Structural comparison of yeast ribosomal protein genes

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

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

ABSTRACT

The primary structure of the genes encoding the yeast ribosomal proteins L17a
and L25 was determined, as well as the positions of the 5'- and 3'-termini of
the corresponding mRNAs. Comparison of the gene sequences to those obtained
for various other yeast ribosomal protein genes revealed several similarities.
In all split genes the intron is located near the 5'-side of the amino acid
coding region. Among the introns a clear pattern of sequence conservation can
be observed. In particular the intron-exon boundaries and a region close to
the 3'-splice site show sequence homology. Conserved sequences were also found
in the leader and trailer regions of the ribosomal protein mRNAs. The
5'-flanking regions of the yeast ribosomal protein genes appeared to contain
sequence elements that many but not all ribosomal protein genes have in
common, and therefore may be implicated in the coordinate expression of these
genes. The amino acid coding sequences of the ribosomal protein genes show a
biased codon usage. Like most yeast ribosomal protein molecules, L17a and L25
are particularly basic at their N-terminus.

INTRODUCTION

The biogenesis of ribosomes is a highly complex process requiring the

coordinate and balanced synthesis of a large number of constituents. In

E. coli the details of the regulation mechanisms underlying this coordinate

synthesis are beginning to emerge. As far as synthesis of ribosomal proteins

is concerned, it has become clear that the organization of ribosomal protein

genes into a limited number of transcriptional and translational units is an

important feature of the regulation mechanism (see Ref. 1 for a recent review).

Regulation of ribosomal protein synthesis in eukaryotic cells so far is by

and large terra incognita. The best-studied organism in this field is yeast,

where about 30 different ribosomal protein genes have so far been cloned by

ourselves as well as others (2-8). In our group we have isolated the genes

for proteins S10, S16A, S24, S31, S33, rp28, L16, L17a, L25, L34, L36 and L46

from a colony bank of HindIII-generated yeast DNA fragments cloned in pBR322

(8). Genomic blotting experiments showed that almost all of these genes are

duplicated, the gene for L25 being the only exception (9). Moreover, both our

own studies and experiments by others have revealed that yeast ribosomal protein genes in general are not clustered (3,8,10). The only exceptions to this rule sofar are the genes encoding proteins S16A and rp28, which are separated by only about 600 bp (Molenaar, C.M.T. and Pearson, N.J., manuscript in preparation) and those encoding S24 and L46 which are about 700 bp apart (R.J. Leer, manuscript in preparation).

Ribosomal protein genes in yeast are somewhat of a class apart since most have been found to contain an intron (3,8,9,11). So far only four ribosomal protein genes [L3 (6), L16 (24), S33 (12) and S24 (R.J. Leer, manuscript in preparation)] do not conform to this rule. In other yeast genes the presence of an intron is the exception rather than the rule. Out of a large number of genes studied as yet only two [the actine gene (13,14) and the MAT1α-gene (15)] are split.

The split nature of the ribosomal protein genes explains the coordinate depression of ribosomal protein synthesis in splicing-defective ts mutants of yeast upon raising the temperature (16,17). We (9) and others (3,11) have shown that under those conditions pre-mRNAs for several ribosomal protein species accumulate that are 300-500 nucleotides longer than their mature counterparts. The ribosomal proteins known to escape this coordinate regulation are precisely the ones mentioned above as containing no intron in their gene. Clearly processing of mRNA is an important regulatory mechanism under restrictive conditions in these ts-mutants.

In wild-type yeast cells coordinate control of ribosomal protein synthesis may take place at both the transcriptional and translational level. For instance, after a temperature shock a sudden coordinate decrease in ribosomal protein gene transcription occurs (11,18). On the other hand the outcome of gene dosage experiments using yeast ribosomal protein genes cloned in multi-copy vectors, indicate that control is also exerted at the translational level (19; J.R. Warner, personal communication).

In order to classify further the molecular mechanisms regulating the coordinate and balanced synthesis of ribosomal proteins in yeast we are carrying out a detailed structural analysis of a number of individual yeast ribosomal protein genes. The aim of these studies is the identification of common sequences that might act as signals in regulatory events. Deletion mapping and site-directed mutagenesis can then be used to ascertain the function of these sequences. In view of the evidence for both transcriptional and translational regulation we have focussed our attention on the untranslated mRNA sequences as well as the DNA sequences lying upstream of the tran-

scription initiation site of the ribosomal protein genes. In this paper we present such an analysis of the genes for the large subunit proteins L17a and L25 and compare the results with previously published data on several yeast ribosomal protein genes.

MATERIALS AND METHODS

DNA preparation and sequence analysis

Plasmids pBMCY111 and pBMCY138 containing the genes coding for L17a and L25 respectively, were purified as described previously (12). Restriction enzyme digestions were performed as recommended by the suppliers (New England Laboratories; Boehringer Mannheim). The cutting sites for the enzymes TaqI, MspI and HaeIII were mapped using the partial digestion procedure of Smith and Birnstiel (20). DNA fragments were subcloned in M13 mp7, mp8 or mp9 and their sequence was determined by the dideoxy chain termination method (21,22). Sequence analysis according to the chemical procedure was performed as described in (23).

### Primer extended sequence analysis of mRNA

Details of primer preparation and RNA sequencing using reverse transcriptase are described elsewhere (24). The primer for the L17a gene was labelled by repair synthesis of the HaeIII-BglII fragment in M13 mp8 (24) (see Fig. 1). After digestion with RsaI a fragment of 107 nucleotides was isolated (24 nucleotides of which originate from the M13 cloning region) (24).The primer for the L25 gene was a 75 nucleotides long DNA fragment obtained by SacI + RsaI digestion of cDNA synthesized using the subcloned HpaI-TaqI fragment in M13 mp9 as a template (24)(see Fig. 2).

S1 nuclease mapping of the 3'-ends of mRNAs

Preparation of the probes and S1 nuclease analysis were carried out as described in detail elsewhere (24). The probe for the transcript of the L17a gene was a 316 nucleotides long DNA fragment obtained by HindIII digestion of cDNA synthesized using the subcloned Taq-Taq fragment in M13 mp9 as a template (24) (see Fig. 1; 47 nucleotides originate from M13).
For mapping the 3'-end of the L25 gene transcript pBMCY138 DNA was digested with HpaI. After limited treatment with exonuclease III and repairing the 3'-end with Klenow polymerase in the presence of α-^{32}P-dATP, the 264 nucleotides long probe was isolated as described elsewhere (24).

RESULTS AND DISCUSSION

From the colony bank of HindIII-generated yeast DNA fragments in pBR322 that has been described in a previous paper (4), a number of ribosomal protein
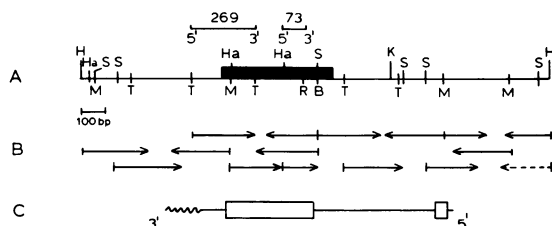
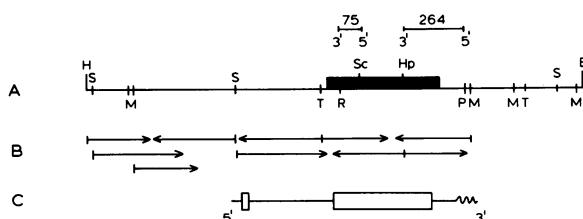Fig. 1. Map of the insert of pBMCY111 and structural analysis of the L17a gene.
The position of some restriction enzyme sites as well as the location of the R-loop indicated in A were published previously (8). H = HindIII, S = Sau3A, M = MspI, T = TaqI, R = RsaI (only RsaI sites are indicated that are used for constructing the primers), Ha = HaeIII, B = BglII, K = KpnI.
In B the sequence strategy is shown. The arrows give the extent of nucleotide analysis according to the chain termination method; the dashed arrow indicates a sequence determined using the chemical procedure. The 269 nucleotides long probe used for the 3'-mapping experiment and the 73 nucleotides long primer used for the primer extension experiment are indicated. The structure of the L17a transcript is given in C.

genes has been isolated (8). Recombinant DNA pBMCY111 carries a gene coding for ribosomal protein L17a (8; R.J. Leer, unpublished results) whereas clone pBMCY138 contains the gene encoding the large subunit protein L25 (8). Southern analysis using genomic digests of yeast DNA revealed that L25 is encoded by a simple locus (data not shown; see Ref. 9) whereas the gene encoding L17a is duplicated on the yeast genome (9).

In Figures 1 and 2 the physical maps of the respective cloned DNA fragments are depicted. The genes were located in the maps by electron microscopic R-loop analysis (8). The strategy followed in sequencing the respective genes is also outlined.

The results of the nucleotide sequence analysis are shown in Figures 3 and 4. The open reading frame of the L17a gene encompasses 411 nucleotides, coding for a protein with a molecular mass of 14,321 D and a net charge of +13. The gene for L25 consists of 414 nucleotides encoding a protein with a molecular mass of 15,225 D and a net charge of +19. The identity of the protein encoded by the L25 gene is confirmed by the amino acid composition deduced from the nucleotide sequence. This composition is in good agreement with corrected data obtained from studies on protein YP42' = YL25 (25; E. Otaka, personal communication) which is identical to L25 (26). No such data so far exist for protein L17a, which is identical to YL32 (26). It is noteworthy that both L25 and L17a are particularly basic at their N-terminus. Neither protein contains acidic residues within the first 55 amino acids. A similar

Fig. 2. Map of the insert of pBMCY138 and structural analysis of L25 gene. The position of some restriction enzyme sites as well as the location of the R-loop indicated in A were published previously (8). For restriction enzyme sites the same symbols are used as in Fig. 1. In addition Sc = SacI, Hp = HpaI, E = EcoRI. In B the sequence strategy is shown. The 264 nucleotides long probe for the 3'-mapping experiment as well as the 75 nucleotides long primer for cDNA synthesis are indicated. The sequence spanning over the Sau3A site at the 5' end of the gene was confirmed by primer extension analysis. In C the structure of the L25 transcript is given.

skewed distribution of charged amino acids is observed in several other, though not all yeast ribosomal proteins. Furthermore L25, which probably is one of the primary assembling ribosomal proteins (27) is characterized by the presence of 5 Lys-Lys doublet sequences, 4 of which are part of a Lys-Lys-Ala tripeptide.

The aminoacid coding sequence of the genes for L17a and L25 reveals a biased codon usage, in agreement with findings for other yeast ribosomal protein genes (see Table I). The pattern of prefered codons in ribosomal protein genes is quite similar to the one observed for highly expressed yeast genes (28) and may reflect an efficient translation of the moderately abundant ribosomal protein mRNAs.

As can be concluded from the sequence data presented in Figures 3 and 4 the open reading frames of both the L17a and the L25 gene are interrupted by an intron of 513 and 415 nucleotides respectively. In this respect the genes for L17a and L25 are similar to most other ribosomal protein genes in yeast. A remarkable feature of all split yeast ribosomal protein genes studied so far is the location of the intervening sequence close to the 5'-end of the gene (cf. Table II). Recent findings (9,35) indicate that in some ribosomal protein genes the intron is present even upstream of the ATG codon. Since a split character is rarely found for non-ribosomal protein genes on the yeast nuclear genome it is tempting to suggest a regulatory function for these introns. Whether the interruption of many yeast ribosomal protein genes at their very 5'-end is implicated in transcriptional or post-transcriptional control, however, is still open to speculation.

```
          -420              -400              -380              -360              -340
AAGCTTGCAATACCAATCACTATTACCATGTTAATGATCCCCCAGATAAAAAGACAAGACAGAGCGTATATCGACACTACTTTAACATCTGTACACCCTCA

          -320              -300              -280              -260              -240
ATTGCTTTGAAAAAACCCATGCAGTTATTATATTTTTCAACGGACCAATTTTTTTTTATCGCTTATCCGGCTCACAGAACGAGCCGTACAACCGTCATCAG

          -220              -200              -180              -160              -140
ACTTGCATGGAAATGTTGAAAAAGAGGTAAAAAAGAAAATATCAAACCCAAACACACTCGGGCTGAAAAATTAATATGGCAGGAAACGAACTGTAGCCAT

          -120              -100              -80               -6C               -40
AATGTACCTGTTCCCGCTAATAGTAATACCATCGTATTGCTAATGTTAGTAGCTAAAAATCTCTTACTTTCTTTATTTTGATTCTTGGGTCTTACAAGCA

        ↓   ↓-2Q                    -1 1          20                    40
ATACAAAACCAACACACCTATATATAACTAATA ATG TCC GGT AAC GGT GCT CAA GGT ACT AAG TTT AGA ATC TCA GTATGTTAAA
                                   met ser gly asn gly ala gln gly thr lys phe arg ile ser

         60                80               100               120               140
ATTTTTATTTTCCACAATGCAATTTGGTTAAATTGATCATAAAGTAAAGTTCCAAGATTTCATTTTGCTGGGTACAACAGAATTAAACAGAGGTTTAAAA

        160               180               200               220               240
GACGCTAAGACGCTAAACACTAAGTGTGAAGCCTGATCATATACTTATCGATGTTATTTTCCATAAGAATGGTAGATGTGCCATCACTGGTACCGCCCCG

        260               280               300               320               340
CCAAAGATGCAATTACATACATAAACTGAGTTTTTAAATGGTTAAGCGGACTGTATTTGCAGATGAAGTGATACCTGTCAATATTCATGGACTCAAAAAT

        360               380               400               420               440
TTCTGGTTTCAAAATCATACACGGCAAGCATATCCTAACTCCCTCTTTTGCCCATACTAATGGATAACCATCGAAATGTTATTTCTACAAAAATCACACT

        460               480               500               520               540
TAACTCTATTTTTCAAGATTAACCACATCTGCTAACTTTCTCCCTATGCTTTTACTAACAAAATTATTCTCACTCCCCGATATTGAGATCTTTTTTAACT

          560               580               600               620
AG TTA GGT CTA CCA GTC GGT GCC ATC ATG AAC TGT GCT GAC AAC AGT GGT GCC AGA AAC TTG TAC ATT ATC GCC
   leu gly leu pro val gly ala ile met asn cys ala asp asn ser gly ala arg asn leu tyr ile ile ala

          640               660               680               700
GTC AAA GGC TCT GGT TCC AGA TTG AAC AGA TTG CCA GCC GCC TCT CTA GGT GAT ATG GTT ATG GCC ACC GTT AAG
val lys gly ser gly ser arg leu asn arg leu pro ala ala ser leu gly asp met val met ala thr val lys

          720               740               760
AAG GGT AAG CCA GAA TTG AGA AAG AAG GTT ATG CCA GCT ATT GTT GTC CGT CAA GCT AAG TCT TGG AGA AGA AGA
lys gly lys pro glu leu arg lys lys val met pro ala ile val val arg gln ala lys ser trp arg arg arg

780               800               820               840
GAC GGT GTC TTT TTG TAC TTC GAA GAC AAT GCT GGT GTC ATC GCT AAT CCT AAG GGT GAA ATG AAG GGT TCC GCC
asp gly val phe leu tyr phe glu asp asn ala gly val ile ala asn pro lys gly glu met lys gly ser ala

        860               880               900               920
ATC ACT GGT CCA GTC GGT AAG GAA TGT GCC GAT TTA TGG CCA AGA GTT GCA TCT AAC TCC GGT GTT GTT GTG TAA
ile thr gly pro val gly lys glu cys ala asp leu trp pro arg val ala ser asn ser gly val val val end

          940               960               980               1000              1020
AAAGGAAATTATTTTGACTCATACTCTTAGATTCTTCAATTATATATATTTTCTTTATAAATAACAATCTCTTAAACAGATTAAAGAAACTCAATAATTC

  .   1040              1060              1080              1100              1120
ATTGCATTTTTAGATGAAAATATTACTATTATCGTTCATCGAGAGAATCACATGATATTGAGCGTGATAGGAGGGTCTATATATTTTTTGGGGTACGT

        1140              1160              1180              1200              1220
CAAAAAAGCAAACAGACACGTTAAGAAGGTAAGAATAACAGATGATAGAATGTTGTGCTAGTGAGAAACAAGTATGCACCAGCCGTACAATATGCGTATT

        1240              1260              1280              1300              1320
TTGACATTTTTCAAGGTAAATTAGTTCCAAGTTAATAGAATAAATCAAGACTAATTGACGTCTCAAAGCCTCGCAAAAAAAGCAGGAAATTCGAAAAAAA

        1340              1360              1380              1400              1420
GCCTTCAAAGAAAAGGGAAATCAGTGTAACATAGACGATGGAGGATCATGGGATTGTAGAAACTTTAAACTTTCTATCATCAACAAAAATCAAAGAGAGA

        1440              1460              1480              1500
AACAATGCTTTAGATGAGCTAACAACAATTTTAAAAGAAGATCCGGAAAGGATACCAACCAAGGCCCTATCTACAACGGCAGAAGCTT
```

Fig. 3. DNA sequence of the gene for yeast ribosomal protein L17a and amino acid sequence of the L17a protein.
The nucleotides are numbered with italics starting from the first nucleotide of the initiation codon. The positions of the 5'- and 3'-ends of the L17a mRNA are indicated by arrows.

```
            -720            -700            -680            -660            -640
AAGCTTGCTATTTGACAGCAAAAGATCTTTACGTAAACAGATAGGGTCACACAGACTATACGAAAACCTGTCAATATAATCTATTCGGATAAGATTATCA

            -620            -600            -580            -560            -540
ACCCCTCCCTGCAAGATTTTTTATTGTTATTATTACAGTTTTTAAACTGGTAGCAGAGCTATCCGGTAAACATGTACTAATCAGTCTAACACCCGGAGGA

            -520            -500            -480            -460            -440
GGCTTATCATTACATATTCTTGTCTCGGCGTTCTTTAAGTTTCTTATCTTGTATGCCCGATATAGCAACCTTGTTGGTACCAATCTAACGGTTTCCGTAT

            -420            -400            -380            -360            -340
TTTGCAATGAAGAGATGAGGAGGTATGGGTCACTGATTTAATATGTACGGGTGTTTACATGGAGTTGCTCTCTTTTTTGCCCTAGCAGTCATTGTGCGCC

            -320            -300            -280            -260            -240
GAAAAAAGAGAAAACCGTGAGCCGAAGTCCACGCTCTGGAGATTAGGCTCTCCCATTACGGAGAGAAGCATTTCCTCAGCCTGGAGCCCGTTGGAACAGT

            -220            -200            -180            -160            -140
CAGGCTAAACTGGCCTTCCTACCCACTGCTTGCTGTTTCTCACTGGACGCACAAGGGGATTTTCTTTCTACTTCGGCTTGCCTCACTGCGTTGGGGCTTC

            -120            -100            -80             -60             -40
CCAATGCAACTTCGTTCGTATGCATACAATCTTTTAGATATTATCTTTTAAAATTATTTTAAAACAATTTTAAATGTATCTCATATGCTTTTCTTCTGCT

         ↓  -20        ↓    ↓   -1 1           20              40
GTTGAAAAGGCTAAACAAAGAAGATCAATAAGATAAA ATG GCT CCA TCT G GTATGTGAACTGCAATATTAATAGCACGAGAAAATTGAGAGGAA
                                       met ala pro ser

   60               80              100             120             140
GATAAATGGGAACTAGTACAGTTGATATTGATGAGATACGAAAATCGCACCGTAAATAAACTATCCGACGACAAGAATAGTGCATTAAGGACTTGATTAA

   160              180             200             220             240
GATATGGTGAGCCAACGTAATTATCGGGCTCAACTTTTTATTAGCAATCGTTTTGATAGAAGCGTTGATGCTGGAAGTTTCCTTTTTTACTACCAACAGC

   260              280             300             320             340
CATTAACAAATCCATTAGAGTGTTCGTTCGTTTTTCCGCTCAGATTCAAGAAAAATTATTTCATCCTCATTATACTTTTTTTCTTCTATTTGGTGCTCCA

   360              380             400             420             440
CGTCGAGGTATCAAGGAACATAGTTTACTAACATTAACAAATTCATCTCCTATGAATTTACTTTTTGTATAG CT AAG GCT ACT GCC GCT AAG
                                                                          ala lys ala thr ala ala lys

      460             480             500             520
AAA GCT GTC GTT AAG GGT ACT AAT GGT AAG AAG GCT TTG AAG GTC AGA ACT TCT GCT ACC TTC AGA CTA CCA AAG
lys ala val val lys gly thr asn gly lys lys ala leu lys val arg thr ser ala thr phe arg leu pro lys

      540             560             580
ACC TTG AAG TTG GCT AGA GCT CCA AAA TAT GCT TCC AAG GCT GTT CCA CAC TAC AAC AGA TTG GAC TCA TAC AAG
thr leu lys leu ala arg ala pro lys tyr ala ser lys ala val pro his tyr asn arg leu asp ser tyr lys

   600             620             640             660
GTC ATT GAG CAA CCA ATC ACT TCT GAA ACC GCT ATG AAG AAG GTT GAA GAT GGT AAC ATT TTG GTT TTC CAA GTT
val ile glu gln pro ile thr ser glu thr ala met lys lys val glu asp gly asn ile leu val phe gln val

   680             700             720             740
TCC ATG AAA GCT AAC AAA TAC CAA ATC AAG AAG GCC GTC AAG GAA TTA TAC GAA GTT AAC ATT TTG GTC AGA CCA
ser met lys ala asn lys tyr gln ile lys lys ala val lys glu leu tyr glu val asn ile leu val arg pro

      760             780             800             820
AAC GGT ACC AAG AAG GCT TAC GTT AGA TTG ACT GCT GAC TAC GAT GCT TTG GAC ATT GCT AAC AGA ATC GGT TAC
asn gly thr lys lys ala tyr val arg leu thr ala asp tyr asp ala leu asp ile ala asn arg ile gly tyr

         840             860             880             900         ↓.      920
ATT TAA TCTAATTGGTTTAATTAATAATTTAATATTATTTTTAAATTTTTCTTTAAATATACAATAAATCTTCATTAACATGTTAAATTCATGATTAA
ile end

         940             960             980             1000            1020
GGTAAATAAAGTGTAGTGGCAAAGTGCACGGGGTTTCCTGTGACTTCCAAAGTAGGTACCAATTTGCGTAGCTGCAGCGATGGTTCCGGTTACTATTATA

         1040            1060            1080            1100            1120
ATAACGTGTTAGTGTACTGCGATTGTATTGAAGCTATAACAAGAAAAGGATTTGTGAAGGTTTTAGAGAATGAGGCCGGTTTTCACAGGGTTATTTTTATC
```

Fig. 4. DNA sequence of the gene for yeast ribosomal protein L25 and amino acid sequence of the L25 protein.
The nucleotides are numbered with italics starting from the first nucleotide of the initiation codon. The positions of the 5'- and 3'-ends of the L25 mRNA are indicated by arrows.

Table I. Codon usage for the genes coding for yeast ribosomal proteins L17a and L25 as compared with other ribosomal protein genes (RP) and several highly expressed yeast genes (Y).

	L17a	L25	RP[a]	Y[b]		L17a	L25	RP[a]	Y[b]
Phe UUU	2	–	2	5	Tyr UAU	–	1	5	2
Phe UUC	1	2	52	63	Tyr UAC	2	7	38	73
Leu UUA	2	1	13	14					
Leu UUG	5	8	94	148					
Leu CUU	–	–	1	–	His CAU	–	–	12	3
Leu CUC	–	–	–	1	His CAC	–	1	30	56
Leu CUA	2	1	8	6	Gln CAA	2	3	50	44
Leu CUG	–	–	–	–	Gln CAG	–	–	1	–
Ile AUU	2	5	39	58	Asn AAU	2	1	3	3
Ile AUC	5	3	42	66	Asn AAC	6	6	46	86
Ile AUA	–	–	–	1	Lys AAA	1	4	26	21
Met AUG	6	3	28	39	Lys AAG	10	17	141	151
Val GUU	7	7	73	112	Asp GAU	2	2	33	37
Val GUC	6	5	66	93	Asp GAC	3	3	40	100
Val GUA	–	–	–	3	Glu GAA	4	4	89	67
Val GUG	1	–	2	1	Glu GAG	–	1	1	3
Ser UCU	4	3	46	79	Cys UGU	2	–	6	20
Ser UCC	4	2	29	73	Cys UGC	–	–	1	2
Ser UCA	1	1	3	1					
Ser UCG	–	–	–	1	Trp UGG	2	–	11	26
Pro CCU	1	–	5	6	Arg CGU	1	–	25	2
Pro CCC	–	–	1	1	Arg CGC	–	–	–	–
Pro CCA	6	6	52	71	Arg CGA	–	–	–	–
Pro CCG	–	–	–	–	Arg CGC	–	–	–	–
Thr ACU	2	5	46	49	Ser AGU	1	–	1	1
Thr ACC	1	4	38	66	Ser AGC	–	–	4	1
Thr ACA	–	–	2	–	Arg AGA	9	7	135	63
Thr ACG	–	–	–	–	Arg AGG	–	–	1	1
Ala GCU	6	17	90	182	Gly GGU	13	5	120	164
Ala GCC	8	2	28	62	Gly GGC	1	–	4	11
Ala GCA	1	–	4	3	Gly GGA	–	–	2	1
Ala GCG	–	–	2	4	Gly GGG	–	–	1	1

a) Ribosomal protein genes encoding S10 (29), S33 (12), rp51 (30), rp28 and S16A (Molenaar, C.M.T. and Pearson, N.J., manuscript in preparation), L16 (24), L34 (31), L3 (6) and L29 (32).

b) From Ref. 33.

Table II. General structure of yeast ribosomal protein genes.

		5' exon			intron (a)		3' exon	
		start	no. of codons		no. of nucleotides		no. of codons	stop
S10	(b)	ATG -	1	←	352 (394)	→	234	- TAA
L34	(c)	ATG -	18	←	349 (421)	→	94	- TAA
S16A	(d)	ATG -	5	←	390 (551)	→	138	- TAA
rp28	(d)	ATG -	36	←	429 (447)	→	149	- TAA
L46	(e)	ATG -	1	←	383	→	49	- TAA
rp51	(f)	ATG -	-	←	325 (398)	→	135	- TAA
L29	(g)	ATG -	15	←	510	→	133	- TAA
L17a		ATG -	13	←	513	→	123	- TAA
L25		ATG -	3	←	415	→	134	- TAA

(a) data for the second gene copy are within brackets;
(b) (29) and R.J. Leer, unpublished results;
(c) (31);
(d) Molenaar,C.M.T. and Pearson,N.J., manuscript in preparation;
(e) R.J. Leer, unpublished results;
(f) (30,34);
(g) (32)

The introns occurring in yeast ribosomal protein genes show several sequence similarities as outlined in Table III. The structures of the 5'- and 3'-splice sites without exception are in perfect agreement with the consensus boundary

Table III. Intron sequences within yeast ribosomal protein genes (a).

```
S10-1      ↓GTATGT - 316 n. - TTTACTAACA - 10 n. - TTTATAACAG↓
S10-2       GTATGA - 343 n. - TATACTAACA - 25 n. - TTTAAAACAG
L34-1       GTATGT - 378 n. - GTTACTAACA - 17 n. - TTTTTAATAG
L34-2       GTATGT - 305 n. - TTTACTAACA - 18 n. - ATTAATATAG
S16A-1      GTACGT - 341 n. - TTTACTAACA - 23 n. - TTTTCTACAG
S16A-2      GTACGT - 503 n. - TTTACTAACA - 22 n. - TTTCAATTAG
rp28-1      GTATGT - 409 n. - GTTACTAACA - 12 n. - TTTTTTTTAG
rp28-2      GTATGT - 375 n. - TTTACTAACA - 28 n. - TTAATCACAG
L46         GTATGT - 314 n. - GTTACTAACA - 43 n. - TTTTAATTAG
rp51-1      GTATGT - 326 n. - TATACTAACA - 46 n. - ATTTTAATAG
rp51-2      GTACGT - 282 n. - TTTACTAACT - 23 n. - ATTTTTATAG
L29         GTATGT - 453 n. - TTTACTAACG - 31 n. - TTTTGTACAG

L17a        GTATGT - 455 n. - TTTACTAACA - 32 n. - TTTTAACTAG
L25         GTATGT - 360 n. - TTTACTAACA - 29 n. - TTTTGTATAG
                                                      T T    aaT  ↓
consensus ↓GTA T Gt -   x   -  TT TACTAACa -   y   - T T tN    AG
              C                GA                    A A   ttC
     (b)
```

(a) For references see Table II
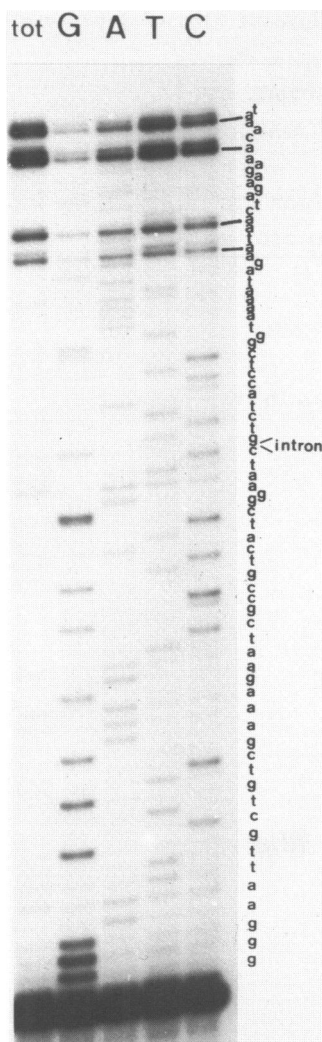(b) a and t are used at positions with a preference for A and T
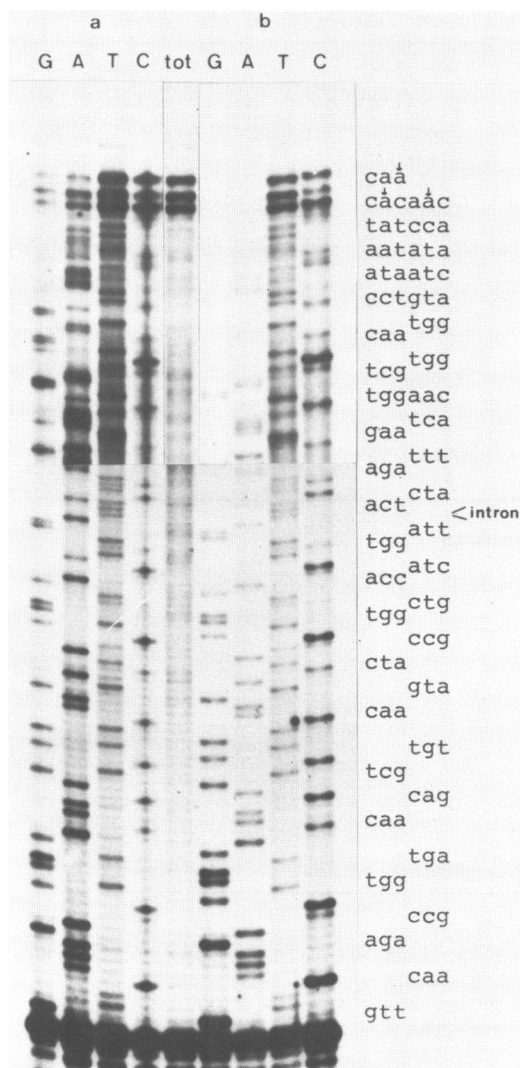
Fig. 5. Sequence analysis of the 5'-end of the L17a mRNA. cDNA synthesis was performed as described in Materials and Methods using the 73 nucleotides long primer depicted in Fig. 1. The arrows indicate the 5'-termini of the L17a mRNA; 'intron' indicates the splice junction.
In b the concentration of ddNTP's was doubled.

signals found in higher eukaryotes (36,37), though in yeast the variability of the bases at the splice sites is even less. In addition to the sequences at the 5'- and 3'-termini of the introns another region of striking sequence homology occurs near the 3'-end. This region contains the TACTAAC-box which has been demonstrated to be essential for splicing in yeast (34,38,39). Apart from these primary structure elements no other homologous sequences nor common secondary structures can be observed in the introns of yeast ribosomal protein genes.

Fig. 6. Sequence analysis of the 5'-end of the L25 mRNA.
cDNA synthesis was performed as described in Materials and Methods using the 75 nucleotides long primer depicted in Fig. 2. The arrows indicate the 5'-termini of the L25 mRNA; 'intron' indicates the splice junction.

The sites at which transcription of the genes encoding L17a and L25 is initiated were determined by primer extended sequence analysis of the respective mRNAs. As can be seen in Fig. 5 with L17a-mRNA two 5'-ends are detected which map at positions -20, -23 and -27. The same type of analysis of the L25 mRNA revealed 4 major 5'-ends located at positions -8, -11, -21 and -25 (see Fig. 6). Multiple capping sites have previously been observed for genes the expression of which is highly regulated (cf. Ref. 40) and therefore may be a feature of the mechanism controlling expression. The RNA sequencing results shown in Figures 5 and 6 also provided definite proof for

Table IV. Transcription initiation sites of several yeast ribosomal protein genes.

```
                        ↓
S10 (a)        ATCCAGTGTGAATAGACGACTGAGCCATC ATG
               ↓     ↓
L16 (b)        GAACCAAGAACATACAAACATAGCCAAAG ATG
                 ↓↓
S33 (c)      GAAAACCAAGCTAGCAATC ATG
              ↓     ↓     ↓
L17a       CAAAACCAACACACCTATATATAACTAATA ATG
            ↓     ↓        ↓   ↓
L25        TAAACAAAGAAGATCAATAAGATAAA ATG
```
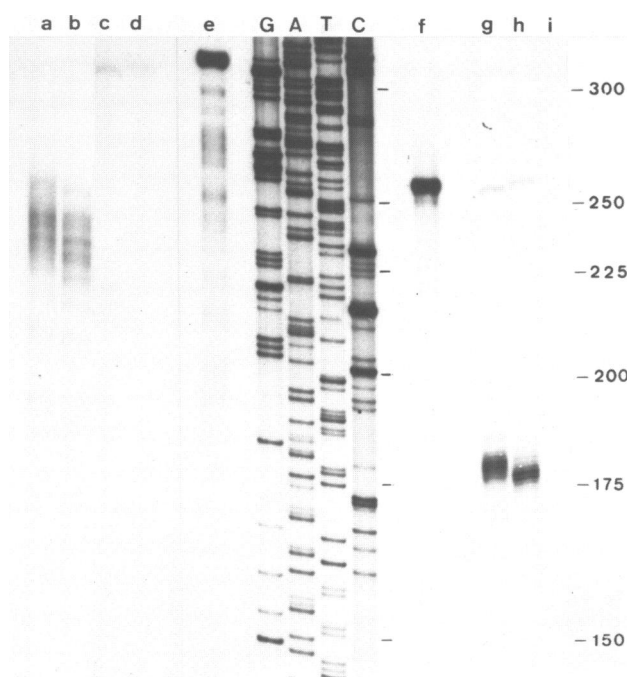
(a) R.J. Leer, unpublished results
(b) (24)
(c) (12)

Fig. 7. Mapping of the 3'-ends of the L17a and L25 mRNAs.
Lanes a-e show the product analysis using the 316 nucleotides long probe for the L17a gene (see Materials and Methods and Fig. 1). Lanes f-i show the product analysis using the 264 nucleotides long probe for the L25 gene (see Fig. 2). Lane a and b: 30 µg polyA containing yeast mRNA with 25 and 100 U S1 nuclease respectively; lanes c and d: 30 µg B. licheniformis RNA as a control with 25 and 100 U S1 nuclease respectively; lanes e and f: untreated probes; lanes g and h: 30 µg polyA-containing yeast mRNA with 25 and 100 U S1 nuclease respectively; lane i: 30 µg B. licheniformis RNA as a control with 25 U S1 nuclease.
The GATC-ladder of a known sequence was used to measure the length of the S1-protected products.

the location of the intron boundaries in both ribosomal protein genes.
The 5'-untranslated parts of the L17a and L25 mRNAs are AT-rich and contain
no or only a few G-residues respectively. This feature of the leader region
holds for most yeast ribosomal protein mRNAs. No obvious secondary structure
elements such as may be involved in an autogenous control of translation in
E. coli (41) could be detected. The transcription start sites of the L17a and
L25 genes map in a region near the sequence PyAAPu, which has been proposed
as a common capping site for RNA polymerase B transcripts (42). The data
summarized in Table IV illustrate that AACCAA forms an additional conserved
element in some but not all ribosomal protein genes. Finally the L17a and L25
mRNAs have an A at position -3, consistent with the prefered surroundings of
the initiation codon observed in other eukaryotic mRNAs (43) and moreover use
the 5'-proximal AUG as the start codon thus fitting into the ribosome-
scanning model for translation initiation (43).

The 3'-ends of the L17a and L25 mRNAs were mapped by S1 nuclease analysis. The
results presented in Fig. 7 show the mature transcript to terminate at
103 ± 10 nucleotides from the stop codon for the L17a mRNA and at approx. 80
nucleotides from the stopcodon for the L25 mRNA. The observed heterogeneity
of the 3'-ends to our opinion reflects the actual situation; it may be the
result of processing events, and is not due to the experimental conditions
used in the mapping experiments. The trailer regions of the two genes contain
a striking sequence homology: a tripartite common element ATTAT---T_4CT_4---
ACAAT can be recognized. A similar sequence is present in the untranslated
3'-terminal region of the rp29 mRNA (35), but it does not occur in any other
yeast ribosomal protein genes so far investigated. Whether this common
sequence fulfils a function in transcription or translation remains to be
examined. In this respect it is notable that the 3'-terminal regions of the
L25 and L17a genes do not contain sequences that are identical to the consensus
sites proposed to be involved in transcription termination in yeast (44). The
sequence AATAAA, which may act as a polyadenylation signal (45), is present 57
nucleotides downstream from the stop codon of the L25 gene but it is not
found in the trailer region of the L17a gene.

A comparative analysis of the 5'-flanking regions of several yeast ribosomal
protein genes has revealed a number of more or less conserved sequence
elements that, however, are not common to all ribosomal protein genes (46).
These boxes may function in the coordinate control of ribosomal protein gene
expression. We searched for similar sequences in the upstream regions of the
genes encoding L17a and L25, (cf. Table V). The most striking homologous

Table V. Homologous sequences upstream of yeast ribosomal protein genes (a).

	consensus	L17a	L25
HOMOL1	AACATC$^{TG}_{CA}$TCA (approx. -300)	AACATCGTACA (-351)	not found
HOMOL2	$^{T}_{A}$CATCTNTA (approx. -300)	ACATCTGTA (-350)	not found
HOMOL3	$^{TC}_{GG}$CTTC$^{T}_{C}$T (approx. -150)	not found	GGCTTCCC (-143) GCCTTCCT (-225)
HOMOL4	$^{T}_{A}$AT$^{T}_{A}$TTNCA (varying positions)	TATTTTTCA (-303)	TATTTGCA (-440)
HOMOL5	TATT$^{TT}_{AA}$ (approx. -80)	TATTTT (-60)	TATTTT (-83)

(a) the consensus sequences for the various boxes are from Ref. 46.

sequences, HOMOL1 and HOMOL2, are present at -351 and -350 respectively in front of the L17a gene, but these boxes are absent from the gene encoding L25. On the other hand HOMOL3 is present at two positions, -143 and -225, upstream of the L25 gene but is not found in the L17a gene. HOMOL4 can be recognized in the 5'-flanking sequences of both ribosomal protein genes though at a rather long distance from the respective initiation codons. HOMOL5 which resembles the TATA box (47) and therefore may fulfil a role in initiation of transcription by RNA polymerase B, occurs at -60 and -83 from the ATG codons of the L17a and L25 genes respectively. Finally the upstream region of the L17a gene contains a CT-rich sequence, which has been suggested to define a high-efficiency yeast promoter (48). A CT-block occurs only occasionally in front of yeast ribosomal protein genes, in addition e.g. those encoding S33 (12) and L29 (32). The significance of all presumed promoter and regulatory sequences mentioned above has to be established by deletion experiments and site-directed mutagenesis. These studies are currently performed.

REFERENCES

1. Dean,D. and Nomura,M. (1982) in The Cell Nucleus (Busch,H. and Rothblum,L. eds) Vol XII, pp. 185-212, Acad. Press, New York.
2. Woolford,J.L., Hereford,L.H. and Rosbash,M. (1979) Cell 18, 1247-1259.
3. Fried,H.M., Pearson,N.J., Kim,C.H. and Warner,J.R. (1981) J. Biol. Chem. 256, 10176-10183.
4. Bollen,G.H.P.M., Cohen,L.H., Mager,W.H., Klaassen,A.W. and Planta,R.J. (1981) Gene 14, 279-182.
5. Fried,H.M. and Warner,J.R. (1982) Nucl. Acids Res. 10, 3133-3148.
6. Schultz,L.D. and Friesen,J.D. (1983) J. Bact. 155, 8-14.
7. Laskin,J.C. and Woolford,J.L. (1983) Nucl. Acids Res. 11, 403-420.
8. 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.
9. Molenaar,C.M.T. (1984) Ph.D. Thesis, Free University, Amsterdam.
10. Woolford,J.L. and Rosbash,M. (1981) Nucl. Acids Res. 9, 5021-5028.
11. Rosbash,M., Harris,P.K.W., Woolford,J.L. and Teem,J.L. (1981) Cell 24, 679-686.
12. 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) Nucl. Acids Res. 11, 7759-7768.
13. Ng,R. and Abelson,J. (1980) Proc. Natl. Acad. Sci. USA 77, 3912-3916.
14. Gallwitz,D. and Sures,I. (1980) Proc. Natl. Acad. Sci. USA 77, 2546-2550.
15. Miller,A.M. (1984) The EMBO J. 3, 1061-1065.
16. Gorenstein,C. and Warner,J.R. (1976) Proc. Natl. Acad. Sci. USA 73, 1547-1551.
17. Warner,J.R. and Gorenstein,C. (1977) Cell 11, 201-212.
18. Kim,C.K. and Warner,J.R. (1983) Mol. Cell. Biol. 3, 457-465.
19. Pearson,N.J., Fried,H.M. and Warner,J.R. (1982) Cell 29, 347-355.
20. Smith,H.O. and Birnstiel,M.L. (1976) Nucl. Acids Res. 3, 2387-2398.
21. Messing,J. (1982) in Genetic Engineering/Principles and Methods (Setlow,J.K. and Hollaender,A. eds) Plenum Press, New York, London, Vol 4, pp. 19-34.
22. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) J. Mol. Biol. 143, 161-178.
23. Maxam,A.M., and Gilbert,W.(1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
24. Leer,R.J., Van Raamsdonk-Duin,M.M.C., Mager,W.H. and Planta,R.J. (1984) FEBS Lett., submitted.
25. Itoh,T., Higo,K-J. and Otaka,E. (1979) Biochemistry 18, 5787-5793.
26. Bollen,G.H.P.M., Mager,W.H. and Planta,R.J. (1981) Mol. Biol. Rep. 8, 37-44.
27. El-Baradi,T.T.A.L., Raué,H.A., De Regt,V.C.H.F. and Planta,R.J. (1984) Eur.J. Biochem., submitted.
28. Bennetzen,J.L. and Hall,B.D. (1982) J. Biol. Chem. 257, 3026-3031.
29. Leer,R.J., Van Raamsdonk-Duin,M.M.C., Molenaar,C.M.T., Cohen,L.H., Mager,W.H. and Planta,R.J. (1982) Nucl. Acids Res. 10, 5869-5878.
30. Teem,J.L. and Rosbash,M. (1983) Proc. Natl. Acad. Sci. USA 80, 4403-4407.
31. Schaap,P.J., Molenaar,C.M.T., Mager,W.H. and Planta,R.J. (1984) Curr. Genet., submitted.
32. Käufer,N.F., Fried,H.M., Schwindinger,W.F., Jasin,M. and Warner,J.R. (1983) Nucl. Acids Res. 11, 3123-3135.
33. Ammerer,G., Hitzeman,R., Hagie,F., Barta,A. and Hall,B.D. (1981) in Recombinant DNA (Walton,A.G. ed) Elsevier Scientific Publishing Co., Amsterdam, pp. 185-197.
34. Pikielny,C.W., Teem,J.L. and Rosbash,M. (1983) Cell 34, 395-403.
35. Mitra,G. and Warner,J.R. (1984) J. Biol. Chem., in press.
36. Breathnach,R. and Chambon,P. (1981) Ann. Rev. Biochem. 50, 349-383.
37. Mount,S.M. (1982) Nucl. Acids Res. 10, 459-527.

38. Langford,C.J. and Gallwitz,D. (1983) Cell 33, 519-527.
39. Langford,C.J., Klinz,F.J., Donath,C. and Gallwitz,D. (1984) Cell 36, 645-653.
40. Thill,G.P., Kramer,R.A., Turner,K.J. and Bostian,K.A. (1983) Mol. Cell. Biol. 3, 570-579.
41. Olins,P.O. and Nomura,M. (1981) Nucl. Acids Res. 9, 1757-1764.
42. Burke,R.L., Tekamp-Olson,P. and Najarian,R. (1983) J. Biol. Chem. 258, 2193-2201.
43. Kozak,M. (1981) Nucl. Acids Res. 9, 5233-5252.
44. Zaret,K.S. and Sherman,F. (1982) Cell 28, 563-573.
45. Fitzgerald,M. and Shenk,T. (1981) Cell 24, 251-260.
46. 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. (1984) Nucl. Acids Res., submitted.
47. Benoist,C. and Chambon,P. (1981) Nature 290, 304-310.
48. Dobson,M.J., Tuite,M.E., Roberts,N.A., Kingsman,A.J., Kingsman,S.M., Perkins,R.E., Conroy,S.C., Dunbar,B and Fothergrill,L.A. (1982) Nucl. Acids Res. 10, 2625-2637.