Gene structure and promoter function of a teleost ribosomal protein: a tilapia (*Oreochromis mossambicus*) L18 gene

Alfredo Molina a, Arati Iyengar b, Luis F. Marins b, Frédéric Biemar a, Sean Hanley c, Norman Maclean b, Terry J. Smith c, Joseph A. Martial a, Marc Muller a,*

a Laboratoire de Biologie Moléculaire et de Génie Génétique, Université de Liège, Institut de Chimie B6, B-4000 Sart-Tilman, Belgium
b School of Biological Sciences University of Southampton, Bassett Crescent East, Southampton, SO16 7PX, UK
c National Diagnostics Center, National University of Ireland, University Road, Galway, Ireland

Received 2 January 2001; received in revised form 18 June 2001; accepted 9 July 2001

Abstract

We have cloned and characterized a tilapia (*Oreochromis mossambicus*) L18 ribosomal protein gene, including the complete transcribed region and 488 bp of upstream regulatory sequences. We have also isolated two L18 cDNAs from another tilapia (*Oreochromis niloticus*) with a few conservative nucleotide differences. Our results suggest the presence of two genes in both species. Reporter constructs were tested for transient expression in CV1 cells and in microinjected zebrafish and tilapia embryos. The tilapia L18 promoter was able to drive expression of the reporter gene in all three experiments, with no apparent preference for a particular tissue. The tilapia L18 promoter is therefore likely to be a powerful tool to drive tissue-independent gene expression in fish. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Ribosomal protein; Promoter; Transcription; L18; Fish; Tilapia

1. Introduction

Ribosomes are composed of 3–4 rRNA molecules and about 60–80 protein subunits. Ribosomal protein (r-protein) genes have been considered housekeeping genes, with a strongly coordinated constitutive expression that is closely linked to that of the different ribosomal RNAs to form the entire ribosome. R-protein expression is adjusted according to the needs of the cell for protein biosynthesis [1].

The coordinated expression of the r-protein genes is controlled at various levels. In prokaryotes, most of them are organized in a comparatively small number of operons [2], which encode a repressor of both transcription and translation that binds to a specific site on the polycistronic RNA [3,4]. In eukaryotes, r-protein gene-specific regulatory elements have been identified in the different promoter sequences, consistent with a coordinated expression. Most eukaryotic r-protein genes have special short promoters devoid of a TATA box. The transcription start site (mostly a C) is part of an oligopyrimidine tract flanked by GC-rich sequences [5–8]. The mRNA thus starts with a 5′-terminal oligopyrimidine tract (5′-TOP) that is required for translational control of r-proteins [9]. Such a 5′-TOP tract was recently described in the mRNA coding for the medaka (*Oryzias latipes*) r-protein gene S3a [10]. Control of expression at the level of RNA splicing and degradation has also been described (for a review, see [11]).

The r-protein L18 in concert with L5 and L25 interacts with the 5S rRNA, which is a component of the large ribosomal subunit [12]. Only a few genomic sequences have been reported in higher eukaryotes (human [13], *Xenopus* [14], *Caenorhabditis elegans* [15]) and only one promoter study in *Xenopus laevis* [16]. Here we report the molecular cloning of a tilapia (*Oreochromis mossambicus*) L18 ribosomal protein gene (tiL18). We analyzed its expression and performed reporter promoter studies, using a reporter gene strategy, by transient assays in cultured cells or in microinjected zebrafish (*Danio rerio*) and tilapia (*Oreochromis niloticus*) embryos. We observed high levels of expression under ideal metabolic conditions suggesting that the L18 regulatory sequences are likely to drive efficient and constitutive expression of a transgene in tilapia.
2. Materials and methods

2.1. Oligonucleotides

Oligonucleotide primers (Eurogentec, Seraing, Belgium) used for PCR procedures were: L18pf1: TCCCTTTT-CGCTCTGAGTCC; L18pr1: TTGGTCCTGCTCATGAACAG; T12MG, T12MA, T12MT and T12MC (M represents G, A, or C); nL18pr1: TTCTTGTAGCCG-CAGCTGGCTC; fL18pr1: TTGGTCCTGCTCATGAA-CAG; pL18f1: AGCCATGGCCCGATTAGAATGCT-TGGTG; pL18r1: GGGAATTCTGGGACTCAGAGC-GAAAAGGG.

All RT-PCR or PCR products were cloned into the pCRII vector and sequenced.

2.2. Library screening

Using a salmon L18 cDNA (GenBank No. CAC36993) as a probe, 2\times10^6 plaques (recombinant λGEM 11) of a tilapia (O. mossambicus) genomic library [17] were screened. Probe labeling, hybridization, phage DNA purification, subcloning and restriction mapping of positive clones were performed according to [18].

2.3. cDNA amplification

Four different anchored oligo(dT) and the L18pf1 primers were used to amplify the L18 cDNA by RT-PCR. Livers from adult tilapia (O. niloticus) were dissected and immediately frozen in liquid nitrogen. Poly(A)^+ RNA was isolated from adult tilapia (O. niloticus) livers using the QuickPrep Micro mRNA Purification Kit (Amersham/Pharmacia Biotech). RT-PCR experiments were performed using the Titan One Tube RT-PCR Kit (Boehringer Mannheim). To confirm the identity of the amplified fragment, a nested PCR reaction was performed using the primers L18pf1 and nL18pr1.

2.4. Northern and Southern blots

Total RNA was isolated from different organs of adult tilapia (O. niloticus) (brain, muscle, intestine and liver) using TRIzol (Gibco BRL). 10 μg were separated on a 1% denaturing agarose gel and transferred to a nylon membrane. The integrity and the amounts of RNA loaded on the gel were verified by ethidium bromide staining. DNA was prepared [19], digested with the appropriate restriction enzymes, separated on a 1% agarose gel and transferred to a nylon membrane. A 192 bp tiL18 cDNA fragment was amplified from clone nilot b (see Section 3) using primers L18pf1 and L18pr1 for use as the probe.

2.5. Cap site determination by primer extension

Total RNA from liver (1 μg) was used to perform primer extension experiments using the 5′-fluorescein labeled fL18pr3 primer. The reaction was treated and analyzed as described [18].

2.6. Reporter gene constructs

A 500 bp PCR fragment (−488/+28), including the appropriate restriction sites at the 5′-end of the pL18f1 and pL18r1 primers, was amplified from the λ5jtiL18 clone. To obtain tiL18-0.5LacZ, this fragment was introduced into the expression vector pGCV-LacZ [20], where the SV40 promoter had previously been deleted [18]. The tiL18-0.2LacZ expression vector was obtained from the tiL18-0.5LacZ by digestion with NcoI and XhoI followed by religation. The carp β-actin promoter construct was previously described [25].

2.7. Cell culture and transfection experiments

CV1 cells were grown in DMEM medium (Gibco-BRL) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin at 37°C in 5% CO₂. For transfections, cells were harvested using trypsin-EDTA and resuspended in the same culture medium (1.5×10^6 cells/ml). 5 pmol of purified plasmid DNA was transfected by the calcium phosphate precipitation procedure as previously described [21]. After 48 h, the cells were harvested [22] and protein concentration and β-galactosidase activity were determined [19]. Triplicate experiments were performed at least twice.

2.8. Microinjection procedure and β-galactosidase assays on injected embryos

Zebrafish care and embryo rearing were as described [23]. Tilapia microinjection and embryo rearing were as described [24]. Fertilized eggs were microinjected into the cytoplasm with 10^6 copies of DNA in 300 pl of NT buffer (88 mM NaCl, 10 mM Tris–HCl, pH 7.5) at the one-cell stage. β-Galactosidase expression was measured in individual zebrafish and tilapia embryos after 24 and 48 h, respectively, using a fluorometric assay in a TTO100 fluorometer (Hoefer). Embryos were homogenized in 45 μl of 10 mM Tris–HCl, pH 7.5, 10 mM NaCl, 0.1% Triton X-100, incubated at room temperature for 10 min, centrifuged briefly to clarify and 40 μl used in the methyl umbelliferone galactoside (MUG) assay as detailed in Hoefer Technical Bulletin No. 129. Few batches of injected 3–4 day old tilapia fry were also used for histochemical staining in order to visualize β-galactosidase in whole mounts [24].
3. Results

3.1. Genomic library screening

A salmon ribosomal protein L18 cDNA clone was used as a probe to screen $2 \times 10^6 \lambda$-phage plaques from a tilapia (O. mossambicus) genomic library. One out of 94 positive clones containing approx. 13 kb of insert, λ5jtiL18, was chosen for restriction mapping and Southern blot analysis (data not shown). A 4.5 kb XhoI fragment hybridizing to the L18 probe was subcloned (pB4.5-tiL18) and sequenced. Primers from the 5' and 3'-ends of the pB4.5-tiL18 insert were used to identify flanking regions by sequencing the λ5jtiL18 DNA. The complete sequence is shown in Fig. 1.

![Fig. 1. The nucleotide sequence of a tilapia L18 gene and the deduced amino acid sequence in the coding region are shown. Upper case letters are used for the coding sequence. The transcription start site is indicated by +1. An asterisk denotes the stop codon. Intron positions are indicated. The putative polyadenylation signals, CAAT box, T-rich regions and the putative K, L, Q or N elements are shadowed. The L27 Box-A is boxed and the GABP inverted binding sites are indicated by arrows. Repeat sequences are in bold letters and italic bold letters show the oligonucleotide used to determine the transcription start site. The sequence reported in this paper has been deposited in the EMBL database under accession No. AF240376.](image-url)
gene sequence of a tilapia L18 gene and about 0.5 kb upstream from the ATG, including the 5'-UTR sequence, was isolated (Fig. 1).

3.2. Structural analysis of the tilapia L18 gene

To determine the position of the L18 introns, RT-PCR reactions were performed on RNA from a tilapia (*O. niloticus*). Two niloticus L18 cDNA clones were obtained (nilot a and nilot b) (GenBank database accession No. AF240374 and AF240375, respectively). Only a few conservative nucleotide changes were detected between the three tilapia sequences: the coding sequence deduced from the *O. mossambicus* gene differs by 1 bp from the *O. niloticus* clone b and 4 bp from the *O. niloticus* clone a.

Seven exons were found, separated by small introns (Fig. 1); the gene structure and the splice positions are conserved as compared to the L18 equivalent gene in *X. laevis*, L14 [14]. Amino acid sequence comparison of the encoded 188 amino acid protein to the GenBank database revealed a high level of homology to other L18 sequences: 85.5% identity with salmon (GenBank No. CAC36993), 85.7% with rat [25] and 85.7% with L14a and 85.1% with L14b (L18 equivalents) from *X. laevis* [16] (Fig. 2A).

By comparing the retention time of a fragment obtained by primer extension of liver total RNA with the retention times of the products from a sequencing reaction performed on the cDNA clone nilot b using the same primer, we were able to locate the start site at 27 bp upstream from the ATG (Figs. 1 and 2B). The poly(A) signal was

---

**Table**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TilapiaM</td>
<td>AAF64459</td>
</tr>
<tr>
<td>TilapiaA</td>
<td>AAF64458</td>
</tr>
<tr>
<td>XenopusA</td>
<td>AF240374</td>
</tr>
<tr>
<td>XenopusB</td>
<td>AF240375</td>
</tr>
<tr>
<td>Salmon</td>
<td>AF240374</td>
</tr>
<tr>
<td>Rat</td>
<td>AF240374</td>
</tr>
<tr>
<td>Mouse</td>
<td>AF240374</td>
</tr>
<tr>
<td>Human</td>
<td>AF240374</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** A) Amino acid sequence comparison of the tilapia (AAF64459 and AAF64458), salmon (CAC36993), *Xenopus* (P02412 and P09897), mouse (BAB26043), human (NP_000970) and rat L18 (NP_112364) sequences. (B) Transcription start site. Primer extension was performed using adult tilapia liver RNA. Comparison of the primer extension products with the products of a sequencing reaction on the cDNA clone nilot b is shown.
3.3. Southern blotting

DNA was prepared from adult tilapia (*O. mossambicus*) livers and Southern blot analysis was performed using a 189 bp tiL18 cDNA probe covering exons 1–3 (Fig. 3A). Two fragments of 4.2 kb and 1.3 kb were detected in DNA digested with XhoI and PstI respectively, corresponding to those expected from the tiL18 genomic sequence (Fig. 3B, bands labeled *). The additional weak PstI fragment might correspond to an adjacent fragment bound by the 5′-end (exon 1) of the probe. In contrast, the second 3.0 kb XhoI fragment revealed by the probe is not explained by the isolated tiL18 sequence. Digestion with SacI or HindIII (one site each in the tiL18 sequence) revealed one and two genomic fragments respectively. These results suggest the presence of at least two genes or alleles homologous to L18.

3.4. tiL18 expression by Northern blotting

Total RNA from different tissues of a tilapia (*O. nilioticus*) was prepared and analyzed by Northern blotting using a 192 bp cDNA fragment as a probe. In all the tissues, a single RNA of approx. 0.6 kb was detected. Although the signal in liver seems to be somewhat higher, no dramatic differences were found (Fig. 4).

3.5. Transfection experiments

Transient expression studies were performed in CV1 cells (derived from monkey kidney). Two expression vectors, tiL18-0.5LacZ and tiL18-0.2LacZ, were used which contain the LacZ reporter gene driven by the complete isolated regulatory sequences and a 300 bp 5′-deleted sequence respectively. A high β-galactosidase activity was obtained after transfection of both constructs, comparable to that observed with the positive control pGCV-LacZ containing the SV40 promoter (Fig. 5).

3.6. Transient expression in zebrafish and tilapia embryos

Transient expression was observed in both zebrafish and tilapia individual embryo extracts (Fig. 6A,B), although the levels of expression seen in tilapia were considerably

Fig. 3. Southern blot. (A) Schematic representation of the tiL18 gene and the deduced cDNA. Introns are indicated by lower case letters in the gene and exons by roman numerals in the cDNA. The probe used is indicated by (^*). Restriction sites are indicated on top of the gene: X, XhoI; H, HindIII; P, PstI; no SacI site in the cloned genomic fragment. (B) Tilapia DNA (10 μg) was digested with the indicated restriction enzymes, separated on a 1% agarose gel and transferred to nylon membrane. Bands corresponding to fragments predicted from the tiL18 sequence are indicated (*).

Fig. 4. Northern blot. Expression analysis in different organs. The 192 bp tiL18 cDNA fragment was used as the probe.
higher than in zebrafish (note different scales for fluorescence units). In tilapia, the overall mean expression of the tiL18-0.2LacZ construct was significantly better ($P < 0.01$) than the tiL18-0.5LacZ construct (respectively $80,000 \pm 10,000$ and $7000 \pm 3000$ fluorescence units), while expression from the carp $\beta$-actin promoter, strongly active in tilapia [24], was only 2-fold higher ($180,000 \pm 60,000$). In zebrafish, expressions from both constructs were not significantly different ($800 \pm 150$ and $1000 \pm 200$ fluorescence units, respectively). Tilapia embryos microinjected with the tiL18-0.2LacZ construct were analyzed by performing histochemical staining on whole mounts. Expression was observed in the YSL (not shown) and in defined patches of cells in the embryo (an example is shown Fig. 6D).

4. Discussion

Although heterologous regulatory sequences are often used to express transgenes in fish with success, a number of researchers have recommended the use of fish-derived sequences for appropriately regulated and high levels of expression. In addition, the unacceptability to the food market of transgenic fish containing heterologous sequences leads to a demand for new fish promoters [26]. We present a new, strong fish promoter that can be used to drive ubiquitous expression of a gene of interest. The tilapia was chosen because it is one of the best candidate fish for the forecasted massive expansion of aquaculture in the tropics.

We have cloned, sequenced and carried out an expres-
sion analysis of a tilapia L18 gene. The complete genomic region including about 500 bp of the 5′-flanking sequence from *O. mossambicus* and two different cDNA sequences from *O. niloticus* are presented. Amino acid sequences, gene structure and the positions of the cap site (+1) and the poly(A) signal (+4933) are conserved as compared to other L18 genes.

In mouse, several non-expressing intron-less pseudogenes for L18 have been described [27]. This observation has also been made for other r-protein genes [5,6]. Typically, mammals have a single expressed r-protein gene and various processed pseudogenes [28]. Our results imply the existence of at least two transcribed L18 genes (or alleles) in *O. niloticus*, as suggested by the two different isolated cDNAs. This is consistent with the presence of at least two genes (or alleles) in *O. mossambicus* as shown by Southern blotting. The presence of two different cDNA clones in *O. niloticus* suggests that both genes are transcribed in tilapia; no evidence for additional L18 processed pseudogenes was found.

The tiL18 promoter is devoid of a TATA box and its transcription start site, located 27 bp upstream from the ATG, comprises the conserved consensus sequence CY-TYTCTYTTYYC (Y = C or T) previously described for r-protein genes [5-8]. The tiL18 5′-region contains two repeat sequences (−355 and −328) and two T-rich regions (−320 and −289). However, as the minimal promoter construct (tiL18-0.2LacZ) lacking these regions is still able to drive a significant expression, we focused our analysis on the more proximal promoter region. In higher eukaryotes, the promoter region of r-proteins is organized into four elements: α (RFX-1), β (GABP), γ (g-factor) and δ (NF-E1) [29]. Potential sites corresponding to these elements were found in the proximal tiL18 regulatory region and their organization was similar to that in other r-protein genes (Fig. 1). Two α (α1, α2) and two δ elements (δ1, δ2) were found. Both δ elements are missing in our reporter constructs (site δ1 overlaps the ATG start codon and δ2 is located in the first intron), suggesting that they are not essential for transcriptional activity. Similarly, the δ site in the first exon of *X. laevis* L1 and L14 r-protein genes was shown to play a marginal role in transient expression of a reporter gene [30]. A region (TCGGCGAGA) at position −20 is identical in sequence and position to the Box-A-like element described in the human and mouse L27 r-protein genes [31] and similar to the Box-A of the mouse L7a r-protein which has been shown to bind nuclear factors [32]. Finally, two inverted GABP binding sites were found around the transcription start site, which are identical in sequence and position to those found in the mouse S16 r-protein and cytochrome c oxidase CO4 and CO5b genes. These sites were demonstrated to cause a significant reduction in the S16 promoter activity [33] and a transcriptional activation in the CO4 and CO5b transcription [34,35]. The exact role of these sites in the tiL18 gene promoter remains to be established. In conclusion, the α, β and γ and probably the CAAT box and the box-A elements are sufficient to drive good levels of expression, although we cannot reject the possibility that the addition of downstream sequences (δ elements) could improve the expression of these promoter sequences further.

The L18 promoter function was tested by transient expression in cell culture and in vivo in zebrafish and tilapia embryos. In cell culture, high level LacZ expression was obtained using both constructs. No significant difference was observed between LacZ expression of the longer tiL18-0.5LacZ construct when compared to the short tiL18-0.2LacZ construct in cell culture and in the zebrafish injections. In contrast, in tilapia embryos, the expression obtained with the short construct was significantly higher than with the longer construct (*P* < 0.01). This is highly suggestive of the existence of a tilapia-specific repressor region in the upstream region of the longer construct. It is possible that a transcription factor responsible for the repression is highly expressed in tilapia embryos.

Expression of the L18 gene was observed in *O. mossambicus* in every tissue tested as shown by Northern blotting. We have also recently demonstrated tissue-independent expression of the L18 gene in *O. niloticus* by semiquantitative RT-PCR [18]. Consistent with these results, our microinjection experiments in zebrafish and tilapia embryos did not reveal any particular preference for a given tissue. Expression was seen in a mosaic fashion throughout the embryo and was often restricted to the YSL, a common feature of transient expression in microinjected fish embryos [36].

In conclusion, we present the structure of a tilapia L18 gene and the first promoter sequence and function analysis of a fish ribosomal protein gene. We show that this promoter is likely to be able to drive high levels of expression of a coupled gene of interest in fish.

Acknowledgements

This work was supported by the ‘Région Wallone (ULg 1815); the ‘Services Fédéraux des Affaires Scientifiques, Techniques et Culturelles’ (PAI P4/30 and ‘Actions de Recherche Concertées’: 95/00-193); the Fonds National de la Recherche Scientifique (FNRS) (3.4537.93 and 9.4569.95), the EU (No. BIO4-CT97-0554). M.M. is a ‘Chercheur qualifié’ at the Fonds National de la Recherche Scientifique (FNRS). A.M. held fellowships from ‘CGRI’ and ‘ULg patrimoine’. L.F.M held a scholarship from CAPES, Brazil—BEX1318/99-3).

References


