CellPress

Review

Illuminating Parasite Protein Production by Ribosome Profiling

Marilyn Parsons[1,2,*] and Peter J. Myler[1,2,3]

While technologies for global enumeration of transcript abundance are well-developed, those that assess protein abundance require tailoring to penetrate to low-abundance proteins. Ribosome profiling circumvents this challenge by measuring global protein production via sequencing small mRNA fragments protected by the assembled ribosome. This powerful approach is now being applied to protozoan parasites including trypanosomes and *Plasmodium*. It has been used to identify new protein-coding sequences (CDSs) and clarify the boundaries of previously annotated CDSs in *Trypanosoma brucei*. Ribosome profiling has demonstrated that translation efficiencies vary widely between genes and, for trypanosomes at least, for the same gene across stages. The ribosomal proteins are themselves subjected to translational control, suggesting a means of reinforcing global translational regulation.

Ribosome Profiling: Technical Aspects

The recently developed technology of **ribosome profiling** (see Glossary) has illuminated various aspects of gene expression [1,2], ranging from definition of coding regions [3] to genome-wide quantitation of protein production [4] to identification of mechanisms underlying translation [5] and its inhibition by drugs [6]. Ribosome profiling provides an unbiased means of identifying those regions of mRNAs associated with assembled ribosomes and hence engaged in protein production. As pioneered by Ingolia and Weissman [7] and summarized in Figure 1A (Key Figure), cell lysates are treated with nuclease to digest RNA that is not protected by protein. Then, the small fragments protected by the ribosome (~28 nt) are purified and used to make an RNA-seq library, employing a procedure that minimizes bias in sequence representation. When the resulting reads are mapped back to the genome (together with standard RNA-seq reads of similar size), one can quantify which mRNAs (and regions thereof) are contributing to protein production in a given condition (Figure 1B). Ribosome profiling provides a large dynamic range and codon-level resolution, making it considerably more sensitive than polysome profiling in terms of quantifying translation or precisely defining CDSs. Thus, ribosome profiling allows differences in protein production between genes or growth conditions to be parsed into changes in mRNA abundance and/or **translation efficiency** (TE, defined here as the ratio of the normalized **ribosome footprint** reads to the mRNA reads). While ribosome profiling has been utilized in several model organisms and human cells (e.g., [7–9]), the approach has only recently seen application in infectious disease biology. Key examples include probing the effects of viral infection on host cells [10], investigating the mechanism of macrolide inhibition [6], and describing developmental regulation of protein production in the protozoan parasites *Trypanosoma brucei* [11,12], *Trypanosoma cruzi* [13], and *Plasmodium falciparum* [14]. In addition, it has been employed in gene discovery in *T. brucei* [15] and viruses [3,16]. This article will focus on the contribution of ribosome profiling to the understanding of gene expression in trypanosomatids, but will also touch on recent data for *P. falciparum*.

Trends

Ribosome profiling technology provides a means to precisely identify and quantitate sequences being translated in a given sample.

Among parasites, ribosome profiling studies have been published for *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Plasmodium falciparum*.

Comparison of ribosome footprint read counts with mRNA abundance reveals the translation efficiency of individual genes, which varies between genes and across conditions.

Switches between stages that have very different rates of protein production are common features of parasite life cycles.

Details of stage-specific regulation of protein synthesis suggest important features of life cycles and transmission.

[1]Center for Infectious Disease Research (formerly Seattle Biomedical Research Institute), 307 Westlake Avenue North STE 500, Seattle, WA 98109 USA
[2]Department of Global Health, Box 357965, University of Washington, Seattle, WA 98195, USA
[3]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA 98195, USA

*Correspondence: marilyn.parsons@cidresearch.org (M. Parsons).

 http://dx.doi.org/10.1016/j.pt.2016.03.005

Ribosome Profiling Contributes to Curation of Parasite Genomes

The genome sequences of *T. brucei*, *T. cruzi*, and *Leishmania major* were published in 2005 [17–19]. Trypanosomatid genes are assembled in long polycistronic arrays separated by transcription initiation and termination sites [20,21]. Individual mRNAs are then generated by *trans*-splicing of a 39 nt **spliced leader** (SL) RNA (transcribed from one or more repeat arrays in the genome) to generate the 5′-end, and by polyadenylation to generate the 3′-end (Figure 1C). In 2005, about one-third of the *T. brucei* CDSs were annotated as encoding proteins without a known function, but having predicted orthologs in one or more of the other trypanosomatids at the time. These predicted proteins were called 'hypothetical protein, conserved'. However, a large number of CDSs with unknown function appeared to encode proteins that are not conserved across these or other species; these gene products were dubbed 'hypothetical protein' or 'hypothetical protein, unlikely'. This situation has now changed significantly, as reported in TriTrypDB [22], because further experimentation and bioinformatic analysis have shown that some proteins thought to be species-specific are actually conserved, and that many conserved hypothetical proteins can be assigned a function.

Ribosome profiling is a useful adjunct to conventional RNA-seq and 5′-end mapping using the SL for evaluating putative CDSs because it can provide support for retaining (or deleting) a CDS or for changing its boundaries. Thus far it has been systematically applied to genome curation only for *T. brucei* [15]. There, we consolidated data from two earlier ribosome profiling studies [11,12] together with other types of data [23–26]. Data visualization using the genome browser and annotation tool Artemis [27] allowed us to eliminate over 600 *T. brucei* CDSs where the ribosome profiling, mRNA, and/or SL RNA-seq evidence data contradicted the gene model. These include cases in which the mRNA reads (and ribosome footprints, if present) aligned to the opposite DNA strand (to that bearing the putative CDS) or in which the putative CDS was in the 3′-untranslated region (UTR) of an adjacent gene and lacked ribosome footprints. In other cases, the putative CDS was split between two mRNAs, and any ribosome footprints mapped to a portion of the downstream gene (Figure 2). This process eliminated 81% of the genes with products annotated as 'hypothetical protein, unlikely' in the *T. brucei* genome sequence (version 5.1) [15]. While many of the remaining 'hypothetical protein, unlikely' CDSs may not be protein-coding, we did not eliminate them solely because of the absence of ribosome footprints because they may be expressed in other developmental stages or conditions. Of the 10 600 annotated transcripts of *T. cruzi*, only 79% have detectable ribosome footprints (>15 per CDS) in at least one of the two stages examined [13], suggesting that some predicted CDSs may not be protein-coding. It is possible that some of the genes without footprints encode proteins expressed in a different stage. In addition, without detailed gene-by-gene visualization of multi-mapping reads and overlapping ORFs, it is difficult to know whether all transcripts with footprints are actually translated.

Ribosome profiling had less to offer with respect to culling incorrectly annotated CDSs in *P. falciparum*. Even though two-thirds of the CDSs were originally annotated in PlasmoDB as 'hypothetical proteins' [28,29], only 40% of the 5398 protein-coding genes are currently annotated as having unknown function. Furthermore, all but 239 of these hypothetical proteins are listed as conserved and are thus likely to be *bona fide* protein-coding genes.

Ribosome profiling also allows experimental confirmation (or revision) of the start codon used for translation. The ribosome profiling data suggested that, in trypanosomatids, the CDS start codon is almost always the first ATG downstream of the major SL site, although some obvious **upstream open reading frames** (uORFs) were observed [11,12]. Mapping of the SL attachment sites suggested that many annotated start codons required revision [24–26]. However, it is not trivial to map the major SL site bioinformatically (especially for mRNAs with low abundance) because it is sometimes difficult to determine if the SL site with the most reads is on the same mRNA as the CDS. Mapping start codons often

Glossary

Bloodstream form of *Trypanosoma brucei* (BF): this is the form that exists within the mammalian host, and here is restricted to the long slender forms, which are actively dividing. Unless noted, BFs are derived from infected mice.

Procyclic cultured form (PCF): the *in vitro* cultured procyclic form of *T. brucei*. This form has been used in ribosome profiling experiments.

Procyclic form (PF): the actively dividing form of *T. brucei* that resides in the tsetse fly midgut.

Ribosome footprints: the short regions of mRNA (~28 nt) that assembled ribosomes protect from nuclease digestion.

Ribosome profiling: a technology that combines sequencing of fragments of mRNA protected by ribosomes and RNA-seq to quantitate protein production and translation efficiency globally.

Spliced leader (SL): the common 39 nt sequence that is spliced onto all mRNAs in trypanosomatids, thereby forming the 5′-end of all mature mRNAs.

Translation efficiency (TE): the relative efficiency of translation of a specific mRNA in a given condition, defined here as the ratio between normalized ribosome footprint read counts and normalized mRNA read counts.

Upstream open reading frame (uORF): an ORF whose start codon occurs before the start codon of the main coding sequence, but is on the same transcript.

Key Figure

Overview of Ribosome Profiling Protocol

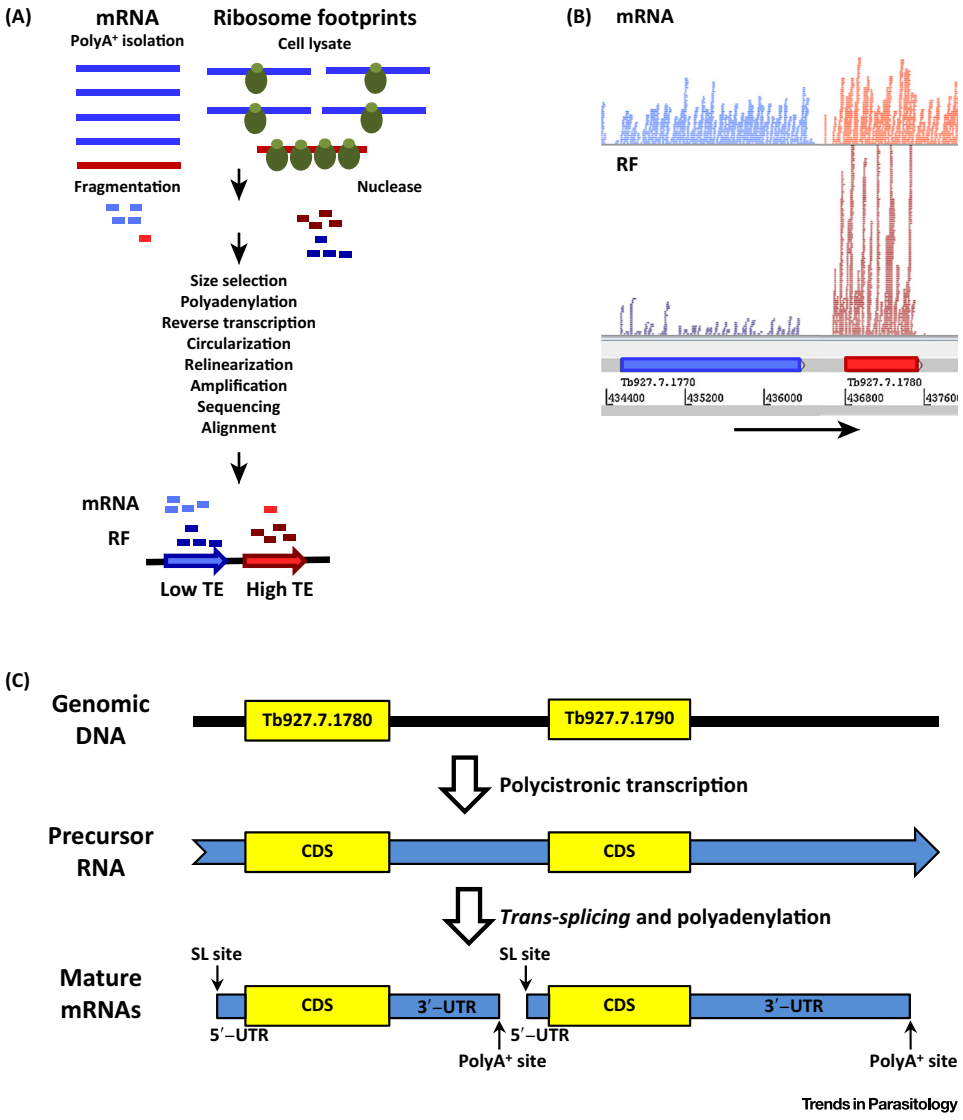

Figure 1. Ribosome profiling provides an unbiased means of identifying the regions of mRNAs that are associated with assembled ribosomes and that are thus engaged in protein production. In addition, it allows quantitation of the extent of ribosome association with each mRNA (or a portion thereof) of interest. (A) Protocol summary. Two mRNA libraries are made from each sample: one from ribosome footprints (RF) produced by nuclease digestion of a cell lysate and the other from fragmented purified polyadenylated (polyA$^+$) mRNA, as detailed [7,67]. Following sequencing and alignment with the relevant genome, read counts for each coding sequence (CDS) are quantified. The green structures in the cartoon represent ribosomes that are associated with the blue and red mRNAs, generating four ribosome footprints for each CDS in this example. However, because the mRNA abundances differ, the corresponding translation efficiencies (TEs) of the mRNAs differ, with the TE for the red mRNA being four fold higher. (B) Two adjacent *Trypanosoma brucei* genes with similar mRNA levels, but different protein production levels. More ribosome footprints correspond to the red CDS, reflecting more protein production and hence a higher TE than for the blue gene. Data for an *in vivo* derived slender bloodstream form (BF) sample are visualized here and elsewhere using Artemis [27]. Each small horizontal bar represents an aligned sequence read, and the arrow shows the direction of transcription. Note that the extended 3′-UTR of the red gene, as defined by mRNA reads (top), lacks ribosome footprints (middle), as expected for an untranslated region of mRNA. A map of the region of

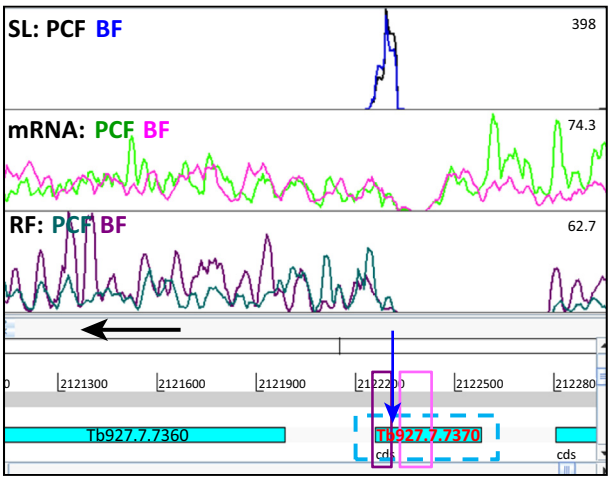

Figure 2. A Putative Coding Sequence Encoding a Hypothetical Protein that Was Eliminated by Analysis of Ribosome Profiling Data. The top three panels depict the RNA-seq reads (using a linear scale) derived from **procyclic cultured forms** (PCF) and bloodstream forms (BF) for the spliced leader (SL), mRNA, and ribosome footprints (RF); each trace is color-coded according to developmental stage, as indicated. Note that traces represent raw read counts and the number at right of each panel indicates the read count of the highest peak. The bottom panel shows the coding sequences (CDSs) aligned across the corresponding region of chromosome 7 of the *Trypanosoma brucei* TREU 927 genome. The CDS for Tb927.7.7370 (enclosed by the broken cyan line) is truncated at the blue arrow by a major SL site (shown in the top panel), with most of the CDS lying in the 3′-untranslated region (UTR) of gene to the right. A region in the middle of the CDS lacks mRNA reads (pink box), while all ribosome footprints that map to Tb927.7.7370 (purple box) lie in the 5′-UTR of the adjacent gene (Tb927.7.7360), which encodes the protein kinase CRK2. Thus Tb927.7.7370 does not lie on a contiguous transcript, and the ribosome footprints do not span from start to stop codon. Therefore a full-length protein cannot be produced. The black arrow shows the direction of transcription.

required time-consuming visual inspection, but resulted in revision of ~400 *T. brucei* start codons, about half of which were moved 5′ (extending the CDS) and half 3′ (shortening the CDS) [15].

Ribosome footprints also mapped to some areas of the *T. brucei* genome that lacked annotated CDSs [11,12,15]. While some of these footprints are the result of repetitive sequences or gene fragments likely to be pseudogenes, others appear to be mapping to unidentified (but *bona fide*) protein-coding genes. Therefore, we visually inspected (using Artemis) hundreds of candidate new CDSs, nominated by virtue of mapped ribosome footprints [11,12] or because of cross-species conservation [23]. Our criteria for designating new CDSs were that the ORF needed to be at least 75 nt (25 aa) in length, on a contiguous transcript, with ribosome footprints that spanned the entire ORF (and that were not attributable to an overlapping CDS), and that ended at the stop codon. These criteria allowed us to identify 225 new CDSs [15], 88% of which were not closely related to known *T. brucei* genes. The median size of the unique new CDSs was only 74 aa. Almost one-third of these unique CDSs were found only in the *T. brucei* subspecies, suggesting that they are young genes, evolving since the separation of different species of African trypanosomes.

chromosome 7 where these genes are located is shown at the bottom. (C) Trypanosomatid mRNA structure. Trypanosomatid genes are organized into clusters with adjacent protein-coding sequences (CDSs, shown as yellow boxes) that are transcribed as a polycistronic precursor RNA, which is processed by *trans*-splicing and polyadenylation to produce mature mRNAs. Each mature mRNA contains 5′- and 3′-untranslated regions (UTRs) that flank the CDS, as well as spliced leader (SL) and polyadenylation (polyA) sites. The SL is donated by a distinct RNA.

Trypanosomatids: Managing a Complex Life Cycle Without Transcriptional Control

Many protozoan pathogens undergo a life cycle during which they transit through distinct environments provided by mammalian and insect hosts (Figure 3). *T. brucei* is extracellular throughout its life cycle: living as slender **bloodstream forms** (BF) in the mammalian host, where the environment is homeostatically maintained at 37°C with abundant glucose and robust adaptive and innate immune surveillance; and as **procyclic forms** (PF) in the tsetse fly midgut, where glucose levels fluctuate and amino acids serve as the primary energy source, and there is a more circumscribed innate immune system. These proliferating stages are separated by non-proliferating forms that are poised for transmission to the alternative host – in *T. brucei* these stages include stumpy BF, which are derived from slender BF in the mammalian host and develop into PF when ingested by the tsetse fly, and metacyclic stages in salivary glands of the tsetse that develop into slender BF when inoculated into the mammal (Figure 3A). Similarly, *Trypanosoma cruzi* and *Leishmania* species alternate between hosts with dividing and

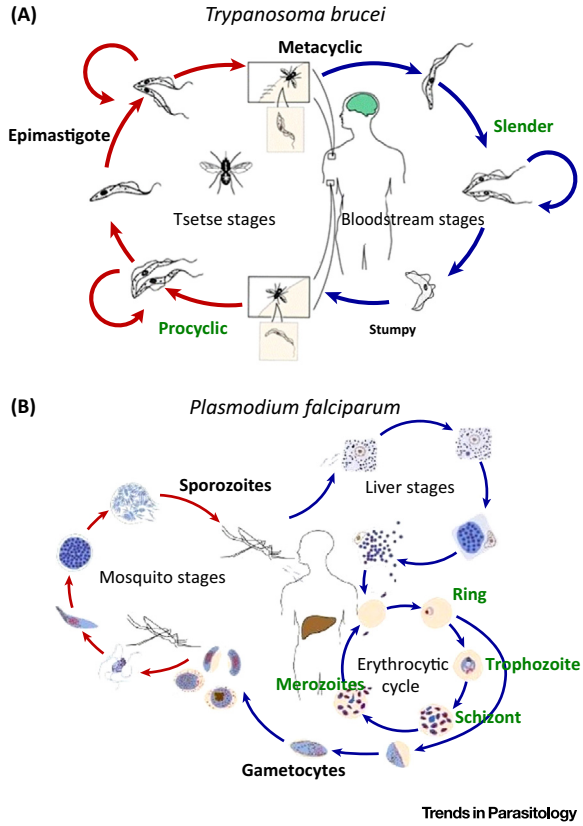

Figure 3. Parasite Life Cycles. *Trypanosoma brucei* and *Plasmodium falciparum* undergo development in both insect and mammalian hosts as shown. The stages that have been examined by ribosome profiling are identified by text in green font. (A) *Trypanosoma brucei* life cycle. Metacyclic forms are transmitted from the tsetse fly to the mammalian host, where they develop into long, slender bloodstream forms that actively divide (as indicated by the curved arrow). Occasionally, parasites will exit the cell cycle and develop into short stumpy bloodstream forms, which can be transmitted to the tsetse fly. There, they become dividing procyclic forms in the fly midgut. These develop into epimastigotes which divide in the salivary gland, ultimately yielding the transmissible metacyclic forms. (B) *Plasmodium falciparum* life cycle. The bite of an infected mosquito injects sporozoites into the human host, where they migrate to the liver. There, a single sporozoite will yield thousands of progeny that subsequently enter red blood cells. The asexual erythrocytic cycle allows continuous parasite replication via an ordered sequence of development from ring stages to merozoites. Occasional parasites exit the asexual erythrocytic cycle to become gametocytes that are then transmissible to the mosquito. In the mosquito, mating, cell division, and development occur to yield numerous sporozoites. Figures were adapted from the Public Health Image Library of the US Centers for Disease Control and Prevention (www.cdc.gov).

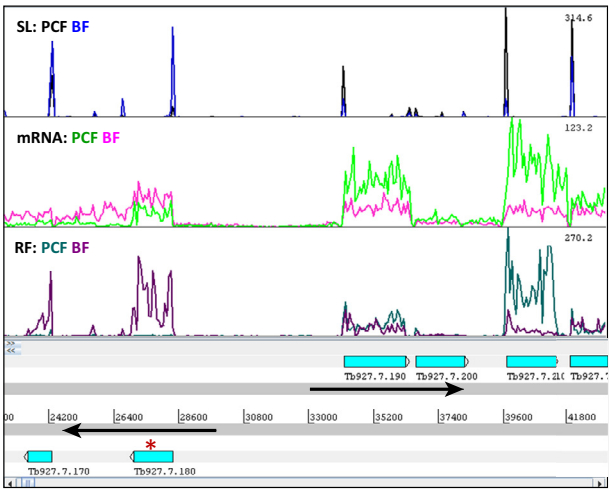

Figure 4. Differential Abundance of mRNA and Ribosome Footprints for Adjacent Genes in Polycistronic Units. Panels are as described in Figure 2. Shown are the first genes in two divergent polycistronic transcription units, which are separated by a region with very low levels of stable transcripts. The direction of transcription is marked by arrows. The asterisk marks a coding sequence (CDS) encoding a conserved hypothetical protein that shows different translation efficiency (TE) in procyclic cultured forms (PCF) and bloodstream forms (BF).

non-dividing stages, but are intracellular in the mammal host and extracellular in the insect. To accommodate these widely differing environments with respect to temperature, nutrients, and immune interactions, numerous differences in protein abundance between developmental stages have been identified [30–37]. Indeed, proteomic analyses suggest that as many as one-third of all proteins show at least a twofold difference in abundance in *T. brucei* cultured BF versus cultured PF (PCF) [30,32]. A key question that has interested the field is how these changes in protein expression are controlled.

In most eukaryotes, including many other protozoan pathogens such as *Plasmodium* spp., transcriptional control is a major means by which changes in expression of individual proteins are regulated across conditions. However, it is by no means the only mechanism involved because differential mRNA turnover, protein production, protein modification, and protein turnover all play roles depending on the gene and condition in question. Surprisingly to many experts on eukaryotic gene regulation, trypanosomatids have almost completely dispensed with transcriptional regulation of gene expression – although this is not unexpected given their polycistronic arrangement of genes. As shown in Figure 4, different transcripts within a polycistronic unit vary in abundance and show differential expression between stages. In large part this is due to differential mRNA stability, although *trans*-splicing efficiency also plays a role. Numerous studies of individual genes show that, in trypanosomatids, the 3′-UTR plays the predominant role in modulating mRNA abundance across stages [38] and in response to conditions such as heat shock [39]. Until recently the contribution of differential translation to regulating gene expression during trypanosomatid life cycles has been examined for only a small number of genes at the individual level, with a few being identified as differentially translated during development [40–45]. However, metabolic labeling of proteins shows that protein synthesis is globally decreased in stumpy BF, accompanied by a collapse of higher-order polysomes [46,47]. Similarly, *Leishmania* amastigotes show a global downregulation of translation [48] and more lower-order polysomes [49] relative to the more rapidly-dividing promastigotes.

What Ribosome Profiling Tells Us about Gene Expression in Parasites

To date, ribosome profiling of parasites in different states/conditions has been restricted to comparisons of developmental stages. In *T. brucei*, animal-derived [11] and cultured BF [12] have been compared to PCF, whereas in *T. cruzi* dividing epimastigotes have been compared to non-dividing metacyclic trypomastigotes (both are cultured insect stages) [13]. In *P. falciparum*,

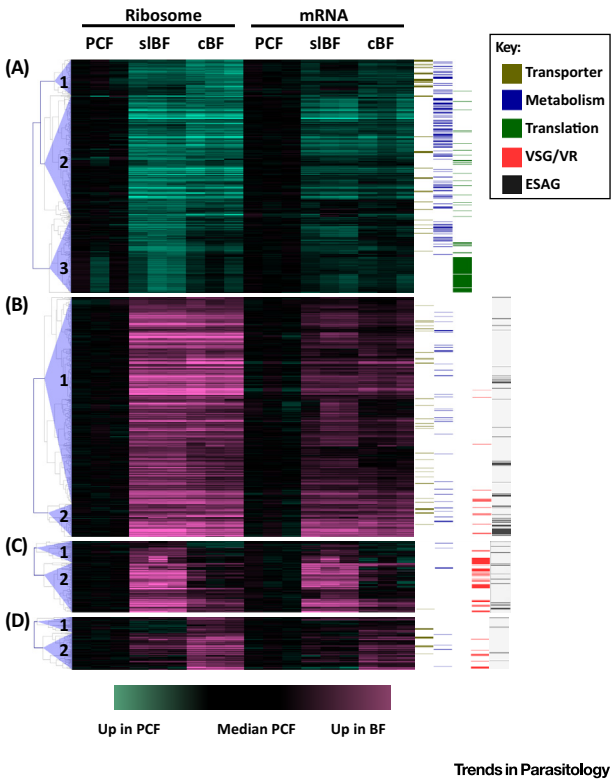

Figure 5. Genes Encoding Translation Machinery in *Trypanosoma brucei* Show Stage-Regulated Protein Production. Clustering analysis groups genes with similar expression patterns, in this case for ribosome footprints (protein production) and mRNA abundance. The 1557 genes showing at least a fourfold difference in ribosome footprint normalized read counts were clustered by K-means and HCL using MEV (for details see [11]). Biological samples included three replicates of PCF, animal-derived slender BF (slBF), and cultured BF (cBF). Each gene is represented by a separate line in the cluster diagram with the magnitude of the fold-change (relative to PCF median) indicated by color intensity (magenta for up and green for down). Bars at right indicate genes within each cluster that can be classified into the listed functional groups. Cluster A3 is highly enriched for proteins involved in translation, particularly structural components of the cytosolic ribosome. Reproduced from [11]. Abbreviations: ESAG, expression site-associated gene; HCL, hierarchical clustering; MEV, multiexperiment viewer; PCF, procyclic cultured form; VR, VSG-related; VSG, variant surface glycoprotein.

the entire asexual erythrocytic cycle (from ring forms that arise shortly after invasion of red blood cells through to merozoites shortly after egress; Figure 3B) has been examined [14].

Protein production is the product of two factors: the mRNA abundance and the TE of that mRNA. In trypanosomatids, changes in protein production are often more extreme than changes in mRNA abundance alone. For example, gene Tb927.7.180 had only 13-fold more mRNA in BF versus PCF, but 70-fold more protein production (Figure 4, gene marked by asterisk) [11]. Indeed, in *T. brucei*, about twice as many genes were detected as stage-regulated between *in vivo*-derived slender BF and PCF by measuring protein production compared to those identified as stage-regulated in terms of mRNA abundance [11]. This indicates that the TE of many genes changes during development. Figure 5 shows a clustering analysis of all *T. brucei* CDSs showing at least a four fold change in protein production between slender BF and PCF as measured by ribosome footprints [11]. This analysis allowed CDSs with similar expression patterns to be grouped together; for example CDSs in the A clusters are upregulated for both mRNA and protein production in PCF. As indicated by the greater color intensity on the ribosome footprint tracks, it is clear that the changes are greater for protein production than for mRNA abundance. Indeed, analyses using generalized linear models indicate that ~1800 of the 2600 genes with a statistically significant (false discovery rate, FDR = $\leq$0.01) two fold or more change in protein production (as measured by ribosome footprints) showed at least some regulation due to changes in TE [11]. However, very few genes appeared to be regulated by changes in TE alone, although this conclusion may be confounded by the complex relationship of translation and mRNA stability [50]. In a study of *T. cruzi* epimastigotes and metacyclic trypomastigotes [13], ~2200 CDSs were translated in epimastigotes, but not in metacyclic trypomastigotes, while 526 CDSs were exclusively translated in metacyclic forms by the same measure (>15 normalized reads). This latter group was enriched for families of surface proteins such as trans-sialidases, gp63 protease,

and MASP proteins that are expressed on metacyclic forms. These data contrast with the analysis of mRNA from the same samples, which showed that more than 95% of the detected transcripts were observed in both stages. Interestingly, ribosome footprints showed a better correlation with proteomic data than did mRNA abundance – but the correlation was much better in epimastigotes than in metacyclic trypomastigotes [13]. Such data suggest that additional levels of control, such as different rates of protein degradation, are particularly important in modulating protein expression in the latter stage. Unfortunately, proteome-wide data on protein turnover in *T. cruzi* or other parasites are not available, although most proteins appear to be relatively stable in *T. brucei* procyclic forms [51]. Perhaps surprisingly, we and others found no evidence that uORFs were associated with stage-regulation of TE [11,12].

Analysis of the functional categories of genes that have a significant component of translational regulation in *T. brucei* revealed that structural proteins of the cytoplasmic ribosome (a group of >150 proteins) are predominantly co-regulated via translational control (Figure 5, cluster A3) [11]. These proteins were more efficiently produced in log-phase PCF than in BF (from infected rats or cultures). Because the strain of *T. brucei* employed for *in vivo* studies (TREU 927) differentiates from slender BF to non-dividing stumpy BF, we suggested that the substantially reduced TE of ribosomal proteins seen in the rat-derived BF may reflect an early marker of commitment to the process of growth arrest in this population. The recent ribosome profiling study in *T. cruzi* extended this concept by directly comparing dividing (epimastigotes) and non-dividing (metacyclic trypomastigotes) insect stages [13]. Ribosomal proteins had lower TEs in the non-dividing metacyclic stage. For both species, co-regulation with some translation factors was also observed. While ribosomal protein mRNAs are highly abundant, their TEs are relatively low and (as indicated above) highly sensitive to growth conditions. Because translation is limited by the abundance of ribosomes rather than mRNA [52], reducing the TE of ribosomal protein mRNAs may serve as a rapid means to curtail not only the production of proteins that themselves require a large fraction of the cell's translational capacity, but also to decrease protein production as a whole.

Translational regulation of ribosomal proteins is also seen in mammalian cells subjected to stresses such as starvation. There, this process is mediated by an oligopyrimidine tract immediately after the mRNA cap [53]. However, the sequences mediating the changes in TE must be at a different location in trypanosomes because their mRNAs share a common 5′-terminus (the SL). Indeed, it may be that these sequence elements are not in the 5′-UTR because several of the trypanosome ribosomal protein mRNAs have 5′-UTRs shorter than 10 nt following the SL. In other systems, RNA-binding proteins have been shown to regulate translation by binding to the 3′-UTR [54], and it would be interesting to pursue this possibility for ribosomal protein mRNAs in trypanosomes.

Translational regulation has also been shown to be important at several steps of the life cycle of *Plasmodium* spp., including the asexual erythrocytic stages [55]. For example, mRNA abundance for many genes peaks before their recruitment to polysomes [56] and the subsequent increase in protein abundance [57,58]. Furthermore, the translational downregulator PK4 (an eIF2α kinase) was shown to be essential for the asexual cycle [59,60]. It is therefore somewhat surprising that ribosome profiling from *P. falciparum* revealed a different picture [14], with the authors indicating that regulation of protein production during the asexual erythrocytic cycle is driven almost exclusively by mRNA abundance. It should be noted that the authors used the term 'translational control' to describe the large range in TE for different genes within each life cycle stage (as described below) rather than changes in TE of a given gene across stages. Nonetheless, they identified a small set of genes (~64), mostly encoding proteins involved in merozoite egress and invasion, that have peak mRNA abundance in late schizonts or merozoites, but higher TE in ring stages. Recently, it has been shown that the RNA binding protein

ALBA1 associates with many of these mRNAs and regulates their translation [61]. Finally, ribosome profiling has not yet been applied to the sporozoite (mosquito salivary gland) or gametocyte stages of *Plasmodium*, where there is striking evidence of translational repression of specific mRNAs [55]. Perhaps the more predictable asexual cycle employs translational regulation as a tool to fine-tune gene expression programmed by mRNA abundance, whereas the unpredictable transitions between host and vector forms need layers of control of specific pre-existing mRNAs that are both rapid and robust.

Insights into Translation and Gene Structure from Ribosome Profiling

A recent study in mammalian cells used pulse-labeling, mass spectrometry, and RNA analysis to measure the contribution of mRNA and protein synthesis and turnover to the steady state abundance of different proteins [62]. It was noteworthy that translation rates had a large effect on the ultimate protein abundance, comparable to the effect of transcription rates. This fits well with the ribosome profiling data from eukaryotes such as yeast [7], mammalian cells [9], and plants [63], which show that the TE of different genes varies widely. Similarly, TEs vary ~70-fold in BF and ~130-fold in PCF of *T. brucei* [11,12] and 100- to 400-fold in *P. falciparum* (depending on the stage) [14], and this likely contributes significantly to the ultimate concentration of individual proteins. While read counts for mRNA and ribosome footprints showed a strong correlation, TEs were not correlated with RNA abundance in *T. brucei* [12].

The mechanisms and sequences that yield such varying TEs across different genes in a single condition have not been comprehensively analyzed in parasites. However, in other eukaryotes, translation of uORFs can decrease the translation of the main CDS because essential translation initiation factors are rapidly lost from the ribosome once translation commences, and therefore reinitiation cannot occur [64]. The two studies of *T. brucei* agree that the main CDSs on transcripts with uORFs tend to be more poorly translated than those without uORFs [11,12]. However, they found different proportions of genes bearing uORFs – 22% of 4909 genes [12] versus 11% of 7331 genes [11] – likely based on the different sets of 5′-end mapping data used for each study. Indeed, it still remains possible that some of these uORFs may actually be on separate transcripts because we observed that less-abundant mRNAs (where mapping of 5′-ends is more challenging) were enriched for uORFs [11]. Thus, it will take additional work to accurately define the subset of CDSs that are preceded by uORFs and to more fully evaluate their role. In addition to careful analysis of contiguity of transcripts, the possibility that non-canonical start codons initiate translation of uORFs, as seen in yeast and mammalian cells [7,9,65], has not been evaluated in *T. brucei*. Such data could help to reveal situations, such as specific stresses, where the uORFs of some genes may regulate translation.

In contrast to *T. brucei*, uORFs were not implicated in the regulation of protein production in *P. falciparum* because there was no difference in the TEs of transcripts with uORFs versus those that lack them [14]. However, it should be noted that *P. falciparum* 5′-UTRs are much longer, with a median length of 736 nt in late trophozoites [14], compared to a median of only 87 nt (excluding the SL) in *T. brucei* [11]. Nonetheless, *var2csa*, a gene that is expressed in placental *P. falciparum*, may be an exception. It has several uORFs [66], and ribosome footprints in the non-placental parasites studied were concentrated over two of these, and were absent from the main CDS [14]. It would be interesting to learn whether parasites expressing *var2csa* show decreased ribosome footprints on these uORFs. If so, this could represent a case of regulation by modulation of uORF translation.

It is worth noting that ribosome footprints can be almost always found at the first ATG on *T. brucei* transcripts, although is not yet clear whether the corresponding uORFs actually produce polypeptides. For example, Figure 6 shows a transcript (Tb7.NT.1) with only a small number of ribosome footprints that localize around the ATG codons of two short ORFs (of 9 and

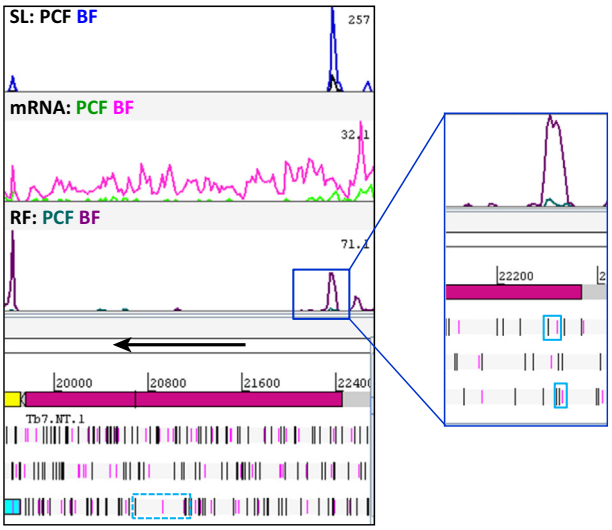

Figure 6. A Transcript with Ribosome Footprints Mapping to the 5′-End While Larger Downstream Open Reading Frames (ORFs) Are Not Translated. This transcript (Tb7.NT.1) has two small ORFs (9 and 27 nt) immediately following the spliced leader (SL) site that are spanned by ribosome footprints (RF) (see enlargement, turquoise boxes). However, downstream ORFs (including one of 468 nt, indicated by the broken box), do not bear ribosome footprints. Black bars represent stop codons and pink bars, start codons. Traces are color-coded as indicated. Abbreviations: BF, bloodstream form; PCF, procyclic cultured form.

Outstanding Questions

What are the features of mRNAs that confer efficient translation of parasite mRNAs?

Do specific uORFs play a role in regulating translation as conditions change?

Which mechanisms regulate translation efficiency across development?

Do stress conditions cause changes in translation efficiency? Do these changes vary according to the stress?

Do drugs and potential therapeutics affect translational regulation?

27 nt) near the 5′-end [24]. The footprints stop after these two ORFs, and the downstream ORFs (including one that would specify a 156 aa protein) appear to be translationally inactive. One possibility is that the function of such very short ORFs is to block translation of downstream ORFs, allowing the sequence to functionally evolve without the danger of producing toxic proteins.

Interestingly, unlike 3′-UTRs (which rarely have ribosome footprints), 5′-UTRs lacking uORFs commonly have ribosome footprints (e.g., Tb927.7.7360 in Figure 2). This is true even when ribosomes were further purified on sucrose gradients and libraries constructed from the 80S fraction (a protocol that should reduce scanning 40S subunits) [7,12,14]. These data suggest that the footprints in the 5′-UTRs are due to assembled ribosomes, even when canonical start codons are not present. Studies of mammalian cells indicate that the size of the footprints in the 5′-UTR matches exactly those of CDSs translated in the cytoplasm, but differ from those translated in mitochondria [65], further suggesting that they reflect assembled ribosomes. Interestingly, in *P. falciparum*, these footprints do not show high preference for mapping to uORFs [14]. In *T. brucei* we saw that when the 5′-UTR lacked a uORF, footprints continued up to the main CDS, regardless of the presence of stop codons [11]. Because functional monosomes are thought to assemble at the start codon, it is possible that such footprints result from non-productively assembled 40 and 60S subunits or ribosomes assembled through a novel pathway. The moderate correlation of the abundance of ribosome footprints in the 5′-UTR with their abundance over the CDS, as was observed in *T. brucei* [11] and in *P. falciparum* [14], could imply a mechanistic connection, such as recycling of ribosomes via the circularization of the mRNA that occurs through the interaction of polyA binding protein and the 5′ cap.

Concluding Remarks

At present many areas concerning translation in protozoan parasites remain to be explored (see Outstanding Questions). Data generated in *E. coli* and yeast [7,67] using the current protocol (which was designed to minimize sequence biases) show that the abundance of ribosome footprints on different CDSs accurately reflects the relative production of the corresponding proteins [4]. If this holds true in parasites, one could calculate the relative production of all proteins on a per molecule basis and determine how much translational capacity each process requires under different conditions. Such data could inform rational selection of drug targets. For

example, ribosome profiling coupled with kinetic data in *E. coli* revealed that MetE (a large and inefficient enzyme) is a bottleneck for methionine biosynthesis [4]. Furthermore, the data predicted that the fitness cost of synthesizing more enzyme was precisely balanced with the requirement for methionine – which was confirmed by overexpression studies. Such an enzyme might make an excellent drug target because it is both rate-limiting and overexpression is disadvantageous.

Expansion of this approach into other organisms and conditions would undoubtedly yield new insights into the systems biology of parasites. For example, the effects of drugs on translation remain relatively unexplored, and could involve differential effects on different CDSs. Similarly, stresses such as heat shock (such as seen in fever) or rapid changes in nutrients (as occur during transmission) could evoke alterations in protein production. Many stages of parasite development remain poorly explored, and some parasites have not yet been studied using this technology. A key roadblock could be the amount of material required to make unbiased libraries that allows comparison between different genes in the same sample. If the question is restricted to analysis of changes in protein production for a gene as conditions change, it may be possible to streamline the protocol and reduce the amount of biological sample required.

Acknowledgments

The authors thank Drs Bryan Jensen, Gowthaman Ramasamy, Elton Vasconcelos, and Nicholas Ingolia for their advice and contributions to our ribosome profiling studies. This work was supported by grants R21AI105769 and R21AI094129 from the National Institutes of Health. The authors are solely responsible for the content.

References

1. Jackson, R. and Standart, N. (2015) The awesome power of ribosome profiling. *RNA* 21, 652–654
2. Brar, G.A. and Weissman, J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664
3. Stern-Ginossar, N. *et al.* (2012) Decoding human cytomegalovirus. *Science* 338, 1088–1093
4. Li, G.W. *et al.* (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635
5. Guydosh, N.R. and Green, R. (2014) Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* 156, 950–962
6. Kannan, K. *et al.* (2014) The general mode of translation inhibition by macrolide antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15958–15963
7. Ingolia, N.T. *et al.* (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
8. Michel, A.M. *et al.* (2012) Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* 22, 2219–2229
9. Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
10. Rutkowski, A.J. *et al.* (2015) Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* 6, 7126
11. Jensen, B.C. *et al.* (2014) Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* 15, 911
12. Vasquez, J.J. *et al.* (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* 42, 3623–3637
13. Smircich, P. *et al.* (2015) Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics* 16, 443
14. Caro, F. *et al.* (2014) Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *Elife* 3, e04106
15. Parsons, M. *et al.* (2015) Advancing *Trypanosoma brucei* genome annotation through ribosome profiling and spliced leader mapping. *Mol. Biochem. Parasitol.* 202, 1–10
16. Arias, C. *et al.* (2014) KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* 10, e1003847
17. Berriman, M. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309, 416–422
18. El-Sayed, N.M. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415
19. Ivens, A.C. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309, 436–442
20. Martinez-Calvillo, S. *et al.* (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell* 11, 1291–1299
21. El-Sayed, N.M. *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309, 404–409
22. Aslett, M. *et al.* (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 38, D457–D462
23. Ericson, M. *et al.* (2014) On the extent and role of the small proteome in the parasitic eukaryote *Trypanosoma brucei*. *BMC Biol.* 12, 14
24. Kolev, N.G. *et al.* (2010) The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog.* 6, e1001090
25. Siegel, T.N. *et al.* (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* 38, 4946–4957
26. Nilsson, D. *et al.* (2010) Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathog.* 6, e1001037
27. Carver, T. *et al.* (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469
28. Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511

29. Aurrecoechea, C. *et al.* (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic. Acids Res.* 37, D539–D543

30. Butter, F. *et al.* (2013) Comparative proteomics of two life cycle stages of stable isotope-labeled *Trypanosoma brucei* reveals novel components of the parasite's host adaptation machinery. *Mol. Cell. Proteomics* 12, 173–179

31. Urbaniak, M.D. *et al.* (2013) Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*. *J. Proteome. Res.* 12, 2233–2244

32. Urbaniak, M.D. *et al.* (2012) Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages. *PLoS ONE* 7, e36619

33. Tsigankov, P. *et al.* (2014) Regulation dynamics of *Leishmania* differentiation: deconvoluting signals and identifying phosphorylation trends. *Mol. Cell. Proteomics* 13, 1787–1799

34. Paape, D. and Aebischer, T. (2011) Contribution of proteomics of *Leishmania* spp. to the understanding of differentiation, drug resistance mechanisms, vaccine and drug development. *J. Proteomics* 74, 1614–1624

35. Queiroz, R.M. *et al.* (2014) Quantitative proteomic and phosphoproteomic analysis of *Trypanosoma cruzi* amastigogenesis. *Mol. Cell. Proteomics* 13, 3457–3472

36. de Godoy, L.M. *et al.* (2012) Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *Proteomics* 12, 2694–2703

37. Brunoro, G.V. *et al.* (2015) Reevaluating the *Trypanosoma cruzi* proteomic map: The shotgun description of bloodstream trypomastigotes. *J. Proteomics* 115, 58–65

38. Clayton, C. and Shapira, M. (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol. Biochem. Parasitol.* 156, 93–101

39. Schwede, A. *et al.* (2012) How do trypanosomes change gene expression in response to the environment? *Protoplasma* 249, 223–238

40. Gale, M.J., Jr *et al.* (1994) Translational control mediates the developmental regulation of the *Trypanosoma brucei* Nrk protein kinase. *J. Biol. Chem.* 269, 31659–31665

41. Parsons, M. and Hill, T. (1989) Elevated phosphoglycerate kinase mRNA but not protein in monomorphic *Trypanosoma brucei*: implications for stage-regulation and post-transcriptional control. *Mol. Biochem. Parasitol.* 33, 215–228

42. Hotz, H.R. *et al.* (1997) Mechanisms of developmental regulation in *Trypanosoma brucei*: A polypyrimidine tract in the 3′-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res.* 25, 3017–3025

43. Furger, A. *et al.* (1997) Elements in the 3′ untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. *Mol. Cell. Biol.* 17, 4372–4380

44. Boucher, N. *et al.* (2002) A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3′ UTR element. *J. Biol. Chem.* 277, 19511–19520

45. David, M. *et al.* (2010) Preferential translation of Hsp83 in *Leishmania* requires a thermosensitive polypyrimidine-rich element in the 3′ UTR and involves scanning of the 5′ UTR. *RNA* 16, 364–374

46. Brecht, M. and Parsons, M. (1998) Changes in polysome profiles accompany trypanosome development. *Mol. Biochem. Parasitol.* 97, 189–198

47. Capewell, P. *et al.* (2013) Regulation of *Trypanosoma brucei* total and polysomal mRNA during development within its mammalian host. *PLoS ONE* 8, e67069

48. Cloutier, S. *et al.* (2012) Translational control through eIF2alpha phosphorylation during the *Leishmania* differentiation process. *PLoS ONE* 7, e35085

49. Lahav, T. *et al.* (2011) Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *FASEB J.* 25, 515–525

50. Roy, B. and Jacobson, A. (2013) The intimate relationships of mRNA decay and translation. *Trends Genet.* 29, 691–699

51. Clayton, C.E. (1988) Most proteins, including fructose bisphosphate aldolase, are stable in the procyclic trypomastigote form of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 28, 43–46

52. Merrick, W.C. (2010) Eukaryotic protein synthesis: still a mystery. *J. Biol. Chem.* 285, 21197–21201

53. Hamilton, T.L. *et al.* (2006) TOPs and their regulation. *Biochem. Soc. Trans.* 34, 12–16

54. Szostak, E. and Gebauer, F. (2013) Translational control by 3′-UTR-binding proteins. *Brief. Funct. Genomics* 12, 58–65

55. Zhang, M. *et al.* (2012) Translational control in *Plasmodium* and *Toxoplasma* parasites. *Eukaryot Cell* 12, 161–167

56. Bunnik, E.M. *et al.* (2013) Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome Biol.* 14, R128

57. Le Roch, K.G. *et al.* (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res.* 14, 2308–2318

58. Foth, B.J. *et al.* (2011) Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *Plasmodium falciparum*. *Mol. Cell Proteomics* 10, M110

59. Solyakov, L. *et al.* (2011) Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*. *Nat. Commun.* 2, 565

60. Zhang, M. *et al.* (2012) PK4, a eukaryotic initiation factor 2alpha (eIF2alpha) kinase, is essential for the development of the erythrocytic cycle of *Plasmodium*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3956–3961

61. Vembar, S.S. *et al.* (2015) The PfAlba1 RNA-binding protein is an important regulator of translational timing in *Plasmodium falciparum* blood stages. *Genome. Biol.* 16, 212

62. Schwanhausser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342

63. Liu, M.J. *et al.* (2012) Widespread translational control contributes to the regulation of *Arabidopsis* photomorphogenesis. *Mol. Syst. Biol.* 8, 566

64. Barbosa, C. *et al.* (2013) Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* 9, e1003529

65. Ingolia, N.T. *et al.* (2014) Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 8, 1365–1379

66. Amulic, B. *et al.* (2009) An upstream open reading frame controls translation of var2csa, a gene implicated in placental malaria. *PLoS Pathog.* 5, e1000256

67. Ingolia, N.T. *et al.* (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550