EXPLORING THE BIOCHEMICAL ROLES OF TbeIF4E-3 AND TbeIF4E-5 IN *T. brucei* CELL CYCLE

by

Jacob Peter Rubus

BS, Biological Sciences, University of Pittsburgh at Greensburg, 2008

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Jacob Rubus

It was defended on

April 14th, 2015

and approved by

Thesis Advisor: Salvador Tarun, Jr., PhD, Former Associate Professor, Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Co-Advisor: Todd Reinhart, ScD, Professor, Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Jeremy Martinson, PhD, Assistant Professor, Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Karen Norris, PhD, Professor, Department of Immunology, School of Medicine, University of Pittsburgh

Salvador Tarun, PhD

Todd Reinhart, PhD

EXPLORING THE BIOCHEMICAL ROLES OF TbeIF4E-3 AND TbeIF4E-5 IN *T. brucei* CELL CYCLE

Jacob Rubus, MS

University of Pittsburgh, 2015

ABSTRACT

African Sleeping Sickness afflicts millions of people in sub-Saharan areas of the continent. The disease causing agent, the unicellular eukaryotic protozoan *Trypanosoma brucei*, is intensely studied, yet no safe cure has been developed and has become a public health issue. Because gene expression control in T. *brucei* exhibits unique patterns, it is attractive for drug targeting. Our lab aims to explore the cellular and molecular role(s) of TbEIF4E-3 (4E-3p) and TbeIF4E-5 (4E-5p), structurally divergent homologs of eukaryotic translation initiation factor 4E (eIF4E), a canonical mRNA 5' cap binding protein involved in post-translational regulation of gene expression. The 4E-3 and 4E-5 proteins are essential in cell cycle in trypanosomes. We hypothesize 4E-3p and 4E-5p to be key regulons, proteins that regulate the translation of distinct sets of multiple, functionally related mRNAs required for normal cell cycle. Evidence from immunofluorescence microscopy (IMF) by other labs and our own revealed cytoplasmic localization of 4E-3p in mRNP granules. Tandem co-immunoprecipitation (co-IP) and mass spectrometry (MS) revealed 4E-3p association with canonical protein partners found in other eukaryotes known to be involved in translation initiation and control such as eIF4G and eIF4A, consistent with earlier studies. Significantly, this present work further identified unique uncharacterized polypeptides, implying both conserved as well as novel functions. Additionally, RNAseq analysis from co-IPs revealed that 4E-3p also associates with various mRNAs, some of

which are involved in the cell cycle, providing evidence towards potential translational regulation of cell cycle. By contrast, 4E-5p appeared more associated with the mitochondria from IMF studies, consistent with findings by others. Combined co-IP and MS analysis showed association with translational regulators, whereas RNAseq revealed association with mRNAs regulating cell division and replication control distinct from those associated with 4E-3p. These findings suggest that both 4E-3p and 4E-5p may act separately to regulate cell cycle, replication, and gene expression in *T. brucei* in a manner consistent with a proposed Parallel RNA Regulon Model. Among potential drug targets being analyzed are translational regulating proteins, of which 4E-3p and 4E-5p are attractive candidates due to their discovered role in normal cell cycle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank Dr. S. Tarun, Jr., formerly of the University of Pittsburgh Graduate School of Public Health, for providing me the opportunity to perform this research, as well as Joseph Rocco, from the University of Pittsburgh School of Medicine, and Dr. A. Tarun, formerly of the University of Pittsburgh Graduate School of Public Health, for assistance with procedures and equipment. Also I would like to acknowledge Dr. Manimalha Balasubramani and Dr. Janette Lamb from the Biomedical Mass Spectrometry Center at the University of Pittsburgh and the University of Pittsburgh Genomics and Proteomics Core, respectively, the mass spectrometry analysis on the samples and for continued assistance with understanding the results. The RNAseq could not have been performed if not for Rahil Sethi from the University of Pittsburgh Bioinformatics Core facility and his supervisor Dr. James Lyons-Weiler. Finally, I would like to thank the committee members who assisted with revisions upon revisions of the document, and assisting me in piecing together the *in silico* analyses after the loss of the lab: Dr. Todd Reinhart, Dr. Jeremy Martinson, and Dr. Karen Norris, all of the University of Pittsburgh Graduate School of Public Health.

The grants used to perform the experiment were the NIH K22 Grant and Pitt T2CP2 Pilot grant awarded to Dr. Salvador Tarun, Jr.

1.0 INTRODUCTION

Trypanosoma brucei is a pathogenic, unicellular eukaryotic protozoan of the kinetoplast group. These specific protozoans maintain only one flagellum and one large mitochondrion containing a kinetoplast, and are closely related to *Trypanosoma cruzi* and the *Leishmania* species. The three protozoans, called "TriTryps," are the cause of millions of infections in humans across the world (WHO 2008), mostly in tropical and sub-tropical areas. *T. brucei,* specifically, is the pathogen responsible for the Human African Trypanosomiasis (HAT), or more commonly called "African Sleeping Sickness" [3][4].

Infection occurs through the bite of a tsetse fly, where the parasite exists in the procyclic form (PF) in the fly's midgut. Upon entering the blood stream, the parasite enters the blood form (BF) stage and travels to the lymph nodes. The PF exhibits very few variable surface glycoproteins (VSGs), while the BF exhibits dense clusters of VSGs (**Figure 1**). The increase in VSGs assist in host immune system evasion, and have recently become attractive targets for control of infection[5][6]. From here the proliferating parasites traverses the blood-brain barrier, eventually causing death [5][6]. HAT has been known to cause symptoms ranging from headaches and joint pain, common to the Early Stage of infection, to more severe neurological disorders once the parasite crosses the blood-brain barrier, during the late stage of infection. Once this occurs, seizures often ensue, followed by coma and eventually death if not treated.

The late stage of infection is characterized by disruptions in sleep patterns. The symptoms do not always show up immediately, often not presenting themselves for months or even years after infection, hindering attempts to treat the disease before the onset of the late stage of infection [7].

HAT is endemic in low-income regions of sub-Saharan Africa, making the disease a neglected tropical disease [8-10]. The drugs developed to target HAT can be severe, in the case

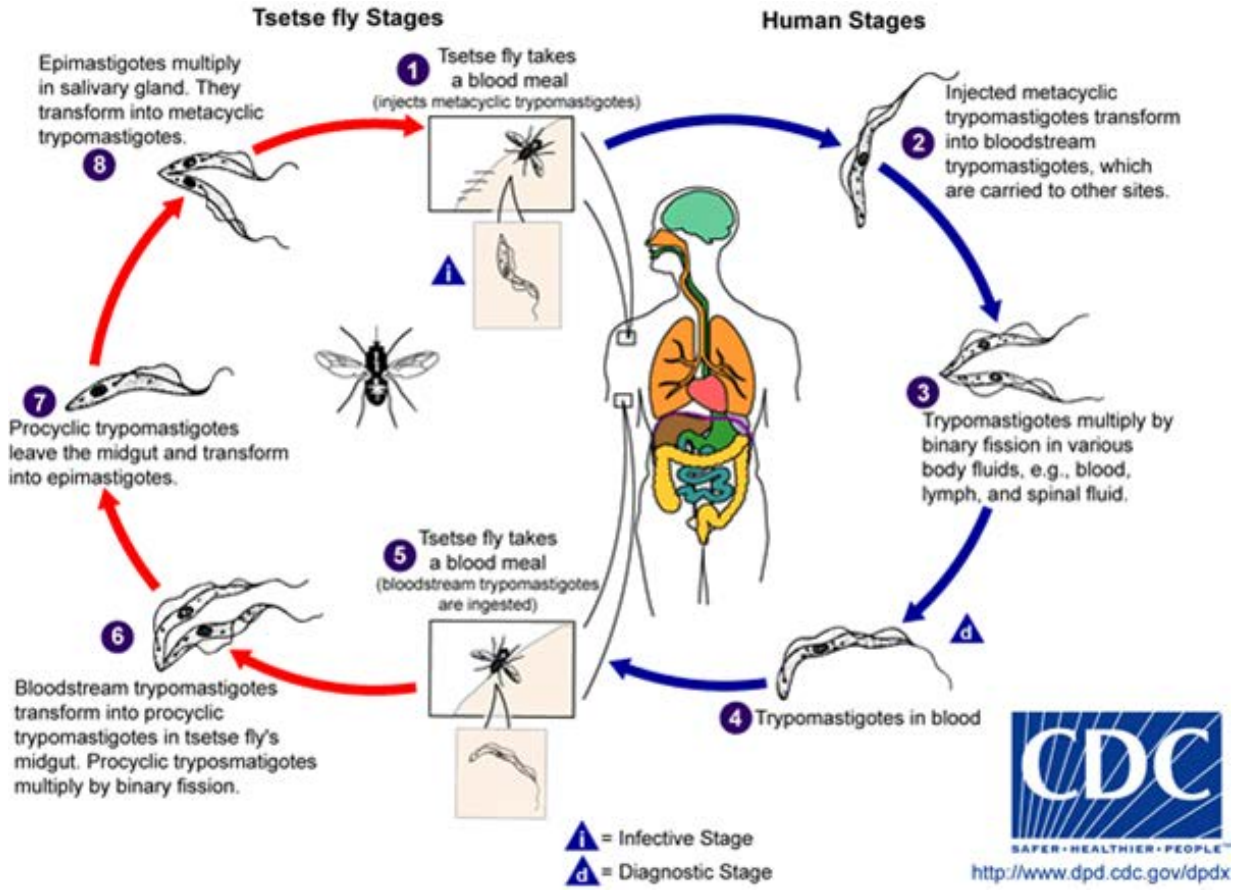


Figure 1. Life Cycle of *T. brucei*.

Procyclic forms of the parasite exist in the midgut of the tsetse fly, becoming the blood form upon entering the bloodstream of the host. This blood form stage exhibits clusters of variable surface glycoproteins (VSGs) that allow for immune system evasion. Adapted from http://www.cdc.gov/dpdx/trypanosomiasisAfrican/index.html

of melarsoprol, which exhibits acute toxicity due to it being an arsenic derivative, or requiring multiple doses, such as eflornithine, which is less severe than melarsoprol, exhibiting reversible

toxic effects but remaining harder to administer in rural Africa [11-15]. While these drugs were developed in the 20th century, with melarsoprol being discovered in 1949 and eflornithine being registered for use on patients with HAT in 1990, there have been only minor steps forward in new drug development, and now strains of trypanosomes are arising with drug resistance [11-13]. A newer drug, DB289, which is a novel aromatic diamidine, was in the final clinical stages, and, while purportedly safer, there is still a 10 day treatment required, a very difficult requirement in rural Africa. The latest data on DB289 now suggest severe liver toxicity and delayed renal toxicity, and research has since been discontinued. However, the aza analogs DB820 and DB829 are currently being investigated as well [16-18].

With the availability of the Tri-Tryps genome and proteome databases (tritrypsdb.org), the opportunity to find genes for new and better drug targets is wide-open. Proteins that act as translational regulators are now high on the list of possible drug targets due to their essential role in the normal cell cycle and their unique structural features, as described below.

1.1 EUKARYOTIC INITIATION FACTOR 4E (EIF4E)

Eukaryotes possess cap-binding proteins that bind to the 5' end of mRNAs and help initiate translation [19]. They are broadly referred to as eukaryotic initiation factor 4E (eIF4E) [20]. Because of the lack of transcriptional regulation in trypanosomes, the *T. brucei* eIF4E's (TbeIF4E's) are expected to play a more prominent role in controlling gene expression by regulating which mRNAs are translated at which time [1][21]. Indeed this 5' cap-binding protein has multiple homologs present in other eukaryotes, potentially allowing for a wide

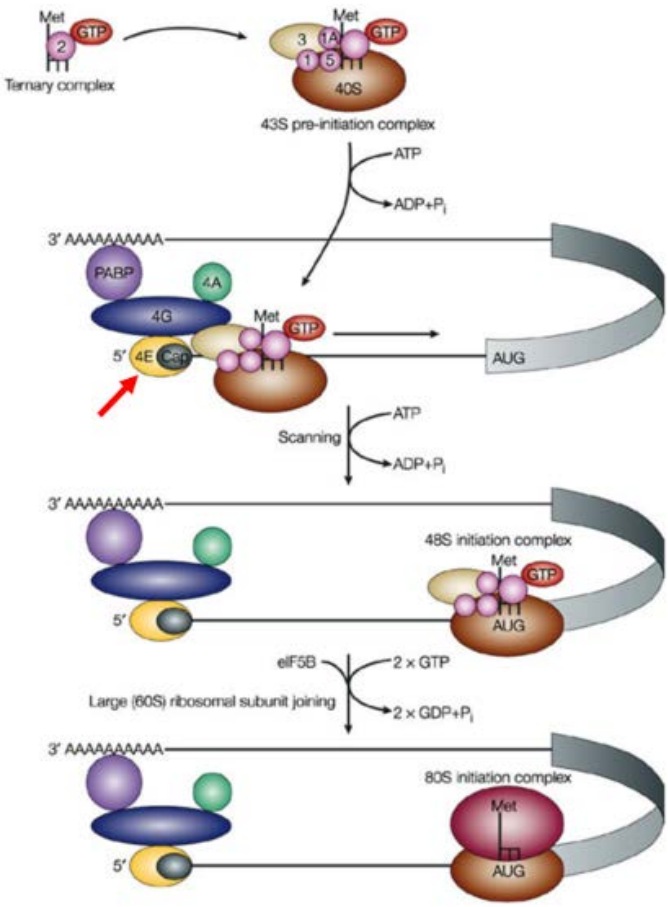


Figure 2. eIf4E-mediated cap-dependent translation initiation.

eIF4E is indicated by a red arrow. (Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology (Ref. 23), copyright (2004).

variety of functions through association with other distinct protein factors, ranging from transporting the processed mRNAs from the nucleus to mRNA localization, stability and translation initiation [22]. In the initiation of translation, the eIF4E (4E) homologs are a key component in the eIF4F (4F) complex that recruits the small ribosomal subunit to the capped mRNA (**Figure 2**) [23]. Aside from 4E, the 4F complex in eukaryotes consists of two other key

proteins, eIF4G (4G) and eIF4A (4A). 4G assists in the binding of other proteins to the 4F complex, such as the mRNA 3' polyA binding protein (PABP), 4A and other initiation factors (e.g. eIF3) that are bound directly to the 40S ribosomal subunit to initiate translation [24]. The most fundamental interaction of 4G is the direct binding of 4E, which is a critical step to allow 4E to initiate translation [25][26]. 4A is a RNA helicase that unwinds the secondary structures of the mRNA to permit 40S ribosomal subunit binding [24][27][28].

Depending on cell growth, 4E exists in a phosphorylated state [29][30]. When the cell is undergoing active growth and proliferation the amount of phosphorylated 4Es are low in proportion to levels of 4E while the cell is not undergoing active growth. Due to this phosphorylation in the cytoplasm by the target of rapamycin (TOR), related to the phosphatidylinositol 3-kinase (PI3K) family, 4E has been proposed to be a rate limiting step in translation initiation and cellular proliferation [19][30]. In mouse models, an overabundance of 4E causes excessive growth and division of cells, leading to the generation of tumors [31][32], while HeLa cells engineered to produce less 4E than controls were shown to have a division rate four to seven times slower [33]. Because of this, 4E plays a key role in translation regulation and is also designated as an oncogene. As such, 4E activities including cap-binding, localization, and 4F assembly are subject to the effects of many different stimuli and growth factors. From yeast to man, there are key activation pathways associated with both growth and stress signals, such as the mTOR and p38 pathways, involving phosphorylation of 4E and/ or other 4F components, leading to an increase in cell proliferation or even apoptosis [34].

1.2 T. BRUCEI EIF4E-3 (TBEIF4E-3) AND EIF4E-5 (TBEIF4E-5) PROTEINS

Similar to other eukaryotes, two of *T. brucei's* eIF4E orthologs, designated TbeIF4E-3 (4E-3) and TbeIF4E-5 (4E-5), were shown through RNAi knockdowns to be critical to cell viability [1] and proper cell division in trypanosomes, causing cell deformations and producing multiple or no kinetoplasts and nuclei [S.Tarun and J. Rocco, data not shown][2]. Partial knockdown of 4E-3 caused rapid loss of general translation based on ^{35}S-methionine incorporation, as well as rapid reduction of mRNA levels of selected cell-cycle related genes, suggesting a role for 4E-3 protein (4E-3p) in mRNA stability related to cell cycle process [2]. Consistent with a possible direct role in translation, trypanosome homologs of 4G and 4A proteins were shown to co-purify with exogenously expressed hemagglutinin (HA)-tagged 4E-3p [1]. 4E-3p shows homology with yeast 4Ep via conserved motifs in the 5' cap binding region. However, it is highly unusual

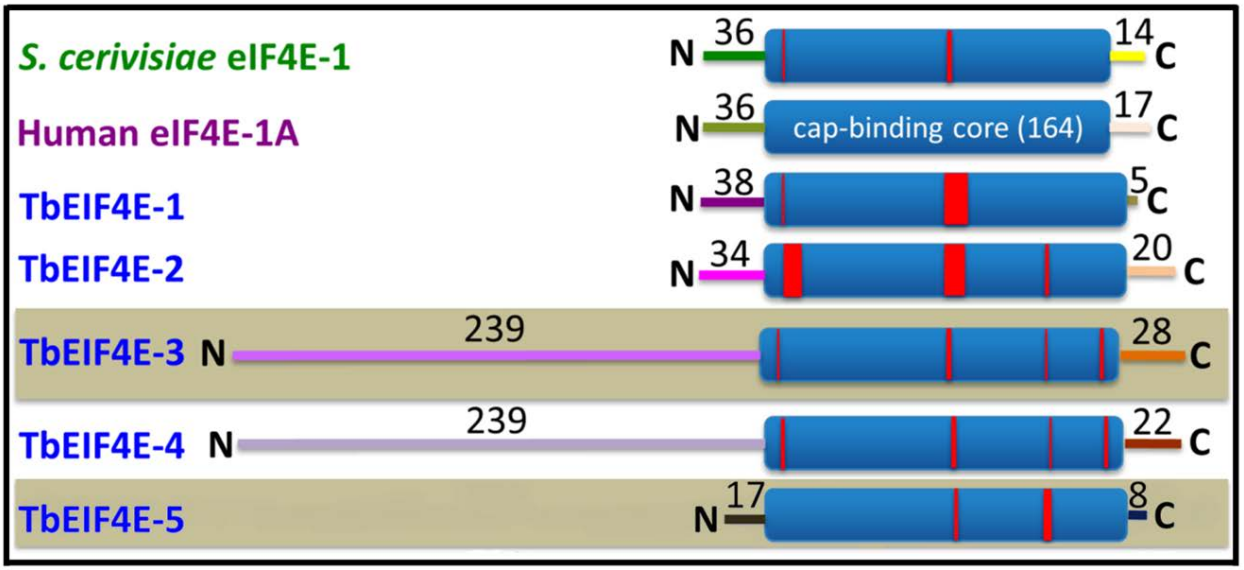


Figure 3. Comparison of 4E homology of cap-binding domains in the five orthologs in trypanosomes with human and yeast forms.

The conserved cap-binding cores are depicted in blue, with the species-specific polypeptide differences depicted in red. The *T. brucei* eIF4Es (TbeIF4Es) show homology with yeast eIF4Es based on these conserved cores. TbeIF4E-3 exhibits a long N-terminal region conserved amonst the TriTryps species and suggest possible new binding interactions.

because of a long N-terminal extension, making it two times larger than typical 4Es **(Figure 3)**. This unique feature is also conserved in related *T. cruzi* and *Leishmania* species [1][21][35], suggesting potentially new binding interactions, function or regulation for this 4E homolog. These properties could make 4E-3p a potential drug target against the parasite.

The protein TbeIF4E-5 (4E-5p) is another 4E homolog based on its conserved cap-binding domain **(Figure 3**). By RNAi knockdown, 4E-5p was also shown to be growth essential affecting cell morphology and cell cycle [2]. Intriguingly, it was identified as part of the *T. brucei* mitochondrial proteome by mass spectrometry (MS) analysis, suggesting an unusual possibility that a 4E-like protein might associate or reside in/on the mitochondrion of the parasite [36]. This has not been observed in other eukaryotes. If verified, this would have profound implication in the evolution of 4E function and eukaryotic translation since the mitochondrion is thought to be a derivative of ancient prokaryotes [37].

2.0 STATEMENT OF THE PROBLEM

The 4E-3p interaction with translational proteins has been analyzed in other studies using *in vitro* hemagglutinin (HA) tagged, exogenously expressed proteins in *T. brucei* to determine its potential biochemical function, whereas 4E-5p remains to be examined in detail [1][21]. Although a role in translation is becoming evident for 4E-3p, we still do not know the precise biochemical interactions and functions of 4E-3p and 4E-5p *in* vivo, much less their fundamental connections to normal cell cycle.

2.1 HYPOTHESIS

Based on functions of 4E and the existence of 4E in other eukaryotic systems, we hypothesize that: (1) 4E-3p binds directly to target cytoplasmic mRNAs involved in the cell cycle and interacts with other post transcriptional proteins to regulate their translation and/or mRNA stability, and that (2) 4E-5p associates with mitochondria and binds mitochondrial-related mRNA to regulate their expression, stability or localization. These hypotheses propose that 4E-3p and 4E-5p act on mRNA regulons, perhaps like 4Es in other eukaryotes that act as mRNA regulons for coordinated nuclear export or translation of functionally-related mRNAs [38][39]. A regulon is defined as multiple mRNAs which are structurally related and/or contribute to the same function regulated by an RNA binding protein (RBP) [40]. In our models, 4E-3p and 4E-5p

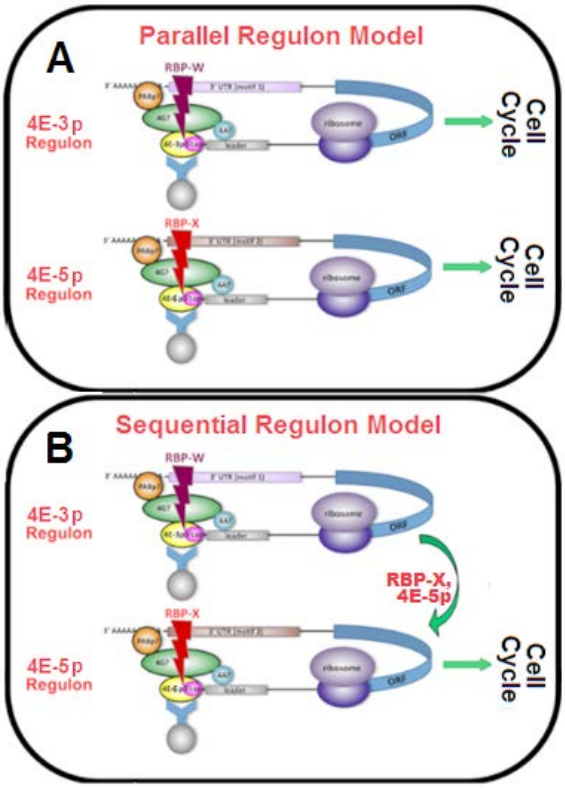


Figure 4. Regulon Models for 4E-3 and 4E-5: Sequential vs Parallel .

We hypothesize that TbeIF4E-3 and TbeIF4E-5 bind to distinct regulons to regulate cell cycle (A) independently on separate mRNA regulons that share similar functions or (B) in sequence on the same mRNA regulons or mRNA regulons in the same pathway. A regulon is a group of functionally related mRNA that are regulated by an RNA binding protein (RBP).

interact with their own unique sets of mRNAs that play key roles in cell cycle and growth regulation, either in parallel or in sequence. In the Parallel Regulon Model, 4E-3p and 4E-5p independently act as regulons for unique sets of functionally related mRNA. In the Sequential Regulon Model, 4E-3p and 4E-5p act as regulons for shared mRNAs at distinct stages of a single post-transcriptional pathway **(Figure 4)**.

2.2 SPECIFIC AIMS

The purpose of these studies in general were to further analyze the biochemical functions of 4E-3p and 4E-5p in trypanosomes, and how their molecular functions impact cell cycle. Thus, the specific aims of this project were:

(1) To determine the intracellular locations of 4E-3p versus 4E-5p using immunofluorescence microscopy.

(2) To develop a coimmunoprecipitation (co-IP) method to capture 4E-3p associated mRNAs and polypeptides.

(3) To determine the associated mRNA and polypeptide identities by RNAseq and mass spectrometry (MS) analyses respectively, in order to begin to elucidate their biochemical functions **(Figure 5)**.

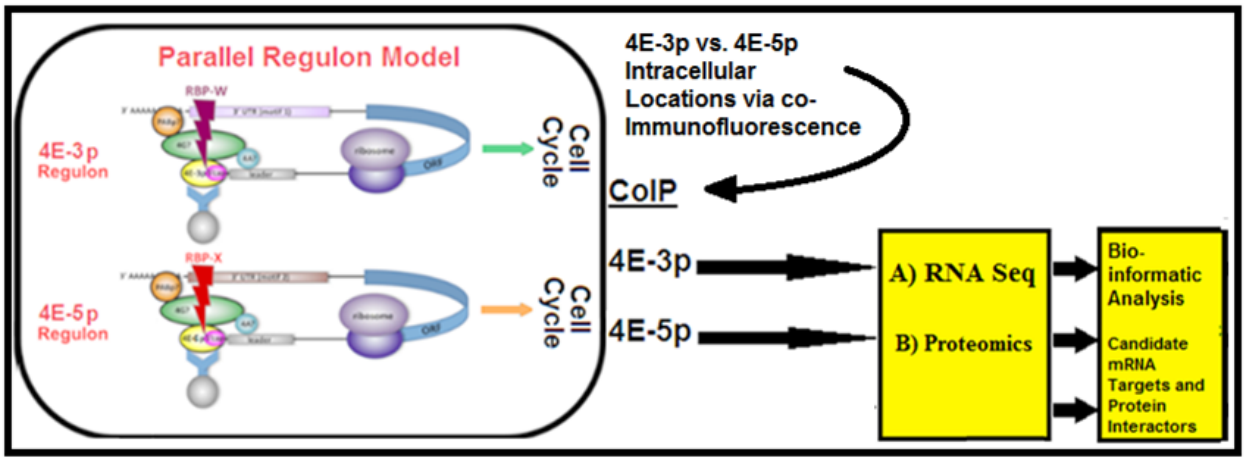


Figure 5. Work flow of experimental plan.

This is a graphical depiction of the aims and the procedures to accomplish them. Aim 1 focused on the determination of the intracellular locations of 4E-3p and 4E-5p using immunofluorescence microscopy. Aim 2 was to develop an optimized co-immunoprecipitation (co-IP) method to capture any proteins or mRNAs associated with 4E-3p and 4E-5p. Aim 3 focused on determining the identities of the mRNA and polypeptides via RNAseq and mass spectrometry (MS) analysis respectively, which would provide possible functions and relations to 4E-3p and/or 4E-5p.

3.0 MATERIALS AND METHODS

3.1 CELL CULTURING

Procyclic forms of the 29.13 cell line of T. *brucei* were used [41]. The 29.13 transgenic cell lines were grown in SDM79 (Gibco) with 10% fetal bovine serum (FBS; GIBCO) treated with 15ug/mL of G-418 (Fisher) and 25ug/mL of hygromycin (Fisher). The RNAi transgenic strains of 29.13 harbored a pZJM plasmid, which contains tetracycline-inducible T7 promoters and gene-specific target sequences against TbeIF4E-3p (4E-3 RNAi), the forward primer being 5'-GAC*CTCGAG*GTTGGGGATGTTGAGTGCTT-3' and reverse primer being 5'-GCC*AAGCTT*TCGACGGTTTTCCTTTCATC-3', and TbeIF4E-5p (4E-5 RNAi), the forward primer being 5'-GAC*CTCGAG*GAAGGATCCCTGGTTTGTGA-3' and reverse primer being 5'-GCA*AAGCTT*CGCCAAAAATAACGTGAGGT-3' [2][42][43]. These allowed for the validation of successful transfection through polymerase chain reaction (PCR) and gel electrophoresis using the restriction enzyme Xhol, the sequence underlined and italicized, and to incur partial knockdown of the mRNA that encode for the target proteins 4E-3p and 4E-5p

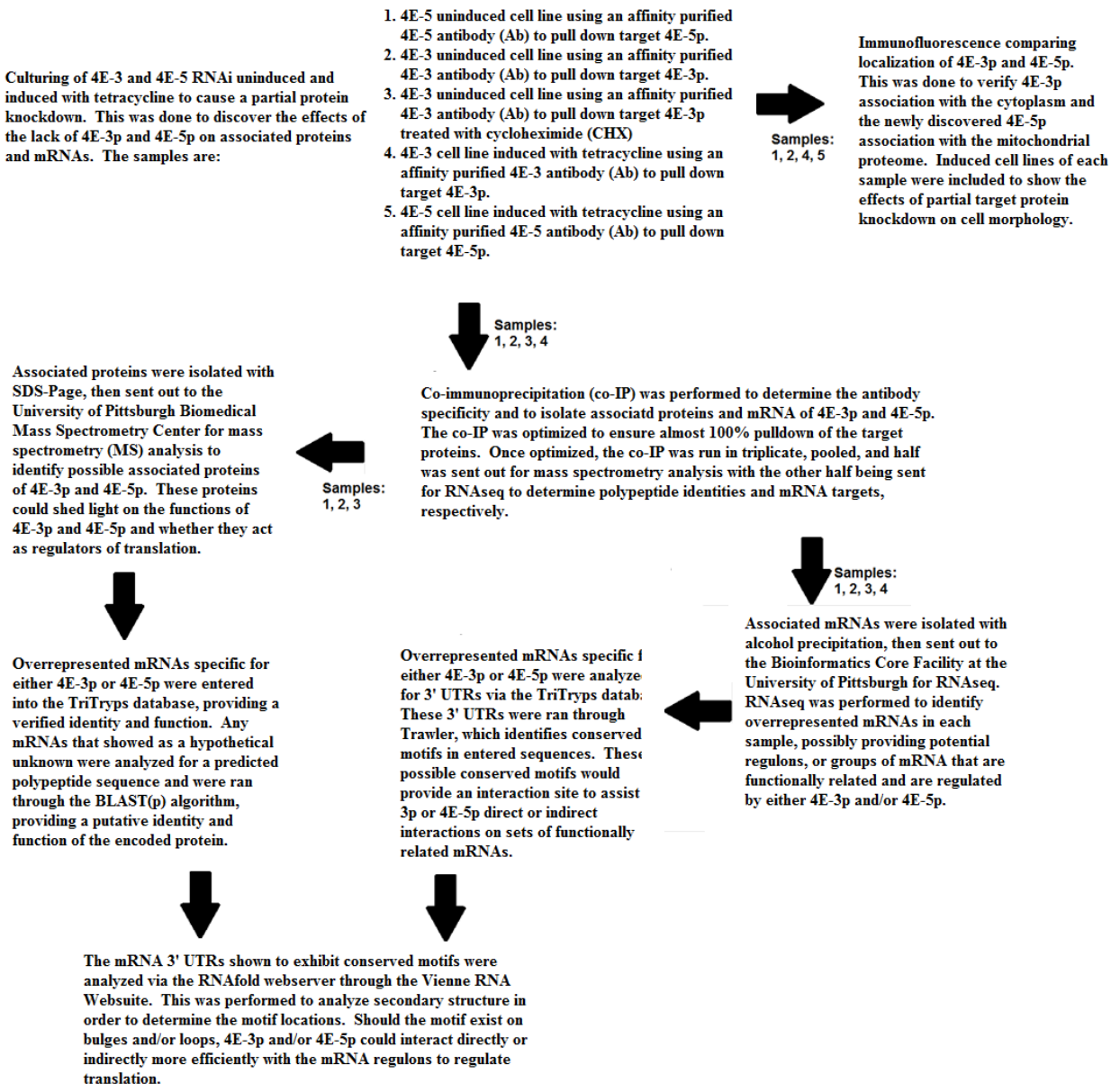


Figure 6. Graphical depiction of methodology.

Cell culturing was designed to create RNAi cell lines that would cause a partial knockdown of 4E-3p and 4E-5p. From here, immunofluorescence was performed to verify the locations of the target proteins. Co-immunoprecipitation was optimized and performed to ensure almost complete pulldown of 4E-3p and 4E-5p using affinity purified antibodies (Ab) targeting the proteins. Half of these samples were sent for mass spectrometry (MS) analysis to determine possible proteins interactors and functions of 4E-3p and 4E-5p as possible regulators of translation. The other half was sent for RNAseq to identify overrepresented mRNAs specific for 4E-3p and/4E-5p. These mRNAs could possibly share functionality or pathways, making them mRNA regulons. The mRNA identities were verified using TriTryps, with hypothetical unknowns being analyzed for predicting polypeptide composition and ran through the BLASTp algorithm for putative function and protein homology. The identified mRNAs were analyzed for conserved motifs in their 3' UTRs, showing a possible site for 4E-3p and/or 4E-5p interaction. These 3' UTRs were then analyzed via RNAfold for secondary structure similarities and motif location. Should the motif reside on loops and/or bulges, that would assist in 4E-3p and/or 4E-5p direct or indirect interaction.

Both the 4E-3 and 4E-5 RNAi lines received the addition of 2.5ug/mL phleomycin (Invitrogen) in the cell culture to select for the successfully transformed cells. The cell cultures were kept at $1x10^6$ cells/mL by monitoring their counts with the Automated Cell Counter (Invitrogen), maintained every two days with fresh SDM79+10% FBS with the required antibiotics. Inducing the 4E-5 and 4E-3 RNAi lines to cause a partial knockdown of the target proteins 4E-5p and 4E-3p required the addition of 1ug/mL tetracycline at a concentration of $1x10^6$ cells/mL.

3.2 IMMUNOFLUORESCENCE STAINING

In collaboration with Joseph Rocco (Pitt Medical Student), approximately $5x10^7$ cells of non-induced and induced 4E-3 and 4E-5 RNAi procyclic (insect stage) cells were harvested and incubated separately in 500nM MitoTracker Red at 37C for 10 minutes to stain the single mitochondrion in each cell (Invitrogen). The induced 4E-3 and 4E-5 RNAi were added to show the effects of their target protein knockdown. After spinning the cells down twice at 4,000 RPM at 4°C for 10 minutes and washing with medium in between, the cells were then fixed in 4% paraformaldehyde (Fisher) for 20 minutes onto separate microscope slides (Fisher) and permeabilized with 1% Triton x-100 (Fisher) for 15 minutes and washed three times with 1x PBS, finally being resuspended with 10% BSA (SIGMA) in PBS. The cells were then incubated with affinity purified primary antibodies (Ab) for two hours using a 1/200 concentration of 4E-3 rabbit Ab for the 4E-3 RNAi line and a 1/200 concentration of 4E-5 rabbit Ab for the 4E-5 RNAi line. The cells were then washed three times with 1% BSA. The customized Abs were designed according to the manufacture's proprietary software (Open Biosystems), with the 4E-3 Ab being

raised from an N-terminal peptide from amino acid (a.a.) positions 4-22 (EAEEFVPKGNRTPGSGGRR) and the 4E-5 Ab being raised from an N terminal peptide from a.a. positions 3-23 (EESHALKDPWFVSYIPQLTTE). The secondary antibody was added, a 1/400 concentration of fluorescein (FITC; Thermo Fisher) as the secondary goat anti-rabbit antibody for the 4E-3 and the 4E-5 strains, incubated for two hours, then washed three times with 1% BSA in PBS **(Figure 7)**. The final stain of 300ng/mL DAPI to stain the DNA was applied for 10 minutes. The samples were then viewed with an Olympus Provis AX70 Fluorescence Microscope with a standard lens, using the U-PHOTO Universal Photo System through the University of Pittsburgh Center for Biologic Imaging.

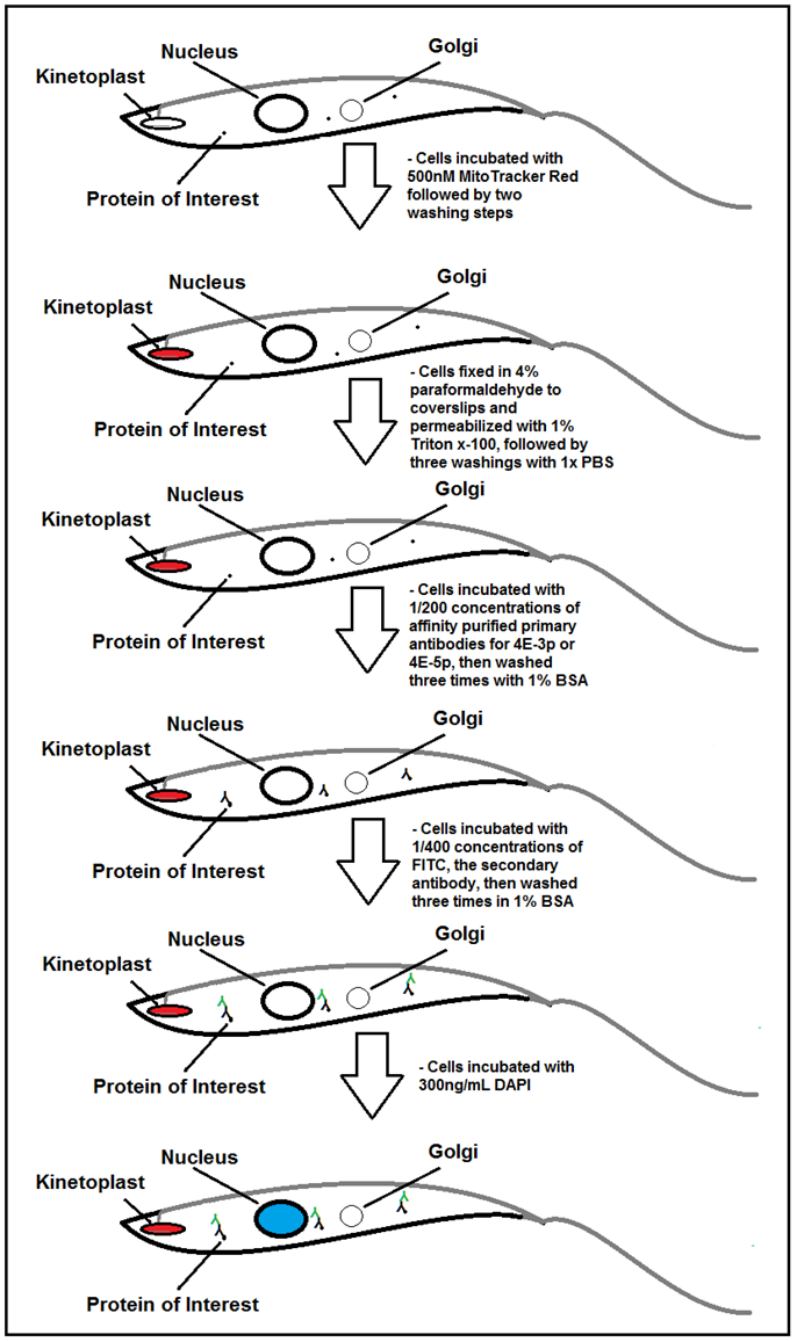


Figure 7. Immunofluorescence methodology.

Graphical depiction of antibody-antibody-protein interaction with staining of the mitochondrion and nucleus using MitoTracker Red and Dapi, respectively. The primary antibodies (Abs) were affinity purified rabbit Abs raised against 4E-3p or 4E-5p. The secondary Ab was an anti-rabbit goat FITC Ab.

3.3 CO-IMMUNOPRECIPITATION

The following co-immunoprecipitation (coIP) protocol was derived from a series of systematic optimization studies. Four cultures were prepared of the 4E-3 RNAi cell line, with one culture receiving Tetracycline (Tet) to induce the 4E-3 RNAi while the others remained untreated (non-induced). Each was harvested at day 2, when complete down-regulation and a loss of 90% of target protein occurs in induced cells without growth defects [2], with one of the non-induced cultures receiving a constant cycloheximide (CHX) (Fisher) treatment of 100ug/mL throughout the rest of the coimmunoprecipitation (co-IP) in order to arrest translation and allow potential polyribosome – associated 4E-3 complexes to be captured. Cell cultures comprised of roughly $1x10^8$ cells were spun down at 4,000 RPM at 4°C and washed gently so as not to disturb the pellet three times. The pellets were then resuspended in a low stringency (10mM) KCl salt solution lysis buffer and subjected to douncing with a glass rod, all the while being kept on ice. Each of the samples were lysed using a low stringency salt solution to preserve protein-protein interaction. The samples were then spun at 14,000 RPM at 4°C to separate the target proteins and their associated complexes from the rest of the sample. The resultant total lysate was then subjected to incubation with optimized amounts of affinity purified Ab raised against 4E-3p and 4E-5p (as specificity control), followed by subsequent incubation with magnetic Protein A Sepharose beads (Invitrogen), again all performed on ice. The resultant protein-Ab-bead complexes were then placed in a magnetic rack and washed with a high stringency (120mM) KCl salt solution to remove any non-specific binding **(Figure 8)**. This entire process was done in triplicate and any subsequent experiments used a pool of the three trials. The samples were as follows: Sample 1 received only the 4E-5 antibody; Sample 2 received only the 4E-3 antibody; Sample 3 received the 4E-3 antibody with a constant treatment of CHX (Acros Organics); and

Sample 4 received the 4E-3 antibody with Tet treatment. The amounts of Abs used were optimized to allow for quantitative pulldown of the target antigen to near complete depletion from the lysate, as assessed by western blot analysis. Samples were also taken at each step to verify the efficacy of the Ab pulldown via western blot analysis.

3.4 MASS SPECTROMETRY

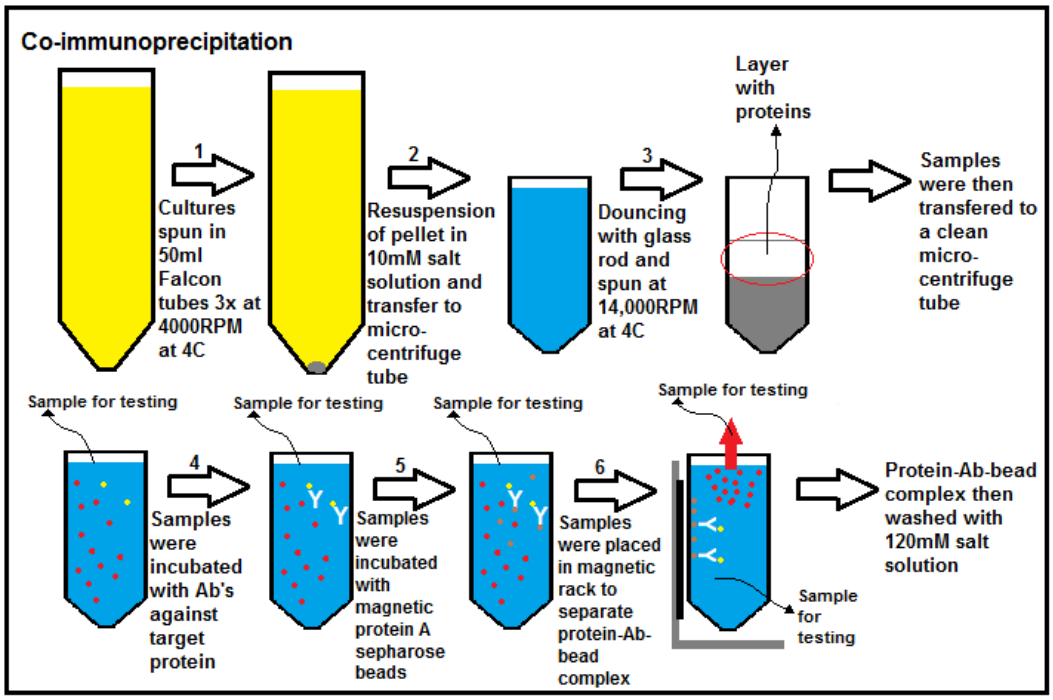


Figure 8. Graphical depiction of anti-4E-3p and anti-4E-5p antibody (Ab) pulldown of target protein with magnetic Protein A Sepharose beads.

Samples were taken to verify specificity of the Ab pulldown and ensure the most efficient capture of the target proteins and their associated complexes and mRNA.

Half of the samples that were prepared using co-IP and confirmed via western blot (Samples 1-3) were then subjected to SDS PAGE treatment to purify the total proteins that were pulled down. These samples were then run on a Mini-Protean TGX Gel (Bio Rad), and stained using Sypro Ruby. The lanes of the gel were then cut, partitioned into thirds, and sent out for MS Analysis at the Biomedical Mass Spectrometry Center at the University of Pittsburgh. The data obtained from Dr. Manimalha Balasubramani at the MS Center were then analyzed using BLAST through NCBI and the TriTryps database to examine the proteins pulled down through coIP and to identify candidate 4E-3 and/or 4E-5 protein interactors.

3.5 RNA ISOLATION FROM IMMUNOPURIFIED COMPLEXES

The other half of the samples prepared using co-IP and confirmed via Western (Samples 1-4) were also subjected to ethanol treatments to precipitate the RNA following the protocol in the Qiagen RNA Prep Handbook (Jan. 2006). In this protocol, the coIP samples were treated with proteinase K (Sigma) at a concentration of 0.1ug/mL, then were heated at 42°C. Sodium acetate (Fisher) was then added to each sample at a concentration of 0.3M and a pH of 5.2. Two volumes of phenol chloroform isoamyl alcohol (PCI) (ACROS) were added next, with the samples then being spun at 14,000 RPM at 4°C. The upper aqueous layer was carefully extracted, then treated with two volumes of chloroform isoamyl alcohol (ACROS) and spun once more at 14,000 RPM at 4°C. The aqueous layer containing the RNA was removed and was treated with glycogen at a concentration of 1ul/sample to assist in RNA precipitation. Each sample received 2.5 volumes of 100% ethanol (Fisher) and were spun at 14,000 RMP at 4°C to

pellet the RNA. The ethanol was poured out and the pelleted RNAs were allowed to dry slightly, then were resuspended in RNase free water (Fisher).

The RNA concentrations were determined using a Nanodrop instrument and then sent out for analysis to determine mRNA identity and quantity by Rahil Sethi and Dr. James Lyons-Weiler at the University Bioinformatics Core Facility. The RNA samples were analyzed using both Burrows-Wheeler Aligner (BWA), which is bidirectional and functions well for ungapped reads, and Bowtie, which performs well with gapped reads, to ensure full coverage of the mRNAs in each sample [44]. It was discovered that each mapper provided different results for the mRNAs specific to each sample, with only a few shared results, so the overlap was used to reduce the chance of having mRNAs specific to neither sample and to ensure the most accurate results. The TREU927 and Lister 427 reference sequences were used to identify each mRNA, again using the overlap to ensure positive identification of the mRNAs present due to each approach yielding different results.

3.6 DISCOVERY OF CONSERVED MOTIFS USING CAGEDA AND TRAWLER

The data returned were then analyzed using caGEDA (helen.genetics.pitt.edu) to discover the overrepresented mRNAs in each sample [45]. The sample comparisons used were the 4E-5 untreated coIP versus the 4E-3 untreated coIP (4E5v4E3), the 4E-3 untreated coIP versus the 4E3+CHX (4E3v4E3+CHX), and 4E-3 coIP versus the 4E-3+Tet treated coIP (4E3v4E3+Tet).

For each sample comparison, normalizations were performed to ensure only overrepresented mRNAs specific for each sample were selected. Each mRNA was then assigned significance based on a mean scaled value that adjusts for a difference in total hits between samples, which represents differential expression in relation to the compared sample, and was used as a threshold to determine sample-specific mRNAs and eliminate mRNAs shared equally between the compared samples. Biological significance could be determined by taking the mean scaled value of a specific mRNA overrepresented in one sample over the mean scaled value of the same mRNA that is underrepresented in the compared sample to determine fold-change, although values were not shown for the compared underrepresented mRNAs. Instead, proportions of the mean scaled values of sample specific mRNAs were used, allowing for a quantitative significance of the mRNAs in each sample. The threshold was used as a baseline to determine the quantitative value of each overrepresented mRNA. The data generated by caGEDA provided an expression grid based on the threshold entered with the value changed to ensure only mRNAs with a p-value above 95%, for specificity in the first sample, or below 5%, signifying a p-value of 95% in the second sample. This allowed for a selection of mRNAs overrepresented in each sample with the highest statistical significance, and eliminated nonspecific mRNAs.

The 4E5v4E3 analysis was normalized to fix varying means between samples, using log base 2 and z transformations. This allows for the determination of sample-specific RNAs. The threshold, which is the cutoff for the number assigned for each gene, denoting specificity and presence in the sample, was set to 4.94, any threshold below 4.94 allowed for nonspecific mRNAs with a lower p-value. The 4E3v4E3+CHX analysis was normalized in the same manner

as the 4E5v4E3 analysis, with the same threshold of 4.94, with the 4E3v4E3+Tet analysis following the previous normalizations of 4E5v4E3, however a threshold value of only 4.3 was needed to allow for the elimination of all nonspecific mRNAs.

The gene identifiers of the overrepresented mRNAs were then compared to the TriTryps database and their encoded proteins were analyzed with NCBI's protein BLAST (BLASTp) algorithm to determine putative function and homology in relation to other eukaryotes and prokaryotes.

The 3' UTRs of the mRNAs were then analyzed using Trawler (EMBL; http://ani.embl.de/trawler/) to examine possible conserved motifs, which would then be used to determine possible regulon properties of 4E-5p and 4E-3p. Trawler examines entered sequences from CHiP experiments for recurring segments 5-20 nucleotides in length that are shared between two or more entries, and provides the location, nucleotide composition, and motif homology with known transcriptional motifs [46][47]. Sequences in FASTA format were generated using the TriTryps database geneID function, selecting the regions of each gene from the translation stop codon to the transcription stop. The background sequence used while analyzing the samples was the entire T. *brucei* genome, used based off of the original paper comparing yeast and mammal binding sites, which uncovered novel binding sites and motifs, and allowed for a more accurate *z*-score threshold [47]. Because Trawler was designed for use with chromatin-immunoprecipitation (ChiP) data, uracils were represented as thymines in the input.

3.7 SECONDARY STRUCTURE PREDICTION OF 3’ UTRS USING RNAFOLD

The three sample comparisons were used as in the previous experiment; 4E5v4E3, 4E3v4E3+Tet, and 4E3v4E3+CHX. Each set of overrepresented mRNAs exhibiting the same 3' UTR motif were first run through the TriTryps database again to obtain their 3' UTRs and to verify the accuracy of the previous run-throughs. These 3' UTRs were analyzed individually using the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) from the Vienna RNA Websuite, which has successfully been used in trypanosome studies to analyze structural elements for non-coding mRNAs (ncRNAs) predicted through QRNA, to determine the most stable 3' UTR structure [48-51]. The minimum free energy (MFE) of the structure and the partition function were used to determine the most stable structure with the best base pair probability.

The secondary structural elements of the mRNA 3' UTRs generated with the RNAfold web server were then analyzed for the specific locations of the shared motifs. Since the RNAfold web server does not number the nucleotides in the secondary structures, the first and last nucleotides were determined using the 3' UTR data the TriTryps database provided. From here, the individual nucleotides were counted, with the specific location of the motifs being designated from the Trawler outputs.

3.8 MOTIF LOCATION USING TRAWLER DATA

The 3’ UTR mRNA motifs were then analyzed for any structural correlations. Should 4E-3p and/or 4E-5p act directly or indirectly to select for specific mRNAs, these shared motifs would need to be accessible. Bulges and loops provide easier access for any proteins that specifically interact with the motifs, so the percent of motif location was determined when the motif sequence contained two or more nucleotides residing on the bulge and loop mRNA secondary structures in relation to the total number of motifs.

4.0 RESULTS

4.1 IMMUNOFLUORESCENCE

A previous study has shown that the hypothetical protein encoded by the gene Tb927.10.5020, later positively identified as TbeIF4E-5 by our lab, associates with the mitochondrial proteome, however further studies have not been conducted to determine its function and confirm the cellular location [2][36]. To address this issue, a coimmunofluorescence staining protocol was developed. Possible 4E-5p co-localization with the mitochondrion was compared with the known 4E-3p co-localization with the cytoplasm to show the contrast between the two localizations by using affinity purified primary antibodies (Ab) specific for 4E-3p or 4E-5p, MitoTracker red to stain for the mitochondria, DAPI to stain for the nucleus, and an anti-rabbit FITC secondary Ab to determine the locations of the two parasite proteins [6][7]. The 4E-3 RNAi and 4E-5 RNAi coimmunofluorescence were performed to show the effects of partial RNAi knockdown of the respective target proteins in comparison to the cell lines expressing 4E-3p and 4E-5p at normal levels.

The immunofluorescence data revealed yellow speckled foci (arrows) suggesting an association between the uninduced 4E-5p and the mitochondria **(Figure 9: 1-5)**. These yellow foci represent the merging of the Mito stain (red) and the FITC stain (green), which show as a

yellow color when the two are co-localized. By contrast, the data from uninduced 4E-3p shows an association with the cytoplasm, consistent with previous studies **(Figure 9: 6-10)** [52]. The RNAi induced 4E-5 and 4E-3 immunofluorescence was shown next to the uninduced immunofluorescence to illustrate the effects of protein knockdown on kinetoplast and nuclear division, which caused multiple kinetoplasts and nuclei deformations in the parasites, evidenced by the multiple nuclei stained by DAPI.

Therefore, these data show evidence for the association of the 4E-5p with the mitochondrion of the parasite, depicted by the yellow coloration in the cells, which shows an overlap of 4E-5p with the mitochondrion. The lack of any of these yellow foci in the 4E-3 slides shows confirmation that 4E-3p acts in the cytoplasm of the parasite. The partial knockdown immunofluorescence supports previous findings from our lab that a lack of 4E-3p and 4E-5p cause deformations in the cell, evidenced by the multiple DAPI-stained nuclei in each cell.

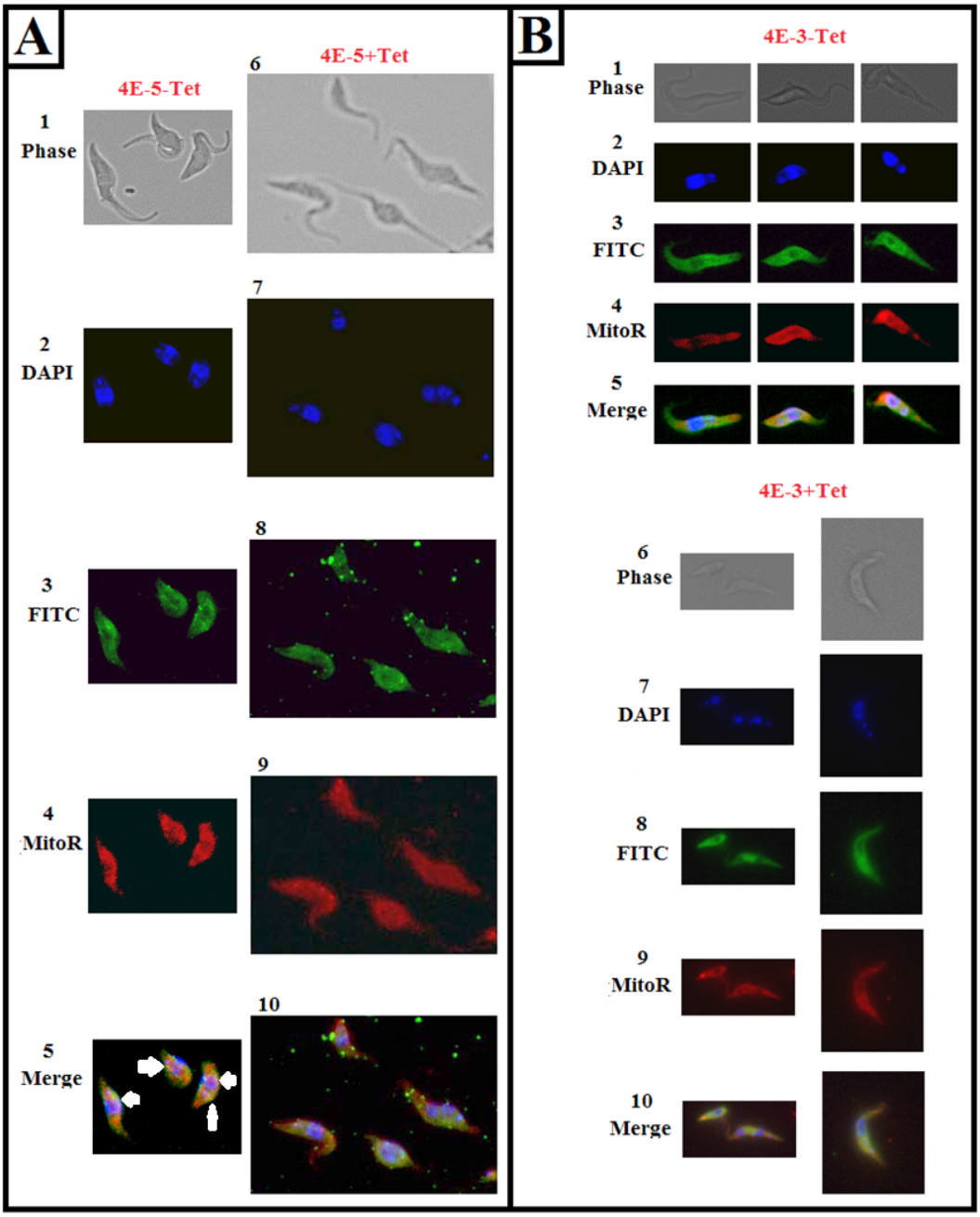


Figure 9. Immunofluorescent images.

(A) Immunofluorescent images depicting 4E-5 association with the mitochondrion in normal cells (-Tet) and RNAi induced cells (+Tet). (1) Procyclic form T. *brucei* light microscopy, (2) DAPI staining of kinetoplast and nucleus, (3) FITC anti-rabbit IgG against anti-4E-5 antibody, (4) MitoTracker Red CMXRos fluorescent dye depicting mitochondrial membrane, (5) merge; yellow foci (arrows) indicate co-localization of 4E-5 and mitochondria. (B) Immunofluorescent images depicting 4E-3 association with the cytoplasm in normal cells (-Tet) and RNAi induced cells (+Tet). (6) PF T. *brucei* light microscopy, (7) DAPI staining of kinetoplast and nucleus, (8) FITC anti-rabbit IgG against anti-4E-3 antibody, the large green masses are clumps of FITC unrelated to the cells, (9) MitoTracker Red CMXRos fluorescent dye depicting mitochondrial membrane, (10) merge; general diffusion of 4E-3 in the cytoplasm.

4.2 CO-IMMUNOPRECIPITATION

To determine the specificities of the affinity purified antibodies in order to ensure the most efficient pulldown of possible functional translation complexes, a co-immunoprecipitation (coIP) procedure was designed and optimized, with western blot analysis used to confirm the complete or near-complete pulldown of the target proteins. A low stringency salt solution was used for dounce homogenization and preservation of protein-protein interactions. The pelleted antibody-protein complexes were washed with a high stringency salt solution to reduce non-specific binding, which was later analyzed to determine the specificities of the affinity purified antibodies.

A second set of coIP experiments was performed to later assess the associations of 4E-3p and 4E-5p with other proteins and with mRNAs. The low stringency salt solution was used for dounce homogenezation, while a moderately stringent salt solution was used to wash the pelleted antibody-protein complexes, hopefully preserving the protein-protein interactions, with an RNase inhibitor to preserve any associated mRNAs. Sample 1 was a pulldown using the affinity purified anti-4E-5 antibody with the 4E-5 RNAi cell line. Sample 2 was a pulldown using the affinity purified anti-4E-3 antibody with the 4E-3 RNAi cell line. Sample 3 was a pulldown using the affinity purified anti-4E-3 antibody with the 4E-3 RNAi cell line, with the addition of cycloheximide to arrest translation. Finally, sample 4 was a pull down using the affinity purified anti-4E-3 antibody with the 4E-3 RNAi cell line, with the addition of tetracycline to partially knock down 4E-3p.

The western blot analysis of the 4E-3p coIP showed there was almost complete pull down of 4E-3p from the total lysate **(Figure 10A)**. The Sypro Ruby analysis of the 4E-3p coIP shows a significant number of proteins associated with the pulldown of 4E-3p containing complexes**(Figure 10B)**.

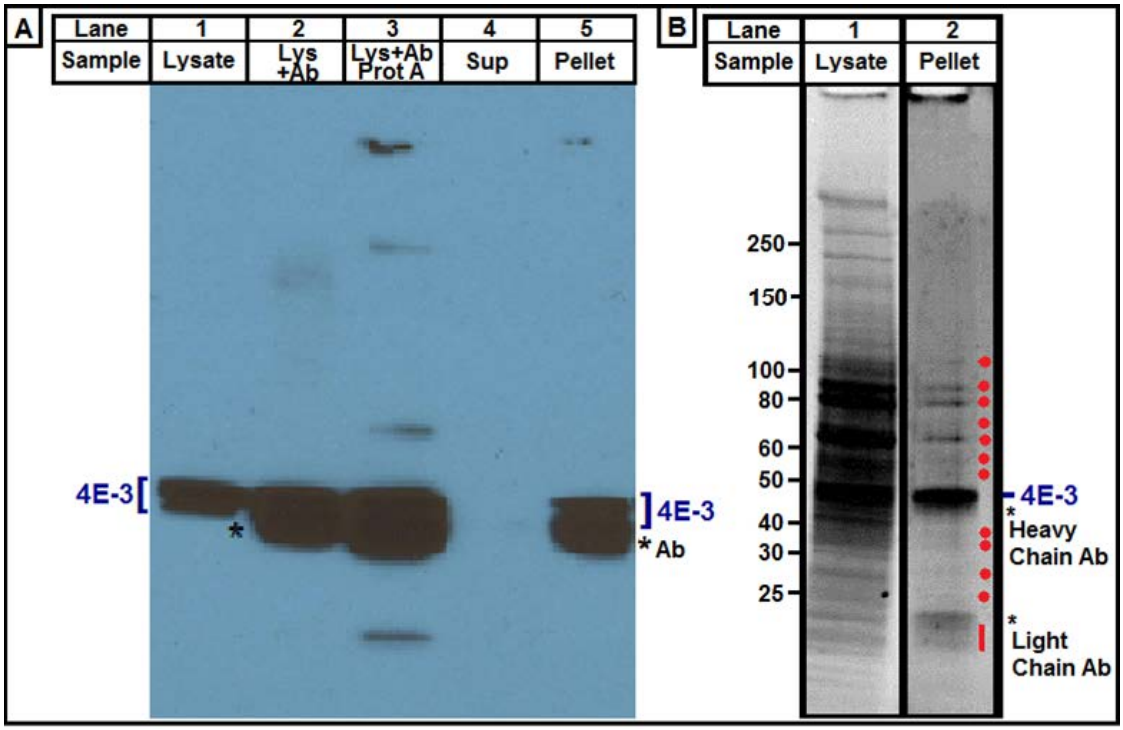


Figure 10. Efficient pulldown of TbeIF4E-3p by rabbit anti-4E-3 Ab and protein A sepharose magnetic beads.

(A) Western blotting for the 4E-3p antibody pulldown. Note immunodepletion of 4E-3p in supernatant, which is after the cell lysates being subjected to antibody (Ab) incubation and pulldown (lane 4). Lane 1 represents the total cell lysates. Lane 2 is after the addition of the antibody, denoted by *. Lane 3 is after the addition of protein A sepharose beads. Lane 4 shows the depletion of 4E-3p with the Ab-bead complex. Lane 5 shows the pelleted 4E-3p-Ab-bead complex. (B) Sypro-Ruby stain (negative image). Note co-IP of multiple proteins (red). Pellets were after 50 mM KCl wash. Lanes are normalized for comparison, each representing 1mL of procyclic form culture (10^7 cells). Lane 1 depicts the total lysate. Lane 2 depicts the 4E-3p-Ab-bead complex.

4.3 MASS SPECTROMETRY

Previous studies have shown the association of the 4E-3p with other proteins in the eIF4F complex. However, these experiments have used HA-tagged 4E-3p, the tag possibly interfering with the natural protein-protein interactions of 4E-3p. Furthermore, a comprehensive analysis of protein interactions has not been performed. Previous studies have shown that partial knockdown of 4E-3p interferes with cell division and nuclear and kinetoplast face, so we hypothesized that 4E-3p associates with other complexes to regulate cell cycle and possibly cell growth. To address this, samples from the optimized coIP experiments were sent out for mass spectrometry analysis. The samples from the coIP used were the Sample 1, denoted 4E-5, as our specificity control, Sample 2, denoted 4E-3, and Sample 3, denoted 4E-3+CHX.

Preliminary analysis of the MS data showed unique proteins being pulled down in each sample. Since background noises can never be eliminated, as a baseline for analysis, it was recommended by the Biomedical Mass Spectrometry Center that a two-fold difference would be accepted as significant difference between samples. Using this criterion, the samples subjected to the 4E-3 antibody (Ab) show a significant pulldown of 4E-3p but not 4E-5p, confirming the specificity of the antibody being used (**Figure 11**). Conversely, the 4E-5 Ab samples showed no presence of 4E-3p, yet contained an uncharacterized protein that was consistent with 4E-5p. Between the three samples analyzed, there were a greater number of proteins represented in the sample treated with the translation inhibitor, cycloheximide (CHX) **(Table 1)**.

Interestingly, under this condition, a homolog of 4G, identified as TbEIF4G-4 (4G-4 accession number Tb927.11.10560) and a homolog of 4A, identified as TbEIF4A-1 (4A-1 accession number Tb09.211.3510), an ATP-dependent DEAD/H RNA helicase, were also present in the

4E-3 Ab sample pulldown treated with CHX **(Figures 12 and 13)**. The 4G-4 peptide was directly identified from the Tri-Tryp database, whereas 4A peptide was identified using BLAST search and peptide alignment, showing a 20% similarity between human eIF4A. In comparison, homology was determined between 4E-3p and yeast 4E protein at 27%. These results are consistent with the hypothesized function of 4E-3p as a canonical eIF4E in translation. This is the first *in vivo* evidence of endogenous 4E-3 association with components of the 4F complex. Intriguingly, of the proteins specifically pulled down in the CHX sample, a majority were metabolic enzymes. The reasons remain unclear but possible explanations are presented in the discussion section.

Ribosomal proteins were also pulled down in all three of the samples in similar quantities, regardless of the Ab used, suggesting either non-specific binding or supporting a role in translation for both 4E-3p and 4E-5p, confirming 4E-3p and 4E-5p as regulators of translation. Of the uncharacterized proteins that were pulled down, two were shown to have similarities with proteins already identified in the TriTryps database. Q587B0 was pulled down in all three of the samples, and is verified to be Eukaryotic RNA Binding Protein 42, which binds to the TATA box on DNA and assists in RNA polymerase II binding, according to the TriTryps database (**Figure 14**). Of the other uncharacterized proteins, two were identified using the BLAST algorithm through NCBI to show homology in some degree to known eukaryotic proteins. Q385C5 was identified in each sample with no significant difference between them, and shows a 27% homology with the eukaryotic mRNA capping enzyme. Q581A4, found specifically in the 4E-5 antibody co-immunoprecipitation sample, was determined using BLAST and the TriTryp

database to be the recently discovered 4E-5p, again supporting the specificity of the antibodies being used (**Figure 15**).

The pull down of metabolic enzymes in the 4E-3 CHX sample supports the function and importance of the homolog 4E-3 in the growth cycle of trypanosomes, due to the need of metabolic enzymes in normal cell processes.

Overall, these data verify the specific pulldown of 4E-5p using the 4E-5 Ab and 4E-3p using the 4E-3 Ab. The association of proteins belonging to the 4F complex when translation is arrested in the 4E-3 pulldown using CHX, specifically 4G-4 and 4A, supports 4E-3p as a bona fide eukaryotic 4E and is important for translation. The general association of ribosomal proteins in all of the samples support both 4E-3p and 4E-5p as regulators in translation. The role of 4E-5p regulating translation while being associated with the mitochondrion would be unprecedented.

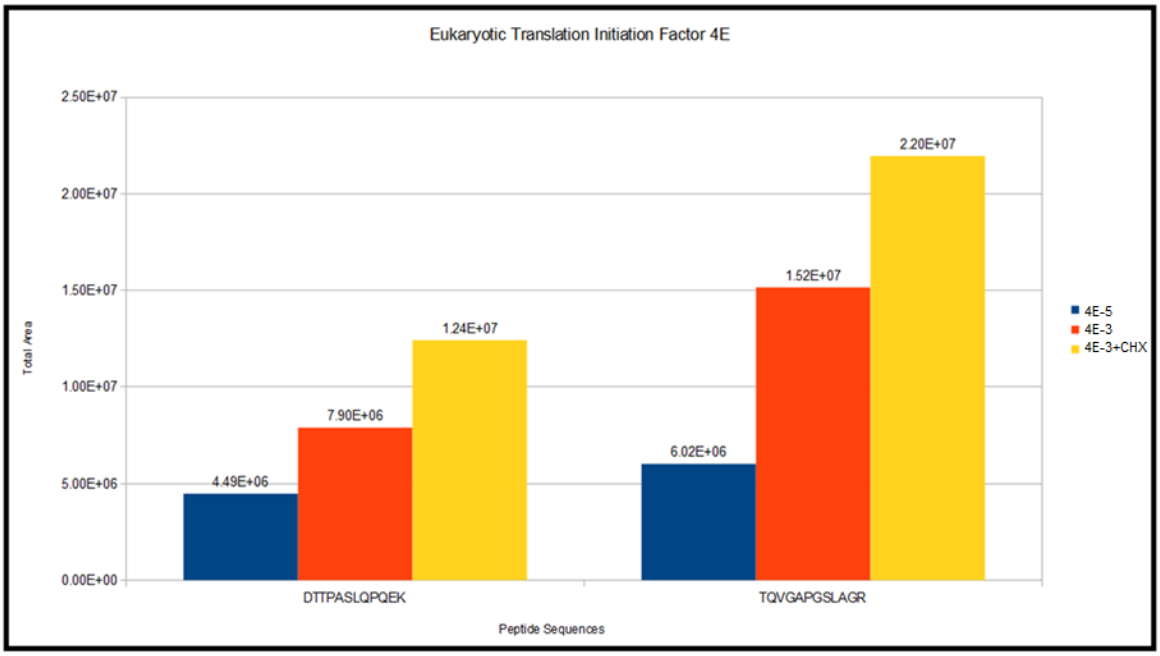


Figure 11. Chart showing relative area derived from mass spectrometry for eIF4E, verified by the TriTryps database to be TbeIF4E-3.

The peptide sequences are the segments the Biomedical Mass Spectrometry Center at the University of Pittsburgh verified to belong to this protein. The area under the peaks for the intensity of the peptide sequences is represented by Total Area. Blue represents the peptide concentrations in the anti-4E-5 antibody (Ab) sample, red represents the peptide concentrations in the anti-4E-3 Ab sample, and yellow represents the anti-4E-3 Ab sample treated with cycloheximide to arrest translation.

Table 1. Table depicting the discovered proteins associated with the target proteins.

Proteins Specific to the 4E-5 Ab Pulldown	Proteins Specific to the 4E-5 Ab Pulldown	Proteins Specific to the 4E-5 Ab Pulldown Treated with Cycloheximide	Proteins Shared Between All Samples
Nucleosome Assembly Protein	TbeIF4E-3	TbeIF4E-3	Glycosomal Malate Dehydrogenase
Retrotransposon Hot Spot		Elongation Factor 2	60S Ribosomal Protein L1
Q581A4 – TbeIF4E-5		Q57W90 – High molecular weight glutenin subunit	40S Ribosomal Protein SA
		Fructose Bisphosphate Aldolase	60S Acidic Ribosomal Subunit
		Elongation Factor 1-alpha-1	Q587B0 RNA Binding Protein 42
		eIF4A	60S Ribosomal Protein L13
		Q383M3 – eIF4G	40S Ribosomal Protein S10
		Succinyl-CoA: 3-ketoacid-coenzyme A transferase	60S Ribosomal Protein L6
		Pyruvate Phosphate Dikinase	Q385C5 – mRNA Capping Enzyme

The first column shows proteins specific for the anti-4E-5 antibody (Ab) pulldown. The second column shows proteins specific for the anti-4E-3 Ab pulldown. The third column shows proteins specific for the anti-4E-3 Ab pulldown treated with cycloheximide (CHX). This is important as CHX arrests translation and provides a more accurate representation of the polypeptides associated with 4E-3p. The fourth column shows proteins having no specificity between the samples.

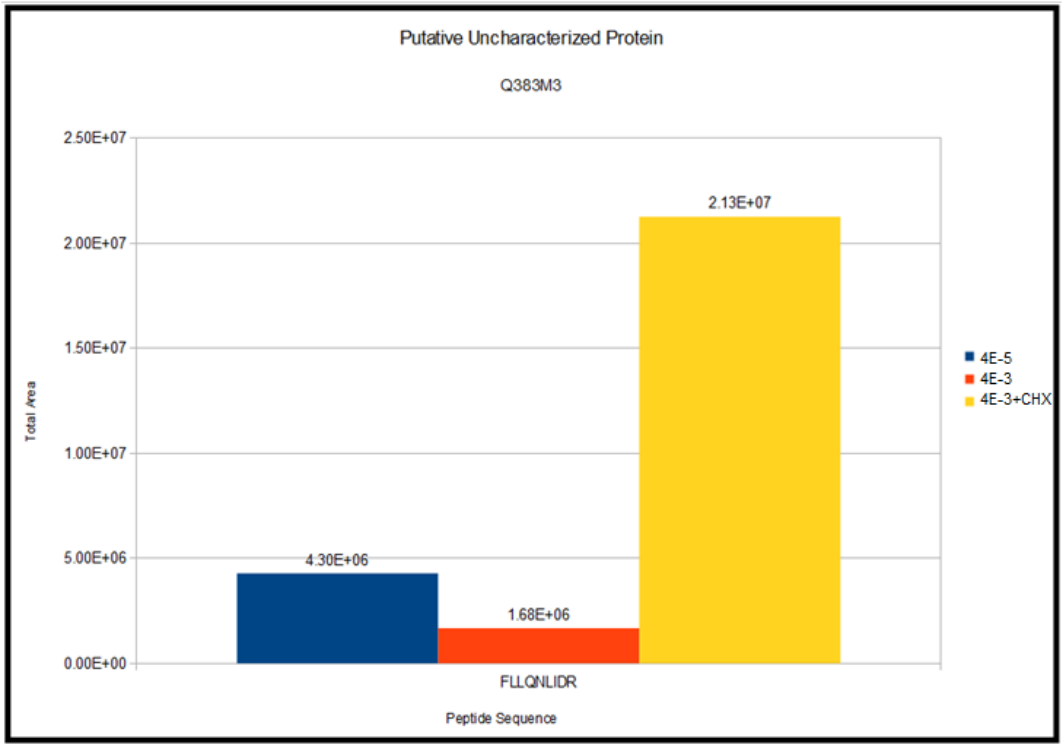


Figure 12. Chart showing relative area derived from mass spectrometry for Q383M3, verified by TriTryps database to be eIF4G.

The peptide sequences are the segments the Biomedical Mass Spectrometry Center at the University of Pittsburgh verified to belong to this protein. The area under the peaks for the intensity of the peptide sequences is represented by Total Area. Blue represents the peptide concentrations in the anti-4E-5 antibody (Ab) sample, red represents the peptide concentrations in the anti-4E-3 Ab sample, and yellow represents the anti-4E-3 Ab sample treated with cycloheximide to arrest translation.

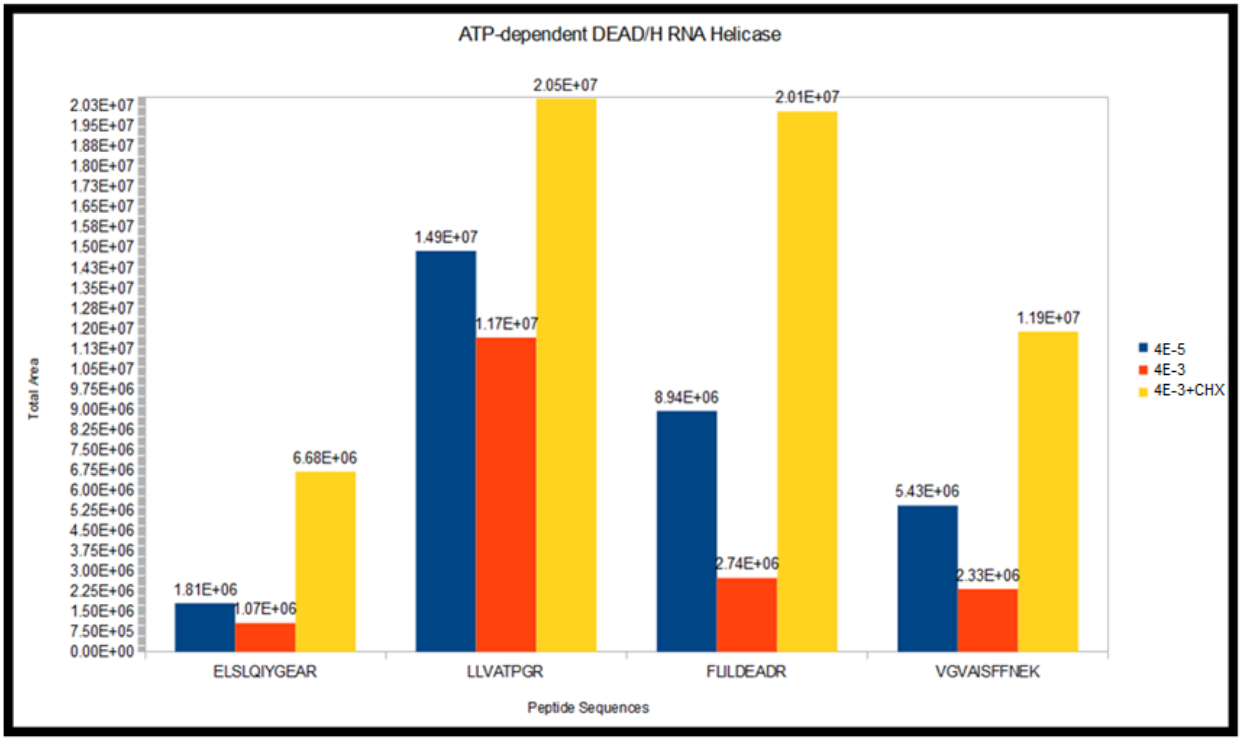


Figure 13. Chart showing relative area derived from mass spectrometry for ATP-dependent DEAD/H RNA Helicase, or eIF4A.

The peptide sequences are the segments the Biomedical Mass Spectrometry Center at the University of Pittsburgh verified to belong to this protein. The area under the peaks for the intensity of the peptide sequences is represented by Total Area. Blue represents the peptide concentrations in the anti-4E-5 antibody (Ab) sample, red represents the peptide concentrations in the anti-4E-3 Ab sample, and yellow represents the anti-4E-3 Ab sample treated with cycloheximide to arrest translation.

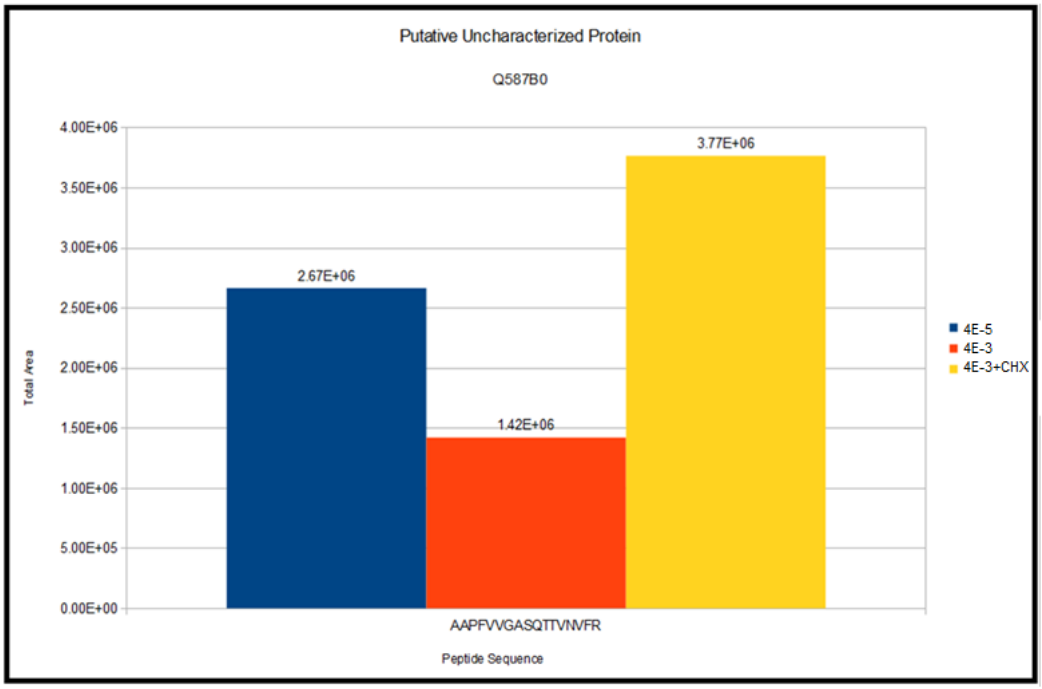


Figure 14. Chart showing relative area derived from mass spectrometry for Q587B0, verified by BLASTp algorithm to be homologous to RNA binding protein 42, which binds to the TATA box on DNA and assists in RNA polymerase II binding.

The peptide sequences are the segments the Biomedical Mass Spectrometry Center at the University of Pittsburgh verified to belong to this protein. The area under the peaks for the intensity of the peptide sequences is represented by Total Area. Blue represents the peptide concentrations in the anti-4E-5 antibody (Ab) sample, red represents the peptide concentrations in the anti-4E-3 Ab sample, and yellow represents the anti-4E-3 Ab sample treated with cycloheximide to arrest translation.

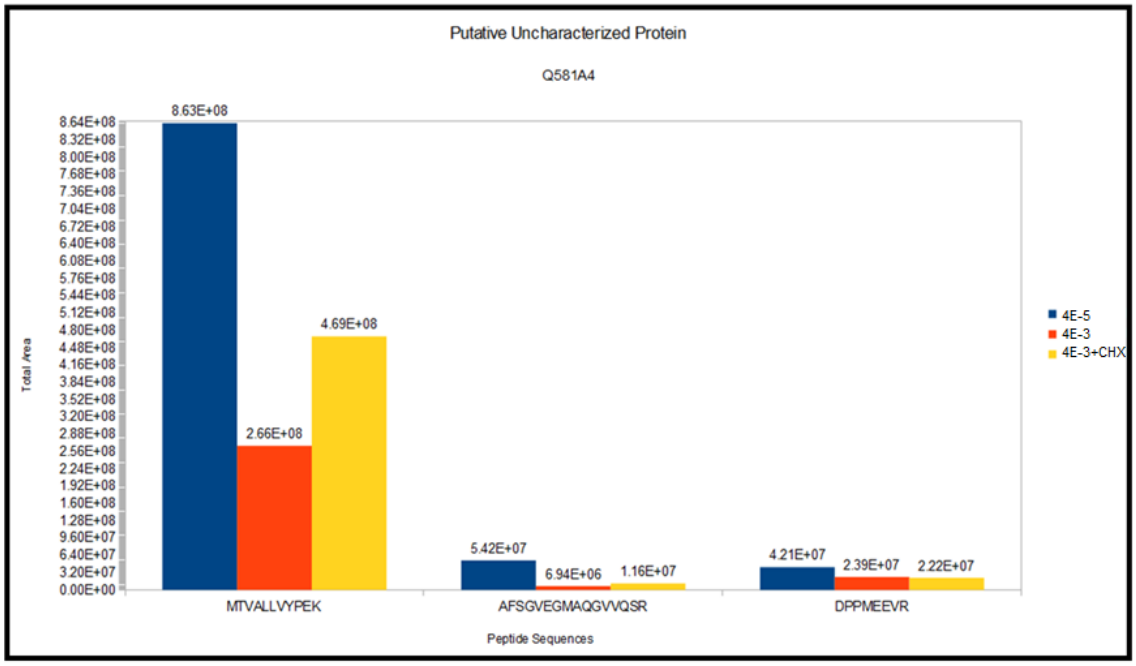


Figure 15. Chart showing relative area derived from mass spectrometry for Q581A4, verified by the TriTryps Database and out lab to be TbeIF4E-5.

The peptide sequences are the segments the Biomedical Mass Spectrometry Center at the University of Pittsburgh verified to belong to this protein. The area under the peaks for the intensity of the peptide sequences is represented by Total Area. Blue represents the peptide concentrations in the anti-4E-5 antibody (Ab) sample, red represents the peptide concentrations in the anti-4E-3 Ab sample, and yellow represents the anti-4E-3 Ab sample treated with cycloheximide to arrest translation.

4.4 RNASEQ ANALYSIS OF 4E-3 AND 4E-5 ASSOCIATED MRNA

Because of the confirmed association of 4E-3p with the components of the eIF4F complex and the confirmed identity of 4E-5p as a eukaryotic initiation factor, an experiment was performed to further determine the roles 4E-3p and 4E-5p play in the T. *brucei* cell viability and proper cell division, specifically through cell cycle regulation and gene expression. Therefore, samples 1 through 4 from the coIP experiment were analyzed using RNAseq. RNAseq quantifies the total mRNA from a given sample and provides known identities, which was important for these purposes to provide a snapshot of what mRNAs 4E-3p and 4E-5p were associated with and which mRNAs were present in higher quantities in each sample, either as general or specific translation initiators.

Preliminary results of the RNAseq performed by Rahil Sethi at the Bioinformatics Core Facility showed a roughly equal amount of unique mRNAs pulled down in each sample. There were two reference strains used to map the mRNA, Treu927 and Lister427. Each reference strain covered different unique sets of mRNA, so both were used to provide the highest amount of coverage. Both Burrows-Wheeler Aligner (BWA), which is bidirectional and functions well for ungapped reads, and Bowtie, which performs well with gapped reads, were used to ensure full coverage of the mRNAs in each sample [44]. All of the mappers used the fragment library type.

Bowtie mapping using the reference TREU927 provided 19,067,853 reads for the 4E-5 Ab pulldown, of which 14,057,831 were mapped and 382,434 were unique. BWA using TREU927 for the 4E-5 Ab pulldown provided 19,067,853 reads, 9,583,154 mapped reads, while

285,465 were unique. Bowtie mapping using Lister427 for the 4E-5 Ab pulldown provided 19,067,853 reads, 14,052,188 mapped reads, and 381,475 uniquely mapped reads. BWA using Lister427 for the 4E-5 Ab pulldown provided 19,067,853 reads, of which 9,622,037 were mapped and 329,076 were unique.

Bowtie mapping using the reference TREU927 provided 20,653,837 reads for the 4E-3 Ab pulldown, of which 13,906,156 were mapped and 405,544 were unique. BWA using TREU927 for the 4E-3 Ab pulldown provided 20,653,837 reads, 9,597,554 mapped reads, while 303,611 were unique. Bowtie mapping using Lister427 for the 4E-3 Ab pulldown provided 20,653,837 reads, 13,910,505 mapped reads, and 405,931 uniquely mapped reads. BWA using Lister427 for the 4E-3 Ab pulldown provided 20,653,837 reads, of which 9,662,124 were mapped and 358,133 were unique.

Bowtie mapping using the reference TREU927 provided 11,927,415 reads for the 4E-3 Ab pulldown treated with the translation inhibitor cycloheximide (CHX), of which 7,786,455 were mapped and 219,817 were unique. BWA using TREU927 for the 4E-3 Ab pulldown treated with CHX provided 11,927,415 reads, 5,092,635 mapped reads, while 155,471 were unique. Bowtie mapping using Lister427 for the 4E-3 Ab pulldown treated with CHX provided 11,927,415 reads, 7,789,011 mapped reads, and 220,012 uniquely mapped reads. BWA using Lister427 for the 4E-3 Ab pulldown treated with CHX provided 11,927,415 reads, of which 5,117,568 were mapped and 180,168 were unique.

Bowtie mapping using the reference TREU927 provided 12,686,034 reads for the 4E-3 Ab pulldown treated with tetracycline (Tet), of which 8,814,488 were mapped and 199,544 were unique. BWA using TREU927 for the 4E-3 Ab pulldown treated with Tet provided 12,686,034 reads, 5,594,119 mapped reads, while 138,332 were unique. Bowtie mapping using Lister427 for the 4E-3 Ab pulldown treated with Tet provided 12,686,034 reads, 8,817,039 mapped reads, and 199,922 uniquely mapped reads. BWA using Lister427 for the 4E-3 Ab pulldown treated with Tet provided 12,686,034 reads, of which 5,633,827 were mapped and 177,423 were unique.

This indicated that there was a very good coverage of reads, although the percent of uniquely mapped segments of RNA were low, indicating a significant amount of ambiguous mapping of reads.

4.5 BIOINFORMATIC ANALYSIS OF RNASEQ DATA

The preliminary RNAseq data were then analyzed to determine if there were overrepresented mRNAs specific to each sample. To do this, the samples were compared against the 4E-3 antibody pulldown as a specificity control in 4E5v4E3 and as negative controls for 4E3v4E3+CHX and 4E3v4E3+Tet. The first comparison, denoted 4E5v4E3, compared the pulldown of 4E-5p, Sample 1, with the 4E-3p pulldown Sample 2, allowing for the determination of overrepresented mRNAs specific to 4E-5p relative to 4E-3p and vice versa. The second comparison, denoted 4E3v4E3+CHX, compared the pulldown of the untreated 4E-3p, Sample 2, with the 4E-3p pulldown treated with cycloheximide, Sample 3, possibly showing overrepresented mRNAs captured during translation inhibition. The third comparison, denoted

4E3v4E3+Tet, compared the pulldown of the untreated 4E-3p, Sample 2, with the 4E-3 RNAi tetracycline-induced pulldown, Sample 4, to show possible reduction of mRNAs that regulate cell cycle and gene expression when 4E-3p is partially knocked down. The caGEDA program identified 313 mRNAs that were overrepresented in all of the samples. The TriTryps database was used to identify verified mRNAs, while using the BLASTp algorithm to assign putative function of mRNAs that were unidentified, comparing the polypeptide sequences retrieved from the TriTryps database encoded by the mRNA to known eukaryotic and prokaryotic proteins, with a few hypothetical unknowns in each sample.

Table 2. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-5 Ab pulldown.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
Cell Cycle	Tb10.26.0300	6%	S-phase kinase-associated protein 1
	Tb09.244.2520	5%	Bloodstream stage alanine rich protein BARP
General Gene Expression	Tb927.6.110	59%	DNA polymerase
	Tb927.4.510	58%	DNA directed RNA polymerase subunit
	Tb11.57.007	37%	RNA Directed DNA Polymerase
	Tb10.70.6250	27%	Bifunctional DNA Primase/Polymerase
	Tb10.100.0045	20%	Type II Restriction Enzyme Kpn1
Misc	Tb11.01.7005	45%	TRAP dicarboxylate transporter DctQ subunit
	Tb927.7.4680	11%	Phosphoprotein lepp12
Misc Signal	Tb927.1.2985	16%	Multisensor signal transduction histidine kinase
Misc Membrane	Tb927.4.140	105%	Lipoporein release system
	Tb10.389.0760	6%	Membrane transporter
Misc Transporter	Tb11.17.0004	153%	Major facilitator transporter
Misc Metabolism	Tb09.142.036	136%	Catabolite activator protein
	Tb927.7.7560	68%	Peptidase S9
	Tb927.8.3780	47%	Ubiquinone/plastoquinone oxidoreductase chain 6
	Tb927.7.3840	25%	Kinesin-like protein fragment
	Tb10.61.3155	24%	Trans-sialidase
	Tb927.1.4950	20%	Glutathione S-transferase
	Tb927.4.3830	17%	Argininosuccinate synthase
	Tb927.3.2260	6%	Hyaluronate lyase
	Tb10.70.2890	4%	Patatin-like phospholipase domain-containing protein 1-like
	Tb927.6.250	2%	Peptidase M28

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 5 variable surface glycoproteins (VSGs) and 12 hypothetical unknowns were also discovered.

The 4E5v4E3 sample showed a larger number of mRNAs that encode for proteins that assist in gene expression and mRNAs that encode proteins for RNA metabolism in the 4E-3 specific sample (**Tables 2-4**). Among these RNA metabolism mRNAs were 4A, which is part of the eIF4F complex, and a putative DNA directed RNA polymerase II subunit, which produces mRNA from the DNA, with less than 1% higher presence than the baseline, meaning that although there is over a 95% confidence that these are specific for the 4E-3 specific sample, although are not highly overrepresented. There were also a number of mRNAs coding for putative mitochondrial proteins, with up to 45% higher presence than the baseline. Among the other mRNAs specific to the 4E-3 sample were a number of putative membrane and transport proteins.

The 4E-5 specific sample contained putative general gene expression mRNAs, but with an emphasis on polymerases, such as DNA polymerase with a 58% higher presence than the baseline, RNA directed DNA polymerase with a 37% higher presence than the baseline, and a bifunctional DNA primase/polymerase with a 27% higher presence than the baseline. Of note are mRNAs that encode for the putative S-phase kinase-associated protein 1, which assists with cell division, and the verified blood stream alaninerich protein (BARP), which are key for the blood form stage of the parasite, at 6% and 5% higher presence than the baseline, respectively. Both samples contained a roughly equal number of cell cycle, metabolic and variable surface glycoprotein (VSG) mRNAs (**Table 2**).

Table 3. Part 1. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-3 Ab pulldown.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
Cell Cycle	Tb927.4.390	18%	Cell Division Protein
	Tb927.7.5780	6%	GO function: Cell redox homeostasis
RNA Metabolism	Tb927.6.1710	16%	DEAD/DEAH Box Helicase
	Tb427.10.7190	12%	D-tyrosyl tRNA deacylase
	Tb927.7.370	10%	tRNA Ligase Class I
	Tb927.3.5500	<1%	DNA Directed RNA Polymerase II Subunit 3C
General Gene Expression	Tb11.57.007	110%	RNA Directed DNA Polymerase
	Tb927.4.120	91%	Type III Restriction Endonuclease Subunit R
	Tb927.2.900	40%	Type III Restriction Endonuclease Subunit R
	Tb10.70.3330	39%	Protein dpy-30 Homolog
	Tb09.211.0485	35%	DNA Mismatch Repair Protein
	Tb10.70.2020	26%	Zinc Finger Protein Family
	Tb927.3.5230	14%	DNA Repair Protein C
	Tb10.61.0490	12%	rRNA Processing Protein
	Tb927.8.2560	10%	Spliceosomal U5 snRNP-specific Protein
	Tb927.2.1250	8%	Type III Restriction Endonuclease Subunit R
	Tb927.1.1410	7%	Histidine Kinase
	Tb10.70.2550	4%	Transcription Elongation Factor S-II-Related Protein L122
Misc-Mitochondrial	Tb10.406.0615	45%	Dynein Light Chain
	Tb927.8.6590	21%	Protein targeting to mitochondrion/mitochondrial outer membrane
	Tb927.7.5910	10%	ATG8 Ubiquitin-like Protein
	Tb927.6.4580	10%	Mitochondrial SSU Ribosomal Protein
	Tb11.02.3065	2%	Mitochondrial Translocase Subunit
	Tb927.4.830	1%	Cytochrome C Oxidase Assembly Factor 5-like
Misc-Flagellar	Tb10.70.7560	17%	TAX-2 Flagellar Protein
Misc-Membrane	Tb927.4.140	143%	Lipoporein Release System
	Tb927.1.1760	38%	EXS Family Protein
	Tb927.3.3700	28%	Sodium/hydrogen Exchanger
	Tb10.61.0870	25%	SNARE Protein
	Tb11.02.0390	15%	Mechanosensitive Ion Channel Protein Msc5
	Tb11.01.7790	8%	Signal Recognition Particle

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 6 variable surface glycoproteins (VSGs) and 24 hypothetical unknowns were also discovered. *(Continued on Table 4).*

Table 4. Part 2. Continuation from Table 3.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
Misc-Transport Protein	Tb10.389.1570	24%	RAB Interacting Protein
	Tb927.4.4780	14%	Ran-Binding Protein 9 Isoform X3
	Tb927.4.2100	5%	Transferrin Receptor Protein 2 Isoform X1
	Tb09.211.1195	5%	Phosphatylcholine:Ceramide Cholinephosphotransferase 2
	Tb09.244.1910	4%	Phosphate ABC Transporter Permease
	Tb927.8.6570	4%	Binding-Protein-Dependent Transport System Inner Membrane Component
Misc-Metabolism Protein	Tb927.1.360	102%	Peptidase U32
	Tb927.7.7560	68%	Peptidase S9
	Tb09.244.2770	54%	Oxysterol-binding protein-related Protein 6
	Tb927.1.110	36%	Acetylneuraminic Acid Synthase
	Tb927.3.4480	36%	ATP Synthase F1 Subunit Gamma
	Tb927.5.1040	32%	Short-Chain Dehydrogenase/Reductase
	Tb927.5.3740	31%	Ring-box Protein 1
	Tb11.02.0865	25%	Thioesterase Superfamily
	Tb11.01.7560	23%	Glutathione Peroxidase
	Tb09.211.4500	22%	Trans-sialidase
	Tb927.6.110	19%	43% Lysozyme
	Tb927.4.1520	15%	FAM206A
	Tb927.3.2970	14%	Pyridoxal-5'-phosphate-dependent Protein Subunit Beta
	Tb927.4.150	12%	Segregation and Condensation Protein A
	Tb10.389.0300	5%	Mannosyl Transferase II
	Tb927.4.3630	3%	Protein Phosphatase 1
	Tb09.354.0020	3%	Glutamate Synthase
	Tb927.1.4490	2%	Acetyltransferase
	Tb927.7.340	2%	Abhydrolase Domain-Containing Protein 11
	Tb927.7.310	1%	Glutathione S-transferase/Glutaredoxin
	Tb11.01.0290	1%	Carbonic Anhydraselike Protein

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).*

Table 5. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-3 Ab pulldown treated with the translation inhibitor cycloheximide in comparison to the untreated anti-4E-3 Ab pulldown.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
RNA Metabolism	Tb927.1.390	8%	RNA Helicase
Misc	N/A	36%	snRNA U2
Misc-Flagellar	Tb927.7.3840	16%	Kinesin-like Protein Fragment - Microtubule-based Movement
Misc-Membrane	Tb927.1.200	4%	Binding-Protein-Dependent Transport Systems Inner Membrane Component
Misc-Transport	Tb927.1.2985	23%	Multisensor Transduction Histidine Kinase
	Tb927.6.3910	1%	STELAR K+ Outward Rectifier Isoform 2
Misc-Metabolism	Tb927.4.510	55%	Molybdopterin Converting Factor Subunit 2
	Tb927.8.1660	35%	Procyclin-Associated Gene
	Tb927.1.2620	35%	Bariochlorphyll 4-vinyl Reductase
	Tb927.6.110	31%	Lysozyme
	Tb927.4.2420	29%	ATP Binding Transporter
	Tb927.8.5550	21%	Fe(3+) Ions Import ATP-Binding Protein FbpC
	Tb927.1.230	9%	Phosphoterase
	Tb927.1.3510	2%	Integral Membrane Sensor Hybrid Histidine Kinase

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 20 variable surface glycoproteins (VSGs) and 4 hypothetical unknowns were also discovered.

Table 6. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-3 Ab pulldown in comparison to the anti-4E-3 Ab pulldown treated with the translation inhibitor cycloheximide.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
General Gene Expression	Tb927.4.120	22%	Type III Restriction Endonuclease Subunit R
	Tb927.5.1360	21%	Nucleoside 2-deoxyribosyltransferase (NDRT)
Misc	Tb927.1.1400	16%	OPI10-like Protein
	Tb927.7.2730	13%	Leucine Rich Repeat Protein
	Tb927.7.1380	8%	Calpain-like Cysteine Peptidase
	Tb927.4.2530	5%	DoxX Protein
Misc-Mitochondrial	Tb927.8.6570	19%	YLR356Wp - uncharacterized mitochondrial protein
Misc-Signal	Tb927.1.1760	21%	EXS Family Protein
Misc-Membrane	Tb927.6.530	1%	Procyclin Associated Gene 3
Misc-Transport	N/A	26%	Got1 Homolog
	Tb927.8.1320	14%	QA-Snare Protein
	Tb927.8.5900	4%	Trafficking Protein Particle Complex 2
Misc-Metabolism	Tb927.5.1040	5%	NADPH-dependent FMN Reductase
	Tb927.4.1520	1%	FAM206A
	Tb927.4.2610	<1%	Phosphoserine Transaminase

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 1 variable surface glycoproteins (VSGs) and 2 hypothetical unknowns were also discovered.

The 4E3v4E3+CHX comparison was denoted as 4E-3+CHX Increase, for the sample that was treated with CHX, versus 4E-3+CHX Decrease, for the sample that was just the pulldown with the 4E-3 Ab, to determine if the addition of CHX captured additional or unique mRNAs (**Tables 5 and 6**). Of note in the 4E-3+CHX Increase sample was the presence of mRNA encoding for a putative RNA helicase, with an 8% higher presence than the baseline, and a large number of putative metabolic mRNAs and VSGs (**Table 5**).

In the sample 4E-3+CHX Decrease sample, there was a large number of miscellaneous mRNAs and two general gene expression mRNAs. The general gene expression mRNAs were a putative nucleoside 2-deoxyribosyltransferase (NDRT), with a 21% higher presence than the baseline, and a verified Type III restriction endonuclease subunit R, with a 22% higher presence than the baseline, the latter of which was also in the 4E-3 sample in the 4E5v4E3 comparison (**Table 6**).

In the comparison 4E3v4E3+Tet, the samples were denoted as 4E-3+Tet Increase, to denote the mRNAs increased in the presence of Tet, and 4E-3+Tet Decrease, to denote the mRNAs decreased in the presence of Tet in the 4E-3 pulldown that was untreated (**Tables 7 and 8**). In the 4E-3+Tet Increase sample, there was a large number of VSG and putative metabolism mRNAs. There was the presence of two putative RNA metabolism mRNAs and mRNAs that encode for putative gene expression proteins, however only one putative general gene expression mRNA was of note, which is the putative RNA-editing complex, with a less than 1% higher presence than the baseline (**Table 7**).

For the sample 4E-3+Tet Decrease, there was the presence of a large number of general gene expression mRNAs, such as the putative DNA-binding helix-turn-helix protein, with a 2% higher presence than the baseline, and putative DNA directed RNA polymerase II subunit mRNAs, with 27% higher presence than the baseline, a few putative mitochondrial mRNAs, and putative cell cycle mRNAs. Of extreme importance was the presence of a verified 4E mRNA, however the identity is undetermined at this time (**Table 8**).

From these data, a general sense of 4E-5p and 4E-3p can be suggested through their specific mRNA targets. The 4E-5p appeared to associate with mRNAs encoding for proteins key for cell division, specifically DNA polymerase and S-phase kinase-associated protein 1, while also associating with mRNAs that encode for transcriptional regulators, such as the RNA polymerase and the DNA primase. In comparison, 4E-3p appeared to associate with mRNAs that encode for proteins that perform a variety of cell functions, such as the 4A and the DNA and RNA polymerases which are key for transcription and translation, and lysozymes and peptidases that assist with cell metabolism and general function.

Table 7. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-3 Ab pulldown treated with tetracycline to partially knockdown 4E-3p in comparison to the anti-4E-3 Ab pulldown.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
RNA Metabolism	Tb927.1.1490	20%	Branched-chain Amino Acid Transport System Carrier Protein
	Tb927.1.2870	17%	Peptidase M20
General Gene Expression	Tb927.6.810	29%	Y53C12B.7
	Tb927.1.2860	11%	Transposon Tn7 Transposition Protein TnsC
	Tb927.8.680	<1%	RNA-Editing Complex (KREPA 5)
Misc	Tb927.7.7550	63%	Potassium Channel Subfamily K Member 16
	Tb927.1.3570	23%	Arabinose Efflux Permease Family Protein
	Tb927.1.3720	18%	Probable Protein Phosphatase 2C 4-like
Misc-Mitochondrial	Tb927.1.1810	44%	Cytochrome C Oxidase Subunit II, Partial
Misc-Flagellar	Tb927.1.1860	24%	Flagellar Biosynthetic Protein FlhB
	Tb927.7.3840	15%	Kinesin-like Protein Fragment
Misc-Membrane Protein	Tb927.4.140	67%	Lipoporein Release System
	Tb927.1.200	42%	Binding-Protein-Dependent Transport System, Inner Membrane Component
	Tb927.1.3510	5%	Integral Membrane Sensor Hybrid Histidine Kinase
Misc-Transport	Tb927.1.4270	17%	Protein Transport Protein SEC13
	Tb927.1.4950	14%	ABC Transporter, ATP-binding Protein
Misc-Metabolism	Tb927.1.3750	89%	Exopolysaccharide Biosynthesis Protein
	Tb927.7.6870	77%	Exopolysaccharide Biosynthesis Protein
	Tb09.v2.0450	60%	Exopolysaccharide Biosynthesis Protein
	Tb.927.1.3610	55%	Nitrite Reductase
	Tb927.1.3350	53%	Halocyanin
	Tb927.1.110	39%	Acetylneuraminic Acid Synthetase
	Tb927.8.1660	37%	Procyclin-Associated Gene
	Tb927.6.110	32%	Lysozyme
	Tb927.1.4670	15%	Ca++-ATPase
	Tb927.6.260	14%	Long-Chain-Fatty-Acid-CoA Ligase
	N/A	10%	Ferrodoxin
	Tb927.1.3630	9%	GTPase-activator Protein for Ras Family GTPase
	Tb11.40.0010	7%	Glycosyltransferase
	Tb927.4.3830	4%	Argininosuccinate Synthase
	Tb927.7.6020	3%	Alpha-Amylase Family Protein

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 32 variable surface glycoproteins (VSGs) and 11 hypothetical unknowns were also discovered.

Table 8. Overrepresented mRNAs discovered by RNAseq and caGEDA in the anti-4E-3 Ab in comparison to the anti-4E-3 Ab pulldown treated with tetracycline to partially knockdown 4E-3p.

Predicted Encoded Protein Function	Gene Identifier	Quantity Greater Than Threshold	Predicted Encoded Protein
Cell Cycle	Tb927.8.2320	30%	Dedicator of Cytokinesis 11-like
	Tb927.6.5020	8%	Cyclin 7
RNA Metabolism	Tb927.6.420	55%	Non-ribosomal Peptide Synthase
	N/A	8%	4E
General Gene Expression	Tb927.8.4830	47%	U1 snRNP
	N/A	27%	DNA-Directed RNA Polymerase II Subunit
	Tb927.8.4130	23%	Class I Transcription Factor A Subunit
	N/A	9%	snRNP
	Tb927.5.4030	3%	U6 snRNA-associated Sm-like Protein Lsm7p
	Tb927.6.590	2%	DNA-binding Helix-turn-helix Protein
Misc	Tb927.7.5010	19%	Arabinose Efflux Permease Family Protein
	Tb927.8.4630	16%	Rho Guanine Nucleotide Exchange Factor 5
	Tb927.7.260	11%	Chemotaxis Protein
	Tb927.4.3960	9%	Relaxase
Misc-Mitochondrial	Tb927.7.2690	43%	Mitochondrial Part
	Tb927.8.6190	16%	Mitochondrial Conserved Eukaryotic Protein
	Tb927.5.3740	15%	E3 Ubiquitin Ligase Protein
Misc-Membrane	Tb927.1.650	36%	Membrane Protein
Misc-Transport	N/A	24%	Protein Transport Protein Sec61 Gamma Subunit
	Tb927.7.5900	21%	ATG8/AUG7/APG8/PAZ2
	Tb927.7.5910	20%	ATG8/AUG7/APG8/PAZ2
Misc-Metabolism	Tb927.6.2120	52%	Beta-Glucosidase-like Glycosylhydrolase
	Tb927.6.2990	50%	Methylenetetrahydrofolate Dehydrogenase
	N/A	46%	Dolichyl-phosphate Beta-D-Mannosyltransferase
	Tb927.1.1770	31%	Glutaredoxin
	Tb927.8.4130	23%	Histidinol-Phosphate Aminotransferase
	Tb927.8.5320	4%	L-arabinose Utilization Protein, Glycerol Dehydrogenase

The mRNA identities were verified via the TriTryps database. Any mRNAs not identified by TriTryps were then analyzed for their predicted encoded polypeptide sequences, which were examined for homology with other known eukaryotic and prokaryotic proteins via the BLASTp algorithm. The first column shows the predicted encoded protein function, based off of either the verified protein's function or the predicted homologous protein's function. The second lane is the gene identifier provided by the Bioinformatics Core Facility at the University of Pittsburgh that identifies which mRNAs were overrepresented. The quantity greater than threshold is the amount of the overrepresented mRNA above the threshold. The threshold was the cutoff value that ensured only overrepresented mRNAs were selected in each sample by caGEDA. Any value lower than that threshold represented a nonspecific mRNA. The predicted encoded protein is the verified or putative protein encoded by the overrepresented mRNA. *(See text for in depth analysis).* A total of 0 variable surface glycoproteins (VSGs) and 1 hypothetical unknown were also discovered.

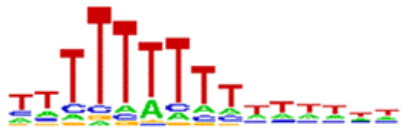
4.6 COMPARISON OF 3’ UTRS TO DETERMINE COMMON MOTIFS

Once overrepresented mRNAs specific to each sample were determined and verified through RNAseq, analyzing the common motifs was crucial to determine whether 4E-3p and 4E-5p act on regulons by interaction with mRNA 3' untranslated regions (UTR), separately or in sequence, for cell cycle and cell growth. Since all of the genes are transcribed at once in trypanosomes, a majority of the transcripts are stored in transcription vesicles, activated later through environmental factors [55]. This would require hypothetical common motifs shared amongst functionally relevant mRNA 3' UTRs, allowing for recognition by translation initiation proteins or ribonucleic proteins (RNPs). The key argument behind the search for common motifs was based on previous evidence that 4E homologs are required for mRNA transportation out of the nucleus, translation initiation and regulation, and mRNA storage [24][53][54]. With 4E-3p and 4E-5p being shown to be key to cell survival in previous papers [S.Tarun and J. Rocco, data not shown], the mass spectrometry (MS) results showing association of 4E-3p with other members of the eIF4F complex, and RNAseq showing the association of 4E-3p and 4E-5p to distinct mRNAs encoding for key proteins related to cell survival and proteins production, 4E-3p and 4E-5p may interact with their own unique, conserved motifs present in the 3’ UTR of mRNA to assist in recognition and selection of transcripts required by the cell.

The program Trawler (EMBL)(http://ani.embl.de/trawler/) was selected to analyze functional sites in the 3’ UTR of the mRNAs based on previous experiments in trypanosomes [46][47]. Trawler’s intended use was to analyze large scale chromatin-immunoprecipitation data (ChIP), however RNAseq data can be used as well. Because of this, any uracils present in the mRNA are represented by thymines in the motif data presented. The key inputs for Trawler

needed to be in FASTA formatting, obtained via the TriTryp.db website. The gene identifier of each overrepresented mRNA was entered into the geneID search, grouped by which sample the mRNA was overrepresented in, and then the 3' UTRs were selected from the translation stop to the transcription stop codons. These data were then analyzed using Trawler against the reference sequence, also in FASTA formatting, of the entire T. *brucei* genome. The parameters for motif analysis were left with the default settings, and are as follows: the minimum occurrence of the motif in the sample was 10; the motif length was a minimum of 6 nucleotides; and the strand selection was left at single-stranded since this was mRNA.

A



Family 1 4E-5 Specific

B

4E-5 Specific Family 1			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
Cell Cycle	Tb10.26.0300 From 19-178nt	S-phase Kinase -Associated Protein 1	Putative
	Tb09.244.2520 From 7-468nt	**Bloodstream stage alaninerich protein BARP**	**Verified**
General Gene Expression	Tb10.100.0045 From 84-90nt	Type II Restriction Enzyme KpnI	Putative
	Tb10.70.6250 From 1-65nt	Bifunctional DNA Primase/Polymerase	Putative
	Tb927.4.510 From 13-297nt	DNA Directed RNA Polymerase Subunit	Putative
Misc	Tb927.7.4680 From 22-155nt	Phosphoprotein lepp12	Putative
	Tb11.01.7005 From 0-130nt	TRAP Dicarboxylate Transporter DctQ Subunit	Putative
Misc-Signal Protein	Tb927.1.2985 From 0-262nt	Multisensor Signal Transduction Histidine Kinase	Putative
Misc-Metabolism Protein	Tb927.8.3780 From 44-373nt	Ubiquinone/plastoquinone Oxidoreductase Chain 6	Putative
	Tb10.61.3155 From 10-167nt	Trans-sialidase	Putative
	Tb927.3.2260 From 4-32nt	Hyaluronate Lyase	Putative

Figure 16. Motifs specific for mRNA overrepresented in 4E-5.

(A) Graphical representation of mRNA motif nucleotide composition for mRNAs overrepresented in the anti-4E-5 Ab pull down in comparison to the anti-4E-3 Ab pull down. (B) List of mRNAs pulled down with their function, gene identifier, name, and whether they are experimentally **Verified** or Putative.

The output from Trawler contained a graphical depiction of each motif family specific for the 3' UTRs from mRNA overrepresented in each sample, the possible identity of the motif based off of motifs in other organisms, or percent relevance, and the motif location for each mRNA. The possible motif identity was not relevant to this study due to the identifications

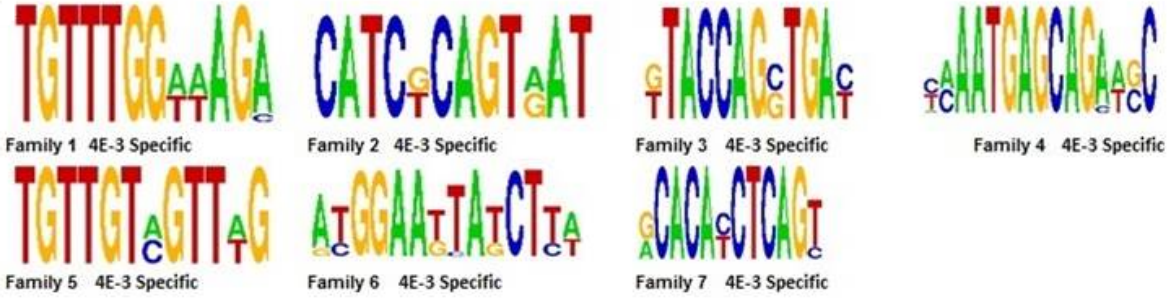
being DNA motifs, however they could reveal associations of ribonucleic proteins (RNPs) or eIF4Es with the double stranded DNA, and have been included in the supplement. Each motif length was designated 1 or -1, with the former representing the motif being from the 5'-3' strand, and the latter representing the 3'-5' strand, due to Trawler comparing the single stranded sequences to the double stranded background sequence and providing both directional strands for reference. Since the samples were single-stranded mRNA, only motifs exhibiting a positive value were selected. While not every mRNA that was determined to be overrepresented in each sample was present in the Trawler output and shared common 3' motifs, a large majority were.

The 4E-5 specific sample, from the 4E5v4E3 comparison, showed only one common motif family, denoted 4E-5 Specific Family 1 (**Figure 16A**). This motif family was heavily composed of uracils, again depicted by Trawler using thymines. The motif was also commonly found near the beginning of the 3'UTR, with multiple copies existing from the beginning to the middle of the 3' UTR in each mRNA (**Figure 16B**).

The 4E-5 Specific Family 1 motif was identified by Trawler to be similar to a wide range of DNA motifs, with a few being irrelevant and related to embryogenesis. Of note, the motif identity with the highest divergence was MA0054, having no known function aside from transcription factor binding [56]. The motif also had a divergence of 0.77 with a nuclear receptor, allowing for the up or down regulation of the mRNA through interaction with nuclear proteins [57]. The DNA motif identities with a divergence higher than 0.600 were all related to cell proliferation, differentiation, cell cycle control, cell migration, apoptosis, or expression regulation (**Table S1**). The proteins encoded by the mRNAs that shared the 4E-5 Specific

Family 1 motif had predicted functions ranging from cell cycle regulation to general gene expression to metabolism. Of note are the putative type II restriction enzyme KpnI, a putative bifunctional DNA primase/polymerase, a putative DNA directed RNA polymerase subunit, and a putative ubiquinone/plastiquinone oxidoreductase chain 6 (**Table S1**).

A



B

4E-3 Specific Families 1-7			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
General Gene Expression	Tb927.2.1250 From 577-647nt F1	Type III Restriction Endonuclease Subunit R	Putative
	From 731-764nt F2		
	From 469-546nt F3		
	From 16-325nt F4		
	From 682-872nt F5		
	From 110-559nt F6		
	From 148-728nt F7		
	Tb927.4.120 From 573-643nt F1	Type III Restriction Endonuclease Subunit R	Putative
	From 731-764nt F2		
	From 465-543nt F3		
	From 16-323nt F4		
	From 678-868nt F5		
	From 110-555nt F6		
	From 148-724nt F7		
	Tb927.2.900 From 577-647nt F1	Type III Restriction Endonuclease Subunit R	Putative
	From 731-764nt F2		
	From 469-546nt F3		
	From 16-325nt F4		
	From 682-872nt F5		
	From 110-559nt F6		
	From 148-728nt F7		

4E-3 Specific Family 6			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
Misc-Mitochondrial Protein	Tb927.7.5910 From 200nt	ATG8 Ubiquitin-like Protein	Putative

Figure 17. Motifs specific for mRNA overrepresented in 4E-3.

(A) Graphical representation of mRNA motif nucleotide composition for mRNAs overrepresented in the anti-4E-3 Ab pull down in comparison to the anti-4E-5 Ab pull down. (B) List of mRNAs pulled down with their function, gene identifier, name, and whether they are experimentally **Verified** or Putative.

The 4E-3 specific sample, from the 4E5v4E3 comparison, showed seven different families of shared motifs, each having a roughly equal composition of purines and pyrimidines (**Figure 17A**). Three different mRNAs, each encoding for a verified type III restriction endonuclease subunit R, were found to have each family of motifs located in the same area and within the same range in each mRNA. A fourth mRNA, encoding for an ATG8 ubiquitin-like protein, contained only 4E-3 Specific Family 6's motif (**Figure 17B**).

Trawler discovered only one possible DNA motif related to 4E-3 Specific Family 1, Sox-5, which has embryonic functions and plays a role in cell fate determination, with a divergence of 0.808 [58]. 4E-3 Specific Family 2 had two possible DNA motif matches, Ma0095 and MA0094, both with unknown functions aside from transcription factor binding, and divergences of 0.661 for the former and 0.585 for the latter [56]. 4E-3 Specific Family 3 had three possible DNA motif matches, the one with the highest divergence being bZIP-CREB at 0.898, which plays a role in regulation and promotion of expression [59]. The next highest, at a divergence of 0.862, was c-FOS, which converts extracellular signals into changes of gene expression [60-62]. The third and final possible identity for Family 3 was MA0086, at a divergence of 0.637, however having no known function aside from transcription factor binding [56]. 4E-3 Specific Family 6 had two possible DNA motif matches for motif identification, Ubx and S8. Ubx, at a divergence of 0.571, allows for ubiquitin protein binding and regulates alternative splicing, while S8, with a divergence of 0.415, which binds to the S section of the 30S ribosomal subunit (**Table S2**) [63][64].

A

Family 1 4E-3 + CHX Increased Specific

B

4E-3 + CHX Increased Specific Family 1			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
Misc-Transport Protein	Tb927.1.2985 From 204-278nt	Multisensor Transduction Histidine Kinase	Putative
	Tb927.6.3910 From 11-257nt	STELAR K+ Outward Rectifier Isoform 2	Putative
Misc-Metabolism Protein	Tb927.4.510 From 12-139nt	Molybdopterin Converting Factor Subunit 2	Putative
	Tb927.8.5550 From 153-631nt	Fe(3+) Ions Import ATP-Binding Protein FbpC	Putative

Figure 18. Motifs specific for mRNA overrepresented in 4E-3 in the presence of CHX.

(A) Graphical representation of mRNA motif nucleotide composition for mRNAs overrepresented in the anti-4E-3 Ab pull down treated with cycloheximide (CHX) in comparison to the anti-4E-3 Ab pull down. (B) List of mRNAs pulled down with their function, gene identifier, name, and whether they are experimentally **Verified** or Putative.

The 4E-3 sample treated with CHX, from the 4E3v4E3+CHX comparison and representing the mRNAs increased when treated with CHX, showed only one common motif family, which was rich in uracils and adenines (**Figure 18A**). Sharing this motif, 4E-3 +CHX Specific Family 1, were only four mRNAs, encoding for a putative multisensory transduction histidine kinase, a putative STELAR K+ outward rectifies isoform 2, a putative molybdopterin converting factor subunit 2, and a putative Fe(3+) ion import ATP-binding protein FbpC. The mRNAs encoding for the putative multisensory transduction histidine kinase and putative Fe(3+) ion import ATP-binding protein FbpC had the motif located roughly in the center of the 3' UTR, while the mRNAs encoding for the putative STELAR K+ outward rectifier isoform 2 and putative molybdopterin converting factor subunit 2 had the motif located near the start of the 3' UTR (**Figure 18B**).

A large number of possible DNA motif identities were provided by Trawler, with varying divergences. Of note were: bZIP- CEBP, which assists in regulation and promotion of expression, at a divergence of 0.769; TRP MYB, which assists in secondary metabolism and cell shape and fate, at a divergence of 0.727; homeo box, which encodes for transcription factors that typically turn on other genes, at a divergence of 0.578; MA0033, with no known function, at a divergence of 0.574; HMG-box, which allows for high mobility group (HMG) protein binding to affect formation and function of transcription complexes, at a divergence of 0.546; and Ubx, which allows for ubiquitin protein binding to regulate alternative splicing, at a divergence of 0.483 (**Table S3**) [56][59][63][65-67].

A

Family 1 4E-3 + CHX Decreased Specific

B

4E-3 + CHX Decreased Specific Family 1			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
General Gene Expression	Tb927.4.120 From 7-877nt	Type III Restriction Endonuclease Subunit R***	Putative
	Tb927.5.1360 From 8-776nt	**Nucleoside 2-deoxyribosyltransferase (NDRT)**	**Verified**
Misc	Tb927.7.1380 From 16-143nt	Calpain-like Cysteine Peptidase	Putative
	Tb927.4.2530 From 47-674nt	DoxX Protein	Putative
Misc-Mitochondrial	Tb927.8.6570 From 3-1052nt	YLR356Wp - uncharacterized mitochondrial protein	Putative
Misc-Membrane Protein	**Tb927.6.530 From 3-9nt**	**Procyclin Associated Gene 3**	**Verified**
Misc-Transport Protein	Tb927.8.1320 From 1-1075nt	QA-Snare Protein	Putative
	Tb927.8.5900 From 19-384nt	Trafficking Portein Particle Complex 2	Putative
Misc-Metabolism Protein	Tb927.5.1040 From 9-125nt	NADPH-dependent FMN Reductase	Putative

Figure 19. Motifs specific for mRNA overrepresented in 4E-3 in the absence of CHX.

(A) Graphical representation of mRNA motif nucleotide composition for mRNAs overrepresented in the anti-4E-3 Ab pull down in comparison to the anti-4E-3 Ab pull down treated with cycloheximide (CHX). (B) List of mRNAs pulled down with their function, gene identifier, name, and whether they are experimentally **Verified** or Putative.

The 4E-3 sample left untreated from the 4E3v4E3+CHX comparison, which denoted the mRNAs that were decreased in the presence of CHX, showed only one common motif family (**Figure 19A**). This common motif, 4E-3 –CHX Specific Family 1, had a large uracil presence

with very little else. The three mRNAs of note in this sample encoded for a verified type III restriction endonuclease subunit R, a putative nucleoside 2-deoxyribosyltransferase, and YLR356Qp, which is a putative uncharacterized mitochondrial protein (**Figure 19B**).

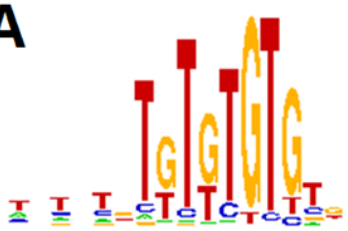
There were a large number of possible hits for DNA motif identification, with a majority functioning as expression regulators or cell cycle regulators. Of note are: REL Class motif at a divergence of 0.8, which regulates transcription factors; bZIP CEPB at a divergence of 0.799, which assists in regulation/promotion of expression; homeo box at a divergence of 0.73, which encodes for transcription factors that turn on other genes; MADS box at a divergence of 0.518, which assists in cell proliferation and is common to eukaryotes; cEBP at a divergence of 0.428, which assists in regulation/promotion of expression; TRP MYB at a divergence of 0.337, which assists with secondary metabolism; and HMG at a divergence of 0.287, which, with the binding of HMG proteins, affects the formation and function of transcription complexes (**Table S4**) [59][66-71].

The 4E-3 sample treated with Tet in the 4E3v4E3+Tet comparison, which denoted the mRNAs increased in the presence of Tet, showed no shared 3' UTR motifs in their mRNA. This is due to the fact that most of the overrepresented mRNA lacked 3' UTRs larger than one nucleotide, which showed no indication of being conserved.

The 4E-3 sample left untreated in the 4E3v4E3+Tet comparison, which denoted the mRNAs decreased in the presence of Tet, showed only one common motif family that was rich in alternating uracils and guanines (**Figure 20A**). A location pattern did not specifically exist for

the mRNAs found to share the common motif, 4E-3 –Tet Specific Family 1, however there was a general trend for the motif to be located towards the beginning or the middle of the 3' UTR (**Figure 20B**). The key mRNAs decreased in the presence of Tet with this shared motif encode for mitochondrial, general gene expression, and cell cycle regulating proteins, such as a putative DNA-binding helix-turn-helix protein and a putative mitochondrial conserved eukaryotic protein (**Figure 20B**).

A



Family 1 4E-3 +Tet Decrease Specific

B

4E-3 +Tet Decrease Specific Family 1			
Predicted Function	**Gene Name**	**Predicted Protein**	**Putative/Experimentally Verified**
Cell Cycle	Tb927.8.2320 From 157nt	Dedicator of Cytokinesis 11-like	Putative
	Tb927.6.5020 From 69-150nt	Cyclin 7	Putative
General Gene Expression	Tb927.6.590 From 248-440nt	DNA-binding Helix-turn-helix Protein	Putative
Misc	Tb927.7.5010 From 7-290nt	Arabinose Efflux Permease Family Protein	Putative
	Tb927.8.4630 From 14nt	Rho Guanine Nucleotide Exchange Factor 5	Putative
	Tb927.4.3960 From 50-391nt	Relaxase	Putative
Misc-Mitochondrial Protein	Tb927.5.3740 From 114-1059nt	E3 Ubiquitin Ligase Protein	Putative
	Tb927.8.6190 From 59-568nt	Mitochondrial Conserved Eukaryotic Protein	Putative
Misc-Membrane Protein	Tb927.1.650 From 124nt	Membrane Protein	Putative
Misc-Metabolism Protein	Tb927.7.5910 From 45-831nt	ATG8/AUG7/APG8/PAZ2	Putative
	Tb927.7.5900 From 158-762	ATG8/AUG7/APG8/PAZ2	Putative

Figure 20. Motifs specific for mRNA overrepresented in 4E-3 in the presence of Tet.

(A) Graphical representation of mRNA motif nucleotide composition for mRNAs overrepresented in the anti-4E-3 Ab pull down treated with tetracycline (Tet) in comparison to the anti-4E-3 Ab pull down. (B) List of mRNAs pulled down with their function, gene identifier, name, and whether they are experimentally **Verified** or Putative.

There were a few possible DNA motif identities of note, such as: Mad box at a divergence of 0.858, which binds MADS transcription factors and plays a role in cell proliferation; bHLH at a divergence of 0.821, which allows dimerization, DNA binding, and

assists in the binding of transcription factors; TRP MYB at a divergence of 0.537, which is responsible for secondary metabolism; and HMG at a divergence of 0.518, which, with the binding of HMG proteins, affects the formation and transcription complexes (**Table S5**) [56][57][71][72].

4.7 SECONDARY STRUCTURE ANALYSIS FOR MOTIF LOCATION

The 3' UTRs of groups of mRNAs exhibiting conserved motifs were analyzed using the RNAfold web server through the Vienna RNA Websuite to determine possible shared structural elements. In general, each secondary structure contained an abundance of hair-pin loops and pseudo-knots. Provided alongside the secondary structures were the percent frequency, which is the likelihood that one structure will exist at 37°C in comparison to the overall structure. For each secondary structure analyzed, the percent frequency was zero or close to zero, meaning that the structure is highly unlikely to exist as a whole, however the 3' UTR may form individual microstructures. Also provided was a color scale for each secondary structure, with red representing the highest probability that a nucleotide is in the right structure, and purple representing the lowest probability.

The first set of 3' UTRs analyzed were the TbeIF4E-5 (4E-5) specific mRNAs from the TbeIF4E-5 pulldown using the 4E-5 antibody (4E-5 ab) comparison with the TbeIF4E-3 (4E-3) pulldown using the 4E-3 antibody (4E-3 ab) (4E5v4E3). A majority of the secondary structures were comprised of stem-loops based around a central loop, with a higher base-pair probability the further the nucleotides were from the general center of the structure, denoted by the yellow to

red colors. The shared motifs had a heavy presence on the protruding stem-loops, with a total of 99.6% existing on bulges and loops, with the starting nucleotide in the motif beginning on the double-helix stem, in most cases, while terminating at or near the end of a single-strand bulge **(Figure 21, S1-11).** There did appear to be subsets of mRNA 3' UTRs in the 4E-5 sample, such as Tb927.7.4680 and Tb927.10.150, which share similar overall design (**Fig. 21 D & H, S4 & 8**).

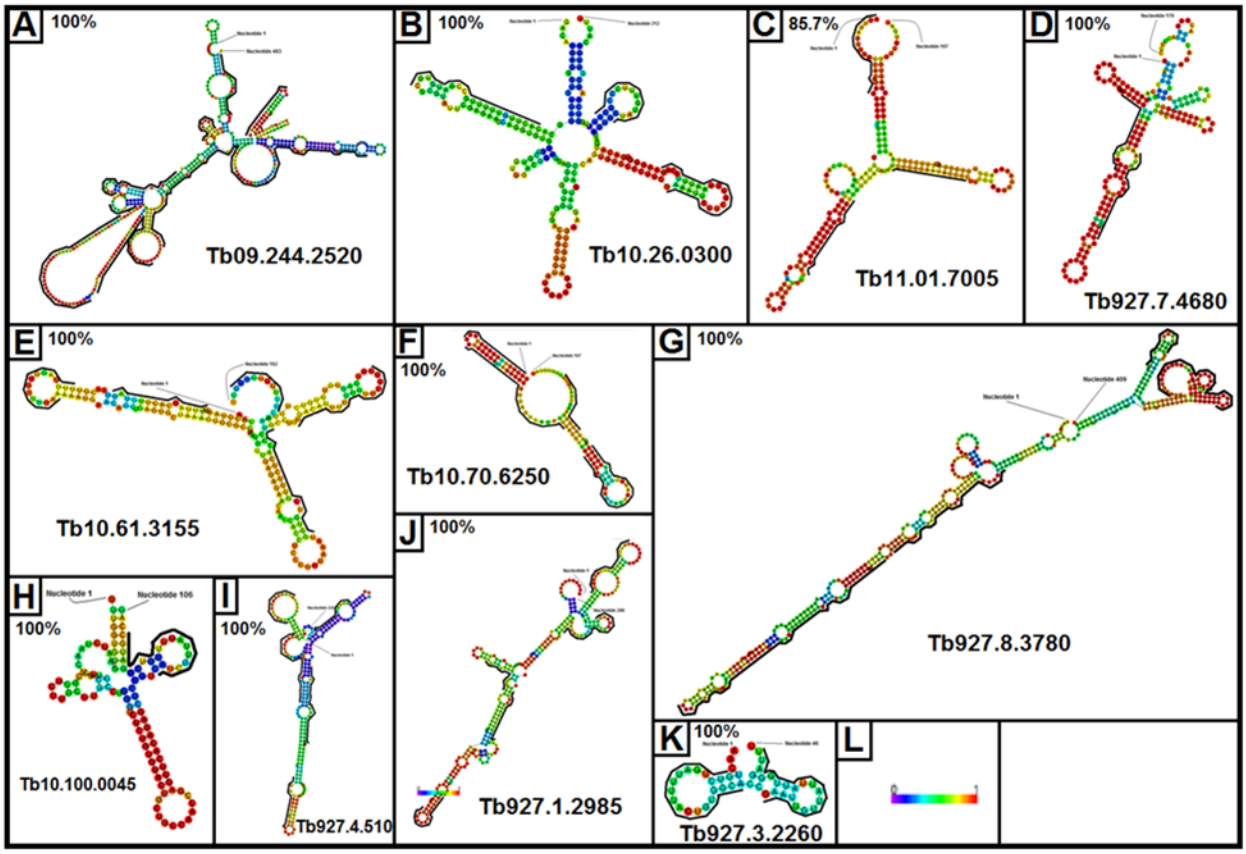


Figure 21. 4E-5 specific 3' UTR secondary structure.

(A-K) Visual representation of 3' UTRs for mRNA sharing the common motif in the 4E-5 specific untreated sample. Each 3' UTR structure is depicted in the minimum free energy configuration. Black bars represent motif location, in each case extending from 3'-5'. The first and last nucleotides are labeled for each structure. (L) Color guide representing the base-pair probability, where purple is the least likely and red is the most likely a nucleotide is in the proper place. The percent of motifs existing on bulges/loops is depicted in the upper right corner of each panel.

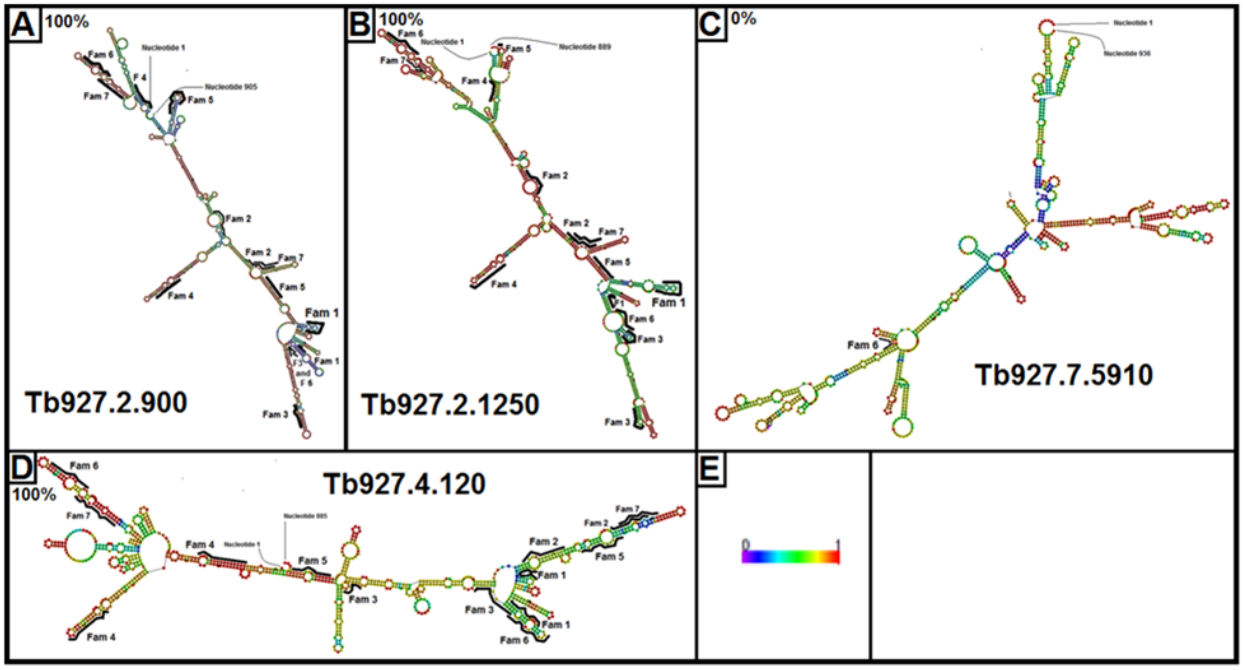


Figure 22. 4E-3 specific 3' UTR secondary structure.

(A-D) Visual representation of 3' UTRs for mRNA sharing the common motif in the 4E-3 specific untreated sample. Each 3' UTR structure is depicted in the minimum free energy configuration. Black bars represent motif location, in each case extending from 3'-5'. The first and last nucleotides are labeled for each structure. (E) Color guide representing the base-pair probability, where purple is the least likely and red is the most likely a nucleotide is in the proper place. The percent of motifs existing on bulges/loops is depicted in the upper right corner of each panel.

The mRNA sharing a common motif from the 4E-3 sample in the 4E5v4E3 comparison have a general overall structure comprised of numerous bulges and pseudo-knots, with the highest base-pair probability existing primarily in the stems (**Figure 22, S12-15**). The 3' UTRs of the mRNAs encoded by Tb927.2.900, Tb927.2.1250 and Tb927.4.120 shared the closest similarities, with all seven of the motif families being represented in generally the same locations, with a total of 97.7% of all motifs existing on bulges and loops (**Figure 22 A-B & D**). A majority of the motif families existed on the stem-loop structures projecting from pseudo-knots, with the shared motifs consisting of one or more bulges. The 3' UTR encoded by

Tb927.7.5910 is the only outlier, having only the Family 6 motif represented, however the structure remains similar (**Figure 22 C**).

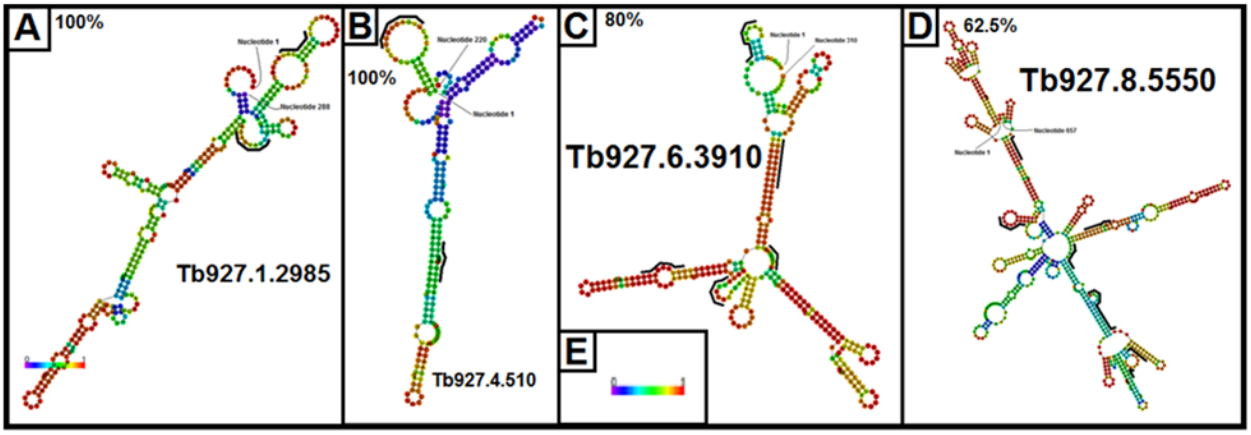


Figure 23. 4E-3+CHX Increase specific 3' UTR secondary structure.

(A-D) Visual representation of 3' UTRs for mRNA sharing the common motif in the 4E-3 specific CHX treated sample. Each 3' UTR structure is depicted in the minimum free energy configuration. Black bars represent motif location, in each case extending from 3'-5'. The first and last nucleotides are labeled for each structure. (D) Color guide representing the base-pair probability, where purple is the least likely and red is the most likely a nucleotide is in the proper place. The percent of motifs existing on bulges/loops is depicted in the upper right corner of each panel.

The sample comparison between 4E-3 untreated and 4E-3 treated with cycloheximide (CHX)(4E-3+CHX) was analyzed individually, with the 4E-3 untreated sample representing a decrease in mRNA and the 4E-3+CHX sample representing an increase in mRNA (4E3v4E3+CHX) (**Figure 23, S16-19, 24, S20-28**). The 4E-3+CHX sample mRNAs shared few commonalities between each other, but still contained the pseudo-knots and stem-loop structures, with a total of 76.5% existing on bulges and loops **(Figure 23, S16-19).** The base-pair stability lacked a shared theme, being most stable on the stems, loops and bulges. The shared 3' UTR motifs consisted of these loops and bulges, extending on to double helical stem structures.

The structure of the 3' UTRs in the 4E-3 mRNA that were left untreated in the 4E3v4E3+CHX comparison exhibit a wide range of complexity with few overall structural similarities, containing pseudo-knots and stem-loops. The highest base-pair probabilities existed primarily on the stem structures, while 73.4% of the common motif is comprised mainly of loops and bulges (**Figure 24, S20-28**).

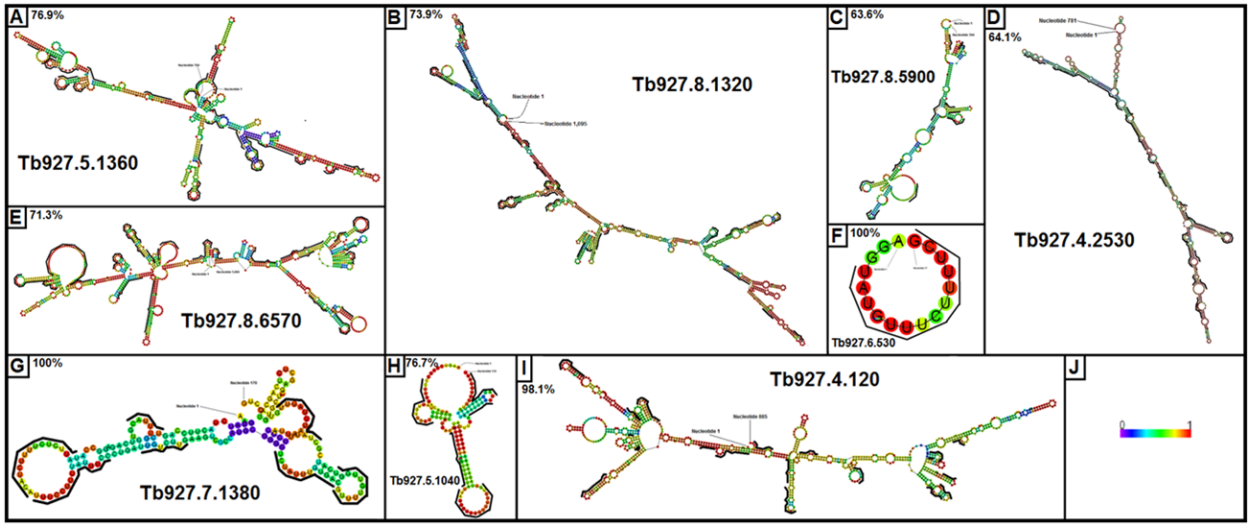


Figure 24. 4E-3+CHX Decrease specific 3' UTR secondary structure.

(A-I) Visual representation of 3' UTRs for mRNA sharing the common motif in the 4E-3 specific CHX untreated sample. Each 3' UTR structure is depicted in the minimum free energy configuration. Black bars represent motif location, in each case extending from 3'-5'. The first and last nucleotides are labeled for each structure. (J) Color guide representing the base-pair probability, where purple is the least likely and red is the most likely a nucleotide is in the proper place. The percent of motifs existing on bulges/loops is depicted in the upper right corner of each panel.

Due to the lack of shared motifs amongst the tetracycline (TET) induced 4E-3 sample in comparison to the 4E-3 untreated (4E3v4E3+TET), the 4E-3 untreated sample had the only mRNAs available for structural analysis (**Figure 25, S29-37**). The 3' UTR secondary structures shared complexity, exhibiting pseudo-knots, stems, bulges and loops. The main bulk of stem-loop structures were off-shoots of pseudo-knot structures, while also maintaining the highest base-pair probabilities. 99.3% of the shared 3' UTR motifs primarily consisted of loops and bulges, extending on to stem structures. In two mRNA 3' UTRs, Tb927.5.3740 and Tb927.7.5900, the shared motifs consisted of pseudo-knot structures (**Figure 25 C-D, S29-37**).

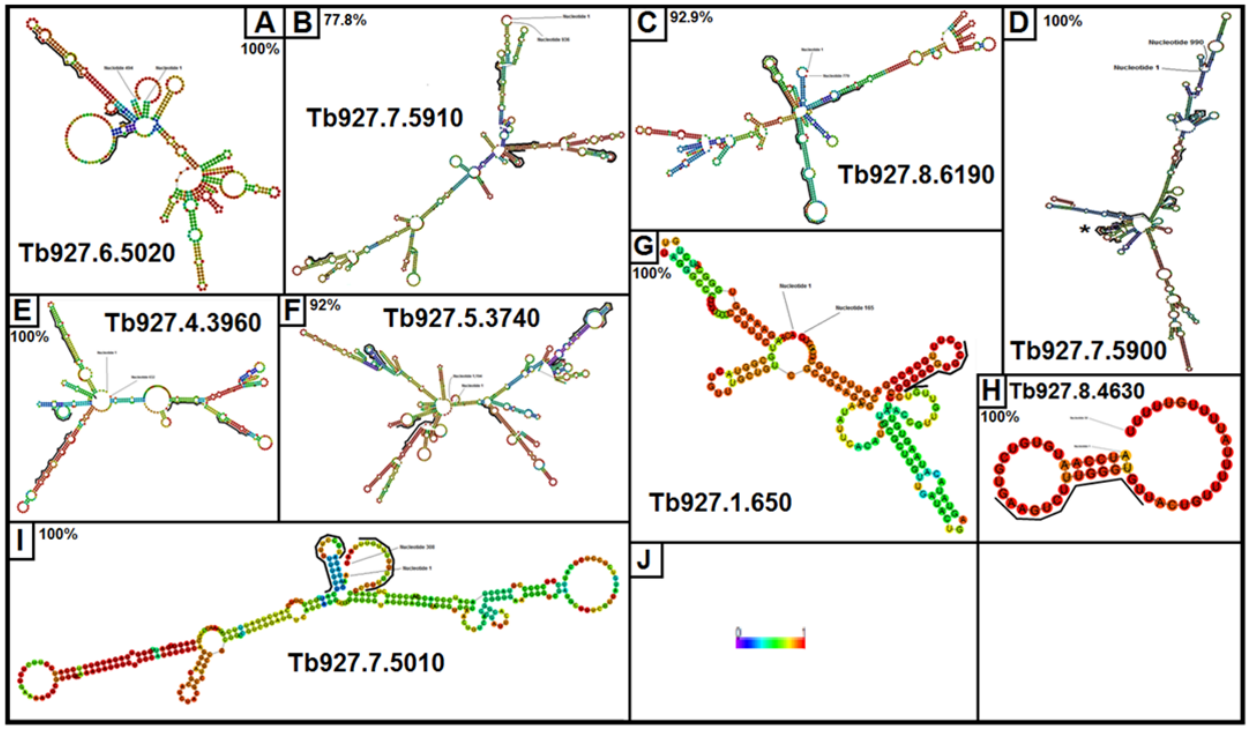


Figure 25. 4E-3+Tet Decrease specific 3' UTR secondary structure.

(A-D) Visual representation of 3' UTRs for mRNA sharing the common motif in the 4E-3 specific Tet untreated sample. Each 3' UTR structure is depicted in the minimum free energy configuration. Black bars represent motif location, in each case extending from 3'-5'. The first and last nucleotides are labeled for each structure. (D) Color guide representing the base-pair probability, where purple is the least likely and red is the most likely a nucleotide is in the proper place. The percent of motifs existing on bulges/loops is depicted in the upper right corner of each panel.

5.0 DISCUSSION

The immunofluorescence results from the 4E-3p experiments show a general localization in the cytoplasm, while the 4E-5p shows co-localization foci/speckles with the mitochondria. This latter observation is consistent with the finding that 4E-5p is part of the mitochondrial proteome and is unprecedented as the 4E proteins in other eukaryotes are localized in the cytoplasm and the nucleus, but not mitochondria. This has profound implication on the evolution of 4Ep function since mitochondria are presumed to be prokaryotic in origin.

Through MS analysis, the co-IP experiments were confirmed to be specific in their pull down of 4E-3p. The identities of the associated proteins suggest 4E-3p association with the translation machinery, particularly the 4F complex, metabolic enzymes (potential feedback regulators of their own translation), and proteins with no known homology. The specific pull down of 4G-4 and 4A1 by 4E-3 Ab coincide with previous experiments in the procyclic forms showing direct binding of 4E-3 with 4G-4 and 4A1 [1], and with other studies in other eukaryotes [73]. These results support the role of 4E-3p in translation and suggest a potential role in the regulation of the expression of metabolic enzymes, as well as association with unique proteins to perform still undetermined function(s).

The pulldown of the 4E-5p in the sample treated with the 4E-5 Ab verified the specificity of the 4E-5 Ab. While there were no exceptional proteins specific to the 4E-5 sample, the shared translational proteins between the 4E-3 pulldown and 4E-5 pulldown support that 4E-5p is associated with translational machinery, however there was no evidence that 4E-5p interacts with the 4F complex.

The RNAseq analysis and the caGEDA comparison indicated the association of multiple mRNAs unique to 4E-3p and 4E-5p. The RNAseq analysis of the 4E-5p depicted a number of putative general gene expression mRNAs, such as the putative DNA polymerase, putative RNA directed DNA polymerase and the putative bifunctional DNA primase/polymerase. The proteins encoded by these mRNAs either assist in or directly perform transcription, leading to a possibility that 4E-5p acts during the S-phase of cytokinesis. Furthermore, the pull down of mRNAs encoding for a kinesin-like protein fragment support this assertion, as the putative kinesin-like proteins are important for cytokinesis [74][75]. There was also a pull down of mRNAs encoding for a putative S-phase kinase-associated protein 1, also a key component for cytokinesis [76]. The presence of a verified bloodstream stage alaninerich protein BARP, which is present in large numbers in the blood form stage, shows a possible association of 4E-5p with the life cycle and pathogenesis of trypanosomes [77]. Taken in concert with the immunofluorescence data and 4E-5p association with the membrane of the mitochondria, there is evidence that 4E-5p acts as a regulator of kinetoplast division during cytokinesis.

In comparison to the 4E-5p pull down in the RNAseq analysis, 4E-3p showed a large number of mRNAs encoding for proteins that assist with gene expression, RNA metabolism, and a host of other general mRNAs that assist in metabolism regulation and other general functions required by the cell. Specifically, the mRNA encoding for 4A-1, a member of the 4F complex, was present. The MS data showed the association of 4E-3p with 4A-1 and 4G-4, forming the possible 4F complex in trypanosomes, while these new data suggest a possible regulatory element possessed by 4E-3p in relation to the production of the 4F complex. There was also a pull down of mRNAs encoding for a putative DNA directed RNA polymerase, which produces the primary transcript of DNA, suggesting that 4E-3p assists in transcription regulation. The general gene expression mRNAs pulled down encoded for various proteins, such as a putative DNA repair protein C and a putative DNA mismatch repair protein, suggesting that 4E-3p assists in regulating DNA maintenance. The presence of a large number of metabolic mRNAs coincides with the MS data to show a possible involvement of 4E-3p with metabolism. Previous experiments have shown that with tetracycline induction of 4E-3 RNAi, the cells become deformed and shriveled, with slower growth, which could mean that 4E-3p also plays a role in metabolic regulation, and without 4E-3p the organism cannot produce proteins required to continue cell respiration.

In the 4E3v4E3+CHX comparison, the 4E-3+CHX sample showed a large presence of mRNAs that encode for proteins that regulate metabolism and cell respiration. Since CHX arrests translation and we assumed that the cells used in this sample were not stressed, then any mRNAs trapped during translation should be from normal cell function. The large presence of metabolic mRNAs further suggested that 4E-3p plays a role in metabolism regulation. Among

the mRNAs pulled down in the sample treated with CHX was a putative RNA helicase, showing further evidence for translation regulation as well, as the RNA helicase interacts with the 4F complex to unwind mRNA for translation. The untreated 4E-3 sample in the 4E3v4E3+CHX comparison exhibits mRNAs decreased in the presence of CHX, however there should be no decrease of cellular functions or translational regulation. As the mRNAs pulled down in this sample awere not outwardly important for cell cycle or cell growth, it can be assumed that these are general mRNAs not being translated at the moment of the CHX addition, and could possibly be transported from the nucleus or to storage vesicles as well.

The comparison 4E3v4E3+Tet is quite important, as this revealed mRNAs decreased with the reduction of 4E-3p levels. The mRNAs increased in expression when the RNAi cells were induced with tetracycline are largely metabolic in nature, with mRNAs encoding for putative RNA-editing complexes and a putative transposon Tn7 transposition protein being of any note in relation to gene expression. In the untreated 4E-3 sample, showing a reduction of mRNA expression in the presence of Tet, there is a reduction of key transcription and translation factors, as well as cell cycle factors. Of note is the reduction of 4E mRNAs, as well as mRNAs encoding for a putative DNA-binding helix-turn-helix, a putative DNA-directed RNA polymerase II subunit, and a putative dedicator of cytokinesis 11-like protein. There was also a reduction of mRNAs encoding for putative metabolic proteins, as well as mRNAs encoding for putative mitochondrial proteins. As 4E-3p associates with the cytoplasm, as shown with the immunofluorescence data, this suggests that 4E-3p is acting as a regulator of gene expression and at least some parts of the cell cycle.

The analysis of common motifs present in the 3’ UTRs of the 4E-5 sample from the 4E5v4E3 comparison show a heavy concentration of uracils. The mRNAs noted before in the 4E-5 sample, those encoding for a verified bloodstream stage alaninerich protein BARP, a putative S-phase kinase-associated protein 1, and the general gene expression proteins, share this motif, suggesting that 4E-5p does act in cell cycle regulation. Combined with the immunofluorescence data, showing an association with the mitochondrion of trypanosomes, there is strong evidence that 4E-5p assists in the kinetoplastic replication during the S-phase of trypanosoma cell growth.

The analysis of common motifs present in the 3’ UTR of mRNAs targeted by 4E-3p shows a common trend amongst all of the sample comparisons with 4E-3p. The evidence points to 4E-3p preferentially targeting mRNAs with uracil-guanine sequences in the 3’ UTR, shown in the sample comparisons 4E5v4E3 and 4E3v4E3+Tet. Again, it is assumed that the addition of CHX to the 4E-3 antibody pull down does not change the specificity of 4E-3p, but just captures the translational machinery at a different time point than seen in the sample without CHX. The addition of tetracycline causes a reduction in 4E-3p levels, also causing a reduction in the mRNAs 4E-3p preferably targets, as supported by the untreated 4E-3 pulldown in comparison to the tetracycline induced sample. The 4E3v4E3+Tet comparison shows a common motif rich in uracils and guanines in the 4E-3 +Tet Decrease Sample, representing those mRNAs reduced in the presence of tetracycline, however the mRNAs overexpressed in the tetracycline induced sample had no common motif due to the 3' UTRs containing only one nucleotide on average. This shows the possibility that although 4E-3p regulates general gene expression and metabolism mRNAs exhibiting a common motif, while also interacting with general mRNAs. The sample

comparison 4E3v4E3+CHX also has a large presence of uracils in the common motifs, with small amounts of other nucleotides.

All of the sample specific mRNA 3' UTR secondary structures show a significant tendency of the shared motif to exist on bulges and loops. This supports the regulatory role of 4E-5p and 4E-3p through direct or indirect interaction with the 3' UTR of mRNAs, with the motifs differing between the 4E-3 and 4E-5 samples. However, the 3' UTR secondary structures are unstable as a whole, requiring further experiments to be performed to verify their importance for 4E-3p and 4E-5p interaction.

While the immunofluorescence data show an association of 4E-5p with the mitochondrion, the RNAseq data do not show any association with mitochondrial mRNAs, nor do the MS data show any association with mitochondrial proteins. However, the data instead suggests that 4E-5p acts as a cell cycle regulator and assists in either replication or transcription. Since there are no mRNAs associated with 4E-5p that directly encode for parts of glycolysis or the mitochondrial proteome, we propose that 4E-5p plays a role in kinetoplast replication during the S-phase of trypanosome division, which would be unprecedented for a 4E homolog associating with an organelle of possible prokaryotic origin [37].

The immunofluorescence data support the 4E-3p translational role in the cytoplasm, with the MS data further supporting 4E-3p being a bona fide unit in the TbeIF4F complex with the association of 4A-1 and 4G-4. The RNAseq data provides evidence for 4E-3p's role in regulating the same 4F complex that this homolog belongs to, while also regulating transcription

and parts of the metabolic cycle. Since, 4E proteins in other eukaryotes act to transport mRNAs from the nucleus to the translation complex, or from the nucleus to storage vacuoles for later translation or degradation, it can be deduced that 4E-3p acts to recognize and transport specific mRNAs from the nucleus to the translation complex as well [20][54][78].

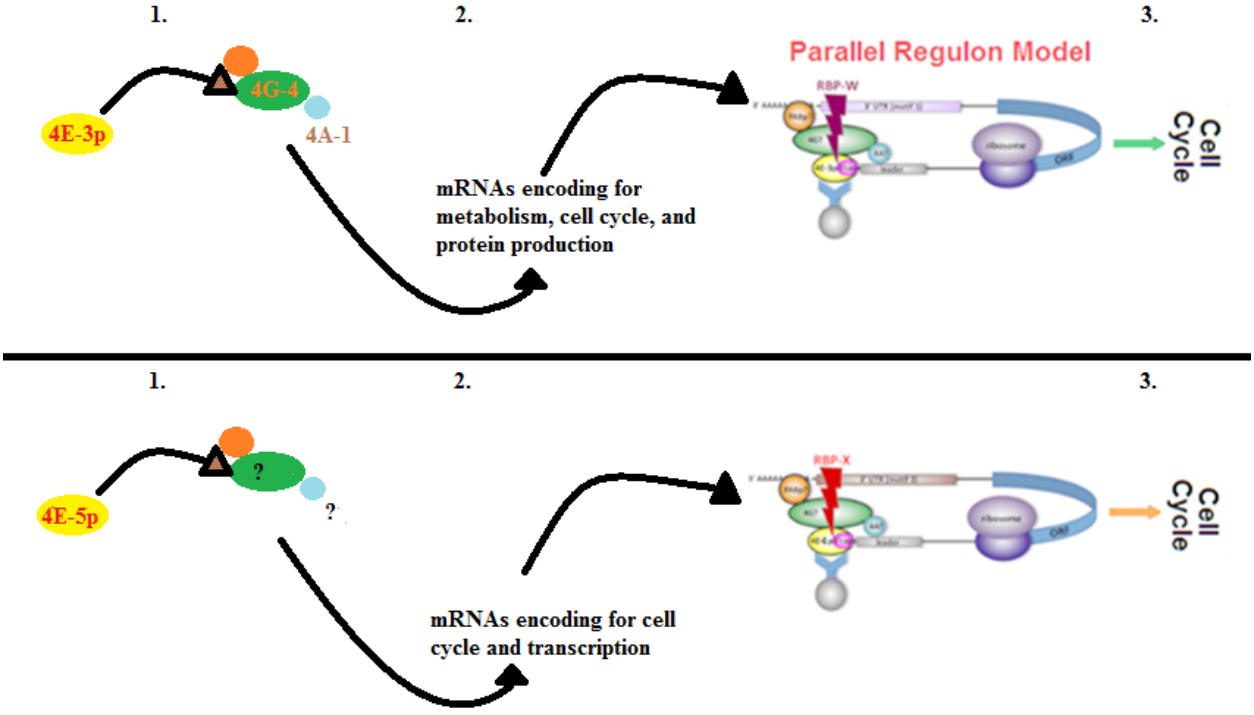


Figure 26. Parallel Regulon Model.

1. The mass spectrometry data showed association of 4E-3p with bona fide parts of the eIF4F (4F) complex, which are required for translation initiation. Although 4E-5p showed no association with a 4F complex, ribosomes responsible for translation were present, supporting the previous evidence that 4E-5p is a homolog of eIF4E. 2. The RNAseq and caGEDA analysis showed unique mRNAs overrepresented in both the 4E-3 and 4E-5 antibody pulldowns. The overrepresented mRNAs in 4E-3 encoded for putative and verified cell cycle regulating proteins, proteins that regulate metabolism and cell function, and proteins that regulate transcription and translation. While the overrepresented mRNAs in 4E-r encoded for proteins of similar function, there were no shared pathways. 3. The findings show the possibility that 4E-3p and 4E-5p act on distinct regulons, or groups of mRNAs that share similar function, in order to regulate cell function and cell cycle.

Overall, the motif data suggest that 4E-3p and 4E-5p interact with unique yet functionally related mRNAs (**Figure 26)**. The immunofluorescence confirms the association of 4E-3p with the cytoplasm, with the MS data supporting 4E-3p associating with translational machinery. The RNAseq data show 4E-3p association with a wide variety of mRNAs that encode for proteins that assist with a host of cell functions, with Trawler showing 4E-3p regulating mRNAs sharing common motifs in their 3' UTR and possibly belonging to a regulon, a regulon being a functionally related group of mRNA regulated by a protein or proteins. These mRNAs sharing the motifs encode for possible transcriptional proteins, translational proteins, and proteins that assist with metabolism and cell respiration. This would mean 4E-3p works to not only control specific gene expression functions, but to also control general cell cycle and survival. The immunofluorescence confirms 4E-5p association with the mitochondrion. The MS data reveal no specific interaction of 4E-5p to the eIF4F complex, however the presence of ribosomes that assist in translation show the possibility of 4E-5p acting to regulate protein function while being associated with the mitochondrion, which would be unprecedented. The RNAseq data show the possible association of 4E-5p with mRNAs that encode for cell cycle regulators and transcriptional proteins. The Trawler analysis of these mRNAs show a common motif shared between these mRNAs, supporting the possibility that 4E-5p also acts upon its own regulons. This supports our Parallel Regulon Theory, with 4E-3p and 4E-5p interacting independently with functionally related mRNA that belong to separate pathways to regulate cell cycle and gene expression.

These data could be used to provide a foundation for the understanding of *T. brucei's* 4E-3p and 4E-5p, and act as a launching point for further studies in to the specific interactions of

these proteins with translation machinery and mRNA. The evidence of 4E-5p interacting with the mitochondrion is unprecedented, and should be further analyzed in other early protozoans, as well as the members of the TriTryps, to determine consistency. The association of 4E-3p with the cytoplasm and translation machinery confirms previous studies pertaining to 4E-3p function and localization. Effort can now be focused on confirming the identities of the 4E-3p protein interactors and associated mRNAs, as well as their specific function in cell cycle.

6.0 FUTURE DIRECTIONS

The findings presented here represent only the tip of the iceberg. While great effort was used to perform *in silico* analyses to make up for the lack of experimental evidence to confidently propose functions for 4E-3p and 4E-5p, further experimental evidence is required to extend the findings presented here.

Further co-immunoprecipitation pulldowns with associated mass spectrometry of 4G-4 and 4A-1 would provide solid evidence of 4E-3p's association with the trypanosomal 4F complex, and to functionally validate anticipated 4E-3 and 4E-5 mRNA targets via qPCR of polysome-derived mRNAs with and without 4E-3 RNAi to determine their loss of translation recruitment during RNAi. The association of 4E-5p with replication and cell cycle machinery, while also being directly associated to the mitochondria are unprecedented, and would require further evidence to fully support the evidence provided by us in these studies.

Mass spectrometry focusing on 4E-5p and the associated complexes would assist in verifying the role of 4E-5p in translation, and confirm association with possible eIF4F subunits. Running the RNAseq experiments with comparisons between a 4E-5p pull down against a tetracycline induced 4E-5 RNAi pull down would show which mRNAs 4E-5 specifically targets, with a comparison with 4E-5 pull downs treated with cycloheximide to arrest translation would

show which mRNAs are being translated at that exact moment and whether 4E-5p acts to transport mRNAs. While previous studies have shown that the protein we identified to be 4E-5 is part of the mitochondrial proteome, there was little evidence of mRNAs that target the mitochondria. However, the large amount of mRNAs encoding for hypothetical unknowns could provide polypeptides unique to trypanosomes, and may show a degree of importance.

The motif data showed conserved motifs in the 3' UTRs of mRNA overrepresented in both 4E-3p and 4E-5p. An *in silico* meta-analysis of the entire *T. brucei* genome could provide information on how prevalent those motifs are. While they maintained high z scores, there was no way to properly judge their prevalence. Provided alongside these motifs were possible identities. These identities are specific for only DNA motifs. Since *T. brucei* is polycistronic, an effort should be made to determine whether these motifs in the mRNA are also present on the chromosomes, and whether they assist in proper cleavage of the pre-mRNA. This could be performed using Trawler on individual genes in the genome, while also knocking out, through point mutations, the motif sequences in the chromosomes to determine whether the presence of the transcribed mRNA is decreased. The importance of the secondary structure to 4E-3p or 4E-5p interaction with the 3' UTR of mRNA would require point mutations or deletions where the motif is located, and then analyzed for a decreased interaction or a decreased production of the encoded protein.

APPENDIX: SUPLEMENTAL FIGURES AND TABLES

The motif results provided by Trawler included possible motif identities with their divergence, or percent relevance. However, these identities are for motifs found on the DNA, and would not be critical to mRNA selection or function. They were included here for informational purposes, and could possibly hint at the importance of the motif during the cleaving of the polycistronic pre-mRNA (**Tables S1-S5)**. The rest of the figures show enlarged versions of the mRNA 3' UTR secondary structures generated by the RNAfold webserver, through the Vienne RNA Websuite, to allow for a better understanding and analysis of motif location and overall structure (**Figures S1-S37**).

Table S1. Motif name, divergence, and function for each 4E-5 specific family, organized by divergence.

Motif Name	Divergence	Motif Function
MA0054	0.89	Unknown
Nuclear Receptor	0.77	Up/down regulation of the gene through interaction with nuclear proteins
Bzip Cebp	0.729	CEBP's bind to this motif and regulate/promote expression
ETS Class	0.709	Cell differentiation, cell cycle control, cell migration, cell proliferation, apoptosis.
MAD Box	0.625	Binds MADS transcription factors. Common to Eukaryotes. Cell proliferation.
MA0039 (Gklf)	0.544	Unknown
MA0033	0.506	Unknown
Homeo Box	0.476	*Encodes for transcription factors that typically turn on other genes.
TRP MYB	0.341	Secondary metabolism, cell shape/fate
Forkhead	0.299	Key for embryogenesis? Could help with metabolism and cell proliferation
HMG	0.264	HMG proteins bind to affect the formation and function of transcription complexes

The first column lists the possible identity of the motif discovered by Trawler. The second column shows the divergence, or percent relevance, of these identities. The third column shows the possible function of the motif identity, discovered via literature searches.

Table S2. Motif name, divergence, and function for each 4E-3 specific family, organized by divergence.

Motif Name	Divergence	Motif Function
Fam 1 - Sox-5	0.808	Embryonic function and determination of cell fate
Fam 2 – MA0095	0.661	Unknown
Fam 2 – MA0094	0.585	Unknown
Fam 3 – bZIP CREB	0.898	CEBP's bind to this motif and regulate/promote expression
Fam 3 – c-FOS	0.862	Nuclear phosphoprotein - converts extracellular signals into changes of gene expression
Fam 3 – MA0086	0.637	Unknown
Fam 6 – Ubx	0.571	Ubx protein regulates alternative splicing
Fam 6 – S8	0.415	Binds to the S section of 30S tibosomal subunit

The first column lists the possible identity of the motif discovered by Trawler. The second column shows the divergence, or percent relevance, of these identities. The third column shows the possible function of the motif identity, discovered via literature searches.

Table S3. Motif name, divergence, and function for each 4E-3+CHX Increase specific family, organized by divergence.

Motif Name	Divergence	Motif Function
Irf-2	0.881	*Interferon Regulatory Factor 2
bZIP CEBP	0.769	CEBP's bind to this motif and regulate/promote expression
TRP MYB	0.727	Secondary metabolism, cell shape/fate
Forkhead	0.612	Key for embryogenesis. Could help with metabolism and cell proliferation
Homeo box	0.578	*Encodes for transcription factors that typically turn on other genes.
MA0033	0.574	Unknown
HMG	0.546	HMG proteins bind to affect formation and function of transcrip. complexes
Ubx	0.483	Ubx protein regulates alternative splicing

The first column lists the possible identity of the motif discovered by Trawler. The second column shows the divergence, or percent relevance, of these identities. The third column shows the possible function of the motif identity, discovered via literature searches.

Table S4. Motif name, divergence, and function for each 4E-3-CHX Increase specific family, organized by divergence.

Motif Name	Divergence	Motif Function
MA0044	0.875	Unknown
EN-1	0.847	Neuron development
REL Class	0.8	Regulates transcription factors
bZIP CEBP	0.799	CEBP's bind to this motif and regulate/promote expression
HFH-2	0.753	Key for formation of a neural crest in embryo
Homeo Box	0.73	Encodes for transcription factors that typically turn on other genes.
SRY	0.666	Initiation of male organ formations
ETS Class	0.657	Cell differentiation, cell cycle control, cell migration, cell proliferation, apoptosis.
Androgen	0.607	Binds with male hormones
MA0033	0.607	Unknown
Gklf	0.521	Unknown
MAD Box	0.518	Binds MADS transcription factors. Common to Eukaryotes. Cell proliferation.
cEBP	0.428	CEBP's bind to this motif and regulate/promote expression
TRP MYB	0.337	Secondary metabolism, cell shape/fate
Forkhead	0.334	Key for embryogenesis. Could help with metabolism and cell proliferation
HMG	0.287	HMG proteins bind to affect the formation and function of transcription complexes

The first column lists the possible identity of the motif discovered by Trawler. The second column shows the divergence, or percent relevance, of these identities. The third column shows the possible function of the motif identity, discovered via literature searches.

Table S5. Motif name, divergence, and function for each 4E-3-Tet Increase specific family, organized by divergence.

Motif Name	Divergence	Motif Function
MAD Box	0.858	Binds MADS transcription factors. Common to Eukaryotes. Cell proliferation.
bHLH	0.821	Allows dimerization, DNA binding, and binds transcription factors
Gklf	0.779	Unknown
MA0068	0.732	Unknown
Forkhead	0.599	Key for embryogenesis. Could help with metabolism and cell proliferation
TRP MYB	0.537	Secondary metabolism, cell shape/fate
HMG	0.518	HMG proteins bind to affect formation and function of transcription complexes

The first column lists the possible identity of the motif discovered by Trawler. The second column shows the divergence, or percent relevance, of these identities. The third column shows the possible function of the motif identity, discovered via literature searches.

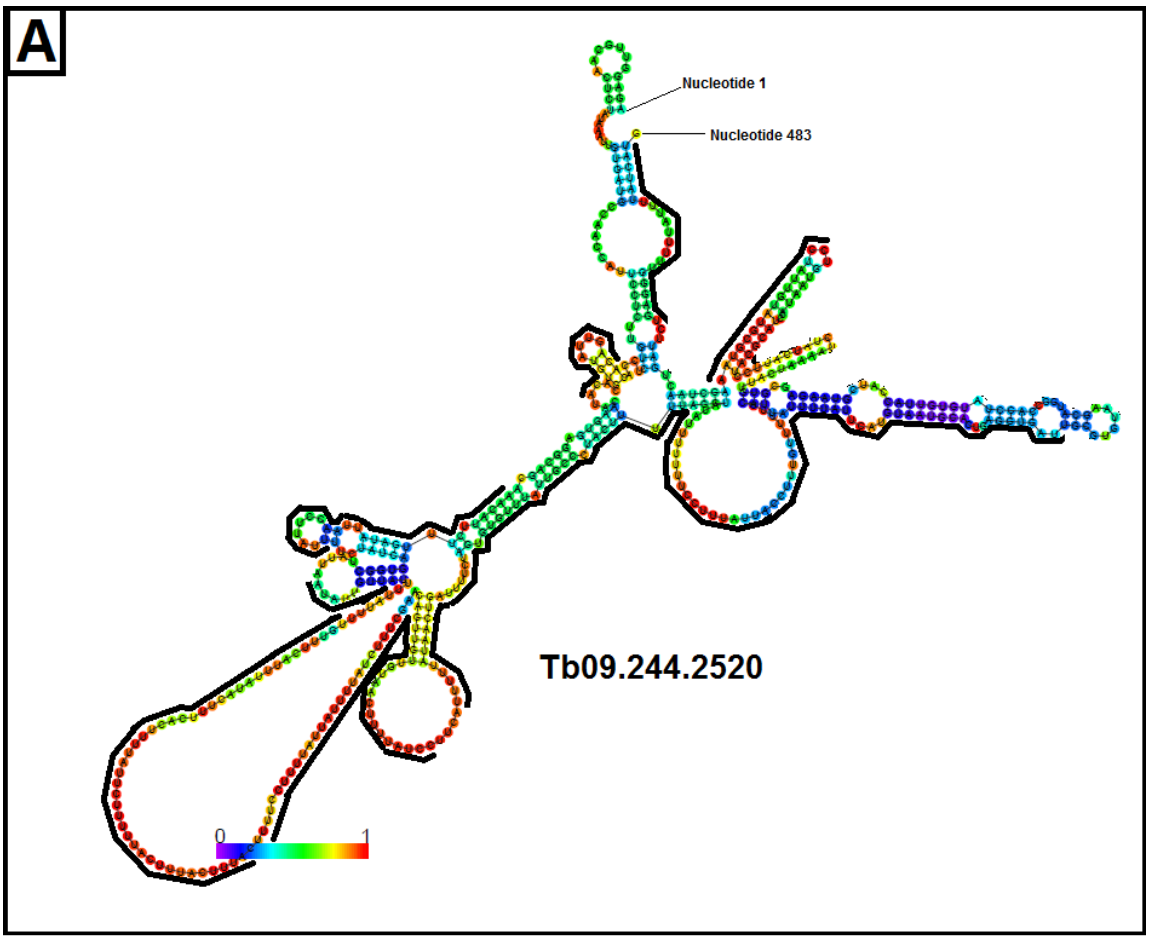


Figure S1. Enlarged version of Figure 21A.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

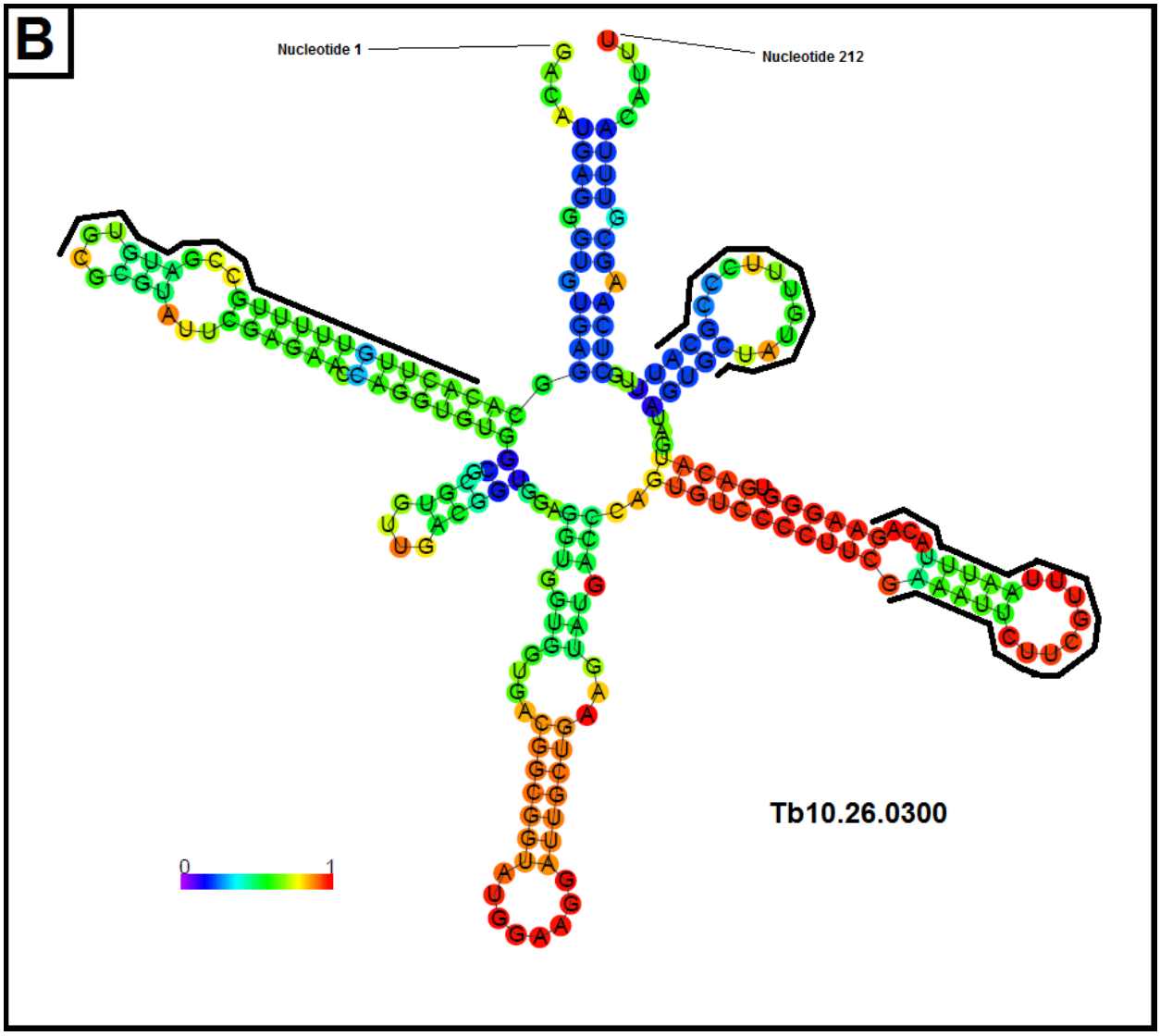


Figure S2. Enlarged version of Figure 21B.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

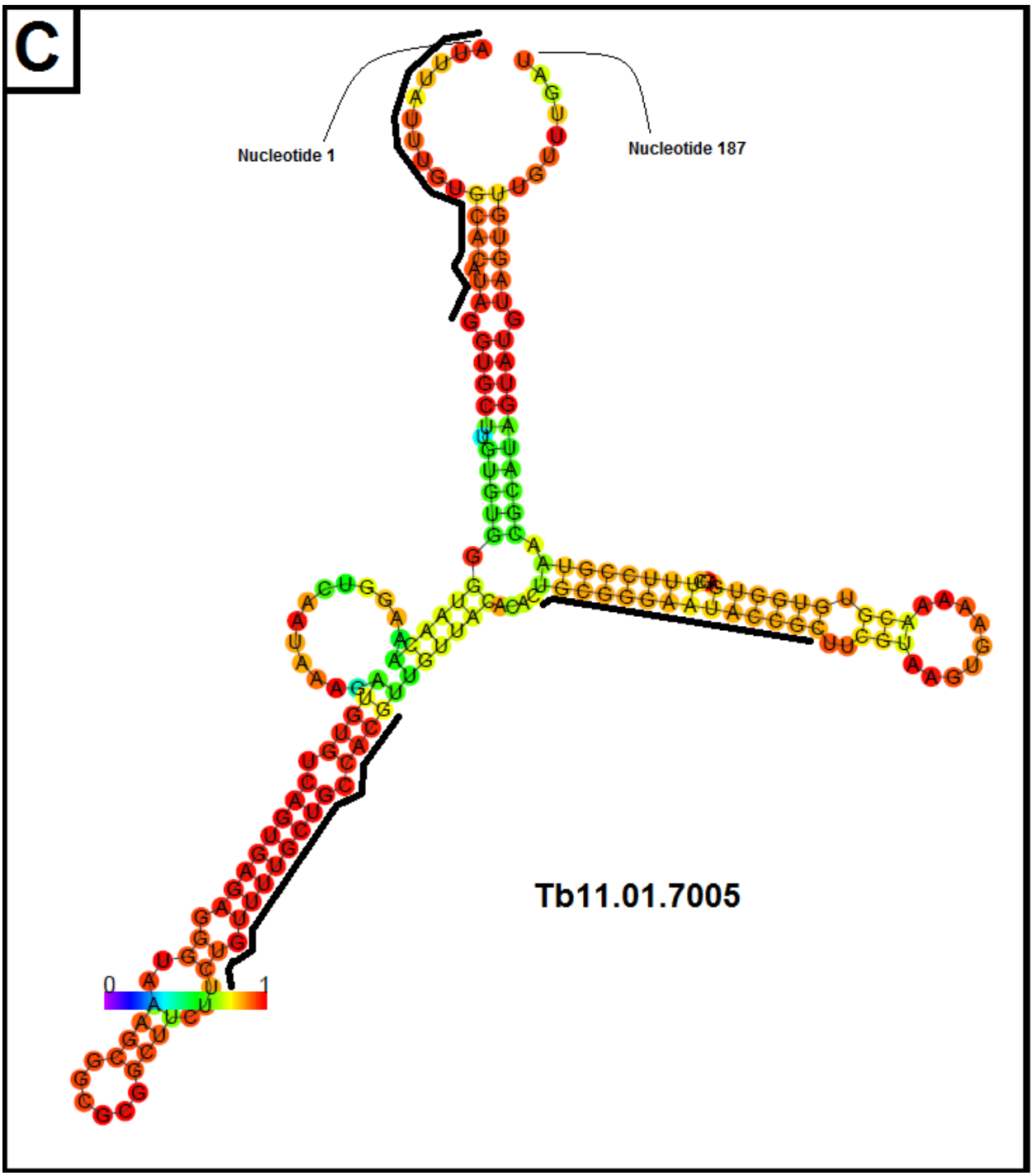


Figure S3. Enlarged version of Figure 21C.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

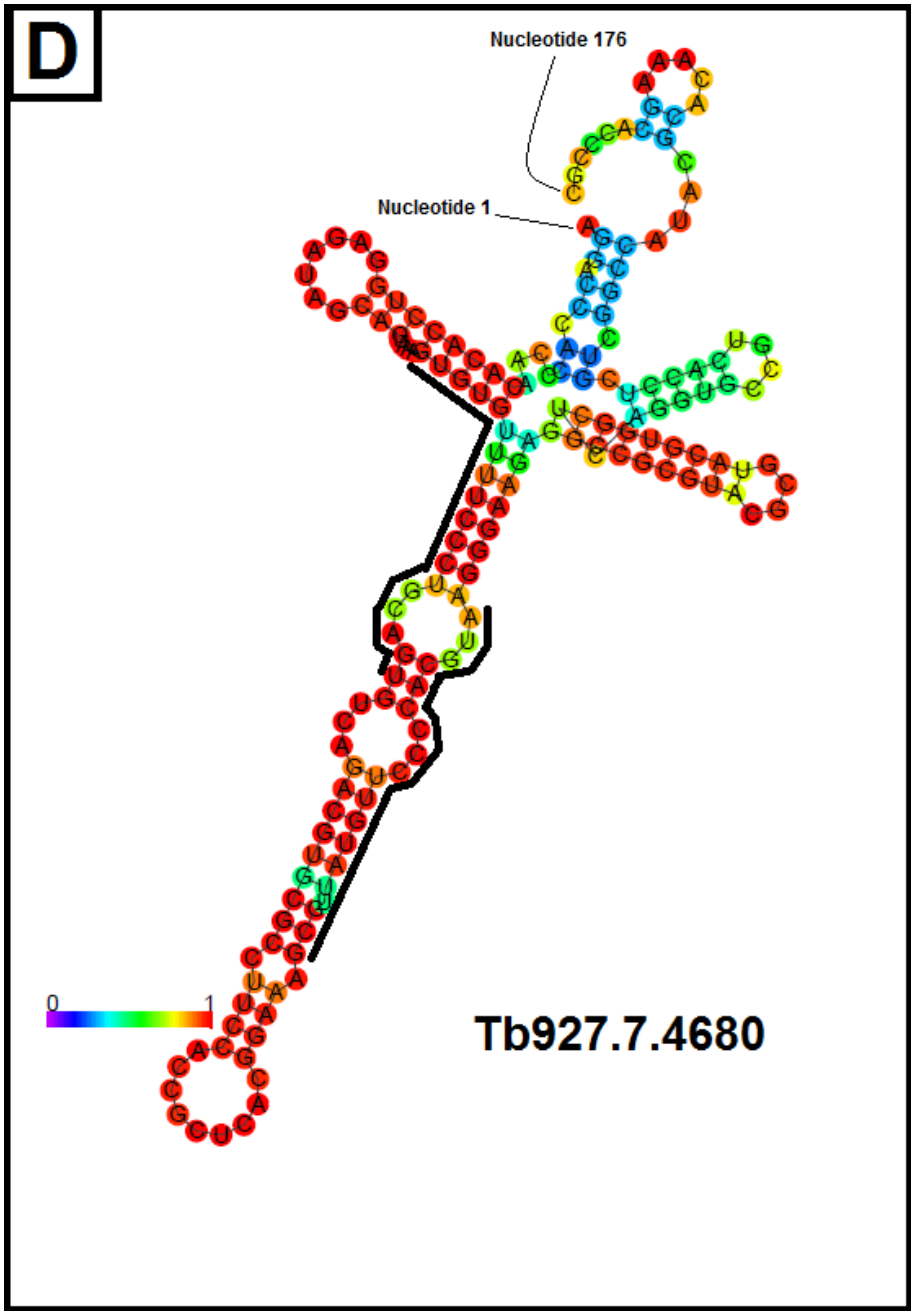


Figure S4. Enlarged version of Figure 21D.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

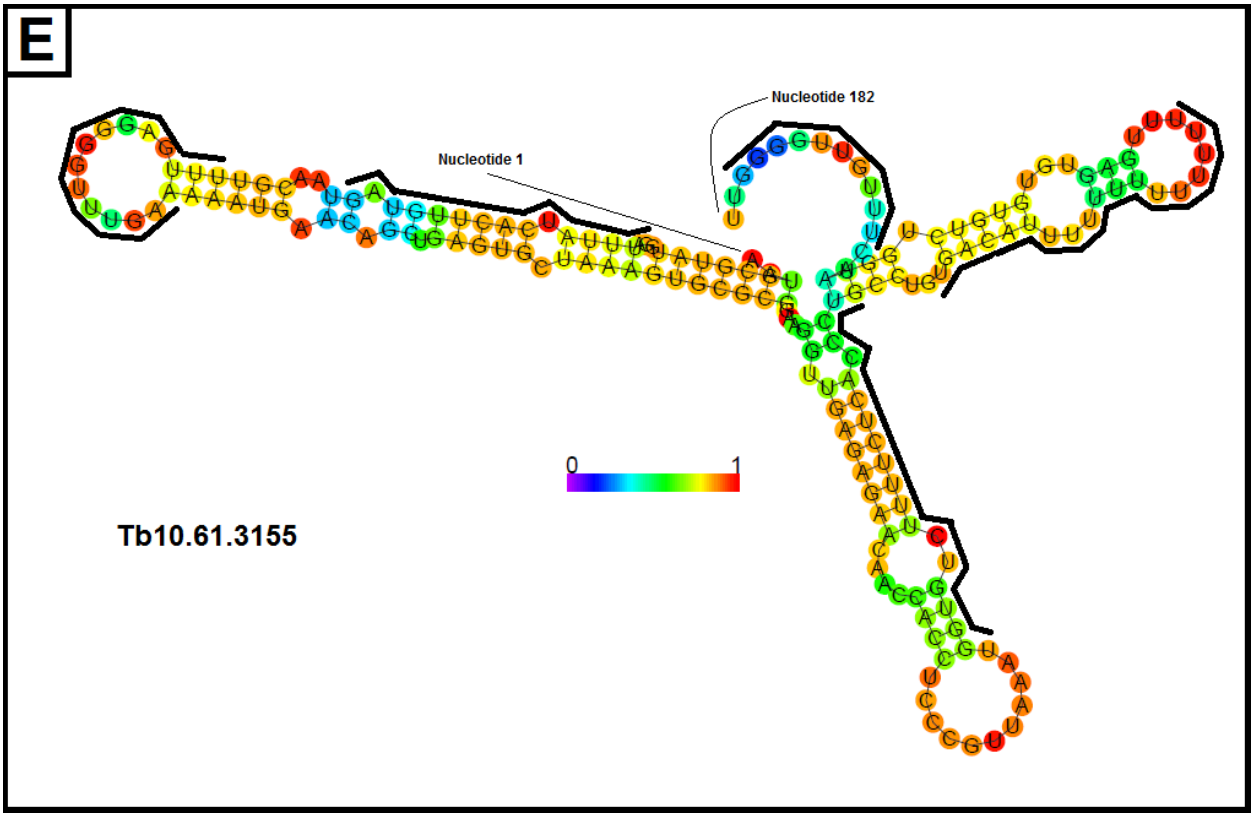


Figure S5. Enlarged version of Figure 21E.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

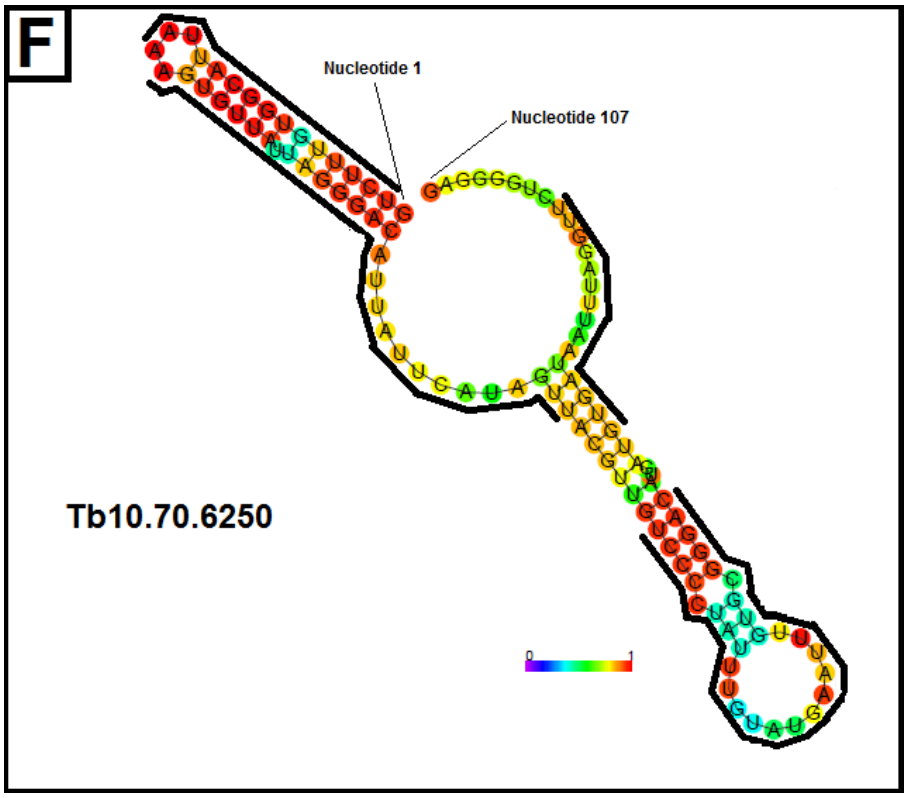


Figure S6. Enlarged version of Figure 21F.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

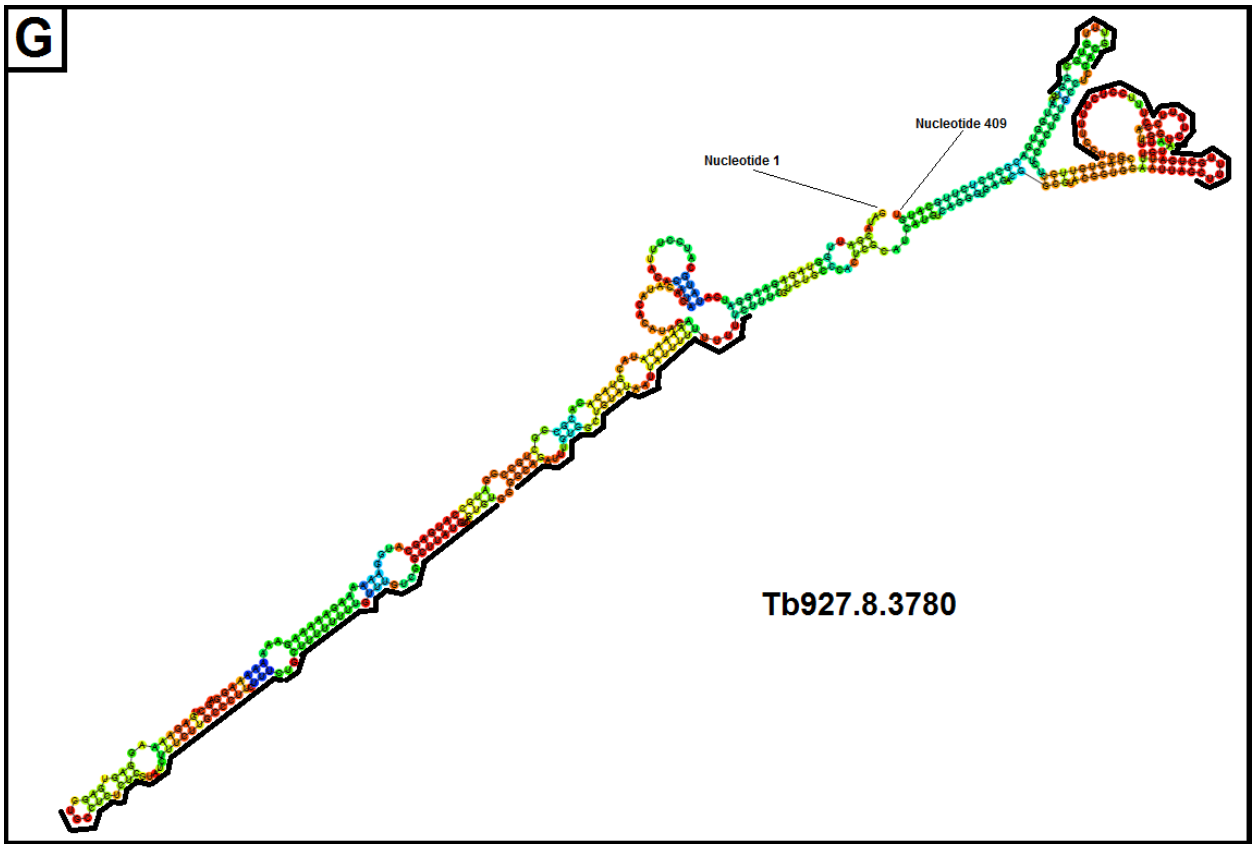


Figure S7. Enlarged version of Figure 21G.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

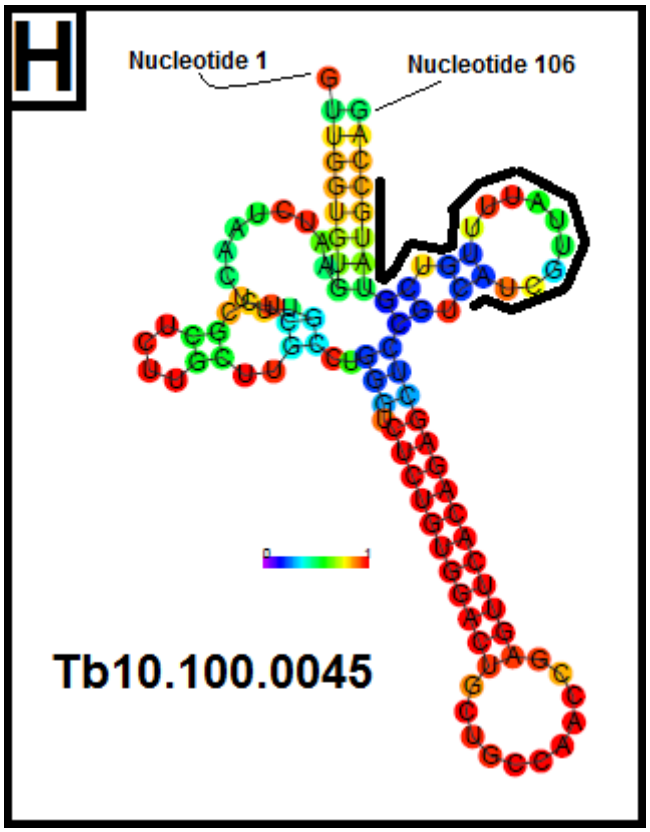


Figure S8. Enlarged version of Figure 21H.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

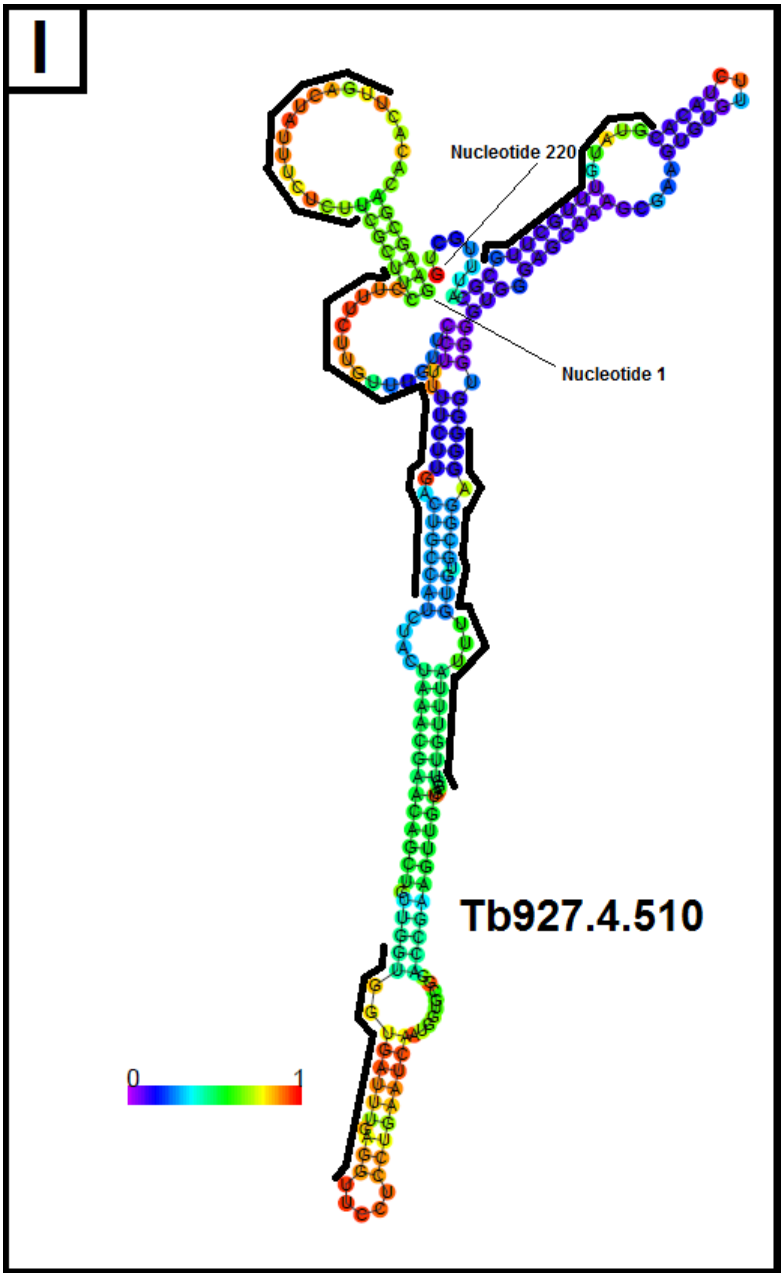


Figure S9. Enlarged version of Figure 21I.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

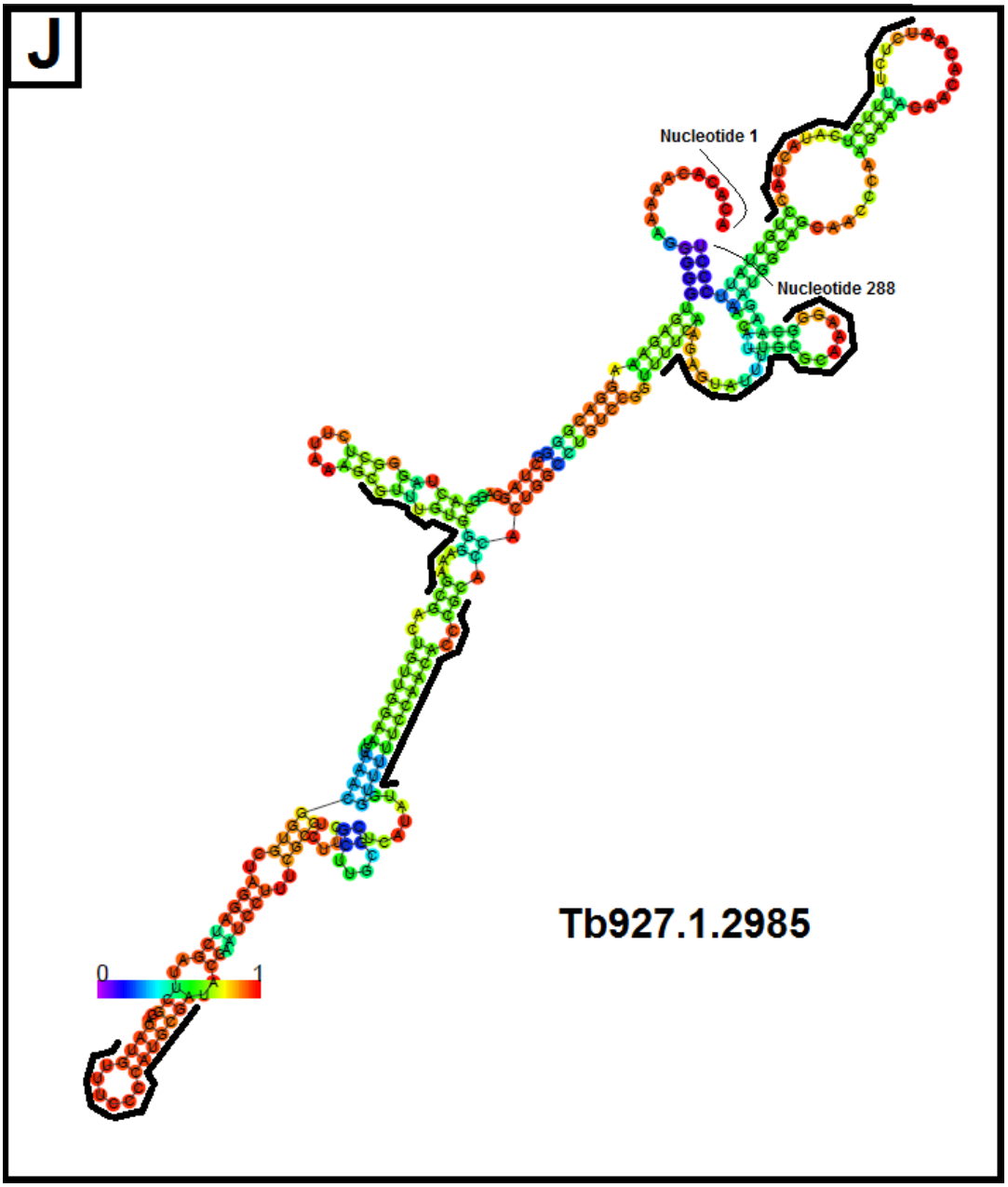


Figure S10. Enlarged version of Figure 21J.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

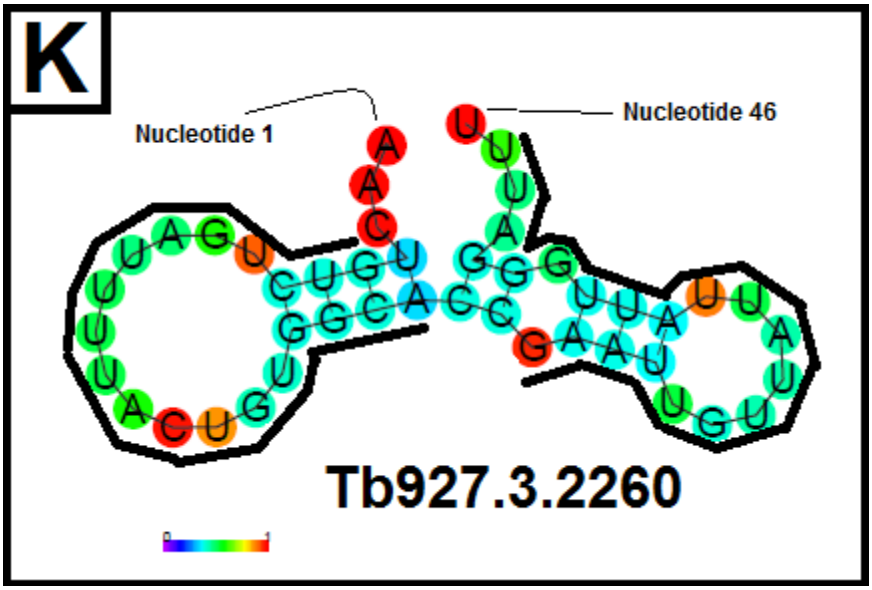


Figure S11. Enlarged version of Figure 21K.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

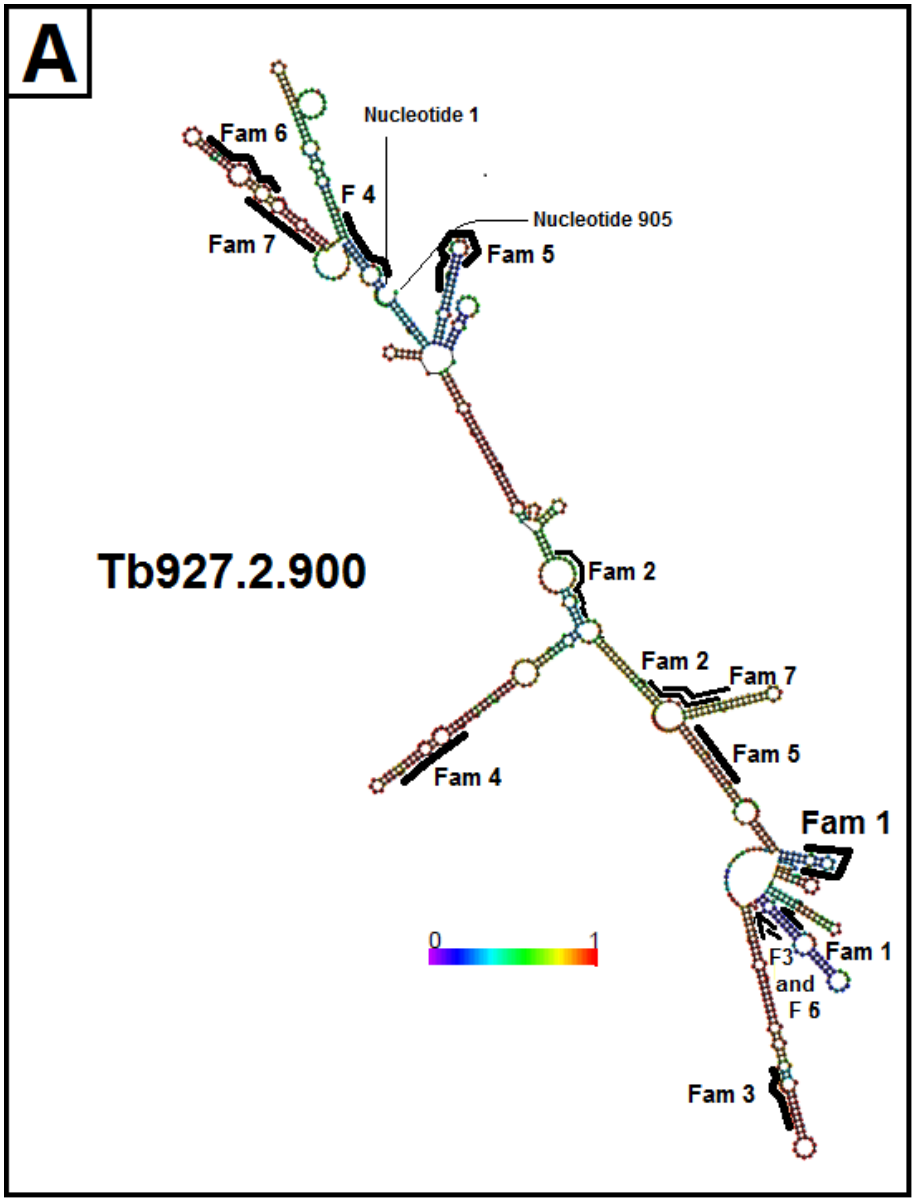


Figure S12. Enlarged version of Figure 22A.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

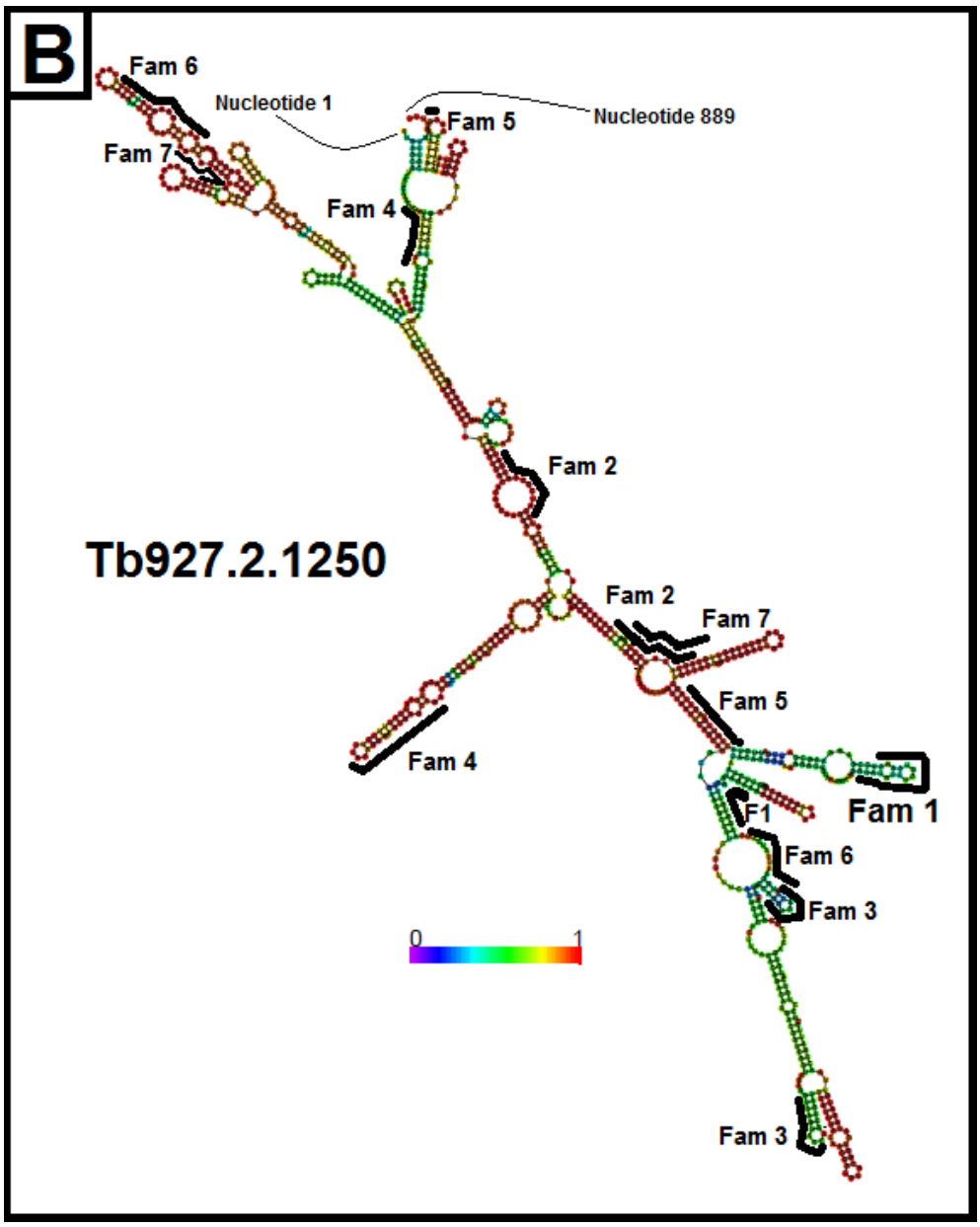


Figure S13. Enlarged version of Figure 22B.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

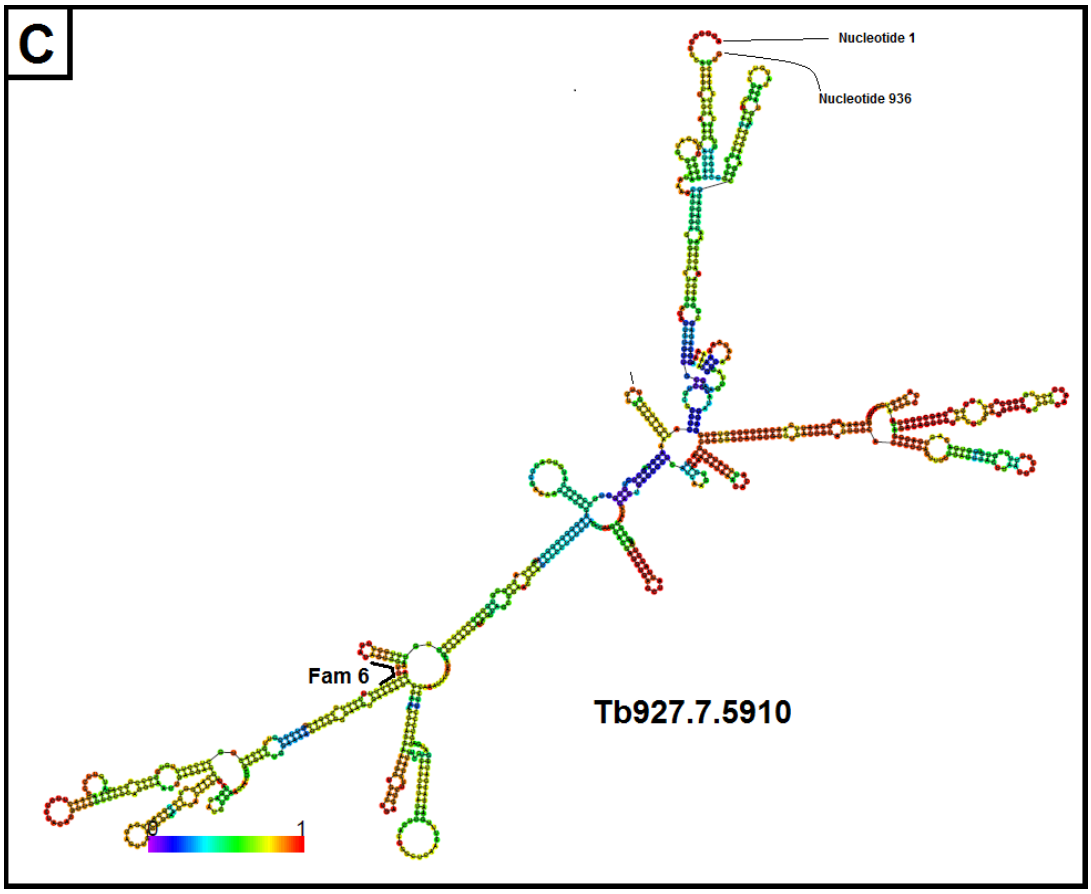


Figure S14. Enlarged version of Figure 22C.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

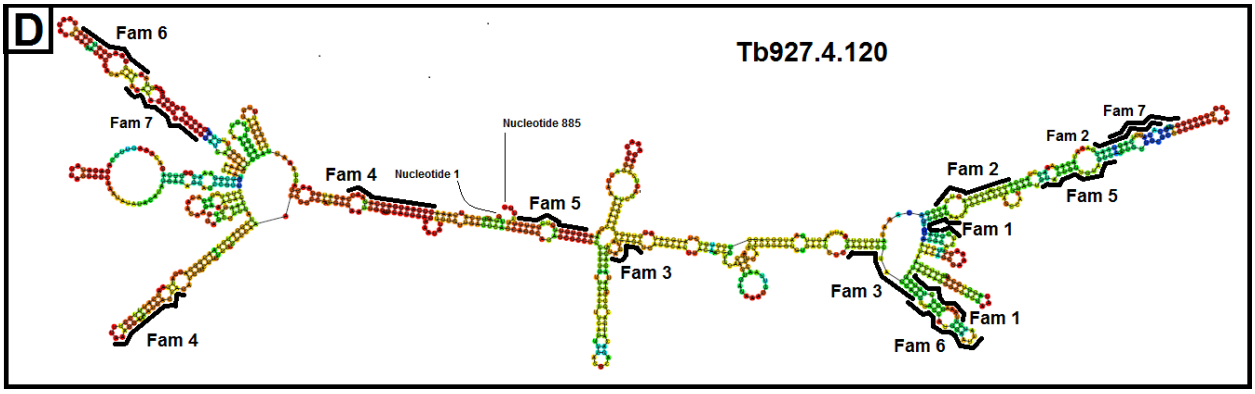


Figure S15. Enlarged version of Figure 22D.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

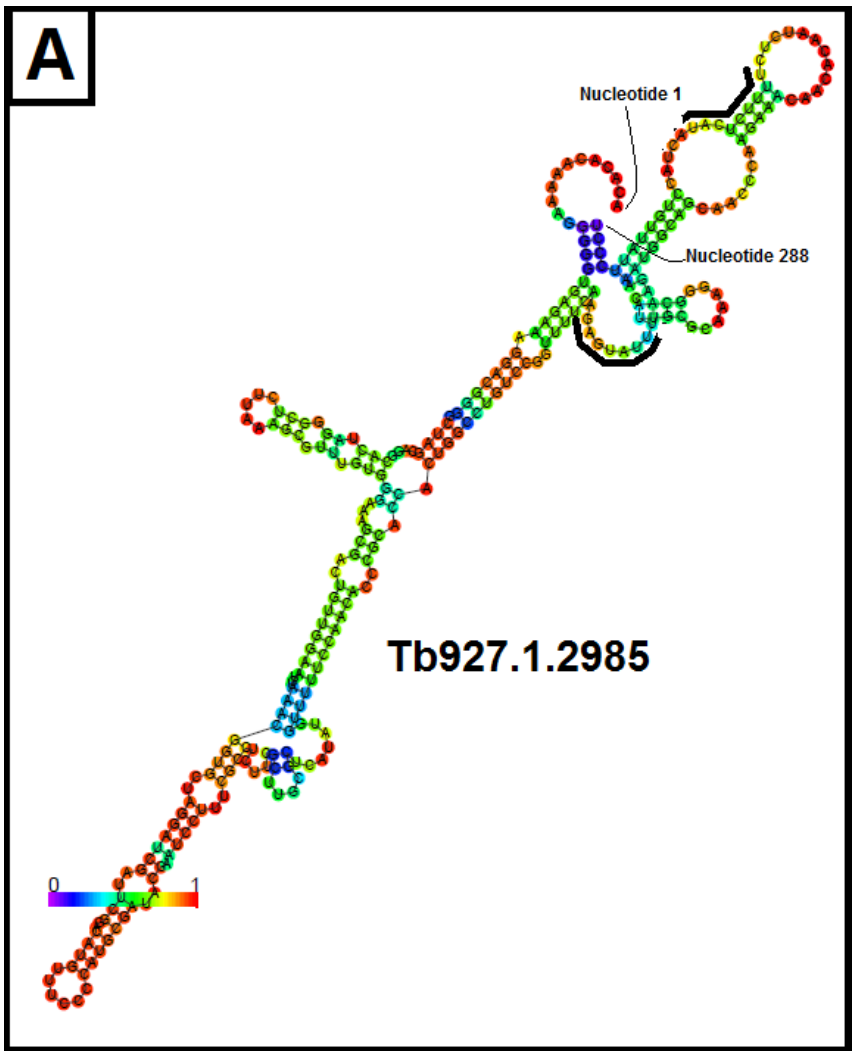


Figure S16. Enlarged version of Figure 23A.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

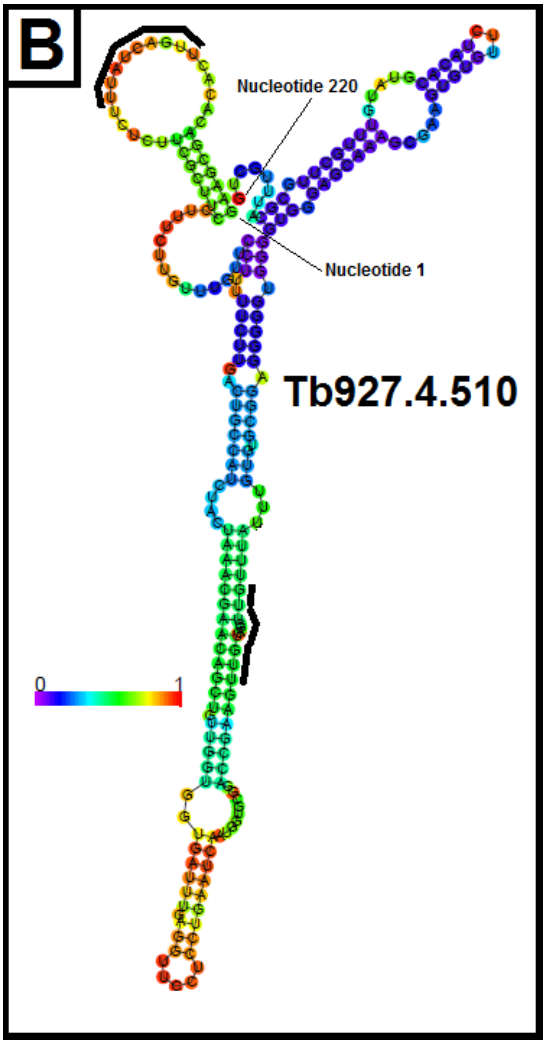


Figure S17. Enlarged version of Figure 23B.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

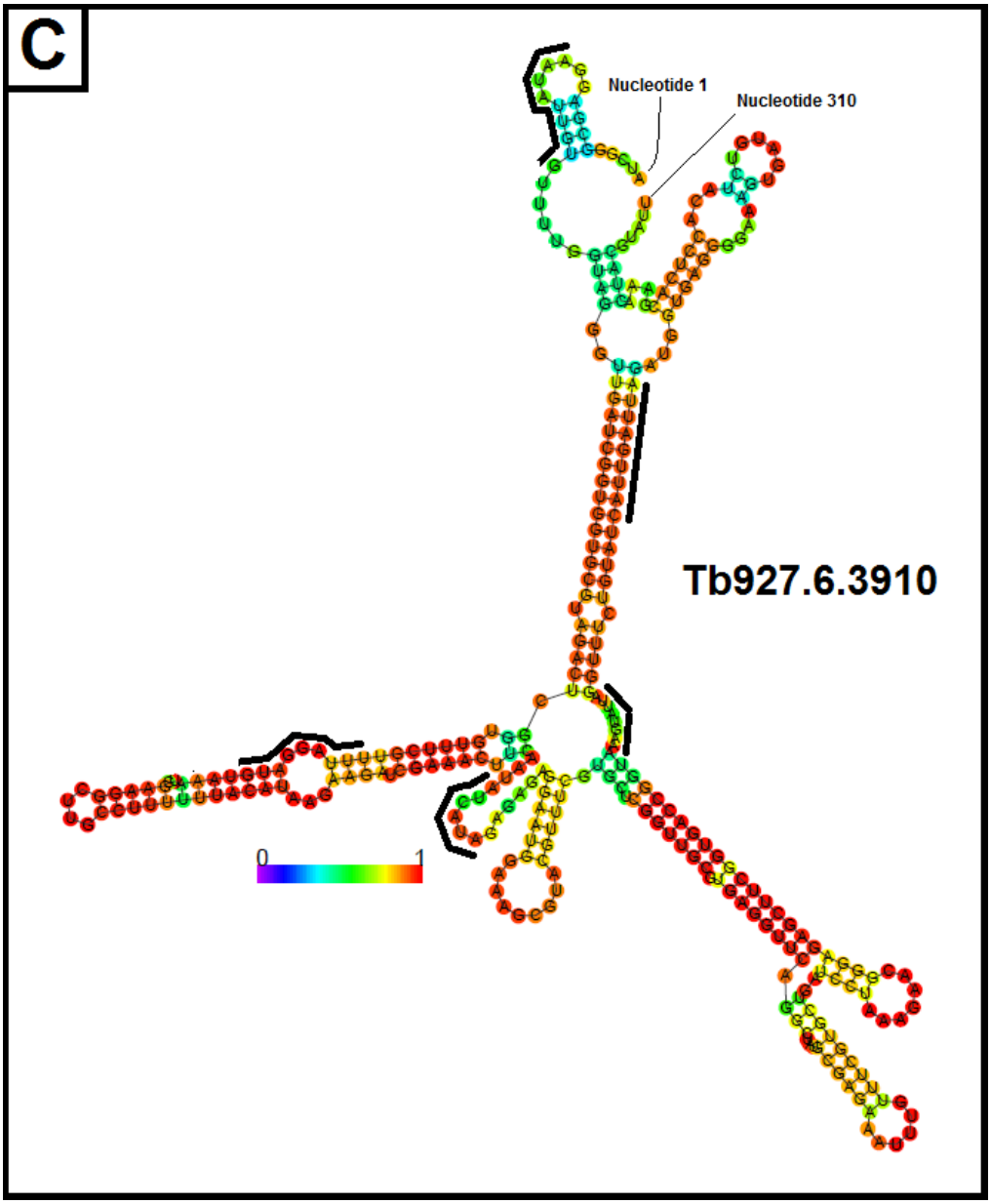


Figure 18. Enlarged version of Figure 23C.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

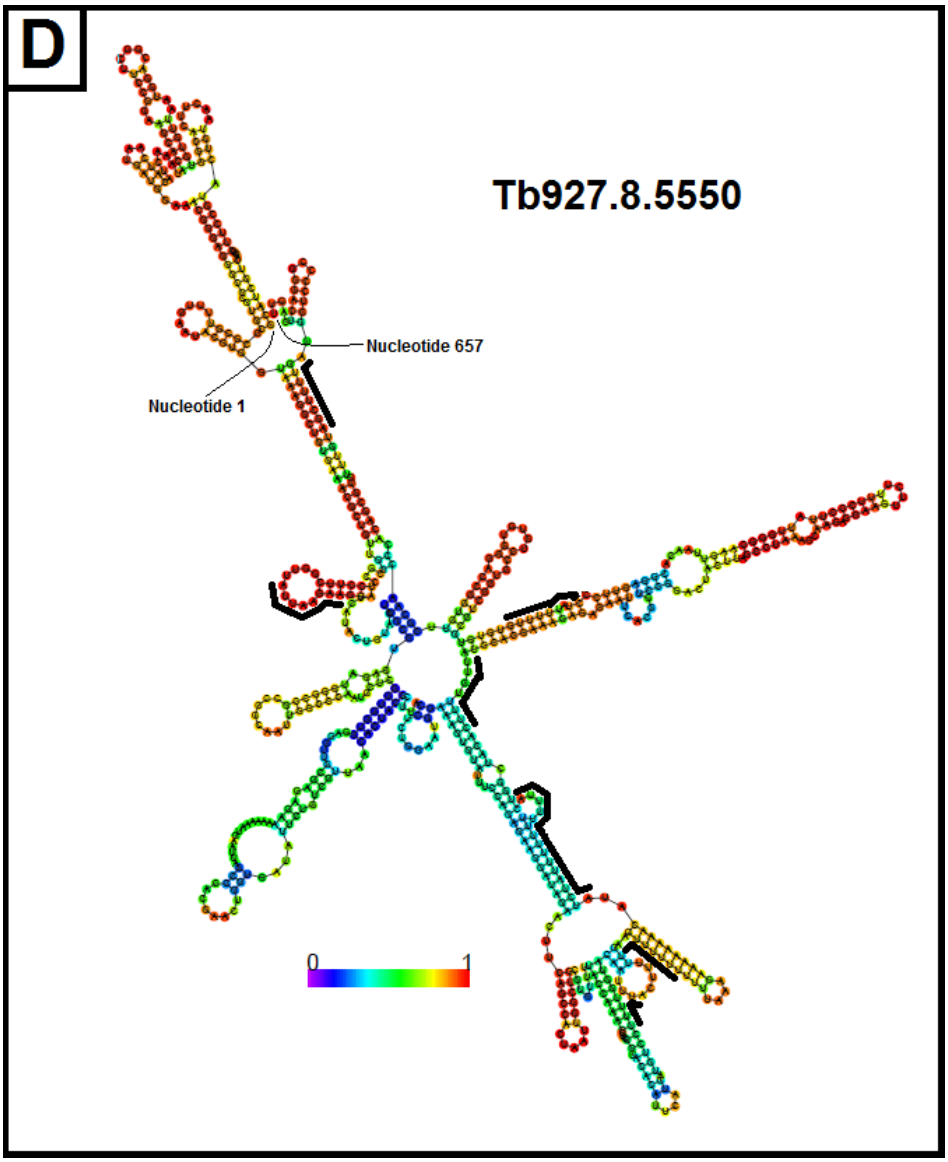


Figure S19. Enlarged version of Figure 23D.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

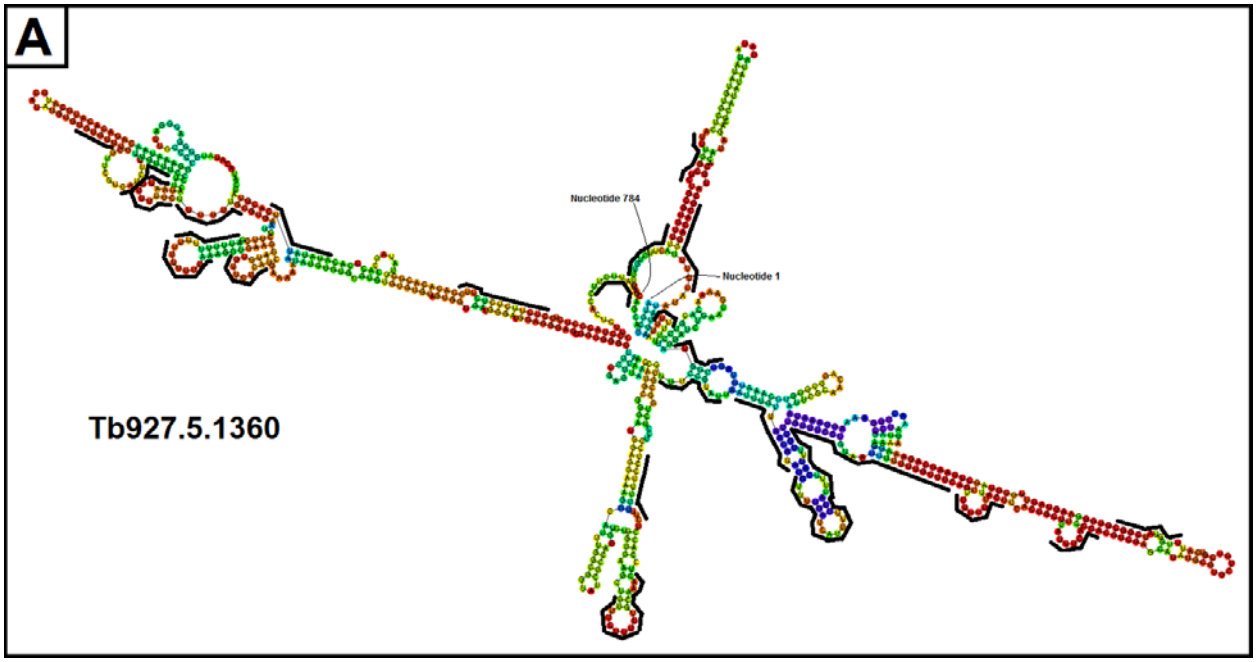


Figure S20. Enlarged version of Figure 24A.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

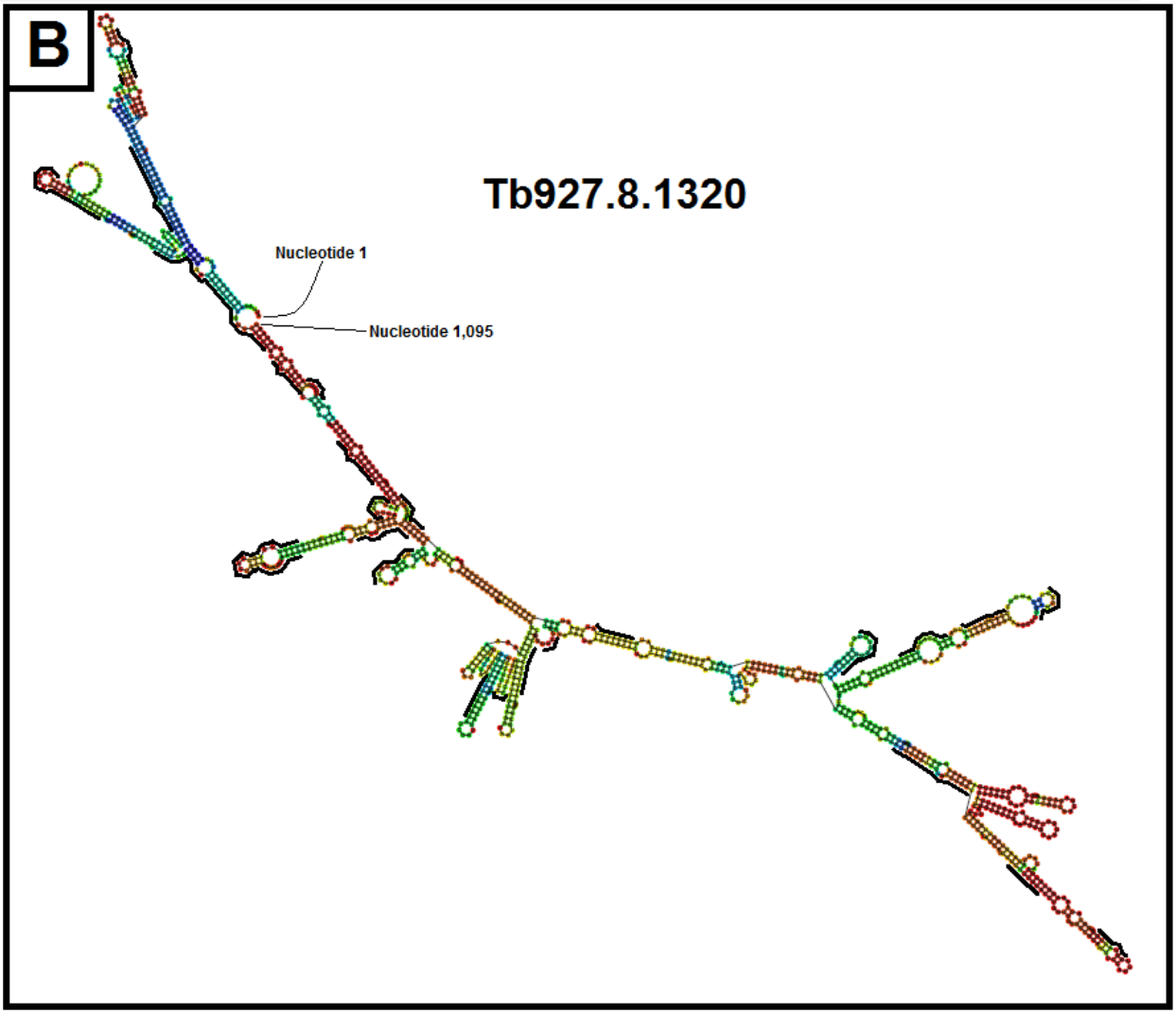


Figure S21. Enlarged version of Figure 24B.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

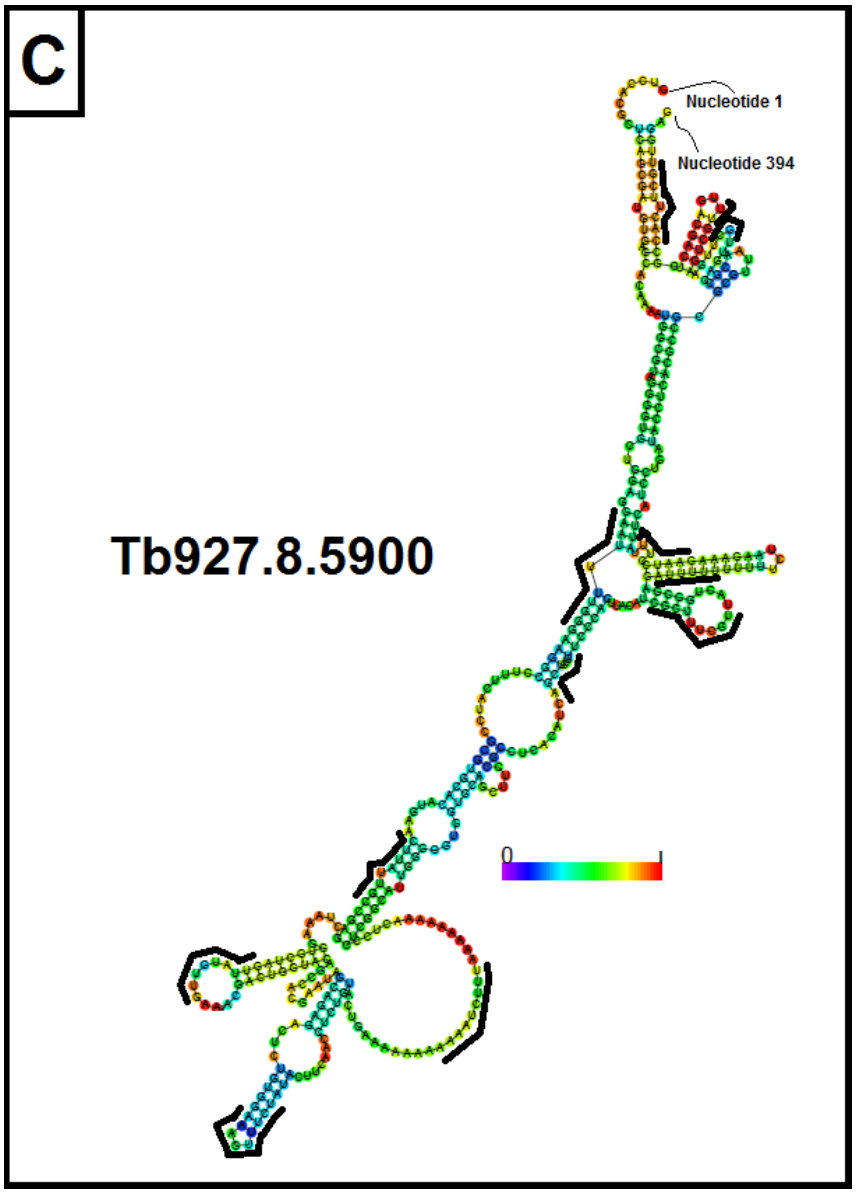


Figure S22. Enlarged version of Figure 24C.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

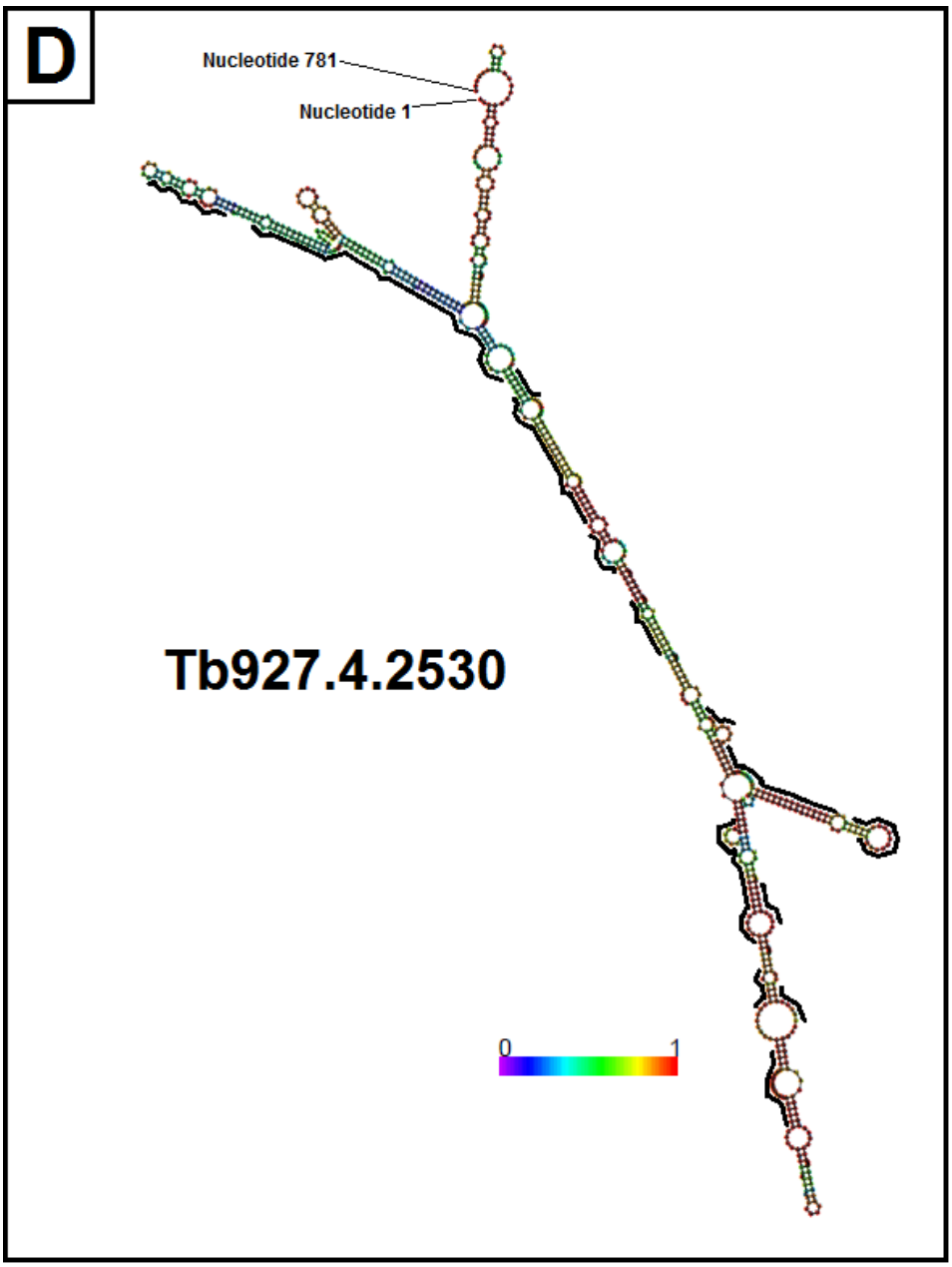


Figure S23. Enlarged version of Figure 24D.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

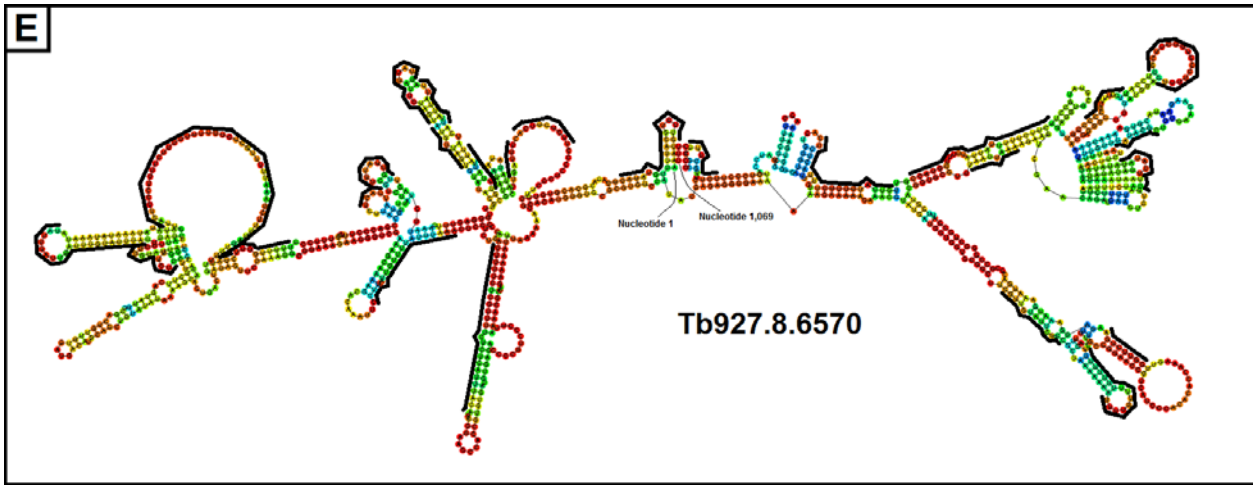


Figure S24. Enlarged version of Figure 24E.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

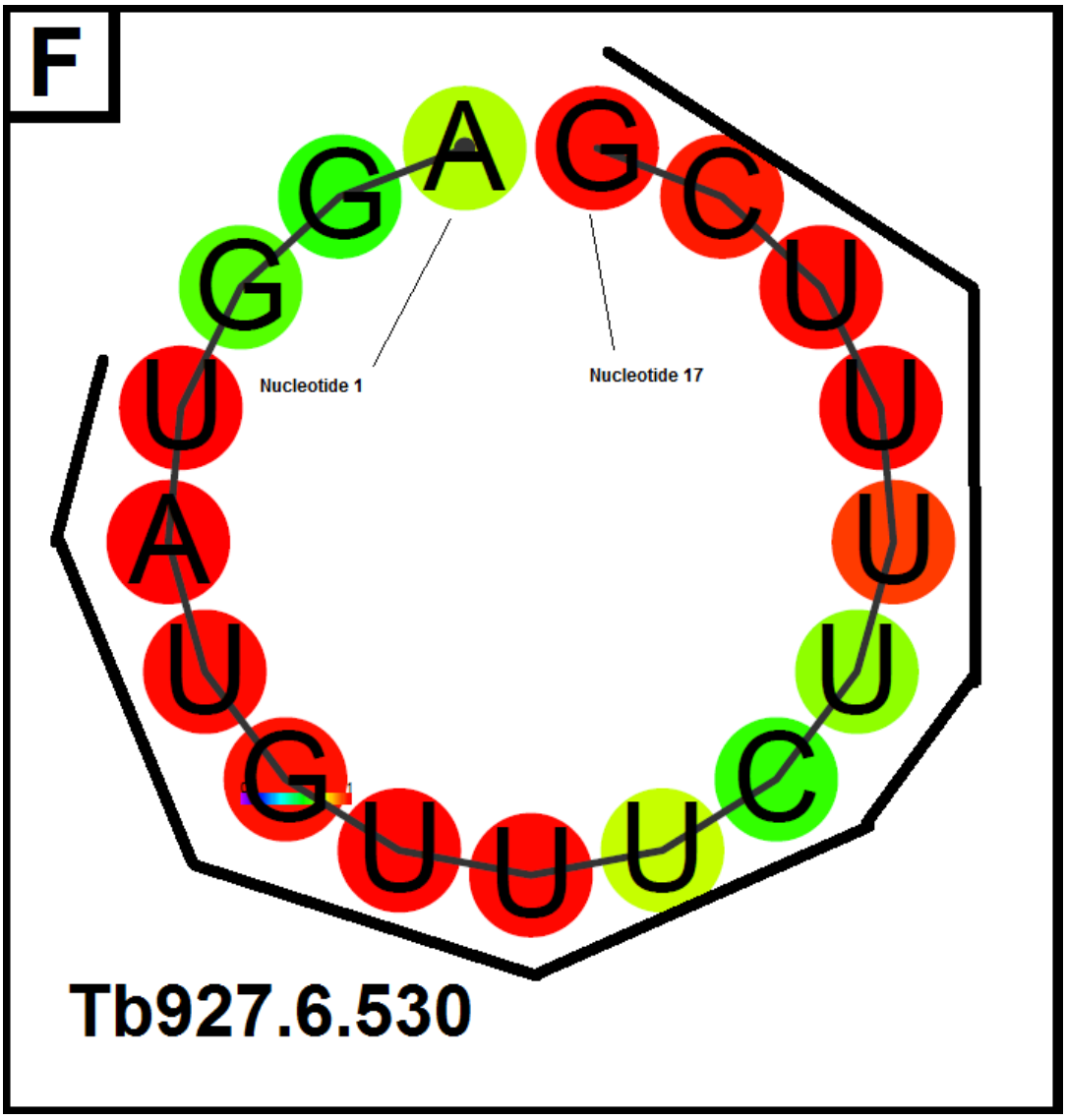


Figure S25. Enlarged version of Figure 24F.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

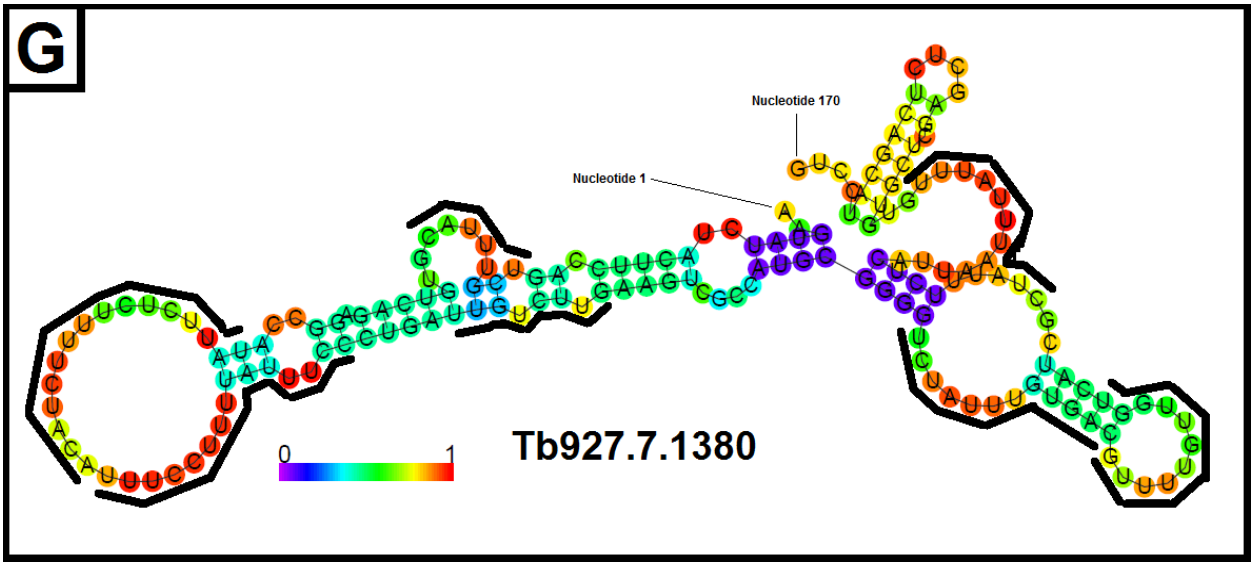


Figure S26. Enlarged version of Figure 24G.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

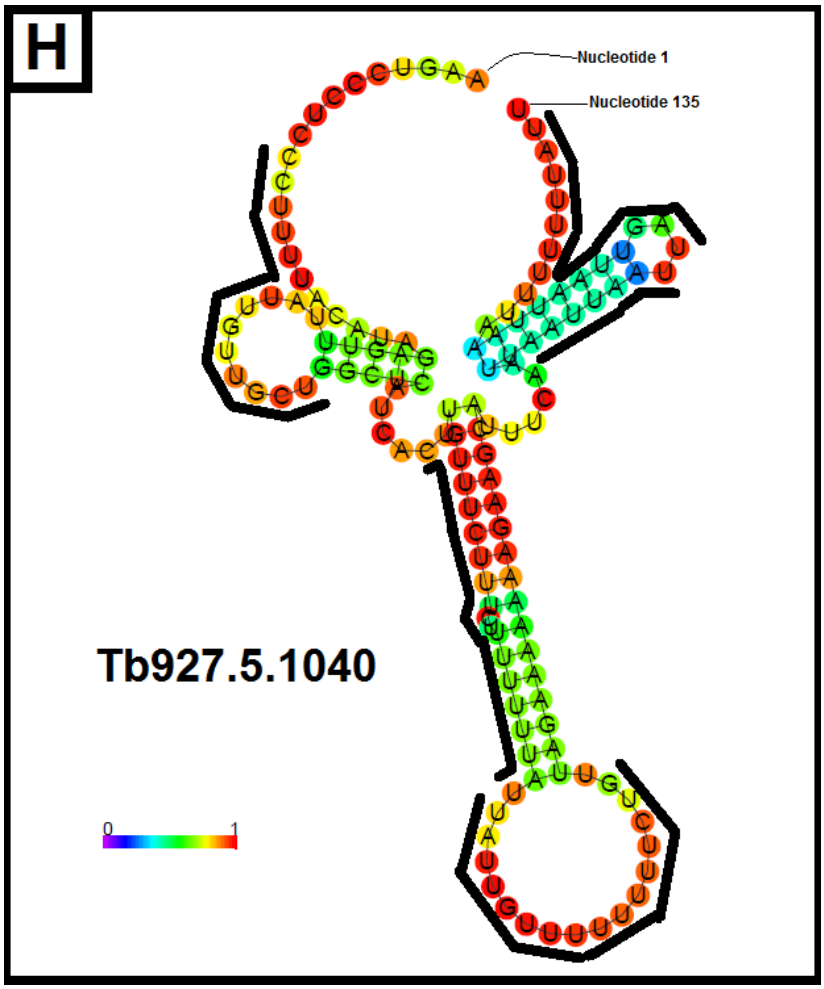


Figure S27. Enlarged version of Figure 24H.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

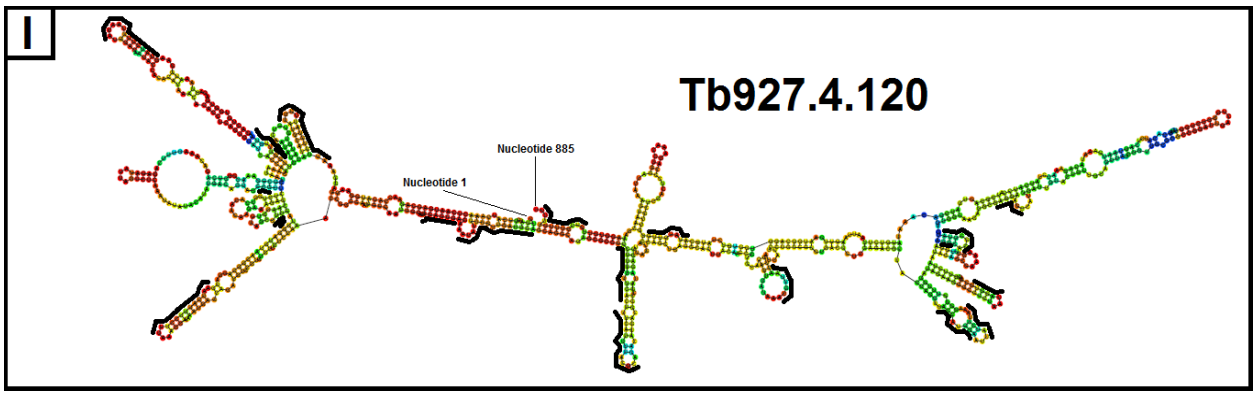


Figure S28. Enlarged version of Figure 24I.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

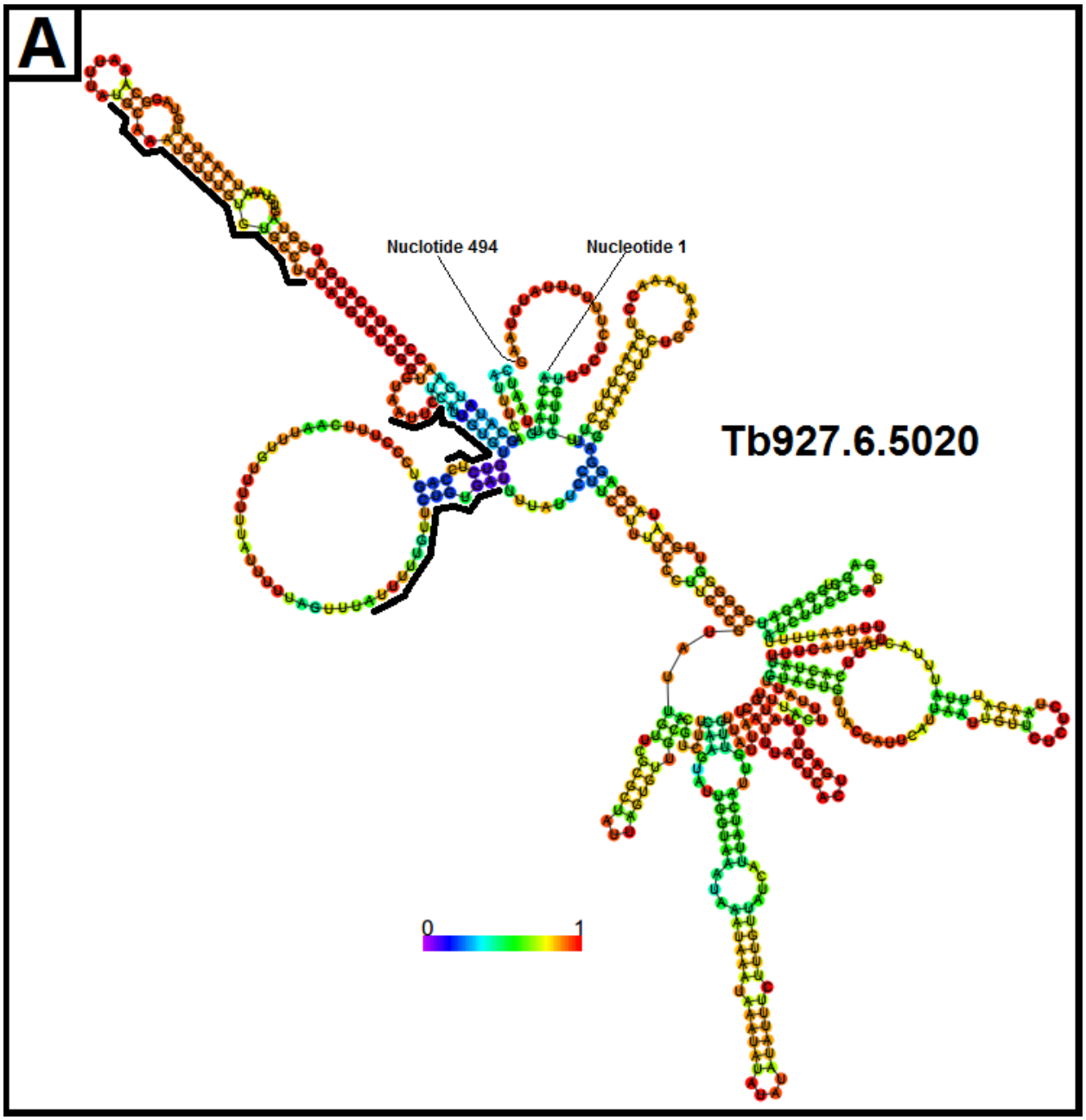


Figure S29. Enlarged version of Figure 25A.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

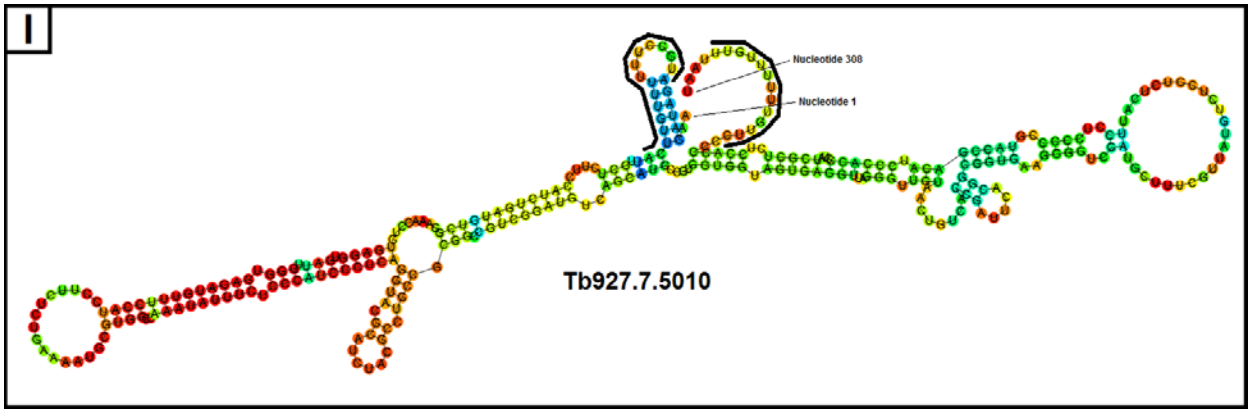


Figure S30. Enlarged version of Figure 25B.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

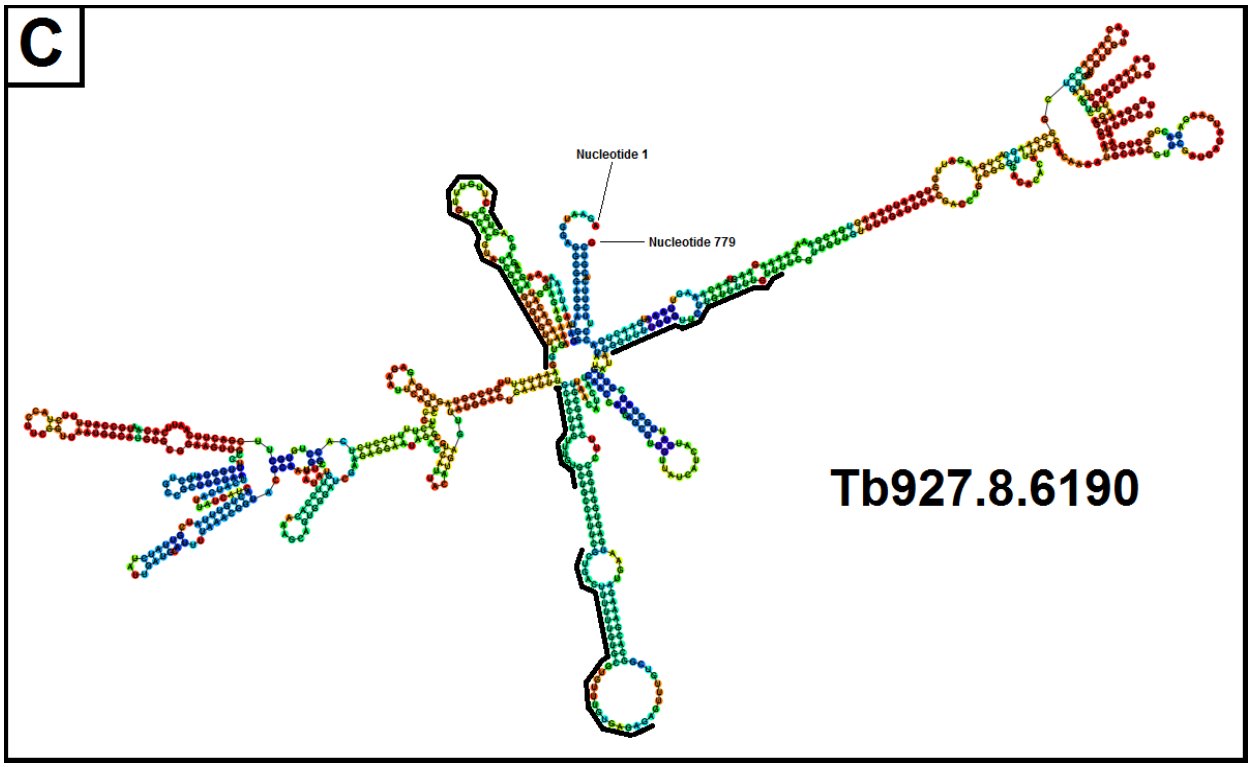


Figure S31. Enlarged version of Figure 25C.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

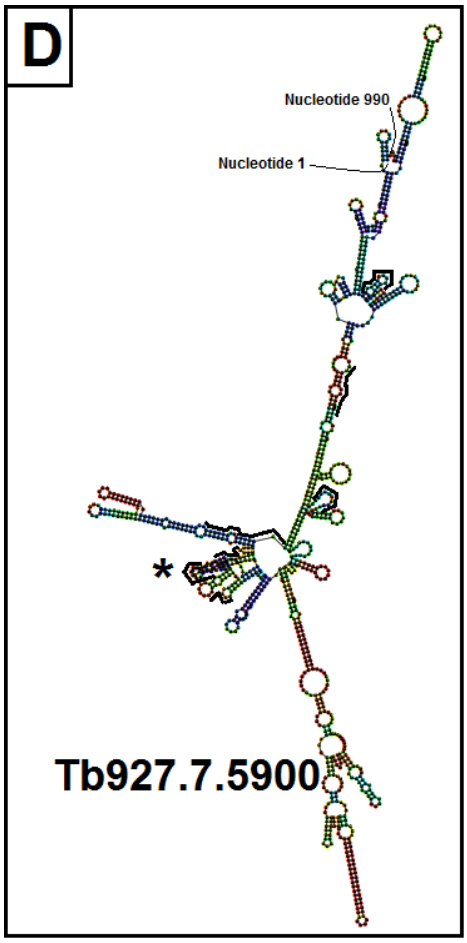


Figure S32. Enlarged version of Figure 25D.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

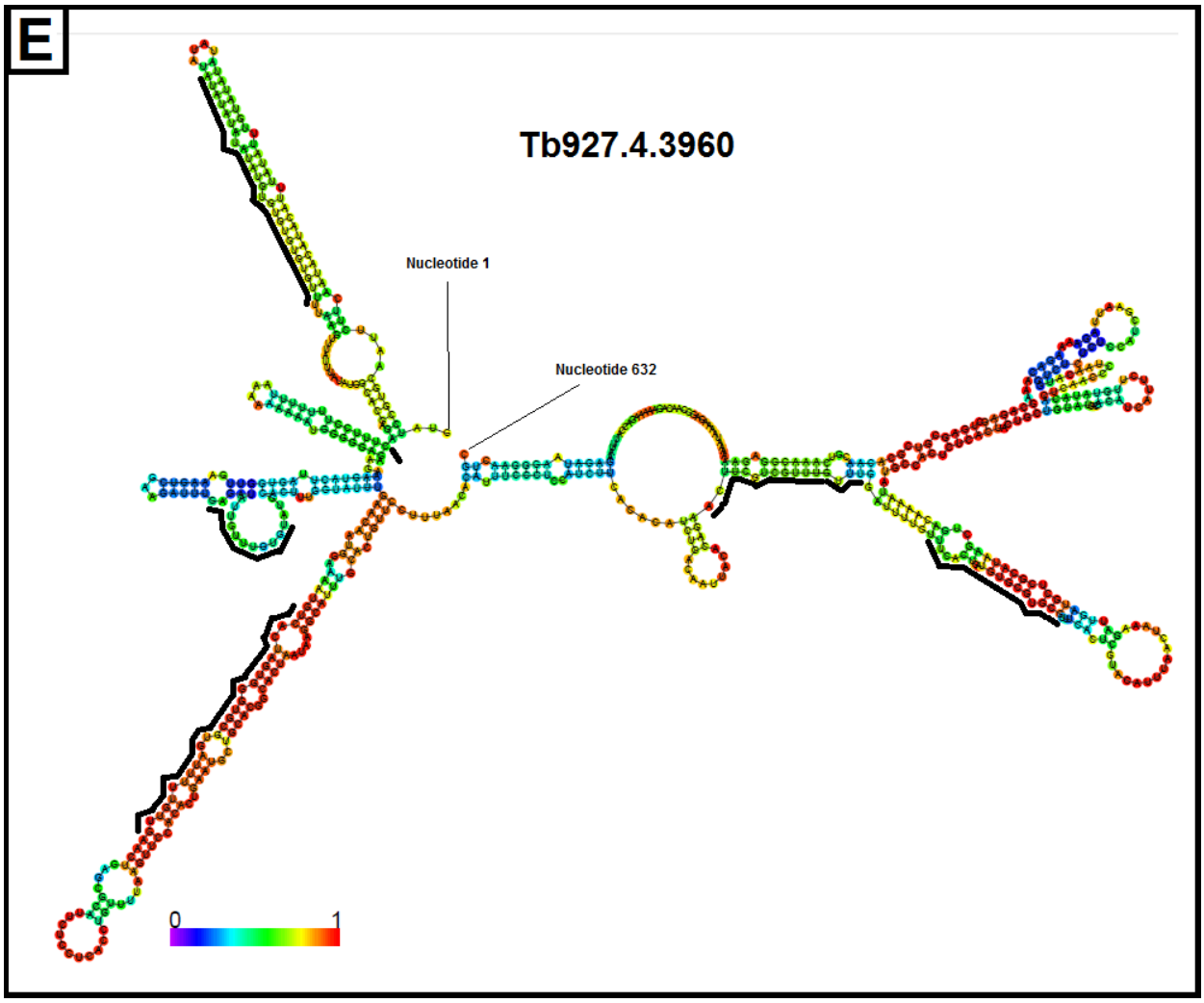


Figure S33. Enlarged version of Figure 25E.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

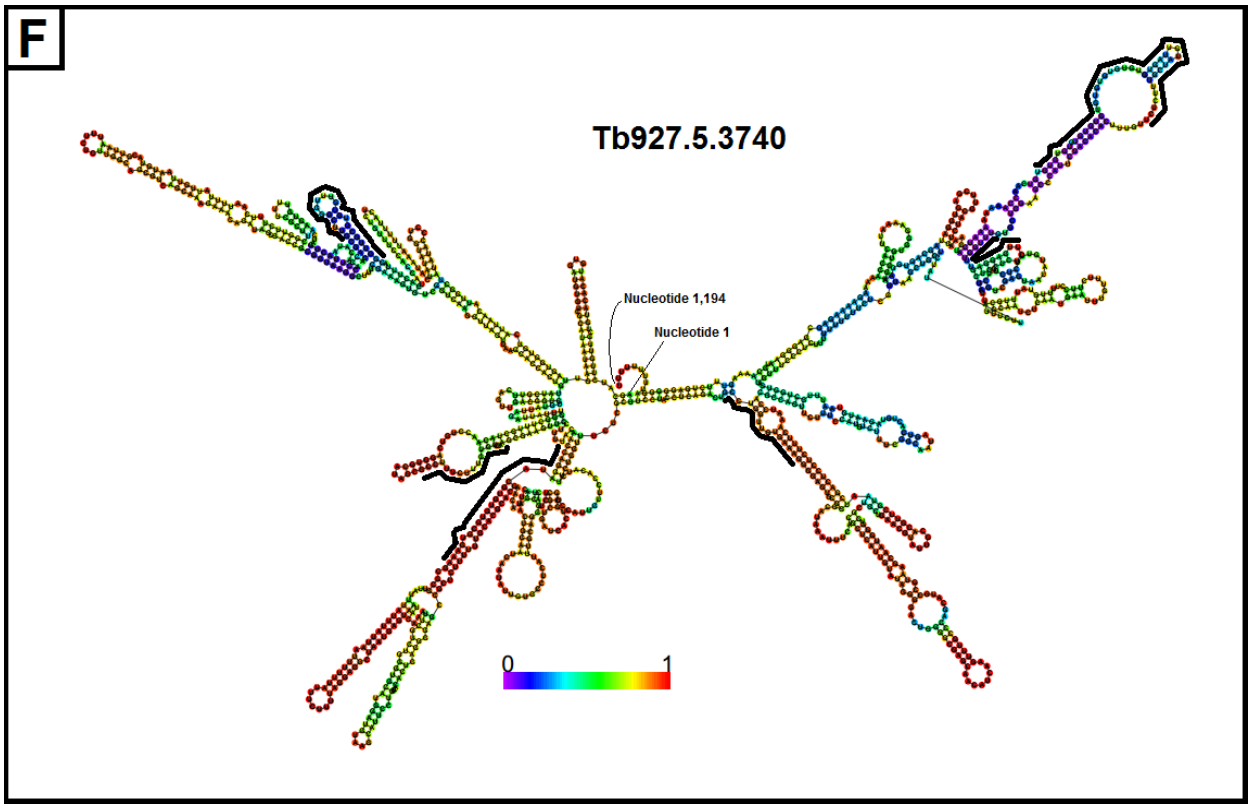


Figure S34. Enlarged version of Figure 25F.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

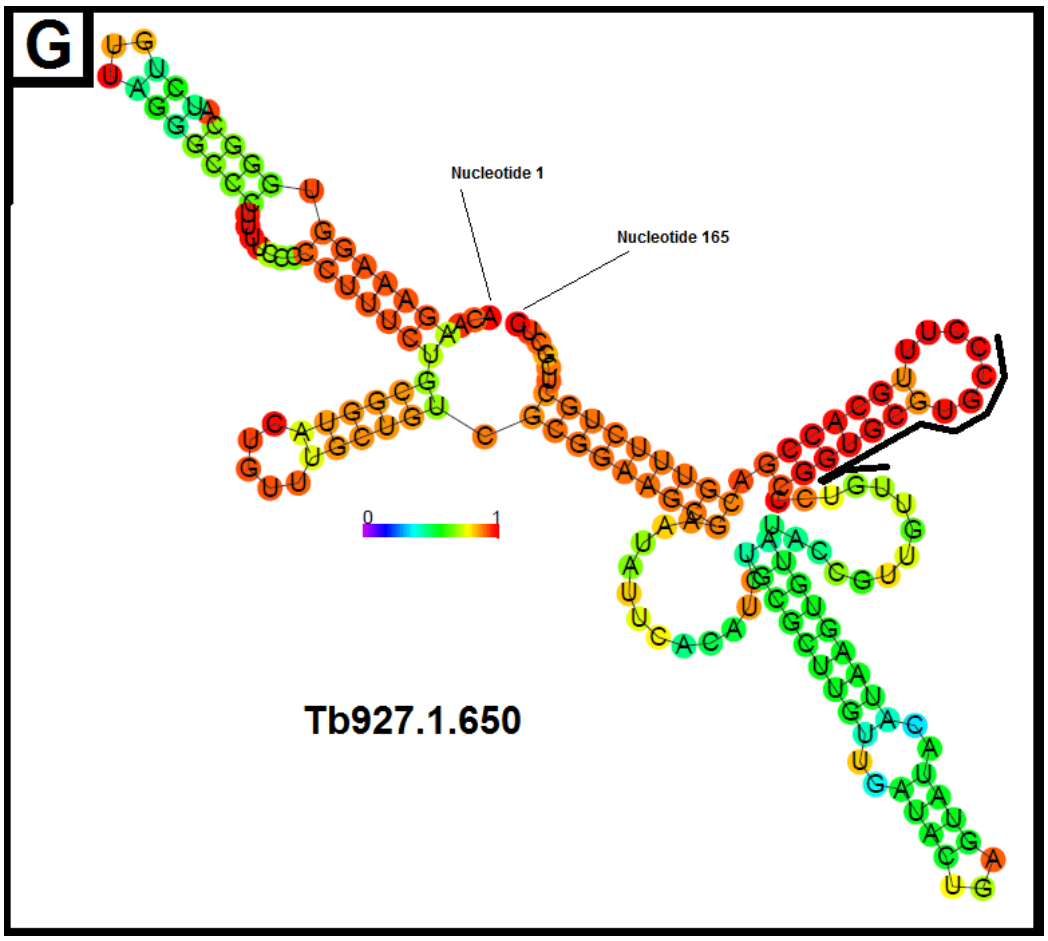


Figure S35. Enlarged version of Figure 25G.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

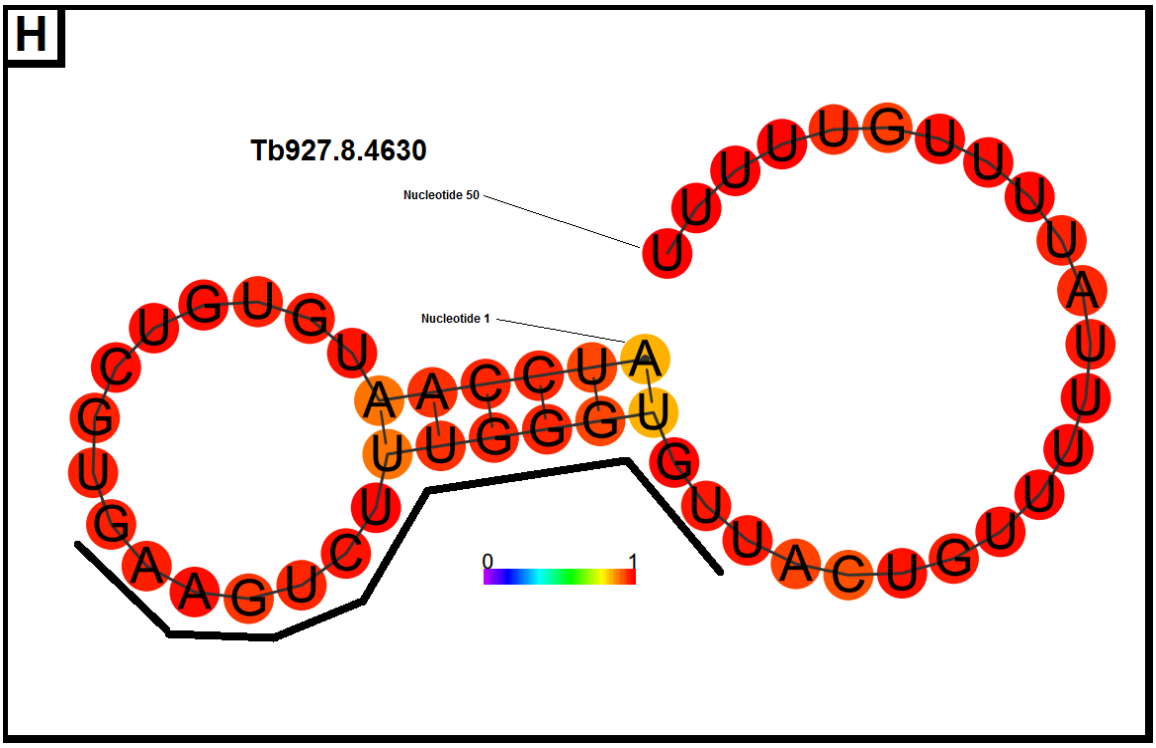


Figure S36. Enlarged version of Figure 25H.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

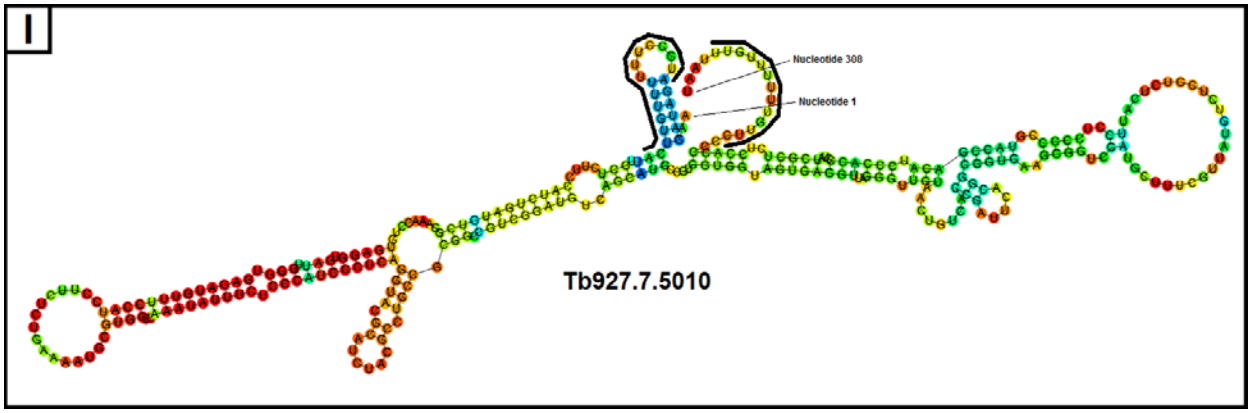


Figure S37. Enlarged version of Figure 25I.

This is a model of the mRNA 3' UTR secondary structure. Base pair probabilities are represented, with red being the highest and purple being the lowest. The gene identifier for the mRNA provided by the RNAseq has been included. Black bars represent the motif occurrence in the nucleotide sequences.

BIBLIOGRAPHY

1. Freire, E. R., R. Dhalia, D. M. Moura, T. D. Da Costa Lima, R. P. Lima, C. R. Reis, K. Hughes, R. C. Fiqueiredo, N. Standart, M. Carrington, and O. P. De Melo Neto. "The Four Trypanosomatid EIF4E Homologues Fall into Two Separate Groups, with Distinct Features in Primary Sequence and Biological Properties." *Molecular and Biochemical Parasitology* 176 (2011): 25-36.

2. Tarun, Salvador Z., Jr., Pamela Buchheit, Katie S. Eckman, Christopher J. Stairiker, Kenneth Stuart, and Alice S. Tarun. "TbEIF4E-3 Is a Unique Trypanosoma Brucei EIF4E Homolog Required for Translation and for Stabilization of MRNAs Essential in Cell Cycle and Morphology." Pending Publication in Experimental Parasitology (2013).

3. Godfrey, D.g., and Veronica Kilgour. "Enzyme Electrophoresis in Characterizing the Causative Organism of Gambian Trypanosomiasis." Transactions of the Royal Society of Tropical Medicine and Hygiene 70.3 (1976): 219-24.

4. Berriman, M. "The Genome of the African Trypanosome Trypanosoma Brucei." Science 309.5733 (2005): 416-22.

5. Barrett, M. P., D. W. Boykin, R. Brun, and R. R. Tidwell. "Human African Trypanosomiasis: Pharmacological Re-engagement with a Neglected Disease." British Journal of Pharmacology 152.8 (2007): 1155-171.

6. Barrett, Michael P., Richard Js Burchmore, August Stich, Julio O. Lazzari, Alberto Carlos Frasch, Juan José Cazzulo, and Sanjeev Krishna. "The Trypanosomiases." The Lancet 362.9394 (2003): 1469-480.

7. Simarro, Pere P., Abdoulaye Diarra, Jose A. Ruiz Postigo, José R. Franco, and Jean G. Jannin. "The Human African Trypanosomiasis Control and Surveillance Programme of the World Health Organization 2000–2009: The Way Forward." Ed. Serap Aksoy. PLoS Neglected Tropical Diseases 5.2 (2011): E1007.

8. Ekwanzala, F., C. Laveissiere, I. Maudlin, N. Van Meirvenne, D. H. Molyneux, V. Nantulya, M. Odiit, and L. Rombo. "Control and Surveillance of African Trypanosomiasis." WHO. 1998.

9. Fèvre, Eric M., Beatrix V. Wissmann, Susan C. Welburn, and Pascal Lutumba. "The Burden of Human African Trypanosomiasis." Ed. Simon Brooker. PLoS Neglected Tropical Diseases 2.12 (2008): E333.

10. Engels, Dirk, and Lorenzo Savioli. "Reconsidering the Underestimated Burden Caused by Neglected Tropical Diseases." Trends in Parasitology 22.8 (2006): 363-66.

11. Bisser, Sylvie, François-Xavier N'Siesi, Veerle Lejon, Pierre-Marie Preux, Simon Van Nieuwenhove, Constantin Miaka Mia Bilenge, and Philippe Büscher. "Equivalence Trial of Melarsoprol and Nifurtimox Monotherapy and Combination Therapy for the Treatment of Second-StageSleeping Sickness." The Journal of Infectious Diseases 195.3 (2007): 322-29.

12. Wang, C. C. "Molecular Mechanisms and Therapeutic Approaches to the Treatment of African Trypanosomiasis." Annual Review of Pharmacology and Toxicology 35.1 (1995): 93-127.

13. Milord, F., J. Pépin, L. Ethier, F. Milord, L. Loko, L. Ethier, and B. Mpia. "Efficacy and Toxicity of Eflornithine for Treatment of Trypanosoma Brucei Gambiense Sleeping Sickness." The Lancet 340.8820 (1992): 652-55.

14. Griffin, Constance A., Milan Slavik, Shu Chean Chien, Joachim Hermann, George Thompson, Oscar Blanc, Gordon D. Luk, Stephen B. Baylin, and Martin D. Abeloff. "Phase I Trial and Pharmacokinetic Study of Intravenous and Oral α-Difluoromethylornithine." Investigational New Drugs 5.2 (1987): 177-86.

15. Friedheim, E. A. "Mel B in the Treatment of Human Trypanosomiasis." The American Journal of Tropical Medicine and Hygiene 29.2 (1949): 173-80.

16. Pholig, G., S. Bernhard, J. Blum, C. Burri, A. Mpanya Kabeya, J.-P. Fina Lubaki, A. Mpoo Mpoto, B. Fungula Mununguu, G. Kambau Manesa Deo, P. Nsele Mutantu, F. Mbo Kuikumbi, A. Fukinsia Mintwo, A. Kayeye Mununguni, A. Dala, S. Macharia, C. Miaka Mia Bilenge, V. Kande Betu Ku Mesu, J. Ramon Franco, N. Dieyi Dituvanga, and C. Olson. 2008. Phase 3 trial of pafuramidine maleate (DB289), a novel, oral drug, for treatment of first stage sleeping sickness: safety and efficacy, abstract. 542. 57th Meeting of the American Society of Tropical Medicine and Hygiene, New Orleans.

17. Wenzler, T., D. W. Boykin, M. A. Ismail, J. E. Hall, R. R. Tidwell, and R. Brun. "New Treatment Option for Second-Stage African Sleeping Sickness: In Vitro and In Vivo Efficacy of Aza Analogs of DB289." Antimicrobial Agents and Chemotherapy 53.10 (2009): 4185-192.

18. Wenzler, T., D. W. Boykin, M. A. Ismail, J. E. Hall, R. R. Tidwell, and R. Brun. "New Treatment Option for Second-Stage African Sleeping Sickness: In Vitro and In Vivo Efficacy of Aza Analogs of DB289." *Antimicrobial Agents and Chemotherapy* 53.10 (2009): 4185-192.

19. Shatkin, A. "Capping of Eucaryotic MRNAs." *Cell* 9.4 (1976): 645-53.

20. Flynn, A., and C. G. Proud. "The Role of EIF4 in Cell Proliferation." Cancer Surv. 27 (1996): 293-310.

21. Dhalia, Rafael, Christian R.S. Reis, Eden R. Freire, Pollyanna O. Rocha, Rodolfo Katz, Joao R.C. Muniz, Nancy Standart, and Osvaldo P. De Melo Neto. "Translation Initiation in Leishmania Major: Characterisation of Multiple EIF4F Subunit Homologues." Molecular and Biochemical Parasitology 140.1 (2005): 23-41.

22. Rhoads, R. E. "EIF4E: New Family Members, New Binding Partners, New Roles." Journal of Biological Chemistry 284.25 (2009): 16711-6715.

23. Gebauer, Fátima, and Matthias W. Hentze. "Molecular Mechanisms of Translational Control." Nature Reviews Molecular Cell Biology5.10 (2004): 827-35.

24. Goyer, Charles, Michael Altmann, Hans Trachsel, and Nahum Sonenberg. "Identification and Characterization of Cap-binding Proteins from Yeast." *The Journal of Biological Chemistry* 254.13 (1989): 7603-610.

25. Mader, S., H. Lee, A. Pause, and N. Sonenberg. "The Translation Initiation Factor EIF-4E Binds to a Common Motif Shared by the Translation Factor EIF-4 Gamma and the Translational Repressors 4E-binding Proteins." *Molecular and Cellular Biology* 15.9 (1995): 4990-997.

26. Lamphear, Barry J., Regina Kirchweger, Tim Skern, and Robert E. Rhoads. "Mapping of Functional Domains in Eukaryotic Protein Synthesis Initiation Factor 4G (eIF4G) with Picornaviral Proteases." *The Journal of Biological Chemistry*270.37 (1995): 21975-1983.

27. Edery, I., M. Humbelin, A. Darveau, K. A. Lee, S. Milburn, J. W. Hershey, H. Trachsel, and N. Sonenberg. "Involvement of Eukaryotic Initiation Factor 4A in the Cap Recognition Process." *The Journal of Biological Chemistry* 258.18 (1983): 11398-1403.

28. Rozen, F., I. Edery, K. Meerovitch, T. E. Dever, W. C. Merrick, and N. Sonenberg. "Bidirectional RNA Helicase Activity of Eucaryotic Translation Initiation Factors 4A and 4F." *Molecular and Cellular Biology* 10.3 (1990): 1134-144.

29. Pyronnet, Stephane, Hiroaki Imataka, Anne-Claude Gingras, Rikiro Fukunaga, Tony Hunter, and Nahum Sonenberg. "Human Eukaryotic Translation Initiation Factor 4G (eIF4G) Recruits Mnk1 to Phosphorylate EIF4E." The EMBO Journal 18 (1999): 270-79.

30. Jung, Hosung, Christos G. Gkogkas, Nahum Sonenberg, and Christine E. Holt. "Remote Control of Gene Function by Local Translation." Cell 157.1 (2014): 26-40.

31. Lazaris-Karatzas, Anthoula, and Nahum Sonenberg. "The MRNA 5' Cap-Binding Protein, EIF-4E, Cooperates with V-myc or ElA in the Transformation of Primary Rodent Fibroblasts." *Molecular and Cellular Biology* 12.3 (1992): 1234-238.

32. Ruggero, David, Lorenzo Montanato, Li Ma, Wei Xu, Paola Londei, Carlos Cordon-Cardo, and Pier Paolo Pandolfi. "The Translation Factor EIF-4E Promotes Tumor Formation and Cooperates with C-Myc in Lymphomagenesis." *Nature Medicine* 10 (2004): 484-86.

33. Benedetti, A. De, S. Joshi-Barve, C. Rinker-Schaeffer, and R. E. Rhoads. "Expression of Antisense RNA against Initiation Factor EIF-4E MRNA in HeLa Cells Results in Lengthened Cell Division Times, Diminished Translation Rates, and Reduced Levels of Both EIF-4E and the P220 Component of E*IF-4F." Molecular and Cellular* Biology 11.11 (1991): 5435-445

34. Morley, Simon J., and Linda McKendrick. "Involvement of Stress-activated Protein Kinase and P38/RK Mitogen-activated Protein Kinase Signaling Pathways in the Enhanced Phosphorylation of Initiation Factor 4E in NIH 3T3 Cells." *The Journal of Biological Chemistry* 272.28 (1997): 17887-7893.

35. Yoffe, Y., J. Zuberek, A. Lerer, M. Lewdorowicz, J. Stepinski, M. Altmann, E. Darzynkiewicz, and M. Shapira. "Binding Specificities and Potential Roles of Isoforms of Eukaryotic Initiation Factor 4E in Leishmania." *Eukaryotic Cell* 5.12 (2006): 1969-979.

36. Panigrahi, Aswini K., Yuko Ogata, Alena Zíková, Atashi Anupama, Rachel A. Dalley, Nathalie Acestor, Peter J. Myler, and Kenneth D. Stuart. "A Comprehensive Analysis of Trypanosoma Brucei Mitochondrial Proteome." *Proteomics* 9.2 (2009): 434-50.

37. Dayhoff, M. O., and R. M. Schwartz. "Evidence On The Origin Of Eukaryotic Mitochondria From Protein And Nucleic Acid Sequences." *Annals of the New York Academy of Sciences* 361.1 Origins and E (1981): 92-104.

38. Culjkovic, B., I. Topisirovic, L. Skrabanek, M. Ruiz-Gutierrez, and K. L.b. Borden. "EIF4E Is a Central Node of an RNA Regulon That Governs Cellular Proliferation." The Journal of Cell Biology 175.3 (2006): 415-26.

39. Culjkovic, Biljana, Ivan Topisirovic, and Katherine L.b. Borden. "Controlling Gene Expression through RNA Regulons: The Role of the Eukaryotic Translation Initiation Factor EIF4E." Cell Cycle 6.1 (2007): 65-69.

40. Maas, Werner K., and A.J. Clark. "Studies on the Mechanism of Repression of Arginine Biosynthesis in Escherichia ColiII. Dominance of Repressibility in Diploids." *Journal of Molecular Biology* 8.3 (1964): 365-70.

41. Wirtz, E., Leal, S., Ochatt, C., and Cross, G. A., 1999. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol Biochem Parasitol 99, 89-101.

42. Wang, Z., Morris, J. C., Drew, M. E., and Englund, P. T., 2000. Inhibition of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promoters. J Biol Chem 275, 40174-40179.

43. Redmond, S., Vadivelu, J., and Field, M. C., 2003. RNAit: an automated web-based tool for the selection of RNAi targets in Trypanosoma brucei. Mol Biochem Parasitol 128, 115-118.

44. Medina-Medina, N., A. Broka, S. Lacey, H. Lin, E. S. Klings, C. T. Baldwin, M. H. Steinberg, and P. Sebastiani. "Comparing Bowtie and BWA to Align Short Reads from a RNA-Seq Experiment." *6th International Conference on Practical Applications of Computational Biology & Bioinformatics*. Vol. 154. Berlin: Springer, 2012. 197-207.

45. Patel, Satish, and James Lyons-Weiler. "CaGeda: A Web Application for the Integrated Analysis of Global Gene Expression Patterns in Cancer." *Applied Bioinformatics* 3.1 (2004): 49-62.

46. Archer, Stuart K., Van-Duc Luu, Rafael A. De Queiroz, Stefanie Brems, and Christine Clayton. "Trypanosoma Brucei PUF9 Regulates MRNAs for Proteins Involved in Replicative Processes over the Cell Cycle." Ed. Elisabetta Ullu. *PLoS Pathogens* 5.8 (2009): E1000565.

47. Ettwiller, Laurence, Benedict Paten, Mirana Ramialison, Ewan Birney, and Joachim Wittbrodt. "Trawler: De Novo Regulatory Motif Discovery Pipeline for Chromatin Immunoprecipitation." *Nature Methods* 4.7 (2007): 563-65.

48. Mathews, D. H. "Incorporating Chemical Modification Constraints into a Dynamic Programming Algorithm for Prediction of RNA Secondary Structure." Proceedings of the National Academy of Sciences 101.19 (2004): 7287-292.

49. Gruber, A. R., R. Lorenz, S. H. Bernhart, R. Neubock, and I. L. Hofacker. "The Vienna RNA Websuite." Nucleic Acids Research 36.Web Server (2008): W70-74.

50. Lorenz, Ronny, Stephan H. Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. "ViennaRNA Package 2.0." Algorithms for Molecular Biology 6.1 (2011): 26.

51. Mao, Yuan, Hamed Najafabadi, and Reza Salavati. "Genome-wide Computational Identification of Functional RNA Elements in Trypanosoma Brucei." BMC Genomics 10.1 (2009): 355.

52. Lejbkowicz, Flavio, Charles Goyer, Andre Darveau, Sonia Neron, Real Lemieux, and Nahum Sonenberg. "A Fraction of the MRNA 5' Cap-binding Protein, Eukaryotic Initiation Factor 4E, Localizes to the Nucleus." *PNAS* 89 (1992): 9612-616.

53. Cassola, Alejandro, Javier G. De Gaudenzi, and Alberto C. Frasch. "Recruitment of MRNAs to Cytoplasmic Ribonucleoprotein Granules in Trypanosomes." *Molecular Microbiology* 65.3 (2007): 655-70.

54. Lejbkowicz, F. "A Fraction of the MRNA 5' Cap-Binding Protein, Eukaryotic Initiation Factor 4E, Localizes to the Nucleus." *Proceedings of the National Academy of Sciences* 89.20 (1992): 9612-616.

55. Ferraiuolo, M. A. "A Role for the EIF4E-binding Protein 4E-T in P-body Formation and MRNA Decay." *The Journal of Cell Biology* 170.6 (2005): 913-24.

56. Burden, Heather E., and Zhiping Weng. "Identification of Conserved Structural Features at Sequentially Degenerate Locations in Transcription Factor Binding Sites." *Genome Informatics* 16.1 (2005): 49-58.

57. Näär, Anders M., Jean-Marle Boutin, Steven M. Lipkin, Victor C. Yu, Jeffrey M. Holloway, Christopher K. Glass, and Michael G. Rosenfeld. "The Orientation and Spacing of Core DNA-binding Motifs Dictate Selective Transcriptional Responses to Three Nuclear Receptors." *Cell* 65.7 (1991): 1267-279.

58. Connor, Frances, Peter D. Cary, Christopher M. Read, Nicola S. Preston, Paul C. Driscoll, Paul Denny, Colyn Crane-Robinson, and Alan Ashworth. "DNA Binding and Bending Properties of the Postmeiotically Expressed Sry-related Protein Sox-5." *Nucleic Acids Research* 22.16 (1994): 3339-346.

59. Santiago-Rivera, Zulma I., David G. Gorenstein, John S. Williams, and Ourania M. Andrisani. "Bacterial Expression and Characterization of the CREB BZip Module: Circular Dichroism and 2DH-NMR Studies." *Protein Science* 2.9 (1993): 1461-471.

60. Gonda, Thomas J., and Donald Metcalf. "Expression of Myb, Myc and Fos Proto-oncogenes during the Differentiation of a Murine Myeloid Leukaemia." *Nature* 310.5974 (1984): 249-51.

61. Sukhatme, Vikas P., Xinmin Cao, Louise C. Chang, Chon-Hwa Tsai-Morris, Dorothy Stamenkovich, Paulo C.p. Ferreira, Donna R. Cohen, Steven A. Edwards, Thomas B. Shows, Tom Curran, Michelle M. Le Beau, and Eileen D. Adamson. "A Zinc Finger-encoding Gene Coregulated with C-fos during Growth and Differentiation, and after Cellular Depolarization." *Cell* 53.1 (1988): 37-43.

62. Greenberg, Michael E., and Edward B. Ziff. "Stimulation of 3T3 Cells Induces Transcription of the C-fos Proto-oncogene." *Nature* 311.5985 (1984): 433-38.

63. Qian, Su, Maria Capovilla, and Vincenzo Pirrotta. "The Bx Region Enhancer, a Distant Cis-control Element of the Drosophila Ubx Gene and Its Regulation by Hunchback and Other Segmentation Genes." *The EMBO Journal* 10.6 (1991): 1415-425.

64. Davies, Christopher, V. Ramakrishnan, and Stephen W. White. "Structural Evidence for Specific S8–RNA and S8–protein Interactions within the 30S Ribosomal Subunit: Ribosomal Protein S8 from Bacillus Stearothermophilus at 1.9 å Resolution." *Structure* 4.9 (1996): 1093-104.

65. Lipsick, Baluda JS. "The Myb Oncogene." *Gene Amplification and Analysis* 4 (1986): 73-98.

66. Mcginnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. J. Gehring. "A Conserved DNA Sequence in Homoeotic Genes of the Drosophila Antennapedia and Bithorax Complexes." *Nature* 308.5958 (1984): 428-33.

67. Rajeswari, Moganty R., and Aklank Jain. "High-mobility-group Chromosomal Proteins, HMGA1 as Potential Tumour Markers." *Current Science* 82.7 (2002): 838-44.

68. Ruben, S. M., J. F. Klement, T. A. Coleman, M. Maher, C. H. Chen, and C. A. Rosen. "I-Rel: A Novel Rel-related Protein That Inhibits NF-kappa B Transcriptional Activity." *Genes & Development* 6.5 (1992): 745-60.

69. Schwarz-Sommer, Z., P. Huijser, W. Nacken, H. Saedler, and H. Sommer. "Genetic Control of Flower Development by Homeotic Genes in Antirrhinum Majus." *Science* 250.4983 (1990): 931-36.

70. West, A. G., P. Shore, and A. D. Sharrocks. "DNA Binding by MADS-box Transcription Factors: A Molecular Mechanism for Differential DNA Bending." *Molecular and Cellular Biology* 17.5 (1997): 2876-887.

71. Johnson, P. "Eukaryotic Transcriptional Regulatory Proteins." *Annual Review of Biochemistry* 58.1 (1989): 799-839.

72. Blackwood, E., and R. Eisenman. "Max: A Helix-loop-helix Zipper Protein That Forms a Sequence-specific DNA-binding Complex with Myc." *Science* 251.4998 (1991): 1211-217.

73. Neff, Carrie L., and Alan B. Sachs. "Eukaryotic Translation Initiation Factors 4G and 4A from Saccharomyces Cerevisiae Interact Physically and Functionally." Molecular and Cellular Biology 19.8 (1999): 5557-564.

74. Mcdonald, Heather B., Russell J. Stewart, and Lawrence S.b. Goldstein. "The Kinesin-like Ncd Protein of Drosophila Is a minus End-directed Microtubule Motor." *Cell* 63.6 (1990): 1159-165.

75. Enos, A. "Mutation of a Gene That Encodes a Kinesin-like Protein Blocks Nuclear Division in A. Nidulans." *Cell* 60.6 (1990): 1019-027.

76. Zhang, Hui, Ryuji Kobayashi, Konstantin Galaktionov, and David Beach. "Pl9skp1 and P45skp2 Are Essential Elements of the Cyclin A-CDK2 S Phase Kinase." *Cell* 82.6 (1995): 915-25.

77. Nolan, D. P. "Characterization of a Novel Alanine-rich Protein Located in Surface Microdomains in Trypanosoma Brucei." *Journal of Biological Chemistry* 275.6 (2000): 4072-080.

78. Andrei, M. A. "A Role for EIF4E and EIF4E-transporter in Targeting MRNPs to Mammalian Processing Bodies." *Rna* 11.5 (2005): 717-27.