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

Sigifredo G.Arevalo* and Jonathan R.Warner*

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 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($MATa/\alpha$, 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 DH5 α 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 (α -³²P)dCTP (Amersham Corp., 3000 Ci/mmol.) and DNA polymerase. Oligonucleotide probes were 5'-end labeled with (γ -³²P)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 RPLAA

RPLAA 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

+Author to whom correspondence should be addressed

^{*} Present address: Facultad de Ciencias Quimicas, Universidad Autonoma de Chihuahua, Apartado Postal 669, Chihuahua, Chih., Mexico

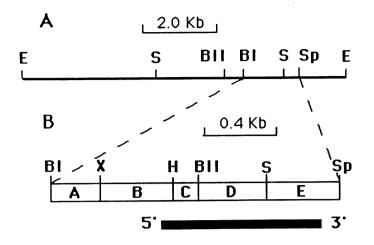


Figure 1: A. Restriction map of the 9.2 EcoRI fragment that contains *RPL4A*. B. Restriction map and subclones A-E. The 3.5 Kb SalI-SalI and 1.75 Kb SalI-EcoRI fragments were isolated from the lambda clone A83,(2) subcloned into fragments A-E that were then sequenced. The solid bar represents the open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

		~~~					
-626 Tattgc	-620 Agccaatatg	-610 GCTTTCAAGG	-600 Atgtgaacga	-590 GCGCCCATTA	-580 Tggaagacc		
-570	-560	-550	-540	-530	-520		
TACACGATGT	TAAGAAATGG	GTCTTTTACA	TTTTACA ATGGAATGGA GAAAAACGAA CAAATAT				
-510	-500	-490	-480	-470	-460		
CAGAGGCTCA -450	TCCAAACCTT -440	AAGATATTCA -430	GAAGTGAATC -420	CGACGATGCT -410	GAAAGGAGAA -400		
AGCATTGGAA	GTATTGGGAA	GTTGCCGTGG	AAAGATCCAA	AGGTTGGCTG	AAGGACATAG		
-390	-380	-370	-360	-350	-340		
AAGGTGAACA	CGAACAGGTT	стадалалст	TCCAATAACA	асаталатал	TTTCTATTAA		
-330 С <b>алтсталтт</b>	-320 TCCATAATTT	-310 TATATTCCTC	-300 TCCACCTTCT	-290 Attgcatcat	-280 GTACTATTCA		
-270	-260	-250	-240	-230	-220		
AATGACTGTA	ACACTAGTAT	татдаадааа	ACACCCAAAC	ATATCTAGGC	CATCAGATTT		
-210	-200	-190	-180	-170	-160		
TTTTTTTTTTC -150	ATTTTTCATT -140	TTTTTCTCAT -130	TTTCTTATTT -120	ATTTTTATTG -110	AAAAATAATA -100		
ACCGACGCAA	ACAAATTGGA	AAAACCAACG	САЛАЛАЛАЛА	AGACGCTAAA	TTGTTTATAA		
-90	-80	-70	-60	-50	-40		
AGGCGAGGAA -30	TTTGTATCTA -20	TCAATTACTA -10	TTCCAGTTGT	CAGTTTACAT 15	TGCTTACCCT		
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		
		TCA ACT AAG Ser Thr Lys					
90	ory and bys	105	ber Abn Djo	120	135		
		TTC GGT ATT					
His Ser Thr	Pro Lys Asn 150	Phe Gly Ile	Giy Gin Ala 165	Val Gin Pro 180	Lys Arg Asn		
TTG TCC AGA		TGG CCA GAA			CAA AAG AAG		
	Tyr Val Lys	Trp Pro Glu		Val Gln Arg			
195	100 1C1 000	210 AAG GTT CCT	225	CCT CNN TTC	240		
Ile Leu Ser		Lys Val Pro					
	255	270		285			
		GCC GAA ACC					
Leu Asp Arg 300	Asn Thr Ala 315	Ala Glu Thr	330	Phe Ash Lys	345		
		AAG GAA AGA		GAA CGT GCC			
		Lys Glu Arg		Glu Arg Ala			
360		375		390	405		
		GAT GCT TCT Asp Ala Ser					
GIG GIY LYS	420	YPD YIG DEL	435	450	5,5 1,1 01,		
		TTG ATC GAA					
	Val Val Ala	Leu Ile Glu	Asn Lys Lys 495	Ala Lys Leu	Val Leu Ile 510		
465 CCT AAC CAT	GTC GAC CCA	480 ATT GAA TTG		TTG CCA GCT			
Ala Asn Asp	Val Asp Pro	Ile Glu Leu	Val Val Phe	Leu Pro Ala	Leu Cys Lys		
-	525	540		555			
AAG ATG GGT		GCC ATT GTC					
		Ala Ile Val	Lys Gly Lys	Ala Arg Leu			
570	585	GCC GTT GCC	600		615 600 GAA GAC		
Val Asn Gln	Lys Thr Ser	Ala Val Ala	Ala Leu Thr	Glu Val Arg	Ala Glu Asp		
630		645		660	675		
		TTG GTT TCT Leu Val Ser					
GIG MIG MIG	690	Det var Ser	705	720	VIG VOD DA2		
		CAC TGG GGT					
Tyr Asp Glu 735	Val Lys Lys	His Trp Gly 750	Gly Gly Ile 765	Leu Gly Asn	Lys Ala Gln 780		
	GAC ANG AGA	GCT ANG ANC	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 850	GCATAAAAAA 860	TTTATTTTTT 870	<b>AATAATAATA</b> 880	TTAATTTTCA 890	AACGTTTTAA 900		
TATAATAATG	ACCTTAAATA	ACCTTTTAAG	GTGATTGCTT	TGTTGAAGCT	TTGATACTTG		
910	920	930	940	950	960		
GACCATGGGA 970	AGTACCTCAT	татаддалал	TCATATCGAC	TTATTTTGAA	TATCACAATA		
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 Ball 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⁺ 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  $Ura^+$  cells showed that the ratio of L4 to other

	сон	PARISON	OF HUN	IAN L7a V	ITH S.C	EREVISIJ	E L4A		
	10		20	30		0	50		60
SCL4A	MAPGKKVAPA	PFGAKST	KSNKTRI	PLTHSTP: ::: :	KNFGIGQA	VOPKRNL	SRYVK	PEYVRV	2R
HIIMI.7a	: :::::: KAKGKKVAPA	PAVVKKO	EAKKVVI	PLFEKRP	KNFGIGOD	IOPKRDL	TRFVK	PRYIRL	QR
	10	20		30	40	50		60	
	70	)	80	90	10	0	110	1:	20
SCL4A	QKKILSIRL							ERAAVAE	GK
	:. :: :::	:::.:.	: .::			:: .:::		. :.	::
HUML7a	QRAILYKRLE 70	WPPAINQ 80	FTQALD	RQTATQLL 90	100	110	QRLLA	120	GR
	130		140			50	170		80
SCL4A	SKQDASPKPY	AVKYGLN	HVVALI	ENKKAKLV	LIANDVDE	PIELVVFL	PALCK	KMGVPYA	IV
UTIME 7 a	G-DVPTKRPI	VI.PAGVN	::. TVTTI.V	ENKKAOLV	VTAHDVDI	PTELVVFL	PALCR	KMGVPYC	11 11
nombra	130	140		150	160	170		180	
	190	)	200	210	22	20	230	2	40
SCL4A	KGKARLGTL				AKLVSTI	DANFADKY	DEVKK	HWGGGIL	GN
HUML7a	KGKARLGRL				AKLVEAII 220	RTNYNDRY 230		HWGGNVL 240	GP
	190	200		210	220	230		240	
	250	)							
SCL4A	KAQAKMDKR			57.2%	identit	ty in 25	0 aa	overlap	,
HUML/a	KSVARIAKLI 250	CKAKAKEI 260	ATKLG						

Figure 3: Amino acid sequence homology between *Saccharomyces* L4 and *Homo* L7a (HUML7a) (3). The predicted amino acid sequences of L4 and L7a were compared (18). Identical amino acid residues are indicated by two dots, conservative amino acid substitutions by one dot.

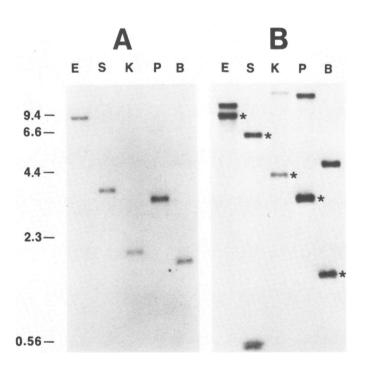


Figure 4: Copy number analysis of genes for L4. A Southern blot of yeast genomic DNA (20 ug) digested either with EcoR1, Sal1, Kpn1, Pst1 or BgII, was hybridized to two kinased oligonucleotides derived from the *RPL4A* gene. One oligonucleotide (A) was from the coding region (+611 to +630); the other (B) was from a non-coding region (-146 to -127). The numbers on the left indicate the size of the lambda HindIII markers. Note that restriction enzyme sites for SalI, KpnI, and BgIII lie between the two probes from *RPL4A*; therefore the sizes of the fragments in A and B are not the same. The bands representing the *RPL4A* gene in B are indicated by asterisks.

ribosomal proteins was near wildtype (data not shown). This result demonstrates that the other copy of the L4 gene, *RPL4B*, is also functional. These results are similar to those obtained with a number of other ribosomal protein genes in yeast, e.g. RP51 (16) which has two active genes each of which is capable of

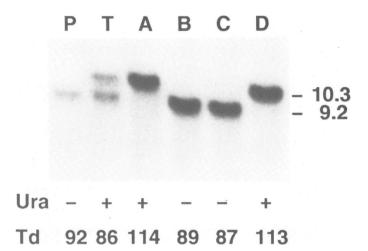


Figure 5: Gene disruption of *RPL4A*. A: Genomic Southern blot of the parental diploid strain (P), a transformed diploid transformed with *RPL4A* into which *URA3* had been inserted (T), and four spores of a tetrad derived from the transformed diploid (A – D). DNA samples were digested with EcoRI and probed with a nick translated 0.77 kb BgIII-XbaI fragment that is upstream of the coding region (see Figure 1). The insertion of the 1.1 kb *URA3* gene increases the size of the EcoRI fragment from 9.2 to 10.3 kb. Its presence in the 10.3 kb fragment was confirmed by reprobing the blot with *URA3* sequences (data not shown). The Ura phenotype and the doubling time (Td) of each strain is shown (see text).

supporting growth, albeit at a reduced rate. It is curious that *S*. *cerevisiae* has this duplication of ribosomal protein genes whereas mammalian organisms have numerous pseudogenes but only a single functional gene (17).

### ACKNOWLEDGEMENTS

This research was supported by Grant # GM25532 of NIGMS and by the Cancer Center Grant # CA13330. The oligonucleotides were supplied by a Facility originally funded from an NSF grant # PCM-8400114. The authors are grateful to Sze-Fong Ng and to Donna Rousso, who carried out early experiments on *RPL4A*.

#### REFERENCES

- 1. Warner, J.R. (1989) Microbiol. Rev. 53, 256-271.
- Fried, H.M., Pearson, N.J., Kim, C.H. & Warner, J.R. (1981) J. Biol. Chem. 256, 10176-10183.
- Kozma, S.C., Redmond, S.M.S., Xiao-Chang, F., Saurer, S.M., Groner, B. & Hynes, N.E. (1988) EMBO J. 7, 147-154.
- 4. Giallongo, A., Yon, J. & Fried, M. (1989) Molec. Cell. Biol. 9, 224-231.
- 5. Johnson, S.P. & Warner, J.R. (1989) Molec. Cell. Biol. 9, 4986-4993.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- Kraft, R., Tardiff, J., Krauter, K.S. & Leinwand, L.A. (1988) Biotechniques 6, 544-547.
- 8. Woolford, J.L. (1989) Yeast 5, 439-457.
- 9. Rotenberg, M.O. & Woolford, J.L. (1986) Mol. Cell. Biol. 6, 674-687.
- Woudt, L.P., Smith, A.B., Mager, W.H. & Planta, R.J. (1986) EMBO J. 5, 1037 – 1040.
- 11. Schwindinger, W.F. & Warner, J.R. (1987) J. Biol. Chem. 262, 5690-5695.
- 12. Warner, J.R. & Gorenstein, C. (1978) Methods Cell Biol. 20, 45-60.
- Lin, A., Chan, Y.-L., McNally, J., Peleg, D., Meyuhas, O. & Wool, I.G. (1987) J. Biol. Chem. 262, 12665-12671.
- Wool, I.G., Endo, Y., Chan, Y.-L. & Gluck, A. (1990) in *Ribosomes*, ed. W. Hill, ASM Press, Washington, in press.
- 15. Rothstein, R.J. (1983) Methods Enzymol. 101, 202-211.
- 16. Abovich, N. & Rosbash, M. (1984) Mol. Cell. Biol. 4, 1871-1879.
- 17. Dudov, K.P. & Perry, R.P. (1984) Cell 37, 457-468.
- 18. Pearson, W.R. & Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444-2448.