Targets of the *Entamoeba histolytica* Transcription Factor URE3-BP

Carol A. Gilchrist¹*, Duza J. Baba¹, Yan Zhang², Oswald Crasta², Clive Evans², Elisabet Caler³, Bruno W. S. Sobral², Christina B. Bousquet¹, Megan Leo¹, Ameilia Hochreiter¹, Sarah K. Connell¹, Barbara J. Mann¹,⁴, William A. Petri Jr¹,⁴,⁵

1 Department of Medicine, University of Virginia, Charlottesville, Virginia, United States of America, 2 Virginia Bioinformatics Institute, Blacksburg, Virginia, United States of America, 3 J. Craig Venter Institute, Rockville, Maryland, United States of America, 4 Department of Microbiology, University of Virginia, Charlottesville, Virginia, United States of America, 5 Department of Pathology, University of Virginia, Charlottesville, Virginia, United States of America

Abstract

The *Entamoeba histolytica* transcription factor Upstream Regulatory Element 3-Binding Protein (URE3-BP) is a calcium-responsive regulator of two *E. histolytica* virulence genes, *hgl* and *fdx*1. URE3-BP was previously identified by a yeast one-hybrid screen of *E. histolytica* proteins capable of binding to the sequence TATTCTATT (Upstream Regulatory Element 3 (URE3)) in the promoter regions of *hgl*5 and *fdx*1. In this work, precise definition of the consensus URE3 element was performed by electrophoretic mobility shift assays (EMSA) using base-substituted oligonucleotides, and the consensus motif validated using episomal reporter constructs. Transcriptome profiling of a strain induced to produce a dominant-positive URE3-BP was then used to identify additional genes regulated by URE3-BP. Fifty modulated transcripts were identified, and of these the EMSA defined motif T[agt][T][T][G][T][G] was found in over half of the promoters (54% p<0.0001). Fifteen of the URE3-BP regulated genes were potential membrane proteins, suggesting that one function of URE3-BP is to remodel the surface of *E. histolytica* in response to a calcium signal. Induction of URE3-BP leads to an increase in tranwell migration, suggesting a possible role in the regulation of cellular motility.

Introduction

The early branching eukaryote *Entamoeba histolytica* is a human parasite that is the etiologic agent of amebic dysentery and liver abscess. Only one of every five infections leads to disease [1], and the parasite and host factors that control the outcome of infection are not well understood.Alteration in transcription of certain crucial genes may contribute to the expression of a virulence phenotype. Distinct gene expression profiles which may be associated with pathogenicity have been identified by comparing transcriptionomes of laboratory-cultured HM-1:IMSS *E. histolytica* to trophozoites growing in vivo, as well as that of less virulent strains and recent clinical isolates [2,3,4,5,6].

Here we have attempted to study the molecular mechanisms involved in the transcriptional regulation of virulence in *E. histolytica* by investigating further the role of the upstream regulatory element 3-binding protein (URE3-BP) transcription factor.

URE3-BP is a calcium regulated transcription factor, that is known to bind to the URE3 motif and thereby modulate transcription of both the Gal/GalNAc-inhibitable lectin *hgl*5 and ferredoxin 1 (*fdx*) genes. Mutation of the URE3 motif within the *hgl*5 and *fdx*1 promoter led to a four-fold rise and a two-fold drop in gene expression respectively, indicating that URE3 may function as a repressor or activator depending on context [7,8].

Previously a yeast one hybrid screen was used to identify an *E. histolytica* cDNA encoding a protein (URE3-BP) that recognized the URE3 DNA motif [9]. The URE3-BP protein was present in the *E. histolytica* nucleus and cytoplasm with an apparent molecular mass of 22.6 KDa. Two EF-hand motifs were identified in the amino acid sequence of URE3-BP. Binding of URE3-BP to the URE3 motif was inhibited in vitro by addition of calcium. Mutation of the second EF hand motif in URE3-BP resulted in the loss of calcium inhibition of DNA binding, as monitored by an electrophoretic mobility shift assay. Chromatin immunoprecipitation experiments confirmed the calcium-dependent interaction of URE3-BP with both the *hgl*5 and *fdx*1 promoter DNA [10].

Because the Gal/GalNac inhibitable lectin is an important virulence factor of *E. histolytica* it may be coordinately regulated at the transcription level with other virulence genes. In this light, it was intriguing that the mRNA of URE3-BP was down regulated [2]. The discovery of direct downstream targets of URE3-BP therefore may identify other genes important in *E. histolytica* pathogenesis and help delineate molecular and cellular mechanisms involved in the expression of virulence.

Position-specific variability in the sequence of transcription factor binding sites renders recognition of valid targets by computational methods alone extremely challenging [11,12]. Most work has been performed in the yeast model organism or the well-studied human transcriptome. The parameters affecting...
transcription regulation in early branching eukaryotes are only beginning to be deciphered [4, 13, 14, 15, 16, 17].

The sequencing of the \textit{E. histolytica} genome identified homologues of most of the RNA polymerase II subunits [18, 19], however the structure of \textit{E. histolytica} core promoter varies from the conventional norm by containing a third regulatory sequence GAAC in addition to the TATA box and INR. This may have an unpredictable impact on the machinery necessary for regulation of transcription [7, 20]. A bioinformatics approach was used by Hackney \textit{et al} to correlate potential \textit{E. histolytica} DNA motifs with high and low gene expression [21]. In our study we have focused on using not only computational but also experimental approaches to discover the gene regulatory network of the \textit{URE3} BP transcription factor.

To identify the consensus binding site sequence, a position weight matrix (PWM) of transcription factor binding to the \textit{URE3} motif was developed. To test the validity of the matrix, selected mutants within the \textit{URE3} motif of the \textit{hgl5} promoter were assessed for promoter activity in an episomal reporter construct. Finally, to identify additional genes regulated by \textit{URE3} BP, genome-wide expression profiling of transcripts from strains over-expressing a calcium insensitive \textit{URE3} BP mutant was performed.

**Methods**

**Cultivation of \textit{E. histolytica} and Nuclear Extract Preparation**

\textit{E. histolytica} strain HM1:IMSS trophozoites were grown at 37°C in TYI-S-33 medium containing penicillin (100 U/ml) and streptomycin (100 µg/ml) (GIBCO/BRL) [22]. Amebae in logarithmic phase growth (~6×10^6) trophozoites/ml) were used for nuclear extract preparation. Crude nuclear extracts were prepared by the method previously described [8, 23] with the following modifications: the protease inhibitors 2 mM (2S,3S)-trans-epoxysuccinyl-L-leucylamido-3-methylbutane and 2 mM 4-(2-aminoethyl) benzenesulfonylfluoride, HCl were added to both cell and nuclear lysis buffers, and dithiothreitol was omitted from the nuclear lysis buffer.

**Transient and stable transfection of \textit{E. histolytica} trophozoites**

Stable transfection of \textit{E. histolytica} trophozoites was achieved by use of the previously described lipofection technique [24, 25]. Briefly, amebaee were washed and suspended (2.2×10^7) amebaae per ml) in Medium 199 (Invitrogen, CA) supplemented with 5.7 mM cysteine, 1 mM ascorbic acid, 25 mM HEPES pH 6.8 (M199) 3 µg of DNA and 15 µl of Superfect (Qiagen) was added. Treated amebaee were left for 3 hours at 37°C, then growth media was added, and incubation at 37°C was continued overnight. The expression of all the recombinant proteins was confirmed by western blotting.

**Electrophoretic Mobility Shift Analysis**

\textit{URE3}-BP, has been shown to bind specifically to the TATTCTATT (\textit{URE3}) DNA motif in Gilchrest \textit{et al} 2001 [9]. In these conditions antibodies raised against \textit{URE3}-BP blocked the formation of the \textit{URE3} DNA-protein complex by native nuclear extracts and competition with a 60 fold excess of the nonspecific oligonucleotide (Olig-1) did not interfere with the formation of the specific complex. EMSA assays were performed with a Klenow-radiolabeled double stranded DNA oligonucleotide that spans the \textit{URE3} motif within the \textit{hgl5} promoter TGTTCCAAAAAGATATTTCTATTGAAAAATAAAAAGAG (\textit{hgl5}-\textit{URE3}). The protein-DNA interaction occurred in band shift buffer (10 mM Tris-HCl [pH 7.9], 50 mM NaCl, 1 mM EDTA, 0.05% nonfat milk powder, 3% glycerol, 0.05 mg of bromophenol blue) to which 0.2 µg of poly(dIdC), 10 fmol of DNA probe, and 2 µg of nuclear extract were added. The reaction mixture was allowed to incubate at room temperature (20°C) for 1 h prior to electrophoresis on a nondenaturing polyacrylamide gel for 2 to 3 h. The gel was then fixed and dried, and the signal from the protein-DNA complex was quantitated after exposure of the gel to a phosphorimage screen as described previously [8]. A ten fold or six fold excess of either cold \textit{hgl5}-\textit{URE3} (wt) or oligonucleotides wherein a base pair alteration within the \textit{URE3} motif had been made were added to the assay and the amount of competition was quantitated using a PhosphorImager. A double stranded oligonucleotide (Olig1) with the sequence AGAAAAAGCTATATTGACGTCA was used as an irrelevant control. Experiments were performed in triplicate, gels scanned (Molecular Dynamics, Model 425) and relative density of the EMSA assessed by use of the ImageQuant program (IQMac v1).

**Stable and Inducible expression vectors**

The stable construct (pHTP.luc) contained the luciferase structural gene under the control of the \textit{E. histolytica} \textit{hgl5} gene [26]. The promoter was mutated at the \textit{URE3} motif as described in results. Inducible vectors were based on the tetracycline inducible gene expression system of Ramakrishnan \textit{et al}. [27]. An N-terminal myc
tag was introduced by the amplification using the oligonucleotide TGCTGATCAATATGAGGAAACAAATTAATTTGAGAGAAA-GATTTA-ATGCAACCAGCTGTAGCTAATTTCC, and a control generated using an oligonucleotide that incorporated two stop codons directly after the myc tag (CTTGTATTTAACAATAGCG-TAAGCATG). Both amplicons were subcloned into the pCR2.1 TOPO expression vector (Invitrogen) and sequenced to confirm the presence of the desired mutations. The DNAs were then subcloned into the tetracycline-inducible gene expression system.

RNA isolation
One ml of Trizol (Invitrogen) was added to 2×10⁶ amebae collected by centrifugation at 900 rpm for 5 min and an initial RNA preparation performed according to the manufacturer’s directions. RNA greater than 200 nucleotides in length was separated from total RNA by the RNeasy protocol (Qiagen). RNA preparation performed according to the manufacturer’s recommendations. All real time amplification reactions were performed in triplicate and the resulting fluorescent values averaged. In all experiments utilizing qRT-PCR the cycle threshold values (CT, the cycle number at which fluorescence exceeds the threshold value) were linked to the quantity of initial DNA after calibration of the effectiveness of the amplifying primer pair [29]. The relatively invariant lgl transcript was used to compensate for the variation in the amount of amebic mRNA isolated.

Hybridization of sample to the Affymetrix E_his-1a520825 custom array
Quality control of RNA samples was performed by use of the Agilent Bioanalyzer Nano Assay. The standard protocol for hybridization of eukaryotic mRNA to Affymetrix arrays was followed (http://www.affymetrix.com/support/technical/manual/expression_manual.affx). Two micrograms of total RNA was used for cDNA and subsequent biotinylated cRNA synthesis. This labeled RNA probe was hybridized to the Affymetrix custom array designed using information generated from the E. histolytica genome sequencing project release date 12/08/04 as previously described [2,18]. The affymetrix probes were mapped to the new Genome Assembly and recognized 6385 of the reannotated open reading frames (78% of E. histolytica Open Reading Frames (ORF) 8197 http://pathema.tigr.org/). The ORF probe sets were preferentially selected from the 600 bases proximal to the 3’ end of the E. histolytica sequences. The arrays were scanned with an Affymetrix Gene Chip scanner 7G and report files were generated to determine the percentage of present calls of each array. The detection calls (present, marginal, absent) for each probe set were obtained using the GCOs system (http://www.affymetrix.com/products/software/specific/gcos.affx). Only genes with at least one “present” call were used in assessment of the data. Raw data from the arrays were normalized at probe level by the gcRMA algorithm and then log2 transformed [30].

Table 1. Presence of the URE3 matrix in the promoters of genes modulated by a dominant positive URE3-BP

| Transcripts significantly modulated by URE3-BP | Promoters which did not contain a URE3 matrix | Promoters containing a URE3 matrix 375-25 bases 5’ of ATG start codon | Total Number of Promoters analyzed | % | Statistical significance |
|-----------------------------------------------|-----------------------------------------------|--------------------------------------------------|----------------------------------|---|--------------------------|
| All E. histolytica promoters                   | 23                                            | 27                                               | 50                               | 54 | p<0.0001                 |
| Promoters with multiple URE3                   |                                               |                                                  |                                  |    |                         |
| Transcripts significantly modulated by URE3-BP | 6527                                          | 1985                                             | 8522                             | 23 |                          |
| All E. histolytica promoters                   |                                               |                                                  |                                  |    |                         |
| Motif Frequency in Promoters                   |                                               |                                                  |                                  |    |                         |
| Transcripts significantly modulated by URE3-BP | 7862                                          | 660                                              | 8522                             | 8  |                          |
| All E. histolytica promoters                   |                                               |                                                  |                                  |    |                         |
| A contingency table χ² test was used to compare the occurrence of the URE3 motif in transcripts significantly modulated by URE3 (the numbers were to large for the Fishers exact test). The motif background in all E. histolytica promoters was determined using a custom motif search script. To clearly illustrate the data we show the analysis on a promoter basis, and motif frequency per sequence nonamer. doi:10.1371/journal.pntd.0000282.t001
Figure 1. URE3 Matrix Discovery. (A) Representative electrophoretic mobility shift assay (EMSA) performed with radioactively labeled hgl5-URE3 double-stranded DNA. The lanes with probe alone are indicated; all other reactions included 2 μg of E. histolytica nuclear extract. EMSA’s were performed with a Klenow-radiolabeled double stranded DNA oligonucleotide that spanned the URE3 motif within the hgl5 promoter TGTTCCAAAAAGATATTCTATTGAAAATAAAGAAG (hgl5-URE3). A ten fold excess of either cold hgl5-URE3 (wt), or an oligonucleotide with a base pair substitution within the URE3 motif, was used as a competition to the wild type oligonucleotide. These are indicated by position and base substitution (i.e. T4C indicates that the T at position 4 in the URE3 motif [TATTCTATT] was changed to a C). The image was generated with a PhosphorImager (Molecular Dynamics model 425) in conjunction with the Adobe PhotoShop software program. (B) DNA-binding profile of URE3-BP derived from the EMSA results. The intensity of an irrelevant control was set as 100% and competition of the wild type oligonucleotide set as 0% (y axis). The position and base changes in the competing oligonucleotides T1A2T3T4C5T6A7T8T9 are shown on the x axis. The results of three independent replicates were averaged and are shown as a mean with standard error. (C) Graphical representation of the URE3 consensus sequence. The percent contribution of each base to the total competition occurring at each position (from each of the four bases) was calculated and shown graphically using the sequence logo program of Crooks et al. doi:10.1371/journal.pntd.0000282.g001
Statistical analysis

Microarray data analysis was performed using the Array Data Analysis and Management System (VBI) (http://pathport.vbi.vt.edu/main/microarray-tool.php). The system uses publicly available tools such as Bioconductor [31] for analysis of the data. Briefly, statistical significance was determined for the microarray data using the Linear Models for Microarray Data (LIMMA) program as described in the results section [32,33]. The statistical significance p values were corrected using the Benjamini and Hochberg false-discovery-rate test (FDR=0.05) [34]. Our comparisons were both between the two strains, and between different time points giving us potentially three control conditions. The most comprehensive comparison was between the test and control strains at 9 h post-induction. Statistical significance was determined for the qRT-PCR results using the students T test and the non-parametric Kruskal-Wallis Test was used to determine significance in the reporter gene assays. URE3 associated promoters were compared to the frequency of motif appearance in all E. histolytica promoters using the chi-squared test (InStat 2.03 program (GraphPad Software)).

Transwell migration assays

Transwell migration assays were performed using 5 mm transwell inserts (8 μm pore size Costar) suspended by the outer rim within individual wells of 24-well plates. Briefly, ameba trophozoites were incubated in serum free growth media containing 2 μg/ml CellTracker Green CMFDA (Molecular Probes) for 1 h [35]. Trophozoites were then washed and suspended at a concentration of 2×10^5/ml in serum free media and 500 μl loaded into the upper chamber. The plates were then placed in anaerobic bags (GasPak 100 Anerobic system; BD Biosciences) and incubated at 37°C for 3 h. Inserts and media were removed and fluorescence measured using a SpectraMax M2 fluorescent plate reader. Fluorescence versus concentration for each sample was determined by using a standard curve. Ameba numbers confirmed in selected experiments by microscopic counting and by use of the Techlab E. histolytica II antigen test used according to the manufacturer’s directions.

Results

URE3 Matrix

Electrophoretic mobility shift analysis (EMSA) was used with base substituted oligonucleotides to define the consensus URE3 motif. The impact of adding an excess of a non-radioactive oligonucleotide with a base pair alteration within the URE3 motif T1A2T3T4C5T6A7T8T9 was measured. A representative gel showing competition with the motif modified at positions 1 (AATTCTATT, GATTCTATT, CATATTATT) or 4 (TATTTATT, TATTTATT, TATTTATT) is shown in Figure 1A. The efficacy of a substituted base in competition assays was compared to the wild type motif (100%) and an irrelevant control (0%), as shown in Figure 1B. The percent contribution of each base to the total competition occurring at each position (from each of the four bases) was then calculated and is shown graphically in Figure 1C. The consensus URE3 motif incorporated base substitutions that maintained at least 15% competition of the gel shifts. The prototypic URE3 motif T1A2T3T4C5T6A7T8T9 as a result was modified to a consensus motif of T1[attg]2T3[gc]2T4[gt]2T5[gt]2. This modified sequence of the lower sensitivity of the episomal reporter assays, the wild type oligonucleotide. We interpreted this as a consequence of the lower sensitivity of the episomal reporter assays, likely due to over-expression of episomal constructs.

Microarray analysis

To further evaluate the physiological relevance of the URE3 matrix, a calcium-insensitive mutant of URE3-BP (EF(2)mutURE3-BP) (Figure 3A), and therefore constitutively active, was inducibly expressed and the changes in gene expression measured by use of an Affymetrix custom array (E_his-1a520285). The array included probes to 6,385 E. histolytica ORFs. Total RNA (12 μg) was isolated before induction (–Tet) and after 9 h of induction (+Tet) from cells carrying the myc-tagged recombinant URE3-BP mutant or the control construct (containing a stop codon immediately after the N terminal myc tag). The expression of the mRNA encoding the recombinant calcium-insensitive dominant positive mutant URE3-BP was induced 10–15 fold at nine hours post induction as indicated by myc specific qRT-PCR (Figure 3B). A western blot of E. histolytica nuclear and cytoplasmic proteins, probed with a myc-specific antibody, confirmed the cytosolic and nuclear distribution of both wild type and recombinant protein (Figure 3C). A calcium insensitive EMSA with hgl5-URE3 occurred only in nuclear extracts prepared from EF(2)mutURE3-BP transformed trophozoites (Figure 3D). In low calcium conditions EF(2)mutURE3-BP and STOP-EF(2)mutURE3-BP had equivalent URE3 binding capacity (data not shown).
Stranded DNA. Other than the lane with probe alone, reactions included 2 μg of *E. histolytica* nuclear extract prepared from either induced trophozoites carrying EF(2)mutURE3-BP or as a control STOP-EF(2)mutURE3-BP. A six fold excess of either cold hgl5-URE3 (wt), or an oligonucleotide with a base pair change which substituted a G for a T at the first position and had no impact on URE3-BP specific band formation (T,G mut) were added as shown.

doi:10.1371/journal.pntd.0000282.g003

### Statistical analysis of microarray data

The complete microarray data (deposited in NCBI’s Gene Expression Omnibus [36] and accessible through GEO Series accession number GSE12188 [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12188]) was normalized using gcRMA and statistical significance determined by LIMMA statistical analysis (Table S2). A total of fifty mRNAs were increased or decreased ≥2-fold at 9h post-induction compared to the induced control strain in which an N-terminal stop codon was present in the EF(2)URE3-BP sequence (Figure 4). The filtered transcripts had a normalized signal intensity of >50 in at least one microarray experiment, a change of greater than 2 fold, and were statistically significant by LIMMA.

### Analysis of modulated transcripts

A total of fifty mRNAs were increased (8) or decreased (42) ≥2-fold at 9h post-induction compared to the induced “control stop” strain, which had a stop codon inserted after the sequence encoding the myc tag. To identify the novel URE3-BP regulated genes, the promoters of transcripts significantly modulated by two-fold or greater were scored for the presence or absence of the URE3 matrix. The DNA Pattern Find program [http://bioinformatics.org/sms/] was used to locate the URE3 matrix in putative promoters of URE3-BP responsive genes (Figure 4 and Table 2). In cases where the probe set represented a ‘family’ of highly similar transcripts the probe set was scored positive if any of the promoters contained a URE3 motif. The three family probe sets are indicated in Table 2.

The URE3 matrix was found in 23% of all predicted promoter regions, however the matrix appeared at a statistically greater frequency (54%) in the URE3-BP modulated transcripts predicted by LIMMA (chi-square test p > 0.0001). Alternative analysis using the motif frequency or requiring the presence of two or more motifs for the positive designation also confirmed the correlation between the URE3 motif and transcript modulation (Table 1). The presence of the URE3 motif in the 3′ UTR regions was not above background values. The breakdown of the motifs found in the promoters of putative URE3-BP targets is shown in Table 2 and a graphical representation of the observed URE3 motifs is shown in Figure 4C. The sequence consensus of the URE3 motifs found 5′ of the modulated transcripts displayed only A or T residues at motif positions 2 and 7. While positions 2 and 7 were found to be the least conserved positions in the URE matrix consensus the predominant substitution of A/T may be a reflection of the AT bias of the *Entamoeba* genome. The other predominant change was a G substitution at position five which was half as effective as the wild type motif in EMSA assays (C5G). InterPro was used to scan the open reading frames of the significantly modulated genes to obtain additional information on protein function, TMpred to predict transmembrane regions, big-PI Predictor to identify Glycosylphosphatidylinositol (GPI) anchored proteins (GPI-anchor) and SignalP to identify signal peptides [37,38,39]. Sequences 150 bp 5′ and 3′ of the annotated ATG start codons were also checked and any additional in-frame peptides also examined for the presence of a signal peptide. On the basis of this information, the majority of the transcripts (47 of 50) could be subdivided into four categories: membrane proteins,
metabolism, cytoskeleton, and transcription & translation. The URE3 associated transcripts are shown in Table 2.

A gene was assigned to the membrane encoding group on the basis of the annotated GO term, the presence of a signal peptide, a GPI-anchor signal, or transmembrane domain. The majority of the membrane gene promoters contained a URE3 matrix (73%\(p<0.0001\)).

The encoded membrane proteins were quite distinct at the protein level. However, a subgroup of these proteins had highly similar promoter, and amino- and carboxyl-terminal sequences (sites of signal peptide and transmembrane domains) (Figure 5). With one exception (EHI_163360), the predicted sizes, pI, and length of the proteins were also quite similar (molecular mass between 29 to 47 kDa, and pI 4.3 to 5.5). In addition, all these
Table 2. URE3 motif in promoters of genes modulated by a dominant positive URE3-BP

| Pathema Loci ID | Annotation                                                                 | Fold Change | Position of URE3 5’ of ATG | Location of Other URE3 motifs | Motif       | Orientation 1 |
|----------------|---------------------------------------------------------------------------|-------------|-----------------------------|-------------------------------|-------------|--------------|
| EHI_062960     | membrane protein, receptor-mediated transport                              | 4.19        | 133                         | 124                           | TATTCTTTT  | –            |
| EHI_010850     | EhCP-A2 &7                                                                | 2.38        | 275                         | TATTCTATT                    | +           |              |
| EHI_046650     | membrane protein, similar to Gal/GalNAc lectin heavy subunit              | –2.01       | 134                         | 113                           | Tt/gTTCATT | –            |
| EHI_166370     | membrane protein                                                          | –2.07       | 182                         | 35                            | TATTCTATTg | –            |
| EHI_183210     | membrane protein                                                          | –2.07       | 209                         | TTTCTATT                     | –           |              |
| EHI_067910     | Competence protein ComEC                                                 | –2.16       | 113                         | TATTCTTT                     | +           |              |
| EHI_084730     | multidrug resistance-associated protein                                    | –2.28       | 147                         | TATTCTTT                     | +           |              |
| EHI_120930     | transmembrane protein kinase                                               | –2.28       | 125                         | 140                           | TATTC/gTATT | +           |
| EHI_057430     | membrane protein, surface antigen arielt family                           | –2.47       | 46                          |                               | TgTTTATT   | –            |
| EHI_145850     | membrane protein                                                          | –2.61       | 125                         |                               | TATTGTATT  | –            |
| EHI_103900     | membrane protein, nucleosome-binding protein 1                            | –3.3        | 46                          | 127                           | TA/tTTGATT | –            |
| EHI_132250     | membrane protein                                                          | –3.53       | 118                         |                               | TATTGTATT  | –            |
| EHI_114000     | membrane protein                                                          | –5.26       | 46                          | 127                           | TA/tTTGATT | –            |
| EHI_109010     | membrane protein                                                           | –5.67       | 46                          |                               | TTTCTATT   | –            |
| EHI_146100     | membrane protein                                                          | –6.11       | 46                          |                               | TTTCTATT   | –            |
| EHI_163360     | membrane protein                                                           | –6.56       | 46                          |                               | TATTGTATT  | –            |
| EHI_154330     | Calponin homology domain                                                   | –3.38       | 261                         | 343                           | TATTCTTTT  | –            |
| EHI_065250     | membrane protein, Lecithin:cholesterol acyltransferase                     | 5.25        | 106                         |                               | TATTCTTTT  | –            |
| EHI_079300     | acyl-CoA synthetase                                                        | 2.98        | 362                         |                               | TATTCTATT  | +           |
| EHI_092490     | Protein with a weak similarity to sulfotransferases2,7                    | 2.58        | 78                          | 217                           | TATTC/gTTT | +           |
| EHI_090430     | Protein with a weak similarity to sulfotransferases2,7                    | 2.52        | 223                         |                               | TTTCTATT   | –            |
| EHI_005060     | Fe-hydrogenase                                                            | 2.35        | 140                         |                               | TTTCTATT   | –            |
| EHI_185240     | long-chain-fatty-acid–CoA ligase                                          | –2.43       | 157                         | 193                           | TTTCTATT   | –            |
| EHI_00480      | basic leucine zipper protein3                                              | 2.36        | 80                          | 89                            | TA/tTTGATT | –            |
| EHI_059690     | chromosome segregation COG 1196                                           | –2.33       | 51                          |                               | TATTCTTTT  | –            |
| EHI_158020     | transcription initiation factor IIIB chain BRF                             | –2.57       | 35                          |                               | TATTCTTTT  | –            |
| EHI_000780     | chromodomain-helicase-DNA-binding protein                                  | –2.94       | 128                         |                               | TATTCTTTT  | –            |

Identity (Pathema Locus number) and gene annotation of significantly modulated genes are shown in conjunction with the observed change between induced EF(2)mutURE3-BP and STOP-EF(2)mutURE3-BP strains. The distance between the URE3 motif and the presumed initiating ATG codon is shown as is the position of other potential URE3 motifs, the promoter consensus URE3 motif, and orientation.

1If the motif is present in the promoter in a 5’ to 3’ direction this is indicated by (+) and on the reverse strand by (–).
2These transcripts were decreased in recent clinical isolates [4]
3Was modestly induced by tetracyclin at below threshold values
4These transcripts were increased in recent clinical isolates [4]
5This probe set recognizes a gene shown to be up regulated in HM-1:IMSS compared to a Rahman strain [3]
6This gene was described by Nixon et al and has been shown to be up regulated in HM-1:IMSS compared to a Rahman strain [3,45]
7A signal peptide was found in either the 5’ extended open reading frame or 3’ of initiating ATG, the position of URE3 motif is shown from the newly designated initiating ATG
8Annotation on the basis of homology to known sulfotransferases
9This transcript was also significantly increased in the EF(2)mutURE3-BPuninduced ameba compared to the uninduced control.
proteins contained a hydrophobic domain at the carboxyl terminus, and an anterior potential GPI anchor cleavage/addition site [40].

Most of the promoters of the small group of genes encoding metabolic enzymes also contained a URE3 matrix (86%, p = 0.0001). The enzymes encoded by these genes were linked to phospholipid metabolism. The opposing regulation of two enzymes that catalyze the addition of Coenzyme A to fatty acids (EHI_079300 and EHI_185240) might reflect different substrate specificities of these enzymes [41]. Both could potentially use the fatty acids, which are produced as a consequence of the breakdown of phospholipids by phospholipid:diacylglycerol acyltransferase (PDAT) (Figure 6) [42]. No URE3 matrix was found upstream of the fourth transcript, fatty acid elongase (EHI_092190), which could also be potentially involved in this potential scavenger cell pathway.

EF(2)mutURE3-BP expression induced migration of amebic trophozoites

To determine whether URE3-BP regulated the promigratory effects of trophozoites, transwell migration assays were performed as described in materials and methods. A two fold increase in migrating trophozoites was observed when comparing ameba induced to express EF(2)mutURE3-BP to uninduced controls (p = 0.04) or to the induced control stop strain transfected with the construct STOP- EF(2)mutURE3-BP (p = 0.02) (Figure 7). No difference was observed in migration when uninduced or induced STOP- EF(2)mutURE3-BP were compared (data not shown).

Discussion

In this work the DNA consensus motif recognized by the URE3-BP transcription factor was experimentally defined, and
then used to identify a subset of E. histolytica transcripts modulated by inducible expression of URE3-BP. URE3-BP had previously been shown to regulate the expression of two virulence factors in the parasite. The current studies provide a more global picture of its role in control of gene expression. The key experimental approach was the inducible expression of a dominant positive URE3-BP mutant and the subsequent identification of uniquely altered transcripts. The majority (42/50) of transcripts were repressed. Over half (54%) of the modulated genes had a URE3 matrix in the promoter region while the other half was comprised of genes presumably downstream of control by URE3.

The URE3 matrix was present in the 5′ sequences of URE3-BP regulated genes involved in fatty acid metabolism and in potential membrane or secreted proteins. The latter suggests that phenotypic changes due to the expression of the dominant positive URE3-BP mRNA could occur most noticeably at the cell surface of E. histolytica trophozoites.

URE3-BP regulated genes, which encoded proteins with an N terminal signal peptide, included the potential virulence factor EhCP-A7, a cysteine protease, an asparagine-rich antigenic surface protein ariel [43,44], a novel lectin-like protein, and a subgroup of genes encoding potential surface proteins which appear to have highly conserved promoters and signal peptides. Most unusually the conservation in this group of potential surface proteins was greater at the DNA rather than the protein level. This may represent a gene duplication followed by functional divergence, or possibly a gene recombination event. A technical limitation of the gene expression analysis was the inability to measure transcript levels of the hgl5 and fdx1 genes that contain URE3 in their promoters, which cannot be distinguished from highly related gene family members that lack URE3-containing promoters. The hgl5 gene belongs to a family of five highly similar genes (up to 99%), and ferredoxin is encoded by two identical ORFs, fdx1, and fdx2 (unpublished data). The presence of the URE3 matrix was not much higher than background in the promoters encoding genes involved in either transcription/translation (25% p = 0.035) or cytoskeletal function (25% p = 0.73).

However while we could not demonstrate changes in the level of the ferredoxin transcript, a URE3 associated Fe-hydrogenase

**Figure 6. Modulation of Transcripts Encoding Enzymes Involved in Phospholipid Degradation and Fatty Acid Biosynthesis.** Transcripts significantly modulated by (EF2)mutURE3-BP are shown in shaded boxes, green indicates a down-regulated transcript and red an up-regulated transcript. As lecithin:cholesterol acyltransferase has homology with phospholipid:diacylglycerol acyltransferase the simpler pathway of triacylglycerol biosynthesis is shown although alternative pathways exist [42,54]. The locus number and the URE3 motif present within the promoter sequences follow the gene name.

doi:10.1371/journal.pntd.0000282.g006

**Figure 7. EF(2)mutURE3-BP expression induced migration of amebic trophozoites.** Expression of the EF(2)mutURE3-BP and STOP-EF(2)mutURE3-BP transcripts was induced by the addition of tetracycline. Analysis of trophozoite migration was then done by a transwell assay. Data shown are representative of assays using two independently transfected trophozoite lines with the relevant expression vectors in 4 separate experiments. The data are shown as mean ± SEM of the number of cells migrated measured using CellTracker Green CMFDA as described in Materials and Methods.

doi:10.1371/journal.pntd.0000282.g007

The URE3-BP transcript or on the genes involved in this pathway, suggesting the lack of feedback inhibition of URE3-BP from the products of this pathway [17].

A limitation of this study was that the microarray analysis measured the steady state mRNA levels and we therefore may have missed changes in newly transcribed RNA, especially for abundant transcripts. Changes occurring in mRNA stability and/or transcript processing may obscure changes occurring at the level of transcription [47,48]. A second limitation is that the high ‘background’ incidence of the URE3 motif (23%) in the promoters of all E. histolytica may indicate that there are other factors not yet identified involved in promoter specific recognition by URE3-BP. Because appreciable levels of wild type URE3-BP were still present, this might have contributed to the failure to observe changes in the roughly 2000 genes with putative URE3-BP binding sites for which no change was seen following induction of EF(2)mutURE3-BP. Because of these issues it is a reasonable conclusion that the 50 changed transcripts are an underestimate of the genes regulated by URE3-BP.

The URE3 matrix was absent in 23 of the regulated promoters. Amebae were harvested at nine hours after the addition of
tetracycline and shortly after appreciable induction of recombinant URE3-BP protein (Figure 3B). Therefore, it is possible that at this time point URE3-BP regulated transcripts may have in turn induced the expression of a set of secondary-response genes [49]. The URE3 associated EHL_004430 ORF encoding a protein with a basic leucine zipper domain, and the EHL_000780 transcript that encodes a potential chromodomain protein, could act as regulators of a secondary response. Among the modulated non-URE3 associated transcripts are members of the virulence associated EhSTIRP family [3,50] and cytoskeletal genes suggesting a potential involvement in attachment or motility [19]. The promigratory impact of URE3-BP overexpression shown in Figure 7 supported this correlation however identifying truly co-promigratory impact of URE3-BP overexpression shown in a potential involvement in attachment or motility [19]. The associated EhSTIRP family [3,50] and cytoskeletal genes suggest that encodes a potential chromodomain protein, could act as inducers of adhesion and chemotaxis of Entamoeba histolytica. BMC Genomics 8: 216.

Supporting Information

Table S1

| Table S1 | At: | Table S2 | At: |
|----------|-----|----------|-----|
|         | doi:10.1371/journal.pntd.0000282.s001 | (0.04 MB DOC) | doi:10.1371/journal.pntd.0000282.s002 | (7.70 MB XLS) |

Acknowledgments

We would like to thank Herman Lorenzi of the J. Craig Venter Institute for his valuable help in using the Pathema Database and Lauren A. Lockhart at the University of Virginia for her technical assistance. Pathema-\Entamoeba is an NIAID Bioinformatics Resource Center (BRC).

Author Contributions

Conceived and designed the experiments: CAG WAPJ. Performed the experiments: CAG DJB CE CBB ML AH SKC. Analyzed the data: CAG YZ EC BJM WAPJ. Contributed reagents/materials/analysis tools: OC BWSS WAPJ. Wrote the paper: CAG BJM WAPJ.

References

1. Haque R, Mondal D, Duggal P, Kabir M, Roy S, et al. (2006) Entamoeba histolytica infection in children and protection from subsequent amebiasis. Infect Immun 74: 904–909.
2. Glichrist CA, Hoopit E, Trangdaize N, Fei Z, Crasta O, et al. (2006) Impact of intestinal colonization and invasion on the Entamoeba histolytica transcriptome. Mol Biochem Parasitol 147: 163–176.
3. MacFarlane RC, Singh U (2006) Identification of differentially expressed genes in virulent and nonvirulent Entamoeba species: potential implications for amebic pathogenesis. Infect Immun 74: 380–391.
4. Ehrenkaufer GM, Haque R, Hackney JA, Eichinger DJ, Singh U (2007) Identification of developmentally regulated genes in Entamoeba histolytica: insights into mechanisms of stage conversion in a protozoan parasite. Cell Microbiol 9: 1426–1444.
5. Davis PH, Schuler J, Stanley SL, Jr (2007) Transcriptomic comparison of two Entamoeba histolytica strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. Mol Biochem Parasitol 151: 110–121.
6. Davis PH, Zhang X, Guo J, Townsend RR, Stanley SL, Jr (2006) Comparative proteomic analysis of two Entamoeba histolytica strains with different virulence phenotypes identifies peroxiredoxin as an important component of amoebic virulence. Mol Microbiol 61: 1523–1532.
7. Purdy JE, Pho LT, Mann BJ, Petri WA Jr (1996) Upstream regulatory elements controlling expression of the Entamoeba histolytica lectin. Mol Biochem Parasitol 78: 91–103.
8. Glichrist CA, Mann BJ, Petri WA Jr (1999) Control of ferredoxin and Gal/GalNAc lectin gene expression in Entamoeba histolytica by a cis-acting DNA-binding sequence. Infect Immun 66: 2383–2386.
9. Glichrist CA, Holm CF, Hughes MA, Schaeferman JM, Mann BJ, et al. (2001) Identification and characterization of an Entamoeba histolytica upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. J Biol Chem 276: 11838–11843.
10. Glichrist CA,ирующ, Hughes MA, Schaeferman JM, Mann BJ, et al. (2001) Identification and characterization of an Entamoeba histolytica upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. J Biol Chem 276: 11838–11843.
11. Kolchansow NA, Merkulova TI, Ignatieva EV, Ananko EA, Oshchepkov DY, et al. (2007) Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes. Brief Bioinform 8: 266–274.
12. Long X, Miano JM (2007) Remote control of gene expression. J Biol Chem 282: 35116–35121.
13. van Noort V, Huynen MA (2006) Combinatorial gene regulation in Plasmodium falciparum. Trends Genet 22: 73–78.
14. Hakimi MA, Deitsch KW (2007) Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation. Curr Opin Microbiol 10: 357–362.
15. Ramakrishnan G, Glichrist CA, Masa H, Torok MS, Grant PA, et al. (2004) Histone acetyltransferases and deacetylase in Entamoeba histolytica. Mol Biochem Parasitol 138: 205–216.
16. Romero-Diaz M, Gomez C, Lopez-Reyes I, Martinez MB, Orozco E, et al. (2007) Structural and functional analysis of the Entamoeba histolytica EhRabB gene promoter. Mol Biochem Parasitol 151: 87–97.
17. Ehrenkaufer GM, Eichinger DJ, Singh U (2007) Trichostatin A affects on gene expression in the protozoan parasite Entamoeba histolytica. BMC Genomics 8: 216.
18. Loftus B, Anderson I, Davies R, Alamark UC, Samuelsen J, et al. (2005) The genome of the protozoan parasite Entamoeba histolytica. Nature 435: 955–968.
19. Clark CG, Alamark UCM, Tazreiter M, Saito-Nakano Y, Ali V, et al. (2007) Structure and content of the Entamoeba histolytica genome. Advances in Parasitology, Vol 63: 51–190.
20. Singh U, Glichrist CA, Schaeferman JM, Rogers JR, Hockensmith J, et al. (2002) Context-dependent roles of the Entamoeba histolytica core promoter element GAAC in transcriptional activation and protein complex assembly. Mol Biochem Parasitol 120: 107–116.
21. Hackney JA, Ehrenkaufer GM, Singh U (2007) Identification of putative transcriptional regulatory networks in Entamoeba histolytica using Bayesian inference. Nucleic Acids Res 35: 2141–2152.
22. Diamond LS (1961) Azoxy cultivation of Entamoeba histolytica. Science 134: 330–337.
23. Gomez C, Perez DG, Lopez-Baygheen F, Lazarri PM, et al. (1997) Stable transfection of Entamoeba histolytica trophozoites by lipofection. Arch Med Res 28: 40–49.
24. Olvera A, Olova F, Vines RR, Reccillas-Targa F, Lizardi PM, et al. (1997) Stable transfection of Entamoeba histolytica trophozoites by lipofection. Arch Med Res 28: 40–49.
25. Ashparpour A, Glichrist C, Baba D, Hamano S, Houpf E (2005) Resistance to intestinal Entamoeba histolytica infection is conferred by innate immunity and Gr+ve cells. Infect Immun 73: 4522–4529.
26. Vines BR, Ramakrishnan G, Rogers JR, Lockhart LA, Mann BJ, et al. (1996) Regulation of adherence and virulence by the Entamoeba histolytica lectin cytoplasmic domain, which contains a beta2 integrin motif. Mol Biol Cell 9: 2099–2079.
27. Ramakrishnan G, Vines RR, Mann BJ, Petri WA Jr (1997) A tetracycline-inducible gene expression system in Entamoeba histolytica. Mol Biochem Parasitol 84: 93–100.
28. Koenen S, Skalskcy H (2000) Primer3 on the WWW for general and for biologist programmers. Methods Mol Biol 132: 356–366.
29. Gingringer DG, Geoffrey TE, Ngo J, Moore DH 2nd, Suzuki S, et al. (2000) Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis. Cancer Res 60: 3465–3469.
30. Inzirra RA, Hobbs B, Collins F, Beazer-Bradley YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.
31. Gentemen RC, Carey VJ, Bates DM, Bolstad B, Deftling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.
32. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116–5121.
33. Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21: 2067–2075.
34. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57: 289–300.
35. Franco E, Vazquez-Prado J, Meca I (1997) Fibronectin-derived fragments as inducers of adhesion and chemotaxis of Entamoeba histolytica trophozoites. J Infect Dis 176: 1397–1602.
35. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207–210.
36. Eisenhaber B, Bork P, Yuan Y, Löffler G, Eisenhaber F (2000) Automated annotation of GPI anchor sites: case study C. elegans. Trends Biochem Sci 25: 340–341.
37. Hofmann K, Stoffel W (1993) TMBase - A database of membrane spanning protein segments. 166 p.
38. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953–971.
39. McCoy JJ, Mann BJ, Verdick TS, Pak Y, Heimark DB, et al. (1993) Structural analysis of the light subunit of the Entamoeba histolytica galactose-specific adherence lectin. J Biol Chem 268: 24223–24231.
40. Knoll LJ, Johnson DR, Gordon JJ (1994) Biochemical studies of three Saccharomyces cerevisiae acyl-CoA synthetases, Faa1p, Faa2p, and Faa3p. J Biol Chem 269: 16348–16356.
41. Stahl U, Carlsson AS, Lenman M, Dahlqvist A, Huang B, et al. (2004) Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. Plant Physiol 135: 1134–1135.
42. Willhoeft U, Buse H, Tannich E (1999) DNA sequences corresponding to the ariel gene family of Entamoeba histolytica are not present in E. dispar. Parasitol Res 85: 787–789.
43. Mai Z, Samuelson J (1998) A new gene family (ariel) encodes asparagine-rich Entamoeba histolytica antigens, which resemble the anaebic vaccine candidate serine-rich E. histolytica protein. Infect Immun 66: 353–355.
44. Nixon JE, Field J, McArthur AG, Sogin ML, Yarlett N, et al. (2003) Iron-dependent hydrogenases of Entamoeba histolytica and Giardia lamblia: activity of the recombinant entamoeba enzyme and evidence for lateral gene transfer. Biol Bull 204: 1–9.
45. Grogan DW, Cronan JE, Jr (1984) Cloning and manipulation of the Escherichia coli cyclopropane fatty acid synthase gene: physiological aspects of enzyme overproduction. J Bacteriol 158: 286–295.
46. Lopez-Camarillo C, Luna-Arias JP, Marchat IA, Oseco E (2003) EhPgp5 mRNA stability is a regulatory event in the Entamoeba histolytica multidrug resistance phenotype. J Biol Chem 278: 11273–11280.
47. Davis CA, Brown MP, Singh U (2007) Functional characterization of spliceosomal introns and identification of U2, U4, and U5 mRNAs in the deep-branching eukaryote Entamoeba histolytica. Eukaryot Cell 6: 940–949.
48. Döllner NR, Sanders MM (2002) Upstream stimulatory factor (USF) is recruited into a steroid hormone-triggered regulatory circuit by the estrogen-inducible transcription factor delta EF1. J Biol Chem 277: 33890–33894.
49. Maclaran RC, Singh U (2007) Identification of an Entamoeba histolytica serine, threonine, isoleucine, rich protein with roles in adhesion and cytotoxicity. Eukaryot Cell.
50. Thomas R, Paredes CJ, Mehta S, Hatzimanikatis V, Papoutsakis ET (2007) A model-based optimization framework for the inference of regulatory interactions using time-course DNA microarray expression data. BMC Bioinformatics 8: 220.
51. Margolin AA, Califano A (2007) Theory and limitations of genetic network inference from microarray data. Ann N Y Acad Sci.
52. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14: 1188–1190.
53. Borggreve SE, De Vries R, Dullaart RP (2003) Alterations in high-density lipoprotein metabolism and reverse cholesterol transport in insulin resistance and type 2 diabetes mellitus: role of lipolytic enzymes, lecithin:cholesterol acyltransferase and lipoprotein lipase. Eur J Clin Invest 33: 1051–1069.