Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular & Biochemical Parasitology 148 (2006) 125–132
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Translational initiation in *Leishmania tarentolae* and *Phytomonas serpens* (Kinetoplastida) is strongly influenced by pre-ATG triplet and its 5′ sequence context

Julius Lukeš [a,*], Zdeněk Paris [a], Sandesh Regmi [a,1], Reinhard Breitling [b], Sergey Mureev [c], Susanna Kushnir [c], Konstantin Pyatkov [d], Milan Jirků [a], Kirill A. Alexandrov [c]

[a] *Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, České Budějovice, Czech Republic*
[b] *Jena Bioscience GmbH, Jena, Germany*
[c] *Max-Planck-Institute for Molecular Physiology, Dortmund, Germany*
[d] *Division of Biology, California Institute of Technology, Pasadena, USA*

Received 2 August 2005; received in revised form 13 March 2006; accepted 17 March 2006
Available online 18 April 2006

Abstract

To investigate the influence of sequence context of translation initiation codon on translation efficiency in Kinetoplastida, we constructed a library of expression plasmids randomized in the three nucleotides prefacing ATG of a reporter gene encoding enhanced green fluorescent protein (EGFP). All 64 possible combinations of pre-ATG triplets were individually stably integrated into the rDNA locus of *Leishmania tarentolae* and the resulting cell lines were assessed for EGFP expression. The expression levels were quantified directly by measuring the fluorescence of EGFP protein in living cells and confirmed by Western blotting. We observed a strong influence of the pre-ATG triplet on the level of protein expression over a 20-fold range. To understand the degree of evolutionary conservation of the observed effect, we transformed *Phytomonas serpens*, a trypanosomatid parasite of plants, with a subset of the constructs. The pattern of translational efficiency mediated by individual pre-ATG triplets in this species was similar to that observed in *L. tarentolae*. However, the pattern of translational efficiency of two other proteins (red fluorescent protein and tetracycline repressor) containing selected pre-ATG triplets did not correlate with either EGFP or each other. Thus, we conclude that a conserved mechanism of translation initiation site selection exists in kinetoplastids that is strongly influenced not only by the pre-ATG sequences but also by the coding region of the gene.

Keywords: Translation; Initiation; Pre-ATG; Kinetoplastida; Leishmania

1. Introduction

Initiation is a critical step of mRNA translation that is regulated by numerous translation initiation factors (eIFs). Additionally, several elements in the mRNA have been shown to control the efficiency of translational initiation and thus protein synthesis [1–4]. Four main elements in eukaryotic mRNAs are involved in regulating translational regulation: (i) the 5′ cap structure; (ii) the position of the AUG codon in regard to the 5′ end of the mRNA; (iii) secondary structure within the 5′ untranslated region (UTR) sequence and (iv) sequences flanking the AUG start codon (for review see [5]). The latter sequences modulate the ability of AUG codon to halt the scanning ribosomal subunit [2,6]. Recognition of the appropriate nucleotide context relies on the ability of eIF1 to change the 43S preinitiation complex into a scanning-competent form, as its addition strongly favors initiation from the correct start codon [7]. The optimal context for initiation of translation in mammals is GCCRCCaugG. In experimental tests, the biggest reduction in efficiency was seen when the purine (R) at the −3 position or G in the +4 position (relative to the +1 adenine of the start codon) was mutated [6,8]. Thus, initiation sites are usually designated “strong” or “weak” based on these two positions.

* Corresponding author. Tel.: +42 38 7775416; fax: +42 38 5310388.
E-mail address: jula@paru.cas.cz (J. Lukeš).
[1] Present address: Biological Sciences, Southern Methodist University, Dallas, USA.

0166-6851/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2006.03.008

Consensus sequences have been postulated for the start codon context of a representative number of genes in several eukaryotic organisms such as yeast [9], *Chlamydomonas* [10], plants and fungi [11]. These sequences appear to be species specific, since the consensus nucleotides in different sites vary from species to species. However, some features of the 5′ adjacent region of the AUG seem to be of universal importance, in particular the occurrence of a purine nucleotide, mostly adenine, at the −3 position [12].

Among lower eukaryotes, kinetoplastid flagellates are notorious for finding unique solutions to general processes of the eukaryotic cell. One of them is the virtual lack of transcriptional control, with regulation of abundance of mRNA and protein operating primarily at the post-transcriptional level [13]. Studies of mRNA levels derived from chromosomes 1 and 3 of *Leishmania major* indicate that transcription by RNA polymerase II (Pol II) initiates bidirectionally from a single region, generating two polycistronic primary transcripts containing dozens of genes [14,15]. The transcription of only few large polycistronic transcripts strongly biases regulation towards the post-transcriptional level. Such regulation may involve cap addition, almost exclusively *trans*-splicing, nucleocytoplasmic export, 3′ end processing, mRNA decay and finally translational initiation and elongation [13,16,17].

On a limited set of 100 genes, it was recently shown that the pre-ATG triplet in *Leishmania* has a G/ACC bias, which is similar to that of vertebrates [18]. This to an extent conflicts with an earlier study of flagellated protozoans showing a very low degree of conservation around AUG with the consensus sequence WNNNNNANNaugNC [19]. Another study suggested that pre-ATG triplets can influence the level of protein expression in *Leishmania* by several orders of magnitude [20].

In addition to the importance for understanding gene expression mechanisms in Kinetoplastida, the knowledge of sequences required for efficient translational initiation could provide useful tool for manipulation of the expression levels of homologous and heterologous genes in these organisms. To experimentally investigate the influence of the pre-ATG triplets on translation and overall protein expression in trypanosomatid flagellates, we undertook an exhaustive in vivo analysis using enhanced green fluorescent protein (EGFP) as a reporter protein. We observed a strong effect on translational efficiency and a similar impact of specific pre-ATG triplets in flagellates *Leishmania tarentolae* and *Phytomonas serpens*. We demonstrate that the different protein expression pattern observed with different target genes resulted from a complex interplay of the pre-ATG triplet and the coding sequence, possibly reflecting the influence of the secondary structure of the coding part of mRNA on translational initiation.

2. Materials and methods

2.1. EGFP plasmid construction

The library of expression plasmids with randomized pre-ATG nucleotides was prepared by PCR amplification of the *EGFP-N1* gene (Invitrogen) with forward primer 882 (ACAGCAGCCAGATCT**NNN**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG) and reverse primer 524 (AGGAGGAGGGCGGCCGCTTTA). The EGFP start codon is underlined and the pre-ATG triplet is in bold. The resulting PCR products were trimmed with restriction enzymes *Bgl*II and *Not*I and inserted into the *Leishmania* expression vector p F4X1.4sat (Jena Bioscience) open with the same enzymes [21]. The obtained *E. coli* clones were screened for the individual pre-ATG triplets by sequencing of the recombinant plasmids. Due to the nucleotide bias in the obtained library, second and third rounds of plasmid construction were performed with forward primers 1160 (ACAGCAGCCAGATCT**CNV**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1161 (ACAGCAGCCAGATCT**ADN**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1162 (ACAGCAGCCAGATCT**GDN**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), or 1163 (ACAGCAGCCAGATCT**THV**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG) and 1316 (GCCAGATCT**CAA**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1317 (GCCAGATCT**CGC**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1318 (GCCAGATCT**CGG**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1319 (GCCAGATCT**GAC**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), 1320 (GCCAGATCT**TCA**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG), or 1321 (GCCAGATCT**TAC**ATGGCTCGAGCGATGGTGAGCAAGGGCGAGGAGCTG) designed to produce missing triplets. After all 64 possible pre-ATG triplets were obtained, the sequence of the entire *EGFP* gene was confirmed for each clone. The resulting plasmids were linearized with *Swa*I and used for transfection of *Leishmania* and *Phytomonas* cells.

Genomic integration of the plasmids into the *SSU* locus was confirmed by diagnostic PCR with primer F2999 (CCTAGTATGAAGATTTCGGTGATC) annealing inside the expression cassette and primer F3002 (CTGCAGGTTCACCTACAGCTAC) annealing to the SSU rRNA sequence not present on the plasmid.

2.2. TET-R and dsRED plasmid construction

To obtain the pre-ATG variants of the tetracycline repressor (*TET-R*) gene, its coding sequence was amplified by PCR with forward primers 1496 (TAATAAAGATCT**ATC**ATGTCTAGATTAGATAAA AGTAAAGTG), 1497 (TAATAAAGATCT**ACC**ATGTCTAGATTAGATAA AAGTAAAGTG), 1498 (TAATAAAGATCT**TCG**ATGTCTAGATTAGATAA AAGTAAAGTG), 1499 (TAATAAAGATCT**CCA**ATGTCTAGATTAGATAA AAGTAAAGTG), and A263 (ACGCGTACACAACACACGGAC) as a reverse primer, using pF4TR1.4hyg as a template [22]. The PCR products were digested with *Bgl*II and *Not*I and subcloned into pF4X1.4hyg vector digested with the same enzymes.

For construction of pre-ATG variants of the red fluorescent protein *dsRED* gene, its coding sequence was PCR amplified with primers 1570 (TAATAAGCTCGAGCGATGAGGTCTTC CAAGAATGTTATCAAGG) and 1571 (TAATAAGGCGGC-

CGCTTTAAAG GAACAGATGGTGGCG) using pDsRed.T3-N1 vector as template [23]. The PCR product was digested with *Xho*I and *Not*I and inserted into the appropriate pre-ATG *EGFP* vector digested with the same enzymes making the use of the *Xho*I site introduced with the *EGFP* forward primers. In the resulting plasmids the *EGFP* gene was replaced by the *dsRED* gene while the 5′ and 3′ UTRs, pre-ATG triplet and first three amino acids remained unchanged.

2.3. Leishmania transfection and cultivation

All the above-described plasmids were linearized with *Swa*I and used for transfection of the *Leishmania* and *Phytomonas* cells [21]. The cultures of *L. tarentolae* UC and *P. serpens* 9T strains were grown in complex LEXSY broth BHI (Jena Bioscience) at 26 °C with slow shaking. Transfections of both species were performed by electroporation [24] in 4 mm cuvettes using a BTX electroporator and two pulses with 10 s pause with the settings 1510 V, 25 μF and 500 Ω. After the pulses, cells were transferred into a fresh medium to which 100 μg/ml of LEXSY NTC (nourseothricin, Jena Bioscience) was added 24 h after electroporation to select for stable transformants. After about a week of cultivation only cells resistant to nourseothricin survived. These cultures were used for EGFP measurement at days 9 and 10 post-transfection.

2.4. Northern blotting

Total RNA was isolated using TRIzol reagent (Sigma) according to the manufacturer's instructions. Approximately 5 μg/ml of RNA per lane was loaded on a 1% formaldehyde agarose gel, blotted and cross-linked following standard protocols. Fragments of the EGFP and α-tubulin genes used as probes were PCR amplified with the following primers: EGFP-F1, ATCTAACTCGAGGACGTAAACGGCCACAAGTTC; EGFP-R1, TCAATACTCGAGCGTCCATGCCGAGAGTGATC; Tub-F1, GCGCCGTTAATTAAGTTTTGGGAGGTGATCTCCG; Tub-R1, GCGCCGAATTCGGATCCCTGCAGCGTGCGGAAGCAAATATCG.

After prehybridization in the NaPi solution (0.5 M Na_2HPO_4 and NaH_2PO_4, pH 7.2; 7% SDS; 1 mM EDTA) for 2 h at 55 °C, hybridization with DNA probes labeled by random priming (Fermentas) with [α^{32}P]dATP (ICN) was performed overnight in the same solution at 55 °C. A wash in 2 × SSC + 0.1% SDS at RT for 20 min was followed by two washes in 0.2 SSC + 0.1% SDS for 20 min each. Northern blots were quantified using a Molecular Dynamics PhoshorImager and ImageQuant software and then also exposed to X-ray film with intensifying screens.

2.5. EGFP expression analysis

EGFP fluorescence emission of recombinant cell lines was measured for 10^8 cells resuspended in 200 μl of 10 mM Tris, pH 8.0 in a Spex Fluorolog-II spectrophotometer at 485 nm excitation and 510 nm emission. The number of cells was determined using the Beckman X2 Cell Counter. A similar approach was used for measuring fluorescence of dsRED, but the excitation and emission parameters were set to 545 and 620 nm, respectively. Three measurements of fluorescence emission were made for each recombinant cell line and the average value was plotted on a graph. The value obtained with the parental cells represents the background (Fig. 2).

For Western blotting protein lysates were prepared as described elsewhere [25], analyzed on 12% Tris–glycine–SDS gels, and blotted. Blots were probed with primary rabbit anti-EGFP antibody (Clontech) and secondary donkey anti-rabbit antibody coupled to horseradish peroxidase (Sevapharma). As a loading control blots were probed with primary monoclonal mouse antibody against the *Trypanosoma brucei* heat shock protein 70 [26] (kindly provided by K. Stuart) and secondary anti-mouse antibody coupled to horseradish peroxidase. For visualization of the TET-R protein, membranes were probed with rabbit anti-TET-R antibody (MoBiTec) and secondary donkey anti-rabbit antibody coupled to horseradish peroxidase. The immunoreactive bands were visualized using the ECL SuperSignal chemiluminiscence system (Amersham Biosciences) according to the manufacturer's protocol.

2.6. Purification of overexpressed EGFP and MALDI-TOF-mass spectrometry

For purification of EGFP, the recombinant strains were inoculated into LEXSY broth BHI. The cells were harvested at a density of approximately 10^8 cells/ml by centrifugation, resuspended in 20 mM Tris, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1 mM PMSF and disintegrated by sonication. The homogenate was cleared by ultracentrifugation for 1 h at 30,000 rpm and the EGFP protein was purified from the supernatant by organic extraction as described elsewhere [27]. The resulting protein was concentrated using Viva Spin filtration unit to a final concentration of 18 mg/ml. MALDI spectra were recorded on a Voyager-DE Pro Biospectrometry workstation (Applied Biosystems). Spectra recording and data evaluation was performed using the supplied Voyager software package.

2.7. L. major database search for pre-ATG triplets

Using the program Artemis [28], the whole genome was downloaded from ftp://ftp.sanger.ac.uk/pub/databases/L.major_sequences/DATASETS/LmjF_V4.0_20040630.artemis. Three bases upstream of the AUG start codon of each predicted gene were retrieved, in addition to their predicted function, where annotated.

3. Results and discussion

3.1. Pre-ATG triplet strongly influences EGFP expression

To study the influence of the sequence of the pre-ATG triplet on translation in the model flagellate *L. tarentolae*, we created a set of 64 plasmids with all possible permutations of this triplet of the *EGFP* reporter gene flanked by untranslated regions of the *L. tarentolae* calmodulin gene. This provides a complete set of point mutations of the pre-ATG sequence in addition to

all possible sequences of this region. This exhaustive series was obtained by inserting the *EGFP* gene cassettes amplified with randomized primers into a *Leishmania* expression vector pF4X1.4sat [21]. In all 64 plasmids, the entire *EGFP* gene and the 5′ and 3′ fusions with the UTRs were verified by sequencing.

These *EGFP* expression constructs were stably integrated into the small subunit ribosomal RNA (*SSU*) gene of the rDNA locus of *L. tarentolae* by homologous recombination [21]. Correct integration in this locus has been confirmed by a diagnostic PCR using one primer annealing inside the expression cassette and one primer annealing to a sequence not present in the plasmid (data not shown). In this configuration, transcription of the cloned gene is under the control of a strong PolI promoter that drives transcription of the ribosomal RNA genes. Transcription of a reporter gene by this polymerase is about 10 times higher than the read-through transcription by Pol II [29]. Although in all so far analyzed cases the transformation of *L. tarentolae* with pF4X1.4sat resulted in the single integration of the construct per genome it cannot be automatically excluded that not all *SSU* loci in the genome were transcribed at the same levels. Therefore, we chose not to obtain clones from the nourseothricin-resistant population but rather to analyze all drug-resistant cells of a cell line resulting from a single transfection. This ensures averaging over many individual integration events, predicted to be in the range of 10^4 on the basis of standard transfection efficiencies obtained with the protocol used [24].

In the recombinant cultures, fluorescence was usually detectable within 1–2 days after electroporation. When the transfectants were inspected under a fluorescence microscope, some differences in the intensity of EGFP expression in individual cells were observed, but most cells fluoresced to the same extent. However, a similar spread was also observed when a clonal population was analyzed, probably reflecting the difference in EGFP expression at different stages of the cell cycle (data not shown).

In all transfected cell lines, the EGFP protein was detected and quantified by measuring its fluorescence in living cells, and in a subset of strains also by estimating the amount of the protein in cell lysates using Western blotting with the anti-EGFP antibodies. The reproducibility of the expression phenotypes was verified with 15 cultures exhibiting either strong, medium or weak EGFP expression (Fig. 1A). The loading was controlled by reprobing the membrane with the anti-Hsp70 antibodies (Fig. 1A). Three independent transfections of the parental *L. tarentolae* with each of the 15 constructs yielded EGFP expression profiles that exhibited high concordance ($r^2 = 0.877$) within the replicates (data not shown). Using anti-EGFP antibody, the target protein was detected in all highly and medium fluorescent cultures but was undetectable in the low expressing cultures, reflecting the limits of the sensitivity of the antibody and non-linear readout of the chemiluminescent detection system (Fig. 1A).

To formally confirm that the observed differences in the EGFP expression were indeed due to the differences in translational initiation, we set out to exclude the possibility that it was influenced by other factors. There are two most obvious alternative explanations of the observed differences: (i) the pre-ATG triplet influences the choice of the first AUG in some cases promoting translational initiation to occur at the next downstream internal methionine; (ii) the changes in the pre-ATG triplet are influencing the stability of the mRNA and thus modulating the protein expression levels.

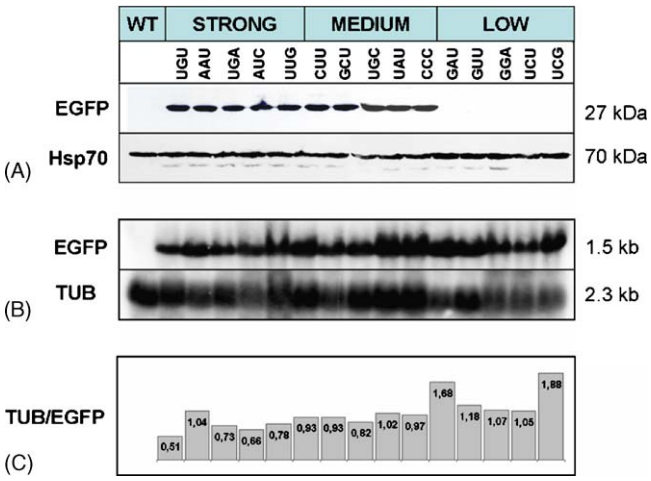

Fig. 1. Levels of EGFP protein but not *EGFP* mRNA, depend on the pre-ATG triplet, as shown in parental (wt) and 15 transfected *L. tarentolae* cultures. The pre-ATG triplet used for transfection is shown above the lanes. (A) EGFP protein levels were analyzed by Western blot analysis in cell extracts. Each lane was loaded with proteins from ~10^7 cells and blots were immunostained using anti-EGFP antibodies, as described in Section 2. Anti-Hsp70 antibody was used as a loading control. The size of the target proteins is indicated. (B) *EGFP* mRNA levels were analyzed by Northern blot analysis of total RNA. As a loading control, the membrane was rehybridized with the probe for α-tubulin. The size of the target mRNAs is indicated. (C) Plotted ratios between the quantified *EGFP* and α-tubulin mRNA signals (shown in B) are displayed.

To address the first possibility, i.e. to test whether translational initiation may occur at the downstream AUG, the EGFP protein was purified from strains bearing UAU, UGA and AUC pre-ATG permutations. This protein was subjected to mass spectrometry as described in Section 2. In all three cases the mass of EGFP was found to be 26,457 Da, which is in good accordance with the calculated mass of 26,322 Da indicating that translation starts invariably from the first methionine (data not shown).

In order to assess the potential influence of pre-ATG permutations on the *EGFP* mRNA stability, we analyzed 15 cultures by Northern blotting using the EGFP probe. Reassuringly, the level of *EGFP* mRNA was similar in all samples, while the transcript was missing from the parental cells (Fig. 1B). To ensure equal loading and introduce an internal standard we rehybridized the membrane with the probe for α-tubulin (Fig. 1B). Intensity of the signals for the α-tubulin and *EGFP* mRNAs of the individual strains were quantified and the plotted ratios are displayed (Fig. 1C). Analysis of the data suggests that although the highly translated mRNAs appear to be slightly more abundant than the poorly translated ones, the average difference of ca. 1.8-folds cannot account for the observed differences in EGFP expression. Therefore, we conclude that the *EGFP* gene was transcribed at the same rate in the individual recombinant cultures, and the stability of the *EGFP* mRNA is not significantly influenced by different pre-ATGs. Minor differences in the levels of mRNAs may reflect their ribosomal load and resulting partial RNase protection observed both in prokaryotes and eukaryotes [30,31].

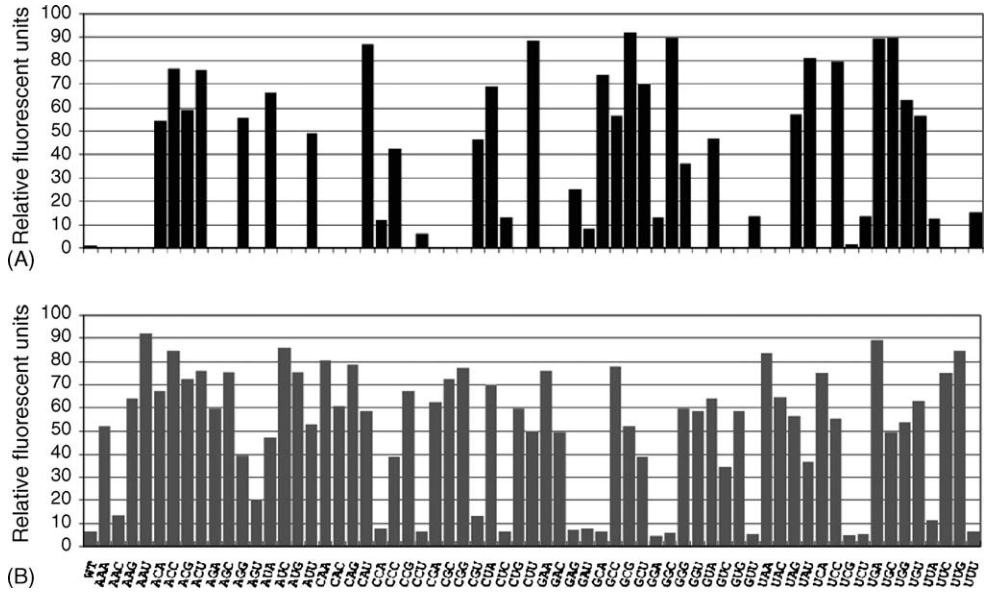

Fig. 2. Fluorescence values of selected *P. serpens* and *L. tarentolae* cell lines. Sequence of individual pre-ATG triplets is shown on the *x*-axis in the same order as they appear in Table 1 (background value of the wild type cells is on the left). The *y*-axis is labeled by number of counts/s. (A) Plotting of the EGFP fluorescence values for 37 constructs in *P. serpens*. (B) Plotting of the EGFP fluorescence values for all 64 constructs in *L. tarentolae*.

The EGFP protein level remained stable even during prolonged periods of cultivation, or upon repeated cycles of freezing and thawing of the cells, confirming its suitability as a reporter gene. Comparison of all 64 constructs that differed by pre-ATG permutations revealed 20-fold differences in *L. tarentolae* (Fig. 2; Table 1). Twenty-five pre-ATGs generated a fluorescence of over 1 million relative fluorescence units (rfu). The highest EGFP expression levels were obtained for pre-ATGs AUC, ACC, UUG, AAU and UGA, reaching up to 1.5 million rfu. Translation in a medium range (200,000–800,000 rfu) was specified by only eight triplets. On the other side of the wide spectrum is a group of 15 constructs, in which the pre-ATG allowed only very weak translation of the *EGFP* mRNA. In several cases (e.g. GUU, GGA, UCG and UCU), synthesis of the EGFP protein was virtually absent, since the obtained values were equivalent to those of the wild type background (Fig. 2A). Good correlation between Western analysis and the emission-based quantification proved the robustness of our quantification protocol for the reporter protein expression.

3.2. The influence of pre-ATG triplet on the expression of EGFP in P. serpens

A point of interest was to consider whether the role of these particularly strong pre-ATGs in the initiation of translation is confined solely to *L. tarentolae*, or is representative of all trypanosomatids. We chose to use for comparison *P. serpens*, which is thought to have diverged from *L. tarentolae* some 340 millions years ago [32]. *Phytomonas* is a cytochrome-mediated respiration-deficient pathogen of economically important plants such as palm trees [33] that, to our knowledge, has not been previously subject of genetic manipulation. Thirty-seven randomly selected constructs were electroporated into this trypanosomatid and a variation in the pattern of EGFP expression profile was

Table 1
In vivo and in silico analysis of all 64 types of pre-ATG triplets

AAA	0.83 (2.79)	GAA	1.21 (1.34)
AAC	0.22 (2.08)	GAC	0.79 (1.88)
AAG	1.02 (4.02)	GAG	0.11 (2.66)
AAU	1.48 (0.44)	GAU	0.12 (0.58)
ACA	1.07 (3.43)	GCA	0.11 (3.31)
ACC	1.36 (6.00)	GCC	1.24 (7.98)
ACG	1.15 (3.47)	GCG	0.83 (3.87)
ACU	1.22 (0.43)	GCU	0.62 (0.79)
AGA	0.95 (1.62)	GGA	0.69 (0.76)
AGC	1.21 (2.70)	GGC	0.10 (1.86)
AGG	0.63 (1.92)	GGG	0.95 (0.92)
AGU	0.32 (0.60)	GGU	0.94 (0.53)
AUA	0.75 0.12)	GUA	1.02 (0.69)
AUC	1.37 (5.53)	GUC	0.55 (3.37)
AUG	1.21 (0.06)	GUG	0.93 (1.78)
AUU	0.85 (0.36)	GUU	0.08 (0.34)
CAA	1.29 (0.87)	UAA	1.34 (0.12)
CAC	0.97 (2.05)	UAC	1.03 (0.57)
CAG	1.26 (1.81)	UAG	0.90 (0.16)
CAU	0.94 (0.47)	UAU	0.59 (0.23)
CCA	0.12 (1.43)	UCA	1.20 (0.82)
CCC	0.62 (1.76)	UCC	0.88 (2.04)
CCG	1.08 (1.87)	UCG	0.08 (1.57)
CCU	0.10 (0.31)	UCU	0.09 (0.46)
CGA	1.00 (1.27)	UGA	1.42 (0.33)
CGC	1.15(2.18)	UGC	0.79 (1.21)
CGG	1.24 (1.07)	UGG	0.85 (0.50)
CGU	0.21 (0.49)	UGU	1.00 (0.45)
CUA	1.12 (0.46)	UUA	0.18 (0.16)
CUC	0.10 (2.09)	UUC	1.20 (1.25)
CUG	0.96 (1.21)	UUG	1.35 (0.74)
CUU	0.79 (0.32)	UUU	0.11 (0.26)

Fluorescence of *L. tarentolae* cell lines (in arbitrary units; average values obtained in vivo from three independent measurements of the same culture). Numbers in brackets are representation of in silico-predicted pre-ATG triplets in the whole genome (=8213 predicted genes) of *L. major* (in %).

observed (Fig. 2B). A comparison between the subset of 37 pre-ATG triplets in *P. serpens* and *L. tarentolae* showed that in 80% of cases, the triplets behave similarly in both species. Yet, significant differences were observed in several cases (in particular triplets AUG, GCA and GGC (Fig. 2)).

Taken together, these data have the following implications. First, the orthologous *SSU* rDNA-targeting region is sufficient for homologous recombination in *P. serpens*. Second, the UTRs flanking the *EGFP* gene, derived from the *L. tarentolae* calmodulin (*cam*) operon gene cluster [21] are able to direct efficient *trans*-splicing, RNA processing and translation in *P. serpens*. Typically, in trypanosomatids high-level protein expression requires the presence of homologous UTRs [34], indicating that the UTRs of the *cam* genes of *L. tarentolae* contain RNA stabilizing sequences conserved in trypanosomatids. This assumption is supported by the finding that the pF4X1.4 expression vectors mediated efficient expression also in other *Leishmania* species, such as *L. major*, *L. mexicana*, *L. donovani* and *L. infantum* (Goyard S., personal communication).

3.3. The pre-ATG triplet strongly influences the expression of TET-R and dsRED, but the influence is gene-dependent

Having demonstrated that the pre-ATG triplet strongly influences the efficiency of *EGFP* mRNA translational initiation in *L. tarentolae* and *P. serpens*, we thought to analyze whether the observed effect would also occur with other protein coding genes. We chose as reporter the tetracycline repressor TET-R [35] and the red fluorescent protein dsRED [36] and as host *L. tarentolae*. To minimize the potential influence of the 5′ end sequence of the open reading frame, in the constructs containing the *dsRED* gene the codons for the first three amino acids of EGFP have been retained.

Based on the EGFP dataset, AUC and ACC were selected as the "strong" triplets, and UCG and CCA as the "weak" ones. The recombinant *L. tarentolae* cell lines were constructed as described in Section 2 and protein expression levels were analyzed by fluorescence measurements (dsRED) or by quantitative Western blotting (TET-R). The results indicate that the expression levels of the selected pre-ATG variants vary among the constructs for both target genes, but correlate neither with the observed EGFP expression levels for the individual pre-ATG variants nor among themselves (Fig. 3). This observation suggests that although the pre-ATG triplet can strongly influence translational initiation in trypanosomatids, it appears to function in concert with a much larger portion of the coding part of mRNA. Since there is no record of translational initiation in eukaryotes being influenced by AUG flanking sequences beyond −4 [6], the observed influence is likely to be a result of secondary structure formation around starting AUG. It has been shown recently that translation can be effectively blocked by the introduction of a hairpin into the 5′ UTR of trypanosome mRNAs [37]. Interestingly, the region exerting such an influence must also involve a sizable portion of the coding region, certainly more than six nucleotides downstream of the start codon, thus coupling the influence of protein sequence and codon usage on expression levels.

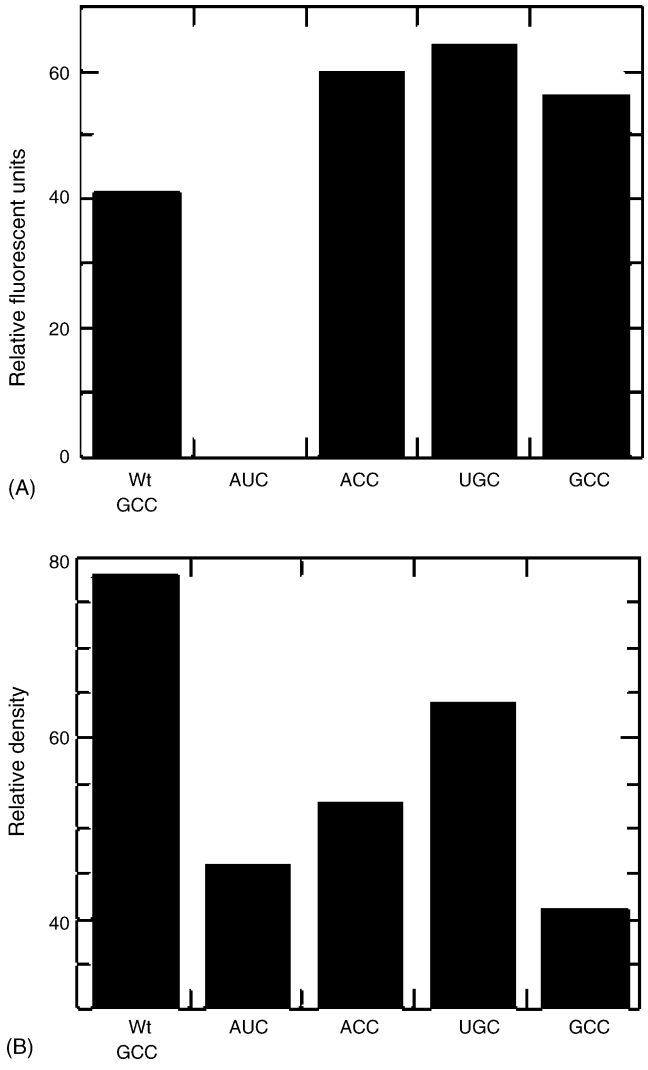

Fig. 3. Levels of dsRED and TET-R proteins vary in dependence on pre-ATG, as shown in wild type and four transfected *L. tarentolae* cell lines. The pre-ATG triplet used for transfection is shown below the lanes. (A) Expression levels of dsRED are shown in relative fluorescence units. (B) Expression levels of TET-R are shown in arbitrary units as relative density of specific bands.

The identification of pre-ATG triplets capable of protein expression modulation potentially provides a tool for engineering kinetoplastid flagellates, where the paucity of identified tunable promoters significantly limits the scope of expression constructs. An alternative is tuning the levels of protein expression by engineering the 5′ and 3′ UTRs, which is highly empirical [34]. Possible alternatives also include manipulation of *trans*-splicing and polyadenylation efficiencies that so far have not been put to practical purposes. Admittedly, the presented alternative requires construction and analysis of pre-ATG variants for each individual protein, a laborious proposition when the entire range of expression levels is desired. However, it may be of importance when one attempts to overexpress a heterologous gene to high levels since, as was shown for EGFP and dsRED, some pre-ATG triplets may lead to a total lack of expression (Fig. 3). In principle, this could be avoided by in silico analysis of mRNA structures. However, our initial efforts to correlate the

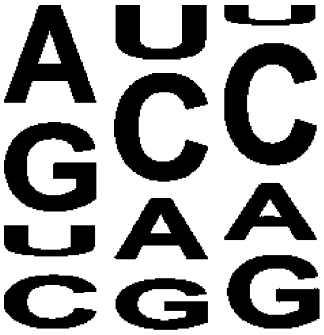

Fig. 4. Consensus of the *L. major* pre-ATG triplet based on in silico analysis.

levels of EGFP expression to predicted secondary structures in the vicinity of AUG did not lead to unambiguous dependence (data not shown).

3.4. Analysis of in silico predicted pre-ATGs shows a lack of consensus in L. major

Recently, the genome sequence of *L. major* has been completed [38]. Genes were predicted using a combination of Hexamer software (http://www.sanger.ac.uk/Software/analysis/hexamer/) and codon preference. In cases when the codon usage plots and hexamer predictions were in disagreement or not clear, start codons have been identified based on the presence of Kozak-like sequences (ACCA/GCCATGGCC) (Ivens A., personal communication).

We have compared our EGFP in vivo dataset with the 8213 in silico-predicted *L. major* pre-ATGs. In fact, all 64 possible combinations are present in this position with only the start and stop codons (AUG, UAA, UAG and UGA) being very rare (Table 1). An in silico-based pre-ATG consensus in *L. major* (Fig. 4) thus seems to be much more complex than the same sequence motif predicted for other eukaryotes.

Although many triplets that drive strong EGFP expression in *L. tarentolae* are more frequently encountered in the genome of *L. major* than the "weak" pre-ATGs, the correlation of frequency to "strength" is not convincing in general. For example AUC, which represents about 5.5% of all in silico-predicted *L. major* pre-ATG triplets, mediated a very strong translation of the EGFP protein in both *L. tarentolae* and *P. serpens*, but shut the dsRED translation down and mediated only weak expression of TET-R in *L. tarentolae* (Figs. 2 and 3). Similarly, triplets AAU and UAA rarely predicted as pre-ATGs in the *L. major* genome (in 0.44% and 0.12% cases, respectively) (Table 1), generated a very strong EGFP fluorescence. All in all, our exhaustive analysis strongly suggests an open reading frame-dependent initiation of translation in trypanosomatid flagellates, which is mediated not only by the nucleotides prefacing the translational start codon but also by the 5′ region of the mRNA beyond position +9.

Acknowledgements

This work was supported by grants from the Grant Agency of the Czech Academy of Sciences (Z60220518 and S60220554), the Ministry of Education of the Czech Republic (MSM6007665801) and Deutsche Forschungsgemeinschaft (AL484/5-3). We thank Al Ivens (Sanger Institute, Cambridge) for help with the analysis of pre-ATG codons in the *L. major* database and comments on the manuscript. Ken Stuart (Seattle Biomedical Research Institute) kindly provided anti-hsp70 antibodies. The help of František Vácha (Institute of Molecular Biology of Plants) with the quantification experiments is acknowledged.

References

[1] Higgs DC, Shapiro RS, Kindle KL, Stern DB. Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol Cell Biol 1999;19:8479–91.
[2] Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene 1999;234:187–208.
[3] Vervoort EB, van Ravenstein A, van Peij NNME, et al. Optimizing heterologous expression in *Dictyostelium*: importance of 5′ codon adaptation. Nucleic Acids Res 2000;28:2069–74.
[4] Wilkie GS, Dickson KS, Gray NK. Regulation of mRNA translation by 5′-and 3′-UTR-binding factors. Trends Biochem Sci 2003;28:182–8.
[5] Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 2005;361:13–37.
[6] Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J 1997;16:2482–92.
[7] Pestova TV, Kolupaeva VG. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev 2002;16:2906–22.
[8] Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 1986;44:283–92.
[9] Yun DF, Laz TM, Clements JM, Sherman F. mRNA sequences influencing translation and the selection of AUG initiator codons in the yeast *Saccharomyces cerevisiae*. Mol Microbiol 1996;19:1225–39.
[10] Esposito D, Hicks AJ, Stern DB. A role for initiation codon context in chloroplast translation. Plant Cell 2001;13:2373–84.
[11] Pesole G, Gissi C, Grillo G, Licciulli F, Liuni S, Saccone C. Analysis of oligonucleotide AUG start codon context in eukaryotic mRNAs. Gene 2000;261:85–91.
[12] Kozak M. Structural features in eukaryotic messenger RNAs that modulate the initiation of translation. J Biol Chem 1991;266:19867–70.
[13] Clayton CE. Life without transcriptional control? From fly to man and back again. EMBO J 2002;21:1881–8.
[14] Johnson PJ, Kooter JM, Borst P. Inactivation of transcription by UV-irradiation of *Trypanosoma brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell 1987;51:273–81.
[15] Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ. Transcription initiation and termination on *Leishmania major* chromosome 3. Eukaryot Cell 2004;3:506–17.
[16] Myung KS, Beetham JK, Wilson ME, Donelson JE. Comparison of the post-transcriptional regulation of the mRNAs for the surface proteins PSA (GP46) and MSP (GP63) of *Leishmania chagasi*. J Biol Chem 2002;277:16489–97.
[17] D'Orso I, De Gaudenzi JG, Frasch ACC. RNA-binding proteins and mRNA turnover in trypanosomes. Trends Parasitol 2003;19:151–5.
[18] Requena JM, Quijada L, Soto M, Alonso C. Conserved nucleotides surrounding the trans-splicing acceptor site and the translation initiation codon in *Leishmania* genes. Exp Parasitol 2003;103:78–81.
[19] Yamauchi K. The sequence flanking translational initiation site in protozoa. Nucleic Acids Res 1991;19:2715–20.
[20] Stanton DJ, Mensa-Wilmot K. Control of protein expression by −3 to −1 nucleotides of *Leishmania* 5′ UTRs: identification of translational enhancers. Mol Biol Cell 2000;11:441A.

[21] Breitling R, Klinger S, Callewaert N, et al. Non-pathogenic trypanosomatid protozoa as a platform for protein research and production. Protein Expr Purif 2002;25:209–18.
[22] Kushnir S, Gase K, Breitling R, Alexandrov KA. Development of an inducible protein expression system based on the protozoan host *Leishmania tarentolae*. Protein Expr Purif 2005;42:37–46.
[23] Bevis BJ, Glick BS. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). Nat Biotechnol 2002;20:83–7.
[24] Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. Mol Biochem Parasitol 2003;128:217–28.
[25] Pasion SG, Hines JC, Aebersold R, Ray DS. Molecular cloning and expression of the gene encoding the kinetoplast-associated type II DNA topoisomerase of *Crithidia fasciculata*. Mol Biochem Parasitol 1992;50:57–68.
[26] Wang BB, Ernst NL, Palazzo SS, Panigrahi AK, Salavati R, Stuart K. TbMP44 is essential for RNA editing and structural integrity of the editosome in *Trypanosoma brucei*. Eukaryot Cell 2003;2:578–87.
[27] Yakhnin AV, Vinokurov LM, Surin AK, Alakhov YB. Green fluorescent protein purification by organic extraction. Protein Expr Purif 1998;14:382–6.
[28] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[29] Biebinger S, Rettenmaier S, Flaspohler J, et al. The PARP promoter of *Trypanosoma brucei* is developmentally regulated in a chromosomal context. Nucleic Acids Res 1996;24:1202–11.
[30] Jack HM, Berg J, Wabl M. Translation affects immunoglobulin mRNA stability. Eur J Immunol 1989;19:843–7.
[31] Braun F, Le Derout J, Regnier P. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA in *Escherichia coli*. EMBO J 1998;17:4790–7.
[32] Fernandes AP, Nelson K, Beverley SM. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa—perspectives on the age and origins of parasitism. Proc Natl Acad Sci USA 1993;90:11608–12.
[33] Marché S, Roth C, Philippe H, Dollet M, Baltz T. Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. Mol Biochem Parasitol 1995;71:15–26.
[34] Clayton CE. Genetic manipulation of Kinetoplastida. Parasitol Today 1999;15:372–8.
[35] Berens C, Hillen W. Gene regulation by tetracyclines. Eur J Biochem 2003;270:3109–21.
[36] Baird GS, Zacharias DA, Tsien RY. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. Proc Natl Acad Sci USA 2000;97:11984–9.
[37] Webb H, Burns R, Ellis L, Kimblin N, Carrington M. Developmentally regulated instability of the *GPI-PLC* mRNA is dependent on a short-lived protein factor. Nucleic Acids Res 2005;33:1503–12.
[38] Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite *Leishmania major*. Science 2005;309:436–42.