

# Ribosomal protein L4 of *Saccharomyces cerevisiae*: the gene and its protein

Sigifredo G. Arevalo\* and Jonathan R. Warner<sup>†</sup>

Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Received December 12, 1989; Revised and Accepted February 20, 1990

EMBL accession no. X17204

## ABSTRACT

**The sequence of a gene for ribosomal protein L4 of *Saccharomyces cerevisiae* has been determined. Unlike most ribosomal protein genes of *S. cerevisiae* this gene has no intron. The single open reading frame predicts that L4 is highly homologous to mammalian ribosomal protein L7a. There appear to be two genes for L4, both of which are active.**

## INTRODUCTION

As part of an ongoing effort to determine the structure of the ribosome of the yeast *Saccharomyces cerevisiae*, we and others have cloned the genes for a number of the ribosomal proteins. (Reviewed in 1) Our hope is that the genetic approaches available in the yeast system will provide opportunities for greater insight into the structure and function of the eukaryotic ribosome. We present here the sequence of the gene for ribosomal protein L4 (formerly called RP6), the third largest protein of the 60S subunit (2), and the predicted amino acid sequence of the protein that is derived from it. L4 is highly homologous to mammalian ribosomal protein L7a. (3,4) There are two genes for L4, both of which appear to be active. The gene studied in this paper is termed *RPL4A*.

## MATERIALS AND METHODS

**Strains and Plasmids.** *Saccharomyces cerevisiae*, strain W303 (*MATa/α*, *ade2-1*, *his3-11,15*, *leu2-3,112*, *trp1-1*, *ura3-1*, *can1-100*) is a homozygous diploid obtained from R. Rothstein (Columbia University). A haploid *MATa* strain was derived from a spore of W303. Yeast strains were grown in synthetic medium and plasmids carrying the *URA3* gene were maintained by growth in media lacking uracil. Recombinant DNA techniques and preparation of nucleic acids were carried out essentially as described in (5). Plasmid pYERP6, containing a 9.2 Kb EcoRI/EcoRI fragment with the *RPL4A* gene, was derived from a lambda clone (A83) described previously (2). *Escherichia coli* strain DH5α was used throughout for the transformation and propagation of the plasmids.

**DNA Preparation, Labeling and Sequencing.** Plasmid DNA preparations, restriction enzyme digestions, and ligation reactions were performed as described (5). DNA was labeled by nick translation using ( $\alpha$ -<sup>32</sup>P)dCTP (Amersham Corp., 3000 Ci/mmol) and DNA polymerase. Oligonucleotide probes were 5'-end labeled with ( $\gamma$ -<sup>32</sup>P)ATP (Amersham Corp., 3000 Ci/mmol) and T4 polynucleotide kinase. DNA fragments were subcloned in the vector pGEM blue (Promega) and their sequence determined by the dideoxy chain termination method (6) applied to DNA minipreps (7).

**DNA Blot Analysis.** Restriction enzyme digested DNA was blotted to Nytran (Schleicher & Schuell) and probed with nick translated DNA fragments or with kinased oligonucleotides.

## RESULTS AND DISCUSSION

### Subcloning and Sequencing of *RPL4A*

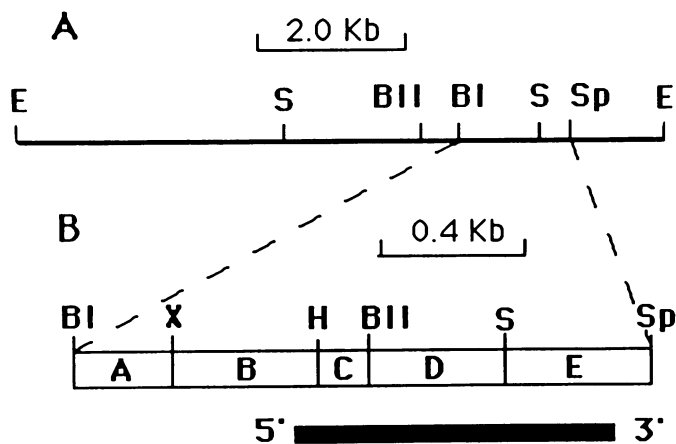
*RPL4A* was originally cloned as a 9.2 kb EcoRI fragment in a lambda vector. The identity of the product of the gene as ribosomal protein L4 was established by 2D gel analysis of the *in vitro* translation of mRNA hybridizing to the DNA fragment (2). Further characterization led to the restriction map shown in Figure 1A. Several of the fragments (A-E in Figure 1B) were subcloned into pGEM vectors. The location and orientation of the transcript (Figure 1B) was determined originally by probing northern blots of yeast RNA with RNA probes transcribed from these subclones (data not shown). Dideoxy sequencing (6) was carried out using either primers from within the vector sequences or primers synthesized according to preliminary sequence data as it was obtained.

The sequence of the *RPL4A* gene is shown in Figure 2. There is a single open reading frame coding for 256 amino acids, suggesting that the gene is uninterrupted by an intron. This view is confirmed by the lack of the canonical sequences (8) characteristic of the 5' splice site (GTATGT) or of the lariat site (TACTAAC), either within or upstream of the open reading frame.

The sequences upstream of the open reading frame are also characteristic of a ribosomal protein, with its tripartite promoter. (9,10,11) There is a putative TATA sequence at -95 (Figure

\* Present address: Facultad de Ciencias Químicas, Universidad Autónoma de Chihuahua, Apartado Postal 669, Chihuahua, Chih., Mexico

<sup>†</sup> Author to whom correspondence should be addressed



**Figure 1:** A. Restriction map of the 9.2 EcoRI fragment that contains *RPL4A*. B. Restriction map and subclones A-E. The 3.5 Kb Sall-SalI and 1.75 Kb Sall-EcoRI fragments were isolated from the lambda clone A83,(2) subcloned into fragments A-E that were then sequenced. The solid bar represents the open reading frame predicted from the nucleotide sequence. BI = BallI, BII = BglII, E = EcoRI, H = HindIII, S = Sall, Sp = SpeI, X = XbaI.

2A), the T-rich region from -162 to -213, and at -227 to -240 a potential RPG box, the upstream activating sequence for transcription of genes for ribosomal proteins, translation factors, RNA polymerase subunits and many others. (Reviewed in 1)

The codon usage for L4, like that for other ribosomal proteins of yeast, is characteristic of an abundant protein. The predicted protein is highly basic, the basic amino acids outnumbering the acidic ones by 49 to 23, also a characteristic of ribosomal proteins. The predicted molecular weight and charge of L4 are entirely consistent with its migration on a two dimensional polyacrylamide gel. (12)

**Yeast L4 is Homologous to Mammalian L7a**

Comparison of the sequence of L4 with several databanks led to the finding that it is highly homologous to the mammalian ribosomal protein L7a, sequenced in human (3) and mouse (4). (See Figure 3) (Note that the mammalian ribosomal proteins L7 and L7a are entirely distinct (13)). Mammalian L7a is slightly longer than yeast L4 at both the N and C terminal ends, but the rest of the sequence is homologous throughout. This finding confirms the extraordinary conservation of the sequence of ribosomal proteins through the evolution of the eukaryotes (reviewed in 1, 14). However, L4 does not have substantial identity with any of the ribosomal proteins of *E. coli*.

A particularly interesting aspect of the sequence is the relatively hydrophobic region of amino acids 156-177, rather uncharacteristic of ribosomal proteins. This region is 95% identical between human and yeast, suggesting that it has an important function.

**There are Two Functional Genes for L4**

Many of the proteins of the yeast translation apparatus, including a majority of the ribosomal proteins, are encoded by two genes. (1) To ask if L4 were in this category, a genomic Southern was probed with an oligonucleotide from just outside the coding region (Figure 4A) and with another from within the coding region (Figure 4B). In the former case we detect only one band, in the latter two, even under stringent washing conditions, suggesting that there are two genes.

**SEQUENCE OF RPL4A OF *S. CEREVISIAE***

|             |             |             |             |             |             |      |            |      |            |
|-------------|-------------|-------------|-------------|-------------|-------------|------|------------|------|------------|
| -626        | TATTGC      | -610        | GCTTCAAGG   | -600        | ATGTGAACGA  | -590 | GCGCCCATTA | -580 | TGGAAGGACC |
| -570        | AGCCAATATG  | -560        | GCTTCAAGG   | -550        | ATGTGAACGA  | -540 | GCGCCCATTA | -530 | TGGAAGGACC |
| TACACGATG   | TAAGAAATGG  | GTCCTTTTACA | ATGGAATGGA  | GAAAAACGAA  | CAAATATCAC  |      |            |      |            |
| -510        | -500        | -490        | -480        | -470        | -460        |      |            |      |            |
| CAGAGGCTCA  | TCCAACCTT   | AGATATTCA   | GAAGTGAATC  | CGACGATGCT  | GAAGGAGAA   |      |            |      |            |
| -450        | -440        | -430        | -420        | -410        | -400        |      |            |      |            |
| AGCATTGGAA  | GTATTGGGAA  | GTTCCCGTGG  | AAAGATCCAA  | AGTTGGCTG   | AAGACATAG   |      |            |      |            |
| -390        | -380        | -370        | -360        | -350        | -340        |      |            |      |            |
| AAGTGAACA   | CGAACAGTT   | CTAGAAAAT   | TCCAATAACA  | ACATAAATA   | TTTCTATTAA  |      |            |      |            |
| -330        | -320        | -310        | -300        | -290        | -280        |      |            |      |            |
| CAATGTAAT   | TCATAATTT   | TATATTCCTC  | TCCACCTTCT  | ATTGCATCAT  | GTACTATTCA  |      |            |      |            |
| -270        | -260        | -250        | -240        | -230        | -220        |      |            |      |            |
| AATGACTGTA  | ACACTAGTAT  | TATGAAGAAA  | ACACCCAAAC  | ATATCTAGGC  | CATCAGATTT  |      |            |      |            |
| -210        | -200        | -190        | -180        | -170        | -160        |      |            |      |            |
| TTTTTTTTTC  | ATTTTTCAAT  | TTTTTCTCAT  | TTTCTTATTT  | ATTTTTATTG  | AAAAATAATA  |      |            |      |            |
| -150        | -140        | -130        | -120        | -110        | -100        |      |            |      |            |
| ACCGACGCAA  | ACAAATTGGA  | AAAACCAACG  | CAAAAAAAA   | AGACCTATA   | TTGTTATAA   |      |            |      |            |
| -90         | -80         | -70         | -60         | -50         | -40         |      |            |      |            |
| AGGGGAGGAA  | TTGTATCTA   | TCAATTAATA  | TTCCAGTTGT  | CAGTTTACAT  | GTCTACCCCT  |      |            |      |            |
| -30         | -20         | -10         |             |             |             |      |            |      |            |
| CTATTATCAC  | ATCAAAACAA  | CTAATTCGAA  | ATG GCC CCA | GGT AAG AAA | GTC GCT CCA |      |            |      |            |
|             |             |             | Met Ala Pro | Gly Lys Lys | Val Ala Pro |      |            |      |            |
| 30          | 45          | 60          | 75          | 90          | 105         | 120  | 135        | 150  | 165        |
| GCT CCA TTC | GGT GCT AAG | TCA ACT AAG | TCT AAC AAG | ACT AGA AAC | CCA TTG ACT |      |            |      |            |
| Ala Pro Phe | Gly Ala Lys | Ser Thr Lys | Ser Asn Lys | Thr Arg Asn | Pro Leu Thr |      |            |      |            |
| 90          | 105         | 120         | 135         | 150         | 165         |      |            |      |            |
| CAC TCT ACT | CCA AAG AAC | TTC GGT ATT | GGT CAA GCT | GTC CAA CCA | AAG AGA AAC |      |            |      |            |
| His Ser Thr | Pro Lys Asn | Phe Gly Ile | Gly Gln Ala | Val Gln Pro | Lys Arg Asn |      |            |      |            |
| 150         | 165         | 180         | 195         | 210         | 225         |      |            |      |            |
| TTG TCC AGA | TAC GTC AAA | TGG CCA GAA | TAC GTC AGA | GTG CAA AGA | CAA AAG AAG |      |            |      |            |
| Leu Ser Arg | Tyr Val Lys | Trp Pro Glu | Tyr Val Arg | Val Gln Arg | Gln Lys Lys |      |            |      |            |
| 195         | 210         | 225         | 240         | 255         | 270         |      |            |      |            |
| ATC TTG TCC | ATC AGA TTG | AAG GTT CCT | CCA ACC ATT | GCT CAA TTC | CAA TAC ACT |      |            |      |            |
| Ile Leu Ser | Ile Arg Leu | Lys Val Pro | Pro Thr Ile | Ala Gln Phe | Gln Tyr Thr |      |            |      |            |
| 255         | 270         | 285         | 300         | 315         | 330         |      |            |      |            |
| TTG GAC AGA | AAC ACC GCT | GCC GAA ACC | TTC AAG TTG | TTC AAC AAG | TAC AGA CCA |      |            |      |            |
| Lys Asp Arg | Asn Thr Ala | Ala Glu Thr | Phe Lys Leu | Phe Asn Lys | Tyr Arg Pro |      |            |      |            |
| 300         | 315         | 330         | 345         | 360         | 375         |      |            |      |            |
| GAA ACT GCT | GCT GAA AAG | AAG GAA AGA | TTG ACC AAA | GAA CGT GCC | GCT GTT GCT |      |            |      |            |
| Glu Thr Ala | Ala Glu Lys | Lys Glu Arg | Leu Thr Lys | Glu Arg Ala | Ala Val Ala |      |            |      |            |
| 360         | 375         | 390         | 405         | 420         | 435         |      |            |      |            |
| GAA GGT AAG | TCC AAG CAA | GAT GCT TCT | CCA AAG CCA | TAC GCT GTC | AAG TAC GGT |      |            |      |            |
| Glu Gly Lys | Ser Lys Gln | Asp Ala Ser | Pro Lys Pro | Tyr Ala Val | Lys Tyr Gly |      |            |      |            |
| 420         | 435         | 450         | 465         | 480         | 495         |      |            |      |            |
| TTG AAC CAC | GTT GTT GCC | TTG ATC GAA | AAC AAG AAG | GCT AAG TTG | GTT ATT ATT |      |            |      |            |
| Leu Asn His | Val Val Ala | Leu Ile Glu | Asn Lys Lys | Ala Lys Leu | Val Leu Ile |      |            |      |            |
| 465         | 480         | 495         | 510         | 525         | 540         |      |            |      |            |
| GCT AAC GAT | GTC GAC CCA | ATT GAA TTG | GTT GTC TTT | TTG CCA GCT | TTG TGT AAG |      |            |      |            |
| Ala Asn Asp | Val Asp Pro | Ile Glu Leu | Val Val Phe | Lys Pro Ala | Leu Lys Lys |      |            |      |            |
| 525         | 540         | 555         | 570         | 585         | 600         |      |            |      |            |
| AAG ATG GGT | GTT CCA TAC | GCC ATT GTC | AAG GGT AAG | GCT AGA TTG | GGT ACC TTG |      |            |      |            |
| Lys Met Gly | Val Pro Tyr | Ala Ile Val | Lys Gly Lys | Ala Arg Leu | Gly Thr Leu |      |            |      |            |
| 570         | 585         | 600         | 615         | 630         | 645         |      |            |      |            |
| GTT AAC CAA | AAG ACC TCT | GCC GTT GCC | GCT TTG ACT | GAA GTC AGA | GCC GAA GAC |      |            |      |            |
| Val Asn Gln | Lys Thr Ser | Ala Val Ala | Ala Leu Thr | Glu Val Arg | Ala Glu Asp |      |            |      |            |
| 630         | 645         | 660         | 675         | 690         | 705         |      |            |      |            |
| GAA GCT GCT | TTG GCT AAG | TTG GTT TCT | ACC ATT GAC | GCT AAC TTC | GCT GAC AAA |      |            |      |            |
| Glu Ala Ala | Leu Ala Lys | Leu Val Ser | Thr Ile Asp | Ala Asn Phe | Ala Asp Lys |      |            |      |            |
| 690         | 705         | 720         | 735         | 750         | 765         |      |            |      |            |
| TAC GAT GAA | GTC AAG AAG | CAC TGG GGT | GGT GGT ATC | CTT GGT AAC | AAG GCT CAA |      |            |      |            |
| Tyr Asp Glu | Val Lys Lys | His Trp Gly | Gly Ile Lys | Gly Lys Pro | Ala Lys Gln |      |            |      |            |
| 735         | 750         | 765         | 780         | 795         | 810         |      |            |      |            |
| GCC AAG ATG | GAC AAG AGA | GCT AAG AAC | TCC GAC TCC | GCT TAA     | ATTGAAAAT   |      |            |      |            |
| Ala Lys Met | Asp Lys Arg | Ala Lys Asn | Ser Asp Ser | Ala *       |             |      |            |      |            |
| 790         | 800         | 810         | 820         | 830         | 840         |      |            |      |            |
| GAGAAATTTT  | GCATAAAAAA  | TTTATTTTTT  | AATAAATAA   | TTAATTTTCA  | AACGTTTTAA  |      |            |      |            |
| 850         | 860         | 870         | 880         | 890         | 900         |      |            |      |            |
| TATAATAATG  | ACCTTAAATA  | ACCTTTTAAAG | GTGATGCTT   | TGTTGAGCTT  | TTGACTTTTG  |      |            |      |            |
| 910         | 920         | 930         | 940         | 950         | 960         |      |            |      |            |
| GACCATGGGA  | AGTACCTCAT  | TATAGGAAAA  | TCATATCGAC  | TTATTTTGA   | TATCACAAATA |      |            |      |            |
| 970         |             |             |             |             |             |      |            |      |            |
| ATTTTGCATG  |             |             |             |             |             |      |            |      |            |

**Figure 2:** Nucleotide sequence of the *RPL4A* gene and predicted amino acid sequence of yeast ribosomal protein L4 (EMBL Accession # X17204). The nucleotides are numbered with respect to the translation initiation codon, designated +1.

To ask about the function of *RPL4A*, it was disrupted by insertion of a 1.1 kb HindIII fragment carrying the *URA3* gene into a BallI site downstream of codon 52 in the plasmid pYEPL4A. (See Figure 2) The 3.5 kb Sall fragment derived from the resulting construct was transformed into a diploid cell, where it could recombine with one of the two authentic *RPL4A* genes (15). Selection on -Ura medium yielded transformants that have the expected 1.1 kb insert in one of their two *RPL4A* genes. (Figure 5) On sporulation of this diploid, and subsequent dissection of the tetrads, we observed four viable spores, two of which gave rise to significantly smaller colonies. Figure 5 shows that two of the spores had no wild type copy of *RPL4A*. These were Ura<sup>+</sup> and grew slowly, with a doubling time 30% longer than wild type cells. Therefore *RPL4A* must be functional.

Two dimensional polyacrylamide gel analysis of the ribosomal proteins of the Ura<sup>+</sup> cells showed that the ratio of L4 to other

