Current Genetics (1984) 9:47–52

Current Genetics

The primary structure of a gene encoding yeast ribosomal protein L34

Peter J. Schaap, Constance M. T. Molenaar, Willem H. Mager, and Rudi J. Planta

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

Summary. Sequence analysis revealed that a gene coding for yeast ribosomal protein L34 comprises an amino acid coding region of 339 nucleotides which is interrupted by an intron after the 19th codon. Like for other yeast ribosomal protein genes analyzed thus far a strong codon bias was observed. The flanking and intervening sequences of this gene encoding L34 show several elements that are conserved in a number of split ribosomal protein genes in yeast. Northern blot analysis using an intron-specific probe demonstrated that the sequenced gene copy coding for L34 is transcribed in vivo.

Key words: Yeast – Ribosomal protein gene – Sequence analysis – Northern blot

Introduction

Ribosome biogenesis in yeast occurs continuously during the cell cycle, and the rate of ribosome formation is directly related to cellular growth rate (see Warner 1982; Planta and Mager 1982 for reviews). Under almost all growth conditions a perfect balance is maintained in the production of all ribosomal constituents. The molecular mechanisms responsible for: i) regulation of ribosome formation under various growth conditions, ii) the balance between rRNA and ribosomal protein synthesis and iii) coordination of ribosomal protein synthesis, are largely unknown so far. With the aim to study the regulation of ribosomal protein synthesis in yeast a large number of ribosomal protein genes have been isolated by molecular cloning (Woolford et al. 1979; Fried et al. 1981; Bollen et al. 1981; Fried and Warner 1981, 1982; Schultz and Friesen 1983; Laskin and Woolford 1983). In general, yeast ribosomal protein genes are not clustered and occur duplicated on the yeast genome (Molenaar 1984). In search of common structural elements that may be involved in the coordinate expression of these genes we undertook a detailed comparative analysis of the structure of several ribosomal protein genes and their flanking sequences. This paper deals with the primary structure of a gene encoding the large subunit protein L34. In addition we present evidence that the sequenced copy of the L34 gene is expressed in the vegetatively growing yeast cell.

Offprint requests to: W. H. Mager

Materials and methods

Isolation of nucleic acids. Recombinant plasmids were purified from Triton-treated bacterial spheroplasts by CsCl-ethidium bromide density gradient centrifugation (Clewell and Helinski 1969).

PolyA-containing RNA was isolated from the yeast *Saccharomyces cerevisiae rna2* (ts 368^-) according to a procedure described previously (Bollen et al. 1980; Molenaar 1984).

Restriction site mapping. Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Laboratories and used as recommended by the supplier.

DNA sequence analysis. DNA sequence analysis was performed using the chain termination method (Sanger et al. 1977). Single-stranded templates were obtained by transforming JM101 cells with recombinant bacteriophage M13RF DNA (Sanger et al. 1980). The M13 vectors used were M13 mp8, M13 mp9 (Messing 1982) or M13 mp11 (Messing and Vieira 1982).

Northern blot analysis. PolyA-containing RNA was fractionated on 1.6% agarose gels after denaturation in 1 M glyoxal and 50% (v/v) dimethylsulphoxide (McMaster and Carmichael 1977), transferred onto nitrocellulose (Thomas 1980) and hybridized with radioactively labelled DNA probes.

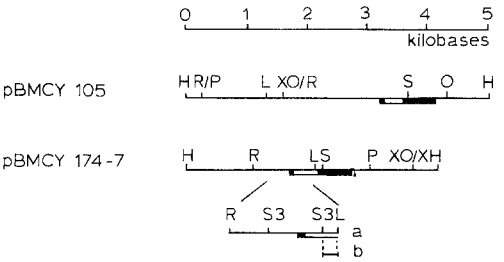

Fig. 1. Map of the inserts of clone pBMCY174-7 (Molenaar 1984) and clone pBMCY105 (Bollen et al. 1981). H = HindIII, R = EcoRI, P = PstI, L = SalI, S = SacI, O = BglI, XO = XhoI, X = XbaI, S3 = Sau3A. *Black* and *open boxes* represent exon and intron sequences respectively. The bars indicated by a and b represent the DNA fragments used for the hybridization experiments described in Fig. 4

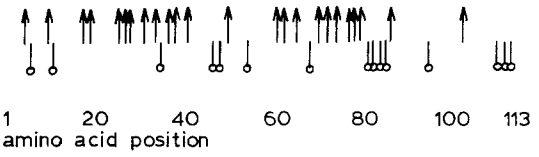

Fig. 3. Distribution of basic (Lys, Arg) and acidic (Glu, Asp) amino acids over the L34 ribosomal protein molecule. ↑: Arg or Lys; ↓: Glu or Asp. The net charge of the protein is +9

Preparation of radioactively labelled DNA probes. Single-stranded recombinant bacteriophage M13 DNA molecules were isolated and subsequently used as a template for Klenow polymerase directed synthesis in the presence of (α-^{32}P)dATP (Amersham). A universal primer complementary to the 3' region flanking the multiple cloning site of M13 was used. Occasionally from the newly-synthesized DNA molecule a fragment was isolated by cutting with proper restriction enzymes and subsequent fractionation of the digest on a 5% polyacrylamide gel.

Results and discussion

The gene encoding the large subunit protein L34 is duplicated on the yeast genome (Molenaar 1984). Both gene copies were selected from a bank of HindIII-generated DNA fragments as described previously (Bollen et al. 1981) and the location of the genes on the cloned fragments was determined by electron microscopic R-loop analysis (Molenaar 1984; see Fig. 1).

We sequenced one copy of the L34 gene (indicated as gene copy-1; on clone pBMCY174-7) as well as part of its flanking sequences taking advantage of the restriction enzyme sites indicated in Fig. 1. The nucleotide sequence of gene copy-2 (on clone pBMCY105) was analyzed only partially. As can be concluded from the data presented in Fig. 2, the gene coding for L34 comprises an amino acid coding sequence of 339 nucleotides, interrupted by an intron of 421 nucleotides (The intron within gene copy-2 is 357 nucleotides long; result not shown). The apparent molecular weight of the L34 protein calculated on basis of the deduced amino acid sequence amounts 12.936 D. In Fig. 3 we depicted the distribution of basic and acidic amino acids over the protein molecule. It turned out that the charged residues are more or less clustered possibly reflecting functional domains of this ribosomal protein. The amino acid coding region of the gene demonstrates a highly preferent

Table 1. Codon usage for the genes coding for ribosomal proteins L34 (this paper) and S10 (Leer et al. 1982) and several other highly expressed yeast genes (Y; Ammerer et al. 1981)

		L34	S10	Y			L34	S10	Y			L34	S10	Y			L34	S10	Y
Phe	UUU	–	–	5	Ser	UCU	1	8	79	Tyr	UAU	–	–	2	Cys	UGU	–	1	20
Phe	UUC	3	9	63	Ser	UCC	3	2	73	Tyr	UAC	3	5	73	Cys	UGC	–	–	2
Leu	UUA	3	–	14	Ser	UCA	–	–	1										
Leu	UUG	6	19	148	Ser	UCG	–	–	1						Trp	UGG	1	–	26
Leu	CUU	–	–	–	Pro	CCU	1	–	6	His	CAU	–	–	3	Arg	CGU	2	4	2
Leu	CUC	–	–	1	Pro	CCC	–	–	1	His	CAC	3	2	56	Arg	CGC	–	–	–
Leu	CUA	2	1	6	Pro	CCA	3	8	71	Gln	CAA	2	12	44	Arg	CGA	–	–	–
Leu	CUG	–	–	–	Pro	CCG	–	–	–	Gln	CAG	–	1	–	Arg	CGG	–	–	–
Ile	AUU	3	5	58	Thr	ACU	3	6	49	Asn	AAU	–	–	3	Ser	AGU	–	–	1
Ile	AUC	1	7	66	Thr	ACC	1	4	66	Asn	AAC	4	7	86	Ser	AGC	–	–	1
Ile	AUA	–	–	1	Thr	ACA	–	–	–	Lys	AAA	2	1	21	Arg	AGA	9	24	63
Met	AUG	2	2	39	Thr	ACG	–	–	–	Lys	AAG	11	27	151	Arg	AGG	–	–	1
Val	GUU	6	8	112	Ala	GCU	6	14	182	Asp	GAU	3	8	37	Gly	GGU	6	15	164
Val	GUC	7	10	93	Ala	GCC	4	4	62	Asp	GAC	2	4	100	Gly	GGC	–	1	11
Val	GUA	–	–	3	Ala	GCA	–	–	3	Glu	GAA	10	17	67	Gly	GGA	1	–	1
Val	GUG	–	–	1	Ala	GCG	–	–	4	Glu	GAG	–	–	3	Gly	GGG	–	1	1

```
-570                                                                    -500
  GAATTCTATA GCACTTGTCG CCGGAATGGA GATGCTTATT CTTGTGTTCT ACATTCATGA AGTGCTGATA CTATATTCTG

  TTTTCCGACC CTCTCCCCTT CTGGGCCCTT TTTCCATTAT TCACAATATT TTTCATACCT TTACCTCCGT ACACCAATCT
          -400
  TTATTTTCAC ACCCATACCT TTGTATTCTG AAATTGGGAA ATGATTTGTG GTGCTTTTTT TTTTTGGACA ATCTGCCTTC
                                  -300
  GTGGACTGTT TTGAAACGCT AGGCGACAGT TTTGCAACGG AAGGAAGTGG GAGAGCCCAG AAAGGCTTTC TCTCGATGGT
                                                          -200
  GGATCATTCT TGCGCAGGCG GAGGATGGGA AAGCTCCGTT CAGGTTTGCG GCTTTTCTTT CTGGCATTTT CGTTCTCCCA
                                                                        -100
  ACTGCGCAGG CAGACATAGC TTGACTTTAC TCATGCTCGC ATTAGGCGGG TTGAATCTCA TAGCTTGAAA GATAGTATTG

  AATTATATTG AAGATTTTAT AACTAGTGAT AATTTCATTC AAGTTTTAAA GAAAGAATTA ATAAGCAAAC AGAACTCAAT

  CAAAGGAAAA  ATG GCC GGT TTG AAA GAC GTT GTC ACT CGT GAA TAC ACC ATT AAC TTG CAC AAA AGA
              met ala gly leu lys asp val val thr arg glu tyr thr ile asn leu his lys arg
               1                                  10

  GTATGTGAAA AAATAAAAAT ATAAGTCCCA AGGGCTTGTC ATCTGAATGG TTAATTAAGA AGATATTTCT ATAGTCAAGG

  AACACAATAT CAAGAACTTG TACCGCGGAA ACCTATCAAC AAAGATCATT GATCAAGGTA TATTTTATAA TACGGAATTT

  GAAGACCATA ATCAATATTA TAACATCCAA GAGGGGAAAG ATACATTTAG GCTGGACTTA AAGAGCGATG AAGACAGAAA

  TAGTTATCTT AAATAGTGGA ATCCTAGATT TTAAGTCGAC AATTATCCAT ATCCCAAGTT TGACTAAATA TGAATCAGAA

  CAAATTGGGT CTCCAATATA CAAAACCATT CATCAAGATT ACACTAATAT TAAATTGAAC AATCGTTACT AACAAAACAT

  TTACCATTTT ATTTTTAATA G TTG CAC GGT GTC TCC TTC AAG AAG AGA GCT CCA AGA GCT GTC AAG GAA
                        leu his gly val ser phe lys lys arg ala pro arg ala val lys glu
                        20                                  30

  ATT AAG AAG TTC GCC AAG TTA CAC ATG GGT ACT GAT GAT GTC CGT CTA GCT CCA GAA TTG AAC CAA
  ile lys lys phe ala lys leu his met gly thr asp asp val arg leu ala pro glu leu asn gln
                  40                                  50

  GCT ATC TGG AAG AGA GGT GTC AAG GGT GTT GAA TAC AGA TTA AGA TTG AGA ATT TCC AGA AAG AGA
  ala ile trp lys arg gly val lys gly val glu tyr arg leu arg leu arg ile ser arg lys arg
          60                                  70

  AAC GAA GAA GAA GAC GCC AAG AAC CCA TTG TTC TCC TAC GTT GAA CCT GTC TTA GTT GCT TCT GCC
  asn glu glu glu asp ala lys asn pro leu phe ser tyr val glu pro val leu val ala ser ala
  80                                  90                                  100

  AAG GGT CTA CAA ACT GTT GTT GTC GAA GAA GAT GCT TAA     ATCAATTAAA TAAAAATAAT ACTCATTTTA
  lys gly leu gln thr val val val glu glu asp ala end
                                  110
                    +50                                                 +100
  TATTAATAAA TATAAAGTAA ACATTTTCGA TTTGTATAAA TCTCTTGTAC TCAAGCTAAG ACTTCTGATC ATCGCTCTGA

  ATGATAGGAA ATATAT
```

Fig. 2. DNA sequence of the gene for ribosomal protein L34 and amino acid sequence of the L34 protein. The nucleotides are numbered starting from the first nucleotide of the initiation codon

codon usage as was also observed for other frequently transcribed genes in yeast (see Table 1).

The genes encoding L34 are split by an intron after the 19th codon. Their general structure is in agreement with the usual pattern of coding and intervening sequences observed for all split yeast ribosomal protein genes as is illustrated in Table 2. The intervening sequences are bounded by two conserved splice sites (viz. GTATGT at the 5′-end and TAG at the 3′-end) and moreover contain the internal conserved sequence TACTAACA. This common box has been found in all yeast introns analyzed so far (see Table 3) and appears

Table 2. General structure of yeast ribosomal protein genes

	5′ exon	intron	3′ exon
S10[a]	ATG – 1 codon	← 352 n →	234 codons – TAA
S16A-1[b]	ATG – 5 codons	← 390 n →	138 codons – TAA
S16A-2[b]	ATG – 5 codons	← 551 n →	138 codons – TAA
rp28-1[b]	ATG – 36 codons	← 447 n →	149 codons – TAA
rp28-2[b]	ATG – 36 codons	← 432 n →	149 codons – TAA
rp51[c]	ATG – –	← 398 n →	135 codons – TAA
L29[d]	ATG – 15 codons	← 510 n →	133 codons – TAA
L34-1	ATG – 18 codons	← 421 n →	94 codons – TAA
L34-2	ATG – 18 codons	← 349 n →	94 codons – TAA

[a] Leer et al. (1982)
[b] Molenaar and Pearson, manuscript in preparation
[c] Teem and Rosbash (1983)
[d] Käufer et al. (1983)

Table 3. Intron sequences within yeast ribosomal protein genes

S10[a]	↓GTATGT –	316 n	–	TTTACTAACA	–	10 n	– TTTATAACAG↓
S16A-1[b]	GTACGT –	341 n	–	TTTACTAACA	–	23 n	– TTTTCTACAG
S16A-2[b]	GTACGT –	503 n	–	TTTACTAACA	–	22 n	– TTTCAATTAG
rp28-1[b]	GTATGT –	410 n	–	GTTACTAACA	–	12 n	– TTTTTTTTAG
rp28-2[b]	GTATGT –	378 n	–	TTTACTAACA	–	28 n	– TTAATCACAG
rp51[c]	GTATGT –	326 n	–	TATACTAACA	–	46 n	– ATTTTAATAG
L29[d]	GTATGT –	453 n	–	TTTACTAACG	–	31 n	– TTTTGTACAG
L34-1	GTATGT –	378 n	–	GTTACTAACA	–	17 n	– TTTTTAATAG
L34-2	GTATGT –	305 n	–	TTTACTAACA	–	18 n	– ATTAATATAG
consensus	↓GTA$^{T}_{C}$GT –	x	–	$^{TT}_{GA}$TACTAAC$^{A}_{G}$	–	y	– $^{T}_{A}$T$^{T}_{A}$NNN$^{AC}_{TT}$AG↓

[a] Leer et al. (1982)
[b] Molenaar and Pearson, manuscript in preparation
[c] Teem and Rosbash (1983)
[d] Käufer et al. (1983)

to be required for proper processing of intron-containing transcripts in yeast (Langford and Gallwitz 1983; Pikielny et al. 1983; Langford et al. 1984).

The DNA sequence upstream from the initiation codon of the gene encoding L34 contains several elements that are frequently found in front of yeast genes. The sequence CAAA is repeated near the translation start and may function as a capping site (cf. PyAAPu proposed by Dobson et al. 1982). In addition several TATA-like structures (Benoist and Chambon 1981) can be recognized, e.g. TATAA at –73 and TATAT at –87, which in general are present at a distance of 60–100 nucleotides from the initiation codons of yeast genes. More notably, computer analyses of various yeast ribosomal protein genes have revealed some common sequence elements that have been suggested to play a role in (coordinate regulation of) ribosomal protein gene expression (Teem et al. 1984). Similar boxes present in the 5′-flanking region of the sequenced L34 gene-copy are listed in Table 4 as well as an additional sequence element occurring in the vicinity of HOMOL1 (Leer RJ, unpublished results). Deletion experiments and site-directed mutagenesis are required to elucidate the significance of these sequence elements for expression of the pertinent genes. Finally, in the DNA region flanking the 3′ end of the L34 coding sequence two possible polyadenylation signals (AATAAA at +9 and +35; Fitzgerald and Shenk 1981) are found but no apparent transcription termination site.

The temperature-sensitive yeast mutant *rna2* is defective in splicing at the restrictive temperature (Rosbash et al. 1981; Bromley et al. 1982; Fried et al. 1981; Leer et al. 1982). Probing RNA isolated from mutant cells grown at the permissive as well as the restrictive temperature with DNA fragments containing the gene coding for yeast ribosomal protein L34 therefore permits analysis of precursor and mature transcripts from this gene

Table 4. Conserved sequences upstream from yeast ribosomal protein genes

	common sequence element	L34-1	
HOMOL1:	AACATC$^{TG}_{CA}$T$^{A}_{G}$CA[a]	TACCTCCGTACA	at −429
HOMOL3:	$^{TC}_{GG}$CTTC$^{T}_{C}$T[a]	GCCTTCGT	at −336
	ACCCA$^{T}_{C}$AC$^{C}_{A}$TT$^{T}_{A}$[b]	ACCCATACCTTT	at −400

[a] Teem et al. (1984)
[b] Leer RJ, unpublished results

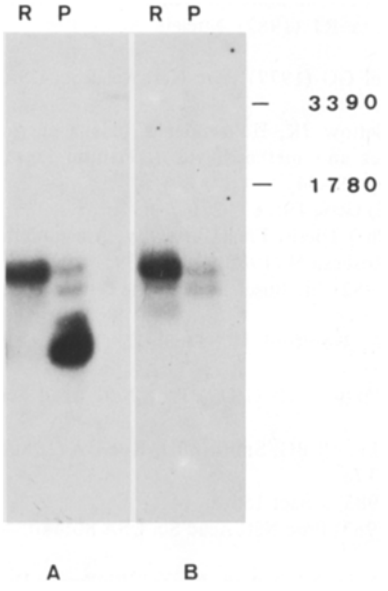

Fig. 4. Northern blot analysis of transcripts from the genes encoding yeast ribosomal protein L34. PolyA-containing RNA was isolated from yeast *rna2* grown at the permissive (23°; P) and restrictive (36 °C; R) temperature. Radioactivity labelled DNA probes were isolated as described in Materials and methods and as indicated in Fig. 1. In A: the probe used was DNA synthesized in vitro from the SalI site within intron to the left (probe a, see Fig. 1); in B: an intron-specific probe (SalI plus Sau3A-generated fragment; probe b, see Fig. 1). The positions of 26S and 17S ribosomal RNA used as size markers are indicated

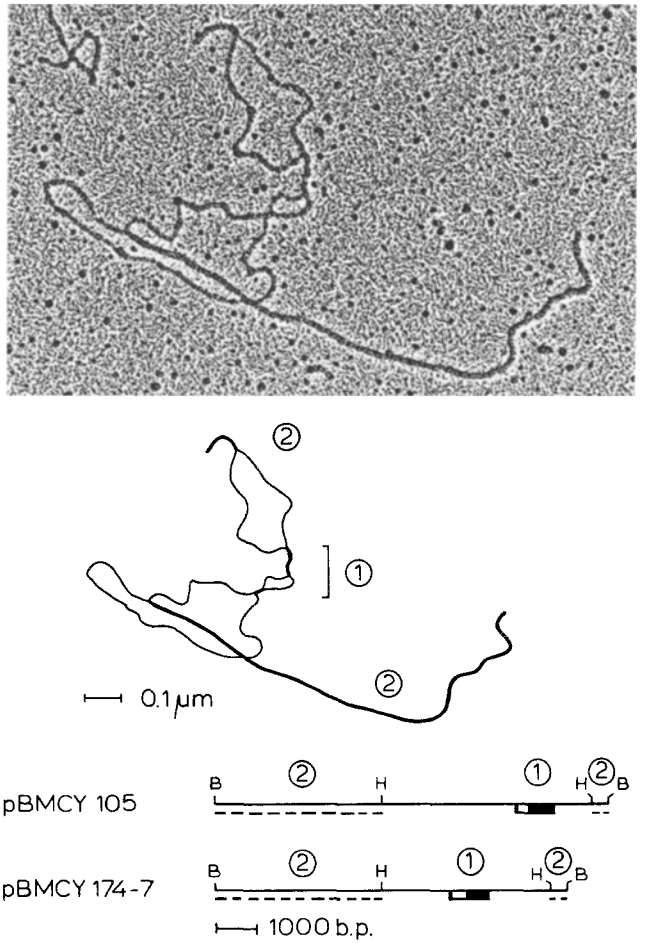

Fig. 5. Heteroduplex analysis of the duplicate genes encoding L34. Recombinant plasmids were linearized by cleavage with BamHI. Only the amino acid coding regions of the L34 genes (1) and regions corresponding to the (identical) vector DNA (pBR322; 2) form stable duplex structures

and enabled us to examine whether the gene is transcribed. We used newly-synthesized DNA flanking the 5′-side of the SalI-site within the intron (see Fig. 1) as a probe in Northern blot hybridization. The results are shown in Fig. 4A.

Hybridization with RNA isolated from mutant cells grown at the permissive temperature (23 °C; P in Fig. 4) revealed the presence of four discernable transcripts: one major which most likely represents a mature L34 mRNA and three minor RNA products. Due to the presence of a large proportion of precursor-specific viz. intron sequences in the probe presumed pre-mRNAs give relatively strong signals. Obviously at the elevated temperature (36 °C; R) the longest RNA accumulates indicating that at least this RNA product is an intron-containing precursor transcript. It cannot be excluded that one more RNA band visible in A represents a precursor mRNA.

To establish whether the L34 gene copy-1 is actually expressed in the vegetatively growing yeast cell we subsequently made use of a DNA probe only consisting of intron-sequences (probe "b" in Fig. 1). This DNA fragment was expected to hybridize only with intron-containing transcripts from gene copy-1 since heteroduplex analysis revealed that the structure of the introns interrupting the duplicate L34 gene copies are strongly different (see Fig. 5). Sequence analysis confirmed the complete divergence of both intervening sequences (result not shown). The intron-specific character of the probe is clear from the hybridization results shown in Fig. 4B: no signals arose at positions into which mature L34 mRNAs migrate. Hybridization only occurred with the two large RNA products, at least the longest of which accumulates at the restrictive temperature (R). These Northern blotting date clearly indicate that the sequenced gene-copy encoding yeast ribosomal protein L34 indeed is transcribed in the vegetatively growing yeast cell. Strikingly, the result obtained with RNA isolated from cells grown at 23 °C (P) suggests that under these growth conditions two size-different pre-mRNAs are transcribed from this gene-copy.

Acknowledgements. The authors are indebted to Drs. R. J. Leer and L. P. Woudt for advice and stimulating discussions, 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

Ammerer G, Hitzeman R, Hagie F, Barta A, Hall BD (1981) In: Walton AG (ed) Recombinant DNA. Elsevier Scientific Publishing Co, Amsterdam, pp 185–197

Benoist C, Chambon P (1981) Nature 290:304–310

Bollen GHPM, Mager WH, Jenneskens LW, Planta RJ (1980) Eur J Biochem 105:75–80

Bollen GHPM, Cohen LH, Mager WH, Klaassen AW, Planta RJ (1981) Gene 14:279–287

Bromley S, Hereford L, Rosbash M (1982) Mol Cell Biol 2: 1205–1211

Clewell DB, Helinski DR (1969) Proc Natl Acad Sci USA 62: 1159–1166

Dobson MJ, Tuite MF, Roberts NA, Kingsman AJ, Kingsman SM, Perkins RE, Conroy SC, Dunbar B, Fothergrill LA (1982) Nucleic Acids Res 10:2625–2637

Fitzgerald M, Shenk T (1981) Cell 24:251–260

Fried HM, Warner JR (1981) Proc Natl Acad Sci USA 78:238–242

Fried HM, Warner JR (1982) Nucleic Acids Res 10:3133–3148

Fried HM, Pearson NJ, Kim CH, Warner JR (1981) J Biol Chem 256:10176–10183

Käufer NF, Fried HM, Schwindinger WT, Jasin M, Warner JR (1983) Nucleic Acids Res 11:3123–3135

Langford CJ, Gallwitz D (1983) Cell 33:519–527

Langford CJ, Klinz FJ, Donath C, Gallwitz D (1984) Cell 36: 645–653

Laskin JC, Woolford JL (1983) Nucleic Acids Res 11:403–420

Leer RJ, Van Raamsdonk-Duin MMC, Molenaar CMT, Cohen LH, Mager WH, Planta RJ (1982) Nucleic Acids Res 10: 5869–5878

McMaster GJ, Carmichael GG (1977) Proc Natl Acad Sci USA 74:4835–4838

Messing J (1982) In: Setlow JK, Hollaender A (eds) Genetic engineering: principles and methods, vol 4. Plenum Press, New York London, pp 19–34

Messing J, Vieira J (1982) Gene 19:269–276

Molenaar CMT (1984) PhD Thesis, Free University, Amsterdam

Pikielny CW, Teem JL, Rosbash M (1983) Cell 34:395–403

Planta RJ, Mager WH (1982) In: Busch H (ed) The cell nucleus, vol XII, pp 213–226

Rosbash M, Harris PKW, Woolford JL, Teem JL (1981) Cell 24:679–686

Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463–5467

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe RA (1980) J Mol Biol 143:161–178

Schultz L, Friesen JD (1983) J Bact 155:8–14

Teem JL, Rosbash M (1983) Proc Natl Acad Sci USA 80:4403–4407

Teem JL, Abovich N, Käufer NF, Schwindinger WF, Warner JR, Levy A, Woolford J, Leer RJ, Van Raamsdonk-Duin MMC, Mager WH, Planta RJ, Schultz L, Friesen JD, Rosbash M (1984) Nucleic Acids Res (in press)

Thomas PS (1980) Proc Natl Acad Sci USA 77:5201–5205

Warner JR (1982) In: Strathern J, Jones E, Broach JR (eds) The molecular biology of the yeast *Saccharomyces*. Cold Spring Harbor, New York Press

Woolford JL, Hereford L, Rosbash M (1979) Cell 18:1247–1259

Communicated by C. P. Hollenberg

Received June 15 / July 25, 1984