Transcriptional Control in *Entamoeba*: Something Old, Something New

Shreyasee Hazra and Dipak Manna

Abstract

Entamoeba histolytica is an extracellular protozoan parasite and is a global health problem that kills approximately 100,000 people annually. The life cycle of *Entamoeba* consists of two stages – infective trophozoite form and dormant cyst form. *Entamoeba* infection begins with the entry of cysts into our body with contaminated hands touching to our mouth or intake of contaminated water and food. Trophozoites emerged from the resting cysts in the intestine and relocate to the colon where they multiply by binary fission and can cause invasive or noninvasive disease. The key aspects of their host cell-killing activities are engulfing small/dead host cells by the process called "phagocytosis," nibbling bigger/live host cells by the process called "trogocytosis," or inducing apoptotic death of the host cell. It is evident that the pathogenesis, virulence, and development of *Entamoeba* are controlled by the fine tuning of the process called transcription; however, not much is known about the transcriptional regulation and gene expression in this parasite. Transcription regulatory networks play a key role in global gene expression which control a vast range of biological processes and mostly are well characterized in model organisms like yeast, *Drosophila*, and mammals; however, these processes are not well understood in a non-model organism like *Entamoeba histolytica.* In *Entamoeba* only a few transcription factors (TFs) and DNA motifs have been characterized so far. In this chapter we give an overview of transcriptional regulation features in *Entamoeba*, summarizing all transcription factors identified up to date and their significant roles in *Entamoeba* biology.

S. Hazra · D. Manna (✉)
School of Biological Sciences, Ramakrishna Mission Vivekananda Educational and Research Institute (RKMVERI), Kolkata, West Bengal, India

B. Mukherjee et al. (eds.), *Pathobiology of Parasitic Protozoa: Dynamics and Dimensions*, https://doi.org/10.1007/978-981-19-8225-5_8

Keywords

Entamoeba · Transcription factor · Development · Stress response · Virulence · Encystation

1 Introduction

Many unusual features in the transcriptional regulation were demonstrated in this early branching protozoan parasite *Entamoeba* including the following: (1) an atypical RNA polymerase present in *Entamoeba* that is resistant to alpha-amanitin [1]; (2) during mitosis chromatin doesn't condense; (3) histone H3 and H4 comprise a variable N-terminal tail, and the TATA box that is present upstream of the (Inr) initiator region is unusual [2–5]; (4) very short untranslated regions (UTRs) [3, 6–8]; (5) a GAAC element (AATGAACT) or GAAC-like element comprises different locations in the main promoter [6, 8–10]; and (6) an Inr element (AAAAATTCA) present adjacent the transcription start position [6, 10]. Furthermore, the putative TATA-binding protein in *E. histolytica* (EhTBP) shows notable sequence divergence from the TATA-binding protein of other higher eukaryotes [11]. Conjointly, it comes out that the transcriptional regulation in *Entamoeba* is controlled by unusual mechanisms. The core promoters of *Entamoeba* consist of three elements: (a) putative TATA element (GTATTTAAA) at approximately 30 nt upstream of the transcription start site, (b) GAAC element (AATGAACT) with different locations in the core promoter, and (c) an Inr element (AAAAATTCA) adjacent the transcription start site [6, 10, 12]. A GAAC-like motif (EiCPM-GL) (GAACTACAAA) that shows high similarity with GAAC element has also been identified in *E. invadens* which create similar diarrheal disease in reptiles.

The unavailability of genetic manipulations in *Entamoeba* was the main hurdle to identify and characterize the transcription factors in this parasite for a long time. Once Nickel and Tannich developed the transfection protocols to introduce plasmid DNA into *Entamoeba*; it gives a new dimension to study and characterize the transcription factors [13–15]. The development of several *Entamoeba* vectors with reporter genes was helpful for the characterization of several *Entamoeba* cis-regulatory elements and core promoter. Additionally, a putative TATA box was identified in this parasite which is 30 nt upstream from the transcription start site along with an Inr element (adjacent to the transcription initiation site) and a GAAC element [6]. In protozoan system, this GAAC element is unique and is capable of controlling the transcription initiation independent of either the TATA box or the Inr element [16]. Further, in silico analysis of gene promoter along with biochemical approaches identified few more TFs (e.g., EhCudA, HRM-BP, ERM-BP) in this parasite characterized through deletion and replacement analysis.

1.1 *Entamoeba* Genome Sequencing and Transcriptomic Data Improves the Field

E. histolytica genome sequencing was an important advancement in understanding the *Entamoeba* biology, and further refinement of *Entamoeba* genomic features was achieved by reassembly of the genome [17, 18]. The whole genome size of *Entamoeba* is predicted to be ~20 Mbp comprising the following characteristics relevant to gene structure and transcription: (1) ~ 8200 gene codes for protein along with a median gene length of 1260 bp, (2) small number of intron (~ 24% genes carry introns), (3) a unique RNA polymerase II having several especial features comprising a highly variant α-amanitin-binding region, which explains why this organism shows resistance to this drug, (4) TATA-binding proteins encoded by three genes in *Entamoeba* [19, 20], (5) Myb domain containing proteins in *Entamoeba* which are greatly expanded [21], (6) histone acetyl transferases and histone deacetylases (these two histone-modifying proteins are identified in this parasite) [22], (7) demethylase domains containing protein not identified in this parasite, and (8) presence of one DNA methyltransferase (cytosine-5) protein in *Entamoeba* [23].

Entamoeba genome sequencing and genome annotation provide an inauguration platform for many studies in this parasite. For an example 32 Myb domain-containing proteins were identified by comparative in silico analysis and are further classified into three families [21]. Family I consists of two Myb domains and structurally resembles the plant Myb domain proteins. Moreover, the individual domains of *Entamoeba* Myb share closest homology with human c-Myb. On the other hand, families II and III both comprise a single Myb domain.

Despite the effort from many groups, there are no methods developed yet to study the encystation in *E. histolytica*, and *E. invadens* which is a closely related parasite in reptiles has been developed as a model for the study of stage conversion in this parasite.

The genome sequence of both *E. histolytica* and *E. invadens* is extremely repetitive, and it appears that only 50% of the genome size is accounted for genic and intergenic sequence. The genome of *E. invadens* accounted for 11,549 predicted genes compared to 8306 in *E. histolytica.* The genome analysis in *Entamoeba* showed that the length of the genes of *E. histolytica* and *E. invadens* is very similar; however, the intergenic regions in *E. histolytica* tend to be shorter compared to *E. invadens*.

In *E. invadens* out of 11,549 predicted genes, 9865 showed a BLASTP (E-value $<10^{-5}$) hit to 7216 genes (out of 8306 predicted genes in *E. histolytica*), and among those 5227 are putative orthologs. Alignment of orthologs showed an average amino acid identity around 69%, indicating that these species are distantly related. Of the *E. invadens* genes which do not have orthologs in *E. histolytica*, 77% (4815/6218) have at least considerate RNA-seq support, compared to 98% (5206/5331) of genes that shared with *E. histolytica* [24]. This result indicates that a fraction of these genes may be false positive predictions; additionally this is also consistent that many of these genes are not constitutively expressed.

To point out the conserveness between the genes in these two *Entamoeba* species, all collinear gene pairs that were adjacent in both *E. histolytica* and *E. invadens* were analyzed. This analysis showed only 561 genes that preserved their neighboring gene in both species (out of 5227 total genes). Hence it is quite clear that there has been extensive rearrangement in the genome between these two species and most of the biological processes are also conserved.

1.2 Gene Expression Profiling and Transcriptional Regulation

Expression profiling with advanced technologies like microarray and RNA-seq has revolutionized the research field in understanding transcriptional regulation on a genome-wide scale [24]. These approaches have been significantly used in studying *Entamoeba* gene expression in different conditions as well as throughout the different stages of development [24]. Different types of microarray platforms in *Entamoeba* have been developed such as that generated from genomic DNA, short oligonucleotides and long oligonucleotides. These tools help in identifying transcript abundance in different developmental time points as well as during stress and host invasion. Moreover, many factors that are responsible for virulence and pathogenicity have been identified by comparative gene expression studies, genes that are upregulated in virulent strain but downregulated in non-virulent *Entamoeba*. All together, these findings have provided insights on molecular aspects of important amebic biology, e.g., stage conversion and pathogenic potential, and allow researcher with the first intuition to identify prospective novel drug targets against amebic disease.

2 Transcription Factors in Cellular Function

The fundamental step in gene expression is "transcription" process, where an mRNA is synthesized from a DNA template, followed by the second step "translation" that strings the amino acid together to make protein. Developmental studies have shown the upregulation and downregulation of sets of transcript level during the different stages of development as well as at different growth conditions. The transcription process is controlled by the orchestrated function of several proteins, e.g., a protein can bind DNA, and this DNA-binding proteins may involve in regulation in gene expression. Most of our knowledge on the basic elements of transcription regulation is achieved from early work on prokaryotic systems, where genes are arranged in sets of contiguous genes that comprise regulatory sequences and structural genes. A classic example is the lactose (*lac*) operon of *E. coli*. The transcription in eukaryotes is much more complex than in prokaryotes. First, the prokaryotes utilize only one RNA polymerase; however, in eukaryotes there are three different RNA polymerases: I, II, and III. Second, the eukaryotic RNA polymerases require additional proteins called general transcription factors (TFs) to position them at the correct start site. However, during transcription in prokaryotes, RNA polymerases

also require accessory polypeptides called sigma factors (σ), which are considered as a subunit of the RNA polymerase. On the other hand, a large, multi-subunit transcription initiation complex is formed in eukaryotic transcription initiation. For example, RNA polymerase II requires a multi-subunit complex of seven general transcription factors to constitute the initiation complex, and each of the subunits must be added in an orchestrated way.

Transcription factors normally have three structural features: a domain that binds to DNA, a transcription-activating domain, and a domain that binds to a ligand. The DNA-binding domain binds to a specific DNA sequence through the formation of hydrogen, ionic, and hydrophobic bonds, although the particular combination and spatial distribution of such interactions are distinctive for each sequence. In silico analysis of many DNA-binding proteins guided the identification of a number of highly conserved DNA-binding structural motifs; these are (1) HTH (helix-turn-helix) motif, (2) ZnF (zinc finger) motif, (3) HLH (helix-loop-helix) motif, (4) leucine zipper motif, and (5) basic zipper motif.

Cellular responses consist of a cascade of events both in prokaryotes and in eukaryotes which involves many intracellular signaling pathways (e.g., PKA, MAPks, JAKs, PKCs) that control the fine tuning of gene regulation by many transcription factors. Transcription factors in bacteria are generally classified by comparison of amino acid sequence with prototypic members of families of DNA-binding proteins, such as LysR-like and AraC-like protein families. TFs are often classified based on the structural motifs that constitute their binding domains, for example, TBP (TATA-binding protein), TBP-associated factors (TAF), and recently identified p300/CBP coactivator family. There are several families of TFs that exist, and each of which shows structural and functional features. The examples of such families are helix-turn-helix (e.g., Oct1), helix-loop-helix (e.g., E2A), zinc finger (e.g., GATA proteins, TFIIIA), leucine zipper motif (cAMP, CREB, AP-1), and beta-sheet motif (e.g., nuclear factor-kB) [25].

In eukaryotes, there is a class of transcription factors called GTFs (general transcription factors) involved in basal transcription regulation which includes TFIIA, TFIIB, TFIID, TFIIE, and THIIH. Jump-start of different transcription factors, for example p53, NF-κB (nuclear factor-κB), AP-1 (activated protein-1), Nrf2 (nuclear erythroid-derived 2-related factor 2), and CREB (cAMP-responsive element-binding) protein associate with various cellular function like p53 and NF-kB are involved in cellular damage response. NF-κB family play critical roles in immunity, inflammation, differentiation, cell proliferation, and survival [26]. AP-2 family transcription factors are evolutionarily conserved that bind to the DNA consensus sequence GCCNNNGGC and upregulate target gene expression. In mammals, four different isoforms of AP-2 have been identified, termed AP-2 α, β, γ, and δ. Studies have identified the role of AP-2 TF in *Plasmodium* ApiAP2 transcription factor (PfAP2-EXP2) – controlling the gene expression in the intraerythrocytic developmental cycle of plasmodium parasite. AP-1 on the other hand participates in control of proliferation, senescence, differentiation, and apoptosis [27]. Sp1 is a member of transcription factors which include Sp2, Sp3, and Sp4 playing a role in DNA repair. CREB is a phosphorylation-dependent nuclear

transcription factor that is involved in different important cellular functions including apoptosis and cell proliferation. The cAMP-CRP protein is considered as lying between the conventional transcription regulators and histone-like proteins, and it can bind specifically to a consensus DNA sequence. Another TF is FOXO3a protein, a fork-head transcription factor that is a member of FOXO subfamily and mediates a variety of cellular process including proliferation, cell cycle progression, DNA damage, and apoptosis [28]. The next important TF is E2F that is activated by E1A protein that is a viral oncoprotein and needed for adenovirus gene expression. E2F transcription factors are recognized as key players in controlling the cell cycle, transformation, and differentiation, and it has been found that the E2F/pRB pathway acts as a key regulator on cell cycle and development. Quite a few important TFs and DNA motifs have been characterized in protozoan parasites. For instance, a member of the HMGB was identified in *Entamoeba histolytica*, and some Myb family members were characterized from *Trichomonas vaginalis*, and a cell cycle-dependent ApiAP2 transcription factor, TgAP2IX-5, was found in *Toxoplasma gondii* [29]. The list of transcription factors identified so far in *Entamoeba* is shown in Table 1, and their functions are depicted in the schematic in Fig. 1. However, the biological role of many TFs in this parasite is still poorly understood, and further characterization is needed for better understanding.

2.1 TATA-Box-Binding Protein

In the past two decades, important improvements have been achieved in terms of molecular biology techniques that expanded our perception of transcriptional regulation *in E. histolytica.* Several groups have identified a number of TFs and the core promoter region in *Entamoeba*. However, very little is known about the transcription machinery and especially transcription regulation during the development of this parasite.

In the late 1990s or early 2000s, the approaches used by different groups to identify the transcription factors in this parasite were mainly based on comparative amino acid sequence analysis of known transcription factors, present in other systems [30], or yeast one-hybrid assay [31] or deletion or replacement analysis of consensus motifs in the promoter region [31, 32]. Among the earlier approaches, comparative analysis of amino acid sequence TATA-box-binding protein from *Acanthamoeba castellanii* identified the *Entamoeba* transcription factor as TATA-box-binding protein (EhTBP) [30]. The EhTBP is more unselective compared to higher eukaryotes and binds a wide variety of *E. histolytica* TATA-box sequence [11, 19]. Later on, genome sequencing, gene expression profiling, and proteomics approaches advanced to study the transcriptional networks and help identify novel transcription factors [12, 20, 21, 33–36]. Subsequently, the sequencing and annotation of *Entamoeba* genome provide identification of two more amoebic TATA-binding proteins (TBP) [20]. TBP and TRF1 transcription factors in *E. histolytica* are GAAC-box-binding proteins that represent distinctive expression of genes under

Table 1 *Entamoeba* transcription factors, their representative DNA-binding motifs, and information gained

Transcription factors	DNA-binding motif	Information gained	References
EhTBP TATA-binding protein	TATTTAAA	Shows specific binding in vitro to TATA box EhTBP is more unselective compared to TBP of human and yeast	Luna-Arias et al. [30] de Dios-Bravo et al. [19]
EhTRF1 TBP-related factor 1	GAAC-box-binding proteins	Two EhTRF identified. TRF1 binds to GAAC box and displays distinctive expression of genes under stress response and during host-*Entamoeba* interaction	Castañon-Sanchez et al. [20]
URE3-BP Upstream regulatory element-binding protein	TATTCTATT (URE3)	DNA motif identified. URE3-BP identified. Identification of target genes	Purdy et al. [6] Gilchrist et al. [56] Gilchrist et al. [59] Gilchrist et al. [58]
EhEBP1 and EhEBP2 Enhancer-binding proteins	AAAAATGAATGGAAAAATGAA (URE4)	DNA motif identified. Identification of EhEBP1 and 2	Schaenman et al. [32] Schaenman et al. [31]
EhMyb10	TAACGG	32 Myb domain proteins identified	Meneses et al. [21]
EhMyb-dr	CCCCCC	EhMyb-dr identifiedEhMyb-dr DNA motif was identified	Ehrenkaufer et al. [50]
ECudA	AGAATTTTCT	CudA homolog identified in *Entamoeba*	Yamada et al. [34]
C/EBP	TGTTTGGTAGTTGAATTGGAAAGAA	C/EBP binding DNA motif was identified. Partial purifications of proteins	Marchat et al. [36]
Ehp53	GGACATGCCCGGGCATGTCC	Shows specific binding in vitro	Mendoza et al. [39]
HMGB1 (high mobility group box protein 1)	DNA structure, e.g., stem-loops	Proof of DNA bendingHMGB1 increase DNA binding of protein p53. Genes regulated by HMGB1 were identified	Gilchrist et al. [33] Abhyankar et al. [40]
EhHSTF	5′-GAA-3′ motif into heat shock elements of the EhPgp5 gene	Play an important function in multidrug resistance activity in *E. histolytica*	Bello et al. [62]
HRM-BP	AAACCTCAATGAAGA	Transcriptional control of genes involved in oxidative stress response	Pearson et al. [35]

(continued)

Table 1 (continued)

Transcription factors	DNA-binding motif	Information gained	References
EhPC4	–	Induce cell migration, multinucleation, and polykaryon formation promoted by EhPC4	Cruz et al. [37, 38]
ERM-BP (encystation regulatory motif-binding protein)	CAACAAA	An NAD^+ regulated TF. Upregulated during encystation. Control cyst-specific gene expression	Manna et al., [51, 52]
NF-Y complex (nuclear factor-Y)	CCAAT	CCAAT motif specifically binds to *Entamoeba* NF-Y complex Control encystation NF-Y works downstream of ERM-BP transcriptional control	Manna et al. [53, 54]
EhGATA-TF	GATA	Modulates genes involved in phagocytosis	Huerta et al. [61]

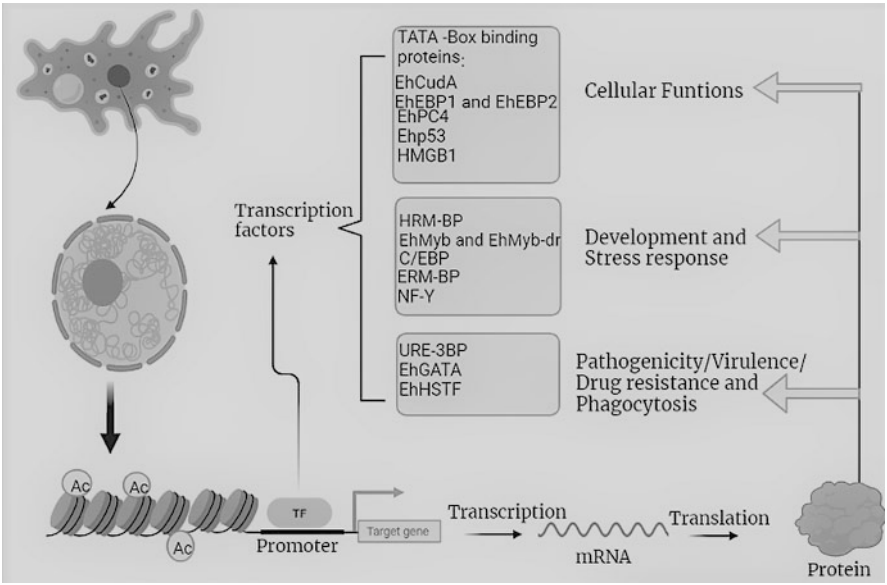

Fig. 1 Transcription factors and their roles in *Entamoeba*

stress response and during the interaction of *Entamoeba* with mammalian cells. However, the biological role of these two new TBP is yet to be determined.

2.2 EhCudA

The transcription factor EhCudA was identified by a comparative in silico approach by utilizing *Dictyostelium* CudA as a query [34]. In *Dictyostelium* this protein is necessary for pre-spore-specific gene expression and has significant homology in *Entamoeba* protein [34]. Yamada et al. expressed CudA protein in bacteria and used recombinant protein and were able to identify the DNA-binding motif AGAATTTTCT which shows specific interaction with CudA in vitro; however the functional characterization of *Entamoeba* CudA is yet to be determined [34].

2.3 EhEBP1 and EhEBP2

Two enhancer-binding proteins (EhEBP1 and EhEBP2) which specifically bind to the URE4-binding domain were discovered by using nuclear extracts from amoeba and DNA affinity chromatography followed by mass spectrometry [31]. Some unique features were reported in these TFs; both EhEBP1 and EhEBP2 comprise

an RNA recognition motif RRM; however; no recognizable DNA-binding domain was identified.

2.4 EhPC4

Analysis of genome-wide microarray data from virulent trophozoites isolated from hamster liver abscesses identified a transcription factor, EhPC4 (*E. histolytica*-positive cofactor), which significantly upregulated during the infection [37]. The author has reported the potential role of EhPC4 in liver abscess formation by controlling the expression of vital genes involved in cytoskeleton dynamics, cell migration, and invasion [37]. The transcription factor EhPC4 also possesses important role in regulating DNA replication and genome stability [38].

2.5 Ehp53

A p53-like *E. histolytica* protein (Ehp53) was identified which binds to the human p53-binding consensus DNA sequence confirmed by human p53 antibodies [39]. It has been reported that monoclonal antibody against human p53 protein could recognize the recombinant *Entamoeba* Ehp53 suggesting that in *Entamoeba* this protein may be evolutionarily conserved [39]. In mammalian cells, p53 takes part in several cellular processes like cell cycle regulation, DNA repair, precluding uncontrolled cellular division, and apoptosis; however the functional characterization of this TF is yet to be determined.

2.6 HMGB1

The TF HMGB1 (high mobility group box protein 1) was identified by the analysis of genome-wide transcriptome data during *Entamoeba* colonization and invasion to the intestine [40]. HMGB proteins can bind a diverse sequence of DNA in a conformation-dependent way which includes stem-loops, palindromes, four-way junctions, B-Z junctions, and even single-stranded or cruciform DNA [41]. This protein contains one or more units of the HMG box DNA-binding motif, and it is observed that it can increase DNA binding in a sequence-specific manner. This protein is involved in many important cellular functions, e.g., transcription, recombination, and repair. In a recent study, it was shown that *Entamoeba* when in contact with macrophage induced the secretion of HMGB1 which functions as a pro-inflammatory cytokine and can also act as a chemoattractant during the *Entamoeba* infection [42].

3 TF in Development and Stress Response

Development needs the conscientious orchestration of many biological episodes in order to generate an entire multicellular organism, and in case of unicellular organisms, this orchestration is equally important throughout the different stages of development. Many transcription factors (TFs) involved in the development are conserved evolutionarily from yeast to humans. For example, there are four TF families that play a determining role and have been characterized immensely both during the development of embryo and in cancer. These are (1) GATA, (2) the high mobility group box (HMG), (3) paired box (PAX), and (4) basic helix-loop-helix (bHLH) [43–45]. Living organisms constantly face diverse types of physiological and environmental stress. To survive with the detrimental consequences of stress or to protect against further exposure to the same or other forms of stress, cells have evolved rapid molecular responses to repair the damage. TF plays an important role by upregulation or downregulation of set of genes which makes the organism more resistant in the adverse condition, and many stress-controlled transcription factors have been discovered and characterized in different systems [46, 47].

Transcription factor activation is a complicated process that may involve numerous signal transduction pathways, including several kinases, e.g., PKA, MAPKs, JAKs, and PKCs, which are activated by cell-surface receptors [48]. Major TF families, together with WRKY, MYB, NAC, and AP2/ERF, are important regulators of diverse genes associated with various stressors. WRKY, as one of the most well-studied plant TFs, regulates a wide range of developmental, physiological, and metabolic activities. The WRKY family has been recognized as a major group of transcription factors in many plant species. These function as activators, repressors, or corepressors of essential pathways, such as the generation of alkaloids, terpenes, and other specialized metabolites, and have been proven to be significant in the activation of diverse immune response pathways, making them important in biotic stress. WRKY transcription factors were found to be useful in relieving infection stress produced by biotic or abiotic agents via self-regulation or hormone-mediated signal transduction pathways. In *Entamoeba*, a few transcription factors were identified which control the expression of important genes pertinent to several important facets of *Entamoeba* biology which includes stage conversion and oxidative stress (Table 1 and Fig. 1) [35, 49].

3.1 HRM-BP

A novel H_2O_2 stress-responsive motif HRM was identified by in silico analysis of the promoter sequences of genes that are upregulated in H_2O_2 stress, and the transcription factor HRM-binding protein (HRM-BP) was identified by biochemical analysis and mass spectrometry [35]. The interaction of HRM-BP with the HRM motif is very specific, and alteration of HRM-BP expression either by silencing or by overexpression in *Entamoeba* showed changes in the basal expression of stress responsiveness or H_2O_2-responsive genes [35].

3.2 EhMyb and EhMyb-dr

A set of 32 Myb domain-containing proteins were identified in Eh by analyzing the c-Myb protein sequences as the query from human [21]. Electrophoretic mobility shift assays (EMSA) by using recombinant *Entamoeba* Myb10 (family I) showed that this Myb10 protein could bind the canonical Myb-binding motif (TAACGG) as reported in other eukaryotes and in *Entamoeba*, EhHSP70 gene promoter comprises a Myb DNA-binding motif suggesting its important role in heat shock gene expression in stress response.

Studies in recent years evidenced that transcriptional control has an important contribution in stage conversion in *Entamoeba*, and three transcription factors were identified. Myb transcription factor (EhMyb-dr) is a SHAQKY family Myb gene which binds to a hexa-nucleotide motif CCCCCC; upregulating this protein in *E. histolytica* trophozoites results a transcriptional profile that highly resembles with the transcriptome profile of amoebic encystation [50]. The interplay between the EhMyb-dr protein and the DNA sequence is eventually confirmed by EMSA as well as by chromatin immunoprecipitation (ChIP) analysis, and it is evident that EhMyb-dr regulates a set of cyst-specific genes [50].

3.3 ERM-BP

An encystation regulatory motif (ERM) that is a hepta-nucleotide sequence was (CAACAAA) identified in the promoter of 131 cyst-specific genes in *E. invadens* which is used as model system for developmental studies. Electrophoretic mobility shift assay showed specific binding of *Entamoeba* cyst protein only, not by the trophozoite protein suggesting that the protein bind to ERM may be specifically expressed in cyst only. ERM-binding protein (ERM-BP) was identified by electrophoretic mobility shift assay followed by mass spectrometry. Metabolic cofactor NAD^+ positively regulates the binding of recombinant ERM-BP with ERM, and downregulation of ERM-BP significantly decreased encystation efficiency, and ghost-like abnormal cysts with defective cyst wall are produced suggesting that ERM-BP plays an important role in encystation [51]. The ERM-BP is conserved among other *Entamoeba* species, and upregulating ERM-BP in *E. histolytica* (EHI_146360) produced quadri-nucleate cyst-like structures and makes the parasite more resistant due to heat stress, supporting the idea that heat stress response and encystation might have a potential overlap and some interconnection and share common signaling pathways [52, 53].

3.4 NF-Y (Nuclear Factor Complex)

Nuclear factor complex (NF-Y) is made up of three subunits, namely, NF-YA, NF-YB, and NF-YC, that very specifically bind to a pentanucleotide motif CCAAT and this TF complex conserved throughout evolution [54]. NF-Y plays

crucial roles in higher eukaryotes, controlling many cellular processes (e.g., cell cycle regulation, development, response to growth, stress, DNA damage, and apoptosis) by regulating the expression of genes that comprise CCAAT promoter motif [54].

In *E. invadens* the expression of NF-YA is constitutive; however NF-YB and NF-YC are expressed during encystation. Silencing of the NF-YC subunit in *Entamoeba* showed significant reduction in DNA-binding ability of the NF-Y complex and also reduced encystation efficiency [54].

4 Transcription Factors in Pathogenicity, Virulence, Drug Resistance, and Phagocytosis

Transcription factors (TFs) are central components which play a critical role in the gene expression. A little change in the TF expression and specificity can alter the entire gene expression. During the infection, pathogenic organisms upregulate or downregulate many genes those are downregulated by their TFs which helps in the adaptation of host or tissue specific environment and adaptation of various physiological changes and in the activation of virulence and pathogenicity. The main aim of the identification of TFs is to block the virulence factors in any pathogenic organism. For developing in-depth knowledge about host-pathogen interaction, it is necessary to identify the interplay of signal exchange mechanism which will be helpful to identify the virulence factor and outcome of the infection. Very little information was known regarding the transcriptional switch that helps cell to adjust in response to immune signals and infection. In 2016 Gray et al., identified Fcγ receptor that helps TFEB transcription factor to enhance lysosome-based degradation and killing bacteria [55]. So, it is uncovered thereafter that IgG immune complexes instruct macrophages to transform it as super killers by the upregulated activation of the lysosomes through a transcriptional circuit. It is evident that in *Entamoeba* pathogenesis, virulence and development are controlled transcriptionally.

4.1 URE-3BP

The upstream regulatory element DNA sequence motif TATTCTATT (URE3) was first discovered in the promoter region of the heavy chain subunit of the lectin gene *hgl5* in *E. histolytica* and later on also found in the promoter of ferredoxin (*fdx*) 1 gene [6, 40, 49, 56, 57]. Upstream regulatory element-binding protein (URE3-BP) was identified through a yeast one-hybrid screen by using URE3 as bait [30]. It was reported that the promoter activity increases due to the mutation in URE3 motif in the promoter of *hgl*5 lectin; on the contrary mutation in URE3 motif in the *fdx* 1 gene promoter decreases the promoter activity by twofold of the reporter gene activity, suggesting that URE3 can act as either a negative or positive regulator in gene expression [57]. This transcription factor comprises two calcium-binding motifs (EF hands), and URE3-BP detach from URE3 DNA in the presence of higher

level of calcium, suggesting that calcium acts as negative regulator [58, 59]. The transcription factor URE3-BP is regulated by calcium and controls the expression of two virulence genes in *Entamoeba*, the Gal/GalNac lectin and ferredoxin. It also has been reported that upregulation of URE3-BP leads to the changes in the morphology of trophozoite and boosts parasite invasion in different organs like the colon and liver, suggesting that transcription factor URE3-BP plays a salient role in *Entamoeba* virulence [60].

4.2 EhGATA

The GATA transcription factors are conserved and a part of the DNA-binding domain (ZFBD) family that contains zinc finger and recognizes the consensus DNA sequence (A/T)GATA(A/G). This ZFBD superfamily TF moderates a wide range of cellular functions.

In 2020, Huerta et al. reported the existence of a single *gata* gene in *E. histolytica* (*Ehgata*) by bioinformatic analysis and the GATA domain ensured in 80% similarity to the GATA protein of human [61]. *Ehgata* codes for a noncanonical EhGATA transcription factor that contains an AT-Hook motif and only one zinc finger DNA-binding domain. Bioinformatic prediction showed the presence of GATA-binding sequence over 1600 gene promoters in *Entamoeba* genome [61]. Electrophoretic mobility shift assay with the bacterially expressed and purified EhGATA protein, additionally with trophozoite nuclear extracts, showed binding to the consensus GATA-DNA sequence. Moreover, Huerta et al. showed that EhGATA especially binds to the promoters of *Ehadh* and *Ehvps32* genes in vivo and eventually controls EhADH and EhVps32 gene expression in the course of phagocytosis. Additionally, overexpressing of EhGATA in trophozoites showed significant changes in morphology, alteration in cell proliferation, change in adherence efficiency, and change in rate of phagocytosis. These findings suggest that EhGATA TF is capable to bind DNA and fine-tune the expression of several genes those involved in cell proliferation, adhesion to surface, and phagocytosis [61].

4.3 EhHSTF

When bacteria or any other organisms are exposed to a certain drug or antibiotic, they alter their cellular mechanism to survive. Continuous and excessive exposure of any drug can lead to the rise of a drug-resistant population of cells. For decades the main drug of choice against amebiasis is metronidazole, but due to emergence of drug resistance (DR) in most of the pathogens, it is really alarming that DR will cause a major public health problem worldwide. It has been reported that methionine γ-lyase (*EhMGL*) gene silencing resulted in resistance to trifluoromethionine, revealing a novel mechanism of drug resistance in *E. histolytica.*

In *Entamoeba* it has been observed that emetine stress induces the expression of the multidrug resistance *EhPgp5* gene [62]. Bello et al. showed that the transcription

factor EhHSTF7 recognizes the 5′-GAA-3′ motif into the heat shock element of EhPgp5 gene and is involved in the transcriptional activation of the EhPgp5 gene [62].

5 Summary and Conclusions

The most constructive way to understand the functions of different genes in an organism is by genetic manipulation, and most of the genetic analyses are achieved by alteration in the transcriptome level. In *Entamoeba*, gene expression and their fine tuning by transcriptional controls are still not well understood, and the majority of genes or proteins are hypothetical. Recent advancement in RNA-seq analysis, proteome analysis, and gene editing by CRISPR/Cas9 opens the avenue to analyze their expression pattern during different stages of development, stress condition, and also the differential gene expression between pathogenic and nonpathogenic strains. For example, during oxidative stress it is reported that 57 genes are upregulated in response to H_2O_2 exposure and the expression of these genes is controlled by transcription factor HRM-BP. A Myb domain protein EhMyb-dr binds to CCCCCC motif and upregulates a set of genes during encystation, and another transcription factor ERM-BP that binds to CAACAAA motif and 131 cyst-specific genes which were upregulated was identified having this motif. It has been seen that the TFs that bind to cis-regulatory sequence can either positively or negatively regulate the transcription regulation. In *Entamoeba* TF URE3-BP is reported to regulate the transcription in both ways. URE3-BP positively regulates the expression of lectin heavy chain and negatively regulates ferredoxin 1 gene. However, only a few TFs have been characterized in this parasite till now, and definitely there is an urgency to extend this line of research for better understanding of many unrevealed area of amoebic biology.

Acknowledgments The authors thank the Institute of Health Sciences, Presidency University, for their support. D.M. also thanks the Department of Biotechnology (DBT), India, for the Ramalingaswami Re-entry Fellowship.

References

1. Lioutas C, Tannich E. Transcription of protein-coding genes in Entamoeba histolytica is insensitive to high concentrations of alpha-amanitin. Mol Biochem Parasitol. 1995;73:259–61. https://doi.org/10.1016/0166-6851(95)00101-6.
2. Torres-Guerrero H, Peattie DA, Meza I. Chromatin organization in Entamoeba histolytica. Mol Biochem Parasitol. 1991;45:121–30. https://doi.org/10.1016/0166-6851(91)90033-3.
3. Bruchhaus I, Leippe M, Lioutas C, Tannich E. Unusual gene organization in the protozoan parasite Entamoeba histolytica. DNA Cell Biol. 1993;12:925–33. https://doi.org/10.1089/dna.1993.12.925.
4. Födinger M, Ortner S, Plaimauer B, et al. Pathogenic Entamoeba histolytica: cDNA cloning of a histone H3 with a divergent primary structure. Mol Biochem Parasitol. 1993;59:315–22. https://doi.org/10.1016/0166-6851(93)90229-q.

5. Binder M, Ortner S, Plaimauer B, et al. Sequence and organization of an unusual histone H4 gene in the human parasite Entamoeba histolytica. Mol Biochem Parasitol. 1995;71:243–7. https://doi.org/10.1016/0166-6851(94)00044-n.
6. Purdy JE, Pho LT, Mann BJ, Petri WA. Upstream regulatory elements controlling expression of the Entamoeba histolytica lectin. Mol Biochem Parasitol. 1996;78:91–103. https://doi.org/10.1016/s0166-6851(96)02614-x.
7. Hon C-C, Weber C, Sismeiro O, et al. Quantification of stochastic noise of splicing and polyadenylation in Entamoeba histolytica. Nucleic Acids Res. 2013;41:1936–52. https://doi.org/10.1093/nar/gks1271.
8. Manna D, Ehrenkaufer GM, Singh U. Regulation of gene expression in the protozoan parasite Entamoeba invadens: identification of core promoter elements and promoters with stage-specific expression patterns. Int J Parasitol. 2014;44:837. https://doi.org/10.1016/j.ijpara.2014.06.008.
9. Singh U, Rogers JB, Mann BJ, Petri WA. Transcription initiation is controlled by three core promoter elements in the hgl5 gene of the protozoan parasite Entamoeba histolytica. Proc Natl Acad Sci U S A. 1997;94:8812–7. https://doi.org/10.1073/pnas.94.16.8812.
10. Singh U, Gilchrist CA, Schaenman JM, et al. Context-dependent roles of the Entamoeba histolytica core promoter element GAAC in transcriptional activation and protein complex assembly. Mol Biochem Parasitol. 2002;120:107–16. https://doi.org/10.1016/s0166-6851(01)00441-8.
11. McAndrew MB, Read M, Sims PF, Hyde JE. Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely a+T-rich human malaria parasite plasmodium falciparum. Gene. 1993;124:165–71. https://doi.org/10.1016/0378-1119(93)90390-o.
12. Hackney JA, Ehrenkaufer GM, Singh U. Identification of putative transcriptional regulatory networks in Entamoeba histolytica using Bayesian inference. Nucleic Acids Res. 2007;35:2141. https://doi.org/10.1093/nar/gkm028.
13. Nickel R, Tannich E. Transfection and transient expression of chloramphenicol acetyltransferase gene in the protozoan parasite Entamoeba histolytica. Proc Natl Acad Sci U S A. 1994;91:7095–8. https://doi.org/10.1073/pnas.91.15.7095.
14. Purdy JE, Mann BJ, Pho LT, Petri WA. Transient transfection of the enteric parasite Entamoeba histolytica and expression of firefly luciferase. Proc Natl Acad Sci U S A. 1994;91:7099–103. https://doi.org/10.1073/pnas.91.15.7099.
15. Hamann L, Nickel R, Tannich E. Transfection and continuous expression of heterologous genes in the protozoan parasite Entamoeba histolytica. Proc Natl Acad Sci U S A. 1995;92:8975–9. https://doi.org/10.1073/pnas.92.19.8975.
16. 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;273:21663–8. https://doi.org/10.1074/jbc.273.34.21663.
17. Loftus B, Anderson I, Davies R, et al. The genome of the protist parasite Entamoeba histolytica. Nature. 2005;433:865–8. https://doi.org/10.1038/nature03291.
18. Lorenzi HA, Puiu D, Miller JR, et al. New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information. PLoS Negl Trop Dis. 2010;4:e716. https://doi.org/10.1371/journal.pntd.0000716.
19. de Dios-Bravo G, Luna-Arias JP, Riverón AM, et al. Entamoeba histolytica TATA-box binding protein binds to different TATA variants in vitro. FEBS J. 2005;272:1354–66. https://doi.org/10.1111/j.1742-4658.2005.04566.x.
20. Castañon-Sanchez CA, Luna-Arias JP, de Dios-Bravo MG, et al. Entamoeba histolytica: a unicellular organism containing two active genes encoding for members of the TBP family. Protein Expr Purif. 2010;70:48–59. https://doi.org/10.1016/j.pep.2009.12.007.
21. Meneses E, Cárdenas H, Zárate S, et al. The R2R3 Myb protein family in Entamoeba histolytica. Gene. 2010;455:32–42. https://doi.org/10.1016/j.gene.2010.02.004.

22. Ramakrishnan G, Gilchrist CA, Musa H, et al. Histone acetyltransferases and deacetylase in Entamoeba histolytica. Mol Biochem Parasitol. 2004;138:205–16. https://doi.org/10.1016/j.molbiopara.2004.09.002.
23. Bernes S, Siman-Tov R, Ankri S. Epigenetic and classical activation of Entamoeba histolytica heat shock protein 100 (EHsp100) expression. FEBS Lett. 2005;579:6395–402. https://doi.org/10.1016/j.febslet.2005.09.101.
24. Ehrenkaufer GM, Weedall GD, Williams D, et al. The genome and transcriptome of the enteric parasite Entamoeba invadens, a model for encystation. Genome Biol. 2013;14:R77. https://doi.org/10.1186/gb-2013-14-7-r77.
25. Hellweg CE, Spitta LF, Henschenmacher B, et al. Transcription factors in the cellular response to charged particle exposure. Front Oncol. 2016;6 https://doi.org/10.3389/fonc.2016.00061.
26. Oeckinghaus A, Ghosh S. The NF- B family of transcription factors and its regulation. Cold Spring Harb Perspect Biol. 2009;1:a000034. https://doi.org/10.1101/cshperspect.a000034.
27. Shang X, Wang C, Shen L, et al. PfAP2-EXP2, an essential transcription factor for the Intraerythrocytic development of plasmodium falciparum. Front Cell Dev Biol. 2021;9:782293. https://doi.org/10.3389/fcell.2021.782293.
28. Liu Y, Ao X, Ding W, et al. Critical role of FOXO3a in carcinogenesis. Mol Cancer. 2018;17:104. https://doi.org/10.1186/s12943-018-0856-3.
29. Khelifa AS, Guillen Sanchez C, Lesage KM, et al. TgAP2IX-5 is a key transcriptional regulator of the asexual cell cycle division in toxoplasma gondii. Nat Commun. 2021;12:116. https://doi.org/10.1038/s41467-020-20216-x.
30. Luna-Arias JP, Hernandez-Rivas R, de Dios-Bravo G, et al. The TATA-box binding protein of Entamoeba histolytica: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy. Microbiology. 1999;145(Pt 1):33–40. https://doi.org/10.1099/13500872-145-1-33.
31. Schaenman JM, Gilchrist CA, Mann BJ, Petri WA. Identification of two Entamoeba histolytica sequence-specific URE4 enhancer-binding proteins with homology to the RNA-binding motif RRM. J Biol Chem. 2001;276:1602–9. https://doi.org/10.1074/jbc.M006866200.
32. Schaenman JM, Driscoll PC, Hockensmith JW, et al. An upstream regulatory element containing two nine basepair repeats regulates expression of the Entamoeba histolytica hgl5 lectin gene. Mol Biochem Parasitol. 1998;94:309–13. https://doi.org/10.1016/s0166-6851(98)00081-4.
33. Gilchrist CA, Houpt E, Trapaidze N, et al. Impact of intestinal colonization and invasion on the Entamoeba histolytica transcriptome. Mol Biochem Parasitol. 2006;147:163–76. https://doi.org/10.1016/j.molbiopara.2006.02.007.
34. Yamada Y, Wang HY, Fukuzawa M, et al. A new family of transcription factors. Development. 2008;135:3093–101. https://doi.org/10.1242/dev.026377.
35. Pearson RJ, Morf L, Singh U. Regulation of H2O2 stress-responsive genes through a novel transcription factor in the protozoan pathogen Entamoeba histolytica. J Biol Chem. 2013;288:4462. https://doi.org/10.1074/jbc.M112.423467.
36. Marchat LA, Gómez C, Pérez DG, et al. Two CCAAT/enhancer binding protein sites are cis-activator elements of the Entamoeba histolytica EhPgp1 (mdr-like) gene expression. Cell Microbiol. 2002;4:725–37. https://doi.org/10.1046/j.1462-5822.2002.00220.x.
37. de la Cruz OH, Muñiz-Lino M, Guillén N, et al. Proteomic profiling reveals that EhPC4 transcription factor induces cell migration through up-regulation of the 16-kDa actin-binding protein EhABP16 in Entamoeba histolytica. J Proteome. 2014;111:46–58. https://doi.org/10.1016/j.jprot.2014.03.041.
38. Cruz OH de la, Marchat LA, Guillén N, et al (2016) Multinucleation and Polykaryon formation is promoted by the EhPC4 transcription factor in Entamoeba histolytica. Sci Rep 6:19611. https://doi.org/10.1038/srep19611.
39. Mendoza L, Orozco E, Rodríguez MA, et al. Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53. Microbiology. 2003;149:885–93. https://doi.org/10.1099/mic.0.25892-0.

40. Abhyankar MM, Hochreiter AE, Hershey J, et al. Characterization of an Entamoeba histolytica high-mobility-group box protein induced during intestinal infection. Eukaryot Cell. 2008;7:1565–72. https://doi.org/10.1128/EC.00123-08.
41. Bustin M, Reeves R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. Prog Nucleic Acid Res Mol Biol. 1996;54:35–100. https://doi.org/10.1016/s0079-6603(08)60360-8.
42. Begum S, Moreau F, Leon Coria A, Chadee K. Entamoeba histolytica stimulates the alarmin molecule HMGB1 from macrophages to amplify innate host defenses. Mucosal Immunol. 2020;13:344–56. https://doi.org/10.1038/s41385-019-0233-6.
43. Roy N, Hebrok M. Regulation of cellular identity in cancer. Dev Cell. 2015;35:674–84. https://doi.org/10.1016/j.devcel.2015.12.001.
44. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68:820–3. https://doi.org/10.1073/pnas.68.4.820.
45. Huilgol D, Venkataramani P, Nandi S, Bhattacharjee S. Transcription factors that govern development and disease: an Achilles heel in cancer. Genes (Basel). 2019;10 https://doi.org/10.3390/genes10100794.
46. Ruis H, Schüller C. Stress signaling in yeast. BioEssays. 1995;17:959–65. https://doi.org/10.1002/bies.950171109.
47. Estruch F. Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. FEMS Microbiol Rev. 2000;24:469–86. https://doi.org/10.1111/j.1574-6976.2000.tb00551.x.
48. Gilmore TD. Introduction to NF-kappaB: players, pathways, perspectives. Oncogene. 2006;25:6680–4. https://doi.org/10.1038/sj.onc.1209954.
49. Pearson RJ, Singh U. Approaches to characterizing Entamoeba histolytica transcriptional regulation. Cell Microbiol. 2010;12:1681–90. https://doi.org/10.1111/j.1462-5822.2010.01524.x.
50. 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. https://doi.org/10.1111/j.1462-5822.2009.01300.x.
51. Manna D, Lentz CS, Ehrenkaufer GM, et al. An NAD+-dependent novel transcription factor controls stage conversion in Entamoeba. eLife. 2018; https://doi.org/10.7554/eLife.37912.
52. Manna D, Lozano-Amado D, Ehrenkaufer G, Singh U. The NAD+ responsive transcription factor ERM-BP functions downstream of cellular aggregation and is an early regulator of development and heat shock response in Entamoeba. Front Cell Infect Microbiol. 2020;10:363. https://doi.org/10.3389/fcimb.2020.00363.
53. Manna D, Ehrenkaufer GM, Lozano-Amado D, Singh U. Entamoeba stage conversion: progress and new insights. Curr Opin Microbiol. 2020;58:62–8. https://doi.org/10.1016/j.mib.2020.09.005.
54. Manna D, Singh U. Nuclear factor Y (NF-Y) modulates encystation in Entamoeba via stage-specific expression of the NF-YB and NF-YC subunits. mBio. 2019;10 https://doi.org/10.1128/mBio.00737-19.
55. Gray MA, Choy CH, Dayam RM, et al. Phagocytosis enhances lysosomal and bactericidal properties by activating the transcription factor TFEB. Curr Biol. 2016;26:1955–64. https://doi.org/10.1016/j.cub.2016.05.070.
56. Gilchrist CA, Holm CF, Hughes MA, et al. Identification and characterization of an Entamoeba histolytica upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. J Biol Chem. 2001;276:11838–43. https://doi.org/10.1074/jbc.M007375200.
57. Gilchrist CA, Mann BJ, Petri WA. Control of ferredoxin and gal/GalNAc lectin gene expression in Entamoeba histolytica by a cis-acting DNA sequence. Infect Immun. 1998;66:2383–6. https://doi.org/10.1128/IAI.66.5.2383-2386.1998.
58. Gilchrist CA, Baba DJ, Zhang Y, et al. Targets of the Entamoeba histolytica transcription factor URE3-BP. PLoS Negl Trop Dis. 2008;2:e282. https://doi.org/10.1371/journal.pntd.0000282.

59. Gilchrist CA, Leo M, Line CG, et al. Calcium modulates promoter occupancy by the Entamoeba histolytica Ca2+-binding transcription factor URE3-BP. J Biol Chem. 2003;278:4646–53. https://doi.org/10.1074/jbc.M211271200.
60. Gilchrist CA, Moore ES, Zhang Y, et al. Regulation of virulence of Entamoeba histolytica by the URE3-BP transcription factor. mBio. 2010;1 https://doi.org/10.1128/mBio.00057-10.
61. Huerta M, Reyes L, García-Rivera G, et al. A noncanonical GATA transcription factor of *Entamoeba histolytica* modulates genes involved in phagocytosis. Mol Microbiol. 2020;114:1019–37. https://doi.org/10.1111/mmi.14592.
62. Bello F, Orozco E, Benítez-Cardoza CG, et al. The novel EhHSTF7 transcription factor displays an oligomer state and recognizes a heat shock element in the Entamoeba histolytica parasite. Microb Pathog. 2022;162:105349. https://doi.org/10.1016/j.micpath.2021.105349.