

Control of gene expression in Trypanosomatidae

S.M.R. Teixeira

Departamento de Bioquímica e Imunologia,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Abstract

The study of mechanisms which control gene expression in trypanosomatids has developed at an increasing rate since 1989 when the first successful DNA transfection experiments were reported. Using primarily *Trypanosoma brucei* as a model, several groups have begun to elucidate the basic control mechanisms and to define the cellular factors involved in mRNA transcription, processing and translation in these parasites. This review focuses on the most recent studies regarding a subset of genes that are expressed differentially during the life cycle of three groups of parasites. In addition to *T. brucei*, I will address studies on gene regulation in a few species of *Leishmania* and the results obtained by a much more limited group of laboratories studying gene expression in *Trypanosoma cruzi*. It is becoming evident that the regulatory strategies chosen by different species of trypanosomatids are not similar, and that for these very successful parasites it is probably advantageous to employ multiple mechanisms simultaneously. In addition, with the increasing numbers of parasite genes that have now been submitted to molecular dissection, it is also becoming evident that, among the various strategies for gene expression control, there is a predominance of regulatory pathways acting at the post-transcriptional level.

Key words

- *Trypanosoma*
- *Leishmania*
- Gene expression
- RNA
- Antigenic variation
- VSG

Correspondence

S.M.R. Teixeira
Departamento de Bioquímica e
Imunologia, ICB, UFMG
Av. Antonio Carlos, 6627
30161-970 Belo Horizonte, MG
Brasil
Fax: +55-31-441-5963
E-mail: santuzat@icb.ufmg.br

Research supported by CNPq
and FAPEMIG.

Received May 26, 1998
Accepted August 26, 1998

Introduction: general mechanisms of gene expression in trypanosomatids

In the Trypanosomatidae family, the genera *Trypanosoma* and *Leishmania* consist of several species of protozoa that are unicellular and unflagellated. These organisms are digenetic parasites whose life cycles present multiple differentiation forms that alternate between invertebrate and vertebrate hosts. *T. brucei*, *T. cruzi* and a few species of *Leishmania* are the model systems used in the studies described here. They belong to an

ancient group of lower eukaryotes which, although we tend to consider them to be closely related organisms, are in fact highly divergent evolutionarily (1).

The medical relevance of this group of parasites continues to be evident according to recent World Health Organization surveys estimating that about 550 million people in the developing world are at risk of contracting leishmaniasis, sleeping sickness or Chagas' disease (The World Health Report, 1997, <http://www.who.ch>). As exemplified by the "Initiative of the Southern Cone Countries", which

resulted in drastic reduction of vectorial and transfusional transmission in five Latin American countries, prevention and control of these infectious diseases by standard methods of breaking chains of transmissions is still an attainable goal. Nevertheless, there continues to be a need for basic research to develop new prophylactic tools and to determine more specific targets for therapy.

In addition to being of medical importance, trypanosomes and *Leishmania* have attracted the attention of molecular parasitologists because of recent studies revealing peculiar aspects of their biology. Knowing that the Trypanosomatidae family diverged very early in the eukaryotic lineage, we should understand their “aberrant behaviors”: their rules for molecular and cellular biology are quite distinct when compared to higher eukaryotes. As most textbooks describe, in eukaryotes protein-coding genes are transcribed into monocistronic pre-mRNA transcripts containing exons (coding sequences) and introns (mostly non-coding sequences) that are processed into mature mRNAs through “*cis*-splicing” reactions. RNA polymerase II is the enzyme responsible for the transcription of protein-coding genes, whereas RNA polymerase I transcribes ribosomal RNA. In trypanosomatids, however, transcription is polycistronic, there are no introns and, therefore, no *cis*-splicing reactions. Processing of pre-mRNA into single-gene units is effected by *trans*-splicing reactions, a process that has been found to operate only in trypanosomatids, *Euglena* and in nematode and trematode worms (2). In addition to these differences, other cellular and molecular novelties were discovered in trypanosomes: i) during cell division, chromosome condensation and nuclear membrane disruption do not occur; ii) peculiar organelles are found, such as glycosomes which concentrate the enzymes and the intermediates of the glycolytic pathway; iii) an extensive post-transcriptional modification of mitochondrial RNA known as RNA editing is

required for the correct expression of mitochondrial enzymes, and, finally, iv) transcription of protein-coding genes can be achieved by RNA polymerase I (or an enzyme with similar properties). Because of these differences between trypanosomatids and higher eukaryotes, we shall first see how the general process of gene transcription and mRNA maturation takes place before studying how gene regulation occurs.

Since the primary transcripts of most protein-coding genes that have been studied so far are polycistronic, cleavage of the pre-mRNA has to occur in the nucleus in order to produce monocistronic mRNAs. Furthermore, for efficient translation by the ribosomes, the addition of a methylated G nucleotide or cap at the 5' end and the poly(A) tail at the 3' end of each mRNA is essential. In trypanosomatids, no consensus signal sequence for polyadenylation has been identified and the cap has a unique, highly methylated structure (3). In fact, the 5' cap is part of the 39-nucleotide miniexon (or spliced leader, SL) that is joined to the 5' end of every mRNA in a transesterification reaction called *trans*-splicing. Recent studies have provided several lines of evidence demonstrating that miniexon/cap addition and polyadenylation are not independent events, but, instead, are part of a “cut-and-paste” mechanism that occurs concomitantly or immediately after transcription (Figure 1). Coupling between *trans*-splicing and polyadenylation was first described by LeBowitz et al. (4) who demonstrated that poly(A) selection is governed by the location of the splice acceptor site of the downstream gene in the polycistronic primary transcript of the DHFR-TS locus of *Leishmania major*. Soon thereafter other groups showed that this was also the case for *hsp70*, procyclic acid repetitive protein (PARP) and tubulin genes in *T. brucei* (5-7). Since no consensus polyadenylation signal has been found in trypanosome genes, the model proposing the existence of a conserved distance between splice acceptor site

and the site for polyadenylation became accepted as a solution to this problem. Several experiments using deletion mutants have tested this hypothesis and have shown the importance of intergenic regions in governing *trans*-splicing and polyadenylation in *T. brucei*. In addition to the correct distance, the presence of a polypyrimidine-rich motif is also crucial, since only AG dinucleotides situated downstream from a polypyrimidine tract are used as splice acceptor site (Figure 1) (4-6,8).

Early transfection experiments of trypanosomes and *Leishmania* were performed in the absence of information regarding the sequences required for gene expression in these organisms (9). For a recent review on genetic transformation in protozoa, see ref-

erence 10. Based on the studies described above, the primary rules are now well established: in order to have a foreign gene expressed in trypanosomatids, SL addition sites must be present both upstream (for *trans*-splicing) and downstream (for polyadenylation) from the gene. On the other hand, the need for a promoter may depend on which trypanosomatid is involved. While in *T. brucei*, studies of genes encoding two abundant surface proteins, variant surface glycoprotein (VSG) and PARP (see below), have provided strong promoters which were used to create powerful expression vectors, in *T. cruzi* and some species of *Leishmania*, we and others have been very successful in obtaining expression of foreign genes without promoters (10).

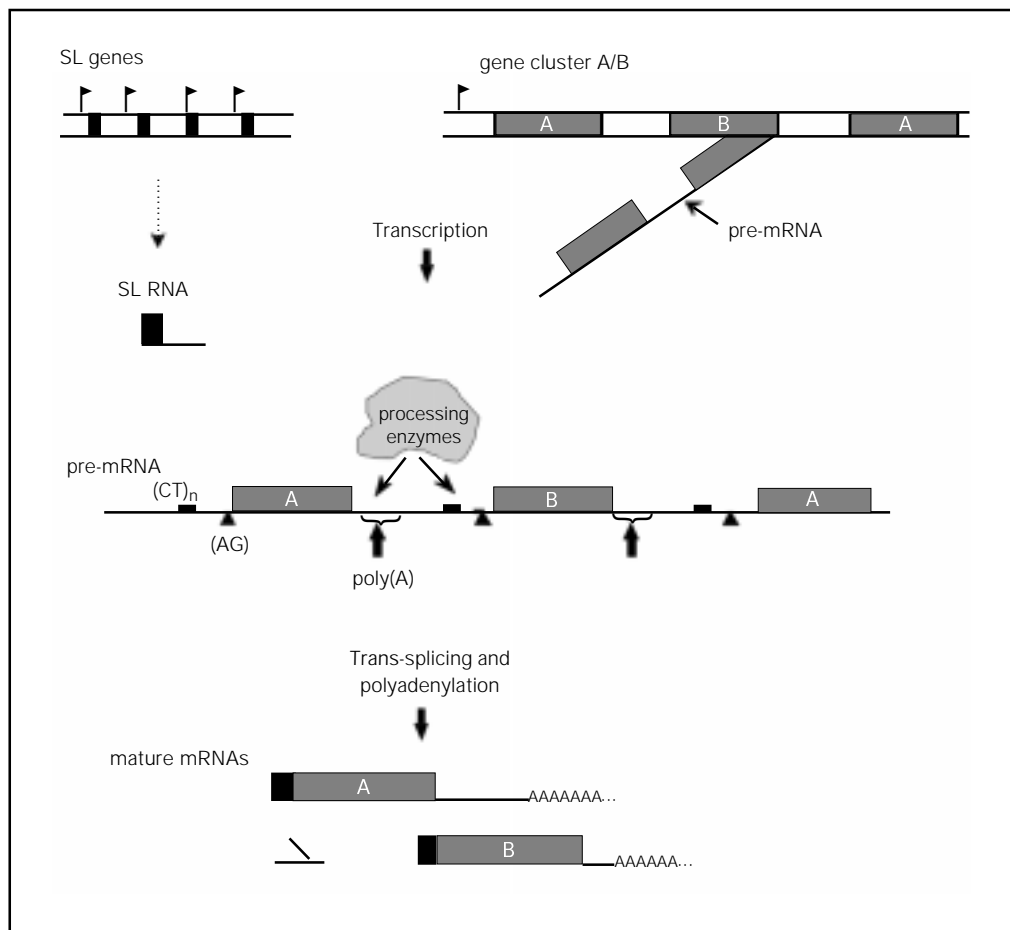


Figure 1 - The general process of mRNA synthesis in trypanosomatids. Most trypanosome and *Leishmania* genes characterized thus far are organized in tandem repeats as indicated by the shaded boxes (representing two copies of an A gene and one copy of a B gene). At another chromosomal location there are several hundred tandem direct repeats encoding SL RNAs (dark boxes). In some cases, a single promoter is present upstream of the first gene in the cluster whereas in the SL gene cluster each repeat is presumed to bear a transcriptional promoter (small flags). After transcription, the polycistronic pre-mRNA is processed by trans-splicing and polyadenylation, which require signals present within the intergenic regions: a polypyrimidine-rich sequence or (CT)_n (small boxes), the spliced leader addition site or AG (arrowheads), and a polyadenylation addition site or poly(A) (arrows). Mature mRNAs which consist of the coding region and 5'UTR (shaded boxes), the SL sequence (dark boxes), 3'UTR and poly(A) tail (thin lines with small "A" letters) are represented by the diagram at the bottom of the figure.

Antigenic variation as a model system to study gene expression control in *T. brucei*

Most of the studies on the molecular biology of trypanosomatids were conducted by parasitologists in the United States and in Europe who were curious about the process of antigenic variation. Antigenic variation, perhaps one of the most interesting recent discoveries in the parasitology field, is a powerful survival strategy devised by African trypanosomes which allows them to escape the immunological attack of the host. The *T. brucei* life cycle, unlike *T. cruzi* and *Leishmania*, does not have an intracellular stage. It has two developmental forms: procyclics, which multiply in the insect midgut, and bloodstream forms, which multiply in the mammalian host where they are able to survive because of antigenic variation.

Antigenic variation started to be understood at the molecular level about 20 years ago. However, the first observations pointing to this phenomenon were made in 1910 (11) when the number of parasites was determined in blood samples from patients with sleeping sickness: there was a dramatic rise and fall in the number of trypanosomes, with one peak of parasitemia followed by another peak every one to two weeks. In 1975, George Cross (12) isolated a surface protein known to be the single major protein present in the bloodstream forms of *T. brucei* and named it variant surface glycoprotein (VSG). When the N-terminal sequences of VSGs isolated from four cloned trypanosome populations derived from a single rabbit infected with *T. brucei* were compared, it became clear that each peak of parasitemia corresponded to a trypanosome population expressing a different VSG on their surface (13). The conclusion was obvious: after host antibodies against one type of VSG molecule are produced and the immune system destroys 99% of the trypanosome population, one of the few surviving parasites

changes the type of VSG it expresses and, with a new surface coat, begins reproducing as if it had never been seen by host immune defenses. This game continues to be played with the immune system always running one step behind the parasite. Each time a VSG switch occurs, 10^7 identical new proteins are added as a molecular shield to the surface of the parasite. Estimated switching rates as high as 10^{-2} cells per generation have been reported in recently transmitted trypanosome isolates (14). Recently, various aspects of antigenic variation have been reviewed (14, 15).

Among several questions involving VSG expression, two of them have fascinated trypanosome molecular biologists: 1) how does a trypanosome switch from the expression of VSG gene A to gene B, while keeping all the other VSG genes silent? and 2) why does VSG expression occur exclusively in bloodstream forms? The number of VSG genes in the diploid genome of *T. brucei* has been estimated to be approximately 1,000 (14,15). It was also shown that they are scattered throughout the hundreds of minichromosomes and about twelve pairs of large chromosomes found in *T. brucei* (14,15). The work of Williams and collaborators in 1979 (16) provided the first evidence of rearrangements of VSG genes in the parasite genome: a restriction fragment containing a VSG gene expressed in clone A may have a different size in the genome of a trypanosome than clone B, which does not express this particular VSG. With the cloning of the first VSG genes various investigators were able to compare their chromosomal environment and to identify some of the differences between active and silent VSG genes. These studies demonstrated that the expressed VSG gene is always located near a chromosome end, about 1-3 kb from the start of the repetitive hexamer present in the telomere (14,17) and, interestingly, VSG genes are not alone (Figure 2). Instead, they are part of a long polycistronic operon, called an expression site or

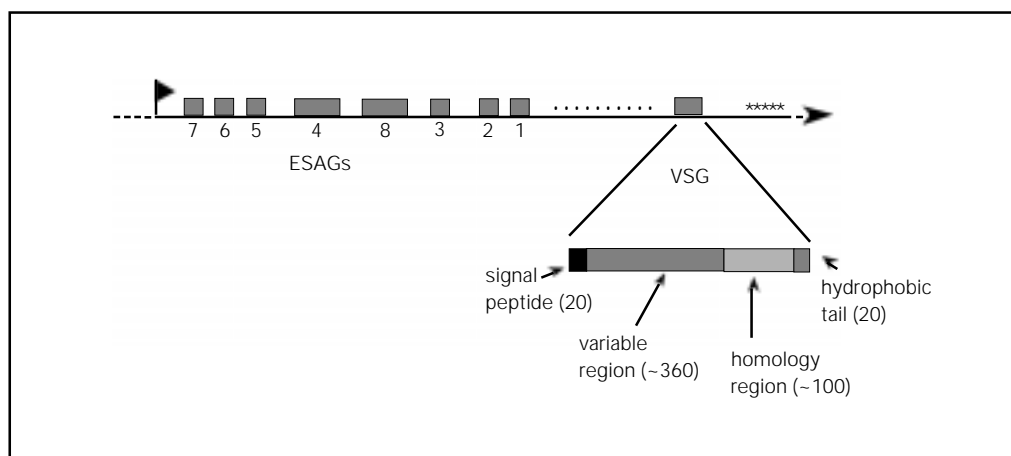
ES, that can extend for 50-60 kb and contains about 10 other genes (18-20). The products of these genes, called ESAG for expression site-associated genes, have no functional relationship to VSG, but an important function has been assigned to at least two of them: the *T. brucei* heterodimeric transferrin receptor is encoded by ESAG6 and ESAG7 (21). In every telomere that has been characterized so far, a resident VSG gene has been found. Since some VSG genes are also found internally in the chromosomes, some type of DNA rearrangement must occur (such as the ones originally described by Williams) in order to activate these internal genes. But because not all VSGs located in telomeres are being expressed in one cell, there must also exist some type of *in situ* gene activation/inactivation to differentiate these telomeric ES. So, it appears that a VSG gene can be turned on in two basic ways: either it is already in a telomeric ES that needs to be activated *in situ* (concomitantly with the inactivation of another ES), or it is an internal gene that needs to be copied and moved near a telomere where it can replace an active gene.

Understanding these mechanisms, particularly the *in situ* switch, has been a difficult task. A model to explain the activation/inactivation of one particular ES in bloodstream trypanosomes, comparable to silencing mechanisms described in yeast, proposes

a significant role for telomeric sequences (22). Moreover, the role of a newly identified nucleotide β -D-glucosyl-hydroxymethyluracil, called "J", that is found predominantly in telomeric repeats near silent VSG genes, is also being investigated (17).

The cloning of promoter regions from active VSG genes, which is crucial for the understanding of the mechanism of VSG regulation, was not an easy experiment: the region researchers were looking for was in fact located 60 kb upstream of a VSG gene (14,15,23). Before a VSG promoter was cloned, the work from Borst's group (23) provided the first evidence for the presence of a distant single promoter transcribing several genes in what is now known as a bloodstream form VSG transcription unit (Figure 2). More recently, it has been found that in a special class of VSG genes, i.e., those that are expressed by metacyclic trypanosomes, the promoter is located immediately upstream of the VSG gene, resulting in an unusual trypanosome monocistronic mRNA (24). After sequencing several VSG promoters, another unusual discovery was made: they do not possess any of the typical sequences present in eukaryotic RNA pol II promoters, such as the TATA box. Moreover, nuclear run-on transcription assays showed that VSG gene promoters are functionally more related to RNA pol I promoters since they are 100% active in the presence of 1 mg/ml α -

Figure 2 - The *T. brucei* VSG molecule and a VSG gene expression site. Studies of several vsg loci indicate that expression sites of bloodstream forms are organized similarly to the schematic representation shown, in which shaded boxes represent one VSG gene and several ESAGs. The flag indicates the promoter, the numbers below the boxes define each ESAG, the small dots and asterisks represent the 70-bp (or 76-bp) repeats and telomeric hexanucleotide repeats, respectively, and the arrowhead indicates the chromosome terminus. The distance between the VSG gene and the start of the repetitive telomeric sequence is 1.3 kb. The enlargement below features a diagram with the different regions of a nascent VSG (numbers in parentheses indicate the number of amino acids in each portion of the molecule).



amanitin (3). Thus, it appears that *T. brucei* has a unique system that allows transcription of protein-coding genes to be initiated by RNA polymerase I or an RNA polymerase I-like enzyme. This is only possible in trypanosomes because maturation of the 5' end (and the addition of cap) is independent of the RNA polymerase. Contrary to the general process of eukaryote transcription, where the enzymes required for the capping process are thought to be associated with RNA polymerase II, in trypanosomes the *trans*-splicing mechanism will provide the SL sequence that already contains a cap structure at its 5' end.

It has been only recently that experiments with stable transfected trypanosomes have begun to provide convincing evidence that the control of VSG gene expression occurs at several levels including transcription initiation. By inserting a VSG promoter linked to a selectable marker into a chromosomal location, instead of episomes, or by tagging an expression site with a reporter gene, it was possible to follow the activity of a given VSG promoter when the trypanosome is induced to differentiate, or when it spontaneously undergoes a VSG switch (22,25,26). The control of VSG gene expression during the parasite life cycle has been more extensively investigated using nuclear run-on assays. Several groups have shown that, although VSG mRNAs are not detected in procyclic forms, transcription initiation does occur at a low level from many, if not all, ES promoters (27-29). This is consistent with the observation of CAT or luciferase expression under the control of VSG gene promoters cloned in episomal vectors in transiently transfected procyclic parasites (29,30). However, when VSG promoters were further analyzed within their chromosomal context, instead of plasmid vectors, it was found that the transcription elongation step is subjected to stage-specific regulation, i.e., the active VSG gene is fully transcribed in bloodstream parasites but most transcrip-

tional activity stops within 1 kb of the VSG promoter in procyclics (27-29).

Post-transcriptional mechanisms as the main method of controlling gene expression in *T. brucei* and *Leishmania* spp

In *T. brucei*, in addition to VSG genes, a second group of genes encoding highly abundant surface glycoproteins has been used as a model to explain the mechanisms of gene expression control. When bloodstream trypanosomes differentiate into procyclics, the VSG coat is rapidly replaced by a new surface coat composed of small, acidic proteins named procyclins or procyclic acid repetitive proteins (PARP) (for a review, see Ref. 31). PARP genes comprise a polymorphic gene family that is found at two genetic loci (α and β) in the genome of several strains of *T. brucei*. Two or three PARP genes are found at each locus, together with a PARP-associated gene (PAG). Like the majority of trypanosome genes, PARP genes are part of a polycistronic transcription unit, and like VSG genes, PARP genes are transcribed by an α -amanitin-resistant RNA polymerase (32). However, sequence comparison shows that the PARP promoters and VSG promoters that have been characterized so far bear little resemblance to each other, or to the classical RNA polymerase I rRNA promoter (3).

The work of several groups has revealed what Isabel Roditi calls the "philosophy behind the regulation of gene expression by trypanosomes: trust is good, control is better" (33). Stage-specific control of PARP gene expression occurs at several levels: induced expression of procyclin mRNA seen 2 h after bloodstream forms differentiate into insect forms is due to a 5- to 10-fold increase in transcription initiation as well as transcription elongation of PARP genes (31). Although PARP mRNA is not detectable in bloodstream trypanosomes, PARP genes are

transcribed at low levels in bloodstream parasites indicating that some post-transcription mechanisms must take place in order to prevent the accumulation of PARP mRNA. Moreover, although PARP and PAG genes are derived from the same polycistronic transcription unit, there is a 100-fold difference in the steady-state levels when PARP and PAG mRNAs are compared (31). Several possibilities have been considered to explain this difference in the accumulation of mRNAs, including differences in *trans*-splicing efficiency and the presence of regulatory sequences in the 3'-untranslated regions (UTR) affecting the steady-state levels of these mRNAs. The study of regulatory elements derived from the 3'UTR has been extensively explored (28,31,34,35). Using transient transfections, Roditi's group (35) has demonstrated that a 16-mer sequence which is predicted to form a stem-loop structure in the 3'UTR of PARP mRNAs can confer regulation of a CAT reporter gene on procyclics. Deletion of this element causes a 12-fold drop in CAT expression but no significant difference in the levels of total RNA, suggesting control at the translational level. Interestingly, the same sequence, with the same predicted secondary structure, has also been found in the analogous gene of another trypanosome species, *T. congolense*. More recently, a systematic analysis of the entire 297-base 3'UTR revealed three additional elements which are involved in post-transcriptional regulation of PARP genes, all three of them having an effect on the rate of mRNA turnover (34). It seems that, by employing several layers of regulation, these parasites can ensure that the rapid changes associated with transmission between insect vector and mammalian host are followed by an instant reprogramming of genetic expression.

When studies performed with *T. brucei* are compared with the results obtained with other trypanosomatids, the basic observations remain. However, there are no reports

describing transcriptional regulation affecting gene expression in *Leishmania* or *T. cruzi*. In fact, there is not a single report describing promoter elements for protein-coding genes in these organisms. Most of the work in *Leishmania* species has been focused on one of the following gene families, the first three of which are developmentally regulated during the parasite life cycle: the α and β tubulin genes (36), the surface metalloprotease gp63 or *msp* genes (37), the amastigote-specific A2 genes (38) and heat-shock protein genes (39). The initial studies performed by Dyann Wirth's group (9) describing a transfection system for *L. enrietti* used a plasmid containing the intergenic region of the α -tubulin cluster flanking a CAT gene. Their prediction was that there might exist at least two potential sites for transcription initiation in a gene cluster: one initiation site adjacent to each gene, or a single promoter upstream from the first gene in the cluster. When tests for the presence of intergenic promoters in the tubulin cluster were performed using transient transfections there was a surprise: only a very short sequence derived from the tubulin locus was necessary for CAT expression and that sequence basically contained an SL addition site (9,40). It was proposed that the promoter function was being provided by plasmid DNA sequences and the tubulin sequences were just providing SL acceptor sites for the processing of the mRNA. Since then, additional attempts to identify promoter regions within *Leishmania* genes have failed. A current model proposes that, in *Leishmania* species and probably in *T. cruzi* as well, the RNA polymerase has the ability to land in any stretch of DNA and initiate transcription. This is postulated to be possible because the genome of trypanosomatids is very dense, with short intergenic sequences. As illustrated in Figure 1, the pre-mRNAs generated by random initiation sites are then cleaved and processed by *trans*-splicing and polyadenylation.

Gene manipulation in *T. cruzi*: the development of vectors and transfection protocols

The field of molecular genetics was progressing rapidly, with several new reports about gene expression in *T. brucei* and *Leishmania* being published, when, in 1991, the first transfection experiment of *T. cruzi* was reported. Using a vector containing a segment of the SL gene placed upstream of the bacterial CAT gene, Lu and Buck (41) were able to detect CAT activity in transfected epimastigotes. Although those experiments were later called “transient transfections”, CAT expression reached a maximum after 48 h but remained unchanged at 72 h and 96 h, indicating that the plasmid was being maintained with some stability inside the cells. In 1992, Kelly et al. (42) described an improved vector named pTEX which was constructed using flanking sequences derived from the two tandemly repeated glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) genes. To help cloning and to allow selecting transformants, a multiple cloning site and the *neo* gene as a drug resistance marker were included. As was reported in similar experiments with *T. brucei* and *Leishmania*, the plasmid replicates episomally in the form of large concatemers and remains stable in the presence of drug selection. More importantly, similar to the results observed in several *Leishmania* species, the copy number and the level of expression can be modulated by changing drug concentration (42).

The results obtained by Kelly's group and later by other investigators using the pTEX vector suggested that, regarding the mechanisms of gene expression following plasmid transfections, *T. cruzi* has more in common with *Leishmania* than with *T. brucei*. Whereas stable transfectants of *T. cruzi* and *Leishmania* were easy to obtain using episomal vectors, with *T. brucei* this was rarely achieved. In fact, pTEX can be used to transfect several species of *Leishmania*, but not *T.*

brucei or *Crithidia fasciculata* (42). More surprisingly, expression of foreign genes in *T. cruzi* and *Leishmania* was achieved using episomal vectors containing no promoter elements. As is the case for *Leishmania*, we can also speculate that transcription initiates randomly within the plasmid vector and the long transcript generated is processed according to the rules of *trans*-splicing and polyadenylation. And once again, similar to *Leishmania*, attempts to identify an RNA polymerase II promoter in *T. cruzi* using the plasmid transfection approach have failed. In a series of experiments aimed at identifying the promoter region of the *T. cruzi* amastin gene cluster (see below), we constructed a variety of plasmids containing various DNA fragments derived from the intergenic region between two copies of amastin genes as well as the 5'-flanking region of the cluster placed in front of the luciferase reporter gene (Figure 3) (43). When luciferase expression in epimastigotes transfected with these constructs was compared, the results were quite similar, about 700-fold above background levels. Moreover, the levels of luciferase activity resulting from transfections with these constructs were also similar to the levels found in cells transfected with the plasmid pLST, where most of the trypanosome sequences were deleted. In this deletion, only a 73-bp fragment containing basically *trans*-splicing signals of amastin mRNA was left upstream from the luciferase gene.

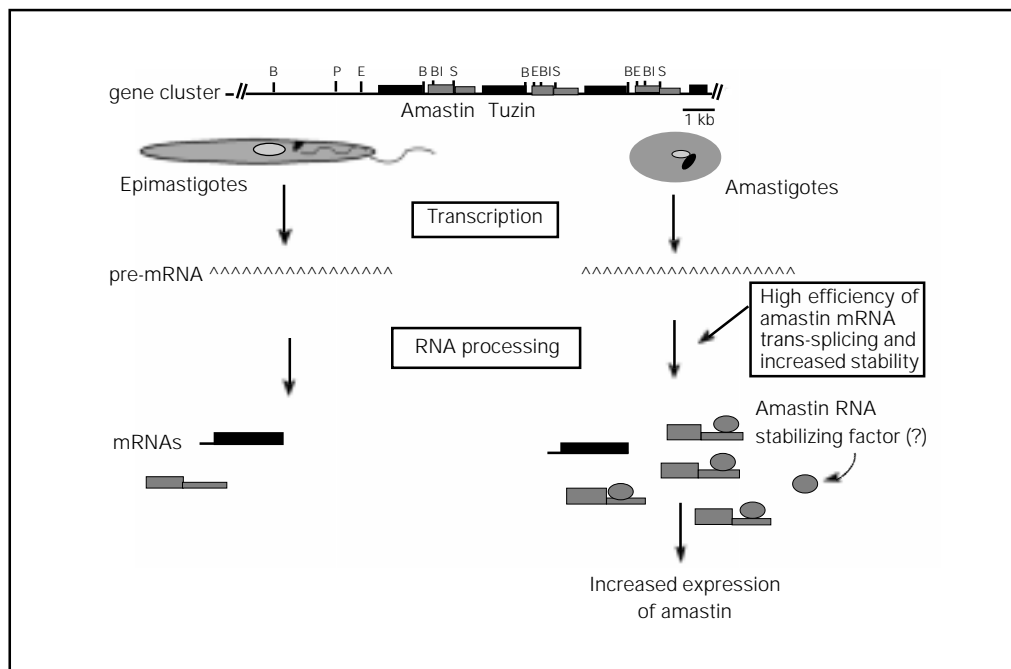
Even in the absence of a promoter, using pLST and various other luciferase plasmids, we were able to detect high levels of luciferase activity in transfected epimastigotes but not in amastigotes. Since the promoter for amastin genes could not be identified, we decided to construct a vector with another type of promoter as a way to improve expression in amastigotes. Initially, we tested the *T. brucei* PARP promoter and SL acceptor site, but this construction resulted in luciferase levels in epimastigotes that were similar to those obtained with pLST (43). At

the time, a group of investigators was in the process of characterizing the *T. cruzi* rRNA promoter. Dietrich et al. (44) have determined the sequence of the intergenic spacer of an rRNA gene and mapped the putative transcription start site. When we inserted a 580-bp fragment containing this sequence in front of the amastin SL addition site, in the proper orientation within the pLST vector, luciferase activity resulting from transfected epimastigotes was 270,000-fold above background (or 2,000-fold greater than that generated with pLST). With this improved vector we were able, for the first time, to obtain luciferase expression in transiently transfected amastigotes (43). At the same time, Tyler-Cross et al. (45) also described a transient expression system mediated by the *T. cruzi* rRNA promoter and showed that the promoter fragment is highly efficient at driving CAT expression in transiently transfected epimastigotes. More recently, Martínez-Calvillo et al. (46) described a pTEXT derivative, called pRIBOTEX, where the upstream region of *gapdh* was replaced by another rRNA gene spacer fragment bearing a previously mapped transcription start site. *T. cruzi*

cultures transiently transfected with pRIBOTEX express a CAT activity 16,000-fold greater than the activity generated with the original pTEXT vector. More importantly, since that vector contains a selectable marker that allows generation of stable transfectants, the presence of a strong promoter permitted a more rapid drug selection time, with G418-resistant clones being obtained 2 weeks earlier.

The only other *T. cruzi* promoter that has been characterized is the SL gene promoter, but the activity of this promoter, as described by Nunes et al. (47), was unexpectedly low. However, these investigators reported a major finding about the activity of the SL and rRNA promoters: promoters isolated from a *T. cruzi* strain classified as group I, as defined by Souto et al. (48), are efficiently expressed only in *T. cruzi* isolates classified as belonging to the same group (49). The description of two major lineages of *T. cruzi* strains, based on the analysis of several nuclear markers, has now been strongly reinforced by the functional specificity of the SL and rRNA promoters. Given the well-documented species-specific pattern of the rRNA

Figure 3 - A model for the regulation of amastin gene expression in *T. cruzi*. The amastin/tuzin gene cluster is shown, with dark and gray boxes representing copies of tuzin and amastin genes, respectively. Intergenic regions are shown as thin lines (The diagram including the restriction sites was modeled after Teixeira et al. (43)). Polycistronic transcription, which occurs at the same rates in amastigotes and epimastigotes, generates a pre-mRNA (wavy lines) that needs to be processed in order to produce mature mRNA encoding amastin (gray box) and tuzin (dark box) proteins. The differential accumulation of these mRNAs in the two forms of the parasite's life cycle may be dependent upon at least two factors: higher efficiency of amastin mRNA processing and increased stability of the mRNA in amastigotes. An RNA-binding protein, which is more abundant in amastigotes, is proposed to be involved in the regulatory pathway through specific binding at the 3'UTR of amastin mRNA (gray line).



promoter in higher eukaryotes, these results represent functional evidence for ancient evolutionary divergence among *T. cruzi* strains.

In addition to episomal transfections, stable transformants of *T. cruzi* can also be generated with high efficiency as a result of integration of the plasmid vector by homologous recombination. Using either episomal vectors or DNA integration protocols of transfection, several groups have reported the expression of foreign genes in *T. cruzi*, including mammalian cytokines such as interleukin-2 and gamma-interferon (50), the *T. brucei* GPI-phospholipase C (51), and the green fluorescent protein (GFP) from algae (Teixeira SMR, unpublished data). The use of GFP-fusion proteins or epitope tagging has great potential as a method for subcellular localization, as recently reported (52). Likewise, stable transfectants were also generated in order to overexpress *T. cruzi* genes such as cruzipain, a cysteine proteinase (53). With this methodology, gene disruption could also be achieved by targeted integration of a DNA segment containing a selectable marker. A few groups have exploited this new technology as a tool for functional analysis of *T. cruzi* genes. That number remains rather low not only for *T. cruzi* but for trypanosomatids in general, mainly because in these organisms genes are usually present as multicopy families, making gene knock-out experiments very difficult. In 1993, just after Hariharan et al. (54) described plasmid integration into the polyubiquitin gene locus, disrupting one of the polyubiquitin genes, Otsu et al. (55) and Cooper et al. (56) reported on target gene replacement resulting in disruption of single copy genes of *T. cruzi*. Deletion of the gene encoding the immunodominant *T. cruzi* surface glycoprotein GP72 by target integration caused a dramatic morphological change in which the normal attachment of the flagellum to the parasite cell membrane was lost (56). Although this experiment has established that gene knock-out is a useful tool for functional

studies of genes in *T. cruzi*, other methodologies, such as the use of anti-sense RNA and, more recently, ribozymes, may also contribute to addressing these questions in cases where multicopy genes are present.

Regulatory elements controlling the expression of stage-specific genes in *T. cruzi*

One of the beauties of working with *T. cruzi* as well as other trypanosomatids is that they represent excellent models to study the genetic control of cell differentiation. Compared to other trypanosomatids, regulatory mechanisms of gene expression during the life cycle of *T. cruzi* have been much less explored. Albeit the molecular tools are now available for in depth studies of life cycle stage-specific regulation of gene expression in *T. cruzi*, only two groups have taken on the task of identifying regulatory elements within *T. cruzi* genes. The strategies are quite straightforward: after defining which DNA fragments are to be analyzed, we can use reporter genes and transfection protocols similarly to the approaches that have been successfully applied in the studies of *T. brucei* and *Leishmania* genes.

Working with John Donelson at the University of Iowa, a few years ago I decided to investigate the processes responsible for the control of genes that are expressed preferentially in amastigotes, the intracellular stage of the *T. cruzi* life cycle. Using the differential screening approach, we cloned and characterized a gene cluster containing several copies of amastins, a gene that is preferentially expressed in amastigote forms (57). Together with tubulin genes, for which Ürményi et al. (58) have started to investigate the changes in mRNA levels during metacyclogenesis, amastin genes constitute excellent models for studying stage-specific expression in *T. cruzi*. Afterwards, we discovered that amastin genes are organized in alternating repeats with a second group of

genes, which we named tuzin (43,57) (Figure 3). DNA sequence analysis predicted that amastins are hydrophobic proteins of 174 amino acids localized at the surface of amastigotes. Northern blot analysis indicated that amastin mRNAs are 60-fold more abundant in amastigotes compared to their levels in the epimastigote or trypomastigote stage. Analysis of similar Northern blots probed with tubulin genes showed that tubulin expression is also under a strict stage-specific control, being expressed at much higher levels in epimastigotes compared to other forms (Teixeira SMR, unpublished data). In contrast to amastin and tubulin genes, expression of tuzin genes, which code for a much less abundant hydrophilic protein, is not regulated during the parasite's life cycle. Although amastin and tuzin genes are transcribed by the same RNA polymerase into the same polycistronic pre-mRNA, mature tuzin mRNAs accumulate at low, constant levels, equivalent to the levels of amastin mRNA in epimastigotes (43). As a first step towards the identification of elements involved in the control of expression of the amastin/tuzin gene cluster we demonstrated, using nuclear run-on experiments, that transcription initiation remains constant when nuclei from epimastigotes and amastigotes are compared. To investigate if the 3' or 5'UTRs and intergenic regions were involved, fragments containing different regions of the cluster were examined using transient transfections. With luciferase as a reporter gene we compared luciferase expression in transfected epimastigotes and amastigotes, normalized to a control plasmid. Our results indicated a significant regulatory role of the amastin 3'UTR: luciferase activity appeared to be 36-fold higher in amastigotes when these cells were transfected with a construct containing this fragment inserted downstream of the luciferase gene (43). We further examined the presence of regulatory elements located in that region using an approach similar to the linker scanner assay where we

sequentially replaced contiguous 70-bp segments of the 630-bp 3'UTR with random sequences derived from synthetic oligonucleotides. We also obtained evidence for the existence of an RNA-binding protein that recognizes these sequences in the amastin 3'UTR (59). More recently, we have obtained evidence indicating that differences in 5'UTR are also important for determining the levels of amastin and tuzin mRNAs (Teixeira SMR, Kirchhoff LV and Donelson JE, unpublished data). Taken together, these results are consistent with the regulatory model shown in Figure 3: several factors acting at the post-transcriptional level may be responsible for determining the differential expression of amastin and tuzin genes in *T. cruzi*. After we described the effect of the amastin 3'UTR, Nozaki and Cross (60) reported a similar study where the 3'UTR from various stage-specific and constitutive *T. cruzi* genes were tested using stable transfection experiments with the luciferase reporter gene. Their results confirmed our findings with the amastin 3'UTR, i.e., luciferase expression increases 16-fold in amastigotes while decreasing 14-fold when epimastigotes were transfected. In contrast, the presence of the 3'UTR from two other stage-specific genes, *gp72* and *gp85*, showed a much smaller modulatory activity than amastin 3'UTR. Various exciting experiments are currently in progress and we expect that we will soon have a better understanding of the mechanisms involving the 3'UTR sequences and other factors that might be responsible for determining the level at which a particular mRNA accumulates during the parasite's life cycle. So far, preliminary evidence indicates the efficiency of the processing step and the mechanisms determining the mRNA half-life.

Concluding remarks

Undoubtedly, parasitologists now have the most powerful instrument for dissecting

their favorite parasite since the electron microscope was invented. Genetic transformation, or DNA transfection, has produced not just an extraordinary amount of new information about the biology of these organisms but has given us the possibility of manipulating a significant part of their genomes. Moreover, the new era of molecular parasitology can go hand-in-hand with the fast growing developments achieved with parasite genome projects. Above all, the excitement that each new report or technical breakthrough brings to the field is especially welcome because it

can drive an important fraction of the future generation of researchers into a field that needs all the help we can get. As everyone agrees, understanding post-transcriptional control in trypanosomes may not solve Chagas' disease, but will certainly contribute to making us more prepared to fight it.

Acknowledgments

I thank Dr. Gregory Kitten and Dr. Maria de Fátima Horta for discussion and comments on the manuscript.

References

1. Fernandes AP, Nelson K & Beverley SM (1993). Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. *Proceedings of the National Academy of Sciences, USA*, 90: 11608-11612.
2. Donelson JE & Zeng W (1990). A comparison of trans-RNA splicing in trypanosomes and nematodes. *Parasitology Today*, 6: 327-334.
3. Lee MGS & Van der Ploeg LHT (1997). Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annual Review of Microbiology*, 51: 463-489.
4. LeBowitz JH, Smith HQ, Rusche L & Beverley SM (1993). Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes and Development*, 7: 996-1007.
5. Matthews KR, Tschudi C & Ullu E (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes and Development*, 8: 491-501.
6. Schurch N, Hehl A, Vassella E, Braun R & Roditi I (1994). Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. *Molecular and Cellular Biology*, 14: 3668-3675.
7. Huang J & Van der Ploeg LHT (1991). Maturation of polycistronic pre-mRNA in *Trypanosoma brucei*: analysis of trans-splicing and poly(A) addition at nascent RNA transcripts from the hsp70 locus. *Molecular and Cellular Biology*, 11: 3180-3190.
8. Huang J & Van der Ploeg LHT (1991). Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from immediately adjacent 3' splice acceptor site. *EMBO Journal*, 10: 3877-3885.
9. Laban A & Wirth D (1989). Transfection of *Leishmania enriettii* and expression of chloramphenicol acetyltransferase gene. *Proceedings of the National Academy of Sciences, USA*, 86: 9119-9123.
10. Kelly JM (1997). Genetic transformation of parasitic protozoa. *Advances in Parasitology*, 39: 227-270.
11. Ross R & Thomson D (1910). A case of sleeping sickness studied by precise enumerative methods. Regular periodic increase in the parasites disclosed. *Proceedings of the Royal Society of London*, 82: 411-415.
12. Cross GAM (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, 71: 393-417.
13. Bridgen PJ, Cross GAM & Bridgen J (1976). N-terminal amino acid sequences of variant-specific antigens from *Trypanosoma brucei*. *Nature*, 263: 613-614.
14. Cross GAM (1996). Antigenic variation in trypanosomes: secrets surface slowly. *BioEssays*, 18: 283-291.
15. Borst P, Rudenko G, Taylor MC, Blundell PA, Van Leeuwen F, Bitter W, Cross M & McCulloch R (1996). Antigenic variation in trypanosomes. *Archives of Medical Research*, 27: 379-388.
16. Williams RO, Young JR & Majiva PAO (1979). Genomic rearrangements correlated with antigenic variation in *Trypanosoma brucei*. *Nature*, 282: 847-849.
17. Van Leeuwen F, Taylor MC, Mondragon A, Moreau H, Gibson W, Kieft R & Borst P (1998). Beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proceedings of the National Academy of Sciences, USA*, 95: 2366-2371.
18. Cully DF, Ip HS & Cross GAM (1985). Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. *Cell*, 42: 173-182.
19. Gibbs CP & Cross GAM (1988). Cloning and transcription analysis of a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 28: 197-206.
20. Kooter JM, Van der Speck HJ, Wagter R, Oliveira CE, Van der Hoeven F, Johnson PJ & Borst P (1987). The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *Trypanosoma brucei*. *Cell*, 51: 261-272.
21. Salmon D, Geuskens M, Hanocq F, Hanocq-Quertier J, Nolan D, Ruben L & Pays E (1994). A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell*, 78: 75-86.
22. Rudenko G, Cross M & Borst P (1998). Changing the end: antigenic variation orchestrated at the telomeres of African trypanosomes. *Trends in Microbiology*, 6: 113-116.
23. Johnson PJ, Kooter JM & Borst P (1987). Inactivation of transcription by UV irradiation

- tion of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. *Cell*, 51: 273-281.
24. Alarcon CM, Son HJ, Hall T & Donelson JE (1994). A monocistronic transcript for a trypanosome variant surface glycoprotein. *Molecular and Cellular Biology*, 14: 5579-5591.
 25. Rudenko G, Blundell PA, Dirks-Mulder A, Kieft R & Borst P (1995). A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in *T. brucei*. *Cell*, 83: 547-553.
 26. Horn D & Cross GAM (1995). A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei*. *Cell*, 83: 555-561.
 27. Graham SV (1995). Mechanisms of stage-regulated gene expression in kinetoplastida. *Parasitology Today*, 11: 217-223.
 28. Vanhamme L & Pays E (1995). Control of gene expression in trypanosomes. *Microbiological Reviews*, 59: 223-240.
 29. Pays E, Coquelet H, Tebabi P, Pays A, Jefferies D, Steinert M, Koenig E, Williams RO & Roditi I (1990). *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *EMBO Journal*, 9: 3145-3151.
 30. Jefferies D, Tebabi P & Pays E (1991). Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression at the post-transcriptional level. *Molecular and Cellular Biology*, 11: 338-343.
 31. Roditi I (1996). The VSG-procyclicin switch. *Parasitology Today*, 12: 47-49.
 32. Rudenko G, Le Blancq S, Smith J, Gwo-Shu-Lee M, Rattray A & Van der Ploeg LHT (1990). Procyclic acidic repetitive protein (PARP) genes located in an unusually small α -amanitin-resistant transcription unit: PARP promoter activity assayed by transient DNA transfection of *Trypanosoma brucei*. *Molecular and Cellular Biology*, 10: 3492-3504.
 33. Roditi I, Schurch N, Furger A & Ruep S (1996). Trust is good, control is better: the philosophy behind the regulation of gene expression by trypanosomes. *Memórias do Instituto Oswaldo Cruz*, 91 (Suppl): 50-51.
 34. Furger A, Schurch N, Kurath U & Roditi I (1997). Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect form of *Trypanosoma brucei* by modulating RNA stability and translation. *Molecular and Cellular Biology*, 17: 4372-4380.
 35. Hehl A, Vassella E, Braun R & Roditi I (1994). A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences, USA*, 91: 370-374.
 36. Coulson RMR, Connor V, Chen JC & Ajioka J (1996). Differential expression of *Leishmania major* β -tubulin genes during the acquisition of promastigote infectivity. *Molecular and Biochemical Parasitology*, 82: 227-236.
 37. Ramamoorthy R, Swihart K, McCoy JJ, Wilson ME & Donelson JE (1995). Intergenic regions between tandem gp63 genes influence the differential expression of gp63 RNAs in *Leishmania chagasi*. *Journal of Biological Chemistry*, 270: 12133-12139.
 38. Charest H, Zhang W & Matlashewski G (1996). The developmental expression of *Leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3' UTR region. *Journal of Biological Chemistry*, 271: 17081-17090.
 39. Argaman M, Aly R & Shapira M (1994). Expression of heat shock protein 83 in *Leishmania* is regulated post-transcriptionally. *Molecular and Biochemical Parasitology*, 64: 95-110.
 40. Curotto de Lafaille MA, Laban A & Wirth D (1992). Gene expression in *Leishmania*: analysis of essential 5' DNA sequences. *Proceedings of the National Academy of Sciences, USA*, 89: 2703-2707.
 41. Lu H & Buck GA (1991). Expression of exogenous gene in *Trypanosoma cruzi* epimastigotes. *Molecular and Biochemical Parasitology*, 44: 109-114.
 42. Kelly JM, Ward HM, Miles MA & Kendall G (1992). A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania donovani*. *Nucleic Acids Research*, 20: 3963-3969.
 43. Teixeira SMR, Kirchhoff LV & Donelson JE (1995). Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*. *Journal of Biological Chemistry*, 270: 22586-22594.
 44. Dietrich P, Soares MB, Affonso MHT & Floeter-Winter L (1993). The *Trypanosoma cruzi* ribosomal RNA-encoding genes: analysis of promoter and upstream intergenic spacer sequences. *Gene*, 125: 103-107.
 45. Tyler-Cross RE, Short SL, Floeter-Winter L & Buck GA (1995). Transient expression mediated by the *Trypanosoma cruzi* rRNA promoter. *Molecular and Biochemical Parasitology*, 72: 23-31.
 46. Martínez-Calvillo S, López I & Hernández R (1997). pRIBOTEX expression vector: a pTEX derivative for a rapid selection of *Trypanosoma cruzi* transfectants. *Gene*, 199: 71-76.
 47. Nunes LC, Carvalho MRC, Shakarian AM & Buck GA (1997). The transcription promoter of the spliced leader gene from *Trypanosoma cruzi*. *Gene*, 188: 157-168.
 48. Souto RP, Fernandes O, Macedo AM, Campbell DA & Zingales B (1996). DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 83: 141-152.
 49. Nunes LC, Carvalho MRC & Buck GA (1997). *Trypanosoma cruzi* strains partition into two groups based on the structure and function of the spliced leader RNA and rRNA gene promoters. *Molecular and Biochemical Parasitology*, 86: 211-224.
 50. La Flamme AC, Buckner FS, Swindle J, Ajioka J & Van Voorhis WC (1995). Expression of mammalian cytokines by *Trypanosoma cruzi* indicates unique signal sequence requirements and processing. *Molecular and Biochemical Parasitology*, 75: 25-31.
 51. Garg N, Tarleton RL & Mensa-Wilmot K (1997). Proteins with GPI signal sequences have divergent fates during a GPI deficiency. GPIs are essential for nuclear division in *Trypanosoma cruzi*. *Journal of Biological Chemistry*, 272: 12482-12491.
 52. Tibbetts RS, Klein KG & Engman DM (1995). A rapid method for protein localization in trypanosomes. *Experimental Parasitology*, 80: 572-574.
 53. Tomas AM, Miles MA & Kelly JM (1997). Overexpression of cruzipain, the major cysteine proteinase of *Trypanosoma cruzi*, is associated with enhanced metacyclogenesis. *European Journal of Biochemistry*, 244: 596-603.
 54. Hariharan S, Ajioka J & Swindle J (1993). Stable transformation of *Trypanosoma cruzi*: inactivation of the PUB12.5 polyubiquitin gene by targeted gene disruption. *Molecular and Biochemical Parasitology*, 57: 15-30.
 55. Otsu K, Donelson JE & Kirchhoff LV (1993). Interruption of a *Trypanosoma cruzi* gene encoding a protein containing 14-amino acid repeats by targeted insertion of the neomycin phosphotransferase gene. *Molecular and Biochemical Parasitology*, 57: 317-330.
 56. Cooper R, de Jesus AR & Cross GAM

- (1993). Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *Journal of Cell Biology*, 122: 149-156.
57. Teixeira SMR, Russell D, Kirchhoff LV & Donelson JE (1994). A differentially expressed gene family encoding "amastin", a surface protein of *Trypanosoma cruzi* amastigotes. *Journal of Biological Chemistry*, 269: 20509-20516.
58. Ürményi T, de Castro FT, Carvalho JF, de Souza W & Rondinelli E (1992). Transcriptional and post-transcriptional control of tubulin gene expression in *Trypanosoma cruzi*. *DNA and Cell Biology*, 11: 101-109.
59. Teixeira SMR, Coughlin BC, Kirchhoff LV & Donelson JE (1996). 5' and 3' regulatory RNA sequences in the developmentally regulated amastin/tuzin gene cluster of *Trypanosoma cruzi*. *Memórias do Instituto Oswaldo Cruz*, 91 (Suppl): 49.
60. Nozaki T & Cross GAM (1995). Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 75: 55-67.