CellPress

Deciphering RNA regulatory elements in trypanosomatids: one piece at a time or genome-wide?

Vahid H. Gazestani[1], Zhiquan Lu[1], and Reza Salavati[1,2,3]

[1] Institute of Parasitology, McGill University, 21,111 Lakeshore Road, Sainte Anne de Bellevue, Montreal, Quebec H9X3V9, Canada
[2] McGill Centre for Bioinformatics, McGill University, Duff Medical Building, 3775 University Street, Montreal, Quebec H3A2B4, Canada
[3] Department of Biochemistry, McGill University, McIntyre Medical Building, 3655 Promenade Sir William Osler, Montreal, Quebec H3G1Y6, Canada

Morphological and metabolic changes in the life cycle of *Trypanosoma brucei* are accomplished by precise regulation of hundreds of genes. In the absence of transcriptional control, RNA-binding proteins (RBPs) shape the structure of gene regulatory maps in this organism, but our knowledge about their target RNAs, binding sites, and mechanisms of action is far from complete. Although recent technological advances have revolutionized the RBP-based approaches, the main framework for the RNA regulatory element (RRE)-based approaches has not changed over the last two decades in *T. brucei*. In this Opinion, after highlighting the current challenges in RRE inference, we explain some genome-wide solutions that can significantly boost our current understanding about gene regulatory networks in *T. brucei*.

Genomic features of trypanosomatids

As members of a highly divergent group of eukaryotes, trypanosomatids are unique in that non-related genes are constitutively cotranscribed into polycistronic units and are processed into individual mRNAs by a combined *trans*-splicing and polyadenylation reaction [1]. Thus, regulation of gene expression in trypanosomes occurs almost exclusively through post-transcriptional mechanisms. Such processes controlling mRNA localization, turnover, and translation are mediated in a large part by the dynamic interactions of RBPs (see Glossary) with specific subpopulations of mRNAs [2]. The RNA regulon model suggests that a system analogous to prokaryotic DNA operons exists at the RNA level, whereby functionally related messages are regulated by RBPs that bind to short RNA sequences or structural elements known as RREs [3].

Recent studies in trypanosomes have characterized the role of several RBPs in various cellular processes, from cell cycle progression to differentiation between life stages [4]. Pumilio/fem-3 binding factor 9 (PUF9), for example, is necessary for the function of the replicative processes in the early G2 phase of the cell cycle [5]. RBP10 regulates bloodstream form-specific genes, and its depletion causes increases in mRNAs associated with the early stages of differentiation [6]. Overexpression of RBP6 induces transformation from insect-form procyclic cells to infective metacyclic forms [7]. Although an abundance of putative RBPs have been identified in trypanosomes, based on conserved RNA-binding domains, less is known about the RREs that are recognized in the RNA targets.

RBP-mediated regulatory maps in the genome can be studied using two main methods: RBP- and RRE-based approaches. In the RBP-based approach the binding site of a specific RBP is defined using various experimental methods, such as individual nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) and RBP-immunoprecipitation coupled to high-throughput sequencing (RIP-seq) [8,9]. Although these valuable methods have permitted detailed analyses of binding-site information, the procedures are labor-intensive and limited to the study of functional elements for one RBP at a time. In this case, the RBP of choice must be selected based on preliminary knowledge about its requirement or functionality in a biological process of interest. However, in many cases, this type of information is not available: of the more than 150 putative RBPs in trypanosomes [10–12] only a small proportion has been characterized. Consequently, many RREs are also uncharacterized. The RRE-based approach circumvents this issue by trying to identify biologically relevant RREs outside the context of a pre-specified RBP.

Recurrent challenges in studying RREs

Current RRE-based studies in trypanosomatids are mostly limited to gene-based approaches, which search for RREs in the 3′-untranslated region (UTR) of a limited number of genes (usually a single gene) at a time. By contrast, genome-wide approaches search for common regulatory sequences in sets of coregulated genes. The most popular experimental RRE-based approaches in trypanosomatids are composed of serial deletions or mutational analyses on

Corresponding author: Salavati, R. (reza.salavati@mcgill.ca).
Keywords: regulatory networks; RNA binding proteins; RNA regulatory elements; trypanosomes.

1471-4922/$ – see front matter
 http://dx.doi.org/10.1016/j.pt.2014.02.008

Glossary

Actinomycin D: an antibiotic commonly used to inhibit mRNA synthesis; intercalates into DNA and blocks transcription by RNA polymerase.
Alignment-based approach: a computational approach that uses whole-genome multiple alignments to find conserved RREs.
Alignment-free approach: a computational approach that looks for conserved RREs from unaligned sequences by comparing only the functional regions of the 3′-UTRs of orthologous genes between related organisms.
Adenylate/uridylate-rich element (ARE): a sequence of nucleotides located in the 3′-UTRs of mRNAs that is rich in adenosine and uridine bases; generally consists of several AUUUA pentamers, overlapping or within a short distance from each other; regulates the stability of host mRNA.
Bloodstream form: the *Trypanosoma brucei* developmental stage that proliferates extracellularly in the bloodstream of the mammalian host.
Conserved Structural Motif Search Tool (COSMOS): a program that identifies RNA motifs based on the assumption that the regulatory network of RREs is conserved across different species. This alignment-free method searches for linear and structural motifs in the 3′-UTRs of multiple genomes.
Deletion analyses: a method used to narrow down regulatory elements in 3′-UTRs. A series of 3′-UTR deletion fragments are fused to a reporter gene; changes in the expression level of the reporter gene suggest a regulatory region of interest.
Double-stranded RNA-binding domain proteins (DRBD): in trypanosomes, RNA-binding proteins containing two RNA recognition motifs.
EP1 procyclin 26-mer element (EP1 procyclin ARE): a U-rich motif found in the 3′-UTR of EP1 procyclin that affects its stability and translation in bloodstream forms.
Expression-based approach: an approach to finding RREs based on the assumption that the regulatory network of RREs is reflected in the transcriptome; searches in transcriptome data for enriched motifs in sets of coexpressed genes.
Expression site-associated genes (ESAGs): genes that are cotranscribed with variant surface glycoprotein (VSG) genes from telomeric expression sites.
Individual nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP): a method that provides nucleotide resolution of RNA binding sites. RNA is UV cross-linked to RBPs *in vivo*, treated with RNases, and made into partial cDNAs from stalled reverse transcriptase of crosslinked RNA. Sites of crosslinking are then mapped based on the stall sites.
Metacyclic form: the *T. brucei* developmental stage that travels from the salivary gland of the tsetse fly into the mammalian host; cellular division is arrested in preparation for transmission to the mammalian host.
Parallel analysis of RNA structure (PARS): a genome-wide measurement of RNA secondary structure based on deep sequencing of RNA fragments generated with structure specific enzymes.
Procyclic form: the *T. brucei* developmental stage that proliferates in the tsetse fly midgut.
RNA-binding protein immunoprecipitation and high-throughput sequencing (RIP-seq): a method in which the RBP is immunoprecipitated to isolate associated RNAs. Isolated RNAs are reverse transcribed to cDNA and sequenced to identify target genes.
RNA-binding proteins (RBPs): proteins that bind to double- or single-stranded RNA via one or more RNA-binding domains; involved in a wide range of processes from alternative splicing to RNA degradation.
RNA-binding protein purification and identification (RaPID): a method to identify ribonucleoprotein components associated to RNA-aptamer tagged mRNA *in vivo* by mass spectrometry; also allows detection of different RNA species captured in the same ribonucleoprotein complex by RT-PCR.
RNA pulse labeling: a method used to measure RNA decay. Cells are incubated with radioactive RNA precursors. RNA is then isolated and fractionated into newly synthesized (radioactive) and pre-existing (nonradioactive) fractions, which are sequenced and quantified. The mRNA half-lives are calculated based on the ratio of newly synthesized RNA to pre-existing RNA.
RNA regulatory elements (RREs): sequence or structural elements, located predominantly in the 3′-UTR of mRNA, that mediate binding to RBPs.
RNA recognition motif (RRM): an RNA-binding domain comprised of about 90 amino acids that mediates binding to single-stranded RNA.
RNA regulon model: a model in which *trans*-acting factors coordinate the regulation of multiple functionally related transcripts.
Self-transcribing active regulatory region sequencing (STARR-seq): genome-wide fragmentation of genomic DNA with subsequent cloning into reporter vectors to identify enhancers that act as regulators of gene expression.
Upstream open reading frames (uORF): short open reading frames located upstream of coding sequences with a role in translational regulation of transcripts.

the 3′-UTRs of target genes to narrow-down gradually the regions containing regulatory elements [13–19]. The effect on the expression of the target gene or a reporter construct is measured following truncations or random deletions of 3′-UTR regions. Although deletion analyses can yield a stretch of nucleotides in which regulatory elements reside, they fall short of defining the exact consensus sequence of RREs. For example, MacGregor and Matthews narrowed-down gene regulatory signals responsible for the gene expression of *PAD1* ('protein associated with differentiation', carboxylate transporter) to distal and proximal repression elements in the 3′-UTR, but were not able to identify the sequence or structure of the regulatory signals [15]. In some cases, the same approach led to the identification of a relatively short regulatory element (between 16 and 34 nt) in the 3′-UTRs of *Trypanosoma brucei* [14,16,18]. However, a large-scale study on RBPs in a wide range of organisms has indicated that RBPs usually bind to even shorter sequence patterns, around 7 nt in length, and show some level of degeneracy in their consensus sequences [20]. In addition, crystal structures have confirmed that the actual lengths of RNA sequences recognized by individual RNA-binding domains span only a few nucleotides: 4 nt for a CCCH zinc finger [21], 1 nt for each pumilio repeat [22], and 2–8 nt for an RNA recognition motif (RRM) [23]. Evidence suggests that this might also be the case in trypanosomatids. For example, Archer *et al.* identified a 7 nt motif that maintains cell cycle-dependent expression of a PUF9 target gene [5]. Another 8 nt RRE within the EP1 procyclin 26-mer element putatively targets a wide range of *T. brucei* transcripts [13].

Given that coexpressed genes may contain common regulatory motifs, gene-based approaches could uncover regulatory networks beyond the gene of interest. However, this can be complicated by the presence of higher-order structures and multiple regulatory regions in the 3′-UTR. The search for single regulatory motifs is further hindered by the notion that RREs are generally 10 nt or less and that trypanosomatid 3′-UTRs average 400 nt. This gives a virtually endless number of combinatorial possibilities for deletion experiments. In the case of a 34 nt regulatory region identified in the 3′-UTR of expression site-associated gene, *ESAG9-EQ*, which controls its expression in the bloodstream form, the regulatory element was not common to other developmentally coregulated transcripts of the ESAG9 family [14]. Without high resolution mapping of the elements present in each regulatory region, or the possibility of extending a regulatory element found in one gene to other coregulated genes of the same family or function, it is difficult for gene-based approaches to contribute to the understanding of regulatory networks.

Alternative genome-wide approaches

Genome-wide mapping of RREs is essential for a systems-level understanding of the *T. brucei* gene regulatory network. The availability of genome sequence for many organisms, together with current advances in high-throughput technologies, especially deep sequencing, have revolutionized the field and led to the development of various accurate, genome-wide experimental strategies for finding functional elements in the genome [24–26]. Genome sequencing has also contributed to the development of various powerful computational tools for comprehensive annotation of RREs [27–30]. Computational approaches,

when accurate, are extremely beneficial, mainly because they need minimal experimental requirements for the validation of predictions. A large-scale study has shown that the results of these computational approaches are highly reliable, and the newly predicted motifs can be recognized by RBPs in the genome [31]. In addition, several genome-wide studies have led to the successful prediction of RREs in *T. brucei*, reflecting the potential power of these approaches in this organism [32,33]. We discuss here some of the recently developed genome-wide computational and experimental methods for predicting RREs, and argue that they can be used as reliable starting points for understanding the mechanisms that underlie gene regulation in *T. brucei*.

Comparative genomics

Several computational approaches seek conserved RREs in sets of orthologous genes by assuming that part of the regulatory circuits are conserved among closely related organisms [34]. These approaches can be categorized into two different classes: alignment-based and alignment-free approaches. Alignment-based approaches try to find conserved regions in the 3′-UTRs by performing whole-genome multiple alignments. Studies have shown that these conserved regions are enriched for the RREs [30,35]. To test the efficiency of this approach in trypanosomatids, we extracted whole-genome multiple alignments from TriTrypDB version 5 [36], considering 16 different trypanosomatid organisms. The results indicated that the 3′-UTRs in these organisms are highly diverged over the course of evolution (Figure 1). Repeating the same analysis considering only eight trypanosome organisms did not change the overall multiple alignment performance. The reliability of alignment-based approaches is highly dependent on the quality of the underlying multiple alignments. This is because the functional sites are very small (usually less than 10 nt) compared with the total 3′-UTR length (median length of 400 nt in *T. brucei* [37]). Therefore, highly diverged 3′-UTRs can easily lead to erroneous multiple alignments and, consequently, lack of identification of functional elements. To alleviate this issue, alignment-free approaches have been developed [33,38]. These approaches examine whether orthologous genes in related organisms tend to have a specific motif, disregarding the location of the motif's occurrence (Figure 2). Based on this concept, a novel approach identified 222 linearly- and 166 structurally conserved RREs. Some of the predicted motifs (both structural and linear) overlap with previously known RREs in *T. brucei*, reflecting the conservation of many regulatory interactions among trypanosomatids [33].

Expression-based approaches

The underlying assumption in expression-based approaches is that regulatory interactions are reflected in the whole-genome transcriptome data. Based on this assumption, these approaches seek enriched motifs in sets of coordinately expressed genes [27,28] (Figure 3). The advantage of expression-based approaches over the comparative-based approaches is that the former is organism-specific and does not need to assume the conservation of gene regulatory networks in a set of organisms. Relaxation of this constraint is beneficial because biological studies indicate huge differences among trypanosomatids in terms of host, lifestyle, and

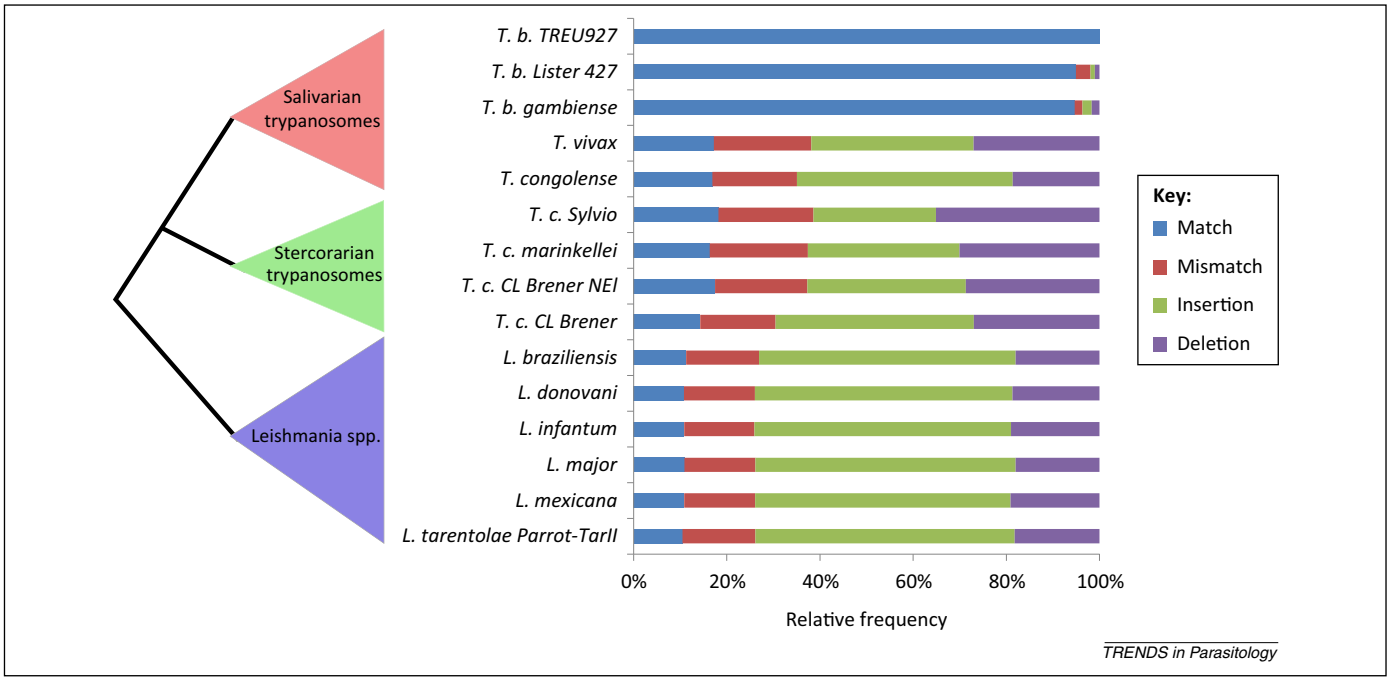

Figure 1. Conservation analysis of 3′ untranslated regions (3′-UTRs) among 16 trypanosomatids. The 3′-UTRs of the *T. brucei TREU927* genome were defined as the median length reported in [37]. For genes with no identified 3′-UTRs, 400 nt downstream of the translational stop codon were chosen as the 3′-UTRs. Considering *Trypanosoma brucei TREU927* as the reference genome, whole-genome multiple sequence alignments corresponding to the 3′-UTRs were extracted from TritrypDB version 5. To improve the accuracy of multiple sequence alignments, the extracted regions were realigned using the ClustalO program [67]. Although 3′-UTR conservation among *T. brucei* sub-species (*T. brucei TREU927, T. brucei Lister 427, T. brucei gambiense*) is very high, alignment of the 3′-UTRs of *T. brucei TREU927* with other trypanosomatids revealed that the 3′-UTRs are poorly conserved, with many insertions and deletions in these regions. However, as expected, the 3′-UTRs of *T. brucei* show higher similarity to other trypanosome organisms (*Trypanosoma vivax, Trypanosoma congolense*, and *Trypanosoma cruzi*) compared with those of *Leishmania*. The phylogenetic relationships were extracted from [68].

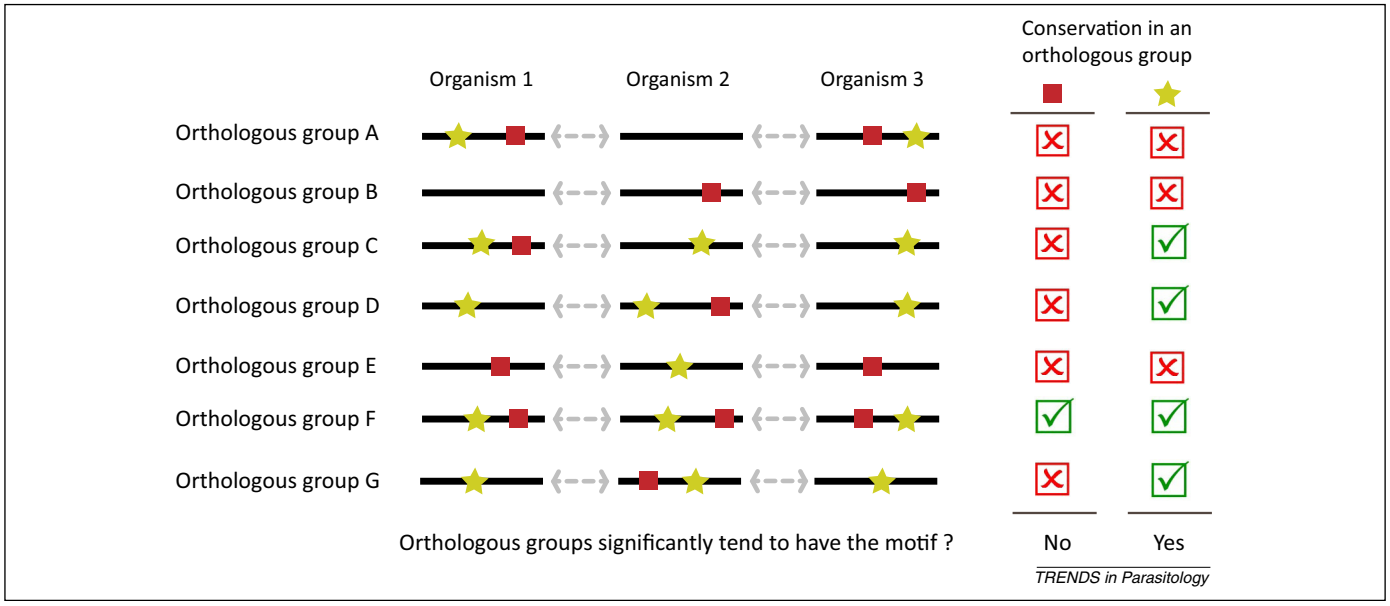

Figure 2. Conceptual representation of alignment-free comparative genomics approaches. After grouping orthologous genes together (grouped via two-sided arrows in this figure), alignment-free approaches search for conserved RNA regulatory elements (RREs) within the 3′ untranslated regions (3′-UTRs), regardless of their position. The gene regulatory network conservation assumption implies that each orthologous group shares a common set of RREs. Therefore, we expect to observe a significant number of orthologous groups that contain a functional motif (represented by a yellow star). By contrast, the distribution for non-conserved sequences is expected to be random (represented by a red square).

developmental stages. Supported by both principle [39] and experiment [40], gene regulatory network rewiring serves as a major source of evolutionary innovation [41]. Interestingly, binding preference is conserved among orthologous transcription factors [42] and RBPs [20]. Therefore, the most likely source of divergence in gene regulatory networks is gain or loss of functional binding sites in the genome [43]. Thus, only looking at the conserved regulatory map utilizing comparative genomics approaches will lose a considerable number of RREs.

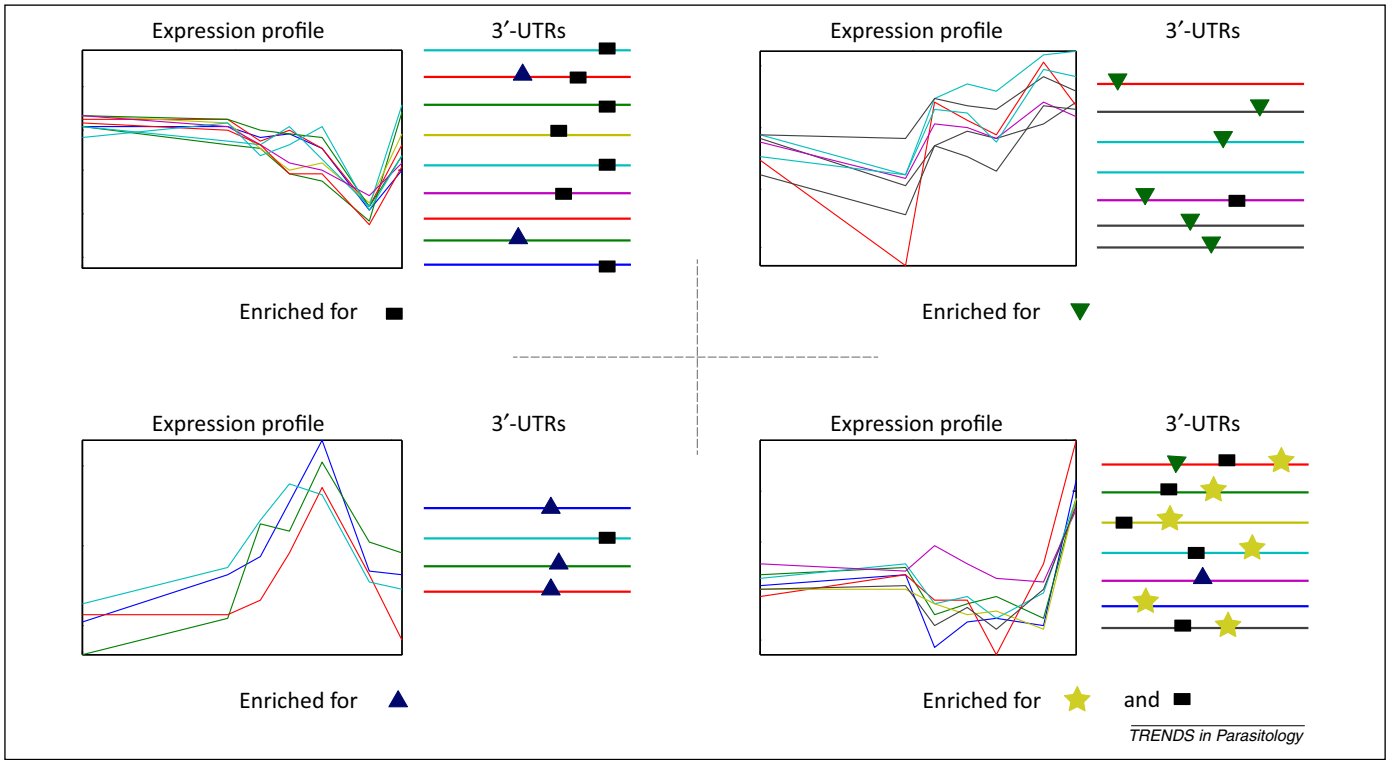

Figure 3. Schematic representation of the framework employed in most expression-based approaches. Although some expression-based approaches make predictions based on one single cell state [29,69], most of these approaches make predictions by specifying clusters of genes that are coregulated with each other in a wide range of conditions [27,28]. To find these clusters, the latter approaches benefit from the fact that the coregulation of a set of genes will lead to their coexpression. Therefore, the regulatory regions of coexpressed genes are enriched for the binding site(s) of regulator(s). Regardless of the details, as illustrated in this figure, the coexpression-based approaches search for motifs (represented as star, rectangle, and triangle) that are over-represented in sets (four clusters in this figure) of coherently expressed genes. The expression pattern and 3′ untranslated regions (3′-UTRs) of genes are represented as colored lines.

One variant of expression-based approaches was successfully applied to identify RREs involved in the cell cycle progression of *T. brucei* [44]. However, there have been many more successful applications of these approaches in other organisms. One of the limitations of these approaches, which have hampered their wider applications to *T. brucei*, is their need for comprehensive transcriptome data for the inference. To address this issue, RREs were identified computationally by integrating three available transcriptome datasets from *T. brucei*. This analysis led to the prediction of 14 significant RREs [45]. Comparison of predicted RREs with current experimental data [20] indicated that three of the predicted RREs were significantly similar to the experimental motifs. Although limited, this comparison suggested that the integration procedure was successful to some extent. The power of expression-based approaches is highly dependent on the expression data that is used for inference. Ideally, the expression data should have sufficient resolution for discriminating various gene regulatory circuits from each other. Naturally, this amount of data can be obtained by extensive cell perturbation studies in different life stages and detailed temporal monitoring of gene fluctuations during the developmental processes.

In addition, genome-wide proteomics studies provide evidence on extensive regulation of genes in steps other than transcript stability, such as translational rate and protein decay rate [46–48]. Additionally, a recent ribosome profiling study highlighted the crucial role of translation regulation in the *T. brucei* gene regulatory program [49]. Although part of translational regulation is governed by upstream open reading frames (uORFs) in the 5′-UTR regions, some reports indicated the role of RBPs and associated RREs in this process [2]. To infer such regulatory maps based on the expression-based approaches, more high-resolution and genome-wide data from both proteomics and ribosome profiling techniques are needed.

To circumvent the need for a comprehensive transcriptome dataset, an elegant study in humans sought to identify RREs by monitoring mRNA decay rates at only 15 different time points [27]. Intriguingly, an independent study confirmed that the regulatory effect of many RBPs could be predicted correctly based on this dataset [20], demonstrating that the regulatory network was truly reflected in the transcriptome data. The key feature that made the above-mentioned study possible was the newly developed technique, RNA pulse labeling [50]. In contrast to the traditional approaches based on chemical treatment (i.e., actinomycin D), which have severe side effects, pulse labeling is a non-disruptive approach that allows the decay rate to be measured without triggering stress responses in the cell [51]. Considering the success of the above-mentioned approach in identifying experimentally validated motifs, it is not inconceivable to expect that the application of similar computational approaches to whole-genome measurements of RNA decay rates in different life stages of *T. brucei* would lead to the identification of stage specific RREs.

STARR-seq

Recently, a genome-wide approach, termed self-transcribing active regulatory region sequencing (STARR-seq), has been developed to identify functional regions in the genome [26]. In this approach, which was originally developed to identify functional enhancers in the genome, a library of genomic sequences is created by random shearing of a genome into relatively small fragments of several hundreds of nt in length. The fragments are then inserted into the 3′-UTR of a reporter gene. The authors showed that, because the functions of enhancers are mostly independent of their location in the genome, the regulatory effect of each inserted region will be reflected in the expression level of the cognate reporter gene, which is monitored by sequencing techniques. This approach has several advantages over the expression-based approaches. For example, because of its design, this approach can search the complete genome for finding RREs whereas expression-based approaches usually search only in the 3′-UTRs. Recent CLIP-seq data in *T. brucei* demonstrated that RREs can reside inside the coding sequence [52]. Besides, unlike the motifs predicted by expression-based approaches, there is no need to use a secondary tool to discriminate the functional instances of the motif from the non-functional fraction. Application of a similar but adapted approach in different life stages of *T. brucei* could help to create context-specific maps of functional RREs.

Profiling RNA secondary structure

Accumulating evidence highlights the importance of RNA secondary structure on its function and regulation [53]. Some RBPs can recognize and bind to specific secondary structures in RNA targets [54,55]. Additionally, a large-scale analysis of human RBPs has revealed that RBPs usually bind to single-stranded regions in the transcripts [56]. These findings are also supported in *T. brucei*. Although some RREs are associated with conserved secondary structures, the others are mostly functional in the single-stranded regions. Besides, fluctuation in temperature can lead to changes in the RNA secondary structure and, consequently, activation of some RREs in *Leishmania* [57]. Advances in sequencing technology have allowed the development of several genome-wide approaches for profiling the RNA secondary structures [58,59]. These approaches benefit from the fact that some RNases are able to distinguish between single- and double-stranded RNA. For example, in one of these approaches, termed parallel analysis of RNA structure (PARS), the transcriptome library is treated separately with two distinct RNases that show preferential digestion toward either the single- or double-stranded RNA. After sequencing the treated samples, the structure for each nucleotide is determined by comparing its frequency in the single-stranded with the double-stranded sample. Although these approaches cannot capture the dynamic structure of RNA under *in vivo* conditions, they can enhance our knowledge of the structural properties of the *T. brucei* transcriptome.

Validating RRE predictions

Predicted RREs must be followed up with experimental and computational analyses not only to confirm the functionality but also to help to uncover the RBP related to each RRE.

To achieve this goal one may need to consider all possible sources of evidence for prediction of RREs. For example,

application of a novel alignment-free approach, the Conserved Structural Motif Search Tool (COSMOS), identified 388 potentially conserved RREs [33]. Subsequent use of the available expression data led to the identification of those RREs that were more likely to contribute to mRNA abundance and stability. This procedure limited the number of candidates to 35 high-confidence RREs among the 388 elements. For experimental validation of newly found motifs, the authors selected a highly conserved adenylate/uridylate-rich element (ARE), AUUUAUU, which was predicted to be functional by both conservation and expression criteria [33]. Downstream bioinformatics analysis revealed that the predicted motif by COSMOS was generally associated with transcripts upregulated in late insect stages and downregulated in mammalian bloodstream stages, suggesting a regulatory role similar to the EP1 procyclin ARE [60,61] but in contrast to another ARE involved in heat-shock response [62]. Further bioinformatics analyses uncovered three possible *trans*-acting factors that may be involved in the regulation of this ARE: RBP6 and the double-stranded RNA-binding domain proteins DRBD12 and DRBD13 [33]. Experiments that followed these predictions provided support for the functional interaction of these three proposed RBPs with the motif [33]. In particular, expression of each protein was inhibited by RNA interference (RNAi) and activated by overexpression. Consistent effects were observed in these two types of experiments; target transcripts that were downregulated after protein inhibition were upregulated after protein activation, and vice versa, suggesting specific effects of inhibition/activation. These experiments revealed a stabilizing role for DRBD13 and a destabilizing role for RBP6 and DRBD12. Additionally, RBP immunoprecipitation followed by sequencing (RIP-Seq) confirmed that one of the candidate proteins, DRBD13, is associated with trypanosomatid ARE-containing transcripts *in vivo* [33]. Alternatively, associated regulatory protein factors can be identified using RREs as ligands to pull down ribonucleoprotein complexes. Experimental approaches for isolating associated proteins involve the exogenous expression of transcripts consisting of small RNA tags, or aptamers, fused to the RNA sequence of interest [63,64]. These RNA aptamers are then recognized and pulled down by a protein or small molecule that recognizes the aptamer. For example, the recently developed technique of RNA-binding protein purification and identification (RaPID) exploits the specific interaction between the *E. coli* bacteriophage coat protein (MS2-CP) and its cognate RNA. The MS2-CP is fused to a streptavidin-binding protein tag, which allows affinity purification using streptavidin-conjugated beads and subsequent analysis by mass spectrometry [65]. A similar method successfully identified several known and novel ARE-binding proteins in mammalian cells using a streptavidin-binding aptamer fused to mRNA containing the ARE of mouse tumor necrosis factor α [66].

Concluding remarks

The essential point in understanding the biology of trypanosomatids is uncovering the mechanisms by which the coordination of RREs and RBPs orchestrate the expression of target genes. Large-scale identification of RREs and their cognate protein factors would contribute to the construction of gene regulatory networks. Functional characterization of these networks and their role in cellular processes is essential for ultimately identifying points at which these networks can be manipulated pharmacologically to interfere with parasite development and transmission.

Acknowledgments

Research in the laboratory of R.S. is supported by the Canadian Institutes of Health Research (CIHR) grant 119464 and the Natural Sciences and Engineering Research Council of Canada (NSERC) grant 328186. V.H.G. is supported by a CIHR Systems Biology Fellowship.

References

1 Kramer, S. (2012) Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. *Mol. Biochem. Parasitol.* 181, 61–72
2 Kramer, S. and Carrington, M. (2011) Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol.* 27, 23–30
3 Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* 8, 533–543
4 Clayton, C. (2013) The regulation of trypanosome gene expression by RNA-binding proteins. *PLoS Pathog.* 9, e1003680
5 Archer, S.K. *et al.* (2009) *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog.* 5, e1000565
6 Wurst, M. *et al.* (2012) Expression of the RNA recognition motif protein RBP10 promotes a bloodstream-form transcript pattern in *Trypanosoma brucei*. *Mol. Microbiol.* 83, 1048–1063
7 Kolev, N.G. *et al.* (2012) Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science* 338, 1352–1353
8 Konig, J. *et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915
9 Zhao, J. *et al.* (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40, 939–953
10 Caro, F. *et al.* (2006) *Trypanosoma cruzi*: analysis of the complete PUF RNA-binding protein family. *Exp. Parasitol.* 113, 112–124
11 De Gaudenzi, J. *et al.* (2005) RNA-binding domain proteins in kinetoplastids: a comparative analysis. *Eukaryot. Cell* 4, 2106–2114
12 Kramer, S. *et al.* (2010) Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*. *BMC Genomics* 11, 283
13 Mayho, M. *et al.* (2006) Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res.* 34, 5312–5324
14 Monk, S.L. *et al.* (2013) A short bifunctional element operates to positively or negatively regulate ESAG9 expression in different developmental forms of *Trypanosoma brucei*. *J. Cell Sci.* 126, 2294–2304
15 MacGregor, P. and Matthews, K.R. (2012) Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in *Trypanosoma brucei*. *Nucleic Acids Res.* 40, 7705–7717
16 Hehl, A. *et al.* (1994) A conserved stem–loop structure in the 3′ untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 370–374
17 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
18 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, 3017–3026
19 Jefferies, D. *et al.* (1991) Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression at the posttranscriptional level. *Mol. Cell. Biol.* 11, 338–343
20 Ray, D. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177

21 Hudson, B.P. *et al.* (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.* 11, 257–264
22 Wang, X. *et al.* (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110, 501–512
23 Clery, A. *et al.* (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.* 18, 290–298
24 Zinzen, R.P. *et al.* (2009) Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462, 65–70
25 Lee, D. *et al.* (2011) Discriminative prediction of mammalian enhancers from DNA sequence. *Genome Res.* 21, 2167–2180
26 Arnold, C.D. *et al.* (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339, 1074–1077
27 Goodarzi, H. *et al.* (2012) Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* 485, 264–268
28 Elemento, O. *et al.* (2007) A universal framework for regulatory element discovery across all genomes and data types. *Mol. Cell* 28, 337–350
29 Foat, B.C. and Stormo, G.D. (2009) Discovering structural cis-regulatory elements by modeling the behaviors of mRNAs. *Mol. Syst. Biol.* 5, 268
30 Xie, X. *et al.* (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* 434, 338–345
31 Hu, S. *et al.* (2009) Profiling the human protein–DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 139, 610–622
32 Archer, S.K. *et al.* (2011) The cell cycle regulated transcriptome of *Trypanosoma brucei*. *PLoS ONE* 6, e18425
33 Najafabadi, H.S. *et al.* (2013) Global identification of conserved post-transcriptional regulatory programs in trypanosomatids. *Nucleic Acids Res.* 41, 8591–8600
34 Meireles-Filho, A.C. and Stark, A. (2009) Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information. *Curr. Opin. Genet. Dev.* 19, 565–570
35 Liu, Y. *et al.* (2004) Eukaryotic regulatory element conservation analysis and identification using comparative genomics. *Genome Res.* 14, 451–458
36 Aslett, M. *et al.* (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 38, D457–D462
37 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
38 Chan, C.S. *et al.* (2005) Revealing posttranscriptional regulatory elements through network-level conservation. *PLoS Comput. Biol.* 1, 564–578
39 Carroll, S.B. (2005) Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245
40 Kunarso, G. *et al.* (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634
41 Weirauch, M.T. and Hughes, T.R. (2010) Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* 26, 66–74
42 Noyes, M.B. *et al.* (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 133, 1277–1289
43 Wittkopp, P.J. and Kalay, G. (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69
44 Queiroz, R. *et al.* (2009) Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics* 10, 495
45 Shateri Najafabadi, H. and Salavati, R. (2010) Functional genome annotation by combined analysis across microarray studies of *Trypanosoma brucei*. *PLoS Negl. Trop. Dis.* 4
46 Gunasekera, K. *et al.* (2012) Proteome remodelling during development from blood to insect-form *Trypanosoma brucei* quantified by SILAC and mass spectrometry. *BMC Genomics* 13, 556
47 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, 172–179
48 Urbaniak, M.D. *et al.* (2012) Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages. *PLoS ONE* 7, e36619
49 Vasquez, J.J. *et al.* (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt1386
50 Friedel, C.C. *et al.* (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res.* 37, e115
51 Schwanhausser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
52 Das, A. *et al.* (2012) The essential polysome-associated RNA-binding protein RBP42 targets mRNAs involved in *Trypanosoma brucei* energy metabolism. *RNA* 18, 1968–1983
53 Wan, Y. *et al.* (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655
54 Oberstrass, F.C. *et al.* (2006) Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat. Struct. Mol. Biol.* 13, 160–167
55 Tadros, W. *et al.* (2007) SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev. Cell* 12, 143–155
56 Li, X. *et al.* (2010) Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA* 16, 1096–1107
57 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
58 Kertesz, M. *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107
59 Underwood, J.G. *et al.* (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001
60 Quijada, L. *et al.* (2002) Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* increases the abundance of mRNAs containing AU-rich regulatory elements. *Nucleic Acids Res.* 30, 4414–4424
61 Haile, S. *et al.* (2003) A role for the exosome in the in vivo degradation of unstable mRNAs. *RNA* 9, 1491–1501
62 Droll, D. *et al.* (2013) Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathog.* 9, e1003286
63 Walker, S.C. *et al.* (2008) RNA affinity tags for the rapid purification and investigation of RNAs and RNA–protein complexes. *Methods Mol. Biol.* 488, 23–40
64 Windbichler, N. and Schroeder, R. (2006) Isolation of specific RNA-binding proteins using the streptomycin-binding RNA aptamer. *Nat. Protoc.* 1, 637–640
65 Slobodin, B. and Gerst, J.E. (2010) A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. *RNA* 16, 2277–2290
66 Leppek, K. and Stoecklin, G. (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* 42, e13
67 Sievers, F. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539
68 Hamilton, P.B. *et al.* (2012) Parasites reveal movement of bats between the New and Old Worlds. *Mol. Phylogenet. Evol.* 63, 521–526
69 Foat, B.C. *et al.* (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17675–17680