READERS' OPINION AND DISCUSSION

OPINION

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

David A Campbell[+], Nancy R Sturm

Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, US

The abundance of polycistronic transcription for protein-coding genes in the kinetoplastid protozoa has influenced profoundly the way in which we think about regulating levels of protein expression in eukaryotes. In contrast to most eukaryotes where the primary level of control may be exerted at transcription initiation, kinetoplastids show little regulation at this checkpoint. Genes transcribed at equal rates can yield mRNA(s) that differ in abundance in lifestage- and cell cycle-specific patterns. The most abundant mechanism of control is post transcriptional (Clayton 2002), affecting the stability of the mRNA through interactions of RNA-binding proteins with cis-elements in the 5'- and 3'- untranlslated regions (UTRs) of the mRNA. Regulation of mRNA levels at the level of trans-splicing is indicated by intergenic sequences that do not end up in stable RNA. The perspective by Brandão highlights the wealth of information that remains to be interpreted in the three kinetoplastid genome sequences.

The current rush of activity is focused downstream from the conceptual translations of protein sequence, as these should give insights into the more immediately tangible aspects of the parasites' biology. Broader interest will return inevitably to understanding the basic mechanisms of regulating protein levels that are mediated through signals embedded in the RNA structure. The forward approach to understanding regulated expression lies in the identification of regulatory elements in mRNAs (Mahmood et al. 1999, Boucher et al. 2002) and subsequent purification of binding activities (Mittra & Ray 2004). The 'reverse genetic' approach defines the targets and biological effects of binding by proteins identified in silico. The kinetoplastids possess an abundance of proteins related to RNA-binding proteins (D'Orso et al. 2003, De Gaudenzi et al. 2005, Caro et al. 2006) that are candidates for regulatory functions.

Brandão and Fernandes' approach to analyzing the UTRs of the *Trypanosoma cruzi* calmodulin genes (Brandão & Fernandes 2006) illustrates practical applications of UTR gazing, classification of isolates and a means to predict clinical outcomes of infection with different parasites. The former goal is feasible for defining at least five of the six subdivisions recognized currently in *T. cruzi* (Brisse et al. 2000, Westenberger et al. 2005). The latter goal remains a major and worthy challenge for the clinicians and molecular diagnosticians, for whom the function of the diagnostic gene is not relevant. Differences in the intergenic regions chosen for study are unlikely to show direct correlations with biological patterns. To relate the primary sequence differences to pathogenesis, molecular biologists may have to focus on mRNA whose protein products relate to the multiple steps of entry into the vertebrate and invertebrate hosts, replication in diverse anatomical environments, survival through the onslaught of simple and complex immune systems, and transmission to the next host in the cycle.

How does the field proceed from here? The direct experimental approaches work in both directions, either defining cis elements in RNA with the subsequent identification of the cognate binding protein(s), or bioinformatic identification of RNA-binding protein families followed by the search for their nucleic acid targets. The reverse approach is limited by known homologues in other organisms, and the forward approach has the greatest likelihood of identifying unrecognized RNA-binding proteins unique to the kinetoplastids, and possibly new RNA recognition strategies. The apparent simplicity of protozoa with only two or three morpholigical lifestages does not diminish the enormity of the task ahead. As pointed out by Brandão, a critical early step in the annotation of the completed genomes is the specification of trans-splicing acceptor and polyadenylation sites to define the 5' and 3'-UTRs respectively. In silico attempts to predict these processing sites are a tremendous asset to the community, though ultimately empirical data validating specific UTRs will be necessary. Much of the primary data are lying fallow in EST databases. An ordered gene-by-gene mapping project, like the systematic chromosomal RNAi project in *T. brucei* (Subramaniam et al. 2006), or high-thoughput transcript mapping, as described for trans-splicing in *Caenorhabditis elegans* (Hwang et al. 2004), will provide complete coverage. The widespread use of alternative trans-splicing sites introduces an additional level of complexity to the analysis.

The future of understanding regulated protein expression in trypanosomatids lies in deciphering the inherent codes in the UTRs and the open reading frames for mRNA turnover and destruction, and in identifying the regulatory proteins that signal the fate of the bound RNAs.

[+]E-mail: dc@ucla.edu

Brandão's perspective (Brandão 2006) is a call to the community to annotate the existing kinetoplastid chromosomal sequences with data in existing EST libraries, and to generate new initiatives to define the stable transcriptomes of these important pathogens. The great reduction in costs of the new technologies, such as tiled-oligonucleotide microarrays, will undoubtedly make this task easier and faster than the approaches of nuclease-protection mapping or cDNA sequencing.

REFERENCES

Boucher N, Wu Y, Dumas C, Dube M, Sereno D, Breton M, Papadopoulou B 2002. A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element. *J Biol Chem 277*: 19511-19520.

Brandão A 2006. The untranslated region of genes of *Trypanosoma cruzi*: perspectives for functional characterization of strains and isolates. *Mem Inst Oswaldo Cruz 101*: 775-777.

Brandão A, Fernandes O 2006. *Trypanosoma cruzi*: mutations in the 3' untranslated region of calmodulin gene are specific for lineages *T. cruzi*, *T. cruzi II*, and the zimodeme III isolates. *Exp Parasitol 112*: 247-252.

Brisse S, Barnabé C, Tibayrenc M 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol 30*: 34-44.

Caro F, Bercovich N, Atorrasagasti C, Levin MJ, Vázquez MP 2006. *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. *Exp Parasitol 113*: 112-124.

Clayton C 2002. Life without transcriptional control? From fly to man and back again. *EMBO J 21*: 1881-1888.

D'Orso I, De Gaudenzi JG, Frasch AC 2003. RNA-binding proteins and mRNA turnover in trypanosomes. *Trends Parasitol 19*: 151-155.

De Gaudenzi J, Frasch AC, Clayton C 2005. RNA-binding domain proteins in kinetoplastids: a comparative analysis. *Eukaryot Cell 4*: 2106-2114.

Hwang BJ, Müller HM, Sternberg PW 2004. Genome annotation by high-throughput 5' RNA end determination. *Proc Natl Acad Sci USA 101*: 1650-1655.

Mahmood R, Hines JC, Ray DS 1999. Identification of cis and trans elements involved in the cell cycle regulation of multiple genes in *Crithidia fasciculata. Mol Cell Biol 19*: 6174-6182.

Mittra B, Ray DS 2004. Presence of a poly(A) binding protein and two proteins with cell cycle-dependent phosphorylation in *Crithidia fasciculata* mRNA cycling sequence binding protein II. *Eukaryot Cell 3*: 1185-1197.

Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D, Field MC 2006. Chromosome-wide analysis of gene function by RNA interference in the african trypanosome. *Eukaryot Cell 5*: 1539-1549.

Westenberger SJ, Barnabé C, Campbell DA, Sturm NR 2005. Two hybridization events define the population structure of *Trypanosoma cruzi. Genetics 171*: 527-543.