Volume 13 Number 3 1985 Nucleic Acids Research

The genes for yeast ribosomal proteins S24 and L46 are adjacent and divergently transcribed

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

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

Received 4 December 1984; Accepted 7 January 1985

ABSTRACT

Unlike most yeast ribosomal protein genes studied so far the genes coding for S24 and L46 are adjacent on the genome. Sequence analysis showed that the two genes are transcribed divergently, their initiation codons being 630 bp apart. Taking the respective ATG translation start sites as reference points, the 5'- end of L46 mRNA was mapped at position -26, while the S24 mRNA showed two major 5'-ends mapping at positions -13 and -16 respectively. Unlike most other yeast ribosomal protein genes, the gene for S24 does not contain an intron. Its coding region encompasses 390 nucleotides encoding a protein of 14762 D. The gene for L46 on the other hand is split by an intron of 386 nucleotides starting after its second codon. This gene encodes a small, very basic protein having a molecular weight of 6334 D. Yeast ribosomal proteins S24 and L46 show striking homologies with ribosomal proteins from other organisms. In particular, yeast L46 is clearly the evolutionary counterpart of rat liver L39. A search of the intergenic region for sequence elements previously identified as common to most yeast ribosomal protein genes, revealed the presence of a single conserved box (RPG-box) roughly equidistant from the transcription initiation sites of both genes. We suggest that this box acts as a regulatory signal in either orientation and thus influences the expression of both genes simultaneously.

INTRODUCTION

With the aim to elucidate the molecular mechanisms underlying the coordinate regulation of ribosomal protein gene expression in yeast we have characterized several cloned yeast ribosomal protein genes (1-6). It turned out that most ribosomal protein genes share some general characteristics: they are not clustered, occur duplicated on the yeast genome and contain an intron near their 5'-end. In some cases two different ribosomal protein genes appear to be physically linked. For instance the split genes encoding S16A and rp28 are separated by only 600 bp and are transcribed in the same direction (5). In this paper we describe a second pair of linked yeast ribosomal protein genes. The split gene coding for L46 and the unsplit gene coding for S24 were found to be located adjacent to each other in a head-to-head arrangement. The intergenic region consists of about 600 bp and must contain the

signals controlling the expression of both genes.

MATERIALS AND METHODS

DNA preparation and sequence analysis

Plasmid pBMCY135 containing the genes coding for S24 and L46 was purified as described previously (2). Restriction enzyme digestions were performed as recommended by the suppliers (New England Laboratories; Boehringer Mannheim). Appropriate DNA fragments were subcloned in M13 mp8 or mp9, and sequence analysis was carried out according to the dideoxy chain termination method (7).

Primer extended sequence analysis of mRNA

RNA sequencing using reverse transcriptase was performed as described elsewhere (4). The primer for the S24 gene was labelled by repair synthesis of the subcloned BglII-BglII fragment in M13 mp8. Subsequent digestion of the labelled fragment with Sau3A yielded a 162 nucleotides long primer (see Fig. 1). The primer for mapping the 5'-end of the L46-transcript was a 55 nucleotide fragment obtained by digestion with Sau3A plus HinfI of cDNA synthesized using the subcloned HaeIII-HaeIII fragment in M13 mp8 (see Fig. 1) (9).

RESULTS AND DISCUSSION

a. Structure of the genes for yeast ribosomal proteins S24 and L46

From a colony bank of HindIII-generated yeast DNA fragments in pBR322 several recombinants containing yeast ribosomal protein genes have been isolated (2). Hybrid-selected in vitro translation followed by two-dimensional gel electrophoresis revealed that recombinant plasmid pBMCY135 carries at least the genes encoding yeast ribosomal proteins S24 and L46 (9). The pertinent physical map is presented in Fig. 1A. Electron microscopic R-loop analysis, successfully used in locating a number of ribosomal protein genes on cloned DNA fragments (9), in this case led to only a rough estimate of the position of the two genes, because of the complexity of the R-loop structures observed (9). However, on the basis of the known N-terminal amino acid sequence of ribosomal protein S24 (which is identical to YS22 - [10]) as well as the preferent codon usage observed within yeast ribosomal protein genes (2), we could predict a BglII site at the very 5'-end of the S24 coding sequence. Starting from these assumptions the sequencing strategy for both ribosomal protein genes as outlined in (the legend to) Fig. 1B was developed. The results of the nucleotide analysis are shown in Fig. 2. The gene coding

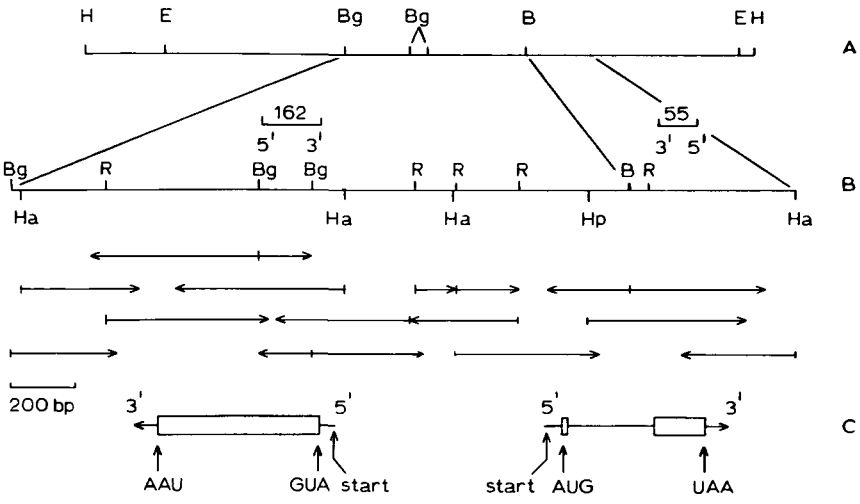


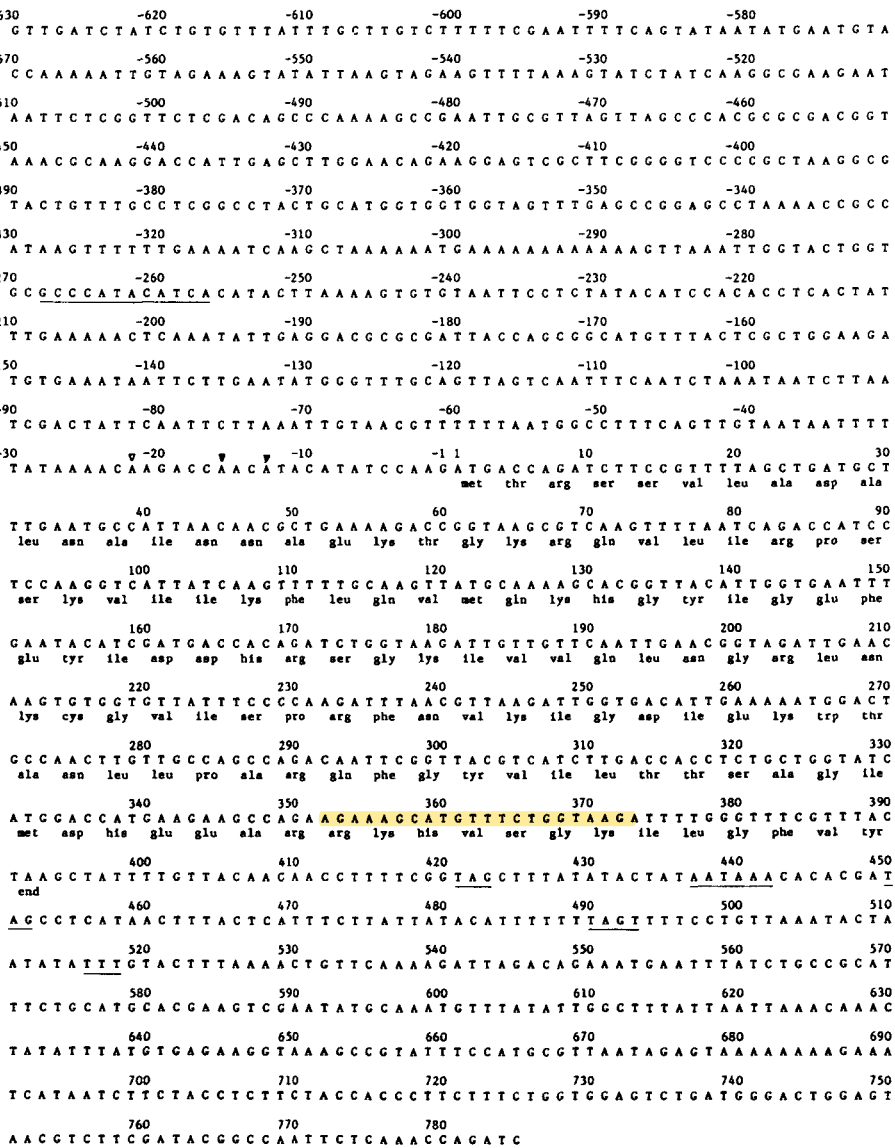
Fig. 1. Map of the insert of pBMCY135 and structural analysis of the genes coding for S24 and L46.
The position of some restriction enzyme sites indicated in A was published previously (9). H = HindIII, E = EcoRI, Bg = BglII, B = BamHI, Hp = HpaI, Ha = HaeIII, R = RsaI.
In B the sequence strategy is shown. The arrows give the extent of nucleotide analysis. The 162 and 55 nucleotides long primers used for the cDNA synthesis are indicated.
In C the structures of the S24 and L46 transcripts are given.

for S24 contains an uninterrupted reading frame of 390 nucleotides, whereas the coding region of the L46 gene comprises 153 nucleotides which is interrupted after the second codon by an intervening sequence of 386 nucleotides. In contrast to the clustered rp28 and S16A genes which are transcribed in the same direction (5), the transcription of the genes for S24 and L46 is divergent. Their translation initiation sites are 630 bp apart. Southern analysis suggested that the L46 gene is unique whereas the S24 gene may be duplicated on the yeast genome (result not shown; [11]).

b. Mapping of the 5'-ends of the mRNAs

The cap sites of both the L46 and S24 gene were determined by primer-extended cDNA synthesis. Using a Sau3A-Sau3A primer fragment for reverse transcription (see Fig. 1) two major 5'-ends of the S24 mRNA were mapped at position -13 and -16. In addition to these major transcription start sites a minor site at -22 was observed (see Fig. 3A). Heterogeneous 5'-ends have also been observed for several other yeast ribosomal protein mRNAs (e.g. Ref. 3). Mapping of the 5'-end of the L46 transcript, by use of a HinfI-Sau3A primer fragment (Fig. 1), revealed a dominant transcription start site located at

26 nucleotides upstream from the translation start codon. A weak primer extended product was found at -22 nucleotides (see Fig. 3B). In this case the mRNA sequence determination also provided definite proof for the presence of an intron within the L46 gene. The nucleotide structure of the transcription initiation regions is in good agreement with the structure of other cap sites previously found for other yeast ribosomal protein genes (3).



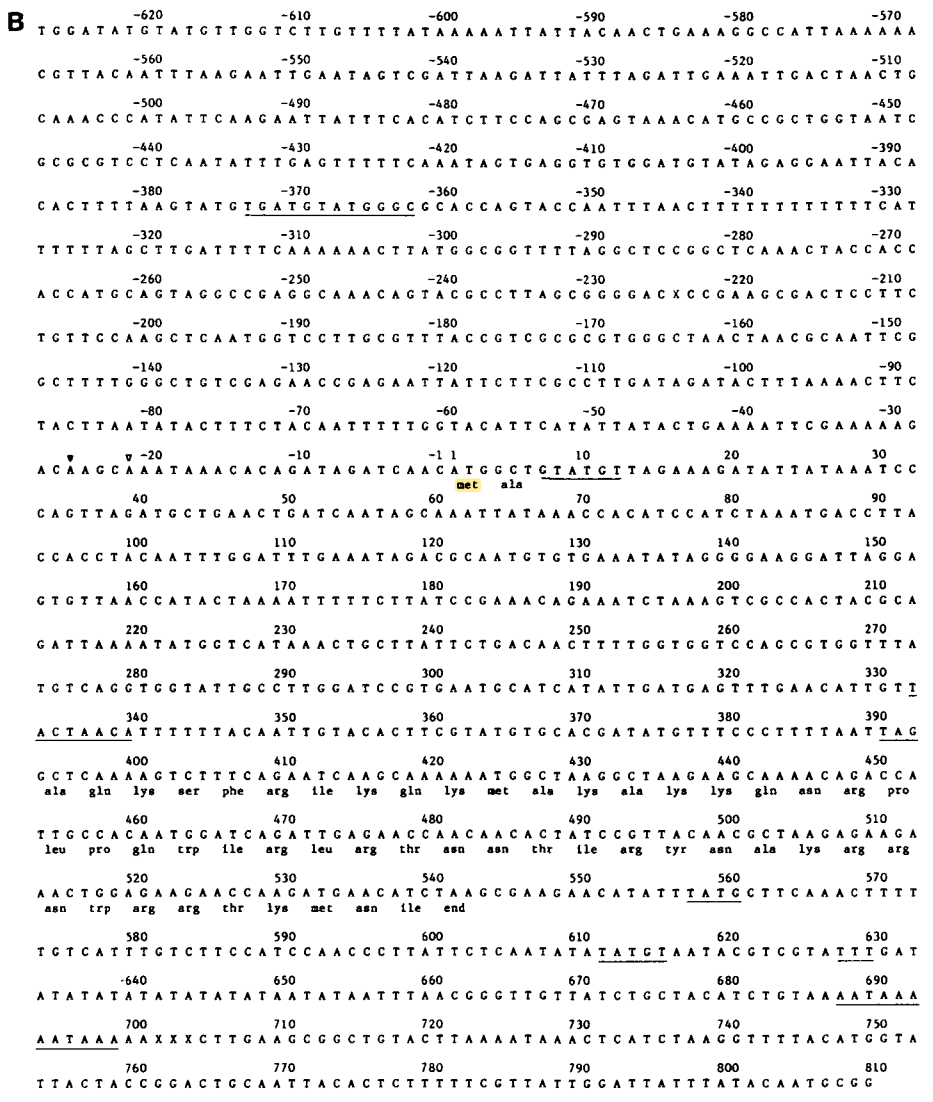


Fig. 2. DNA sequence of the genes for yeast ribosomal proteins S24 (A) and L46 (B). The nucleotides are numbered starting from the first nucleotide of the respective initiation codons. The positions of the 5'-ends of the S24 and L46 mRNA are indicated by triangles (▲ major-, Δ minor start site).
In addition conserved sequence elements occurring in the non-coding regions are underlined.
For the sake of clarity the intergenic region is shown in both DNA sequences.

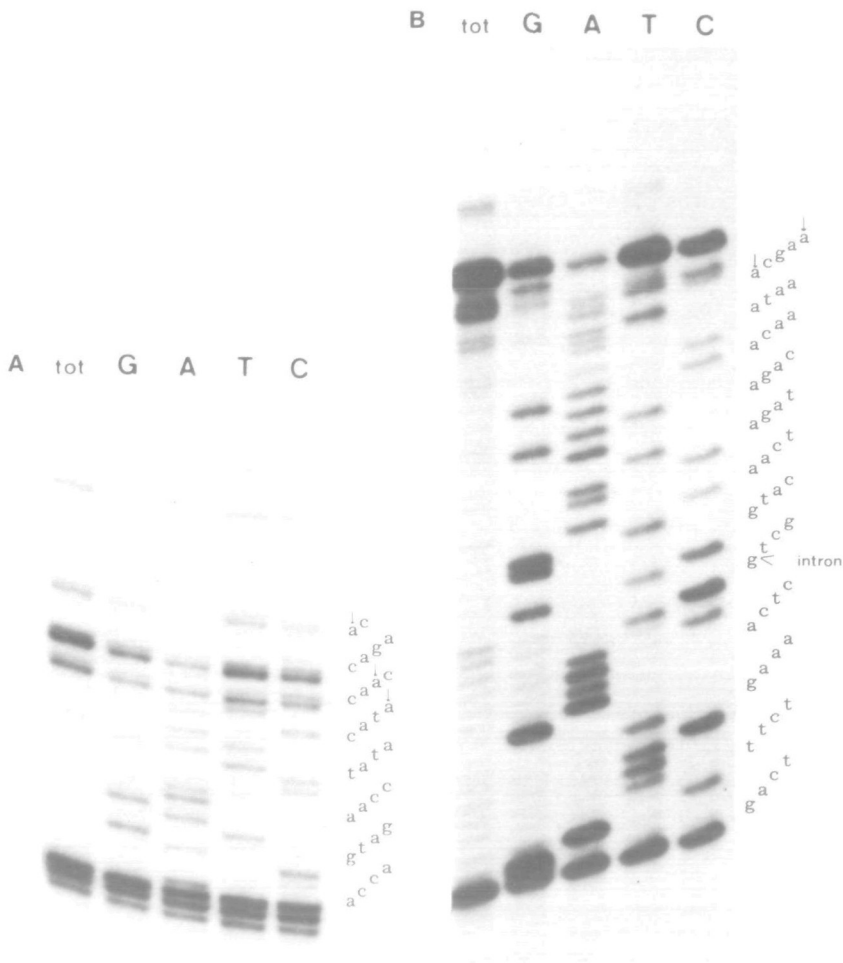


Fig. 3. Sequence analysis of the 5'-ends of the S24 (A) and L46 (B) mRNAs. cDNA synthesis was performed as described in Materials and Methods using the 162 and 55 nucleotides long primers depicted in Fig. 1. The arrows indicate the 5'-termini of the respective mRNAs; 'intron' indicates the splice junction in the L46-gene.

c. The structure of ribosomal proteins S24 and L46

Yeast ribosomal protein S24 as deduced from the nucleotide sequence of its gene is a protein of 14762 D with a net positive charge of +9, while L46 is a very small and extremely basic protein with a molecular weight of 6334 D and a net charge of +17.

The amino acid sequences of S24 and L46 as deduced from the nucleotide

sequence are in perfect agreement with the N-terminal amino acid sequences determined by Otaka et al. (12,13) for the proteins YS22 and YL40, which are identical to S24 and L46 respectively (10,14). The amino acid sequence of yeast S24 is partially homologous to that of HS20 from Halobacterium cutirubrum (12), while yeast L46 shows a strong similarity to L36 from Schizosaccharomyces pombe (13). Even more striking is the extended sequence homology of yeast L46 and rat liver L39 ([15]; see Fig. 5). This finding indicates the existence of evolutionary constraints on the structure of at least some ribosomal proteins which suggests a fundamental role of these proteins in the structure and function of the ribosome.

d. Conserved sequences in the non-coding regions

The intergenic DNA region of 588 nucleotides between the transcription initiation sites of the divergently transcribed L46 and S24 genes is expected to contain the signals controlling the expression of both ribosomal protein genes. Previous computer analysis has revealed several conserved boxes upstream of most yeast ribosomal protein genes that might act as regulatory signals (16). The most striking of these boxes is a 12-nucleotide long element (HOMOL 1) having the consensus sequence AACATC$^{TG}_{CA}$T$^{G}_{A}$CA. Neither strand of the intergenic region, however, contains this sequence element. On the other hand, another, recently detected, conserved sequence element (17), is present in the intergenic region. This element, the RPG-box, having the consensus sequence ACCCATACATTT, occurs once on the non-coding strand of the S24 gene at position -268 relative to the translation start site, which corresponds to position -363 relative to the translation start of the L46 gene (Fig. 2A,B).

In the latter case the RPG-box of course is located on the coding strand and its orientation relative to the gene is reversed. Since the RPG-box is found at about the same relative position upstream of most other yeast ribosomal protein genes (17) sometimes in one and sometimes in the other orientation, we suggest that the single box present in the intergenic region

```
S24      ATAATCTTAATCGA-CTATTCAATTCTTAAATTGTAACGTTT
         ... . ....  ..  ..... ... .  ... ........
L16-1    ATACTGTTAAGAGAGGCATTCATTTCGTGTATTATAACGTTT

S24      -12 n-TCAGTT-16 n-AACA-ACCAACATACATATCCAAG ATG
              ......      .... .. .. ....  . ....
L16-1    - 4 n-TCAGTT-14 n-AACATACAAAAATACGCGTTCAAG ATG
```

Fig. 4. Sequence homology in the upstream regions of the genes for S24 and L16-1 (16). The positions of identical bases are indicated by a dot.

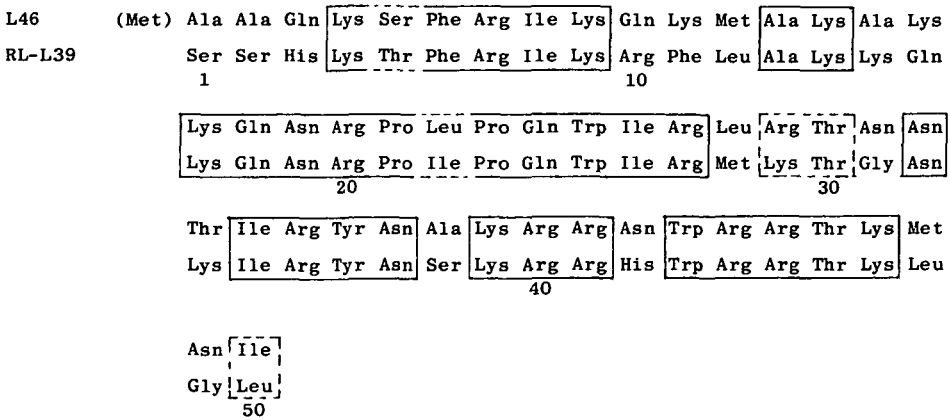


Fig. 5. Comparison of the primary structure of yeast ribosomal L46 with ribosomal protein L39 from rat liver (15). Identical amino acids are boxed, functional identical ones are indicated by dotted lines.

of the L46 and S24 genes is involved in controlling the expression of both genes.

The upstream sequence of the S24 gene shows a remarkable homology with the 5'-flanking region of a gene encoding yeast ribosomal protein L16 (see Fig. 4). Similar homologies have been found for other pairs of different yeast ribosomal protein genes (17). It is therefore tempting to speculate that these regions are involved in the coordinate transcription of the respective genes. No sequence conservation could be detected comparing the upstream sequence of the L46-gene with any other of the known yeast ribosomal protein gene sequences.

The sequences 3' to the coding regions contain the conserved boxes implied in termination and/or polyadenylation (Fig. 2A,B; Ref. 18,19).

Finally the intron of the L46-gene contains all conserved sequence elements found in the other split yeast nuclear (ribosomal) protein genes (16,17,20,21,22), viz. GTATGT at 5'-splice site, TAG at the 3'-end and TACTAACA which in the L46 intron is located 61 nucleotides from the 3'-splice junction. The latter sequence element has recently been suggested to represent the target-site for generating a lariat-like structure intermediate in the splicing of pre-mRNAs in yeast (J. Abelson cited in 23;24).

ACKNOWLEDGEMENTS

The authors are grateful to Drs L.P. Woudt and R.T.M. Nieuwint for stimulating discussions, to Dr. H.A. Raué for critical reading of 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. 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.
2. 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.
3. 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.
4. Leer,R.J., Van Raamsdonk-Duin,M.M.C., Mager,W.H. and Planta,R.J. (1984) FEBS Lett. 175, 371-376.
5. Molenaar,C.M.T., Woudt,L.P., Jansen,A.E.M., Mager,W.H., Planta,R.J., Donovan,D. and Pearson,N.J. (1984) Nucl. Acids Res. 12, 7345-7358.
6. Schaap,P.J., Molenaar,C.M.T., Mager,W.H. and Planta,R.J. (1984) Curr. Genet., in press.
7. Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) J. Mol. Biol. 143, 161-178.
8. Bollen,G.H.P.M., Cohen,L.H., Mager,W.H., Klaassen,A.W. and Planta,R.J. (1981) Gene 14, 279-287.
9. 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.
10. Bollen,G.H.P.M., Mager,W.H. and Planta,R.J. (1981) Mol. Biol. Rep. 8, 37-44.
11. Molenaar,C.M.T. (1984) Ph.D. Thesis, Free University, Amsterdam.
12. Otaka,E., Higo,K. and Osawa,S. (1982) Biochemistry 21, 4545-4550.
13. Otaka,E., Higo,K. and Itoh,T. (1983) Mol. Gen. Genet. 191, 519-524.
14. Michel,S., Traut,R. and Lee,J. (1983) Mol. Gen. Genet. 191, 251-256.
15. Lin,A., McNall,J. and Wool,I.R. (1984) J. Biol. Chem. 259, 487-490.
16. 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., in press.
17. Leer,R.J., Van Raamsdonk-Duin,M.M.C., Mager,W.H. and Planta,R.J. (1985) submitted.
18. Fitzgerald,M. and Schenk,T. (1981) Cell 24, 251-260.
19. Zaret,K.S. and Sherman,F. (1982) Cell 28, 563-573.
20. Miller,A.M. (1984) The EMBO J. 3, 1061-1065.
21. Langford,C.J. and Gallwitz,D. (1983) Cell 33, 519-527.
22. Langford,C.J., Klinz,F.J., Donath,C. and Gallwitz,D. (1984) Cell 36, 645-653.
23. Ruskin,B., Krainer,A.R., Maniatis,T. and Green,M.R. (1984) Cell 38, 317-331.
24. Padgett,R.A., Konarska,M.M., Grabowski,P.J., Hardy,S.F. and Sharp,P.A. (1984) Science 225, 898-903.