Biochemical and Biophysical Research Communications 382 (2009) 30–34

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

# Proteomic analysis of the *Trypanosoma cruzi* ribosomal proteins

Maximiliano Juri Ayub [a,b], James Atwood [c], Arthur Nuccio [c], Rick Tarleton [d], Mariano J. Levin [a,*]

[a] Laboratorio de Biología Molecular de la Enfermedad de Chagas, LaBMECh, Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI-CONICET), Buenos Aires, Argentina
[b] Laboratorio de Biología Molecular, Cátedra de Ingeniería Genética, Universidad Nacional de San Luis, Avda. Ejército de los Andes 950, 5700 San Luis, Argentina
[c] Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, USA
[d] Center for Tropical and Emerging Global Diseases and Department of Cellular Biology, University of Georgia, Athens, Georgia, USA

ARTICLE INFO

*Article history:*
Received 7 February 2009
Available online 24 February 2009

*Keywords:*
*Trypanosoma cruzi*
Protein synthesis
Mass spectrometry
Ribosome

ABSTRACT

*Trypanosoma cruzi* is a parasite responsible for Chagas disease. The identification of new targets for chemotherapy is a major challenge for the control of this disease. Several lines of evidences suggest that the translational system in trypanosomatids show important differences compared to other eukaryotes. However, there little is known information about this. We have performed a detailed data mining search for ribosomal protein genes in *T. cruzi* genome data base combined with mass spectrometry analysis of purified *T. cruzi* ribosomes. Our results show that *T. cruzi* ribosomal proteins have ~50% sequence identity to yeast ones. Nevertheless, some parasite proteins are longer due to the presence of several N- or C-terminal extensions, which are exclusive of trypanosomatids. In particular, L19 and S21 show C-terminal extensions of 168 and 164 amino acids, respectively. In addition, we detected two 60S subunit proteins that had not been previously detected in the *T. cruzi* total proteome; namely, L22 and L42.


## Introduction

*Trypanosoma cruzi* is a protozoan kinetoplastid parasite responsible for Chagas disease. This is a widespread disease in Latin America, where about 18 million people are infected and around 100 million inhabit endemic areas [1]. No vaccines are available at present, and drugs used for treatment show undesirably side effects. The identification of new targets for chemotherapy is a major challenge in the control of this disease.

Recently, the sequencing of the genome of this parasite has been completed. Moreover, a proteomic analysis of the different stages of *T. cuzi* has been carried out, providing important evidences of stage-specific expression of numerous proteins [2]. The approach is an invaluable tool for the study of global pattern of gene expression. The application of this method in trypanosomatids is particularly important because these organisms do not use transcription initiation as a regulatory step to control gene expression.

Trypanosomatids are also among those few lower eukaryotic organisms that possess unusually structured mRNAs, as products of a special transsplicing RNA editing mechanism in post-transcription [3–5]. As a result, every trypanosome mRNA has a 39 nucleotide long identical sequence at the 5′ terminus, named spliced leader (SL). Besides the universally conserved 7-methyl guanosine cap, the first four nucleotides of the 39-nt SL are all methylation-modified. This unusually modified cap, known as the cap-4 structure [6–9], is the most modified cap structure of all eukaryotic cells. The functionality of the specially constructed mRNAs in trypanosomatids and the reason for their occurrence in only a few organisms are as yet unknown. Counting on the important role of the mRNA 5′ end during eukaryotic translation initiation [10], the SL sequence and its associated cap structure are expected to play a significant part in translation initiation, and possibly to be involved in a unique mechanism of initiation. In kinetoplastid *Leishmania*, the cap-4 structure and the first two thirds of the SL sequence have been found to be critical for the association of mRNAs with polysomes [11]. More recently, we have shown that *T. cruzi* ribosomes are highly resistant to Trichosanthin, a ribosome inactivating protein that strongly inhibits mammalian ribosomes [12,13].

All these observations suggest the importance of studying the ribosomal particle in trypanosomatids, a field that has been poorly investigated up to date. In the present work we performed a detailed data mining of the putative ribosomal proteins from *T. cruzi*, complemented with MS analysis of highly purified ribosomes. Several parasite-specific features were found, which could be of great interest for the identification of novel chemotherapy targets, as well as for studies concerned with the evolution of eukaryotic cells.

* Corresponding author.
*E-mail address:* mlevin@dna.uba.ar (M.J. Levin).

0006-291X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2009.02.095

## Materials and methods

*Data mining of T. cruzi ribosomal protein genes*. We used the *Saccharomyces cerevisiae* ribosomal proteins sequences (available from the site http://bo.expasy.org/cgi-bin/lists?ribosomp.txt) as probes for searching in the *T. cruzi* genome database (www.genedb.org/genedb/tcruzi) using the TBLASTN option. Multiple alignments were performed using the CLUSTALW program.

*Ribosome purification*. Ribosomes were purified from *T. cruzi* epimastigotes as described [14]. This protocol has shown to yield highly homogeneous and pure ribosomes as demonstrated by cryo EM images [14]. Proteins were specifically extracted from this preparation using Trizol reagent (GIBCO BRL) following the manufacturer instructions and dried under vacuum.

*Preparation of ribosome peptides*. The ribosomal proteins were dissolved in 50 mM $NH_4HCO_3$ and reduced (25 mM DTT, pH 8.5) for 30 min at 55 °C followed by carboxyamidomethylation with iodoacetamide (90 mM) for 1 h at room temperature in the dark. The samples were then digested overnight at 37 °C with sequencing grade porcine trypsin (1:50, Promega, Madison, WI). The digest was dried under vacuum, then reconstituted in 100 μl of 0.1% TFA, and filtered prior to separation.

*First dimension reverse phase chromatography (RP)*. The peptides mixture was injected onto a 1× 150 mm Zorbax $C_{18}$ column (Agilent, Palo Alto, CA) at a flow rate of 50 μl/min, and separated into 10 fractions by reverse phase chromatography as previously described [15]. Each fraction was dried under vacuum and resuspended in 8 μl of 0.1% formic acid for LC–MS/MS analysis.

*RPLC–MS/MS*. Each fraction was analyzed independently using a Agilent CapLC (Agilent, Palo Alto, CA) interfaced directly to a linear ion trap mass spectrometer (LTQ) (Thermo Electron, Waltham, MA). Mobile phase A and B were $H_2O$/0.1% formic acid and ACN/0.1% formic acid, respectively. Five microliters of each fraction was loaded onto the column (15 cm × 150 μm, Microtech, Vista, CA) at a flow rate of 1 μl/min. In order to maximize separation, RP gradients were designed to correspond with the percentage of organic at which each peptide fraction eluted from the first dimension RP as previously described [15].

*MS/MS parameters*. The mass spectrometer was operated in the data-dependent acquisition mode with dynamic exclusion enabled for 160 s. The instrument was set to acquire a MS scan from 500 to 2000 *m*/*z* followed by up to nine MS/MS scans from 50 to 2000 *m*/*z*. Raw mass spectra were processed into .dta format and merged into a single file prior to database searching.

*Database searching*. Mascot (version 1.9) searches were performed against a database of approximately 25,000 *T. cruzi* gene annotations provided by *T. cruzi* Sequencing Consortium (TSK-TSC). The following parameters were used during the database search: a specified trypsin enzymatic cleavage with one possible missed cleavage, peptide tolerance of 500 parts-per-million, fragment ion tolerance of 0.6 Da, variable modifications due to carboxyamidomethylation of cysteines, mono and di methylation of lysine and arginine, N-terminal acetylation, and acetylation of lysine and arginine.

## Results and discussion

Cryo EM of highly purified *T. cruzi* ribosomes has recently shown important differences in comparison with the corresponding organelles of model organisms like *S. cerevisiae* and mammals [14]. Some of them were attributed to large expansions in the primary sequence of the ribosomal RNA molecules. On the other hand, the presence of specific features due to ribosomal proteins was difficult to demonstrate. Therefore, we decided to use the *T. cruzi* genome database to identify all the putative ribosomal proteins of *T. cruzi*. Using the *S. cerevisiae* ribosomal protein sequences [16] as probes, we searched in the *T. cruzi* genome database for homologues genes. Results are shown in Supplementary Tables 1 and 2 for proteins from the large and small subunits, respectively. The average amino acid identity between the *S. cerevisiae* and *T. cruzi* ribosomal proteins was remarkably low (≅50%), taking into account the highly degree of conservation of 80S ribosome through evolution.

Our search revealed that *T. cruzi* 60S ribosomal proteins are, in average, 20 amino acids longer than the corresponding *S. cerevisiae* proteins. For the 40S subunit, *T. cruzi* proteins have an average 10 amino acids longer chain. The extra regions in *T. cruzi* ribosomal proteins are generally at the N- or C-terminal ends. The most intriguing examples of these terminal extensions are TcL19 and TcS21, showing C-terminal extensions of 168 and 164 amino acids, respectively, compared to yeast. Figs. 1 and 2 show sequence alignments of *T. cruzi*, *T. brucei*, *Arabidopsis thaliana*, *S. cerevisiae*, and *Homo sapiens* L19 and S21 proteins, showing that these extensions are only present in kinetoplastids, although its length varied among species. MS analyses of *T. cruzi* ribosomes confirmed the presence of peptides matching to TcL19 and TcS21 (Supplementary Tables 1 and 2, respectively), strongly suggesting that these genes correspond to the functional ribosomal components. The functional meanings of these extensions, as well as the molecular mechanisms that generated them over time, constitute interesting fields for future studies.

It is interesting to note that *S. cerevisiae* L19 protein has been described forming part of the polypeptide chain exit channel [17]. Moreover, several of the ribosomal proteins implicated in the formation of this channel showed important extensions in *T. cruzi* (Table 3). This fact could be related to the absence of a flat surface on this region in *T. cruzi* 80S ribosome [14].

Another interesting observation concerned the L24 protein. Two putative ORF encoding to this protein were found in the *T. cruzi* genome database; TcL24A (Tc00.1047053510101.240) and TcL24B (Tc00.1047053503611.20) (Supplementary Fig. 3). The amino acids identity between these *T. cruzi* ORFs was very low (21.5%). The amino acids identity of *S. cerevisiae* L24 is 27.2% and 27.0% with TcL24A and TcL24B, respectively. These almost identical values do not allow predicting which of these ORFs correspond to the functional homologue of yeast L24. The MS analysis of purified ribosomes confirmed the presence of only one of the putative TcL24 gene product; TcL24B. This observation strongly suggests that TcL24B is the *T. cruzi* functional orthologue of *S. cerevisiae* L24. Because TcL24A was not detected in the ribosome preparation nor in the total *T. cruzi* proteome [2], it may be that this ORF is not expressed in *T. cruzi*. The genes encoding for the corresponding L24A and L24B proteins are also present in *T. brucei* and *Leishmania major* (not shown) showing that the duplication of the L24 gene took place in a common ancestor of trypanosomatids.

A similar case was observed for the S12 protein. Two putative homologue genes showing 65% of amino acids identity; TcS12A (Tc00.1047053506181.59) and TcS12B (Tc00.1047053508231.20), were detected in the *T. cruzi* genome, both of them automatically annotated as S12 putative proteins and detected in the total proteome from *T. cruzi* epimastigotes. TcS12A is slightly closer to the yeast S12 (34.5% of amino acids identity) than TcS12B (30.8% of amino acids identity). Both genes are expressed at the protein level [2] but only TcS12B was detected in the proteomic analysis of purified ribosomes (Supplementary Fig. 4). Interestingly, we have found S12A and S12B genes in *T. brucei* (Tb11.02.1830 and Tb10.6k15.2050, respectively) but only one in *L. major* (LmjF13.0570), suggesting gene duplication into the trypanosomatid lineages.

In other eukaryotes like mammals and yeast, ribosomal proteins S27A and L40 are synthesized as C-terminal fusion to ubiquitin. By

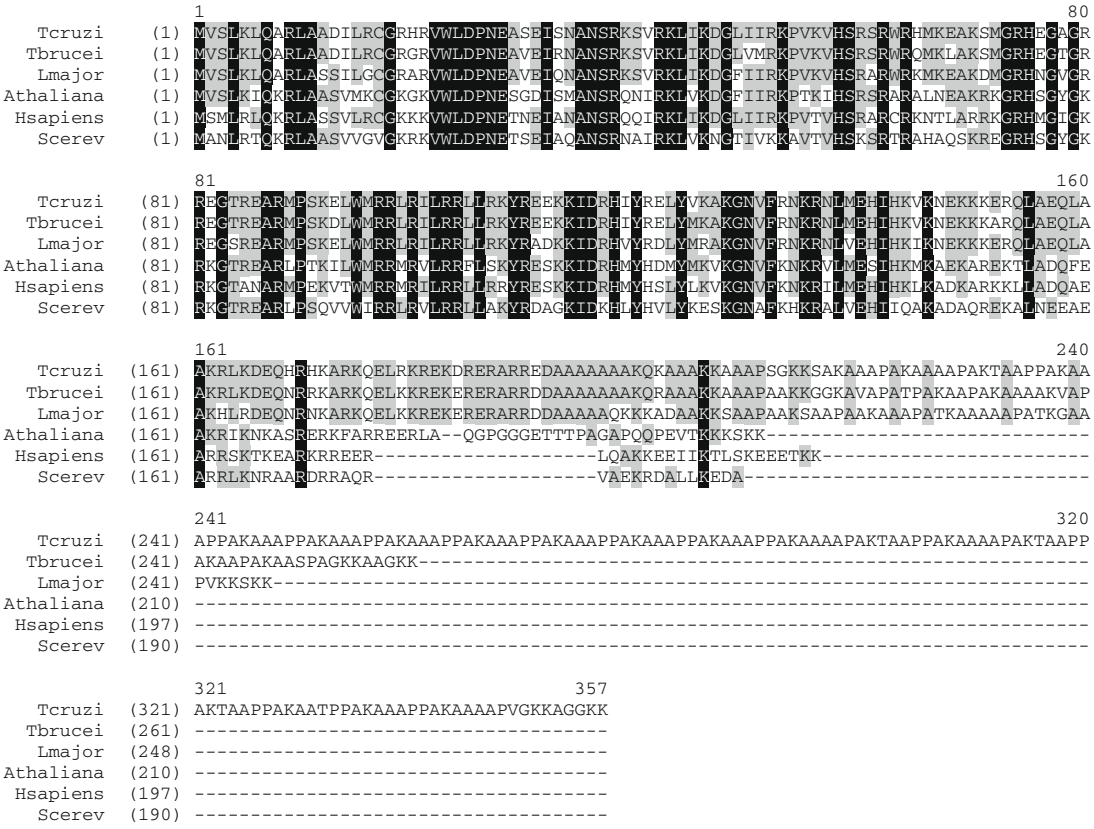


**Fig. 1.** Sequence alignment of L19 proteins from different organisms, including *T. cruzi*.

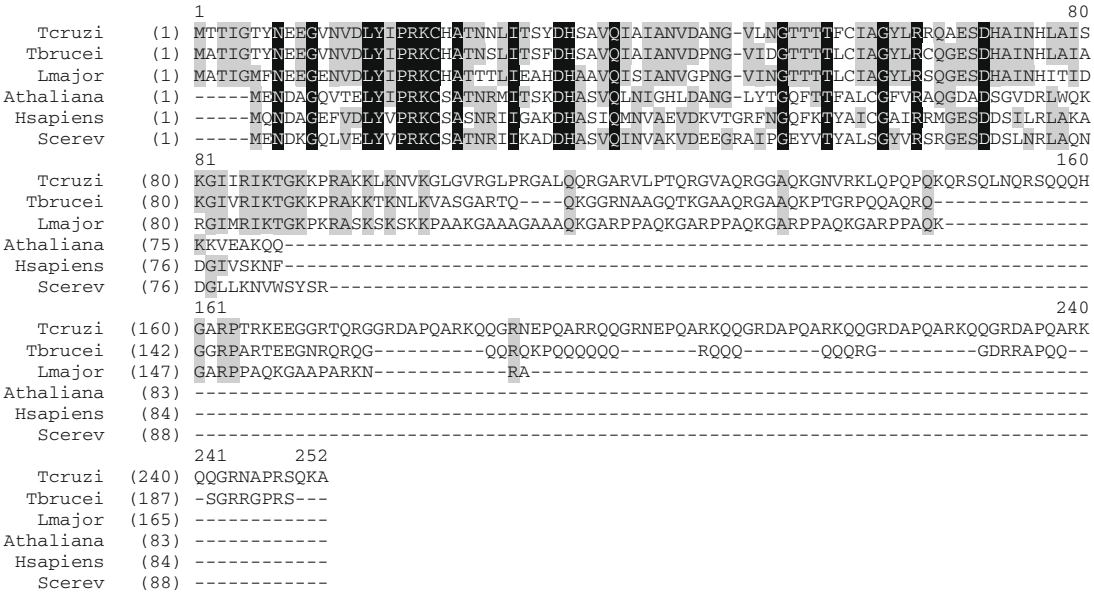


**Fig. 2.** Sequence alignment of S21 proteins from different organisms, including *T. cruzi*.

data mining revealed similar fusion genes in the *T. cruzi* genome (Supplementary Tables 1 and 2). The ubiquitin/L40 gene had not been annotated as ribosomal protein in the *T. cruzi* genome database. From these ubiquitin fusion proteins, only TcS27A was detected by mass spectrometry on pure ribosomes.

The mammalian L28 protein is the only ribosomal component, which had no counterpart in *S. cerevisiae*. Our analysis showed that the *T. cruzi* genome has a putative L28 coding gene, showing 28.7% amino acids identity to human L28. This prediction is supported by the detection of several peptides matching to this sequence (Supplementary Fig. 5 and Supplementary Table 1). The presence of L28 in trypanosomatids and mammals suggests that the corresponding gene has been lost in *S. cerevisiae*. This prediction is strongly supported by the presence of a gene encoding L28 in

**Table 3**
Comparison of ribosomal proteins forming the polypeptide exit channel.

| Protein | Length (amino acids) | | % Identity | *T. cruzi*-specific features |
|---|---|---|---|---|
| | *S. cerevisiae* | *T. cruzi* | | |
| L17 | 184 | 166 | 54 | C-terminal deletion (20 aa) |
| L19 | 189 | 357 | 50 | C-terminal extension (170 aa) |
| L25 | 142 | 226 | 45 | N-terminal extension (90 aa) |
| L26 | 127 | 143 | 57 | C-terminal extension (21 aa) |
| L31 | 113 | 188 | 42 | N-terminal extension (75 aa) |
| L35 | 120 | 127 | 41 | – |

*Aspergillus fumigatus* (Protein ID: XP_747557) and *Schizosaccharomyces pombe* (Protein ID: BAA31554), two organisms phylogenetically close to *S. cerevisiae*.

In summary, we found 32 proteins with significant sequence identity to *S. cerevisiae* 40S proteins. Peptides matching to 29 of these proteins were detected by MS of *T. cruzi* ribosomes, including S13 (Tc00.1047053511291.90) and S28 (Tc00.1047053506413.20), which were not detected in the proteome of *T. cruzi* [2]. Peptides matching to S22 (Tc00.1047053506297.330), S25 (Tc00.1047053509233.190) and S30 were not detected in our analyses. S30 was not detected in MS studies on total extracts of *T. cruzi* [2].

For the large subunit, we have found genes encoding putative homologues for all the 48 yeast proteins, excepting L41 (GenBank code X16065), a short peptide, 25 amino acids long. MS detected these predicted proteins, excepting L1 (Tc00.1047053511277.160), L35 (Tc00.1047053509979.90), L39 (Tc00.1047053511217.145), L40 (Tc00.1047053507483.4), P1 and P2. From these, L1, L35, P1α, P2α, and P2β were previously detected in epimastigote crude extracts [2]. The presence of all the P1 and P2 proteins in our ribosome preparation has previously been confirmed by *Western blot* with monoclonal antibodies [18]. On the other hand, we detected two large subunit proteins that were not previously detected in the *T. cruzi* total proteome; L22 (Tc00.1047053504147.120) and L42 (Tc00.1047053507105.40).

When MS analyses of *T. cruzi* total extracts were performed, 6.19% (72 from 1162) of the identified proteins corresponded to ribosomal proteins [2]. In our ribosome preparation, 45.6% (68 from 149) of the total hits were ribosomal proteins. From the non-ribosomal proteins identified in our sample, several corresponded to abundant proteins like chaperones and cytoskeletal proteins, two putative RNA binding proteins with no significant sequence similarity to any protein with known function (Tc00.1047053511727.290/Tc00.1047053511727.270 and Tc00.1047053510755.120), proteins involved in translation like initiation and elongation factors (IF4A, EF1 alpha, EF1 beta, EF2) as well as RNA helicases, a few mitochondrial proteins and several proteins with unknown function. Supplementary Table 4 shows all the non-ribosomal proteins detected in our sample.

Recently, twelve new *S. cerevisiae* proteins have been described to be associated to ribosomal complexes [19]. We did not detect in our sample any peptide matching to *T. cruzi* proteins with homology to those yeast genes.

The presence of mitochondrial ribosomes co-sedimenting with 80S ribosomes could be a possible contaminant in our preparation. Recently, the mitochondrial ribosomes of *L. tarentolae* have been analyzed by MS, identifying 20 mitochondrial proteins [20]. Our sample showed only one peptide corresponding to one of the *T. cruzi* orthologue proteins (mtS15; Tc00.1047053507949.50), showing that our preparation is virtually free of mitochondrial particles.

It is interesting to note that MS analyses detected RACK1 (Tc00.1047053511211.130), a protein tightly associated to the small ribosomal subunit in eukaryotes, and apparently involved in the regulation of the translation initiation [21,22]. This fact is in apparent conflict with cryo EM of *T. cruzi* ribosomes showing no densities corresponding to RACK1. This inconsistency could be caused by a weaker interaction between RACK1 and the ribosome in *T. cruzi*, comparing to other species. This could results in too low amounts of RACK1 in purified ribosomes for visualization by cryo EM. However, these low amounts could be enough for detection by MS analyses.

Our results gave the confirmation of the expression at the protein level, as well as their association with ribosomal particles, for virtually all the predicted ribosomal proteins. In addition, for those cases where assignation of the right homologue gene could be difficult, our MS data allow to propose which of these genes are expressed and associated with ribosomes. Finally, we have identified peptides matching to four ribosomal proteins, which had not previously detected in MS of *T. cruzi* extracts (TcS13, TcS28, TcL22 and TcL42). In summary, our results give a convenient start point for the study of any ribosomal protein in *T. cruzi*, a field that has been poorly investigated up to date. Some particular cases (e.g., TcS21, TcL19) merit particular attention.

## Acknowledgments

This work was partially supported by grants from Howard Hughes Medical Institute (Chevy Chase, USA), the South-South Initiative of the TDR programme of the World Health Organization, and project PICT 2006 No 2439 of the Agencia Nacional de Promoción de Ciencia y Tecnología, Argentina, to MJL. M.J.A. and M.J.L. are members of the career of scientific investigator of CONICET.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2009.02.095.

## References

[1] Control of Chagas Disease, World Health Organ Tech. Rep. Ser. 905 (2002) i–vi, 1–109, back cover.
[2] J.A. Atwood 3rd, D.B. Weatherly, T.A. Minning, B. Bundy, C. Cavola, F.R. Opperdoes, R. Orlando, R.L. Tarleton, The *Trypanosoma cruzi* proteome, Science 309 (2005) 473–476.
[3] W.J. Murphy, K.P. Watkins, N. Agabian, Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing, Cell 47 (1986) 517–525.
[4] R.E. Sutton, J.C. Boothroyd, Evidence for trans splicing in trypanosomes, Cell 47 (1986) 527–535.
[5] J.A. Walder, P.S. Eder, D.M. Engman, S.T. Brentano, R.Y. Walder, D.S. Knutzon, D.M. Dorfman, J.E. Donelson, The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNA's, Science 233 (1986) 569–571.
[6] J.D. Bangs, P.F. Crain, T. Hashizume, J.A. McCloskey, J.C. Boothroyd, Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides, J. Biol. Chem. 267 (1992) 9805–9815.
[7] M.S. Freistadt, G.A. Cross, H.D. Robertson, Discontinuously synthesized mRNA from *Trypanosoma brucei* contains the highly methylated 5′ cap structure, m7GpppA A C(2′-O)mU A, J. Biol. Chem. 263 (1988) 15071–15075.
[8] M. Lewdorowicz, Y. Yoffe, J. Zuberek, J. Jemielity, J. Stepinski, R. Kierzek, R. Stolarski, M. Shapira, E. Darzynkiewicz, Chemical synthesis and binding activity of the trypanosomatid cap-4 structure, RNA 10 (2004) 1469–1478.
[9] K.L. Perry, K.P. Watkins, N. Agabian, Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader, Proc. Natl. Acad. Sci. USA 84 (1987) 8190–8194.
[10] T. Preiss, M.W. Hentze, Starting the protein synthesis machine: eukaryotic translation initiation, Bioessays 25 (2003) 1201–1211.
[11] G.M. Zeiner, N.R. Sturm, D.A. Campbell, The *Leishmania tarentolae* spliced leader contains determinants for association with polysomes, J. Biol. Chem. 278 (2003) 38269–38275.
[12] M. Juri Ayub, K.W. Ma, P.C. Shaw, K.B. Wong, *Trypanosoma cruzi*: high ribosomal resistance to trichosanthin inactivation, Exp. Parasitol. 118 (2008) 442–447.
[13] M. Juri Ayub, C.R. Smulski, K.W. Ma, M.J. Levin, P.C. Shaw, K.B. Wong, The C-terminal end of P proteins mediates ribosome inactivation by trichosanthin but does not affect the pokeweed antiviral protein activity, Biochem. Biophys. Res. Commun. 369 (2008) 314–319.

[14] H. Gao, M. Juri Ayub, M.J. Levin, J. Frank, The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components, Proc. Natl. Acad. Sci. USA 102 (2005) 10206–10211.
[15] D.B. Weatherly, J.A. Astwood 3rd, T.A. Minning, C. Cavola, R.L. Tarleton, R. Orlando, A heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results, Mol. Cell. Proteomics 4 (2005) 762–772.
[16] W.H. Mager, R.J. Planta, J.G. Ballesta, J.C. Lee, K. Mizuta, K. Suzuki, J.R. Warner, J. Woolford, A new nomenclature for the cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*, Nucleic Acids Res. 25 (1997) 4872–4875.
[17] R. Beckmann, C.M. Spahn, N. Eswar, J. Helmers, P.A. Penczek, A. Sali, J. Frank, G. Blobel, Architecture of the protein-conducting channel associated with the translating 80S ribosome, Cell 107 (2001) 361–372.
[18] M. Juri Ayub, C.R. Smulski, B. Nyambega, N. Bercovich, D. Masiga, M.P. Vazquez, C.F. Aguilar, M.J. Levin, Protein–protein interaction map of the *Trypanosoma cruzi* ribosomal P protein complex, Gene 357 (2005) 129–136.
[19] T.C. Fleischer, C.M. Weaver, K.J. McAfee, J.L. Jennings, A.J. Link, Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes, Genes Dev. 20 (2006) 1294–1307.
[20] D.A. Maslov, M.R. Sharma, E. Butler, A.M. Falick, M. Gingery, R.K. Agrawal, L.L. Spremulli, L. Simpson, Isolation and characterization of mitochondrial ribosomes and ribosomal subunits from *Leishmania tarentolae*, Mol. Biochem. Parasitol. 148 (2006) 69–78.
[21] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Direct analysis of protein complexes using mass spectrometry, Nat. Biotechnol. 17 (1999) 676–682.
[22] J. Sengupta, J. Nilsson, R. Gursky, C.M. Spahn, P. Nissen, J. Frank, Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM, Nat. Struct. Mol. Biol. 11 (2004) 957–962.