Gene, 29 (1984) 113–124
Elsevier

GENE 1024

Structure and function of the yeast *URA3* gene: expression in *Escherichia coli*

(Recombinant DNA; DNA sequence polymorphisms; OMP decarboxylase; promoters; *Saccharomyces cerevisiae*; plasmid and phage vectors)

Mark Rose, Paula Grisafi and David Botstein

Department of Biology, 56-344, Massachusetts Institute of Technology, Cambridge, MA 02139 (U.S.A.) Tel. (617) 253-7240

(Received October 12th, 1983)
(Revision received February 2nd, 1984)
(Accepted February 7th, 1984)

SUMMARY

The expression of the cloned *Saccharomyces cerevisiae URA3* gene in *Escherichia coli* on both plasmid and phage vectors was studied. Isolates of the gene from two different laboratory strains of yeast differ in their ability to be expressed in *E. coli* in the absence of external adjacent promoters of transcription. The DNA sequence of the two genes was determined and revealed several differences in the DNA flanking the structural gene. One base change alters the "Pribnow-box" of an *E. coli* promoter present in the yeast sequences. Three amber alleles of the yeast gene were also cloned from yeast. Two of the alleles could be suppressed in *E. coli* by a tRNA suppressor mutation. One of the amber alleles was determined to be a mutation in the seventh codon of the structural gene, thereby establishing the reading frame and extent of the coding sequence. The initiator codon of the reading frame encoding the *URA3* structural gene is preceded by two other ATG codons in a different reading frame 61 and 79 bp away. The nearer ATG begins an open reading frame that overlaps the structural gene sequences by 17 bp. With the DNA sequence of the *URA3* gene many of the common yeast vector plasmids are now completely known at the level of DNA sequence.

INTRODUCTION

The enzyme OMP decarboxylase (orotidine-5′-phosphate carboxy-lyase, EC 4.1.1.23) is specified in *S. cerevisiae* (yeast) by the *URA3* gene. Expression is regulated at the level of transcription (Bach et al., 1979). Increased intracellular levels of dihydro-orotate lead to increased levels of the enzyme (Lacroute, 1968) and the *URA3* mRNA (Bach et al., 1979). A fragment of yeast DNA which expresses this enzyme in *E. coli* was isolated by complementation of the corresponding *E. coli pyrF* mutation (Bach et al., 1979). The mode of expression of the yeast gene in *E. coli* was obscure, partially due to the lack of knowledge of plasmid promoters and partially due to unexpected changes in the level of functional enzyme expressed under different growth conditions and on different plasmids.

Abbreviations: bp, base pairs; kb, 1000 bp; LB, see MATERIALS AND METHODS, section **b**; m.o.i., multiplicity of infection; OMP, orotidine-5′-phosphate.

To clarify the means by which the gene is expressed in *E. coli*, we have recloned the gene onto a phage λ vector. λ vectors have relatively well characterized promoters and stable copy number in lysogens. We chose to reclone the gene directly from yeast for two reasons. First, the gene was originally isolated on DNA shear fragments (Bach et al., 1979) which contained relatively little yeast DNA sequence flanking the region which expresses OMP decarboxylase. Second, the source of the DNA that had been used for the original isolation of the gene was the genetically heterogeneous and poorly characterized diploid strain + D4 (Petes et al., 1978). The genetics of the *URA3* gene had, on the other hand, been carefully studied in a series of isogenic strains derived from strain FL100. Therefore, we recloned the *URA3* gene from strain FL100 as well as several *ura3*$^-$ nonsense mutants derived from FL100.

Analysis of expression in *E. coli* revealed fundamental differences between the *URA3* genes obtained from FL100 and + D4. DNA sequence differences in the region upstream of the genes can explain the observed differences in their expression in *E. coli*.

MATERIALS AND METHODS

(a) Strains

(1) Bacterial strains

The standard strain employed was DB6566 ($trp_{(am)}$ $lacZ_{(am)}$ *pyrF*: : Mu-1 *hsdR*$^-$ *hsdM*$^+$) described by Bach et al. (1979). DB6660 was derived from DB6566 by P1 transduction to SupF from a strain carrying *suIII*.

(2) Yeast strains

FL100 (*MAT***a**, a wild type) and the three amber mutants *ura3*-3, *ura3*-18 and *ura3*-25 which were derived by mutation from FL100 were all the generous gifts of Francois Lacroute.

(3) Bacteriophages

λgt-7*ara6* (Davis et al., 1980) was provided by R. Davis. λNM816 (λ*plac*5 srIλ3° *imm*21 *c*Its *srI*λ4° *nin*R5 *srI*λ5°) was obtained from N. Murray. λ*imm*434 *c*Its and λ*imm*434 *c*Its p4 were obtained from J. Weil (Fiandt et al., 1971; Weil et al., 1972). λ*J*am6 p*lac*5 *c*I857 was obtained from M. Lichten.

(4) Plasmids

"Clone 1" and "clone 2" are pMB9-derived plasmids carrying shear fragments of yeast DNA containing the *URA3* gene. "Clone 6" is a pBR322-derived plasmid containing the 1.1-kb *URA3* *Hin*dIII fragment. All of the preceding plasmids were described in Bach et al. (1979). pBR322 plasmid DNA was the generous gift of D. Shortle.

(b) Media

E. coli cells were generally grown in LB (Miller, 1972) broth or on LB agar. Cells used for OMP-decarboxylase assays were grown in M9 salts (Miller, 1972) supplemented with 0.2% glucose or maltose and 1% casamino acids (Difco, technical grade, charcoal-filtered before use). λ phage stocks were grown either as plate stocks (Davis et al., 1980) or in liquid (Blattner et al., 1977). Ampicillin was used at a concentration of 100 μg/ml (Bristol Laboratories). Tetracycline (Sigma) was used at a concentration of 15 μg/ml.

(c) Nucleic acid techniques

(1) Enzymes

All restriction enzymes, *E. coli* DNA polymerase I, and T4 DNA ligase were purchased from New England Biolabs and used according to the recommendations of the manufacturer. *Micrococcus luteus* DNA polymerase was purchased from Miles. Bacterial alkaline phosphatase and T4 polynucleotide kinase were purchased from Boehringer-Mannheim.

(2) 3′-end labeling

We routinely used 1–2 units of *M. luteus* DNA polymerase in a 40-μl reaction containing 70 mM Tris · HCl pH 8.0, 7 mM $MgCl_2$, 1 mM β-mercaptoethanol, 25 pmol [α-^{32}P]dNTP and 5–10 pmol of DNA ends.

(3) 5′-end labeling and DNA sequencing

Conditions for phosphatising, phosphorylating DNA ends and sequencing were essentially those of Maxam and Gilbert (1980).

(4) Plaque filter hybridization and nick translation

Hybridizations were done essentially as described by Benton and Davis (1977). Nick-translated probe

DNA was prepared by a modified method of Rigby et al. (1977) as described in Davis et al. (1980).

(d) DNA preparation

(1) λ DNA

Large-scale DNA preparations were made from liquid lysates as described by Blattner et al. (1977). Phage were purified by polyethylene glycol precipitation (Yamamoto et al., 1970) and centrifugation through CsCl step-gradients (Davis et al., 1980). DNA was prepared by phenol extraction, followed by extensive dialysis to remove residual phenol. Small-scale DNA preparations were made by a modification of the method of Davis et al. (1980) for small-scale liquid lysates.

(2) Yeast DNA

High M_r yeast DNA was prepared essentially as described by Petes et al. (1978).

(3) Plasmid DNA

Large-scale plasmid DNA preparations were prepared by the cleared lysate method as described by Clewell and Helinski (1969). Small scale preparations were prepared as previously described (Rose et al., 1981).

(e) DNA transformation, transfection and in vitro λ packaging

E. coli cells were transformed with plasmid DNA according to the method of Mandel and Higa (1970). Transfection with λ DNA was performed according to the modifications of the plasmid transformation procedure described by Davis et al. (1980). Packaging extracts were prepared and used according to the method of Hohn and Murray (1977) using the strains of Sternberg et al. (1977).

(f) OMP-decarboxylase assays

OMP decarboxylase was assayed in *E. coli* by modification of the method of Lieberman et al. (1955). 25 ml of exponentially growing cells were quick-chilled in an ice-salt bath, centrifuged and resuspended in 1 ml of 0.1 M Tris · HCl pH 8.0, 20% (v/v) glycerol and 1 mM dithiothreitol. Expression from λ promoters was assayed at various times after infection of the indicated *E. coli* strains at a multiplicity of infection of 2. Cells were broken open by sonication and extracts were clarified by centrifugation for 5 min in an eppendorf centrifuge. Activity was determined in 1 ml reaction mixtures containing 0.1 M Tris · HCl pH 8.0, 20 μg pyridoxal phosphate, 10 mM $MgCl_2$ and 0.1 mM OMP. Radioactive assays also contained 0.05 μCi of [^{14}C]OMP. $^{14}CO_2$ was collected by absorption into 0.2 ml of 2 N NaOH contained in a plastic reservoir within a 10-ml reaction chamber (Kontes). The reaction was stopped by injection of 0.5 ml of 2 M H_2SO_4 and incubation at 37 °C for 180 min. The reservoir was removed and dropped into 5 ml of Aquasol (New England Nuclear) and counted in the ^{14}C window of a Beckman Scintillation counter. The activity thereby measured was standardized relative to the non-radioactive spectrophotometric assay in which a decrease in absorbance at 285 nm of 1.38 corresponds to a conversion of 1 μmol/ml of OMP to UMP. Protein concentration, when measured, was determined by the method of Bradford (1976) using bovine serum albumin as a standard and reagents purchased from Bio-Rad.

RESULTS

(a) Isolation of the *URA3* gene

A pool of recombinant phages was constructed by ligation of *Eco*RI-digested yeast DNA into λgt7 (Davis, 1980) vector DNA. Approx. 1×10^5 independent plaques of hybrid phages were obtained and screened by plaque filter hybridization (Benton and Davis, 1977) for homology to *URA3* specific DNA using clone 6 (Fig. 1) as probe. Twelve plaques hybridized, of which nine were purified for further study. Two other phages were also obtained by lysogenic complementation of an *E. coli pyrF* mutation using λ *J*am6 p*lac*5 *c*I857 helper phage. Restriction enzyme analysis revealed that all of the phages contained a 13-kb *Eco*RI fragment which is the same size as the band observed in gel-transfer hybridization analysis of genomic DNA (not shown).

The position and orientation of the *URA3 Hin*dIII fragment within the *Eco*RI fragment were determined by both heteroduplex and restriction enzyme

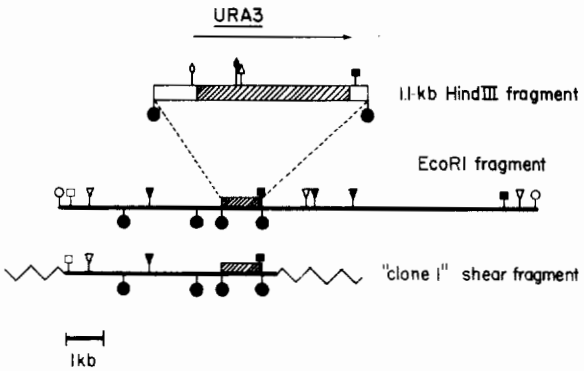

Fig. 1. Restriction maps of various cloned fragments of yeast DNA containing the *URA3* gene. The upper map is of the 1.1-kb *Hin*dIII fragment. The 13-kb *Eco*RI fragment, shown in the middle panel, was cloned from strain FL100. "Clone 1" is a pMB9 plasmid containing a shear fragment of yeast DNA isolated by Bach et al. (1979) from strain + D4. The restriction sites are as follows: ○, *Eco*RI; ●, *Hin*dIII; ▽, *Bam*HI; ▼, *Bgl*II; ⊓, *Sal*I; ■, *Sma*I; ◇, *Pst*I; ◆, *Eco*RV; △, *Nco*I. Some of the restriction sites indicated on the *Hin*dIII fragment (*Pst*I, *Nco*I and *Eco*RV) are not indicated on the other maps.

analysis. These data are summarized in Fig. 1. Nine of the phages containing the *Hin*dIII fragment with the *Pst*I site closer to the long arm (containing the "late" genes) of the λ vector are defined as orientation I phages (λ*URA3*-I). The two remaining phages (λ*URA3*-II) contained the *Pst*I site closer to the right arm of λ ("early" genes). Two of the phages in orientation I contained an extra *Hind*dIII site close to one end of the insert.

(b) Expression of the *URA3* gene from phage promoters

To examine the ability of the cloned gene to be expressed in *E. coli*, cells of the *pyrF* mutant were infected with phages bearing *URA3* in either orientation. OMP decarboxylase was assayed at several times after infection. Orientation I phage produce the enzyme late in infection, whereas orientation II phage produce the enzyme early and transiently (Fig. 2). This is consistent with expression of an unstable enzyme predominantly from phage promoters. Orientation II phage expression would be from the early leftward facing promoter p_L. Orientation I phage would be expressed from the late rightward-facing promoter, p'_R, as a result of circularization of the λ chromosome. The highest level of activity during the infection corresponds to less than 5% of the level of wild-type *E. coli* cells grown without uracil.

To determine the level of expression from the *URA3* gene in the absence of the strong transcription provided by the phage promoters, it was necessary to construct stable lysogens of the phages. As the hybrid phages contain deletions of both the *c*I and *int* genes required for lysogenization we crossed the hybrid phages with phage NM816 (λ *Jam6lac5*-

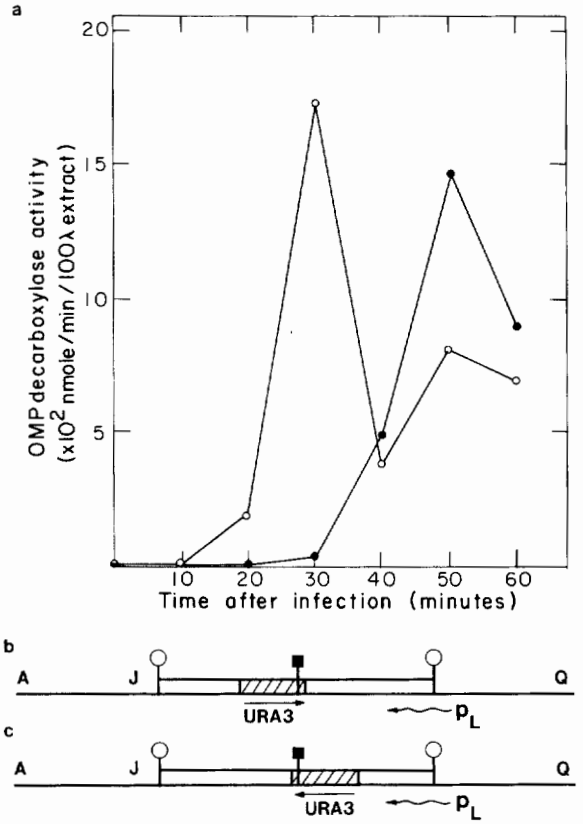

Fig. 2. Kinetics of expression of OMP decarboxylase from the *URA3* gene carried on λgt7 in either orientation. (a) DB6566 cells (*pyrF*⁻) were grown to 3 × 10^8 cells per ml at 37°C and infected with λ*URA3*-I (●) or λ*URA3*-II (○) at an m.o.i. of 2. At the indicated times aliquots were chilled, concentrated 25-fold and frozen at −20°C. Sonicated extracts were prepared and assayed for OMP-decarboxylase activity as described in MATERIALS AND METHODS, section **f**. Activity is expressed as nmol of substrate reacted per min/100 μl of extract. Such extracts usually contain approx. 2 mg protein/ml. Measurements are the mean of two determinations. (b) Schematic structure of λ*URA3*-I. Restriction sites are as in Fig. 1. (c) Schematic structure of λ*URA3*-II. Note that during lytic infection the λ chromosome is circular. Thus the rightward facing late promoter p'_R which lies to the right of the *Q* gene (Fig. 3b) can transcribe the *URA3* gene in late infection.

*imm*21c Its). Turbid (containing the λNM816 immunity region), Lac⁻ (containing the yeast *URA3* sequences) recombinants were picked and screened for *int*⁺ by the red-plaque test of Enquist and Weisberg (1976). Candidate phages were further checked by restriction enzyme analysis for the correct structure. Lysogens of λ*URA3*-I in the *pyrF* mutant were able to grow slowly in the absence of added uracil whereas lysogens of λ*URA3*-II were completely dependent on the addition of uracil to the growth media. Enzyme assays showed that λ*URA3*-I lysogens contained low but detectable levels of OMP decarboxylase whereas the λ*URA3*-II lysogens contained no detectable enzyme activity. The level of enzyme detected in the λ*URA3*-I lysogen was less than 0.5% of the level present in a *pyrF*⁺ strain and about 10% of the maximal level observed during infection. The assymetries of lytic and lysogenic expression suggest that the *URA3* gene is dependent on the phage promoters for expression in *E. coli*.

To verify that the enzyme level is a function of the level of transcription from phage promoters, we used *Q*-transactivation to specifically activate the late operon of the prophage (Thomas, 1966; Dambly et al., 1968; Herskowitz and Signer, 1974). In lysogens superinfected with a heteroimmune phage, which contains the homologous *Q* gene, the late operon of the prophage is specifically activated in the absence of prophage induction. A heteroimmune phage with a heterologous *Q* gene does not cause this "transactivation" showing that the activation is specifically due to the *Q* gene product. Lysogens of λ*URA3* were infected with a λ phage having phage 434 immunity and either the λ *QSR* region or the heterologous region from a cryptic lambdoid prophage (p4). As can be seen in Fig. 3 in the λ*URA3*-I lysogen, OMP-decarboxylase synthesis is induced by superinfection with λ*imm*434 Q^{λ}. The heterologous phage is unable to induce the enzyme and the lysogens bearing the opposite orientation of *URA3* within the prophage are not induced. Thus the enzyme level is a function of transcription from the late promoter p'_R. These experiments imply that the gene as isolated from the yeast strain FL100 does not contain a functional *E. coli* promoter. In addition, they show that the *URA3* gene must be oriented on λ*URA*-I so that it is normally transcribed in the same direction as the late operon. Referring to the *Hin*dIII fragment illustrated in Fig. 1 this means that transcription is from the

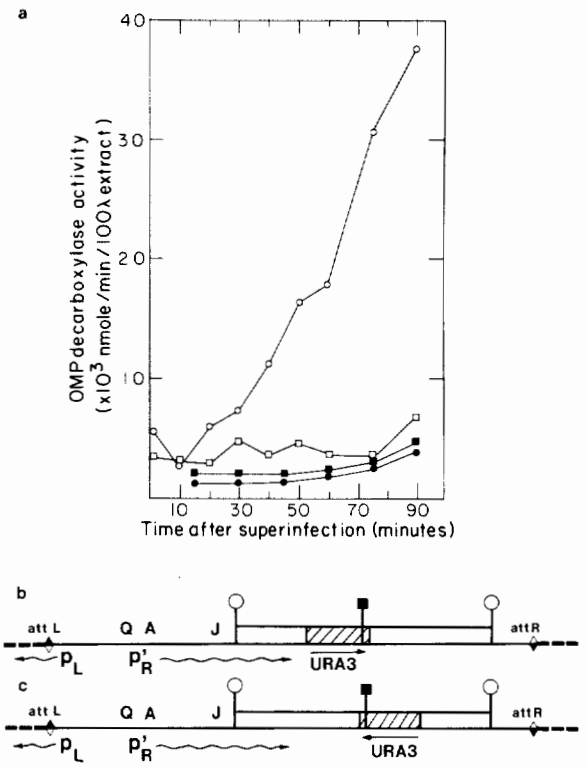

Fig. 3. Expression of OMP decarboxylase from the $\lambda p'_R$ promoter in lysogens by "*Q*-transactivation". (a) Lysogens of λ*URA3*-I (○,□) and λ*URA3*-II (●,■) were grown at 32°C to a density of 2×10^8 cells/ml in the presence of uracil. Cells were infected with λ*imm*434cIts (○,●) and λ*imm*434cItsp4 (□,■) at an m.o.i. of 4. At the indicated times cells were chilled, concentrated 25-fold and stored frozen at −20°C. Sonicated extracts were prepared and assayed for OMP-decarboxylase activity as in Fig. 2. (b) Schematic structure of lysogens of λ*URA3*-I. Note that integration changes the order of the insert DNA relative to the λ promoters. Flanking *E. coli* DNA is indicated by broken lines. Restriction sites are as in Figs. 1 and 2. (c) Schematic structure of lysogens of λ*URA3*-II.

*Pst*I site towards the *Sma*I site. The same conclusion was reached by Hubert et al. (1980).

(c) Expression of the *URA3* gene from plasmid promoters

Previously published results on the *URA3* gene isolated from the yeast strain + D4 indicated that the *Hin*dIII fragment expressed OMP decarboxylase in either orientation at the *Hin*dIII site of pBR322, indicating that the gene is expressed in *E. coli* by a promoter present on the yeast fragment. Our results

indicated that it is unlikely that the *URA3* gene from the strain FL100 contains such a promoter. To resolve this apparent contradiction, the *URA3 Hin*dIII fragment was isolated from *λURA3*-I and ligated into the *Hin*dIII site of pBR322. Ampicillin-resistant transformants of the *pyrF* strain DB6566 were screened for uracil-independent growth and resistance to tetracycline. Of 45 tetracycline-sensitive colonies isolated, 24 were uracil-dependent. Plasmid DNA was isolated from six Ura$^+$ Tet^s and four Ura$^-$ Tet^s colonies. All ten plasmids contained the *URA3 Hin*dIII fragment. The orientation of the *Hin*dIII fragment correlated exactly with the ability to grow without added uracil. The orientation in which the insert confers uracil independence corresponds to transcription of the *URA3* gene in a counterclockwise fashion on pBR322. This is opposite to the direction of transcription of the tetracycline resistance gene whose promoter is disrupted by insertion at the *Hin*dIII site. This is consistent with the existence of a previously identified counterclockwise promoter in the tetracycline resistance region (Stuber and Bujard, 1981). We repeated the subcloning of the *Hin*dIII fragments from the +D4 *URA3* containing plasmid "clone 1" using the same batch of pBR322 DNA. Sixteen Ura$^+$ plasmids analyzed were equally divided between the two orientations of the *URA3* fragment. No difference in growth rate in the absence of uracil was observed amongst transformants carrying the different plasmids. Thus the *URA3* genes from yeast strains +D4 and FL100 differ in their ability to be expressed in the *E. coli* cell.

(d) DNA sequence of the *URA3* gene

To determine the reason for the difference in expression of the genes in *E. coli* and the extent of divergence between them we determined the DNA sequence of both *URA3* genes. Two different strategies were used to sequence the two genes. The *URA3 Hin*dIII fragment from +D4 was sequenced from the naturally occurring restriction sites within the gene using the chemical method of Maxam and Gilbert (1980). The scheme of fragments which were sequenced is shown in Fig. 4. The gene from FL100 was sequenced by making use of a series of plasmids each of which contained a unique *Bam*HI site which had been artificially introduced at random sites in the *URA3* gene (Rose and Botstein, 1983). The position

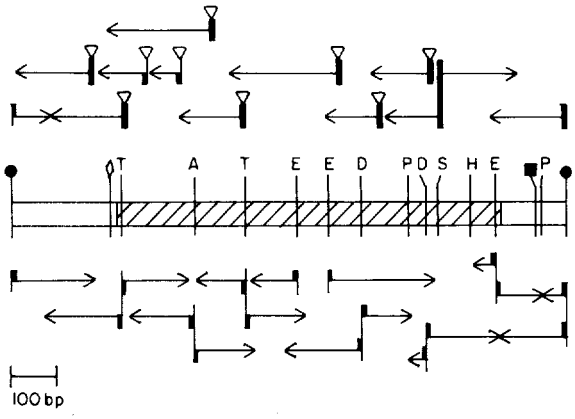

Fig. 4. Strategy for DNA sequencing of the *URA3* 1.1-kb *Hin*dIII fragment. The *Hin*dIII fragment derived from the +D4 background was sequenced from the indicated restriction sites (lower set of arrows). The fragment obtained from the FL100 background (upper set of arrows) was largely sequenced from a set of plasmids containing derivatives of the gene with artificial *Bam*HI sites introduced at various positions within the gene. The positions of the *Bam*HI sites are indicated by the tails of the upper set of arrows which have triangles. The length of the arrow is proportional to the amount of sequence obtained from any one labeled site. The method of labeling and therefore the DNA strand from which the sequence was obtained is indicated by the width of the tail of the arrows. Fragments which were 5′-labeled by polynucleotide kinase are indicated by the thick tails above the arrow shaft. Fragments which were labeled at the 3′ end by DNA polymerase are indicated by thick tails below the shaft. Totally thick tails indicates that both strands were sequenced. The hatched region indicates the *URA3* coding region. Restriction sites are as in Fig. 1 and as follows: T, *Taq*I; A, *Ava*II; E, *Hae*III; D, *Hin*dII; P, *Hpa*II; S, *Sau*3A; H, *Hph*I.

of the *Bam*HI sites used for the sequencing is shown in Fig. 4.

The DNA sequences of the two genes are shown in Fig. 5. A single long open reading frame of 801 bp starts with an ATG codon. The predicted protein is 267 amino acids long, which agrees well with the observed monomer M_r of 27 500 (Brody and Westheimer, 1978). As can be seen, the genes from the two yeast strains are very similar in DNA sequence. Ten differences are found (two deletions, one of 5 bp at position 14 and one of 1 bp within a run of T's starting at position 23, six transitions at positions 33, 49, 88, 146, 637, 1132, and two transversions at positions 22 and 704). Two of the mutations change a total of three restriction sites (at position 49 the presence of T in the +D4 sequence destroys *Hin*fI and *Hpa*II sites, at position 704 the FL100 sequence

```
          10        V       T          T  G      40          C           60          70
   AGCTTTTCAA TTCAATTCAT CATTTTTTTT TTATTCTTTT TTTTGATTTC GGTTTCTTTG AAATTTTTTT
                     *       *        * *          *

          80         G          100         110         120         130         140
   GATTCGGTAA TCTCCGAACA GAAGGAAGAA CGAAGGAAGG AGCACAGACT TAGATTGGTA TATATACGCA
                      *

       G 150         160         170         180         190         200         210
   TATGTAGTGT TGAAGAAACA TGAAATTGCC CAGTATTCTT AACCCAACTG CACAGAACAA AAACCTGCAG
     *

         220                                 T   250
   GAAACGAAGATAAATC ATG TCG AAA GCT ACA TAT AAG GAA CGT GCT GCT ACT CAT CCT AGT
                    Met ser lys ala thr tyr lys glu arg ala ala thr his pro ser

                                       301                                     331
CCT GTT GCT GCC AAG CTA TTT AAT ATC ATG CAC GAA AAG CAA ACA AAC TTG TGT GCT TCA
pro val ala ala lys leu phe asn ile met his glu lys gln thr asn leu cys ala ser

                                       361                                     391
TTG GAT GTT CGT ACC ACC AAG GAA TTA CTG GAG TTA GTT GAA GCA TTA GGT CCC AAA ATT
leu asp val arg thr thr lys glu leu leu glu leu val glu ala leu gly pro lys ile

                                       421                                     451
TGT TTA CTA AAA ACA CAT GTG GAT ATC TTG ACT GAT TTT TCC ATG GAG GGC ACA GTT AAG
cys leu leu lys thr his val asp ile leu thr asp phe ser met glu gly thr val lys

                                       481                                     511
CCG CTA AAG GCA TTA TCC GCC AAG TAC AAT TTT TTA CTC TTC GAA GAC AGA AAA TTT GCT
pro leu lys ala leu ser ala lys tyr asn phe leu leu phe glu asp arg lys phe ala

                                       541                                     571
GAC ATT GGT AAT ACA GTC AAA TTG CAG TAC TCT GCG GGT GTA TAC AGA ATA GCA GAA TGG
asp ile gly asn thr val lys leu gln tyr ser ala gly val tyr arg ile ala glu trp

                                       601                                     631
GCA GAC ATT ACG AAT GCA CAC GGT GTG GTG GGC CCA GGT ATT GTT AGC GGT TTG AAG CAG
ala asp ile thr asn ala his gly val val gly pro gly ile val ser gly leu lys gln

     G                                 661                                     691
GCG GCA GAA GAA GTA ACA AAG GAA CCT AGA GGC CTT TTG ATG TTA GCA GAA TTG TCA TGC
ala ala glu glu val thr lys glu pro arg gly leu leu met leu ala glu leu ser cys
      *

               ala
               G                       721                                     751
AAG GGC TCC CTA TCT ACT GGA GAA TAT ACT AAG GGT ACT GTT GAC ATT GCG AAG AGC GAC
lys gly ser leu ser thr gly glu tyr thr lys gly thr val asp ile ala lys ser asp
               *

                                       781                                     811
AAA GAT TTT GTT ATC GGC TTT ATT GCT CAA AGA GAC ATG GGT GGA AGA GAT GAA GGT TAC
lys asp phe val ile gly phe ile ala gln arg asp met gly gly arg asp glu gly tyr

                                       841                                     871
GAT TGG TTG ATT ATG ACA CCC GGT GTG GGT TTA GAT GAC AAG GGA GAC GCA TTG GGT CAA
asp trp leu ile met thr pro gly val gly leu asp asp lys gly asp ala leu gly gln

                                       901                                     931
CAG TAT AGA ACC GTG GAT GAT GTG GTC TCT ACA GGA TCT GAC ATT ATT ATT GTT GGA AGA
gln tyr arg thr val asp asp val val ser thr gly ser asp ile ile ile val gly arg

                                       961                                     991
GGA CTA TTT GCA AAG GGA AGG GAT GCT AAG GTA GAG GGT GAA CGT TAC AGA AAA GCA GGC
gly leu phe ala lys gly arg asp ala lys val glu gly glu arg tyr arg lys ala gly

                                     1021                                    1055
TGG GAA GCA TAT TTG AGA AGA TGC GGC CAG CAA AAC TAA AAAAC TGTATTATAA GTAAATGCAT
trp glu ala tyr leu arg arg cys gly gln gln asn end

       1065                                1095                                1125
  GTATACTAAA CTCACAAATT AGAGCTTCAA TTTAATTATA TCAGTTATTA CCCGGGAATC TCGGTCGTAA

       C          1145                          1170
  TGATTTTTAT AATGACGAAA AAAAAAAAAT TGGAAAGAAA AAGCT
       *
```

Fig. 5. DNA sequence of the *URA3* 1.1-kb *Hin*dIII fragment. The sequence of the gene as isolated from the strain + D4 is shown. Changes observed in the strain FL100 are indicated by the base changes written above the sequence and by the asterisks. A small deletion of 5 bp is shown beginning at position 14. The exact position is indeterminant because it occurs within a sequence which contains almost three tandem TTCAA repeats. Similarly the extra T indicated at position 30 may lie anywhere within the stretch of T's. Other changes are indicated at positions 22, 33, 49, 88, 146, 637, 704, and 1132. The three ATG codons at the upstream end of the segment are underlined with horizontal brackets.

contains an *Alu*I site that is not present in the + D4 sequence).

Much of the region to the 5′ side of the coding region is composed of long stretches in which one strand is largely composed of purines or pyrimidines. This asymmetry switches strand at one place. This feature may be involved in the regulation or expression of the gene in yeast. In addition a fairly good match is present between the sequence TATATATA beginning at position 129 and the Goldberg-Hogness box that has been postulated to be important in positioning the start of transcription in eukaryotes (Corden et al., 1980). Of some interest in regards to the expression of the gene in *E. coli* is the fairly good correspondence between the sequence starting at position 145 and the conserved −10 region of the *E. coli* promoter. In addition a fairly good match to the −35 region is found with the appropriate spacing at position 125. This is the best match to the consensus sequence that can be made in the region 5′ to the coding sequence. One of the base differences between the two sequences is a change from A to G that would mutate one of the most conserved bases in the −10 region. Indeed mutation of this base has been observed to be a promoter down mutation in several systems (Rosenberg and Court, 1979). It is likely that the difference in expression in *E. coli* observed between the two genes is due to the mutation of a fortuitous promoter sequence.

Two of the base changes between the two yeast genes lie within the coding sequence. One of the coding sequence mutations would result in the change of a serine residue to an alanine. The second coding sequence change does not change the codon assignment. In addition to the difference in expression in *E. coli* that we observed, Lacroute, F. (personal communication) has observed antigenic differences between the OMP-decarboxylase proteins produced by the two yeast strains. This observation is explained by the predicted single amino acid difference between the two proteins.

The ATG at position 227 is not the first ATG found in the sequence. Two additional ATGs are present upstream at positions 142 and 160. The first ATG is followed immediately by an in frame termination codon in + D4 or two codons later in strain FL100. The second ATG begins an open reading frame which does not terminate until 28 codons downstream. This open reading frame overlaps the long open reading frame begun by the third ATG by 17 bp. The presence of ATG codons close upstream of eukaryotic genes is unusual (Kozak, 1980; 1981).

The DNA sequence downstream of the *URA3* coding sequence contains neither the AATAAA transcription termination signal of higher eukaryotes (Benoist et al., 1980) nor the yeast consensus termination sequence identified by Zaret and Sherman (1982). However, the sequence TTTTTATA, identified by Henikoff et al. (1983) as essential for transcription termination of a *Drosophila* gene in yeast, is present downstream of *URA3* at position 1129. Note the presence of a point mutation within this sequence in the FL100 version of the gene.

(e) Amber mutations of the *URA3* gene

Three amber mutations of the *URA3* gene isolated in the FL100 background were cloned onto λgt7, essentially as described for the wild–type gene.

Lysogens of orientation I phages were constructed as before using the *su°pyrF* strain DB6656 and the isogenic *suIII pyrF* strain DB6660. Lysogens of λ*ura3*-18 and λ*ura3*-3 were able to grow in the absence of uracil only in the strain carrying the *suIII* suppressor. Lysogens of the third cloned amber mutation were Ura⁻ regardless of the suppressor mutation. Thus two of the genes behave as amber mutations in *E. coli* as they did in yeast. The mutation *ura3*-25 presumably has a different suppressor spectrum from the other two mutations. Suppression of the two amber alleles was confirmed by subcloning the *Bam*HI fragments containing the *ura3*-3 and *ura3*-18 alleles onto multicopy plasmids.

One of the mutations, *ura3*-3, has been shown to map genetically near one end of the gene (Losson and Lacroute, 1979), between two artificially introduced *Bam*HI sites (Falco, 1983). This led to the prediction that the mutation would lie within a specific enzyme restriction fragment. The appropriate *Taq*I-*Ava*II fragment was isolated and sequenced. The *ura3*-3 mutation was found to be an A to T transversion at position 245 (Fig. 5),which changes the seventh codon from lysine to amber. The autoradiogram of this sequencing gel is shown in Fig. 6. This confirms that the protein starts to the 5′ side of this mutation and not at the next available methionine at position 299.

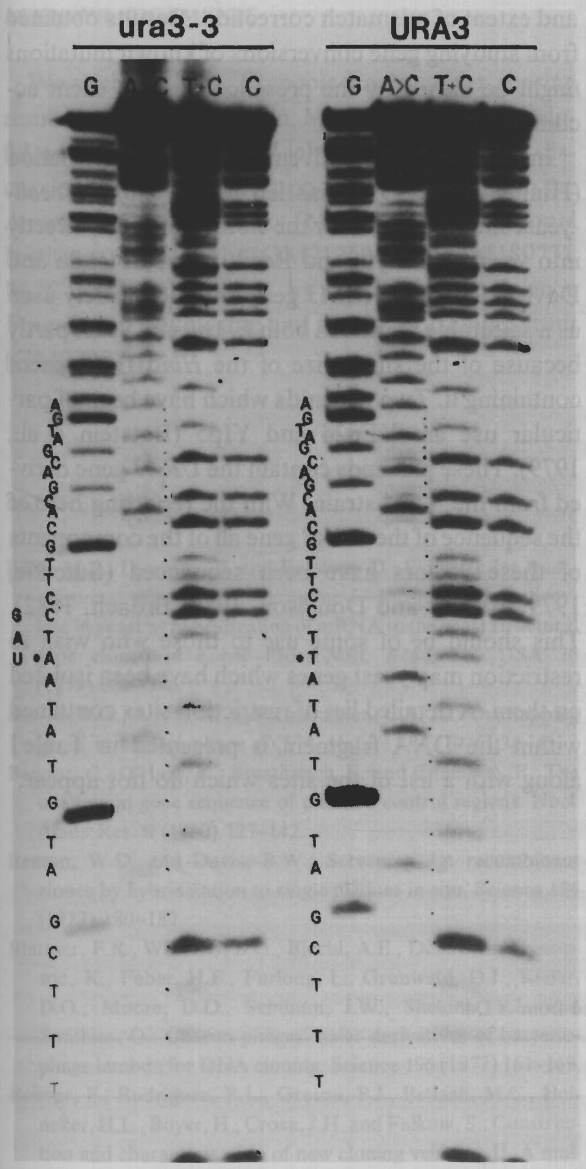

Fig. 6. DNA sequencing gel for the *ura3*-3 amber mutation. Plasmids carrying either the wild-type form of the FL100 *URA3* gene (pRB101) or the *ura3*-3 allele (pRB43) were cleaved with *Taq*I and labeled by filling in a portion of the cohesive end with [α-^{32}P]dCTP. The appropriate *Taq*I fragments were purified and cleaved with *Ava*II. The *Taq*I-*Ava*II fragments were purified and analysed by the chemical DNA sequencing methods of Maxam and Gilbert (1980). A single base change which creates an amber codon on the complementary DNA strand is indicated.

DISCUSSION

The yeast *URA3* gene was isolated from two different laboratory strains of *S. cerevisiae*. Expression of the gene from the yeast strain FL100 in *E. coli* is completely dependent on transcription from bacterial promoters in the surrounding vector sequences. The gene as isolated from the yeast strain + D4 contains sequences which are sufficiently similar to the bacterial promoter to express the gene in *E. coli* in the absence of vector promoters. The expression of the gene from this promoter is not strong; even on the multicopy plasmid pBR322 the enzyme level amounts to only about a third of the level of the analogous gene on the *E. coli* chromosome. Both yeast genes still possess functional eukaryotic promoters (Chevallier et al., 1980; Rose et al., 1981). Thus expression of a yeast gene in *E. coli* is not an inherent property of yeast promoters even for those genes which are fortuitously expressed in *E. coli*.

The DNA sequence of the 5′ region of the two genes differs in several places. Adjacent to the structural gene isolated from + D4 is a sequence which is fairly homologous to the consensus sequence for the *E. coli* promoter. The idea that this sequence is the promoter responsible for expression in *E. coli* is supported by the observation that one of the most conserved bases of the −10 region is different in the gene which is not expressed in *E. coli*. The observation that this base change affects expression of the gene in *E. coli* but not in yeast suggests that the essential promoter elements are not the same in yeast and *E. coli*.

We have isolated the homologous DNA fragment from strains bearing amber alleles of the *URA3* gene. Some of these DNA fragments express functional OMP decarboxylase only in *E. coli* strains which carry an amber suppressor. Direct DNA sequencing of one of the amber alleles revealed that the mutation is a change in the seventh codon of the structural gene. The *ura3*-3 allele was originally assumed to lie near the amino terminus of the structural gene because it showed the most extreme instability of the mRNA as compared to a series of amber mutations mapping throughout the *URA3* gene (Losson and Lacroute, 1979).

The unexpected finding of out-of-frame ATG codons preceding the apparent initiator for the *URA3* coding sequence deserves mention, especially since one of these, if expressed, would constitute an overlapping gene (Barrell et al., 1976; Shaw et al., 1978). Gene fusion experiments suggest that the initiator codon for this small peptide is both transcribed and

translated; these results and their implications are discussed elsewhere (Rose and Botstein, 1983). The existence of numerous apparently silent DNA polymorphisms amongst two different laboratory strains of yeast is a surprising result. It has long been known that different strains commonly contain restriction site polymorphisms (Petes and Botstein, 1977). In this case some 10 different changes appear in about 1100 bases for an average divergence of about 1%. Most of changes are clustered in the 200 bp to the 5′ side of the structural gene. Presumably most of this region contains little essential information and therefore is free to change. Two of the mutations are in the coding sequence and one of these causes a change in the primary structure of the protein. The probable presence of such changes all over the genome must be considered when using strains which are not known to be isogenic. Such base changes would lead to mismatches in the heteroduplex DNA formed during recombination. The mismatches, particularly those involving small deletions might have serious effects on the formation, branch migration and resolution of heteroduplexes as well as on the polarity and extent of mismatch correction. Results obtained from studying gene conversions of known mutations might be biased by the presence of these silent accompanying mutations.

In recent years, the advent of yeast transformation (Hinnen et al., 1978) has led to the use of *E. coli*-yeast shuttle vectors for the cloning of genes directly into yeast (Nasmyth and Reed, 1980; Botstein and Davis, 1982). The *URA3* gene has been widely used as a selectable marker in both *E. coli* and yeast partly because of the small size of the *Hin*dIII fragment containing it. Two plasmids which have been of particular use are YEp24 and YIp5 (Botstein et al., 1979). These plasmids contain the *URA3* gene derived from the + D4 strain. With the reporting here of the sequence of the *URA3* gene all of the components of these vectors have been sequenced (Sutcliffe, 1979; Hartley and Donelson, 1980; Broach, 1982). This should be of some use to those who wish to restriction map yeast genes which have been isolated on them. A detailed list of restriction sites contained within the DNA fragment is presented in Table I along with a list of the sites which do not appear.

TABLE I

Restriction enzyme sites in the *URA3* 1.1-kb *Hin*dIII fragment derived from + D4[a]

Enzyme	Sites	Enzyme	Sites
*Acc*I	551, 1056	*Hin*dII	731, 867
*Alu*I	1, 235, 285, 1078, 1167	*Hin*fI	71, 1111
*Asu*I	380, 601, 602	*Hpa*II	831, 1107
*Ava*I	1106	*Hph*I	968
*Ava*II	380	*Mbo*II	95, 152, 216, 489, 494, 638, 742, 795, 927, 1008
*Ava*III	1050	*Mnl*I	437, 660, 930, 965
*Bbv*I	953, 1011	*Nci*I	830, 1106
*Dde*I	119, 720, 957	*Nco*I	432
*Eco*PI	895	*Pst*I	205
*Eco*RI*	61, 387, 503, 754, 905, 1111	*Rsa*I	342, 475, 538, 726
*Eco*RII	605	*Sau*3A	906
*Eco*RV	413	*Sfa*NI	953, 1011
*Fnu*4HI	254, 278, 451, 632, 1014	*Sma*I	1106
*Hae*I	661	*Stu*I	661
*Hae*III	602, 662, 1016	*Taq*I	230, 492
*Hga*I	857	*Xho*II	905
*Hgi*AI	110		

[a] The following sites do not appear: *Acy*I, *Avr*II, *Bal*I, *Bam*HI, *Bcl*I, *Bgl*I, *Bgl II*, *Bst*EII, *Cla*I, *Eco*B, *Eco*K, *Eco*P15, *Eco*RI, *Fnu*DII, *Hae*II, *Hha*I, *Hpa*I, *Kpn*I, *Mst*I, *Nae*I, *Nar*I, *Nru*I, *Pvu*I, *Pvu*II, *Sac*I, *Sac*II, *Sal*I, *Sph*I, *Xba*I, *Xho*I, *Xma*III, *Xmn*I.

ACKNOWLEDGEMENTS

We wish to thank François Lacroute for sharing results prior to publication, Marion Carlson for helpful advice and Michael Lichten for stimulating discussion. This work was supported by grants from the American Cancer Society (MV90) and the National Institutes of Health (GM21253 and GM18973). M.R. was supported by a National Science Foundation Fellowship and an N.I.H. training grant (GM07287).

REFERENCES

Bach, M.L., Lacroute, F. and Botstein, D.: Evidence for transcriptional regulation of orotidine-5'-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in *E. coli* Proc. Natl. Acad. Sci. USA 76 (1979) 386–390.

Barrell, B.G., Air, G.M. and Hutchinson, C.A. III: Overlapping genes in bacteriophage ϕX174. Nature 264 (1976) 34–41.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P.: The ovalbumin-gene sequence of putative control regions. Nucl. Acids Res. 8 (1980) 127–142.

Benton, W.D. and Davis, R.W.: Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196 (1977) 180–182.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O.: Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science 196 (1977) 161–169.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heineker, H.L., Boyer, H., Crosa, J.H. and Falkow, S.: Construction and characterization of new cloning vehicles, II. A multipurpose cloning system. Gene 2 (1977) 95–113.

Botstein, D., Falco, S.C., Stewart, S.E., Brennan, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis, R.W.: Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. Gene 8 (1979) 17–24.

Botstein, D. and Davis, R.W.: Principles and practice of recombinant DNA research with yeast, in Strathern, J., Jones, E. and Broach, J. (Eds.), The Molecular Biology of the Yeast *Saccharomyces*. Metabolism and Gene Expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Bradford, M.M.: A dye binding assay for protein. Anal. Biochem. 72 (1976) 248–254.

Broach, J.R.: The yeast plasmid, 2μ circle, in Strathern, J., Jones, E.W. and Broach. J.R. (Eds.), The Molecular Biology of the Yeast *Saccharomyces*. Life Cycle and Inheritance. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1981, pp. 445–470.

Brody, R.S. and Westheimer, F.H.: The purification of orotidine-5'-phosphate decarboxylase from yeast by affinity chromatography. J. Biol. Chem. 254 (1978) 4238–4244.

Chevallier, M.R., Bloch, J.C. and Lacroute, F.: Transcriptional and translational expression of a chimeric bacterial-yeast plasmid in yeasts. Gene 11 (1980) 11–19.

Clewell, D.B. and Helinski, D.R.: Supercoiled circular DNA protein complex in *E. coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA 62 (1969) 1159–1166.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P.: Promoter sequences of eukaryotic protein-coding genes. Science 209 (1980) 1406–1414.

Dambly, C., Couturier, M. and Thomas, R.: Control of lysozyme synthesis. J. Mol. Biol. 32 (1968) 67–81.

Davis, R.W., Botstein, D. and Roth, J.R.: A Manual For Genetic Engineering: Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980.

Enquist, L.W. and Weisberg, R.A.: The red plaque test: a rapid method for identification of excision defective variants of bacteriophage lambda. Virology 72 (1976) 147–153.

Falco, S.C., Rose, M. and Botstein, D.: Homologous recombination between episomal plasmids and chromosomes in yeast. Genetics 105 (1983) 843–856.

Fiandt, M., Hradečná, Z., Lozeron, H.A. and Szybalski, W.: Electron micrographic mapping of deletions, insertions, inversions, and homologies in the DNAs of coliphages lambda and phi80, in Hershey, A.D. (Ed.). The Bacteriophage Lambda, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1971, pp. 329-354.

Hartley, J.L. and Donelson, J.E.: Nucleotide sequence of the yeast plasmid. Nature 286 (1980) 860–865.

Henikoff, S., Kelly, J.D. and Cohen, E.H.: Transcription terminates in yeast distal to a control sequence. Cell 33 (1983) 607–614.

Herskowitz, I. and Signer, E.R.: Substitution mutations in bacteriophage λ with new specificity for late gene expression. Virology 61 (1974) 112–119.

Hinnen, A., Hicks, J.B. and Fink, G.R.: Transformation in yeast. Proc. Natl. Acad. Sci. USA 75 (1978) 1929–1933.

Hohn, B. and Murray, K.: Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc. Natl. Acad. Sci. USA 74 (1977) 3259–3263.

Hubert, J.C., Bach, M.L. and Lacroute, F.: Measure of asymmetrical transcription of the yeast OMP decarboxylase gene expressed in yeast or in *E. coli*. Curr. Genet. 2 (1980) 103–107.

Kozak, M.: How do eukaryotic ribosomes select initiation regions in messenger RNA. Cell 15 (1978) 1109–1123.

Kozak, M.: Evaluation of the "Scanning Model" for initiation of protein synthesis in eukaryotes. Cell 22 (1980) 7–8.

Kozak, M.: Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucl. Acids Res. 9 (1981) 5233–5252.

Lacroute, F.: Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. J. Bacteriol. 95 (1968) 824–832.

Lieberman, I., Kornberg, A. and Simmons, E.S.: Eukaryotic

synthesis of pyrimidine nucleotides orotidine 5′ phosphate and uridine-5′-phosphate. J. Biol. Chem. 215 (1955) 403–415.

Losson, R. and Lacroute, F.: Interference of nonsense mutation with eukaryotic messenger RNA stability. Proc. Natl. Acad. Sci. USA 76 (1979) 5134–5137.

Mandel, M. and Higa, A.: Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53 (1970) 159–162.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages, in Grossman, L. and Moldave, K. (Eds.) Methods in Enzymology, Vol. 65. Academic Press, New York, 1980, pp. 499–560.

Miller, J.H.: Experiments in Molecular Genetics. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1972.

Nasmyth, K.A. and Reed, S.I.: The isolation of genes by complementation in yeast: molecular cloning of a cell cycle gene. Proc. Natl. Acad. Sci. USA 77 (1980) 2119–2123.

Petes, T.D. and Botstein, D.: Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. Proc. Natl. Acad. Sci. USA 74 (1977) 5091–5095.

Petes, T.D., Broach, J.R., Wensink, P.C., Hereford, L.M., Fink, G.R. and Botstein, D.: Isolation and analysis of recombinant DNA molecules containing yeast DNA. Gene 4 (1978) 37–49.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P.: Labelling DNA to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113 (1977) 237–251.

Rose, M. and Botstein, D.: Structure and function of the yeast *URA3* gene. Differentially regulated expression of hybrid beta-galactosidase from overlapping coding sequences in yeast. J. Mol. Biol. 170 (1983) 883–904.

Rose, M., Casadaban, M.J. and Botstein, D.: Yeast genes fused to β-galactosidase can be expressed normally in yeast. Proc. Natl. Acad. Sci. USA 78 (1981) 2460–2464.

Rosenberg, M. and Court, D.: Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13 (1979) 319–353.

Shaw, D.C., Walker, J.E., Northrop, F.D., Barrell, B.G., Godson, G.N. and Fiddes, J.C.: Gene K, a new overlapping gene in bacteriophage G4. Nature 272 (1978) 510–515.

Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 (1975) 503–517.

Sternberg, N., Tiemeier, D. and Enquist, L.: In vitro packaging of a λ*D*am vector containing *Eco*RI DNA fragments of *Escherichia coli* and phage P1. Gene 1 (1977) 255–280.

Stuber, D. and Bujard, H.: Organization of transcriptional signals in plasmids pBR322 and pACYC184. Proc. Natl. Acad. Sci. USA 78 (1981) 167–171.

Sutcliffe, J.G.: Complete nucleotide sequence of the *E. coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. 43 (1979) 77–90.

Thomas, R.: Control of development in temperate bacteriophages, I. Induction of prophage genes following heteroimmune super-infection. J. Mol. Biol. 22 (1966) 79–95.

Weil, J., Cunningham, R., Martin, R., Mitchell, R. and Bolling, B.: Characteristics of λp4 a λ derivative containing 9% excess DNA. Virology 50 (1972) 373–380.

Yamamoto, K., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G.: Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. Virology 40 (1970) 734–744.

Zaret, K.S. and Sherman, F.: DNA sequence required for efficient transcription termination in yeast. Cell 28 (1982) 563–573.

Communicated by G.R. Fink.