

## [9] Construction and Use of Gene Fusions to *lacZ* ( $\beta$ -Galactosidase) That Are Expressed in Yeast

By MARK ROSE and DAVID BOTSTEIN

Gene fusion is a powerful tool for the analysis of the expression, regulation, and structure of a gene. Typically the regulatory and/or structural sequences of the gene of interest are covalently linked to the structural sequences of another gene whose protein is easily assayed. Many of the uses for gene and operon fusions in the study of prokaryotic genes are discussed in the excellent review of Bassford *et al.*<sup>1</sup> The use of gene fusions has been extended to the analysis of the genes of the simple eukaryote *Saccharomyces cerevisiae*.<sup>2,3</sup>

The advent of techniques for DNA sequence analysis has raised new problems that are particularly amenable to solution by gene fusion techniques. For example, the discovery of an open translational reading frame raises the question of whether the sequence is functionally expressed *in vivo*. By fusing a known assayable protein to the sequence, it becomes possible to detect its expression. In addition, the hybrid protein provides a means for the purification of the gene product normally encoded by the sequence.<sup>4</sup> Two other uses derive from the novel junction of the DNA segments that are joined in the fusion process. Gene fusions necessarily delete DNA sequences that are internal to the structural gene of interest and therefore can be used to determine the position of any internal sites that might be essential for gene expression. Furthermore, the novel junction can be used to introduce a useful restriction enzyme recognition site, which can in turn be used to accelerate DNA sequence analysis.

The protein most commonly used for gene fusions has been the enzyme  $\beta$ -galactosidase. There are several advantages to this enzyme. The assay for the activity is extremely sensitive and easy both as a plate assay, in which expressing colonies are colored, and as a quantitative enzyme

<sup>1</sup> P. J. Bassford, J. Beckwith, M. Berman, E. Brickman, M. Casadaban, L. Guarente, I. Saint-Girons, A. Sarthy, M. Schwartz, H. Shuman, and T. Silhavy, in "The Operon" (J. H. Miller and W. S. Reznikoff, eds.), p. 245. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1978.

<sup>2</sup> M. Rose, M. J. Casadaban, and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2460 (1981).

<sup>3</sup> L. Guarente and M. Ptashne, *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2199 (1981).

<sup>4</sup> H. A. Shuman, T. J. Silhavy, and J. R. Beckwith, *J. Biol. Chem.* **255**, 168 (1980).

assay in cell extracts.<sup>5</sup> It has been demonstrated that as many as 30 amino acids at the amino terminus of the protein may be replaced with other sequences of varying length without substantially affecting the activity of the enzyme.<sup>6</sup> Various techniques have been devised that select for the presence or the absence of the enzyme in *Escherichia coli*.<sup>5</sup> These facts have led several workers to construct mutations of the *lacZ* gene (which encodes  $\beta$ -galactosidase) in which the protein cannot be expressed unless it is first fused to the beginning of another gene.<sup>7,8</sup> We have made extensive use of one such defective gene (called '*lacZ*') constructed by Casadaban and Cohen,<sup>7</sup> to investigate the expression of the sequences around the yeast *URA3* gene.<sup>9</sup>

### Principles

Two general methods for the construction of sets of gene fusions have been used: *in vivo* selection<sup>2</sup> and *in vitro* manipulation of the DNA. In either case the sequence of interest is placed onto a plasmid containing the unexpressed '*lacZ*' gene. The DNA fragments are oriented so that the putative coding sequences would be read from the same DNA strand, with the  $\beta$ -galactosidase coding sequence on the 3' side of the sense strand. *In vivo* selection depends on spontaneous deletions formed in plasmids residing in *E. coli*. If the original construction has a Lac<sup>-</sup> phenotype, selection for Lac<sup>+</sup> results in fusion of the '*lacZ*' segment to sequences in frame with a functional initiation codon. Since the sequence of interest may not contain a functional *E. coli* promoter, a plasmid promoter must be situated so as to provide transcription of the selected hybrid gene.

The *in vitro* approach relies on the presence of unique restriction sites in the gene of interest and at the beginning of the '*lacZ*' segment. Deletion of material between the gene and '*lacZ*' is accomplished by means of the double-stranded DNA exonuclease *Bal31* acting on the ends produced within or near the gene by an appropriate restriction endonuclease. Ligation using DNA linkers containing the restriction site at the beginning of '*lacZ*' allows fusion between the two genes. Plasmids recovered in *E. coli* are then transferred into yeast for analysis of the expression of the hybrid protein.

<sup>5</sup> J. H. Miller, "Experiments in Molecular Genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1972.

<sup>6</sup> E. Brickman, T. J. Silhavy, P. J. Bassford, Jr., H. A. Shuman, and J. R. Beckwith, *J. Bacteriol.* **139**, 13 (1979).

<sup>7</sup> M. J. Casadaban, J. Chou, and S. N. Cohen, *J. Bacteriol.* **143**, 971 (1980).

<sup>8</sup> L. Guarente, G. Lauer, T. M. Roberts, and M. Ptashne, *Cell* **20**, 543 (1980).

<sup>9</sup> M. Bach, F. Lacroute, and D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.* **76**, 386 (1979).

## Materials and Reagents

*Escherichia coli* strains. Strains M182, (*lacIPOZYA*)  $\Delta X74$  *galU galK strA* (from J. Beckwith) and a derivative containing an insertion of Tn5 in the *pyrF* gene have been used. Any other extensive deletion of the *lac* operon should serve as well.

*Media.* *Escherichia coli* strains containing a plasmid conferring ampicillin resistance are propagated on LB<sup>5</sup> supplemented with 100  $\mu$ g of ampicillin per milliliter. Selection for Lac<sup>+</sup> derivatives of plasmid containing strains is accomplished on M-9 minimal medium<sup>5</sup> containing 0.2% lactose. Strains that produce as few as 10 units of  $\beta$ -galactosidase are able to grow on M-9 lactose plates. Higher levels of  $\beta$ -galactosidase can be distinguished on MacConkey-lactose plates (Difco) containing ampicillin. Intermediate levels of expression can be distinguished by supplementing media with 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-gal; Bachem Inc., Torrance, California) to 40  $\mu$ g/ml. The X-gal is dissolved in *N,N*-dimethylformamide at a concentration of 20 mg/ml and can be stored at  $-20^\circ$ . The intensity of the blue color varies with the level of  $\beta$ -galactosidase in the cell, the rate of cell growth, and the type of medium. The most intense color is obtained on M-9; less color is obtained on the richer media such as LB or MacConkey.

Yeast strains containing plasmids are propagated on SD minimal medium<sup>10</sup> with constant selection for the presence of the plasmid. SD is composed of 6.7 g per liter of yeast nitrogen base (without amino acids, Difco) supplemented to 2% glucose added after autoclaving. Solid SD medium is made by mixing sterile double-strength SD with an equal volume of molten 4% agar. The expression of  $\beta$ -galactosidase activity is observed on agar plates buffered at pH 7.0 and supplemented with X-gal at 40  $\mu$ g/ml. The acid media (such as SD) usually employed in culturing yeasts do not allow color development. Previously we have used SD medium buffered by the addition of potassium phosphate. Better results have been obtained using the buffered minimal medium of Clifton *et al.*<sup>11</sup> Buffered minimal medium contains a final concentration of 0.1 M  $\text{KH}_2\text{PO}_4$ , 0.015 M  $(\text{NH}_4)_2\text{SO}_4$ , 0.075 M KOH, 0.8 mM  $\text{MgSO}_4$ , 2  $\mu$ M  $\text{Fe}_2(\text{SO}_4)_3$ . After autoclaving, sterile glucose is added to 2%, and  $\frac{1}{100}$  volume of a vitamin stock is added. The vitamin stock consists, per milliliter, of 40  $\mu$ g of thiamine, 40  $\mu$ g of pyridoxine, 40  $\mu$ g of pantothenic acid, 200  $\mu$ g of inositol, and 2  $\mu$ g of biotin. Solid medium is made by making up the salts at double

<sup>10</sup> F. Sherman, G. R. Fink, and C. W. Lawrence, "Methods in Yeast Genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1979.

<sup>11</sup> D. Clifton, S. Weinstock, and D. G. Fraenkel, *Genetics* **88**, 1 (1978).

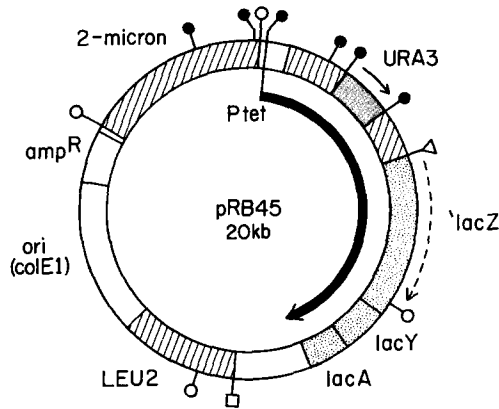


FIG. 1. Structure of plasmid pRB45. Yeast sequences, except for the *URA3* gene, are hatched. The stippled area is derived from the *Escherichia coli lac* operon and was obtained from plasmid pMC874.<sup>7</sup> Other areas are derived from pBR322.<sup>14,15</sup> The tetracycline-resistance gene promoter initiates transcription in a clockwise direction. The *URA3* gene and the '*lacZ*' gene are both oriented so that clockwise transcription is from the sense strand. Restriction enzyme sites are as follows: ○, *EcoRI*; ●, *HindIII*; ▽, *BamHI*; □, *SmaI*.

strength and adding to an equal volume of sterile molten 4% agar. X-gal is added just before pouring the plates, as for the bacterial media.

**Plasmids.** All fusions between the *URA3* gene and *lacZ* were obtained from the plasmid pRB45,<sup>2</sup> shown in Fig. 1. The '*lacZ*' fragment is derived from pMC874.<sup>7</sup> Plasmid pRB45 contains a unique *BamHI* site at the beginning of '*lacZ*' and a unique *SmaI* site near the end of the *URA3* gene. (Other derivatives of the '*lacZ*' segment with different unique restriction sites at the 5' end have been constructed by M. Casadaban<sup>11a</sup> and Guarente *et al.*<sup>8</sup>) The other relevant features are a portion of the yeast 2- $\mu$ m (also called 2-micron) plasmid<sup>12</sup> for autonomous replication in yeast, the yeast *LEU2* gene<sup>13</sup> which serves as a selectable marker in yeast and the pBR322-derived ampicillin resistance gene and *E. coli* origin of replication.<sup>14,15</sup> In *E. coli* the *URA3* gene is transcribed by the tetracycline promoter derived from pBR322 (unpublished results).

**Enzymes and Nucleic Acids.** All the restriction enzymes and T4 DNA ligase were purchased from New England BioLabs and were used according to the recommendations of the manufacturer. The enzyme *Bal31* was

<sup>11a</sup> M. Casadaban, this series, Vol. 100 [21].

<sup>12</sup> J. L. Hartley and J. E. Donelson, *Nature (London)* **286**, 860 (1980).

<sup>13</sup> B. Ratzkin and J. Carbon, *Proc. Natl. Acad. Sci. U. S. A.* **74**, 487 (1977).

<sup>14</sup> F. Bolivar, R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow, *Gene* **2**, 95 (1977).

<sup>15</sup> JG. Sutcliffe, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77 (1979).

purchased from Bethesda Research Laboratories and used according to the conditions described by Legerski *et al.*<sup>16</sup> *Bam*HI linkers<sup>16a</sup> were obtained from Collaborative Research Inc. T4 polynucleotide kinase was the generous gift of O. Uhlenbeck.

### *In Vivo* Selection of Fusions

To express  $\beta$ -galactosidase activity, '*lacZ*' must be fused to a DNA segment so that an initiation codon is provided upstream in the correct reading frame without any intervening nonsense codons. Starting with a plasmid constructed as described under Principles, the simplest way to obtain such fusions is to select the rare cells in which the plasmid suffered a spontaneous deletion event that allows the '*lacZ*' sequence to express  $\beta$ -galactosidase. This can be done by inserting the plasmid into a strain carrying a deletion of the chromosomal *lacZ* gene and then demanding that the strain grow on lactose.

The original unfused construction containing '*lacZ*' behind the gene of interest must not express sufficient  $\beta$ -galactosidase for the cell to grow on lactose. It has been our experience that some such constructions are Lac<sup>+</sup>. This property will prevent selection of fusions to the sequence of interest. One approach to remedy this problem is to change the reading frame to which the '*lacZ*' has been fused by insertion of extra nucleotides at the beginning of the '*lacZ*' segment. When this is done, the phenotype usually becomes Lac<sup>-</sup>.

Nonreverting point mutations of *lacZ* do not suffice as the method of inactivating the chromosomal source of  $\beta$ -galactosidase. The homologous sequences on the plasmid will result in a high frequency of Lac<sup>+</sup> recombinants. It is likely that most deletion events are catalyzed by the *E. coli* recombination system so that the strain should be Rec<sup>+</sup>. The cell must also be *lacY*<sup>+</sup> so that lactose can enter the cell by the permease. Most of the '*lacZ*' fragments include the *lacY* gene so that it is expressed by plasmid promoters along with *lacZ* and is not subject to *lacI* repression.

The second requirement for expression of the  $\beta$ -galactosidase activity is that there be transcription of the hybrid gene in *E. coli*. To ensure this, it is convenient to orient the fragments such that they are transcribed by one of the plasmid promoters. A simple test for whether the plasmid transcription will be sufficient is to construct the analogous plasmid using an intact *lacZ* gene instead of '*lacZ*'. The expression of  $\beta$ -galactosidase from the control plasmid is a measure of the level of transcription impinging on the *lacZ* gene.

<sup>16</sup> R. J. Legerski, J. L. Hodnett, and H. B. Gray, Jr., *Nucleic Acids Res.* **5**, 1445 (1978).

<sup>16a</sup> C. P. Bahl, K. J. Mariani, R. Wu, J. Stawinsky, and S. Narang, *Gene* **1**, 81 (1976).

Prior to selection for fusions many independent clones of the strain carrying the plasmid are grown up to saturation in LB broth containing antibiotics to select for the presence of the plasmid. One milliliter of each culture (around  $10^9$  cells) is harvested by centrifugation and washed free of the medium by repeated resuspension and centrifugation out of M-9 salts without sugar. The actual selection is begun by spreading the resuspended cells on the surface of agar plates containing M-9 salts and 0.2% lactose and incubated at 37°. After 2 days, Lac<sup>+</sup> colonies appear initially at a frequency of around  $10^{-8}$ , but new colonies continue to appear for several days owing both to slow growers and events that occur on the plate. The colonies that appear later are largely independent isolates, and it is most efficient to pick several of these after marking the colonies that appear early. Leaky growth of the bacterial lawn is often observed and is probably due partly to contamination of lactose with other sugars and partly to residual  $\beta$ -galactosidase activity from internal translation initiation events in the 'lacZ' fragment.

The Lac<sup>+</sup> colonies are purified by restreaking on M-9 lactose plates and then tested semiquantitatively for levels of  $\beta$ -galactosidase by spotting onto a series of plates containing various indicators as described in the materials and reagents section. Quantitative assays of *E. coli* stains containing plasmids are performed on exponentially growing cells in LB broth containing ampicillin. Cells are grown to a density of about  $2 \times 10^8$  cells per milliliter, chilled on ice, centrifuged to remove the medium, and resuspended in an equal volume of chilled M-9 salts lacking sugar. Assays are then performed as described by Miller<sup>5</sup> using cells permeabilized with SDS and CHCl<sub>3</sub>. Plasmid DNA is purified from several independent Lac<sup>+</sup> colonies and the DNA is analyzed by restriction enzyme digestion to determine the extent of the deletion.

### *In Vitro* Fusion Construction

Two different *in vitro* approaches have been taken to construct fusions. If the DNA sequence of the region of interest is known, then it is often possible to find a restriction enzyme cleavage site that allows the direct construction of a fusion.<sup>3</sup> However, for small targets or for genes in which the sequence is not known, this approach will not suffice. We have utilized the double-stranded DNA exonuclease *Bal31* and DNA linkers to introduce the requisite restriction enzyme sites at random positions in the region of interest.

A plasmid is constructed using the same criteria as for the *in vivo* method. The general structure of such a plasmid and the scheme for *in vitro* fusion construction are diagrammed in Fig. 2. The requisite features

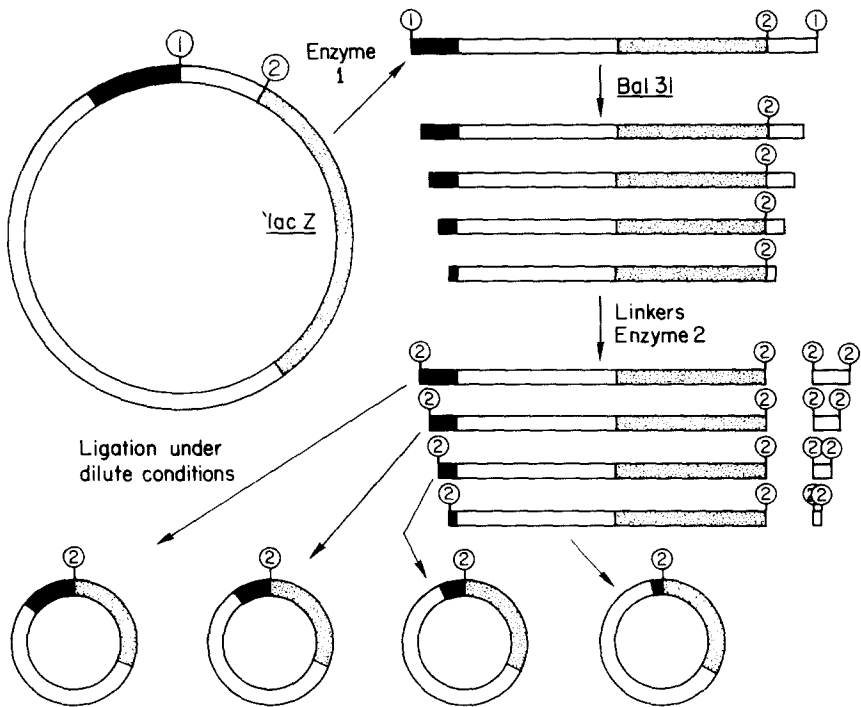


FIG. 2. Scheme for *in vitro* gene fusion construction. The plasmid shown is analogous to pRB45 shown in Fig. 1. The stippled area is the *lacZ* segment. The black area represents the target gene of interest. See the text for details.

of the plasmid are the two unique restriction sites, 1 and 2. Site 2 is the restriction site at the boundary of the *lacZ* segment and may be *Bam*HI, *Eco*RI, *Hind*III, *Sma*I or *Xho*I. Site 1 can be in or near the end of the gene of interest. In the case of pRB45 the site is *Sma*I. The distance from site 1 to site 2 must not be shorter than the distance that must be deleted on the other side of site 1, or portions of *lacZ* may get deleted. If necessary, a DNA fragment can be ligated into site 2 to provide a more favorable spacing.

The first step is to cut with enzyme 1 to completion. Since the later steps in the procedure are not completely efficient, it is advisable to remove the small fraction of uncut supercoils from the population, as many of these are likely to be preexisting deletions and will be selected for later on. This can be done conveniently by chromatography on acridine yellow-ED acrylamide gel<sup>17</sup> (Boehringer Mannheim) or by agarose gel elec-

<sup>17</sup> H. Buenemann and W. Mueller, in "Affinity Chromatography" (O. Hoffmann-Ostenhof *et al.*, eds.), p. 353. Pergamon, New York, 1978.

trophoresis. After restriction enzyme digestion, the DNA is phenol extracted and ethanol precipitated. To perform the chromatography, the DNA pellet is dissolved at 500  $\mu\text{g}/\text{ml}$  in 10 mM sodium citrate, pH 7.0, 0.5 mM EDTA and then brought to 0.5 M NaCl (loading buffer). The dissolved DNA is slowly passed through a Pasteur pipette column containing 0.25 ml of packed gel preequilibrated in the loading buffer. The column is washed with 10 column volumes of loading buffer, and fractions that contain DNA are pooled and ethanol precipitated. Usually essentially all of the linear DNA is contained in the first three column volumes. Under these conditions supercoiled DNA is bound to the column and can later be eluted with 1.0 M NaCl. Any relaxed circular DNA elutes with the linear DNA.

Linear DNA is then digested with *Bal31* using the conditions of Leigerski *et al.*<sup>16</sup> Reactions are carried out in 20 mM Tris-HCl, pH 8.0, 0.45 M NaCl, 1 mM  $\text{Na}_2\text{EDTA}$ , 12.5 mM  $\text{MgSO}_4$ , and 12.5 mM  $\text{CaCl}_2$ . For our experiments, DNA was present at a concentration of 100  $\mu\text{g}/\text{ml}$ , which corresponds (for our plasmid) to a DNA end concentration of around 15 nM. *Bal31* was added to a concentration of approximately 35 units/ml, and reaction mixtures were incubated at 28°. Under these conditions the enzyme digested approximately 100 bp from each end per minute. It is important to titrate the enzyme to determine a rate for the reaction under the conditions used for the experiment. This can be done by cutting with a second restriction enzyme to generate a fragment from one end of the linear molecule whose size can be observed to decrease by as little as 100 bp on agarose gels. The *Bal31* reaction is terminated by addition of  $\text{Na}_2\text{EDTA}$  to 25 mM. To obtain a broad spectrum of fusions, aliquots are removed from the reaction mixture at different times. The distribution of ends is fairly broad, so that close spacing of time points is unnecessary. After the reaction is terminated the DNA is phenol extracted and ethanol precipitated.

Linker DNA molecules are then ligated onto the ends to facilitate making fusions to the enzyme 2 site. Commercially prepared linker molecules containing the enzyme 2 restriction site are first phosphorylated with T4 polynucleotide kinase and then ligated together in the presence of the *Bal31*-treated DNA. We have routinely phosphorylated linkers in 6 mM Tris-Cl, pH 7.5, 10 mM  $\text{MgCl}_2$ , 12.5 mM dithiothreitol, 1 mM spermidine, 1 mM ATP, and 200  $\mu\text{g}$  of bovine serum albumin (BSA) per milliliter with linkers at a concentration of 300  $\mu\text{g}/\text{ml}$  (about 50  $\mu\text{M}$  for a 10-mer) and T4 polynucleotide kinase at 2000 units/ml. The phosphorylation reaction mixture is incubated at 37° for 1 hr and then is diluted by the addition of five volumes of *Bal31*-digested DNA dissolved in the phosphorylation buffer. The final linearized plasmid DNA concentration is ap-



proximately 200  $\mu\text{g/ml}$  (30  $\text{nM}$  in ends). T4 DNA ligase is then added to about 20,000 units/ml and incubated at 13° for 2 hr. Ligation of the linkers can be checked by electrophoresis of a portion of the ligation mixture on a 12% polyacrylamide gel and staining the gel with ethidium bromide. A successful ligation is indicated by the appearance of an extensive ladder of linker polymers. Alternatively, the linkers can be phosphorylated with  $^{32}\text{P}$  at the kinase step, and incorporation of radioactivity into the plasmid DNA can be followed.

It is necessary both to trim away the extra linkers from the linears and to cut the plasmid at the site next to *lacZ*. The presence of the linker polymers will inhibit the restriction enzyme by competing with the plasmid DNA. The plasmid DNA is purified away from the linkers by agarose gel electrophoresis. DNA can be purified out of the agarose by various methods. We have used a modification<sup>18</sup> of the hydroxyapatite chromatography method described by Tabak and Flavell,<sup>19</sup> which works well for purification of large DNA fragments from agarose. The DNA is adsorbed onto hydroxyapatite and sequentially washed with two column volumes of fresh running buffer, two volumes of 10  $\text{mM}$   $\text{KH}_2\text{PO}_4$ , pH 7.0, two volumes of 100  $\text{mM}$  potassium phosphate and then eluted with 400  $\text{mM}$  potassium phosphate. By incorporation of 0.5  $\mu\text{g}$  of ethidium bromide per milliliter into the phosphate buffers, the elution of the DNA can be followed with longwave UV light. The eluted DNA is pooled, 10  $\mu\text{g}$  of carrier tRNA are added and then phenol extracted to remove ethidium bromide. The DNA is dialyzed to remove phosphate and then ethanol precipitated.

The linear plasmid DNA with the linkers ligated onto the ends is then digested with enzyme 2. To remove all of the excess linkers polymerized onto the linear plasmid, it is important to get complete digestion. We have overcut the DNA about 10-fold by adding several aliquots of the restriction enzyme over the course of several hours. The reaction is terminated by the addition of  $\text{Na}_2\text{EDTA}$  to 25  $\text{mM}$  followed by phenol extraction and ethanol precipitation. In order to circularize the molecules, the redissolved DNA is diluted into ligase buffer to a final concentration of 0.5  $\mu\text{g/ml}$ . Under dilute conditions (0.1  $\text{nM}$  ends) circulation is favored over polymerization. Ligase buffer is 50  $\text{mM}$  Tris-Cl, pH 8.0, 10  $\text{mM}$   $\text{MgCl}_2$ , 20  $\text{mM}$  DTT, 1  $\text{mM}$  ATP, and 50  $\mu\text{g}$  of BSA per milliliter. Ligase is added to between 200 and 2000 units per milliliter, and the reaction is incubated overnight at 14°. The ligation mixtures are then used directly to

<sup>18</sup> R. W. Davis, D. Botstein, and J. R. Roth, "A Manual for Genetic Engineering, Advanced Bacterial Genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1980.

<sup>19</sup> H. F. Tabak and R. A. Flavell, *Nucleic Acids Res.* 5, 2321 (1978).

transform *E. coli* to drug resistance. By incorporating X-gal into the selection agar, the "in frame fusions" can be easily recognized.

A high proportion of the plasmids recovered in *E. coli* show varying degrees of  $\beta$ -galactosidase expression. In one experiment, as many as 40% showed higher levels of  $\beta$ -galactosidase than the parent plasmid pRB45.

### Introduction of Fusions into Yeast and Assay of $\beta$ -Galactosidase Activity

Interesting fusion plasmids produced by these methods are introduced into appropriate yeast strains by standard transformation procedures.<sup>20</sup> We have taken the approach of using 2- $\mu$ m plasmid-derived vectors because of the high frequency of transformation and the higher levels of expression from the multicopy plasmid. However, newer techniques of linearizing the vector DNA to obtain increased frequencies of integrative transformation may supersede this approach.<sup>21</sup> Transformants that carry the fusion are easily assayed for their level of  $\beta$ -galactosidase activity by patching the cells onto buffered minimal plates containing X-gal. The neutral pH that is required for efficient color production does not allow regeneration, so that transformants cannot be assayed for  $\beta$ -galactosidase on the transformation selection plates. Cells that express  $\beta$ -galactosidase may turn blue at various rates, from overnight to several weeks, depending on the level of expression. A more quantitative assay (and ultimately quicker for low-level expression), is the direct determination of *o*-nitrophenylgalactoside cleavage by cell extracts.

Yeast cells containing plasmids are grown under selection to about  $2 \times 10^7$  cells per milliliter. Cells are chilled, concentrated about 20-fold by centrifugation and resuspension in 100 mM Tris-Cl, pH 8.0, 1 mM dithiothreitol, and 20% (v/v) glycerol and routinely frozen at  $-20^\circ$ . The frozen cells are thawed on ice, 0.45 mm glass beads (Glasperlen, VWR-Scientific) are added to fill the liquid volume up to the meniscus, and phenylmethylsulfonyl fluoride (Sigma) is added to 1 mM from a 40 mM stock in 95% ethanol stored at  $-20^\circ$ . Cells are broken open by vigorous agitation on a Vortex mixer (six times, 15 sec each) at  $4^\circ$ . An equal volume of the breaking buffer is mixed in, and the solution above the beads is clarified by 15-min centrifugation in an Eppendorf centrifuge.  $\beta$ -Galactosidase is then assayed as for a bacterial extract except that the activity is normalized to the protein concentration of the extract. We have measured the

<sup>20</sup> A. Hinnen, J. B. Hicks, and G. R. Fink, *Proc. Natl. Acad. Sci. U. S. A.* **75**, 1929 (1978).

<sup>21</sup> T. L. Orr-Weaver, J. W. Szostak, and R. J. Rothstein, *Proc. Natl. Acad. Sci. U. S. A.* **78**, 6354 (1981).

protein concentration by the method of Bradford,<sup>22</sup> as this method is not sensitive to high Tris concentrations. Activity normalized to protein was found to be a more reliable measure of specific activity than that normalized to the optical density of the culture, particularly when nonisogenic strains are used.

The range of activities of  $\beta$ -galactosidase we have measured in yeast is from 2 units to 20,000 units. Five milliliters of exponentially growing cells is sufficient to assay this range of activity within a few hours of incubation. The reactions have been carried out for up to 24 hr with less than a 20% loss of activity. Autonomously replicating yeast plasmids are rather unstable, so that recombinants that transfer the selectable marker to the chromosome are at a selective advantage. This can result in fast growing strains that may contain the multicopy plasmid in much less than 10% of the cells in the culture. It is therefore best to assay exponentially growing cells grown from fresh overnights made from fresh colonies of the transformants.

#### Typical Results of the Two Methods

The distribution of the deletion end points produced by the application of the two techniques to pRB45 is shown in Fig. 3. Of the total number of Lac<sup>+</sup> colonies obtained by *in vivo* selection, approximately 80% remained Ura<sup>+</sup> due to deletion hotspots between *URA3* and '*lacZ*'. The remaining plasmids were Ura<sup>-</sup> and contained large deletions that were broadly distributed over a large region of the plasmid. By contrast, the deletion end points constructed by the *in vitro* method were distributed rather uniformly throughout a small region of the plasmid. The larger deletions produced *in vivo* are useful for determining the location of a structural gene within a large DNA fragment. The *in vitro* method of fusion construction produces more tightly clustered deletion end points, which are useful for determining function over small regions of DNA or for obtaining fusions within a specific DNA segment.

The *in vitro* method allowed the recovery of '*lacZ*' fusions to an open reading frame of 84 base pairs that begins with an ATG and overlaps the start of the *URA3* coding sequence. The DNA sequence of one of the fusions to the peptide coding sequence is illustrated in Fig. 4. After transformation of yeast with a derivative of pRB45 containing this fusion, we found that  $\beta$ -galactosidase is expressed at 10% the level of equivalent fusions made to the *URA3* gene. Fusions to the sequences that are not part of yeast genes give no detectable expression of  $\beta$ -galactosidase. As the

<sup>22</sup> M. M. Bradford, *Anal. Biochem.* **72**, 248 (1976).

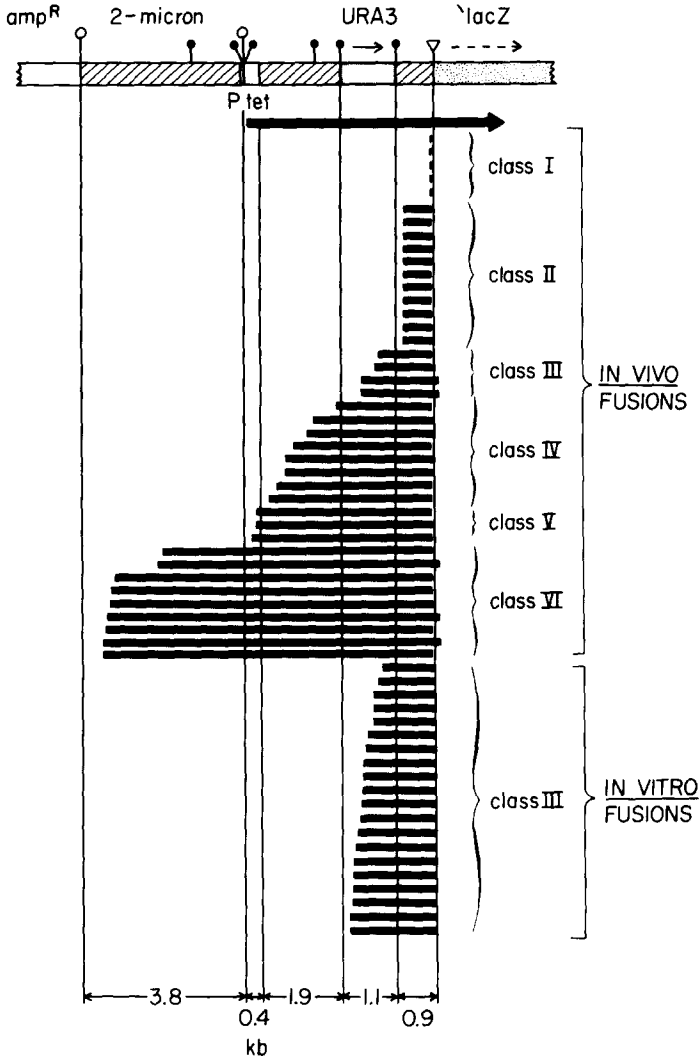


FIG. 3. Deletion map of fusion plasmids. The structure of the relevant portions of pRB45 are shown at the top. Left to right corresponds to clockwise in Fig. 1; the map is drawn approximately to scale. Restriction enzyme site symbols are as in Fig. 1. For the *in vivo* fusions, the deletion end points are assumed to lie close to the *Bam*HI site. Some of the deletions remove the *Bam*HI site and must therefore end in the *lacZ* sequence. All of the *in vitro* fusions necessarily end at the *Bam*HI site. The various fusion classes are defined by the left-hand deletion end point. In yeast, only class II, class III, and some of class VI fusions express significant  $\beta$ -galactosidase activity. All the fusions shown express  $\beta$ -galactosidase activity in *Escherichia coli*.

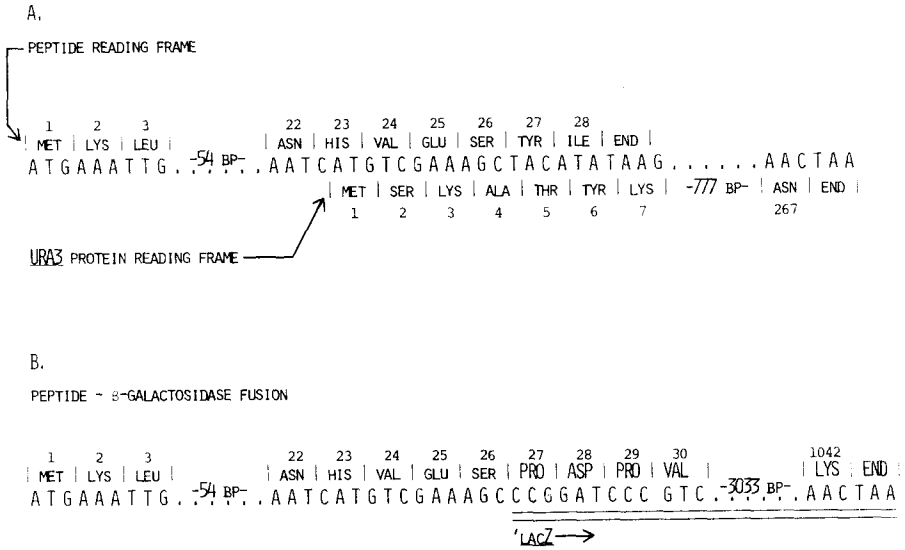


FIG. 4. (A) DNA sequence at the beginning of the *URA3* gene. An open reading frame potentially encoding a peptide of 28 amino acids overlaps with an open reading frame encoding a protein of 267 amino acids. The two reading frames overlap by 17 base pairs. (B) DNA sequence of a fusion of '*lacZ*' to the peptide coding sequence. '*lacZ*' is fused after the 26th codon of the peptide. The '*lacZ*' segment is underlined. This fusion is expressed in yeast and *Escherichia coli*. A derivative of this fusion in which the '*lacZ*' sequence is frameshifted so as to be fused to the beginning of the *URA3* protein coding sequence is also expressed in yeast.

hybrid protein is expressed in yeast, the 84 base pair sequence encoding the peptide must be transcribed. Moreover, since the AUG at the start of this sequence is capable of initiating translation of the fusion protein it must also be capable of initiating translation of the peptide itself. Evidence for the expression of this open reading frame would be difficult to obtain by other methods.

Fusions of '*lacZ*' to the *URA3* gene show a gradient of  $\beta$ -galactosidase expression in yeast. Fusions obtained near the amino terminus of the gene tend to show higher levels of  $\beta$ -galactosidase (as much as 10-fold higher) than fusions obtained near the carboxyl terminus. The difference is apparently not simply due to variation in the inherent specific activity of the proteins as all of the fusions make similar levels of  $\beta$ -galactosidase in *E. coli*. The reason for the differences in expression are not yet understood but invalidate arguments based on absolute levels of expression.

The presence of the defined linker segment in the *in vitro* constructed fusions is extremely useful in determining the exact position and reading

frame of the fusion. The linker is a unique restriction site so that DNA sequencing from this site rapidly yields the exact position of the fusion. This is particularly useful in determining the reading frame of new sequence. If the fusion expresses  $\beta$ -galactosidase then the reading frame of the sequence to the 5' side of '*lacZ*' can be extrapolated by extension of the '*lacZ*' reading frame. If the deletions are constructed in an unsequenced gene, the set of fusion plasmids provides an ordered set of unique restriction sites from which to sequence. This greatly accelerates the rate at which one can sequence a new gene as it does not require knowledge of a detailed restriction map. We have sequenced most of the *URA3* gene by this method.

The *in vivo* selected deletions end at various points 5' to '*lacZ*' resulting in a set of hybrid genes that are ill defined at the fusion junction. Most of the deletions do not remove material from the '*lacZ*' segment, and thus the hybrid gene must also contain some material that was immediately adjacent to the '*lacZ*' segment. The specific activity of the hybrid protein may vary owing to differences in this intervening segment. This problem is avoided with the *in vitro* method as all the fusions must contain the same linker segment.

It is worthwhile to point out that many of the fusions that are functional in *E. coli* are not necessarily true fusions to a yeast gene. We have observed several different events that complicate the analysis of fusions in *E. coli*. Fusions to the yeast sequences to the 5' side of the *URA3* gene are strongly expressed in *E. coli* but are not at all expressed in yeast. Fusions to the peptide sequences are indistinguishable from fusions to *URA3* in *E. coli*, but show much lower levels of expression in yeast and are not regulated by pyrimidine metabolism as is *URA3*. A fusion was obtained to a point within the *URA3* coding segment that made high levels of  $\beta$ -galactosidase activity in *E. coli* undetectable levels in yeast. DNA sequence analysis revealed that the fusion is out of frame with respect to the *URA3* sequence. Whereas *E. coli* is capable of either an internal initiation event, or frameshifting during translation, yeast is unable to express this false hybrid. In the last analysis, the best criterion for a true fusion of *lacZ* to a yeast gene is the observation of expression of the  $\beta$ -galactosidase activity in yeast.

### Acknowledgments

We thank Arlene Wyman, David Shortle, and Michael Lichten for helpful comments on this manuscript. We thank Jack Szostak for recommending the buffered minimal medium of Clifton *et al.* This work was supported by NIH Grants R01-GM21253-08 and R01-GM18973-11 and ACS Grant MV 90C. M. R. was supported by an NSF fellowship and by NIH training Grant GM07287.