Expression of a Micro-protein*

Xiang Yu‡ and Jonathan R. Warner§

From the Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461

The smallest known open reading frame encodes the ribosomal protein L41, which in yeast is composed of only 24 amino acids, 17 of which are arginine or lysine. Because of the unique problems that might attend the translation of such a short open reading frame, we have investigated the properties and the translation of the mRNAs encoding L41. In *Saccharomyces cerevisiae* L41 is encoded by two linked genes, *RPL41A* and *RPL41B*. These genes give rise to mRNAs that have short 5′-leaders of 18 and 22 nucleotides and rather long 3′-leaders of 205 and 210 nucleotides not including their poly(A) tails. The mRNAs are translated exclusively on monosomes, suggesting that ribosomes do not remain attached to the mRNA after termination of translation. Calculations based on the abundance of ribosomes and of L41 mRNA indicate that the entire translation event, from initiation through termination, must occur in ~2 s. Termination of translation after only 25 codons does not subject the mRNAs encoding L41 to nonsense-mediated decay. Surprisingly, despite the L41 ribosomal protein being conserved from the archaea through the mammalia, *S. cerevisiae* can grow relatively normally after deletion of both *RPL41A* and *RPL41B*.

Our conventional view of translation is based on mRNAs that can accommodate several ribosomes to form a polyribosome. During translation, their products will pass through a cavity in the large subunit of the ribosome, before emerging at the bottom of the subunit. This passage, originally identified by Yonath *et al.* (1) and recently analyzed at high resolution (2), will accommodate some 30–40 amino acids (3). Because nearly all proteins are synthesized as polypeptides larger than 50 amino acids, they will naturally emerge from the ribosome during translation and be available for the folding chaperones (4).

An exception to both these conventions is ribosomal protein L41. L41 was originally purified from the ribosomes of *Saccharomyces cerevisiae* and identified as a very small, very basic protein that appeared to have orthologues in the ribosomes of eubacteria, depending on the plasmid markers. The RACE method was adopted from Frohman (17), using three adaptor primers referred to as Q1, Q2, and Q3.

* This work was supported in part by National Institutes of Health Grants GM25532 (to J. R. W.) and CA13330 (to the Albert Einstein Cancer Center). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Current address: Bioinformatics Research Center, Campus Box 7566/1501 Partners II Bldg., North Carolina State Univ., Raleigh, NC 27695-7566.

§ To whom correspondence should be addressed. Tel.: 718-430-3022; Fax: 718-430-8597; E-mail: warner@ecomed.yu.edu.

---

* This paper is available on line at http://www.jbc.org

© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.

**The abbreviations used are:** ORF, open reading frame; UTR, untranslated region; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; nt, nucleotide(s); NMD, nonsense mediated decay; kb, kilobase(s).
ng of Q primers and reverse transcribed by 1 μl (200 units) of MMLV reverse transcriptase in a 10-μl reaction. The reaction mixture was diluted to 100 μl, and then two consecutive rounds of PCR amplification were carried out in a 50-μl PCR mixture (1 μM dNTPs, 1× PCR buffer, 1.5 mM Mg²⁺, 2.5 units Taq polymerase). For the first round PCR, 1 μl of diluted cDNA pool together with 25 pmol of Q primers and 25 pmol of gene-specific primer 1 were used. For the second round, 25 pmol of Q primers and 25 pmol gene-specific primer 2 were used. Oligonucleotides used (See Fig. 1, lower panel) were: 

**GSP1(A)** GAAACGACGACATCGATC 
**GSP2(A) CAGTCAACATGAAATGAGCC** 
**GSP1(B) TACGGTATTTGCTCAAGG** 
**GSP2(B) GCTAAGTGGAGAAGAG** 

5′-RACE—5 μg of total RNA was reverse transcribed using 12.5 pmol of gene-specific primer and MMLV reverse transcriptase in a 10-μl reaction. The 5′ partial cDNA pool was diluted to 100 μl purified, concentrated to 10 μl, added to terminal transferase components (0.5 mM ATP, 1× TdT buffer, 0.5 μl of terminal deoxynucleotidyl transferase) to a final volume of 20 μl, and incubated at 37 °C for 20 min. The product was diluted to 100 μl of which 1 μl was used for amplification by two rounds of PCR similar to 3′-RACE. Oligonucleotides used (See Fig. 1, lower panel) were: 

**GSP1(A) GAGTATTTACTATATAATC** 
**GSP2(A) TCCGCTATTTGAGATCGG** 
**GSP1(B) TAACGGATTTGCTCCTCAAGG** 
**GSP2(B) TCCGCTATTTGAGATCGG** 

**Polycistronic Analysis**—Yeast cells were grown in 50 ml of yeast extract-peptone-dextrose to mid-log phase and chilled by addition of crushed ice immediately following the addition of cycloheximide (50 μg/ml). The cells were harvested and washed twice in 0.1 M NaCl, 0.03 mM MgCl₂, 0.01 mM Tris, pH 7.4, 50 μg/ml cycloheximide, 200 μg/ml heparin and resuspended in 0.5 ml LHB buffer (0.1 M NaCl, 0.03 mM MgCl₂, 0.01 mM Tris, pH 7.4). Cells were lysed by vortexing with glass beads and the lysate centrifuged twice for 15 min. at 15,000 × g. The supernatant was layered onto a 7–47% sucrose gradient in TMN solution (0.05 M Tris acetate, pH 7.0, 0.05 M NH₄Cl, 0.012 M MgCl₂) and centrifuged for 1 hr at 150,000 rpm in a SW41 rotor. Gradients were then collected through an ISCO fractionator into Eppendorf tubes containing 0.1 ml of 10% SDS. Each fraction was extracted with hot phenol and the RNA was collected by ethanol, fractionated on a denaturing agarose gel, and subjected to Northern analysis (16). 

**Primer Extension**—Primer Extension System-AMV Reverse Transcriptase (Promega) was used for primer extension analysis. 10 pmol of specific primer was labeled with [γ-32P]ATP (3000 Ci/mmol) using 10 units of T4 polynucleotide kinase, annealed to 10 μg of total RNA, and reverse transcribed using 1 unit of avian myeloblastosis virus reverse transcriptase in a 20-μl reaction at 42 °C for 30 min. Primer extension products were electrophoresed on a denaturing 8% polyacrylamide gel containing 8 μM urea. The gel was dried and exposed in a phosphorimager cassette. 

**One-step Gene Disruption**—The RPL41A genes were deleted using a PCR-based one-step gene disruption (18). A set of oligonucleotides that contain 45 nt of RPL41A flanking sequence and 23 nt of HIS3 sequence were designed to carry out PCR using the HIS3 gene in pRS305 as the template. An analogous set of oligonucleotides for the RPL41B and URA3 genes were used for PCR using the URA3 gene of pRS306. Amplified DNA fragments were purified and used to transform yeast cells by homologous recombination. Transformants were screened on selective plates. 

**Southern Blot Analysis**—Total yeast DNA was digested with EcoRI, separated by electrophoresis on a 0.8% agarose gel, and transferred onto a Zeta-Probe blotting membrane. 32P-labeled DNA probes were prepared by random primer extension of a fragment containing either the RPL41A or RPL41B gene and flanking sequences within the two EcoRI cutting sites. 

**RESULTS AND DISCUSSION**

**RPL41 mRNA Has Unusually Short 5′-UTRs and Unusually Long 3′-UTRs**—To identify the ends of the RPL41 transcripts, we utilized the RACE method (17). cDNAs representing the region between a single point in a mRNA transcript and its 3′- or 5′-end were amplified using PCR. Because the coding sequences of the RPL41 genes are nearly identical, it was necessary to use gene-specific primers, oriented in the direction of the missing sequence. Extension of the partial cDNAs from the unknown end of the message back to the known region is achieved using primers that anneal to the preexisting or an appended poly(A) tail. 

The amplified cDNA fragments were excised from an agarose gel and sequenced (Fig. 1). Unique sequences were found, suggesting that there is a single site of initiation and termination for each gene. Both RPL41A and RPL41B mRNAs have unusually short 5′-UTRs, 22 nucleotides in the case of RPL41A and 18 nucleotides for RPL41B. By contrast the 3′-UTRs are unusually long, ~210 and ~203nts for RPL41A and RPL41B, respectively. A similarly long 3′-UTR has been observed for the human transcript encoding L41 (19). Because the mRNAs end at a series of A residues in the gene (underlined), the exact site of the cleavage that precedes the addition of poly(A) is indeterminate. With the addition of ~50 poly(A) residues, the sizes described in Fig. 1 would be consistent with a published size of 325 nt based on Northern analysis (6). 

Although the signals for cleavage and polyadenylation in *S. cerevisiae* are less rigid than in mammals, three elements have been identified (reviewed in Refs. 20 and 21) and supplemented with an extensive computer analysis (22). Comparison of the presence of these signals in the two RPL41 genes is illuminating. An “upstream” or “efficiency” element, UAUAUA, is present twice in the A gene but not in the B gene. A “positioning” element, AAUAAA, is present in the B gene but only as a variant, AUAUAA, in the A gene. The poly(A) site, itself, generally Y(A)n, is present in the B gene but as the variant YG(A)n in the A gene. Termination at the latter may be enhanced by its (Ur) element that is reported to facilitate 3′-cleavage and poly(A) addition (22). Thus, the two genes each possess some but not all of the elements used to denote 3′-cleavage in yeast. 

The non-coding sequences of the two genes diverge from position –3 upstream of the ORF and from position +8 downstream of the ORF. Indeed, their putative transcription factor binding sites, from positions –200 to –600, differ substantially (23). Yet, so the sum of the effects of transcription factor binding sites, transcription initiation sites, and 3′-cleavage sites provide just the appropriate amount of mRNA to enable equimolar synthesis of L41 with the other ribosomal proteins (see below). 

**RPL41 mRNA Is Exclusively Translated on Single Ribosomes**—To ask how RPL41 mRNA is translated, we carried out a polycistronic analysis. Yeast ribosomes were separated in a 7–47% sucrose gradient, and total RNA from each fraction was extracted and subjected to Northern analysis (Fig. 1). The positions of 18S rRNA and the 25S rRNA indicate the 40S subunits and the 60S subunits, respectively. *ACT1* mRNA, encoding a 478-amino acid protein, is primarily translated on higher order polycistromes, most of which have run to the bottom of the gradient. *RPL30* mRNA, encoding a 104-amino acid protein, is translated largely on dimer and trimer polyosomes. By contrast, *RPL41* mRNA, encoding its 25-amino acid protein, is translated exclusively on single ribosomes, although there is some forward spreading of the peak. 

The most recent estimate, based on sensitivity to hydroxyl radicals, is that about 58 nucleotides of mRNA are protected by a translating 70S-ribosome of *S. cerevisiae* (24). The larger eukaryotic ribosome would probably protect a somewhat longer stretch. Thus, in mid-translation a single ribosome should occupy essentially the entire L41 ORF. When it reaches the termination codon, however, some 50 nucleotides should be available for the binding of a new 43S initiation complex. The few mRNAs with both a 43S-initiation complex as well as an 80S-ribosome might account for the forward spreading from the 80S-peak observed in Fig. 2. Finally, Fig. 2 suggests that the extensive 3′-UTR does not stably associate with a ribosome that has terminated translation. Nevertheless, it is possible that the long 3′-UTR of the RPL41 mRNAs facilitates
the interaction of proteins bound to the 5'-CAP with those bound to poly(A), as suggested in the "circular" polyribosome model recently proposed (25).

Small RPL41 Transcripts Are Not Subject to Nonsense-medi- ated Decay—Nonasen mediated decay (NMD) is a highly conserved mechanism used to rid the cell of mRNAs that are likely to produce aberrant proteins because of premature stop codons introduced either by mutation or by errors in splicing (reviewed in Ref. 26). UPF1 is one of the genes essential for NMD. Deletion of UPF1 results in accumulation of aberrant mRNAs, such as the unspliced transcripts of the ribosomal protein gene, CYH2 (27) (Fig. 3). We asked whether the termination codon of either L41-encoding transcript, only 25 codons from the initiator, would invoke NMD. As shown in Fig. 3, it does not. There is no detectable difference in the level of L41 mRNA between strains carrying a UPF1 or a upf1::LEU2 allele, while the amount of pre-CYH2 mRNA increases substan- tially. Although this is perhaps to be expected because the L41 mRNAs are "natural", this result emphasizes the extraordinary subtlety that one must invoke to explain NMD. The 3'-UTRs of the two L41 transcripts do not appear to contain a sequence match to the putative downstream elements that activate NMD. Deletion of UPF1 is one of the genes essential for NMD (28, 29) although such an element has been identified in only a small fraction of yeast genes.

Is L41 Essential for Cell Growth?—Because most ribosomal proteins are essential, knockout experiments are usually conducted on diploid strains. On the other hand, in cases such as CYH2 in S. cerevisiae, the two RPL41 genes are tightly linked on chromosome IV, separated by only 8 kb. Therefore, we started with a homozygous diploid strain, first disrupting RPL41A with a HIS3 marker (see "Materials and Methods"). Using this strain we disrupted RPL41B with URA3. We expect that half the URA3 disruptants would be on the same chromosome as the HIS3 disruptants. After the second transformation, several colonies that grew on a —His—Ura plate were subjected to PCR to identify colonies in which both the HIS3 and URA3 genes had integrated into the correct loci (data not shown). These double mutants were sporulated, and the resulting tetrads were analyzed. Unexpectedly, colonies with the genotype His"Ura" survived. A Southern blot (Fig. 4) as well as primer extension (see below) demonstrated that both RPL41A and RPL41B had been successfully deleted. The haploid double mutant grows well on a —His—Ura plate at 30 °C as well as at 23 °C or 37 °C, showing that it is neither heat- nor cold-sensitive. As with most ribosomal proteins, the function of L41 is still unknown. Yet, it is surprising that this highly conserved protein seems almost entirely dispensable for growth. On the other hand, there is a report that L41 may be absent from the proteome of Caenorhabditis elegans (30), suggesting that it is not essential even in higher organisms. Unfortunately, the one high-resolution structure of the large subunit is from Halolcaria marismortui, (2) an archaeon that appears not to have a version of L41.
because the small L41 ORFs are omitted from many of the measurements of genome-wide transcription. In heterozygous diploid strains of genotype $\text{RPL41A}/\text{rpl41A}$ or $\text{RPL41B}/\text{rpl41B}$, the amount of L41A or L41B mRNA is reduced $\sim50\%$, respectively, as expected. Interestingly, in a haploid $\Delta\text{RPL41A}$ strain the amount of L41B transcript increases about 35%, suggesting some measure of dosage compensation as has been shown for $\text{CRY2}$, encoding ribosomal protein S14 (32). No obvious change in the level of L41 mRNA was observed in a haploid $\Delta\text{RPL41B}$ strain, perhaps reflecting the relatively larger amount of $\text{RPL41A}$ transcript that is normally present.
In the double knockout strain, no mRNA-encoding L41 is detected.

Translation of L41—A rapidly growing yeast cell has about 200,000 ribosomes. To accommodate a doubling time of 100 min, it must synthesize 2000 ribosomes and 2000 molecules of L41 each minute (33). Our estimate (Fig. 5) of 50–60 mRNAs/cell encoding L41 means that each L41 mRNA is translated 30–40 times/min. Thus, the entire translational process, from initiation through termination, must occupy no more than about 2 s. Perhaps the extended 3′-UTR is necessary for the short L41 mRNA to circularize during translation (30). However, a recent report that a 3′-UTR is necessary for the short L41 mRNA to circularize during translation through interaction of poly(A) binding protein and eIF4G (25).

The structure of the exit passage of the large ribosomal subunit has been solved to 2.4 Å resolution (2). The passage is relatively narrow and tortuous, lined almost entirely with RNA helices. It is remarkable that L41 traverses this passage despite its short length that should preclude its interaction with chaperonins and interact electrostatically with the walls of the passage, and its high concentration of positive charge that should preclude its interaction with chaperonins at the surface of the ribosome that could assist its exit (4).

REFERENCES

1. Yonath, A., Leonard, K. R., and Wittmann, H. G. (1987) Science 236, 813–816
2. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) Science 289, 905–920
3. Malkin, L. I., and Rich, A. (1967) J. Mol. Biol. 26, 329–346
4. McCallum, C. D., Do, H., Johnson, A. E., and Frydman, J. (2000) J. Cell Biol. 149, 591–602
5. Otaka, E., Higo, K., and Itoh, T. (1984) Mol. Gen. Genet. 195, 544–546
6. Suzuki, K., Hashimoto, T., and Otaka, E. (1990) Curr. Genet. 17, 185–190
7. Chan, Y. L., Olivera, J., and Wool, I. G. (1995) Biochem. Biophys. Res. Commun. 214, 810–818
8. Velculescu, V. E., Madden, S. L., Zhang, L., Lash, A. E., Yu, J., Rago, C., Lal, A., Wang, C. J., Beauchy, G. A., Cirillo, K. M., Cook, B. P., Dufault, M. R., Ferguson, A. T., Gao, Y., He, T. C., Hermeking, H., Hirode, S. K., Hwang, P. M., Lopez, M. A., Luderer, H. F., Mathews, B., Petroziello, J. M., Poyak, K., Zawel, L., Kinzler, K. W., and (1998) Nat. Genet. 23, 387–388
9. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FritzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Gledek, A., Scott, J. L., Geoghegan, N. S., and Venter, J. C. (1996) Science 273, 1058–1073
10. Chol, T., Franceschi, F., Yonath, A., and Wittmann-Liebold, B. (1993) Biol. Chem. Hoppe-Seyler 374, 377–383
11. Kawai, S., Murao, S., Mochizuki, M., Shibuya, I., Yano, K., and Takagi, M. (1992) J. Bacteriol. 174, 254–262
12. Dehoux, P., Davies, J., and Cannen, M. (1993) Eur. J. Biochem. 213, 841–848
13. Mager, W. H., Planta, R. J., Ballesta, J. P., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R., and Woolford, J. L., Jr. (1997) Nucleic Acids Res. 25, 4872–4875
14. Thomas, B. J., and Rothstein, R. (1989) Cell 56, 619–630
15. Vilar dell, J., Chartrand, P., Singer, R. A., and Warner, J. R. (2000) RNA 6, 1573–1578
16. Li, B., Nieras, C. R., and Warner, J. R. (1999) Mol. Cell. Biol. 19, 5393–5404
17. Frohman, M. A. (1994) PCR Methods Applications 4, S40–S58
18. Lorenz, M. C., Muir, R. S., Lim, E., McElver, J., Weber, S. C., and Heitman, J. (1998) Gene 158, 113–117
19. Klaudiny, J., von der, K. H., and Scheit, K. H. (1992) Biochem. Biophys. Res. Commun. 187, 901–906
20. Guo, Z., and Sherman, F. (1996) Trends Biochem. Sci. 21, 477–481
21. Zhao, J., Hyman, L., and Moore, C. (1999) Microbiol. Mol. Biol. Rev. 63, 405–445
22. Graber, J. H., Cantor, C. R., Mohr, S. C., and Smith, T. F. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 14055–14060
23. Lascaris, R. F., Mager, W. H., and Planta, R. J. (1999) Bioinformatics 15, 267–277
24. Huttenhofer, A., and Noller, H. F. (1994) EMBO J. 13, 3892–3901
25. Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B. (1998) Mol. Cell. Biol. 2, 135–140
26. Czaplinski, K., Ruiz-Echevarria, M. J., Gonzalez, C. I., and Peltz, S. W. (1999) BioEssays 21, 685–696
27. He, F., Peltz, S. W., Denahue, J. L., Rosbash, M., and Jacobson, A. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 7034–7038
28. Hagan, K. W., Ruiz-Echevarria, M. J., Quan, Y., and Peltz, S. W. (1999) Mol. Cell. Biol. 19, 809–823
29. Zhang, S., Ruiz-Echevarria, M. J., Quan, Y., and Peltz, S. W. (1995) Mol. Cell. Biol. 15, 2231–2244
30. Wheelan, S. J., Boguski, M. S., Duret, L., and Makalowski, W. (1999) Gene 238, 163–170
31. Holstege, F. C. P., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Lewis, R. L., Guarente, L., Golub, T. R., Lander, E. S., and Young, R. A. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5604–5609
32. Li, Z., Paulovich, A. G., and Woolford, J. L., Jr. (1995) Mol. Cell. Biol. 15, 6454–6464
33. Warner, J. R. (1999) Trends Biochem. Sci. 24, 437–440
Expression of a Micro-protein
Xiang Yu and Jonathan R. Warner

J. Biol. Chem. 2001, 276:33821-33825.
doi: 10.1074/jbc.M103772200 originally published online July 12, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M103772200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 33 references, 13 of which can be accessed free at
http://www.jbc.org/content/276/36/33821.full.html#ref-list-1