J. Mol. Biol. (1983) **170**, 883–904

Structure and Function of the Yeast *URA3* Gene
Differentially Regulated Expression of Hybrid β-Galactosidase from Overlapping Coding Sequences in Yeast

MARK ROSE AND DAVID BOTSTEIN

Department of Biology
Massachusetts Institute of Technology
Cambridge, Mass. 02139, U.S.A.

(Received 27 May 1983, and in revised form 3 August 1983)

Expression of the *URA3* gene of *Saccharomyces cerevisiae* was studied by analysis of *URA3–lacZ* gene fusions constructed *in vitro*. Synthesis of hybrid β-galactosidase by fusions in frame with the coding sequence for orotidine-5′-phosphate decarboxylase (OMPdecarboxylase) was found to be normally regulated even when only 11 nucleotides of *URA3* coding sequence remained, indicating that all transcription initiation and regulatory sites are present at the beginning of the *URA3* gene. An upstream initiator codon that begins a short overlapping coding sequence in another reading frame was also found to be active in producing hybrid β-galactosidase. However this β-galactosidase synthesis showed little or no regulation. Nuclease protection experiments revealed numerous species of *URA3* mRNA. The regulation of these is consistent with the idea that the *URA3* protein and the overlapping peptide are translated from differentially regulated mRNAs of different lengths.

1. Introduction

The *Escherichia coli lacZ* gene has been fused to the yeast *URA3* gene in *E. coli* and produces hybrid β-galactosidase in yeast (Rose *et al.*, 1981). The expression of one *URA3–lacZ* gene fusion was found to be regulated in yeast in essentially the same fashion as the intact *URA3* gene. This result along with the similar demonstration for *lacZ* gene fusions to the cytochrome *c* gene (Guarente & Ptashne, 1981) showed that gene fusions can be used in yeast as an assay for gene expression.

The *URA3* gene encodes the enzyme orotidine-5′-phosphate decarboxylase, which catalyzes the last step in the *de novo* synthesis of pyrimidines. Increased intracellular levels of the precursor dihydro-orotate led to increased levels of the *URA3* mRNA (Bach *et al.*, 1979) and of OMPdecarboxylase†. Higher levels of dihydro-orotate are conveniently obtained in *ura1* mutants, which cannot convert dihydro-orotate to orotate.

† Abbreviations used: OMPdecarboxylase, orotidine-5′-phosphate decarboxylase; kb, 10^3 bases or base-pairs; exoVII, exonuclease VII.

0022–2836/83/320883–22 $03.00/0

In this paper we describe the use of gene fusions to determine whether there are sites within the structural gene sequence that might be essential for the expression or regulation of the *URA3* gene, similar to the internal "promoter" in the 5 S RNA genes of *Xenopus laevis* (Bogenhagen *et al.*, 1980; Sakonju *et al.*, 1980). Internal sites essential for gene expression or regulation will place a limit on the proximity of the fusion junction and the amino terminus of the coding sequence. Gene fusions that remove an internal site will not produce properly regulated β-galactosidase. Expression of the hybrid protein from an external promoter in *E. coli* serves as a control for differences in the specific activity of the hybrid protein. As shown below, functional gene fusions of *lacZ* to *URA3* have been obtained that are 11 base-pairs from the start of the *URA3* coding sequence. Thus, it is unlikely that any internal sequences are required for either expression or regulation of the *URA3* gene.

The DNA sequence of the *URA3* gene has been determined (Rose, 1982). The *URA3* coding sequence begins with an ATG codon and extends for 801 base-pairs before terminating with TAA. The DNA sequence displays a novel situation: an open reading frame beginning with an ATG codon is observed that precedes and overlaps the *URA3* coding sequence; this reading frame could potentially encode a peptide of 28 amino acids (Fig. 1). To determine whether this open reading frame is expressed *in vivo* we have isolated gene fusions that link β-galactosidase to the amino terminus of the overlapping peptide. Expression of β-galactosidase activity indicates that this segment is transcribed and translated *in vivo*. Moreover, we have mapped the 5′ termini of the RNA transcripts from the vicinity of the *URA3* gene. Both the expression of hybrid β-galactosidase fused to the peptide and the presence of RNA transcripts for the peptide support the hypothesis that the open reading frame constitutes a functional overlapping gene in yeast.

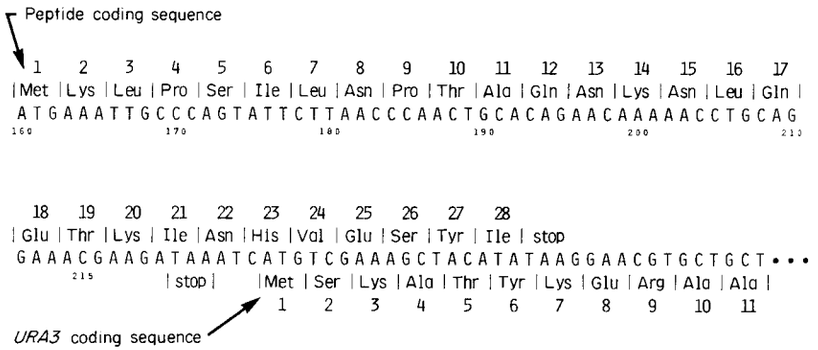

FIG. 1. An open reading frame beginning with ATG overlaps the start of the *URA3* coding sequence. A portion of the DNA sequence at the start of the *URA3* gene is illustrated. An ATG at position 160 begins an open reading frame that extends 84 base-pairs before terminating with a TAA codon at position 244. This open reading frame is designated the +2 or peptide reading frame. The +2 reading frame can potentially encode a peptide of 28 amino acids. For 17 base-pairs the +2 reading frame overlaps the O or protein reading frame that encodes the *URA3* gene. The nucleotide positions indicated are designated counting from the *Hin*dIII site upstream from the *URA3* gene. Hyphens have been omitted from the sequence for clarity.

2. Materials and Methods

(a) *Strains and genetic methods*

E. coli strain DB4572 ((*lacIPOZYA*) ΔX74 *galU galK sstrA pyrF74* : : Tn*5*) was derived from strain M182 (obtained from J. Beckwith) by phage P1 transduction to introduce the *pyrF* mutation. *E. coli* strain GM48 (*thr*$^-$, *leu*$^-$, *thi*$^-$, *lacY*, *galK*, *galT*, *ara*$^-$, *tonA*, *tsx*$^-$, *dam*$^-$, *dcm*$^-$, *supE*44; Marinus, 1973) was obtained from K. Backman and was used to prepare unmethylated plasmid DNA. Strain DB6566 (*trp*$_{am}$ *lac*$_{am}$ *pyrF* : : μ *hsdr*$^-$ *hsdm*$^+$) has been described previously (Bach *et al.*, 1979). Standard methods were used to transform *E. coli* with plasmid DNA (Mandel & Higa, 1970). Strains carrying plasmids were propagated on LB medium (Miller, 1972) supplemented with ampicillin (Bristol Laboratory) to 100 μg/ml. Ability to ferment lactose was determined by growth on M9 minimal medium (Miller, 1972) and MacConkey–lactose agar (Difco) as described by Rose *et al.* (1981). Low levels of β-galactosidase production could be discerned on plates by supplementing LB–ampicillin agar with 40 μg of X-gal/ml (5-bromo-4-chloro-3-indolyl-β-D-galactoside, Bachem Fine Chemicals, Torrance, Ca).

Yeast strains DBY689 (*MATa ura3-50 leu2-3 leu2*-112) and DBY941 (*MATa ura1*-21 *leu2*-3 *ade2*-1) were described by Rose *et al.* (1981) DBY629 (*MATa ura3*-50), DBY633 (*MATa ura1*-21) and DBY634 (*MATa ura2*-60) are all in the isogenic FL100 background (Lacroute, 1968) and were obtained from F. Lacroute. Yeast cells were grown in YEp–glucose (complete) or SD (minimal) media (Sherman *et al.*, 1979).

Yeast diploids were obtained after mating either by selection for complementation of two different auxotrophic markers or by micro-manipulation of zygotes (Sherman *et al.*, 1979). Conditions for sporulation were standard except that diploids containing YEp (2 μm plasmid) vectors were grown up in SD medium lacking leucine prior to shifting onto sporulation plates (Sherman *et al.*, 1979). Spores were either separated by micro-manipulation or by sonication (for strains containing YEp plasmids). Suspensions of asci were digested with a 10% solution of Glusulase (Endo Labs) for up to 5 h, after which they were sonicated for a total of 2 to 3 min using 15 s pulses with chilling in ice-water. Spores were judged to be separated adequately when greater than 95% of the spores were released from the ascal sac. Yeast strains were transformed by the method of Hinnen *et al.* (1978).

(b) *Plasmid isolation, restriction enzyme analysis and gel electrophoresis*

Plasmid DNA was prepared by a modification of the method of Clewell & Helinski (1969).

Most restriction enzymes were purchased from New England Biolabs. Endonuclease *Cla*I was purchased from Boehringer–Mannheim. Enzymes were used according to the recommendations of the manufacturer.

Agarose gel electrophoresis of DNA fragments was performed in 89 mM-Tris · HCl (pH 8·3), 89 mM-boric acid, 2·5 mM-Na_2EDTA (Peacock & Dingman, 1968). Polyacrylamide gel electrophoresis was performed in the same buffer.

(c) *Construction of* Bal*31-generated gene fusions*

pRB45 (60 μg) was cleaved with endonuclease *Sma*I, extracted with phenol and precipitated with ethanol. The DNA was resuspended at a final concentration of 100 μg/ml in the *Bal*31 buffer of Legerski *et al.* (1978) (20 mM-Tris · HCl (pH 8·0), 0·45 M-NaCl, 1 mM-Na_2EDTA, 12·5 mM-$MgSO_4$, 12·5 mM-$CaCl_2$). *Bal*31 was added to an empirically determined concentration such that approximately 100 base-pairs were removed per minute from each end of the linearized plasmid. The reaction was incubated at 28°C and samples were removed at 3 min intervals (15 μg each). The reaction was terminated by addition of Na_2EDTA to 25 mM followed by extraction with phenol and precipitation with ethanol. The DNA was resuspended at a final concentration of 200 μg/ml in 6 mM-Tris · HCl (pH 7·5), 10 mM-$MgCl_2$, 12·5 mM-dithiothreitol, 1 mM-spermidine, 1 mM-ATP and 200 μg

bovine serum albumin/ml. Phosphorylated *Bam*HI linkers (10mers, Collaborative Research) were added to a final concentration of 10 μM. Bacteriophage T4 DNA ligase (New England Biolabs) was added to a concentration of 20,000 cohesive end units per ml and incubated at 13°C for 2 h.

Plasmid DNA was purified away from excess linkers by agarose gel electrophoresis. The DNA was purified from the agarose by a modification of the hydroxyapatite adsorption method of Tabak & Flavell (1978). The purified DNA was resuspended at a final concentration of 200 μg/ml and digested to completion with *Bam*HI. The DNA was extracted with phenol, precipitated with ethanol and resuspended at a final concentration of 0·5 to 2 μg/ml in ligase buffer (50 mM-Tris·HCl (pH 8·0), 10 mM-$MgCl_2$, 20 mM-dithiothreitol, 1 mM-ATP, 50 μg bovine serum albumin/ml). T4 DNA ligase was added to a concentration of 200 cohesive end units per ml and incubated overnight at 14°C. DB4572 was transformed to ampicillin resistance.

(d) *β-Galactosidase assays*

Enzyme assays were performed as described by Rose *et al.* (1981). Protein was measured by the method of Bradford (1976) using reagents purchased from Bio-Rad.

β-Galactosidase activity was checked on plates using a buffered minimal medium (Clifton *et al.*, 1978) containing 40 μg X-gal/ml. Blue color appeared either overnight for strains containing *lacZ* gene fusions on 2 μm plasmid vectors or after 1 week for integrated gene fusions.

(e) *DNA sequence analysis of fusion junctions*

Fusion plasmids were cleaved with *Bam*HI and labeled either on the 3′ end of one strand or the 5′ end of the homologous strand. To label the 3′ end, 2 to 3 pmol of plasmid DNA were digested with *Bam*HI and *Hin*dIII, extracted with phenol, precipitated with ethanol, and dissolved in 40 μl of 70 mM-Tris·HCl (pH 8·0), 7 mM-$MgCl_2$ and 1 mM-β-mercaptoethanol plus 25 pmol of [α-^{32}P]dGTP (2000 to 3000 Ci/mmol, Amersham). A sample (1·7 units) of *Micrococcus luteus* polymerase (Miles) was added and the reaction was incubated at 16°C for 30 min. Alternatively, the DNA was cleaved first with *Bam*HI, dephosphorylated with bacterial alkaline phosphatase or calf intestinal phosphatase and the 5′ end was labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (gift from O. Uhlenbeck) by the method of Maxam & Gilbert (1980). The labeled DNA was then cleaved with *Hin*dIII and the appropriate DNA fragments were purified by polyacrylamide gel electrophoresis. DNA sequencing was performed as described by Maxam & Gilbert (1980).

(f) *Construction of frameshifted protein fusions*

Plasmids pRB71 and pRB72 (5 μg each) were digested to completion with endonuclease *Bam*HI, extracted with phenol and precipitated with ethanol. The DNA, resuspended in 50 μl of 10 mM-sodium citrate (pH 7·0), 0·5 mM-Na_2EDTA, 0·5 M-NaCl, was passed over a 0·25 ml column of acridine yellow ED (Buenemann & Mueller, 1978) (Boehringer-Mannheim) to remove contaminating supercoils. The flowthrough, containing the linear DNA, was collected and precipitated with ethanol. The DNA was resuspended in 50 μl of 10 mM-Tris·HCl (pH 7·5), 10 mM-$MgCl_2$, 1 mM-dithiothreitol, 50 mM-NaCl. All 4 deoxyribonucleoside triphosphates were added to 50 μM along with 0·7 unit of the Klenow fragment of DNA polymerase I (New England Biolabs). After incubation at 16°C for 45 min the reaction was terminated by addition of excess Na_2EDTA, extracted with phenol and precipitated with ethanol. The DNA was resuspended at 2 μg/ml in ligase buffer, T4 DNA ligase was added to a final concentration of 20,000 cohesive-end units per ml and the mixture was incubated overnight at 14°C. DB4572 was transformed selecting for ampicillin resistance. Small-scale preparations of plasmid DNA were made from 30 transformants all of which were found to have lost the *Bam*HI site. Fourteen candidate plasmids were transformed into GM48 and plasmid DNA was prepared from this background. All 14 plasmids contain a new *Cla*I site at the same position as the *Bam*HI site in the parent plasmid, confirming that the manipulations had produced the correct structure.

(g) *Construction of integrating fusion plasmids*

Plasmids pRB72 and pRB73 were digested to completion with *Sal*I and then partially digested with *Eco*RI. Digests that contained large proportions of the desired fusion fragment running from the *Eco*RI site in the pBR322 (Bolivar *et al.*, 1977) portion of the plasmid to the *Sal*I site at the end of the *lacZ* segment were ligated into YIp5 (Struhl *et al.*, 1979; Botstein *et al.*, 1979) DNA. Strain DB6566 was transformed with the mixture selecting for Ura$^+$. Plasmid DNAs prepared from Lac$^+$, Amp^r transformants were analyzed by restriction enzyme cleavage.

(h) *Preparation of yeast RNA*

RNA was prepared by a modification of the method of Carlson & Botstein (1982). Yeast cells containing the plasmid pRB73 in either *ura1*$^-$ or *ura3*$^-$ genetic backgrounds were grown to 1×10^7 to 2×10^7 cells/ml in SD medium containing uracil. The cells (100 ml) were harvested by centrifugation in the cold, resuspended in 4 ml cold breaking buffer (0·5 M-NaCl, 200 mM-Tris·HCl (pH 7·5), 10 mM-EDTA). Eight grams of cold acid-washed siliconized glass beads ("Glasperlen", 0·45 mm diameter, VWR Thomas) was added along with diethylpyrocarbonate to 0·1%. Four ml of phenol/chloroform/isoamyl alcohol (25 : 24 : 1, by vol.) was added and then it was vigorously vortexed for 3 min. Debris and glass beads were removed by centrifugation and the supernatant was repeatedly extracted with phenol/chloroform/isoamyl alcohol (25 : 24 : 1, by vol.) until no interface remained. The nucleic acids were precipitated with ethanol, resuspended in 7·5 M-guanidine hydrochloride (Chirgwin *et al.*, 1979), and again precipitated with ethanol by the addition of 0·025 vol. of 1 M-acetic acid, and 0·5 vol. of absolute ethanol. The RNA pellet was rinsed with ethanol, resuspended in water, insoluble material was removed by centrifugation and then precipitated with ethanol once more before resuspending in sterile glass-distilled water. The yield was between 250 and 500 μg of RNA.

(i) *Mapping the 5′ ends of the* URA3 *transcripts*

The method used was essentially the same as described by Berk & Sharp (1977) and modified by Weaver & Weissman (1979). The 5′ end of the *Bam*HI site at the fusion junction was labeled by polynucleotide kinase as described for DNA sequencing. The *Hin*dIII/*Bam*HI fragment corresponding to the 5′ flanking sequence was labeled, purified by polyacrylamide gel electrophoresis and the DNA strands were separated and purified as described by Maxam & Gilbert (1980). Incorporation of ^{32}P was between 2×10^6 and 3×10^6 disints/min per pmol of end. Single-stranded probe (0·05 pmol) was precipitated with ethanol with 25 μg of total yeast RNA, resuspended in a total of 12·5 μl of hybridization buffer (50 mM-PIPES (pH 6·4), 400 mM-NaCl, 1 mM-EDTA and 50% formamide). The mixture was sealed into a glass capillary tube, heated at 85°C for 3 min and then hybridized at 42°C for 3 h. The reaction was chilled in ice-water and then diluted 10-fold into S_1 buffer (final concentration 0·3 M-NaCl, 0·03 M-sodium acetate (pH 4·6), 1 mM-$ZnSO_4$, 20 μg denatured chicken blood DNA/ml). Various amounts of endonuclease S_1 were added, the mixture was incubated at 30°C for 45 min and the reaction was terminated by the addition of excess Na_2EDTA and chilling on ice. A 5 μg sample of yeast transfer RNA was added and the nucleic acids were precipitated twice with ethanol. The material was resuspended in 5 μl of 80% formamide, 89 mM-Tris·HCl, 89 mM-boric acid, 2·5 mM-EDTA, 0·1% bromophenol blue and xylene cyanol. The products of digestion were

electrophoresed on a 0·5 mm thick sequencing gel in 8·3 M-urea as described by Maxam & Gilbert (1980). DNA size standards consisting of the G and T+C chemical sequencing reactions were prepared from the same labeled DNA strand and run on the gel alongside the S_1 digestion products. For digestion with Mung bean nuclease, the hybridization was carried out as above but the reaction was diluted 10-fold into 0·03 M-sodium acetate (pH 4·6), 0·07 M-NaCl, 1 mM-$ZnSO_4$, and 20 μg denatured chicken blood DNA/ml. For digestion by exonuclease VII the hybridization reactions were diluted 10-fold into 50 mM-KCl, 50 mM-Tris·HCl (pH 7·5) and 10 mM-Na_2EDTA. Both of the latter digestion reactions were run for 90 min at 30°C. The gels were autoradiographed for varying times at −70°C using Kodak XAR-5 film and Dupont Cronex Lightning Plus intensifying screens.

3. Results

(a) In vitro *construction of gene fusions of* E. coli lacZ *to the yeast* URA3 *gene*

Gene fusions between the *lacZ* gene and the yeast *URA3* gene were previously obtained by *in vivo* selection for deletions (Rose *et al.*, 1981). We developed an *in vitro* method to generate numerous fusions within the *URA3* gene, which would have one deletion endpoint uniquely defined by a unique restriction site at the end of *'lacZ*. (The notation *'lacZ* indicates deletion of the 5′ end of the *lacZ* coding sequence.)

Plasmid pRB45 (Fig. 2) (Rose *et al.*, 1981) contains a unique *Sma*I restriction

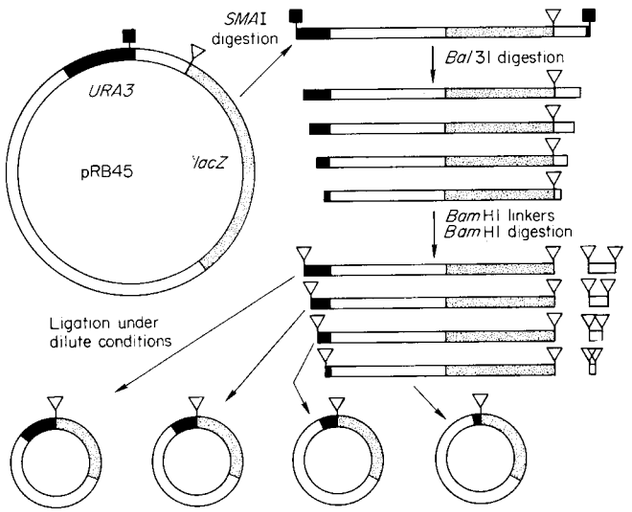

FIG. 2. *In vitro* construction of gene fusions. Plasmid pRB45 (Rose *et al.*, 1981) was treated as described in the text to produce fusions between the *URA3* gene and *'lacZ*. In addition to the sequences indicated, the plasmid also contains a portion of the yeast 2 μm circle (Broach, 1982) and the yeast *LEU2* (Ratzkin & Carbon, 1977) gene for autonomous replication and selection in yeast. The remainder of the plasmid is derived from pBR322 (Bolivar *et al.*, 1977; Sutcliffe, 1979). Both the *URA3* and *'lacZ* segments are oriented with their coding sequences running in a clockwise fashion. The indicated restriction sites are: ▼, *Sma*I; ▽, *Bam*HI.

site just outside the carboxyl end of the coding sequence of the *URA3* gene. A unique *Bam*HI restriction site marks the boundary of the *'lacZ* segment (Casadaban *et al.*, 1980). The *Sma*I site is approximately equidistant from the *Bam*HI site and a *Hin*dIII restriction site present in the 5′ flanking sequence of the *URA3* gene. Digestion with *Sma*I produced linear DNA molecules, which were used as substrate for the double-stranded exonuclease *Bal*31 (Legerski *et al.*, 1978). Synthetic oligonucleotide DNA linkers containing a *Bam*HI site (Bahl *et al.*, 1976) were ligated onto the ends of the shortened linear plasmid molecules. Subsequent digestion with restriction endonuclease *Bam*HI removed both the extra linker molecules that had ligated onto the ends and the remaining yeast DNA flanking the *'lacZ* segment. The restriction endonuclease products were ligated under dilute conditions to promote recircularization of the plasmid molecules. The plasmids were recovered by transformation into *E. coli*, selecting for ampicillin resistance. Transformant colonies were tested for the plasmid-borne abilities to grow in the absence of exogenous uracil (Pyr^+) or on lactose as the sole source of carbon (Lac^+). Out of a total of 568 transformants, 181 (32%) were Lac^+. All but three of the transformants were Pyr^-. Thus virtually all of the plasmids had suffered deletions that disrupt the *URA3* gene, one-third of these had simultaneously activated the *'lacZ* gene.

Plasmid DNA was prepared from 36 representative transformants (including some Lac^- colonies) and analyzed by *Bam*HI and *Hin*dIII restriction endonuclease cleavage. The position of the fusion was determined from the size of the new *Bam*HI/*Hin*dIII "fusion fragment" and the disappearance of a corresponding *Hin*dIII fragment. In all of the plasmids, both the 1·1 kb *Hin*dIII *URA3* fragment and the flanking 0·9 kb *Bam*HI/*Hin*dIII fragment are gone. A new band of variable size is present in each plasmid. Of all the plasmids, 21 had structures consistent with a single deletion extending from the unique *Bam*HI site into the 1·1 kb *Hin*dIII *URA3* gene fragment. Four plasmids had deletions that entered the 0·6 kb *Hin*dIII fragment flanking the *URA3* gene fragment. Eight of the plasmids had suffered extremely large deletions. The remaining plasmids could not be assigned a structure as they either contained two new bands or a new band was not observed. The map of the 21 fusions obtained within the 1·1 kb *Hin*dIII fragment containing the *URA3* gene is shown in Figure 3.

(b) *Expression of* URA3–lacZ *gene fusions in yeast*

Twelve selected fusion plasmids (10 Lac^+, 2 Lac^- in *E. coli*) were introduced into yeast strain DBY689 by DNA transformation, selecting for the plasmid derived yeast *LEU2* gene. The expression of the hybrid β-galactosidase activity in the transformants was assayed and compared to the levels observed from the plasmids in *E. coli* (Table 1). Three classes of fusions were immediately apparent. Nine of the ten fusions that were Lac^+ in *E. coli* also expressed β-galactosidase in yeast (class A). One of the fusions was Lac^+ in *E. coli* but failed to express β-galactosidase in yeast (class B). The two Lac^- fusions failed to express β-galactosidase in either organism (class C). The level of β-galactosidase expressed

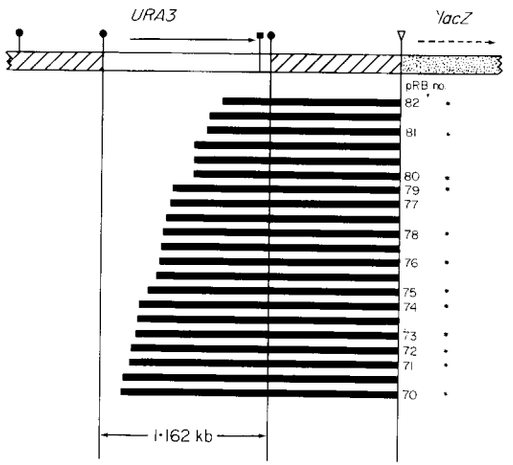

FIG. 3. Map of the deletions generated *in vitro*. The structure of the relevant portions of pRB45 are shown at the top of the Figure. Left to right corresponds to clockwise in Fig. 2. All of the deletion endpoints end at the *Bam*HI site. Deletions marked by asterisks have been defined by DNA sequence analysis of the fusion junction. Deletions with pRB numbers have been analyzed for gene expression in yeast. Restriction sites are: ♦, *Hin*dIII; ▽, *Bam*HI; ■, *Sma*I.

TABLE 1

Expression of in vitro *constructed* URA3–lacZ *gene fusions*

pRB no.	Deletion position	Reading frame	β-Galactosidase activity		Class
			Yeast	*E. coli*	
70	171	+1	<2	<2	C
71	238	+2	112	340	A
72	241	+2	56	380	A
73	275	0	1200	320	A
74	284	0	1670	450	A
75	356	0	570	230	A
76	425	0	405	160	A
78	443	0	115	120	A
79	489	+1	<2	<2	C
80	692	0	220	440	A
81	780	+1	<2	370	B
82	884	0	150	250	A

Extracts were prepared from transformed yeast cells (DBY689) grown in SD minimal medium, supplemented with uracil (40 μg/ml) and the presence of plasmid was selected for by the absence of leucine. *E. coli* cells were grown in LB broth containing ampicillin (100 μg/ml). For both yeast and *E. coli*, assays were performed in duplicate on 2 independent transformants of each plasmid. The reading frame and class are defined in the text. Activities are expressed as nmol of *o*-nitrophenol-β-D-galactoside cleaved per minute per mg protein. For yeast, protein was measured directly; for *E. coli*, the method of Miller (1972) was used and normalized by assuming that a culture of cells at a turbidity of 1·4 O.D. at 600 nm equals 150 μg protein/ml. For all assays, 1 nmol of *o*-nitrophenol corresponds to an $A_{420} = 0{\cdot}0045$.

in yeast varies widely among the members of class A although the levels are fairly uniform within that class in *E. coli*.

In order to clarify the causes for the variation in expression in yeast, we determined the exact DNA sequence of the fusion junctions in the 12 plasmids. The fusion junction is adjacent to the *Bam*HI site derived jointly from the end of the *'lacZ* segment and the *Bam*HI linker introduced within the yeast sequences. The fusion junction is defined as the first nucleotide base-pair derived from the artificial *Bam*HI linker. The *Bam*HI site is always at a fixed position relative to the *lacZ* coding sequence. Therefore, the reading frame into which *'lacZ* has been fused can be determined by counting codons in the 5′ direction starting from one of the invariant base-pairs in the *Bam*HI site. The three possible reading frames within the *Hin*dIII containing the *URA3* gene are defined relative to an ATG codon that begins a long unbroken reading frame of 267 codons without any translation termination codons. This long open reading frame is defined as the "0" frame, henceforth also referred to as the "protein" reading frame. As seen in Figure 1 a second open reading frame beginning with ATG is present in the 5′ flanking sequence of the protein sequence. The second open reading frame could potentially encode a peptide of 28 amino acids before terminating with a TAA codon beginning at the 18th base-pair of the protein reading frame. This second reading frame is in the "+2" frame (i.e. it can be derived from the 0 frame by adding 2 base-pairs) and will be referred to as the "peptide" reading frame.

The expression of hybrid β-galactosidase from the fusion plasmids shows a correlation with the reading frame into which *'lacZ* has been fused (Table 1). The class A fusions, which are expressed in yeast, were obtained in both of the open reading frames. The plasmids pRB71 and pRB72 contain *'lacZ* fused within the 17 base-pairs of overlap between the two reading frames and both place *'lacZ* in frame with the peptide ATG. They are downstream from the protein ATG but not in frame with it. The seven remaining class A fusions all place *'lacZ* into the protein reading frame (0) and are all downstream from the TAA that terminates the peptide. The two class C fusions are not in any open reading frame, thus they do not express β-galactosidase in either organism. The single class B fusion expressed in *E. coli* but not in yeast is not in either of the open reading frames that begin with ATG. However, it is within a 114 base-pair stretch without any termination codons that does contain several TTG and CTG codons that could serve to initiate translation in *E. coli* (Miller, 1974).

(c) *Construction of protein fusions from peptide fusions by shifting the reading frame*

The two class A peptide fusions (pRB71 and pRB72) express hybrid β-galactosidase in yeast at a level corresponding to 5 to 10% of the level of the two nearest class A protein fusions. Lower levels of expression could arise from several sources, particularly if the *URA3* mRNA is spliced so as to start the OMPdecarboxylase protein from the upstream ATG. Many such models can be distinguished by determining whether the peptide fusions have deleted an essential site or sites required for proper expression and regulation of the *URA3*

gene. Therefore, we examined whether *'lacZ* fusions obtained in-frame with the protein ATG but at the same positions as the two peptide fusions would produce high levels of normally regulated β-galactosidase. The construction of these new "frame-shifted" fusions took advantage of the unique *Bam*HI site at the boundary between *'lacZ* and the yeast sequences.

The two peptide fusion plasmids were cleaved with *Bam*HI and the protruding 5′ single-stranded ends were made double-stranded by synthesizing the homologous sequence using DNA polymerase I (Klenow fragment). The resulting blunt ends were ligated together to recircularize the plasmid molecules. Filling in the *Bam*HI site overhanging ends in this manner adds four base-pairs to the fusion junction, shifting the reading frame of *'lacZ* from +2 to 0. Two additional consequences are the creation of a new *Cla*I restriction site at the fusion junction and the destruction of the *Bam*HI site (G-G-A-T-C-C becomes G-G-A-T-C-G-A-T-C-C).

(d) *Regulation of* URA3–lacZ *fusions in yeast*

Plasmids containing the frameshifted *'lacZ* fusions (pRB87 derived from pRB71 and pRB90 derived from pRB72) were introduced into yeast by DNA transformation. To check whether the expression of the hybrid β-galactosidase was under the normal uracil control, the plasmid-bearing strains were crossed to a *ura1* mutant strain (induced for *URA3* gene expression). The resulting diploids were sporulated under conditions selecting for the continued maintenance of the plasmid. Random spores were germinated and their genotypes were determined by standard complementation analysis. Several URA^+, *ura1*, *ura3* and *ura1*, *ura3* strains containing the plasmids were collected and assayed for their level of hybrid β-galactosidase activity.

Table 2 shows the level of hybrid β-galactosidase activity expressed by the plasmids in the different genetic backgrounds, measured before and after uracil starvation. Plasmid pRB73 is a class A fusion of *'lacZ* to the protein reading frame at position +48 with respect to the ATG (outside the region where the two reading frames overlap). In a URA^+ background, pRB73 expresses a fairly high level of β-galactosidase activity, which does not change significantly upon starvation for uracil. The levels expressed in $ura1^-$ strains are 10 to 15-fold higher than the URA^+ basal level. Assuming that the hybrid protein has the wild-type level of β-galactosidase activity (3×10^5 units per mg; Craven *et al.*, 1965) this corresponds to as much as 5% of the soluble cell protein. The levels obtained from $ura3^-$ strains are similar to the URA^+ basal level in the presence of exogenous uracil, but increase about threefold upon starvation for uracil. This pattern of expression parallels the levels of OMPdecarboxylase expressed from the wild-type *URA3* gene, and are similar to the observations for the *URA3–lacZ* gene fusion obtained by *in vivo* methods (Bach *et al.*, 1979; Rose *et al.*, 1981).

The peptide fusion plasmids pRB71 and pRB72 express relatively low levels of β-galactosidase in the wild-type strains, only 5 to 10% of the levels expressed by pRB73. The level of expression shows roughly similar levels in the $ura1^-$ background. The levels of activity show wide variability but the β-galactosidase

TABLE 2

Comparison of the expression and regulation of the protein and peptide fusions

Genotype	β-Galactosidase activity +2 Frame (peptide) fusions		Genotype	β-Galactosidase activity 0 Frame (protein) fusions	
	Grown in uracil	Starved for uracil		Grown in uracil	Starved for uracil
pRB71			pRB87		
URA^+	29 ± 19 (1)	29 ± 24 (1)	URA^+	290 ± 200 (1)	280 ± 160 (1)
$ura1^-$	82 ± 69 (2·5)	99 ± 75 (3)	$ura1^-$	3800 ± 900 (13)	3300 ± 600 (11)
$ura3^-$	57 ± 27 (2)	81 ± 55 (2·5)	$ura3^-$	620 ± 570 (2)	1320 ± 850 (4·5)
$ura1^-, ura3^-$	31 ± 17 (1)	50 ± 53 (1·5)	$ura1^-, ura3^-$	2500 ± 720 (9)	2500 ± 750 (9)
pRB72			pRB90		
URA^+	61 ± 21 (1)	46 ± 22 (1)	URA^+	280 ± 120 (1)	250 ± 110 (1)
$ura1^-$	160 ± 72 (2·5)	145 ± 66 (2·5)	$ura1^-$	5600 ± 1200 (20)	5800 ± 1200 (21)
$ura3^-$	69 ± 36 (1)	54 ± 32 (1)	$ura3^-$	880 ± 370 (3)	1720 ± 650 (6)
$ura1^-, ura3^-$	230 ± 247 (4)	100 ± 100 (2)	$ura1^-, ura3^-$	3030 ± 2700 (11)	2850 ± 2580 (10)
			pRB73		
			URA^+	770 ± 660 (1)	1100 ± 860 (1·4)
			$ura1^-$	8700 ± 3200 (11)	17000 ± 6400 (22)
			$ura3^-$	1700 ± 840 (2)	4900 ± 3100 (6)
			$ura1^-, ura3^-$	6100 ± 1700 (8)	7900 ± 2000 (10)

The indicated plasmids were introduced into the various genetic backgrounds by mating transformants of strain DBY689 with DBY941 as described in Materials and Methods. Extracts were prepared from strains of the indicated genotypes before and after starvation for uracil for 150 min. Specific activities are expressed as in Table 1. Six strains of each genotype were assayed in duplicate and the results reported are the mean values ±S.E.M. pRB87 and pRB90 were derived from pRB71 and pRB72, respectively, by *in vitro* constructed frameshift mutations. pRB73 is a 0 frame fusion outside of the peptide overlap region.

may increase as much as two- to threefold in the *ura1*$^-$ background in contrast to the 10 to 15-fold increase shown by pRB73. Apparently the peptide fusions are not regulated in the same fashion as the nearest protein fusion.

Plasmids pRB87 and pRB90 are the protein fusion plasmids derived from the peptide fusion plasmids pRB71 and pRB72, respectively. Changing the reading frame of the fusion results in a five- to tenfold increase in the basal level of expression of hybrid β-galactosidase in the *URA*$^+$ background. The levels obtained in the *ura1*$^-$ strains are 10 to 20-fold higher than the *URA*$^+$ basal level. These levels do not change upon uracil starvation but do increase in the *ura3*$^-$ background. Thus, shifting the reading frame so that the fusions are expressed from the protein reading frame restores high levels of expression, which are regulated in a manner indistinguishable from that of the other class A protein fusions. Therefore, no site has been deleted that is essential for either the expression or the regulation of the *URA3* gene. The promoter and regulatory sequences must lie upstream from position +11. The peptide reading frame is functional in the sense that it is fully capable of both the transcriptional and translational steps required for the synthesis of the β-galactosidase fusions. The differences in regulation and level of expression suggest that the peptide reading frame is at least partially independent of the *URA3* gene "promoter" and regulatory sites.

(e) *Expression of* URA3–lacZ *fusions integrated into the yeast chromosome*

To determine whether the expression of the fusions is influenced either by the copy number of the yeast vector or by the presence of plasmid sequences upstream from the *URA3* gene, we subcloned the *lacZ* fusions onto an integrating yeast vector, YIp5 (Struhl *et al.*, 1979; Botstein *et al.*, 1979). The structure of one such plasmid is illustrated in Figure 4. The only yeast-derived DNA segments are the 2 kb of DNA upstream from the fusion junction and a copy of the *URA3* 1·1 kb *Hin*dIII fragment, which has been inserted into the *Ava*I site of pBR322 (Bolivar *et al.*, 1977; Sutcliffe, 1979). The plasmids can transform *ura3*$^-$ strains of yeast to uracil prototrophy by integration *via* homologous recombination in either of the two yeast DNA segments. The two modes of integration differ in the resulting topological relationship of the plasmid DNA to the *lacZ* fusion junction. Following integration *via* the upstream yeast segment (illustrated in Fig. 4), only wild-type yeast sequences are present on the 5′ side of *'lacZ*. Integration *via* the YIp5-derived *URA3* gene (not shown) leaves plasmid DNA 5′ to the fusion junction. The two modes of integration were distinguished by means of gel transfer hybridization (Southern, 1975) analysis.

Several transformants of yeast strain DBY629 were obtained in which single copies of the plasmids pRB102 (the integrating derivative of pRB73) or pRB177 (the integrating derivative of pRB71) had integrated *via* the upstream yeast segments. Transformants were subsequently crossed to strain DBY633 (*ura1*-21). Various *ura1*$^-$ and *URA*$^+$ spore colonies that contained the integrated plasmid were collected and assayed for β-galactosidase activity.

Table 3 shows the levels of β-galactosidase activity expressed from the fusions

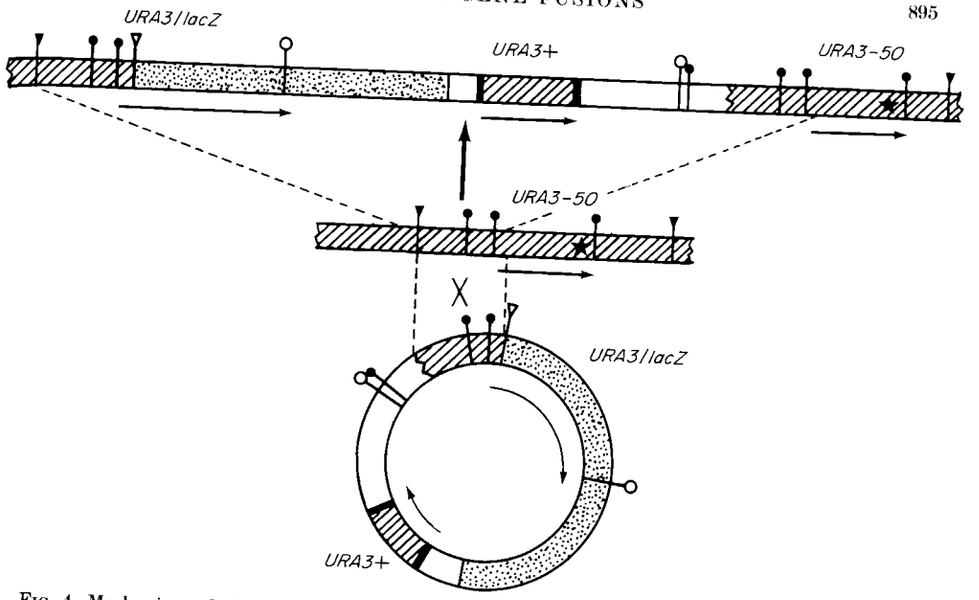

FIG. 4. Mechanism of chromosomal integration of YIp fusion plasmids. An *Eco*RI/*Sal*I fragment carrying all of the *URA3–lacZ* fusion was recombined *in vitro* onto plasmid YIp5. YIp5 contains pBR322 (shown as white) and the *URA3* 1·1 kb *Hin*dIII fragment from yeast strain +D4, which has been inserted into the *Ava*I site. The *Hin*dIII sites flanking the *URA3* fragment have been destroyed in the construction (designated by the black bars). Integration can occur by homologous recombination between the chromosome and either of the yeast sequences on the plasmid. Integration *via* recombination within the segment upstream of the *URA3* gene is shown diagrammatically. Only wild-type yeast sequences are upstream from the gene fusion. Integration *via* recombination within the *URA3* gene (not shown) retains the topology of the plasmid (i.e. plasmid sequences are retained upstream from the gene fusion. The *ura3-50* allele maps near the carboxy terminus and is indicated by a star. Restriction sites are indicated: ▼, *Hin*dIII; ▼, *Bgl*II; ♀, *Eco*RI; ▽, *Bam*HI.

TABLE 3
Expression of peptide and protein fusions integrated into the chromosome

Relevant genotype	β-Galactosidase activity	
	Grown in uracil	Starved for uracil
Protein fusion:		
Wild type	25± 2	37± 6
ura1$^-$	349±31	437±56
Peptide fusion:		
Wild type	3·2± 0·5	3·5± 0·2
ura1$^-$	6·4± 1·2	6·8± 1·9

Yeast strain DBY629 was transformed with pRB102 (protein fusion) and pRB177 (peptide fusion). Transformants were crossed to strain DBY633 (*ura1*$^-$) and spores containing the integrated plasmid were recognized by the color of the colonies formed on X-gal plates. Extracts from 8 *URA*$^+$ and 8 *ura1*$^-$ protein fusion transformants were prepared and assayed as described in Materials and Methods. Six *URA*$^+$ and 8 *ura1*$^-$ peptide fusion transformants were assayed. Activities are reported as nmol *o*-nitrophenol-β-D-galactoside cleaved per min per mg protein ±S.E.M.

integrated into the chromosome. The protein fusion expresses a basal level of expression in wild-type cells, which is some 30-fold lower than the same fusion on the multi-copy 2 μm plasmid. As seen previously on the 2 μm plasmid vector, this level changes only slightly upon starvation for uracil. The levels in the *ura1*$^-$ background are around 15-fold higher than the wild-type basal levels. The peptide fusion also expresses hybrid β-galactosidase activity, again around 30-fold lower than the level observed from the same fusion on the 2 μm plasmid vector. In the *ura1*$^-$ background the level is twofold higher than the basal level in the wild type. Moreover, the peptide fusion is again expressed at about 10% of the level of the protein fusion, as had been observed for expression on the 2 μm vector. Therefore, the expression of the peptide fusions is not attributable to either high copy number or adjacent plasmid sequences upstream.

(f) *Mapping the 5′ ends of the* URA3 *transcripts*

The previous experiments showing the differences in expression and regulation of the *lacZ* fusions suggest that the peptide and the protein reading frames may be expressed from different mRNA species *in vivo*. To test the validity of this hypothesis the 5′ termini of the mRNA transcripts from the *URA3* gene were mapped. In the experiments shown in Figures 5 and 6, total RNA was prepared from a yeast transformant containing plasmid pRB73 and a *ura1*$^-$ strain containing the same plasmid. The RNA was hybridized with a DNA probe prepared from pRB73, which runs from the *Bam*HI site at the fusion junction to the *Hin*dIII site upstream from the *URA3* gene. Single-stranded DNA was prepared with the ^{32}P label at the *Bam*HI end of the molecule, which corresponds to the 5′ end of the anti-sense strand. This end of the probe is derived originally from a DNA linker so that it does not hybridize to the mRNA produced from the chromosomal copy of the *URA3* gene and is therefore specific for the mRNA produced from the fusion on the 2 μm plasmid vector. The RNA : DNA hybrids were digested with the single-strand-specific endonucleases S_1, Mung bean nuclease or the single-strand exonuclease VII. The trimmed hybrids were then denatured and electrophoretically fractionated on a DNA sequencing gel along with the G and T+C tracks from Maxam & Gilbert DNA sequencing reactions performed on the same labeled single DNA strand.

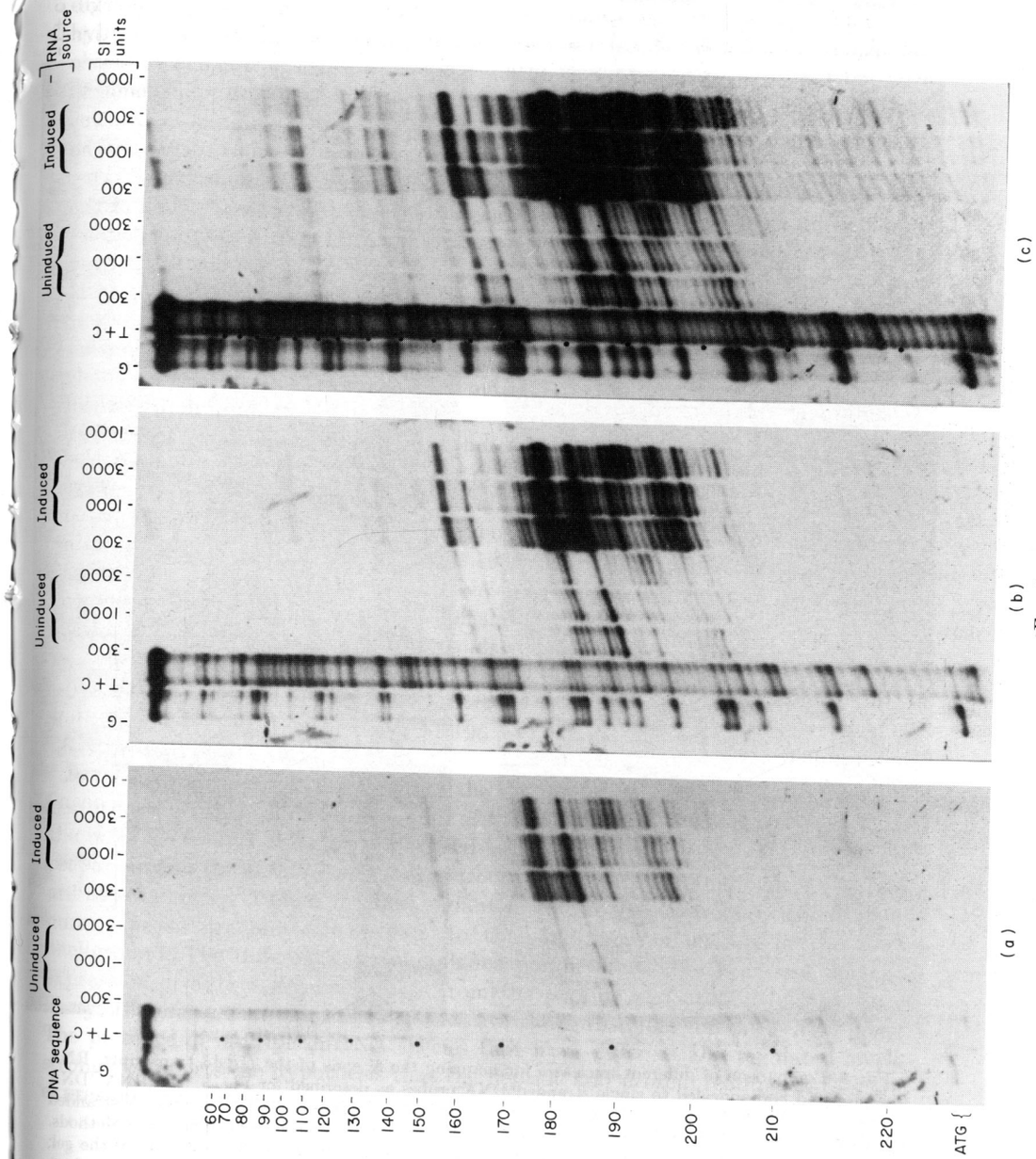

FIG. 5.

FIG. 5. S_1 mapping of the 5′ ends of the *URA3* gene transcripts. Total RNA was prepared from yeast transformants containing plasmid pRB73; 25 μg of RNA was annealed to 0·05 pmol of purified single-stranded *Bam*HI/*Hin*dIII fragment from pRB73. The DNA molecule had been ^{32}P-labeled on the 5′ end (*Bam*HI site) by polynucleotide kinase. After annealing, the hybrid molecules were digested with the indicated quantity of single-strand-specific endonuclease S_1 as described in Materials and Methods. Induced RNA is from a *ura1*$^-$ transformant. Uninduced RNA protected approximately 5% of the label from S_1 digestion, whereas induced RNA protected approximately 50%. Probe incubated with 25 μg of tRNA and treated with 1000 units of S_1 was run in the right-hand lane in each panel. The trimmed RNA : DNA hybrids were fractionated on an 8% (w/v) polyacrylamide gel containing 8·3 M-urea along with the indicated Maxam & Gilbert DNA sequencing reactions performed on the single-stranded DNA probe. The gel was then autoradiographed at −70°C using intensifying screens. (a) A 5-h exposure; (b) a 24-h exposure; (c) a 40-h exposure. The nucleotide positions indicated are designated counting from the *Hin*dIII site upstream from the *URA3* gene. The ATG at the start of the *URA3* coding sequence is indicated, the ATG at the start of the peptide reading frame is at position 160.

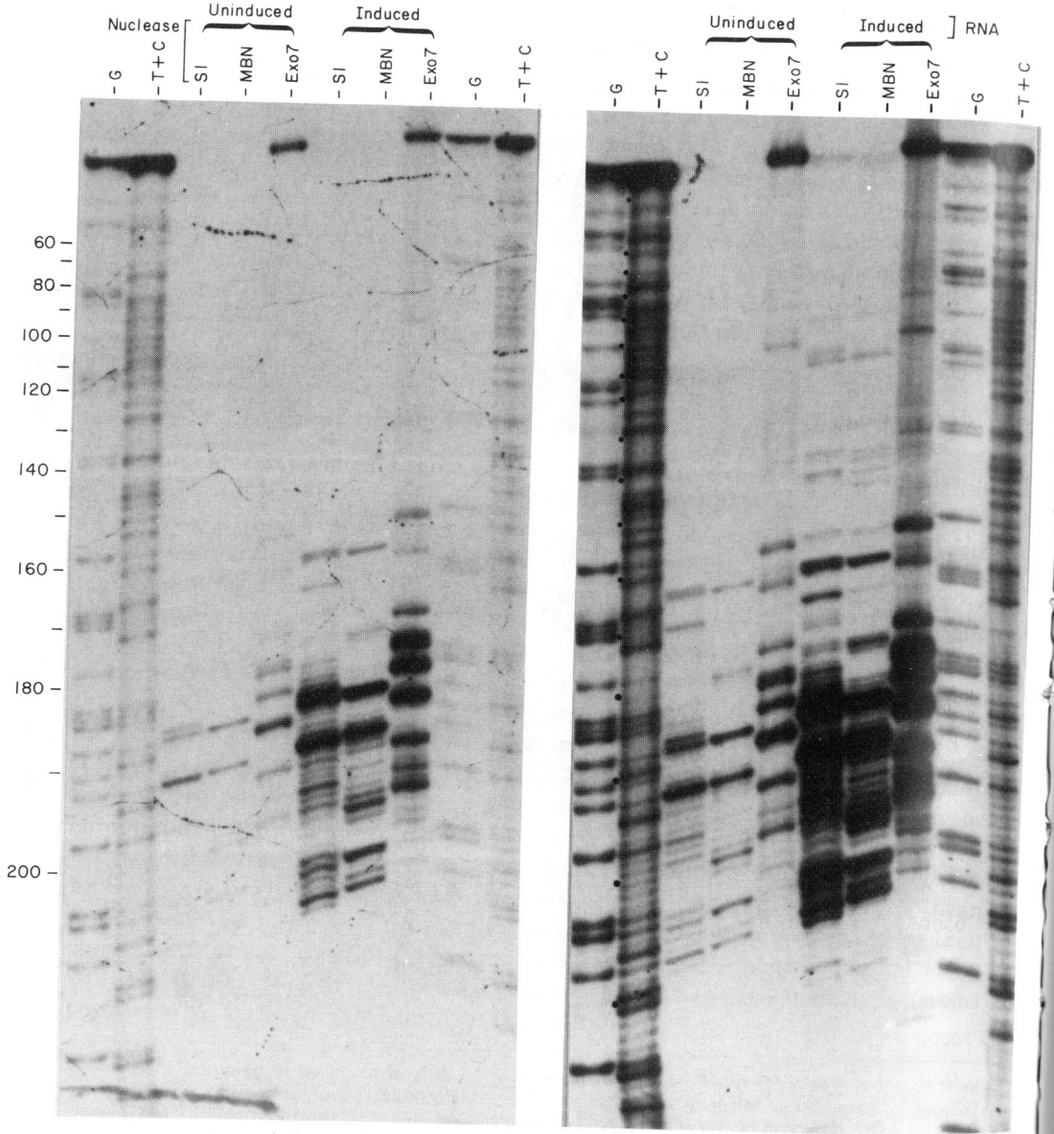

FIG. 6. Comparison of different nucleases for mapping the 5′ ends of the *URA3* transcripts. RNA was prepared and annealed to single-stranded DNA probes as described for Fig. 5. The RNA : DNA hybrids were digested with S_1, Mung bean single-stranded endonuclease (MBN) and single-stranded DNA exonuclease VII (Exo 7). Conditions for the digestions are described in Materials and Methods. Otherwise the experiment is identical to the one described in Fig. 6. (a) A 13-h exposure of the gel; (b) a 48-h exposure. Most of the bands observed from S_1 protection are also seen resulting from protection from Mung bean nuclease. exoVII also shows a similar pattern of protection though in each case the protected band is 4 to 5 base-pairs larger. In the exoVII reactions, a portion of the intact DNA strand is protected from digestion. This protection occurs in the absence of added RNA. This is presumably due to protection of both ends of the molecule either by interstrand annealing or intrastrand hairpin formation.

As can be seen in Figure 5, multiple DNA bands are protected from nuclease S_1 digestion by the RNA species. Increasing the amount of S_1 used to digest the hybrids serves mainly to sharpen the observed bands. Two major bands are observed at positions 189 and 194 (−38 and −33 with respect to the ATG starting the protein reading frame) taking into account the difference in migration between S_1-digested DNA and the products of chemical sequencing (Green & Roeder, 1980). These two major bands are of greatly increased abundance in the *ura1*$^-$ background. A less-abundant species of higher molecular weight is present at position 166, which is close to the canonical 30 bases away from an apparent "TATA" box (Corden *et al.*, 1980) centered at position 133. This species is also of greater abundance in the *ura1*$^-$ background. Longer exposure times of the autoradiographs (Fig. 5(b) and (c)) revealed as many as seven minor RNA species of higher molecular weight that are somewhat more abundant in the *ura1*$^-$ background. The three major species map within the peptide coding sequence, whereas the seven minor species all map upstream from the peptide ATG at position 160. Similar patterns are observed using the enzymes Mung bean nuclease or exonuclease VII (Fig. 6) indicating that the lower molecular weight species are not derived from the higher molecular weight species by RNA splicing. The species protected from exoVII are all four to five base-pairs longer than the S_1 or Mung bean nuclease-protected species, as has been previously observed (Nasmyth *et al.*, 1981; Donahue *et al.*, 1982). Similar S_1-protection patterns are also observed using poly(A)-selected RNA purified from cells containing the intact *URA3* gene carried on the 2 μm plasmid (not shown). In addition to the transcripts mentioned above several bands are observed of lower molecular weight than the prominent band at position 189. Some of these are only observed at increased concentrations of S_1 and may be due to overdigestion at the ends of the hybrids. Others, such as the bands at 196 and 205 may be produced by authentic transcripts.

To determine the relative levels of the different transcripts experiments were performed in which different amounts of the DNA : RNA hybrids prepared from the *ura1*$^-$ strain were electrophoresed adjacent to the DNA : RNA hybrids prepared from the *URA1*$^+$ strain. This revealed that the two major RNA species are approximately 10-fold more abundant in the *ura1* background, whereas the minor species are only about two- to threefold more abundant in the *ura1* background. The differences in abundance parallel the differences seen for the expression of hybrid β-galactosidase from the peptide and protein fusions. It is likely, therefore, that the major species are the transcripts for the *URA3* gene. The presence of multiple mRNA species that have differentially regulated levels strongly supports the hypothesis that the peptide and protein reading frames are expressed from different mRNA species.

4. Discussion

By an *in vitro* method, fusions of the *'lacZ* gene have been obtained throughout the 1·1 kb *Hin*dIII fragment containing the yeast *URA3* gene. Most of the plasmids that produce hybrid β-galactosidase in yeast contain the *'lacZ* segment

fused into a single long open reading frame encoding the enzyme OMPdecarboxylase. Two other plasmids contained fusions of *'lacZ* to an open reading frame that overlaps with the start of the *URA3* coding sequence. These latter fusions also express β-galactosidase in yeast, at about 10% of the level of the nearest *URA3* gene fusions. In yeast, the *URA3–lacZ* gene fusions were regulated in the same fashion as the intact *URA3* gene. Fusions to the overlapping reading frame were slightly elevated in the genetic background that greatly induces *URA3* gene expression. The differences in the levels of expression and regulation appear to be due to differential regulation of a subset of the multiple mRNA species that are transcribed from the *URA3* gene. These data suggest that the yeast *URA3* gene segment contains a small overlapping gene that is expressed in yeast, similar to the overlapping genes found in prokaryotic viruses (Barrell *et al.*, 1976; Shaw *et al.*, 1978).

(a) *All transcription initiation and regulation sites are present at the 5′ end of the* URA3 *gene*

Fusions of the *lacZ* gene to the *URA3* gene have been obtained as close as 11 base-pairs away from the start of the *URA3* coding sequence. These fusions are expressed and regulated in the same fashion as the *URA3* gene. Therefore, it is unlikely that the internal sites are required for the expression and regulation of this gene in yeast. This result for a gene presumably transcribed by polymerase II is in contrast to the result obtained for polymerase III-transcribed genes (e.g. *Xenopus laevis* 5 S RNA; Bogenhagen *et al.*, 1980; Sakonju *et al.*, 1980). A similar result has been obtained for the yeast cytochrome *c* gene (Guarente & Ptashne, 1981). Chevallier *et al.* (1980) have demonstrated that the *URA3* gene is normally regulated from the 1·1 kb *Hin*dIII fragment. Therefore, the *URA3* regulatory sequences are localized to within 240 base-pairs.

(b) *Gradient of β-galactosidase activity according to position of fusion*

One potentially interesting phenomenon is that fusions obtained near the amino terminus of the *URA3* gene express higher levels of β-galactosidase in yeast than fusions obtained near the carboxy terminus. In *E. coli* however, all of the fusions produce about the same level of β-galactosidase implying that the inherent specific activities of the proteins are the same. Several possible differences between the yeast and *E. coli* environments that might lead to this effect are apparent. One is the presence of the mutant OMPdecarboxylase subunits in the cytoplasm that might prevent β-galactosidase tetramer formation by binding to the amino-terminal portion of the hybrid protein. β-Galactosidase is only active as a tetramer (for a review, see Zabin & Fowler, 1980) whereas OMPdecarboxylase is a dimer (Brody & Westheimer, 1978). Another possibility is that the yeast proteases are more efficient in degrading hybrid proteins containing large regions of partially denatured protein. Fusing β-galactosidase into the middle of a folding domain might lead to a bridge of denatured protein, whereas fusing to the amino terminus would allow folding into a protease-resistant structure. Another possibility is that an internal site has been deleted that is responsible for the repression of the gene expression from *URA3*. This repression must be different from the normal regulation of the *URA3* gene, which is intact in the amino-terminal fusions.

(c) *Differentially regulated* URA3 *transcripts*

The regulation of the yeast *URA3* gene has been shown to be largely at the level of transcription (Bach *et al.*, 1979). In this paper we have shown that at least 10 different RNA transcripts are present that are homologous to the beginning of the *URA3* gene as shown diagrammatically in Figure 7. The regulation seems to be due to the increased abundance of a subset of the mRNA species. The most abundant RNAs, which are the most proximal to the *URA3* coding sequence, are also the most amplified in the *ura1*$^-$ background. As the mRNA 5′ ends begin more distally, both their abundance and their induction in the *ura1*$^-$ strains decrease. Thus, the RNAs that are long enough to include the peptide coding sequence are only about twofold more abundant in *ura1*$^-$ strains. The presence of heterogeneous mRNA ends has been described for other yeast genes such as *CYC1* (Faye *et al.*, 1981), *HIS1* (Hinnebusch & Fink, 1983), the *MAT* locus (Nasmyth *et al.*, 1981), and *SUC2* (Carlson & Botstein, 1982). For all of these cases the ends have been examined only by indirect means such as S_1 protection or primer extension. Thus, it is not known whether the observed RNA species represent the primary transcripts or whether they have been modified or degraded. In one case the same RNA preparation was found to yield a single RNA species for one gene (*HIS4*) and multiple species for another gene (*HIS1*; Hinnebusch & Fink, 1983). In the case of *SUC2* the abundance of the longer transcript increases dramatically upon derepression, whereas the shorter transcripts do not change. This leads to

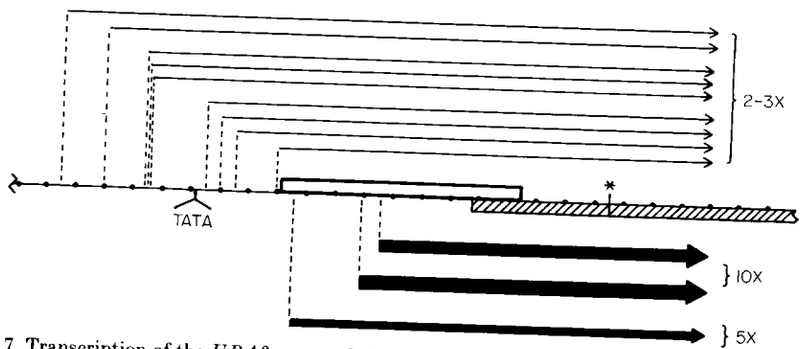

FIG. 7. Transcription of the *URA3* gene and the overlapping peptide sequence. The sequences at the 5′ end of the *URA3* gene are shown diagrammatically. The *URA3* coding sequences are indicated by the crosshatched bar. The overlapping peptide sequence is indicated by the open bar. Major transcripts for the *URA3* gene are indicated below the line, whereas the minor transcripts for the peptide are indicated above the line. The approximate levels of induction for the various transcripts are indicated. Ten base-pair intervals are indicated on the line. The location of the artificial *Bam*HI site used for the S_1 protection experiments is indicated by the asterisk. The position of the putative TATA box is at position 133, as indicated.

the differential expression of proteins containing different amino termini (Carlson & Botstein, 1982). For the other yeast genes mentioned the different transcripts have all been reported to change co-ordinately with regulation. The longer *URA3* transcripts are minor species so that it is not possible to determine whether they are translated to produce OMPdecarboxylase as well as the peptide. For the majority of its expression the *URA3* gene must be translated from the first ATG on the mRNA as expected from the current models of translation initiation (Sherman & Stewart, 1975; Sherman *et al.*, 1980; Kozak, 1981).

One considerable advantage of the *lacZ* gene fusion technique is that it is a relatively simple means of determining whether a given DNA segment is functional for both transcription and translation. In the case of the overlapping reading frame at the start of the *URA3* gene, *lacZ* gene fusions indicate that both transcription and translation initiation occur. Therefore, it is likely that this DNA segment is indeed expressed *in vivo* in the intact state. The pattern of expression of the hybrid β-galactosidases leads to the specific prediction of multiple mRNA species for that region of the DNA that are differentially regulated. That prediction was confirmed by mapping the 5′ ends of the mRNAs. The same pattern of multiple RNA transcripts is present for the intact *URA3* gene and the *URA3–lacZ* fusion when each is present on the 2 μm plasmid (data not shown). In addition, the peptide gene fusions are expressed in single copy number on the chromosome. It is likely, therefore, that expression from the gene fusion is an accurate representation of the expression from the intact *URA3* gene on the chromosome. Of course, direct mapping of the RNA transcripts from the intact *URA3* gene on the chromosome will be necessary to confirm this point. Further confirmation for the expression of the overlapping gene must await either the identification of the protein product or the isolation of nonsense mutants that lead to an observable phenotype. Either procedure will be made much easier by the existence of the *lacZ* fusions. The hybrid β-galactosidase can be used as an immunogen to obtain antibodies against the peptide as was done for an *E. coli* membrane protein (Shuman *et al.*, 1980). In addition, the fusion plasmids can be used to construct the necessary nonsense mutations *in vitro* using the β-galactosidase activity in *E. coli* as a signal.

(d) *Possible functions of the overlapping peptide gene*

The possible functions of the peptide expressed from the overlapping gene are as various as they are unknown. The possibility that the peptide may somehow be involved in the regulation of the *URA3* gene is tantalizing. However, one argument cautions against this possibility. The *URA3–lacZ* fusions that were constructed in the overlap region must be making a mutant peptide, yet they are regulated normally. This point is weakened by the observation that the peptide would essentially only be lengthened by eight amino acids and would therefore differ only in the specific carboxy terminus. Moreover, one copy remains on the chromosome. Thus, it would appear, at least, that a completely wild-type peptide sequence is not needed *in cis* for normal *URA3* gene expression. One possibility is that the peptide serves no present function and is a relic. For example, the enzyme OMPdecarboxylase might once have been produced in two forms differing in their amino-terminal sequences and produced under different conditions. A single frameshift mutation in front of the ATG at the start of *URA3* could have destroyed that situation. An important consideration concerns the likely frequency of expression of the peptide. The yeast *URA3* mRNA is present at about 3·3 copies per cell (Lacroute *et al.*, 1981, cited by Chevallier, 1982) and has a half-life of around 10 minutes (Bach *et al.*, 1979). With a cell-doubling time of 150 minutes the *URA3* mRNA must be transcribed at least 30 times per cell cycle. If we take the relative abundance of the transcripts containing the peptide reading frame to be only 10% of the shorter transcripts and assume similar parameters for the decay then the peptide encoding mRNA must be produced about three times per cell cycle. This would seem to approach a lower limit of frequency of transcription and abundance for genes whose expression is required at least once during the cell cycle. Moreover, expression may be higher under different metabolic conditions, Again, the only means for assessing the functional importance of the peptide lies in the phenotypes obtained from mutants.

We thank David Shortle and Malcolm Casadaban for helpful discussions. M.R. was supported by a National Science Foundation Fellowship and an NIH traineeship (GM07287). This work was supported by a grant from the American Cancer Society (MV90) and the National Institutes of Health (GM18973 and GM21253).

REFERENCES

Bach, M. L., Lacroute, F. & Botstein, D. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 386–390.
Bahl, C. P., Marians, K. J., Wu, R., Stawinsky, J. & Narang, S. (1976). *Gene*, **1**, 81–92.
Barrell, B. G., Air, G. M. & Hutchinson, C. A. III (1976). *Nature (London)*, **264**, 34–41.
Berk, A. J. & Sharp, P. A. (1977). *Cell*, **12**, 721–732.
Bogenhagen, D. F., Sakonju, S. & Brown, D. D. (1980). *Cell*, **19**, 27–35.
Bolivar, F., Rodriguez, R. L., Betlach, M. C. & Boyer, H. W. (1977). *Gene*, **2**, 75–94.
Botstein, D., Falco, S. C., Stewart, S., Brennan, M., Scherer, S., Stinchcomb, D., Struhl, K. & Davis, R. W. (1979). *Gene*, **8**, 17–24.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Broach, J. R. (1982). In *The Molecular Biology of the Yeast Saccharomyces* (Strathern, J., Jones, E. & Broach, J., eds), pp. 445–470, Cold Spring Harbor Laboratory, New York.
Brody, R. S. & Westheimer, F. H. (1978). *J. Biol. Chem.* **254**, 4238–4244.
Buenemann, H. & Mueller, W. (1978). In *Affinity Chromatography* (Hoffman-Ostenhoff, O., *et al.*, eds), pp. 353–356, Pergamon Press, New York.
Carlson, M. & Botstein, D. (1982). *Cell*, **28**, 145–154.
Casadaban, M. J., Chou, J. & Cohen, S. N. (1980). *J. Bacteriol.* **143**, 971–980.
Chevallier, M. R. (1982). *Mol. Cell Biol.* **2**, 977–984.
Chevallier, M. R., Bloch, J. C. & Lacroute, F. (1980). *Gene*, **11**, 11–19.
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979). *Biochemistry*, **18**, 5294–5299.
Clewell, D. B. & Helinski, D. R. (1969). *Proc. Nat. Acad. Sci., U.S.A.* **62**, 1159–1166.
Clifton, D., Weinstock, S. & Fraenkel, D. G. (1978). *Genetics*, **88**, 1–11.
Corden, J., Wasylyk, B., Buchwilder, A., Sassome-Corsi, P., Kedinger, C. & Chambon, P. (1980). *Science*, **209**, 1406–1414.
Craven, G. R., Steers, E. Jr & Anfinsen, C. B. (1965). *J. Biol. Chem.* **240**, 248–2477.
Donahue, T. F., Farabaugh, P. J. & Fink, G. R. (1982). *Gene*, **18**, 47–59.
Faye, G., Leung, D. W., Tatchell, K., Hall, B. & Smith, M. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 2258–2262.

Green, M. R. & Roeder, R. G. (1980). *Cell*, **22**, 231–242.
Guarente, L. & Ptashne, M. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 2199–2203.
Hinnebusch, A. G. & Fink, G. R. (1983). *J. Biol. Chem.* **258**, 5238–5247.
Hinnen, A., Hicks, J. B. & Fink, G. R. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 1929–1933.
Kozak, M. (1981). *Nucl. Acids Res.* **9**, 5233–5253.
Lacroute, F. (1968). *J. Bacteriol.* **95**, 824–832.
Lacroute, F., Bach, M. L., Chevallier, M. R., Hubert, J. C., Losson, R., Botstein, D. & Loison, G. (1981). In *Molecular Genetics of Yeast* (von Wettstein, D., Friis, J., Kielland-Brandt, M. & Stenderup, A., eds), *Alfred Benzon Symp.*, vol. 16, Munksgaard, Copenhagen.
Legerski, R. J., Hodnett, J. L. & Gray, H. B. Jr (1978). *Nucl. Acids Res.* **5**, 1445–1464.
Mandel, M. & Higa, A. (1970). *J. Mol. Biol.* **53**, 159–162.
Marinus, M. G. (1973). *Mol. Gen. Genet.* **127**, 47–55.
Maxam, A. M. & Gilbert, W. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 499–560, Academic Press, New York.
Miller, J. H. (1972). *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Miller, J. H. (1974). *Cell*, **1**, 73–76.
Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. (1981). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 961–980.
Peacock, A. C. & Dingman, C. W. (1968). *Biochemistry*, **7**, 668–674.
Ratzkin, B. & Carbon, J. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 487–491.
Rigby, P. W. J., Dickmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–251.
Rose, M. (1982). Ph.D. thesis, M.I.T.
Rose, M., Casadaban, M. J. & Botstein, D. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 2460–2464.
Sakonju, S., Bogenhagen, D. F. & Brown, D. D. (1980). *Cell*, **19**, 13–25.
Shaw, D. C., Walker, J. E., Northrop, F. D., Barrell, B. G., Godson, G. N. & Fiddes, J. C. (1978). *Nature (London)*, **272**, 510–515.
Sherman, F. & Stewart, J. W. (1975). In *Organization and Expression of the Eukaryotic Genome* (Bernardi, G. & Gros, F., eds), pp. 175–191, North Holland/American Elsevier, Amsterdam.
Sherman, F., Fink, G. R. & Lawrence, C. W. (1979). *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Sherman, F., Stewart, J. W. & Schweingruber, A. M. (1980). *Cell*, **20**, 215–222.
Shuman, H. A., Silhavy, T. J. & Beckwith, J. R. (1980). *J. Biol. Chem.* **255**, 168–174.
Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.
Struhl, K., Stinchcomb, D. T., Scherer, S. & Davis, R. W. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 1035–1039.
Sutcliffe, J. G. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
Tabak, H. F. & Flavell, R. A. (1978). *Nucl. Acids Res.* **5**, 2321–2332.
Weaver, R. F. & Weissman, C. (1979). *Nucl. Acids Res.* **7**, 1175–1193.
Zabin, I. & Fowler, A. V. (1980). In *The Operon* (Miller, J. & Reznikoff, W., eds), pp. 89–121, Cold Spring Harbor Laboratories, Cold Spring Harbor, New York.

Edited by I. Herskowitz