

The untranslated regions of genes from *Trypanosoma cruzi*: perspectives for functional characterization of strains and isolates

Adeilton Brandão

Departamento de Medicina Tropical, Instituto Oswaldo Cruz-Fiocruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brazil

The sequencing of Trypanosoma cruzi genome has been completed and a great deal of information is now available. However, the organization of protozoa genomes is somewhat elusive and much effort must be applied to reveal all the information coded in the nucleotide sequences. Among the DNA segments that needs further investigation are the untranslated regions of genes. Many of the T. cruzi genes that were revealed by the genome sequencing lack information about the untranslated regions. In this paper, some features of these untranslated segments as well as their applications in T. cruzi populations are discussed.

Key words: untranslated regions - genes - genomes - protozoa

The finishing of the *Trypanosoma cruzi* genome sequencing (El Sayed et al. 2005) has disclosed for the community of parasitologists an enormous volume of information about genes and other nucleotide sequences from this protozoan parasite. Paradoxically, this same flow of information constitutes the bottleneck for the revelation of all molecular aspects of the parasite life cycle, from its purely biological component (interaction with the invertebrate and vertebrate hosts) up to the medical and social impact (Chagas disease and its clinical variations). At present, the speed with which nucleotide sequences are produced has led to disequilibrium between the potential of the genomic information and its straight use for the elucidation of the biological phenomena. The reductionism impulse stimulated by the availability of the nucleotide sequences only made clear that the complex network of biological interactions is not explicitly encoded in the order of the genome bases. Indeed, they are the result of a dynamic process in which several live beings define what genomes will survive. In the case of parasites and its hosts, there always is a dispute between two unrelated genomes, which can interact for million years and the net result is an improvement in their skills for survival in hostile environments.

The genes of the eukaryotes present segments in their 5' and 3' ends that are transcribed into RNA but not translated, and therefore they are called untranslated regions or simply 5' and 3' UTR. In these segments, DNA elements involved in regulation of the gene expression and in the interaction with the new types of RNA (e.g., micro-RNA) have already been identified (Mignone et al. 2002, Hughes 2006). For *T. cruzi* (and all trypanosomatids), most of the genes have no introns and the messenger RNA are processed by the mechanism of trans-splicing (Gull 2001, Liang et al. 2003). In principle, the analysis of the transcripts in trypanosomatids could be easily accomplished,

from the very moment that the respective sequences are available. The definition of the "cardinal points" in the UTRs of the genes from *T. cruzi*, i.e. beginning of 5'UTR and the end of 3'UTR, is a process that depends a great deal on the experimental investigation, especially for the 3'UTR, since the polyadenylation signal typical of eukaryotes has not been recognized in the trypanosomatids. For the time being no computational tool has been developed to accurately predict the site of polyadenylation in trypanosomatids, and only with the aid of experimental work can the complete information be extracted from their genes. The absence of canonic polyadenylation signal in trypanosomatids may be related to the mechanism that couples the polyadenylation of upstream gene to trans-splicing of a downstream gene, and it has been shown that the modification of the downstream splice site (dinucleotide AG) shifted the poly (A) site (Lebowitz et al. 1993, Mathews et al. 1994). Nucleotide sequence per se does not inform precisely at which points the transcription of *T. cruzi* genes begins and finishes, but the existence of the cis elements (the dinucleotide AG, poly-pyrimidines rich tracts, adenine branching point) which are involved in the coupling of trans-splicing and polyadenylation provide the starting points for generation of transcription recognition algorithms (Gopal et al. 2005, Benz et al. 2005). Although these algorithms proved to work fairly well in detecting the correct splice site for the organisms studied (> 80% of hits for *T. brucei* and *Leishmania major*) they suffer from limitations imposed on gene prediction algorithms (Brent & Guigo 2004). Experimental approaches can also be used to complement the work of these algorithms and some methods have been developed to explore the canonical signals associated to trypanosomes mRNA processing (Webb et al. 2005, Siegel et al. 2005). Thus, if information generated by cDNA sequencing or transcription assays is not available, computational analysis of sequences allows only an estimate of these "cardinal points". In this context a window of opportunity opens for complementing the information of *T. cruzi* genome and to achieve this goal it is essential that accurate description of their untranslated regions be available to perform thorough analysis. The power of current experimental methods allows one to pinpoint nucleotides in

E-mail: abrandao@fiocruz.br

Received 16 May 2006

Accepted 15 August 2006

these regions and afterwards observe the effects on parameters such as transcription rates, gene expression and regulation, mRNA stability and half-life. Association of sequence motifs in 3'UTR of *T. cruzi* genes to defined cellular and molecular effects has gained sound experimental support (D'Orso & Frasch 2001, Bartholomeu et al. 2002). Though both untranslated segments are expected to influence gene expression and function, most investigations points to 3'UTR playing a major role by the presence of cis-acting elements in *T. cruzi* genes (Nozak & Cross 1995, D'Orso et al. 2003). This can be viewed as the result of structural differences, for the 3'UTR being on average 2-3 times longer than the 5'UTR increases the possibilities to influence mRNA processing (more sequence elements, protein binding sites, and secondary structures). Genes like glycoprotein 72 (gp72), glycoprotein 85 (gp85), glyceraldehyde 1-phosphate dehydrogenase (gapdh), amastin, β -tubulin, mucin, trans-sialidase and, FL-60 have elements (presumptively or accurately described) in their 3'UTR that in different ways affect gene expression in stage-specific process (Nozak & Cross 1995, Abuin et al. 1999, Weston et al. 1999, Coughlin et al. 2000, Di Noia et al. 2000, Bartholomeu et al. 2002).

The bottleneck created by the huge amount of nucleotide sequences offers new routes in the search of models and experimental strategies that translate the DNA code into biological phenomenon. Thus the DNA sequence to biological phenomenon translation is a process of searching for codes behind the code, i.e., the searching of a Meta code. Except for the house keeping genes and the functionally important ones in closely related species, the composition and structural characteristics of UTRs of eukaryotes can be very distinct (Mignone et al. 2005). This at first can be explained by the absence of highly functional pressures, as occurs to the protein coding segments of genes. Inspection of *T. cruzi* sequences shows that untranslated regions in some genes can be good molecular and functional markers for the populations of *T. cruzi*. Minor variations in each UTR (small insertion/deletions) can be the starting points for the establishment of relationships between the gene expressions and population structure. It is not clear at what basis this relationship should be structured, but some hints come from variations detected in the UTR of evolutionary conserved genes in strains of *T. cruzi*. It is worth note that small changes in the UTRs can affect the expression of the gene (Weston et al. 1999, Teixeira et al. 1999), and this can be a smart evolutionary strategy to gain functional advantage without large mutations in genes themselves. The result of these alterations in protozoa parasites is represented by more ability for survival in different environments.

Sequence analysis shows that 3'UTR of the calmodulin gene in the major lineages of *T. cruzi* present mutational events that cause alterations in the predicted secondary structure of the respective messenger RNA (Brandao & Fernandes 2006). This gene is very important for the cell and carries out multiple tasks in the cellular metabolism (Hoefflich & Ikura 2002). Thus, it is very promising that a population group of *T. cruzi*, the so-called Zymodeme III (ZIII), which is predominant in a region that is one of the

biggest repositories of the genetic diversity in the globe (the Amazon region), present a deletion of 13 nucleotides in 3'UTR in comparison with the groups TcI and TcII. Certainly, this different secondary structure of the calmodulin messenger RNA in *T. cruzi* groups means that either the expression or the regulation of this fundamental gene is driving the functional divergence of *T. cruzi* populations. If it is true, the network of gene interactions in *T. cruzi*, and its biological consequences (interaction with the vertebrate and invertebrate hosts) show fluctuations that not only respond to the different environments but create evolutionary alternatives for the species. In the present description of the *T. cruzi* genome (El-Sayed et al. 2005) the ~ 6000 hypothetical genes are presumed to act one by one with their respective untranslated regions, but a completely different picture can emerge if we imagine this same number of genes making interactions with each other. In this last possibility two separated (conceptually, not biological) network should be built: the gene products and the untranslated regions interactions.

The meeting of two very different organisms, a protozoan parasite and its vertebrate host, is a biological phenomenon in which the most complex genome (from the point of view of the content and functions) will be tested in his capacity of dealing with a less complex one. This will promote many events of selection, adaptation and genomic variation. Thus, from an evolutionary perspective, the parasitic David (less complex genomes) will keep on challenging the Goliad contained in the more complex genomes.

REFERENCES

- Abuin G, Freitas-Junior LH, Colli W, Alves MJ, Schenkman S 1999. Expression of trans-sialidase and 85-kDa glycoprotein genes in *Trypanosoma cruzi* is differentially regulated at the post-transcriptional level by labile protein factors. *J Biol Chem* 274: 13041-13047.
- Bartholomeu DC, Silva RA, Galvão LM, el-Sayed NM, Donelson JE, Teixeira SM 2002. *Trypanosoma cruzi*: RNA structure and post-transcriptional control of tubulin gene expression. *Exp Parasitol* 102: 123-133.
- Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL 2005. Messenger RNA processing sites in *Trypanosoma brucei*. *Mol Biochem Parasitol* 143: 125-134.
- Brandão A, Fernandes O 2006. *Trypanosoma cruzi*: mutations in the 3' untranslated region of calmodulin gene are specific for lineages *T. cruzi* I, *T. cruzi* II, and the Zymodeme III isolates. *Exp Parasitol* 112: 247-252.
- Brent MR, Guigo R 2004. Recent advances in gene structure prediction. *Curr Opin Struct Biol* 14: 264-272.
- Coughlin BC, Teixeira SM, Kirchhoff LV, Donelson JE 2000. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3' untranslated region position-dependent cis-element and an untranslated region-binding protein. *J Biol Chem* 275: 12051-12060.
- Di Noia JM, D'Orso I, Sanchez DO, Frasch AC 2000. AU-rich elements in the 3' untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J Biol Chem* 275: 10218-10217.

- D'Orso I, Frasch AC 2001. Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *J Biol Chem* 276: 15783-15793.
- D'Orso I, De Gaudenzi JG, Frasch AC 2003. RNA-binding proteins and mRNA turnover in trypanosomes. *Trends Parasitol* 19: 151-155.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Wortley EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409-15.
- Gopal S, Awadalla S, Gaasterland T, Cross GA. 2005. A computational investigation of kinetoplastid trans-splicing. *Genome Biol* 6: R95.
- Gull K 2001. The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. *Int J Parasitol* 31: 443-452.
- Hoeflich KP, Ikura M 2002. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* 108: 739-742.
- Hughes TA 2006. Regulation of gene expression by alternative untranslated regions. *Trends Genet* 22: 119-122.
- LeBowitz JH, Smith HQ, Rusche L, Beverley SM 1993. Coupling of poly (A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 7: 996-1007.
- Liang XH, Haritan A, Uliel S, Michaeli S 2003. Trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot Cell* 2: 830-840.
- 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 Dev* 8: 491-501.
- Mignone F, Gissi C, Liuni S, Pesole G 2002. Untranslated regions of mRNAs. *Genome Biol* 3: reviews 0004.1-0004.10.
- Mignone F, Grillo G, Licciulli F, Iacono M, Liuni S, Kersey PJ, Duarte J, Saccone C, Pesole G 2005. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 33 Database Issue: D141-6.
- Nozaki T, Cross GA 1995. Effects of 3' untranslated and intergenic regions on gene expression in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 75: 55-67.
- Siegel TN, Tan KS, Cross GA 2005. Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol* 25: 9586-9594.
- Teixeira SM, Kirchhoff LV, Donelson JE 1999. *Trypanosoma cruzi*: suppression of tuzin gene expression by its 5'UTR and spliced leader addition site. *Exp Parasitol* 93: 143-151.
- Webb H, Burns R, Kimblin N, Ellis L, Carrington M 2005. A novel strategy to identify the location of necessary and sufficient cis-acting regulatory mRNA elements in trypanosomes. *RNA* 11: 1108-1116.
- Weston D, La Flamme AC, Van Voorhis WC 1999. Expression of *Trypanosoma cruzi* surface antigen FL-160 is controlled by elements in the 3'untranslated, the 3' intergenic, and the coding regions. *Mol Biochem Parasitol* 102: 53-66.