The primary structure of the gene encoding yeast ribosomal protein L16

Robert J. Leer, Mary M.C. van Raamsdonk-Duin, Willem H. Mager and Rudi J. Planta*

Biochemisch Laboratorium, Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Received 1 July 1984; revised version received 14 August 1984

As part of our studies on the molecular basis for the coordinate expression of ribosomal protein genes in yeast we analyzed the primary structure of the gene encoding protein L16 of the large ribosomal subunit including the flanking sequences. L16 turned out to be a ribosomal protein with a molecular mass of 22662 Da and a net charge of +12. Both the 5'- and the 3'-end of the L16 mRNA were mapped by primer extension and S_1 nuclease analysis. In the DNA regions flanking the coding sequence several conserved elements are present that may be involved in transcription initiation or termination.

Yeast *Ribosomal protein gene* *Sequence analysis* *Transcript mapping*

1. INTRODUCTION

Yeast cells can accurately control the production of their ribosomes under various conditions of growth. Apart from the coordination between ribosome biosynthesis and cellular growth rate, ribosomal protein synthesis appears to be in perfect balance with ribosomal RNA synthesis under almost all conditions (see [1] and [2] for reviews). To get insight into the molecular mechanisms underlying the coordinate control of ribosomal protein synthesis in yeast, we undertook a detailed analysis of the organization and structure of several ribosomal protein genes. Here we describe the analysis of the gene, including its flanking sequences, encoding protein L16 of the large ribosomal subunit.

2. MATERIALS AND METHODS

2.1. *Isolation of nucleic acids*

Recombinant plasmid pBMCY89 [3] was purified from Triton-treated bacterial spheroplasts by CsCl-ethidium bromide density gradient centrifugation [4]. A HindIII–BamHI fragment containing the L16 gene was subcloned in pBR322 (subclone 89). Poly(A)-containing RNA was isolated from the yeast *Saccharomyces carlsbergensis* S74 as described previously [5].

2.2. *Restriction site mapping*

Restriction endonucleases were obtained from Bethesda Research Laboratories or New England Laboratories and used as recommended by the supplier. Cleavage sites for Sau3A, TaqI, MspI and AluI were determined by the partial digestion procedure of [6].

2.3. *DNA sequence analysis*

DNA sequence analysis was performed using the chain termination method [7]. Single-stranded templates were obtained by transforming JM101 cells with recombinant bacteriophage M13 RF DNA [8]. The M13 vector used was M13 mp9 [9].

2.4. *Mapping of the 5'-end of the L16 gene*

For this purpose the base-specific chain termination method using reverse transcriptase (Life Sciences Inc.) was employed [10]. A 29 nucleotides long Sau3A-AluI primer (SA29 in fig.1) was labelled and isolated essentially as described in Method A used for 3'-end mapping (see below).

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

The primer fragment (60 000 cpm) was annealed at 65°C for 30 min to 25 μg of poly(A)-containing RNA in 30 μl 50 mM Tris-HCl (pH 8.3), 6 mM $MgCl_2$ and 50 mM KCl. The base-specific reaction mixtures contained 5.5 μl hybrids, 2 μl 0.5 mM dNTPs, 2 μl of either 0.5 mM ddNTPs, 0.3 μl reverse transcriptase (5 U), 1 μl 100 mM DTT and 0.3 μl actinomycin D (2 mg/ml). No ddNTPs were added in the 'total' reaction mixture. Incubation was performed in eppendorf tubes during 30 min at 42°C. After hydrolysis of the RNA, cDNA was precipitated and analyzed [11].

2.5. *Mapping of the 3'-end of the L16 gene*

Two methods were applied to achieve labelled probes. In Method A the Sau3A-generated fragment ligated into the bacteriophage M13 mp9 was labelled by synthesis of the complementary strand in the presence of Klenow polymerase and [α-^{32}P]-dATP [12]. The reaction was stopped by heating the mixture for 10 min at 70°C. DNA was then digested with KpnI and the strands were separated on a 5% polyacrylamide gel [13]. The single-stranded SK233 (see fig.1) DNA fragment was eluted with 0.5 M ammoniumacetate, 10 mM EDTA and 0.1% SDS, precipitated with ethanol and dissolved in re-distilled H_2O. According to Method B subclone 89 DNA was digested with KpnI. The protruding 3'-ends were removed with Klenow polymerase [14]. After controlled treatment with exonuclease III (to remove 40–50 nucleotides) the 3'-ends were filled in by repair synthesis using Klenow polymerase and [α-^{32}P]-dATP [14]. After digestion with Sau3A the single-stranded SK233 fragment was isolated as described above. S_1 nuclease analysis was performed as described previously [15] with some modifications. The hybridization mixture was incubated for 16 h at a temperature gradually decreasing from 50 to 37°C. S_1 digestion was performed for 2 h at 16°C.

3. RESULTS AND DISCUSSION

The recombinant DNA pBMCY89, selected from a bank of HindIII-generated *Saccharomyces carlsbergensis* DNA fragments in pBR322 as described previously [3], carries the gene encoding protein L16 of the large ribosomal subunit. The map of a HindIII–BamHI subcloned fragment is presented in fig.1A. We sequenced the gene according to the strategy outlined in fig.1B and the results of these nucleotide analyses are shown in fig.2.

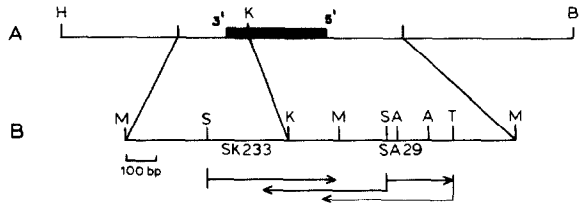

Fig. 1. Map of the insert of subclone 89 and extent of DNA sequencing. The location of the L16 gene and the position of some restriction enzyme sites (indicated in A) were published previously [16]. In B the sequence strategy is shown; the arrows give the direction and extent of sequence analysis. H = HindIII; K = KpnI; B = BamHI; M = MspI; S = Sau3A; A = AluI; T = TaqI.

The coding sequence for yeast ribosomal protein L16 is 522 nucleotides long and gives rise to a protein having a molecular mass of 22 622 Da. The identity of this protein is clear from a comparison of the amino acid composition deduced from the nucleotide sequence with data previously published for protein YL22 [17], which is identical to L16 ([18]; see table 1). As has been observed for other yeast ribosomal protein genes [15], the L16 coding sequence reveals a preferent codon usage which is in agreement with the pattern previously observed for highly expressed yeast genes [19]. The only two rare codons found, viz. CCT for Pro and CGT for Arg, are located in the 5'-part of the gene.

Most ribosomal protein genes studied so far are split near the 5'-end of the coding sequence. The gene for L16, however, is not interrupted by an intron and therefore is one more exception to this rule, the others being the gene coding for the ribosomal proteins L3 [20], S33 [15] and S24 (unpublished). The absence of an intervening sequence from these genes explains why the proteins in question continue to be synthesized in splicing-defective yeast mutants grown at the restrictive temperature [21].

We mapped the transcription initiation site of the L16 gene by the primer extension method. Using the SA29 fragment as the primer (fig.1) we detected two major start sites, at −24 and −29, respectively, as well as a minor start site at −35 (see fig.3). The transcription start site at position

```
-240                -220                -200                -180                -160
GTAATAGACTCACATCCACGTGACCAGTTTCCAATCGAACTTTTTCACTTTGCAGGGAATTATTGTTTCACTAGCAAAGGTAGCCCACTTACCACTCAGC

-140                -120                -100                 -80                 -60
TATGCGGAAAGTTTCATTGTTTGATACATCTTGGTAGTAACCGCAGCTTCTTTCTTAGTTCATATTTTATTGTATTTCAACTAATATTATTTTTTTTTCA

 -40                 -20                  -1 1                   20                          40
GTGCAAGGCAAGGTGAACCAAGAACATACAAACATAGCCAAAG ATG TCT GCC AAA GCT CAA AAC CCT ATG CGT GAT TTG AAG ATC
                                            met ser ala lys ala gln asn pro met arg asp leu lys ile

                     60                          80                         100
GAA AAG TTG GTC TTA AAC ATT TCT GTT GGT CAA TCT GGT GAC AGA TTG ACC AGA GCC TCC AAG GTT TTA GAG CAA
glu lys leu val leu asn ile ser val gly glu ser gly asp arg leu thr arg ala ser lys val leu glu gln

120                      140                         160                         180
TTA TCT GGT CAA ACT CCA GTT CAA TCC AAG GCC AGA TAC ACT GTC AGA ACT TTC GGT ATC AGA AGA AAC GAA AAA
leu ser gly gln thr pro val gln ser lys ala arg tyr thr val arg thr phe gly ile arg arg asn glu lys

       200                         220                         240                         260
ATT GCT GTT CAC GTT ACC GTC AGA GGT CCA AAG GCT GAA GAA ATT TTG GAA AGA GGT TTG AAG GTC AAG GAA TAC
ile ala val his val thr val arg gly pro lys ala glu glu ile leu glu arg gly leu lys val lys glu tyr

      280                         300                         320                         340
CAA TTG AGA GAC AGA AAC TTC TCT GCT ACC GGT AAC TTC GGT TTC GGT ATT GAC GAA CAC ATT GAC TTG GGT ATC
gln leu arg asp arg asn phe ser ala thr gly asn phe gly phe gly ile asp glu his ile asp leu gly ile

                     360                         380                         400
AAG TAT GAC CCA TCC ATC GGT ATT TTC GGT ATG GAT TTC TAT GTC GTC ATG AAC AGA CCA GGT GCT AGA GTC ACT
lys tyr asp pro ser ile gly ile phe gly met asp phe tyr val val met asn arg pro gly ala arg val thr

420                      440                         460                         480
AGA AGA AAG AGA TGT AAG GGT ACC GTT GGT AAC TCC CAC AAG ACA ACT AAG GAA GAC ACC GTC TCT TGG TTC AAG
arg arg lys arg cys lys gly thr val gly asn ser his lys thr thr lys glu asp thr val ser trp phe lys

       500                         520                 540                 560                 580
CAA AAG TAC GAT GCT GAT GTT TTG GAC AAA TAA TTTAATTAGTTTGTGAAGAAATATAATACATTTATATACTCATATCTATGTTTTT
gln lys tyr asp ala asp val leu asp lys end

                600                 620                 640                 660
TTTGTAACCAATTAATAAAATTATGAATTTTAAACCAAGATGAAACTCTATGTTATTGTACATGCGCGCTAAGGACTTCTATGC
```

Fig. 2. DNA sequence of the gene for yeast ribosomal protein L16. The nucleotides are numbered with italics starting from the initiation codon.

−24 is located within the sequence pyAAPu which has been proposed to be a consensus site for initiation of transcription by yeast RNA polymerase II [22]. From the data summarized in table 2 it is obvious that in several cases the capping region can be extended to the sequence AACCAAG found near the 5′-end of the mRNAs. The untranslated 5′-terminal part of the L16 mRNA contains the sequence CATACA, implicated in control of translation efficiency [23] and shows the prefered environment for the initiation codon [27]. Upstream from the 5′-end of the L16 gene the TATA-like structure [28] TAATATTAT is present at position −62. Computer analysis of the 5′-flanking regions of several yeast ribosomal protein genes has revealed some common sequence elements at −150 to −400 from the initiation codon, that may be involved in transcription regulation [29]. No such common boxes are found in the region of the L16 gene extending to position −240.

We also located the 3′-end of the L16 mRNA by S_1 nuclease analysis. As can be concluded from the results presented in fig.4 the 3′-end of the mRNA maps to the region 85–90 nucleotides downstream from the stop codon. The trailer sequence contains several elements that may function in transcription termination [30] (viz. TAGT at +7 and +77 and TATGT at +47), and polyadenylation [31] (viz. AATAAA at +69). The longer exposed lower part of the autoradiogram shown in fig.4 shows some

Table 1

Amino acid composition of yeast ribosomal protein L16

Amino acid	No. of residues	
	Found[a]	Reported[b]
Asp + Asn	18	18.1
Thr	11	10.3
Ser	10	8.4
Glu + Gln	16	16.8
Pro	5	5.8
Gly	15	14.2
Ala	9	9.4
Val	15	15.6
Met	4[c]	2.4[c]
Ile	10	9.9
Leu	11	12.0
Tyr	5	5.3
Phe	7	7.7
His	3	3.1
Lys	17	17.1
Arg	16	15.6

[a] Number deduced from nucleotide sequence
[b] Number determined for yeast ribosomal protein YL22 (= L16; [18]), calculated from an amino acid analysis [17] assuming the total number of residues to be 173
[c] The N-terminal methionine apparently is not present in the mature L16 ribosomal protein

minor signals that can be explained by the presence in yeast cells of two L16 mRNAs transcribed from different copy genes. Indeed, Southern hybridization has revealed the duplicate nature of the gene encoding yeast ribosomal protein L16 [32]. Comparative sequence analysis (J. Woolford, personal communication; this paper) revealed that the trailer sequences of both gene copies have diverged to a large extent, whereas the amino acid coding regions are largely homologous. However, at some positions near the translation stop codon second and third base substitutions have occurred. The position of the minor signals is in good agreement with these sequence differences present in the 3′-part of the coding sequences of both L16 gene copies. At these positions mismatches occur between the L16 DNA sequence and the mRNA transcribed from the other gene copy, giving rise to S_1 nuclease sensitive sites. The difference in intensity between the major and minor signals shows that in vegetatively growing yeast cells L16 mRNA is predominantly derived from the gene copy described in this paper.

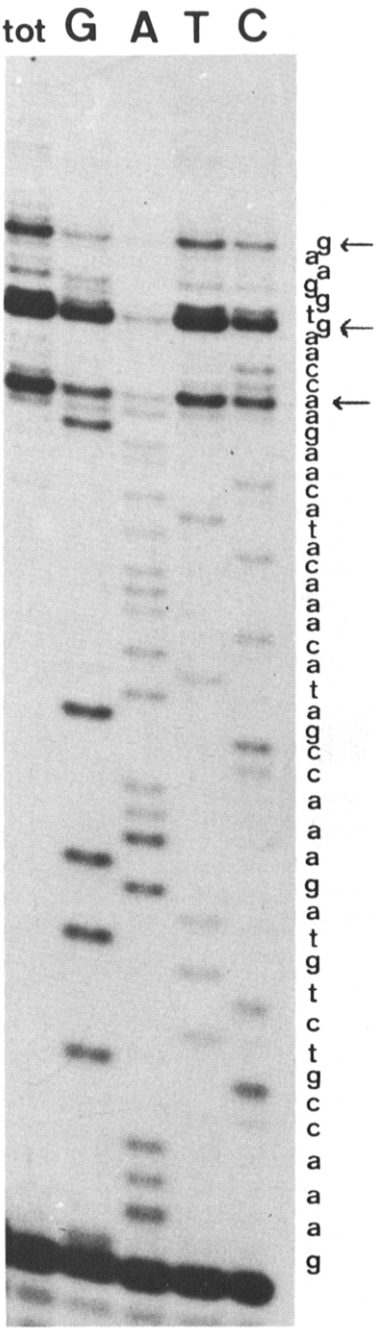

Fig. 3. Sequence analysis of the 5′-end of the L16 mRNA. The positions indicated correspond to the sequence data presented in fig. 2. See section 2. The arrows indicate the mapped 5′-ends of the L16 mRNA.

Table 2

Comparison of leader sequences of several yeast genes

TATTTACAAĠACACCAATC	pyruvate kinase [22]
GATCATCAAGAAGTAATTA	phosphoglycerate kinase [23]
ATATTTCAȦGCTATACCAA	alcoholdehydrogenase I [24]
CTATACCAȦGCATACAATC	
GAATȦTCȦAGCTACAAAAA	alcoholdehydrogenase II [25]
AAAȦȦCCAAGCAACTGCTT	pENO 46 [26]
TȦACACCAAGCAACTAATA	pENO 8 [26]
AAȦAACCAAGAAGTTAGTT	pGAP 63 [26]
AAȦCACCAAGAACTTAGTT	pGAP 491 [26]
CATAAATAȦGAAATTCATC	pGAP 11 [26]
GAAAȦCCAȦGCTAGCAATC	S33 [15]
ĠAACCȦAGAACATACAA	L16 (this paper)

The major transcription start sites are indicated with a dot

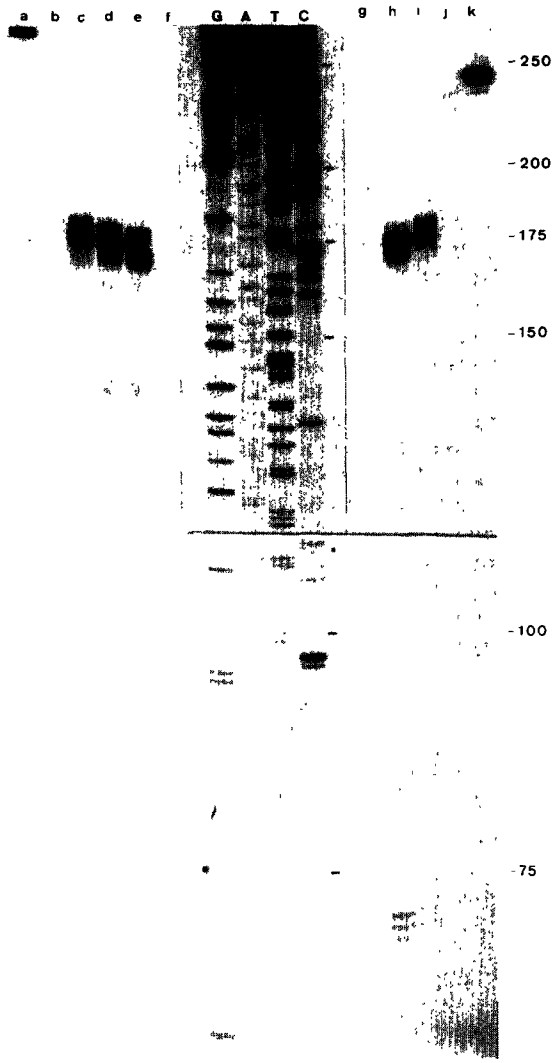

Fig. 4. Mapping of the 3′-end of the L16 mRNA. The results shown in lanes b–f are obtained with the probe labelled according to Method A (see section 2); the untreated fragment is present in lane a. Lanes b and f: *B. Licheniformis* RNA as a control with 25 and 125 nuclease U S_1, respectively; lanes c–e: yeast poly A containing RNA with 25, 75 and 125 U nuclease S_1, respectively. The results shown in lanes g–j are obtained with the probe labelled according to Method B (see section 2); the untreated fragment is present in lane k. Lanes g and j: *B. licheniformis* RNA with 75 and 25 U nuclease S_1, respectively; lanes h and i: yeast poly(A)-containing RNA with 75 and 25 U nuclease S_1, respectively. The lower part of the figure shows a four times longer exposure of this part of the gel (see text). The GATC ladder of a known sequence was used to measure the length of the S^1-protected products.

ACKNOWLEDGEMENTS

This work was supported in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO). The authors are indebted to Dr H.A. Raué for critical reading of the manuscript and to Mrs P.G. Brink for preparing the typescript.

REFERENCES

[1] Planta, R.J. and Mager, W.H. (1982) in: The Cell Nucleus (Bush, H. and Rothblum, L. eds) vol. XII pp. 213–226, Academic Press, New York.

[2] Warner, J.R. (1982) in: The Molecular Biology of the Yeast *Saccharomyces* (Strathern, J.N., Jones, E.W. and Broach, J.R. eds) pp. 529–560, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[3] Bollen, G.H.P.M., Cohen, L.H., Mager, W.H., Klaassen, A.W. and Planta, R.J. (1981) Gene 14, 279–287.

[4] Clewell, D.B. and Helinski, D.R. (1969) Proc. Natl. Acad. Sci. USA 62, 1159–1166.

[5] Bollen, G.H.P.M., Mager, W.H., Jenneskens, L.W. and Planta, R.J. (1980) Eur. J. Biochem. 105, 75–80.

[6] Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387–2398.

[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[8] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.

[9] Messing, J. (1982) in: Genetic Engineering/Principles and Methods (Setlow, J.K. and Hollaender, A. eds) vol. 4, pp. 19–34, Plenum Press, New York.

[10] Sures, I., Levy, S. and Levy, L.H. (1980) Proc. Natl. Acad. Sci. USA 77, 1265–1269.

[11] Levy, S., Sures, I. and Kedes, L.H. (1979) Nature 279, 737–739.

[12] De Haan, M., Van Loon, A.P.G.M., Kreike, J., Vaessen, R.T.M.J. and Grivell, L.A. (1984) Eur. J. Biochem. 138, 169–177.

[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[14] Guo, L.-H. and Wu, R. (1983) Methods Enzymol. 100, 60–96.

[15] Leer, R.J., Van Raamsdonk-Duin, M.M.C., Schoppink, P.J., Cornelissen, M.T.E., Cohen, L.H., Mager, W.H. and Planta, R.J. (1983) Nucleic Acids Res. 11, 7759–7768.

[16] Bollen, G.H.P.M., Molenaar, C.M.T., Cohen, L.H., Van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.J. (1982) Gene 18, 29–37.

[17] Otaka, E., Higo, K.J. and Osawa, S. (1982) Biochemistry 21, 4545–4550.

[18] Bollen, G.H.P.M., Mager, W.H. and Planta, R.J. (1981) Mol. Biol. Rep. 8, 37–44.

[19] Ammerer, G., Hitzeman, R., Hagie, F., Barta, A. and Hall, B.D. (1981) in: Recombinant DNA (Walton, A.G. ed) pp. 185–197, Elsevier Scientific Co., Amsterdam.

[20] Schultz, L. and Friesen, J.D. (1983) J. Bacteriol. 155, 8–14.

[21] Warner, J.R. and Gorenstein, C.G. (1977) Cell 11, 201–212.

[22] Burke, R.L., Tekamp-Olson, P. and Najarian, R. (1983) J. Biol. Chem. 258, 2193–2201.

[23] Dobson, M.J., Tuite, M.E., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Forthergrill, L.A. (1982) Nucleic Acids Res. 10, 2625–2637.

[24] Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018–3025.

[25] Russell, D.W., Smith, M., Williamson, V.M. and Young, E.T. (1983) J. Biol. Chem. 258, 2674–2682.

[26] Holland, J.P., Labieniec, L., Swimmer, C. and Holland, M.J. (1983) J. Biol. Chem. 258, 5291–5299.

[27] Kozak, M. (1981) Nucleic Acids Res. 9, 5233–5252.

[28] Benoist, C. and Chambon, P. (1981) Nature 290, 304–310.

[29] Teem, J.L., Abovich, N., Käufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., Van Raamsdonk-Duin, M.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D. and Rosbash, M. (1981) Nucleic Acids Res., in press.

[30] Zaret, K.S. and Sherman, F. (1982) Cell 28, 563–573.

[31] Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251–260.

[32] Molenaar, C.M.T. (1984) Ph.D. Thesis, Free University, Amsterdam