

---

**Yeast contains two functional genes coding for ribosomal protein S10**

---

Robert J. Leer, Mary M.C. van Raamsdonk-Duin, Constance M.T. Molenaar, Hanneke M.A. Witsenboer, Willem H. Mager and Rudi J. Planta

---

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

---

Received 24 May 1985; Revised and Accepted 2 July 1985

---

**ABSTRACT**

The DNA sequence of the second copy of the gene coding for yeast ribosomal protein S10 was determined and compared with the sequence of the first gene-copy. In addition, the sites at which the transcription of these genes start and terminate are identified. The amino acid coding regions of the two gene copies are virtually identical. The leader and in particular the trailer sequences, however, are significantly different, while the intervening sequences have hardly any homology. Taking advantage of the sequence differences we could establish that both genes are expressed in the vegetatively growing yeast cell; the respective transcripts, however, differ in their relative amounts.

**INTRODUCTION**

Most eukaryotic ribosomal protein genes studied so far are present on the genome in more than one copy. For instance, the haploid genome of Xenopus laevis contains two to five copies of individual ribosomal protein genes (1), while in rodents (2) and in HeLa cells (3) there are 7 to 20 genomic regions homologous to each ribosomal protein gene. Whether all multiple sequences are functional genes or whether they represent pseudogenes can only be determined by transcriptional analyses. In fact, it was shown that most of the reiterated sequences homologous to the genes for mouse ribosomal proteins S16, L7, L18, L30 and L32 are non-functional, processed pseudogenes (4-6). These gene families have only one or a few functional, intron-containing gene(s).

In yeast, most of the cloned ribosomal protein genes have been shown to give rise to two hybridization signals in genomic Southern analyses (7,8). These homologous regions most probably represent two active gene-copies, since so far no pseudogenes have been detected in yeast. In order to obtain experimental support for this assumption the second copy of the gene coding for ribosomal protein S10 was isolated from a yeast cosmid bank (7), and structurally compared with the first copy gene. In addition transcript analyses were performed to examine the expression of both duplicated S10 genes.

### MATERIALS AND METHODS

#### Preparation of nucleic acids and DNA sequence analysis

Recombinant plasmid TJP113 and poly(A)-containing RNA from the yeast *Saccharomyces carlsbergensis* S74 were isolated and purified as described previously (9). Restriction endonuclease digestions were performed according to the recommendations of the suppliers. DNA sequence analysis using the chain termination method was performed as described (9,10).

#### Northern blot analysis

Poly(A)-containing RNA was fractionated on 1,6% agarose gels after prior denaturation in 1 M glyoxal and 50% (v/v) dimethylsulphoxide (11). The RNA was transferred to DBM-paper (12) and hybridized with a 370 bp long Sau3A-Sau3A fragment (codon 52 to codon 175 of the S10-1 gene (9)) labelled by nick translation (13).

#### Heteroduplex analysis

Heteroduplex analysis and electron microscopy were performed essentially as described by Verbeet *et al.* (14).

#### Primer-extended sequence analysis of mRNA

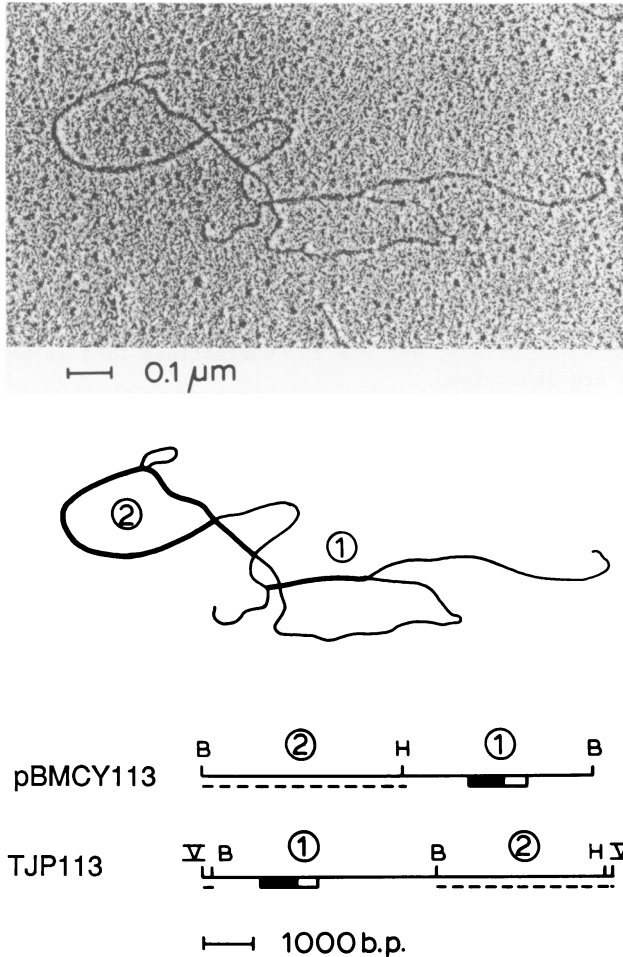
Primer preparation and RNA sequencing using reverse transcriptase were performed as described elsewhere (15). A uniformly labelled 107 nucleotides long EcoRI-HincII primer (see Fig. 2A), was used to determine the 5'-ends of the S10 transcripts. Synthetic oligonucleotide DNA primers were used to distinguish between the transcripts from the duplicate S10 genes. These transcript-specific primers (S10-1 primer 5'-GTTCAACTTCATGATGGC-3' and S10-2 primer: 5'-GTTCAACTTCATGATGTA-3') kindly provided by Dr. J. van Boom, Leiden, were 5'-end labelled using ( $\gamma$ -<sup>32</sup>P)-ATP and T4 polynucleotide kinase. Primer-extended RNA sequencing using reverse transcriptase started at position -7 (see Fig. 3 and Ref. 9) of the pertinent transcripts.

#### S1 nuclease mapping of the 3'-ends of the S10 genes

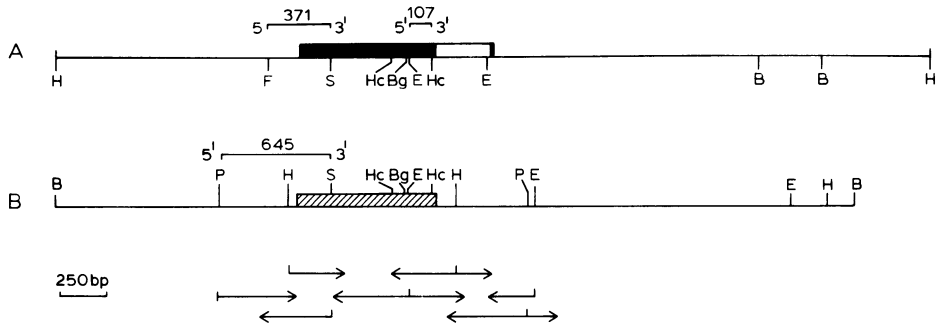
S1 nuclease analysis was performed as described previously (10). The 3'-end of the S10-1 gene was determined using a 371 nucleotides long Sau3A-TaqI fragment (Fig. 2A) labelled at the Sau3A site using Klenow polymerase. Mapping of the 3'-end of the S10-2 gene was performed using a 831 nucleotides long Sau3A-PvuII fragment (645 nucleotides PstI-Sau3A originate from yeast DNA, see Fig. 2B, the additional 186 nucleotides from the M13 cloning region). This probe was labelled at the Sau3A site using Klenow polymerase.

RESULTS AND DISCUSSION

The second copy of the gene encoding ribosomal protein S10 was isolated from a yeast cosmid bank as described elsewhere (7). A 4.5 kb BamHI-generated fragment containing the S10 gene sequences was cloned in the vector pAT153.



**Fig. 1.** Heteroduplex analysis of the duplicate genes encoding S10. Recombinant plasmids carrying the first (pBMCY113) and the second (TJP113) copy of the gene encoding S10 were linearized by cleavage with BamHI (B) and EcoRV (V), respectively. Only coding regions of the S10 gene (1) and regions corresponding to vector DNA (pAT153; pBR322; (2)) form stable duplex structures. In the upperpanels the electronmicrograph is shown as well as the corresponding schematic diagram. In the lower panel the restriction maps of pBMCY113 and TJP113 are depicted; location and orientation of the genes are deduced from Ref. 9.



**Fig. 2.** Physical maps of the pBMCY113 and TJP113 inserts. Restriction enzyme symbols: H = HindIII, Hc = HincII, Bg = BglII, E = EcoRI, B = BamHI, P = PstI, F = FokI, S = Sau3A (only FokI and Sau3A sites used for DNA sequence analysis and for constructing the 3'-probes are shown).

In A the map of the pBMCY113 insert is shown. The position of restriction enzyme sites, as well as the location and the structure of the interrupted S10-1 gene were described in Ref. 9. The 371 nucleotides long probe used for the 3'-end mapping experiment and the 107 nucleotides long primer for the cDNA synthesis are indicated.

In B the region on the TJP113 insert homologous to the 3'-exon of the S10-1 gene (7) is depicted as a dashed box. The arrows indicate the direction and extent of nucleotide sequence analysis. The 645 nucleotides long probe used in the S1 nuclease experiment is shown.

Heteroduplex analysis of the resultant recombinant plasmid (designated TJP113) and pBMCY113 (carrying the first gene-copy) revealed stable hybrids in the region containing the 3'-exon of the S10 gene (see Fig. 1). However, in this analysis the 5'-exon (and therefore the presence of an intron) could not be detected, most likely due to the very short lengths of the 5'-exons ((7) vide infra). On the basis of this heteroduplex study the location as well as the orientation of the second S10-gene on the TJP113 insert could be established. In order to determine the DNA sequence of the pertinent region, a physical map of the TJP113 was constructed (Fig. 2B). The positions of the restriction enzyme sites for EcoRI, BglII and HincII within the homologous regions of the TJP113 and pBMCY113 inserts again indicate that these inserts are virtually identical (see Fig. 2A and 2B). In the lower part of Fig. 2B the sequencing strategy is outlined. The results of the nucleotide sequence analysis are presented in Fig. 3. A comparison with the DNA sequence of the large exon of the S10 gene present on pBMCY113 (described in Ref. 9) demonstrates unequivocally that the 4.5 kb BamHI fragment (TJP113 insert) contains a second S10-coding sequence. We designate the gene-copy present on pBMCY113 (Fig. 2A) as S10-1 and the second copy gene present on TJP113 (Fig. 2B) as

S10-2. The two gene copies code for identical proteins of 236 amino acids. The coding regions of the 3'-exons of both genes are more than 99% homologous, the only differences being four third-base changes. The S10-1 gene as well as the S10-2 gene is interrupted by a single intron between codon 2 and 3. Whereas the S10-1 gene intron is 352 nucleotides long (9), the length of the intron within the S10-2 gene is 394 nucleotides. Apart from the conserved intron-exon boundaries and the sequence TACTAACA (16) which is essential for splicing of yeast pre-mRNAs (17,18) two other homologous elements are found within the introns. The sequence TCAATGAAA at position 247 in the S10-1 gene is found at position 267 in the S10-2 gene. The other conserved sequence TTTAG<sup>G</sup>AAT<sup>A</sup> is present in the intron of both the S10-1 and the S10-2 gene at nucleotide positions 237 and 119 respectively. The introns of the duplicate genes for ribosomal protein S16A and rp28 also appear to contain conserved sequence elements in addition to the generally conserved intron sequences (7,19). Whether these gene-specific conserved sequences have a regulatory function in the expression of the pertinent genes remains to be examined. Comparative analysis of the upstream sequences of both S10-gene copies revealed the presence of two common conserved elements HOMOL1 (20) and RPG-box (21). The postulated function of these homologous sequence blocks is discussed elsewhere (21).

Using a 370 bp DNA fragment of the S10-1 coding region as a probe Northern blot analysis was performed of polyA-containing RNA isolated from *S. carlsbergensis* or *rna 2* mutant cells of *S. cerevisiae* grown at both permissive and restrictive temperatures. As can be concluded from the autoradiogram shown in Fig. 4 obviously two mature mRNAs having different sizes are present as well as two precursor transcripts of different length. If both copy genes encoding S10 are functionally expressed, this result can be explained by assuming that, apart from a different size of the intron, the two S10-transcription units also differ in the length of the leader and/or trailer sequences. Alternatively, either gene (and possibly both) may have multiple transcription initiation and/or termination sites. To map the 5'-ends of the S10 gene transcripts the base-specific chain termination method using reverse transcriptase was employed. Using the 107 nucleotides long EcoRI-HincII fragment from the S10-1 gene as a primer (Fig. 2A) the cDNA sequence obtained by reading the dominant signals (Fig. 5), fully agrees with the DNA sequence upstream of the HincII site (codons 9 and 10) in the S10-1 gene as determined previously (9). Starting from about position -9 (relative to the ATG initiation codon) a different sequence can be discerned as weak signals



interspersed with the strong signals corresponding to the S10-1 leader sequence. These signals correspond to the S10-2 leader sequence. The intensity of the bands corresponding to the S10-2 cDNA sequence suggests that the S10-2 gene transcript is present in a much lower amount than the S10-1 gene transcript. Transcript-specific primers were used to map the 5'-ends of the transcripts of both S10 genes. As can be seen in Fig. 6 the 5'-ends of the S10-1 mRNA are mapped at position -25, -30, -35 and -39, the signal at -25 being the major one. The S10-2 mRNA revealed a major 5' end at -30 and a minor start site at -70.

Comparing the DNA sequence of the S10-1 leader with the DNA sequence in front of the ATG codon of the S10-2 transcript short homologous regions could be detected (Fig. 7). Likewise, the leader regions of the duplicate genes encoding ribosomal proteins L16, rp51, rp28 and S16A show small sequence homologies (Fig. 7). On the other hand, the sequences downstream of the translation stop codons show relatively little homology. Whether the short homologous leader sequences are important in post-transcriptional events (e.g. mRNA stability, translational control of the mutual ribosomal protein mRNAs) is an intriguing but so far unanswered question.

Subsequently, the 3'-ends of the S10 mRNAs were mapped by S1-nuclease analysis. The probes used for the respective transcripts are depicted in Fig. 2A and 2B. The results presented in Fig. 8 show that both genes give rise to transcripts with heterogeneous 3'-ends. The 3'-ends of the S10-1 transcripts map within a region of 70-85 nucleotides downstream of the translation stop codon. The 3'-ends of the S10-2 mRNAs map at 80, 129, 143 and 165 nucleotides downstream of the stop codon, the signal at position 143 being the major one. On the basis of the latter signal for the S10-2 gene as compared with the signals found for the transcripts of the S10-1 gene, and the about equal lengths of the respective leader sequences (see above), we conclude that the mature S10-2 mRNA is at least 60 nucleotides longer than its S10-1 counterpart. As stated above the precursor transcripts of the S10 genes also differ in the length of the respective introns. These differences

**Fig. 3.** DNA sequence of the S10-2 gene and the inferred amino acid sequence of the S10 protein. Third-base changes with respect to the S10-1 gene (9) are indicated with an asterisk. The positions of the 5'-ends of the S10-2 mRNA are indicated by triangles ( $\blacktriangledown$  major,  $\blacktriangledown$  minor start site). Positions of the 3'-ends of the S10-2 mRNAs are underlined. Major signal  $\text{—}$ , intermediate  $\text{-----}$  and minor signals  $\text{—}$ .



**Fig. 4.** Analysis of precursor and mature transcripts from the genes encoding ribosomal protein S10. Poly(A) RNA was isolated from *Saccharomyces carlsbergensis* (wt, wild-type), or from the *rna 2* mutant of *Saccharomyces cerevisiae* grown at 23°C (P, permissive temperature) and 36°C (R, restrictive temperature).

in size of both precursor and mature transcripts found at the nucleotide level are in good agreement with the dual signals in the Northern blot analyses discussed above. The S1-protected bands at position 192 (Fig. 8, lanes a and f) correspond to the end of the open reading frame of the transcript from the other S10 gene-copy, which competes with the probe during the hybridization reaction. Since the S1-analyses were performed using an excess of poly(A)-containing RNA the S1-protected bands obtained are not proportional to the relative amount of the respective S10 mRNAs present in the yeast cell (24).

Within the trailer sequences of both S10 transcripts several elements are present that may function in transcription termination and/or polyadenylation. The S10-1 gene trailer sequence contains the tripartite consensus sequence proposed by Zaret and Sherman (25); *viz.* TAG-3n-TATGT-13n-TTT, at 41 nucleotides downstream of the stop codon. On the other hand in the S10-2 gene sequence only the TATGT element is present at position 130 located between the mapped 3'-ends of the transcripts. In addition, in the S10-trailer sequence elements can be found that fit partially to the TAAATA<sup>A</sup><sub>G</sub> consensus sequence proposed to function in poly(A)-addition (26), although their significance remains to be established, since the consensus of this octanucleotide sequence is not very strong (see Ref. 26). Finally the sequence TTTTATA which has been implicated in transcription termination (27, 28) is found in both S10-gene trailer sequences, at +22 and +97 downstream of



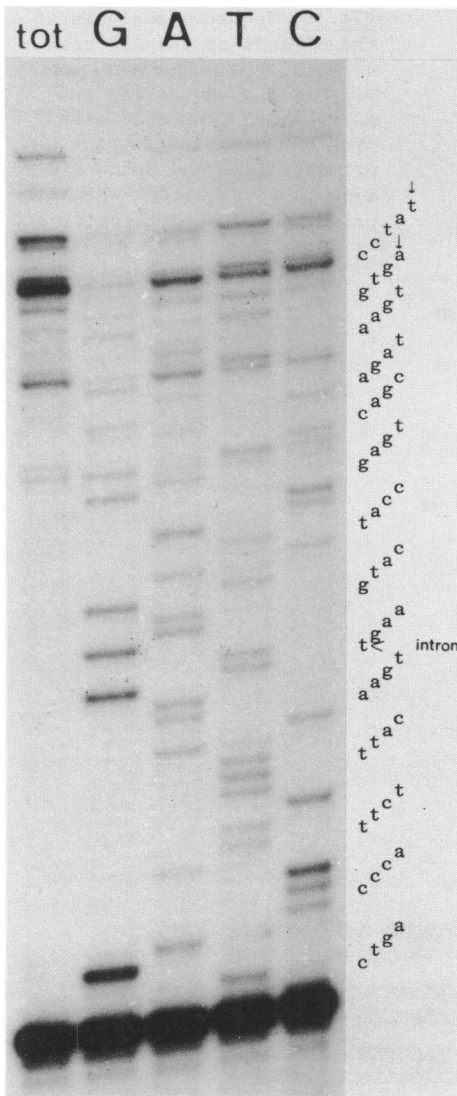
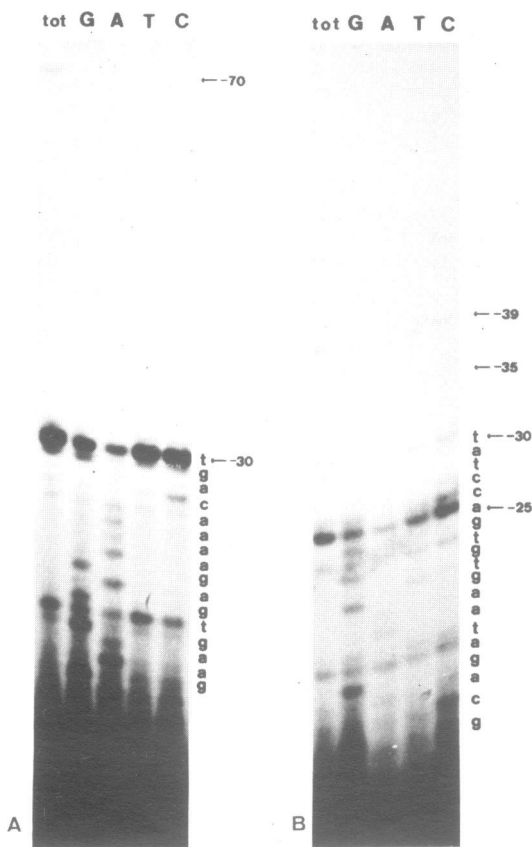


Fig. 5. Determination of the 5'-ends of the S10 transcripts by primer extension. A uniformly labelled EcoRI-HincII fragment of 107 nucleotides (see Fig. 2A) was used to prime cDNA synthesis from poly(A) RNA as described in MATERIALS AND METHODS. The arrows indicate the 5'-ends of the S10-1 transcripts; the position of the intron splice junction is indicated (<).

the stop codons of S10-1 and S10-2, respectively. The 3'-ends of the respective S10 mRNAs map at sites 40 to 70 nucleotides away from this signal, which is in good agreement with previously published data (28).

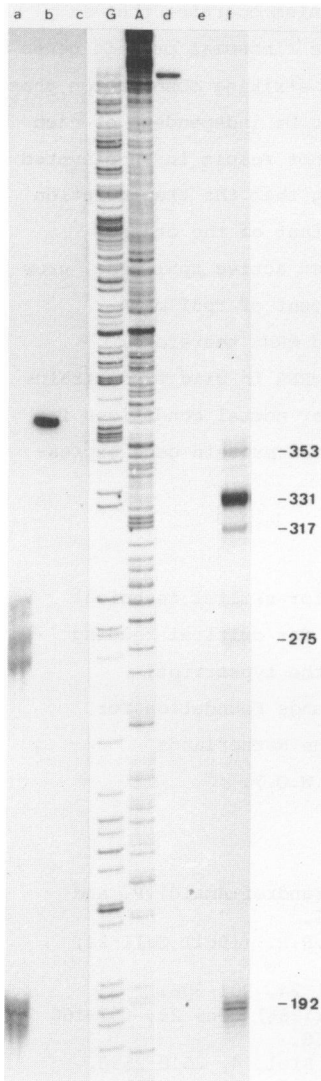
Taking together the results obtained with the mapping experiments and the Northern blot studies (7), we suggest that the transcripts derived from the S10-1 gene are present in a higher amount than those transcribed from gene copy-2. The same picture emerges from the study of other duplicate



**Fig. 6.** Sequence analysis of the 5'-ends of the S10 transcripts. cDNA synthesis was performed as described in MATERIALS AND METHODS using synthetic oligonucleotide DNA primers. In A and B the S10-2 and S10-1 transcript-specific primers were used respectively.

	↓		
S10-1		ATCCAGTCTCAATGAC GACTGAGCCATC	ATG
S10-2	↓	GTGACAAAAGACTCA -AGACAGACT-ATACATC	ATG
L16-1		CCAA---CATATACAAAATACGGGTTCAA-G	ATG
L16-2		CCAAGAACATA--CAACATA-CC---CAAAG	ATG
rp28-1		AAGAAAGAGATCAT---CAAGCTTAC-GGATTCACA	ATG
rp28-1		AAGAAAAA-ATCAATTAGAA GAAAACAAAAACAAA	ATG
S16A-1		ATCCA--CAATAGAG--TAAAA	ATG
S16A-2		ATAGAACCAA--GATAATAANG	ATG
rp51-1		TCAAGTCTCGAGACTACCA-ATACAAA	ATG
rp51-2		AATAGATAAAGAAAAACCAATA--AAA	ATG

**Fig. 7.** Comparative analysis of sequences preceding the start methionine of duplicate ribosomal protein genes. To maximize homology shifts (-) were allowed. Homologous elements consisting of two or more nucleotides are boxed. Arrows indicate mapped (major) 5'-ends of the respective mRNAs. Data: L16-1,2 (15,20); rp51-1,2 (22,23); rp28-1,2 and S16A-1,2 (19).



**Fig. 8.** Mapping of the 3'-ends of the S10-1 and S10-2 mRNAs. Lanes a to c show the product analysis using the 371 nucleotides long probe for the S10-1 gene (MATERIALS AND METHODS and Fig. 2A). Lane a: 36  $\mu$ g poly(A)-containing yeast RNA with 100 u S1 nuclease; lane b: untreated probe; lane c: 36  $\mu$ g *Bacillus licheniformis* RNA as a control with 100 u S1 nuclease. The lanes d to f show the product analysis using the 831 nucleotides long probe (see Materials and Methods and Fig. 2B) for the S10-2 gene. Lane d: untreated probe; lane e: 36  $\mu$ g *B. licheniformis* RNA with 100 u S1 nuclease; lane f: 36  $\mu$ g poly(A)-containing yeast RNA with 100 u S1 nuclease. The GA ladder of a known sequence was used to measure the length of the S1 protected products.

ribosomal protein genes (*viz.* those encoding S16A, L16 and rp51 (7,15,19,22, 23). Although the genes for yeast ribosomal proteins usually are duplicated, a few ribosomal proteins are encoded by a single gene (7,8), and therefore the question arises how the yeast cell can compensate for this unequal gene dosage and maintain an equimolar synthesis of the ribosomal proteins. Kim and Warner demonstrated (29) that ribosomal protein mRNAs, irrespective of their gene number, are present in roughly equimolar amounts. Therefore we expect

that at the transcriptional level a regulation mechanism operates that is responsible for tuning the transcription of duplicate ribosomal protein genes to that of the unique ones. In this respect it was a striking observation that the expression of the duplicate rp51 genes appears to be independent of each other (23). Disruption of one of the rp51 genes did not result in an elevated transcription level of the remaining gene, suggesting that the transcription rate of each of these rp51 genes is not affected by that of the other. Surprisingly, it was found that cells lacking the most active rp51 gene, grow better than one might predict on the basis of the amount of rp51 mRNA supplied by the remaining gene. The remaining limited mRNA therefore, probably is utilized more efficiently than the rp51 mRNA in wild type strains (23). Underutilization of ribosomal protein mRNA under normal conditions is consistent with the results obtained on yeast ribosomal protein gene expression from other laboratories (29-31).

### ACKNOWLEDGEMENTS

The authors are grateful to Mr. A.E.M. Jansen for skilful technical assistance during part of the work, to Dr. H.A. Raué for critical reading the manuscript and to Mrs. P.G. Brink for preparing the typescript.

This work was supported in part by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

### REFERENCES

1. Bozzoni, I., Beccari, E., Luo, Z.X., Amaldi, F., Pierandrei-Amaldi, P. and Campioni, N. (1981) *Nucl. Acids Res.* 9, 1069-1086.
2. D'Eustachio, P., Meyuhas, O., Ruddle, F. and Perry, R.P. (1981) *Cell* 24, 307-312.
3. Monk, R.J., Meyuhas, O. and Perry, R.P. (1981) *Cell* 24, 301-306.
4. Peled-Yalif, E., Cohen-Binder, I. and Meyuhas, O. (1984) *Gene* 29, 157-166.
5. Dudov, K.P. and Perry, R.P. (1984) *Cell* 37, 457-468.
6. Wiederman, L.M. and Perry, R.P. (1984) *Mol. Cell. Biol.* 4, 2518-2528.
7. Molenaar, C.M.T. (1984) Ph.D. Thesis, Vrije Universiteit, Amsterdam.
8. Fried, H.M., Pearson, N.J., Kim, C.H. and Warner, J.R. (1981) *J. Biol. Chem.* 256, 10176-10183.
9. 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.
10. 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.
11. McMaster, G.K. and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4838.
12. Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5350-5354.

13. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
14. Verbeet, M.Ph., Klootwijk, J., van Heerikhuizen, H., Fontijn, R.D., Vreugdenhil, E. and Planta, R.J. (1983) *Gene* 23, 53-63.
15. Leer, R.J., van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.J. (1984) *FEBS Lett.* 175, 371-376.
16. Leer, R.J., van Raamsdonk-Duin, M.M.C., Hagendoorn, M.J.M., Mager, W.H. and Planta, R.J. (1984) *Nucl. Acids Res.* 12, 6685-6700.
17. Langford, C.J. and Gallwitz, D. (1983) *Cell* 33, 519-527.
18. Langford, C.J., Klinz, F., Donath, C. and Gallwitz, D. (1984) *Cell* 36, 645-653.
19. Molenaar, C.M.T., Woudt, L.P., Jansen A.E.M., Mager, W.H., Planta, R.J., Donovan, D.M. and Pearson, N.J. (1984) *Nucl. Acids Res.* 12, 7345-7358.
20. 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., Fried, H. and Rosbash, M. (1984) *Nucl. Acids Res.* 12, 8295-8312.
21. Leer, R.J., van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.J. (1985) *Curr. Genet.* 9, 273-277.
22. Teem, J.L. and Rosbash, M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4403-4407.
23. Abovich, N. and Rosbash, M. (1984) *Mol. Cell. Biol.* 4, 1871-1879.
24. Miller, K.G. and Sollner-Webb, B. (1981) *Cell* 27, 165-174.
25. Zaret, K.S. and Sherman, F. (1982) *Cell* 28, 563-573.
26. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3018-3025.
27. Henikoff, S., Kelly, J.D. and Cohen, E.H. (1983) *Cell* 33, 607-614.
28. Henikoff, S. and Cohen, E.H. (1983) *Mol. Cell. Biol.* 4, 1515-1520.
29. Kim, C.H. and Warner, J.R. (1983) *J. Mol. Biol.* 165, 79-89.
30. Pearson, N.J., Fried, H.M. and Warner, J.R. (1982) *Cell* 29, 347-355.
31. Himmelfarb, H.J., Vassarotti, A. and Friesen, J.D. (1984) *Mol. Gen. Genet.* 195, 500-506.