolchem.2008.03.019.
37. Espinosa N, Hernández R, López-Griego L, López-Villaseñor I. Separable putative polyadenylation and cleavage motifs in *Trichomonas vaginalis* mRNAs. *Gene* 2002, 289:81–86. doi: 10.1016/S0378-1119(02)00476-6.
38. Fuentes V, Barrera G, Sánchez J, Hernández R, López-Villaseñor I. Functional analysis of sequence motifs involved in the polyadenylation of *Trichomonas vaginalis* mRNAs. *Eukaryot Cell* 2012, 11:725–734. doi: 10.1128/EC.05322-11.
39. Hon CC, Weber C, Sismeiro O, Proux C, Koutero M, Deloger M, Das S, Agrahari M, Dillies MA, Jagla B, et al. Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. *Nucleic Acids Res* 2013, 41:1936–1952. doi: 10.1093/nar/gks1271.
40. Clayton C, Michaeli S. 3' processing in protists. *Wiley Interdiscip Rev RNA* 2011, 2:247–255. doi: 10.1002/wrna.49.
41. López-Camarillo C, Orozco E, Marchat LA. *Entamoeba histolytica*: comparative genomics of the pre-mRNA 3' end processing machinery. *Exp Parasitol* 2005, 110:184–190. doi: 10.1016/j.exppara.2005.02.024.
42. López-Camarillo C, Hernández de la Cruz ON, Vivas JG, Retana JF, Valdez MP, Rosas IL, Alvárez-Sanchez E, Marchat LA. Recent insights in pre-mRNA 3'-end processing signals and proteins in the protozoan parasite *Entamoeba histolytica*. *Infect Disord Drug Targets* 2010, 10:258–265. doi: 10.2174/187152610791591575.
43. López-Camarillo C, Marchat LA, Orozco E, Azuara-Liceaga E, García-Vivas J. Canonical and non canonical poly(A) polymerases: from human to the protozoan parasite *Entamoeba histolytica*. In: Takeo Takeyama Ed. *Messenger RNA Research Perspectives*. Hauppauge, NY: Nova Publishers; 2007, 125–144.
44. García-Vivas J, López-Camarillo C, Azuara-Liceaga E, Orozco E, Marchat LA. *Entamoeba histolytica*: cloning and expression of the poly(A) polymerase EhPAP. *Exp Parasitol* 2005, 10:226–232. doi: 10.1016/j.exppara.2005.02.017.
45. López-Rosas I, Orozco E, Marchat LA, García-Rivera G, Guillen N, Weber C, Carrillo-Tapia E, Hernández de la Cruz O, Pérez-Plasencia C, López-Camarillo C. mRNA decay proteins are targeted to poly(A) + RNA and dsRNA-containing cytoplasmic foci that resemble P-bodies in *Entamoeba histolytica*. *PLoS One* 2012, 7:e45966. doi: 10.1371/journal.pone.0045966.
46. Zhang H, Pompey JM, Singh U. RNA interference in *Entamoeba histolytica*: implications for parasite biology and gene silencing. *Future Microbiol* 2011, 6:103–117. doi: 10.2217/fmb.10.154.
47. Abed M, Ankri S. Molecular characterization of *Entamoeba histolytica* RNase III and AGO2, two RNA interference hallmark proteins. *Exp Parasitol* 2005, 110:265–269. doi: 10.1016/j.exppara.2005.02.023.
48. Aravind L, Koonin EV. DNA polymerase-like nucleotidyl-transferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res* 1999, 27:1609–1618. doi: 10.1093/nar/27.5.1223.
49. Pezet-Valdez M, Fernández-Retana J, Ospina-Villa JD, Ramírez-Moreno ME, Orozco E, Charcas-Lopez S, Soto-Sanchez J, Mendoza-Hernandez G, Lopez-Casamichana M, Lopez-Camarillo C, et al. The 25 kDa subunit of cleavage factor Im is a rna-binding protein that interacts with the poly(A) polymerase in *Entamoeba histolytica*. *PLoS One* 2013, 8:e67977. doi: 10.1371/journal.pone.0067977.
50. Coseno M, Martin G, Berger C, Gilmartin G, Keller W, Doublié S. Crystal structure of the 25 kDa subunit of human cleavage factor Im. *Nucleic Acids Res* 2008, 36:3474–3483. doi: 10.1093/nar/gkn079.

51. Rüegsegger U, Beyer K, Keller W. Purification and characterization of human cleavage factor Im involved in the 3′ end processing of messenger RNA precursors. *J Biol Chem* 1996, 271:6107–6113. doi: 10.1074/jbc.271.11.6107.

52. Shimazu T, Horinouchi S, Yoshida M. Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3'-end processing. *J Biol Chem* 2007, 282:4470–4478. doi: 10.1074/jbc.M609745200.

53. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 2010, 11:345–355. doi: 10.1038/nrg2776.

54. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 2011, 12:715–729. doi: 10.1038/nrg3052.

55. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* 2007, 8:113–126. doi: 10.1038/nrm2104.

56. Chen CY, Shyu AB. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2011, 2:167–183. doi: 10.1002/wrna.40.

57. Caponigro G, Parker R. Mechanisms and control of mRNA turnover in *Saccharomyces cerevisiae*. *Microbiol Rev* 1996, 60:233–249.

58. Cevher MA, Zhang X, Fernandez S, Kim S, Baquero J, Nilsson P, Lee S, Virtanen A, Kleiman FE. Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage. *EMBO J* 2010, 29:1674–1687. doi: 10.1038/emboj.2010.59.

59. Parker R, Song H. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* 2004, 11:121–127. doi: 10.1038/nsmb724.

60. Lebreton A, Tomecki R, Dziembowski A, Seraphin B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* 2008, 456:993–996. doi: 10.1038/nature07480.

61. Eulalio A, Behm-Ansmant I, Izaurralde E. P-bodies: at the crossroads of post-transcriptional pathways. *Nat Rev Mol Cell Biol* 2007, 8:9–22. doi: 10.1038/nrm2080.

62. Cassola A, De Gaudenzi JG, Frasch AC. Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. *Mol Microbiol* 2007, 65:655–670. doi: 10.1111/j.1365-2958.2007.05833.x.

63. Holetz FB, Correa A, Avila AR, Nakamura CV, Krieger MA, Goldenberg S. Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochem Biophys Res Commun* 2007, 356:1062–1067. doi: 10.1016/j.bbrc.2007.03.104.

64. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998, 391:806–811. doi: 10.1038/35888.

65. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature* 2004, 431:343–349. doi: 10.1038/nature02873.

66. Pak J, Fire A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 2007, 315:241–244. doi: 10.1126/science.1132839.

67. Halic M, Moazed D. Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell* 2010, 140:504–516. doi: 10.1016/j.cell.2010.01.019.

68. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004, 116:281–297.

69. Vayssie L, Vargas M, Weber C, Guillen N. Double-stranded RNA mediates homology-dependent gene silencing of γ-tubulin in the human parasite *Entamoeba histolytica*. *Mol Biochem Parasitol* 2004, 138:21–28. doi: 10.1016/j.molbiopara.2004.07.005.

70. Boettner DR, Huston CD, Linford AS, Buss SN, Houpt E, Sherman NE, Petri WA Jr. *Entamoeba histolytica* phagocytosis of human erythrocytes involves PATMK, a member of the transmembrane kinase family. *PLoS Pathog* 2008, 4:e8. doi: 10.1371/journal.ppat.0040008.

71. Zhang H, Ehrenkaufer GM, Pompey JM, Hackney JA, Singh U. Small RNAs with 5′-polyphosphate termini associate with a piwi-related protein and regulate gene expression in the single-celled eukaryote *Entamoeba histolytica*. *PLoS Pathog* 2008, 4:e1000219.

72. De S, Pal D, Ghosh SK. *Entamoeba histolytica*: computational identification of putative microRNA candidates. *Exp Parasitol* 2006, 113:239–243. doi: 10.1016/j.exppara.2006.01.009.

73. Tanner NK, Linder P. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* 2001, 8:251–262. doi: 10.1016/S1097-2765(01)00329-X.

74. Linder P, Jankowsky E. From unwinding to clamping—the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 2011, 12:505–516. doi: 10.1038/nrm3154.

75. Marchat LA, Orozco E, Guillen N, Weber C, López-Camarillo C. Putative DEAD and DExH-box RNA helicases families in *Entamoeba histolytica*. *Gene* 2008, 424:1–10. doi: 10.1016/j.gene.2008.07.042.

76. Cordin O, Tanner NK, Doere M, Linder P, Banroques J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 2004, 23:2478–2487. doi: 10.1038/sj.emboj.7600272.

77. Weber C, Guigon G, Bouchier C, Frangeul L, Moreira S, Sismeiro O, Gouyette C, Mirelman D, Coppee JY, Guillén N. Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *Entamoeba histolytica*. *Eukaryot Cell* 2006, 5:871–875. doi: 10.1128/EC.5.5.871-875.2006.

78. Gilchrist CA, Houpt E, Trapaidze N, Fei Z, Crasta O, Asgharpour A, Evans C, Martino-Catt S, Baba DJ,

Stroup S, et al. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Mol Biochem Parasitol* 2006, 147:163–176. doi: 10.1016/j.molbiopara.2006.02.007.

79. Marion S, Laurent C, Guillen N. Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in *Entamoeba histolytica*: a proteomic approach. *Cell Microbiol* 2005, 7:1504–1518. doi: 10.1111/j.1462-5822.2005.00573.x.

80. López-Camarillo C, de la Luz García-Hernández M, Marchat LA, Luna-Arias JP, Hernández de la Cruz O, Mendoza L, Orozco E. *Entamoeba histolytica* EhDEAD1 is a conserved DEAD-box RNA helicase with ATPase and ATP-dependent RNA unwinding activities. *Gene* 2008, 414:19–31. doi: 10.1016/j.gene.2008.01.024.

