
The structure of the gene coding for the phosphorylated ribosomal protein S10 in yeast

Robert J. Leer, Mary M.C. van Raamsdonk-Duin, Constance M.Th. Molenaar, Louis H. Cohen*,
Willem H. Mager and Rudi J. Planta

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

Received 23 July 1982; Revised and Accepted 2 September 1982

ABSTRACT

From previous studies on cloned yeast ribosomal protein genes we obtained evidence that a large number of them contain an intron [Bollen *et al.* (1982) *Gene* 18, 29-38]. In the temperature-sensitive *rna2*-mutant transcription of these genes leads to the accumulation of precursor RNAs at the restrictive temperature. These precursor mRNAs are several hundreds of nucleotides longer than the respective mature mRNAs. The split character of one of these ribosomal protein genes, viz. the gene coding for the major phosphorylated small-subunit protein S10, was further established by sequence analysis. The intervening sequence interrupts the coding sequence after the second codon and has a length of 352 nucleotides. Genomic Southern hybridizations with a DNA fragment carrying part of the S10-gene revealed that this gene is duplicated on the yeast genome. The molecular weight of S10 as deduced from the sequence analysis was estimated to be 31462 dal. Comparison of the N-terminal amino acid sequence of the yeast ribosomal protein S10 with that of ribosomal protein S6 from rat liver revealed a striking homology between both proteins. Moreover, at the C-terminal end of the yeast ribosomal protein the sequence Arg-Ala-Ser-Ser-Leu-Lys is present which is very similar to the phosphorylation site of the rat liver protein S6.

INTRODUCTION

Eukaryotic ribosomes contain a large number of ribosomal proteins which either may play some role in ribosomal assembly and maturation or may fulfil specific functions in the translation process.

In recent years attempts have been made to compare ribosomal protein species from different eukaryotic organisms in order to identify possible functional homologies. Though in general the twodimensional separation of ribosomal proteins from various origins give similar results, an effective comparison is hampered by the diversification of the methods applied [see *e.g.* Ref. 1]. Therefore, a reliable correlation between ribosomal proteins from different cells on basis of their electrophoretic behaviour has to await the adoption of a standard separation and numbering system. Comparative studies of functional homologies between eukaryotic and prokaryotic ribosomal

proteins so far are restricted to the bacterial protein L7/L12 which is involved in EF_2 -dependent GTPase activity [2]. Several lines of evidences exist that mammalian ribosomal protein L40/L41 [see Ref. 3] as well as the yeast protein L44/L45 [4] are related to L7/L12. In most cases, however, no special function can be attributed to a certain ribosomal protein. Then comparative structural analyses are required to discover at least structural homologies. This approach has been applied for some yeast, rat liver and *E. coli* ribosomal proteins [5]. Furthermore, correlations can be made on basis of the nature and the extent of modifications of certain ribosomal proteins. Considerable work has been done on the phosphorylation of the ribosomal protein, designated S6, which is the major phosphorylated 40S ribosomal protein in higher eukaryotic cells [see Ref. 6,7]. On basis of its rapid and reversible phosphorylation the yeast 40S ribosomal protein S10 almost certainly is the analogue of mammalian S6 [8-10].

In a previous report we have described the isolation of a cloned yeast DNA fragment carrying the gene coding for yeast ribosomal protein S10 [11]. The present paper deals with the sequence of this gene. Comparison of the aminoacid sequence predicted by the nucleotide sequence with the primary structure data of rat liver ribosomal protein S6 [12,13] reveals a considerable homology both at the N- and the C-terminal part of the molecule.

The work presented in this paper forms part of our study on ribosomal protein gene expression in yeast. We recently presented evidence that a number of ribosomal protein genes contain an intron [14]. In this paper we demonstrate by Northern blot analyses of RNA isolated from the temperature sensitive *rna2* mutant at both 23^o and 36^oC, that the primary transcripts from these genes have to be spliced. The location of the intervening sequence within the coding sequence of the S10 gene is presented. In addition we show that the yeast genome contains two copies of the S10 gene.

MATERIALS AND METHODS

Isolation of nucleic acids

Recombinant plasmids pBMCY44, pBMCY89, pBMCY113 and pBMCY138 were purified from Triton treated bacterial spheroplasts by CsCl-ethidium bromide density gradient centrifugation [15].

PolyA-containing RNA was isolated from the yeast *Saccharomyces carlsbergensis* S74 and the yeast *Saccharomyces cerevisiae rna2* (ts 368⁻) according to a procedure described previously [16]. Yeast DNA was isolated essentially as described previously [17].

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 Taq I, Sau 3A and Msp I were determined by the partial digestion procedure of Smith and Birnstiel [18].

Northern blot analysis

PolyA-containing RNA was fractionated on 1.6% agarose gels and transferred to DBM-paper [19] which was then hybridized with recombinant DNA fragments labelled in vitro by nick-translation [20].

Southern blot analysis

Total yeast DNA was digested with EcoR I or Hind III. The resulting fragments were electrophoresed on agarose gels, transferred to nitro-cellulose and hybridized with nick-translated fragments for 18 h at 60°C in 2xSSC, 0.2% BSA, 0.2% Ficoll, 0.2% PVP, 0.1% SDS, 1 mM EDTA, 0.1% sodium pyrophosphate and 10 µg/ml low-molecular-weight denatured bacterial DNA. Strips were washed in 2xSSC at 60°C (2x30 min) and then in 0.3xSSC at 50°C (2x20 min).

DNA sequence analysis

DNA sequence analysis was performed using the chain termination method [21]. Single-stranded templates were obtained by transforming JM101 cells with recombinant bacteriophage M13 RF DNA [22]. The M13 vectors used were M13 mp7 [23] or M13 mp8 or M13 mp9 (J. Messing, personal communication). In some cases the sequencing procedure according to Maxam and Gilbert [24] was followed. Each sequence determination was performed at least twice.

Nomenclature of yeast ribosomal proteins

Ribosomal proteins indicated by S and L are numbered according to the standard numbering system proposed in Ref. 35. Proteins indicated by rp are numbered according to Ref. 36.

RESULTS AND DISCUSSION

We have previously described the isolation from a colony bank of Hind III-generated yeast DNA fragments of a number of recombinant DNAs carrying genes coding for ribosomal proteins [11]. Analysis under the electron microscope of R-loop structures formed between various DNA fragments and yeast mRNA suggested that a number of the cloned ribosomal protein genes contain an intervening sequence [14]. The recombinant DNAs were used to probe blots of RNA isolated from the temperature sensitive ma2 mutant [25] at both 23°C and 36°C and separated on agarose gels. The results

of some of these hybridizations are shown in Figure 1. Hybridization with DNA from pBMCY44 (carrying the genes for S16A and rp28), pBMCY113 (carrying the gene for S10) and pBMCY138 (carrying the gene for L25) [11,14], respectively, revealed the accumulation at 36°C (R, restrictive temperature) of complementary RNA molecules several hundreds of nucleotides longer than the mature mRNAs (P, permissive temperature). Obviously, this temperature sensitive strain fails to process precursor mRNAs for S16A, rp28, S10 and L25 at the restrictive temperature. These results are in agreement with data reported by Rosbash *et al.* [26] and Fried *et al.* [27]. No larger transcript is observed in the case of pBMCY89, which carries the gene for L16. This finding is in accordance with the observation that the synthesis of L16 continues at the restrictive temperature [28,29]; apparently no splicing is involved in the maturation of the L16 mRNA.

From the experiments described above we tentatively concluded that the genes for ribosomal proteins S16A, rp28, S10 and L25 are interrupted by an intron of 300-400 nucleotides.

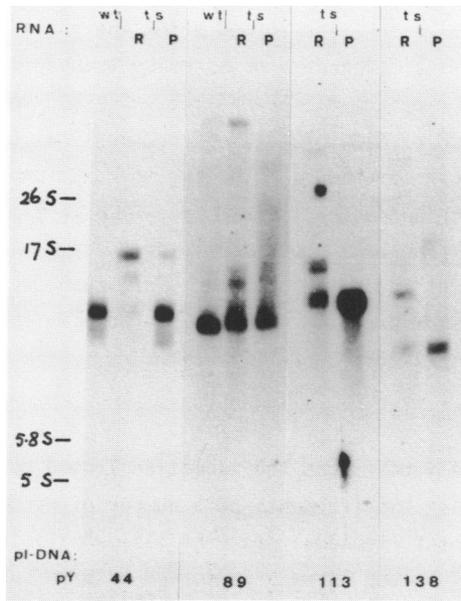


Fig. 1. Northern blot analysis of transcripts from ribosomal protein genes. Recombinant DNAs were labelled *in vitro* by nick-translation and hybridized with RNA isolated from *rna2* cells at 23°C (P, permissive temperature) and 36°C (R, restrictive temperature). wt: RNA isolated from *S. carlsbergensis*.

The structure of the gene coding for S10 was determined by nucleotide sequence analysis. To this end part of the recombinant DNA pBMCY113 was characterized further by cleavage with several restriction endonucleases. The resulting fine map within the physical map of the insert reported previously [14], is depicted in Figure 2. The position of the S10 gene was established by R-loop analysis under the electron microscope [14]. The following subfragments were cloned into the bacteriophage M13 (see Fig. 2):

- a 440 nucleotide fragment generated by digestion with EcoR I (designated as EE440) was cloned into vector mp7;
- a 395 nucleotide fragment generated by double digestion with EcoR I and Sau 3A (designated as SE395) was cloned into the vectors mp8 and mp9;
- a 880 nucleotide fragment generated by digestion with Sau 3A (designated as SS880) was cloned into the vector mp9.

Sequence analysis was performed according to the strategy outlined in Figure 2. In order to identify the yeast S10 coding sequence we took advantage of the suggested homology between S10 and rat liver ribosomal protein S6 [8-10]. Within the EE440 fragment we found a nucleotide sequence from which an aminoacid sequence could be deduced showing strong similarity with the N-terminal part [12] of the corresponding rat liver protein (cf. Figs 3 and 4). However the codons for the N-terminal methionine and lysine residues were not found to be adjacent to the remaining N-terminal part of the sequence (aminoacid 3-33). Since on basis of the above mentioned results we expected an intron within the S10 gene of about 400 nucleotides, we also determined the nucleotide sequence in the large EcoR I-Hind III fragment

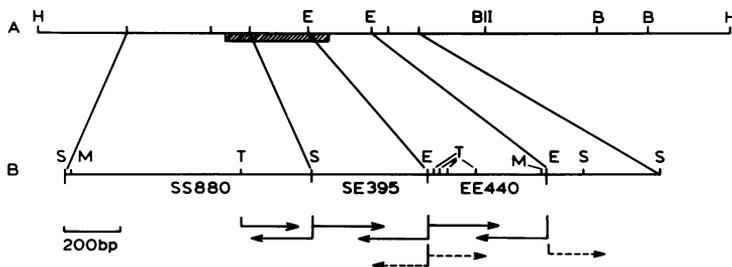


Fig. 2. Map of the insert of pBMCY113 and extent of DNA sequencing. The position of the restriction sites indicated in A as well as the location of the S10 gene were published in a previous paper [14]. In B the sequencing strategy is shown. → : direction and extent of sequencing according to the chain termination method; ---→: direction and extent of sequencing using the chemical procedure. H = Hind III, E = EcoR I, B = Bam HI, B I I = Bgl I, S = Sau 3A, T = Taq I, M = Msp I.

```

-140          -120          -100          -80          -60          -40
TTGTTTGTGGGCACGGTCAATGAAAGCTCTGGGGAAAAAATTCATCTTATGAAAAAGTATCTTAATTTGAATAGTGTGCCCTTCAAAACATTTGGAAAGTCATTAAT

-20          -1          1          2          20          40          60
GAAACTACCGATGTGAATAGACGACTGAGCCATC  ATG  AAG  GTATGTAATATCACCCAAATCCTTTGAATTTCAATGAATCAGCAGCGCTAACCCGGCTGTTCTGCAC
      met  lys
      80          100          120          140          160          180
TGTTTGATAAACGTATACCCACAATTAGTGCATATAATAAAAAATCTCAAGAACCCGTTGTTTAAACGAGATAATCCCTCTAATATACAGTACCAGACACTAGGAA

      200          220          240          260          280
AATATCTCGCTAAGTCAAATTAAGGAAATGAAAAGGAATTTACGAAAAGGGTTTAGGAATATCAATGAAAATAAGAAAACCTGTAACGGAAAGAAAGACAGCAGGGATT

      300          320          340          3          10
CGTTGGAATTTGTCGATATTGGCTTCGGCAACTTTACTAACAAATGGTATTATTATAACAG  TTG  AAC  ATT  TCT  TAC  CCA  GTC  AAC  GGG  TCT  CAA  AAG
      20          30          40
ACC  TTC  GAA  ATT  GAT  GAT  GAA  CAC  CGT  ATT  CGT  GTT  TTC  TTC  GAC  AAG  AGA  ATC  GGT  CAA  GAA  GTC  GAT  GGT  GAA  GCC  GTT  GGT
thr phe glu ile asp asp glu his arg ile arg val phe phe asp lys arg ile gly gln glu val asp gly glu ala val gly
      50          60          70
GAT  GAA  TTC  AAG  GGC  TAC  GTC  TTC  AAG  ATC  TCT  GGT  GGT  AAC  GAC  AAA  CAA  GGT  TTC  CCA  ATG  AAG  CAA  GGT  GTT  TTG  TTG  CCA
asp glu phe lys gly tyr val phe lys ile ser gly gly asn asp lys gln gly phe pro met lys gln gly val leu leu pro
      80          90
ACT  AGA  ATC  AAG  TTG  TTG  TTG  ACC  AAG  AAC  GTT  TCT  TGT  TAC  AGA  CCA  AGA  CGT  GAT  GGT  GAA  AGA  AAG  AGA  AAG  TCC  GTC  AGA
thr arg ile lys leu leu leu thr lys asn val ser cys tyr arg pro arg arg asp gly glu arg lys arg lys ser val arg
      100          110          120
GGT  GCC  ATT  GGT  GGT  CCA  GAT  TTG  GCT  TTC  GCT  ATT  GTC  AAG  AAG  GGT  GAA  CAA  GAA  TTG  GAA  GGT  CTA  ACT  GAC
gly ala ile val gly pro asp leu ala val leu ala leu val ile val lys lys gly glu gln glu leu glu gly leu thr asp
      130          140          150
ACT  ACT  GTT  CCA  AAG  AGA  TTG  GGT  CCA  AAG  AGA  GCT  AAC  AAC  ATC  AGA  AAG  TTC  TTC  GGT  TTG  TCC  AAG  GAA  GAT  GAC  GTT  CGT
thr thr val pro lys arg leu gly pro lys arg ala asn asn ile arg lys phe phe gly leu ser lys leu ser lys glu asp asp val arg
      160          170          180
GAT  TTC  GTC  ATC  AGA  AGA  GAA  GTC  ACC  AAG  GGT  GAA  AAG  ACT  TAC  ACC  AAG  GCT  CCA  AAG  ATC  CAA  AGA  TTG  GTT  ACT  TCT  CAA
asp phe val ile arg arg glu val thr lys gly glu lys thr tyr thr lys ala pro lys ile gln arg leu val thr ser gln
      190          200          210
AGA  TTG  CAA  AGA  AAG  AGA  CAC  CAA  AGA  GCT  TTG  AAG  GTC  AGA  AAC  GCT  CAA  GCT  CAA  AGA  GAA  GCT  GCT  GCC  GAA  TAC  GCT  CAA
arg leu gln arg lys arg his gln arg ala leu lys val arg asn ala gln ala gln arg glu ala ala gln tyr ala gln
      220          230          240          250          260          270          280
TTG  TTG  GCT  AAG  AGA  TTG  TCT  GAA  AGA  AAG  GCT  GAA  AAG  GCC  GAA  ATC  AGA  AAG  AGA  AGA  GCT  TCT  TCT  TTG  AAG  GCT  TAA  GTA
leu leu ala lys arg leu ser glu arg lys ala glu lys ala glu ile arg lys arg arg ala ser ser leu lys ala end
      1070          1090          1110          1130          1150
TGCTTGGACCTGCATAATTTTATACATTTTCCCTTTAGATATATGTACTTTTGGCTTAATTTAATATAATAACTATAACTTACG
    
```

Fig. 3. DNA sequence of the gene for ribosomal protein S10 and amino acid sequence of the S10 protein. Within the protein coding sequence the codons are numbered. Outside this region the nucleotides are numbered with italics starting from the first nucleotide of the initiation codon.

(see Fig. 2) from the EcoR I site to the right. We looked for the sequence ATGAAGGT which would comprise both the N-terminal codons and the consensus 5'-splice junction [30]. The searched octanucleotide was found at a position 24 to 31 nucleotides upstream from this EcoR I site. The complete nucleotide sequence of the S10 gene is given in Figure 3. The length of the intron is 352 nucleotides which is compatible with the extra length of the precursor mRNA as compared to mature mRNA (cf. Fig. 1). The sequences at

	1	5	10
yeast S10	met - lys - leu - asn - ile - ser - tyr - pro		
rat liver S6	met - lys - leu - asn - ile - (ser) - phe - pro		
	15	20	
yeast S10	gly - ser - gln - lys thr - phe glu ile asp - asp - glu -		
rat liver S6	gly - (ser) - gln - lys leu - leu glu val asp - asp - glu -		
	25	30	
yeast S10	his - arg ile arg val phe - phe - asp - lys arg - ile -		
rat liver S6	(arg) - lys leu (arg) X phe - tyr - glu - lys X - met -		

Fig. 4. Comparison of the primary structure of the amino-terminal part of yeast ribosomal protein S10 and rat liver ribosomal protein S6. Both identical (—) and functionally identical (---) residues are boxed.

Thus, strong evidence exists that yeast ribosomal protein S10 is the structural and functional analogue of mammalian S6. The large degree of homology between the primary structures of both the N- and C-terminal parts of the molecules suggest a conservation reflecting the functional role of this ribosomal protein in eukaryotic protein synthesis.

The SE395 fragment which contains only S10 coding sequences was used to probe restriction digests of total yeast DNA to determine the copy number of the S10 gene on the genome. The results shown in Figure 6 demonstrate the occurrence of two copies of the gene for S10. The pertinent copies are located on Hind III-generated fragments of 4.8 and 0.9 kb (lane a) and on EcoR I-generated fragments of 8.2 and 7.5 kb (lane b). The corresponding homologous fragments of DNA from *S. carlsbergensis* differ slightly from those of *S. cerevisiae* (lanes c and d). This strain difference is confirmed by comparison of our pBMCY113 with the recombinant phage C6 isolated by Fried et al. [27].

Further studies are necessary to unravel whether both gene copies are expressed during exponential growth of yeast cells. We are currently determining the 5' and 3' ends of the mature mRNA as well as the sequence of the flanking regions. Comparison of the nucleotide sequence upstream from the initiation codons of the genes for S10 and rp 51 (M. Rosbash, personal

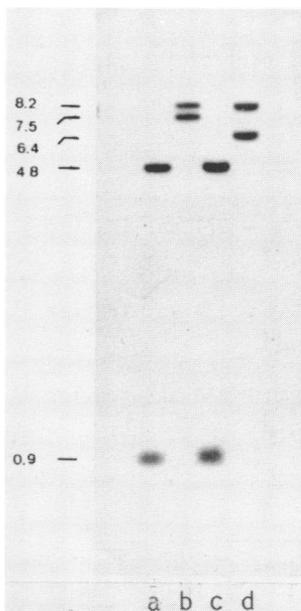


Fig. 6. Copy number analysis of ribosomal protein S10 gene. DNA fragment SE395 was labelled *in vitro* by nick-translation and hybridized with both Hind III (lanes a and c) and EcoR I (lanes b and d) digests of DNA from *S. carlsbergensis* (lanes a and b) and *S. cerevisiae* (lanes c and d).

communication) revealed a homology block of 10 nucleotides - CATTAAAT_C^TGA - at positions -33 to -43 and -45 to -55 respectively. The determination of the 5' flanking sequences of other ribosomal protein genes has to be awaited however, before any solid conclusion can be drawn with respect to the significance of this block. Concerning the codon usage within the S10 gene it is noteworthy that a high preference is observed. Actually almost the same codon bias is seen as has been described for strongly expressed genes like alcoholdehydrogenase and glyceraldehyde-3-phosphate dehydrogenase [33]. An exception can be made for the Arg-triplet CGU which was never used in the synthesis of the other proteins [33]. Preferred triplets tend to be homologous to the anticodons of the major yeast isoacceptor tRNAs [33], which allows rapid and efficient translation of the respective mRNAs. Ribosome formation requires a continuously high level of ribosomal protein synthesis. Thus, an effective though regulated translation of ribosomal protein mRNAs is necessary.

ACKNOWLEDGEMENTS

The authors are indebted to Mrs L.H.P. Caro and Mrs V.C.H.F. de Regt for skillful technical assistance during part of the work.

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.).

*Present address: Gaubius Institute, Division of Health Research TNO, Leiden, The Netherlands

REFERENCES

1. McConkey, E.H., Bielka, H., Gordon, J., Lastick, S.M., Lin, A., Ogata, K., Reboud, J-P., Traugh, J.A., Traut, R.R., Warner, J.R., Welfle, M. and Wool, I.G. (1979) *Mol. Gen. Genet.* 169, 1-6
2. Möller, W. (1974) in *Ribosomes* (Nomura, M., Tissières, A. and Lengyel, P., eds.), pp. 711-731, Cold Spring Harbor Laboratory, New York
3. Vidalis, F.J., Sanchez-Madrid, F. and Ballesta, J.P.G. (1981) *Biochim. Biophys. Acta* 656, 28-35
4. Sanchez-Madrid, F., Reyes, R., Conde, P. and Ballesta, J.P.G. (1979) *Eur. J. Biochem.* 98, 409-416
5. Itoh, T., Higo, K., Otaka, E. and Osawa, S. (1980) in *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes* (Osawa, S., Ozeki, H., Uchida, H. and Yura, T., eds.) University of Tokyo Press, Tokyo
6. Leader, P. (1980) in *Molecular Aspects of Cellular Regulation* (Cohen, P., ed.) Vol. 1, pp. 203-233, Elsevier/North-Holland Biomedical Press, Amsterdam
7. Gordon, J., Nielsen, P., Manchester, K.L., Towbin, H., Jimenez de Asua, L. and Thomas, G. (1982) *Curr. Top. Cell Regul.*, in the press
8. Zinker, S. and Warner, J. (1976) *J. Biol. Chem.* 251, 1799-1807

9. Hébert, J., Pierre, M. and Loeb, J.E. (1977) *Eur. J. Biochem.* 72, 167-174
10. Kruiswijk, T., de Heij, J.T. and Planta, R.J. (1978) *Biochem. J.* 175, 213-219
11. Bollen, G.H.P.M., Cohen, L.H., Mager, W.H., Klaassen, A.W. and Planta, R.J. (1981) *Gene* 14, 279-289
12. Wittmann-Liebold, B., Geissler, A.W., Lin, A. and Wool, I.G. (1979) *J. Supramol. Struct.* 12, 425-433
13. Wettenhall, R.E.H. and Cohen, P. (1982) *FEBS Lett.* 140, 263-269
14. Bollen, G.H.P.M., Molenaar, C.M.Th., Cohen, L.H., Van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.J. (1982) *Gene* 18, 29-38
15. Clewell, D.B. and Helinski, D.R. (1969) *Proc. Nat. Acad. Sci. USA* 62, 1159-1166
16. Bollen, G.H.P.M., Mager, W.H., Jennekens, L.W. and Planta, R.J. (1980) *Eur. J. Biochem.* 105, 75-80
17. Retèl, J. and Planta, R.J. (1972) *Biochim. Biophys. Acta* 281, 299-309
18. Smith, H.O. and Birnstiel, M.L. (1976) *Nucl. Acids Res.* 3, 2387-2398
19. Stark, G.R. and Williams, J.G. (1979) *Nucl. Acids Res.* 6, 195-203
20. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Nat. Acad. Sci. USA* 74, 5463-5467
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. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucl. Acids Res.* 9, 309-321
24. Maxam, A.M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. USA* 74, 560-564
25. Hartwell, L.H., McLaughlin, C.S. and Warner, J.R. (1970) *Mol. Gen. Genet.* 109, 42-56
26. Rosbash, M., Harris, P.K.W., Woolford, J. and Teem, J.L. (1981) *Cell* 24, 679-686
27. Fried, H.M., Pearson, N.J., Kim, C.H. and Warner, J.R. (1981) *J. Biol. Chem.* 256, 10176-10181
28. Gorenstein, C. and Warner, J.R. (1976) *Proc. Nat. Acad. Sci. USA* 73, 1547-1551
29. Warner, J.R. and Gorenstein, C. (1977) *Cell* 11, 201-212
30. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) *Nature* 283, 220-224
31. Wettenhall, R.E.H. and Howlett, G.J. (1979) *J. Biol. Chem.* 254, 9317-9323
32. Kalthoff, H., Darner, D., Towbin, H., Gordon, J. Amons, R., Möller, W. and Richter, D. (1982) *Eur. J. Biochem.* 122, 439-443
33. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3026-3031
34. Ng, R. and Abelson, J. (1980) *Proc. Nat. Acad. Sci. USA* 77, 3912-3916
35. Bollen, G.H.P.M., Mager, W.H. and Planta, R.J. (1981) *Mol. Biol. Rep.* 8, 37-44
36. Warner, J.R. and Gorenstein, C. (1978) *Methods in Cell Biol.* 20, 45-60