[8] Searching for Clones with Open Reading Frames

By Mark R. Gray, Gail P. Mazzara, Pranhitha Reddy, and Michael Rosbash

We have devised a simple strategy to quickly locate and express open reading frames. Using this approach, it is possible to identify protein-coding regions without any extensive information about the products. Typically, in order to identify and characterize a region of DNA, its sequence, RNA, or protein products are used to construct a molecular map.

Frequently, it is not easy to generate such a map. The approach described in this chapter can be used to locate coding regions whose products are either unknown or undetectable, starting with cloned or viral DNA. The same approach can be used to express any portion of either a poorly characterized or a well-characterized DNA or cDNA. Antibodies can be made against polypeptide determinants encoded by the open reading frame sequences of the starting DNA.

Principle

One of the distinguishing features of DNA in protein-coding regions is the presence of at least one open reading frame (ORF). DNA sequences outside of protein-coding regions, such as intergenic regions and introns, rarely have long open reading frames. In some organisms, such as Drosophila, the DNA outside of protein-coding units is enriched for A and T bases, increasing the chance of having one of the three stop codons TGA, TAG, and TAA.¹

The open reading frame cloning strategy discussed here utilizes the pMR series of plasmid vectors; in these plasmids, a lacZ fusion gene is used to select small (100–1000 bp) DNA fragments that have continuous open reading frames.² The enzyme β-galactosidase, encoded by the lacZ gene of Escherichia coli, is often biologically active when an additional polypeptide sequence is attached to its amino terminus.³ The pMR plasmids carry a strong bacterial promoter driving the expression of a

cI::lacIZ fusion gene; a cloning site is located between the cI and the lacIZ portions of the gene. When there is a frameshift mutation downstream from the cloning site, the production of a high level of β-galactosidase activity by bacterial transformants is eliminated. This frameshift can be corrected by the insertion (into the cloning site) of a DNA fragment that restores a continuous open reading frame through the lacIZ portion of the gene. This results in the production of a large amount of β-galactosidase activity by the transformants. When the frame-shifted vector is ligated with appropriate DNA fragments, and the ligated plasmids used to transform λc− bacteria, transformants that produce high levels of β-galactosidase activity often contain plasmids with inserts having continuous open reading frames. The recombinant plasmids can be used as probes to map the genomic location of open reading frames. These transformants express the inserted DNA sequence as part of a β-galactosidase fusion protein. Thus, with this procedure, the protein-coding portion of a gene can be identified, mapped, and expressed.

Materials

All enzymes, chemicals, and apparatus needed to locate and express open reading frames are available from commercial sources as indicated below. The sources of the bacterial strains and DNA fragments used to construct the plasmid vectors have been previously reported. The Ubx DNA was a gift from Welcome Bender; the CPV DNA was a gift from Solon Rhode.

Methods

A diagram of the structure of all of the pMR vectors is shown in Fig. 1A. Each of the plasmids pMR1, pMR2, pMR100, and pMR200 differ from each other only in the region of the cloning site between the cI and lacIZ portions of the fusion gene; the different sequences for each of the plasmids are shown in Fig. 1B. The plasmids pMR1 and pMR200 confer on host bacteria a strong lac+ phenotype; the plasmids pMR2 and pMR100 give a lac− phenotype.

The plasmid pMR2 has a frameshift in the cI::lacIZ fusion gene upstream from the lacIZ part of the gene; it is designed to select open reading frame DNA fragments with HindIII termini. The plasmid pMR1 has no frameshift and causes the production of a high level of full-length and enzymatically active cI::lacIZ fusion protein when used to transform the lac operon deletion strain LG90. The plasmid pMR1 is used to esti-

A

FIG. 1. (A) Diagram of the structure of pMR vectors; all four versions are identical except for the sequence at the cloning site between the cl and lacI portions of the fusion gene. (B) DNA sequences at the cloning sites of pMR1, pMR2, pMR100, and pMR200.

mate the insert frequency in parallel ligations of the same fragments with pMR2; the fraction of the total number of pMR1 transformants that become lac- after ligation with insert fragments can be used to estimate the number of transformants that have inserts in an identical ligation using pMR2.
The plasmid pMR100 was derived from pMR1 by the insertion of a 10-bp Smal/BamHI adapter at the BamHI site (Fig. 1B). This insertion introduces a frameshift in the fusion gene of pMR1, creating a plasmid that can be used for the selection of any blunt-ended fragment that can correct the frameshift by insertion into the Smal site. The plasmid pMR200 is a single base-pair deletion derivative of pMR100 that retains the Smal site and removes the frameshift (Fig. 1B). As with pMR1, LG90/pMR200 transformants produce large amounts of full-length cI::lacIZ fusion protein (Table I; see Fig. 2). The amounts of β-galactosidase activity produced by transformants of each of the pMR plasmids, and their transformant lacZ

<table>
<thead>
<tr>
<th>Clone</th>
<th>MacConkey phenotype</th>
<th>XGal phenotype</th>
<th>Enzyme activity</th>
<th>Insert size</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG90</td>
<td>White</td>
<td>White</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>pMR1</td>
<td>Red</td>
<td>Blue</td>
<td>4471</td>
<td>—</td>
</tr>
<tr>
<td>pMR2</td>
<td>White</td>
<td>Blue</td>
<td>121</td>
<td>—</td>
</tr>
<tr>
<td>pMR100</td>
<td>White</td>
<td>Blue</td>
<td>60</td>
<td>—</td>
</tr>
<tr>
<td>pMR200</td>
<td>Red</td>
<td>Blue</td>
<td>9800</td>
<td>—</td>
</tr>
<tr>
<td>UBX7</td>
<td>Red</td>
<td>Blue</td>
<td>2083</td>
<td>~305</td>
</tr>
<tr>
<td>UBX17</td>
<td>Red</td>
<td>Blue</td>
<td>508</td>
<td>~375</td>
</tr>
<tr>
<td>UBX35</td>
<td>Red</td>
<td>Blue</td>
<td>687</td>
<td>~335</td>
</tr>
<tr>
<td>UBX27</td>
<td>Red</td>
<td>Blue</td>
<td>401</td>
<td>317</td>
</tr>
<tr>
<td>UBX20</td>
<td>Red</td>
<td>Blue</td>
<td>342</td>
<td>305</td>
</tr>
<tr>
<td>UBX10</td>
<td>Red</td>
<td>Blue</td>
<td>1173</td>
<td>~315</td>
</tr>
<tr>
<td>UBX12</td>
<td>Red</td>
<td>Blue</td>
<td>2316</td>
<td>~525</td>
</tr>
<tr>
<td>UBX13</td>
<td>Red</td>
<td>Blue</td>
<td>2091</td>
<td>569</td>
</tr>
<tr>
<td>CPV3</td>
<td>Red</td>
<td>Blue</td>
<td>6126</td>
<td>295</td>
</tr>
<tr>
<td>CPV39</td>
<td>Red</td>
<td>Blue</td>
<td>666</td>
<td>439</td>
</tr>
<tr>
<td>CPV170</td>
<td>Red</td>
<td>Blue</td>
<td>1724</td>
<td>431</td>
</tr>
<tr>
<td>UBX102</td>
<td>Weakly red</td>
<td>Blue</td>
<td>278</td>
<td>503</td>
</tr>
<tr>
<td>UBX103</td>
<td>Red</td>
<td>Blue</td>
<td>461</td>
<td>429</td>
</tr>
<tr>
<td>CPV6</td>
<td>Red</td>
<td>Blue</td>
<td>8301</td>
<td>312</td>
</tr>
<tr>
<td>CPV27</td>
<td>Red</td>
<td>Blue</td>
<td>844</td>
<td>315</td>
</tr>
</tbody>
</table>

* The UBX and CPV plasmids are derived from pMR100 as described in the text. The MacConkey phenotype refers to the color of the colony on MacConkey agar plates (red = lac+); the XGal phenotype refers to the color on XGal/minimal plates (blue = lac+). The enzyme activity is the number of β-galactosidase activity units (Miller units) produced by each cell line. The insert size is the length in base pairs, determined by DNA sequencing or comparison to DNA fragment size standards by analytical gel electrophoresis.
phenotypes on 5-bromo-4-chloro-3-indolyl-β-D-galactoside (XGal) and MacConkey agar plates are listed in Table I.

The methods described below refer to the ligation of blunt-ended DNA fragments into pMR100; the same procedures are applicable to ligations using pMR1, pMR2, and pMR200. In order to map open reading frames in uncharacterized cloned DNA or express portions of a well-characterized gene, the simplest approach is to randomly fragment the DNA, repair the fragment termini to blunt ends, ligate size-fractionated fragments into the SmaI site of pMR100, transform, and analyze lac+ LG90 transformants. Since all restriction and DNA modification enzymes have some substrate specificity (and therefore possible sequence specificity), sonication is the fragmentation method discussed below. Sonication does not cut DNA entirely at random; AT-rich DNA fragments are disrupted by sonication faster than GC-rich fragments. Many eukaryotic sequences within protein-coding exons may be GC-rich, such as in those of Drosophila; in these cases, sonication should disrupt exon sequences less quickly than AT-rich intron and spacer sequences.

The random fragmentation approach has several advantages over using restriction enzymes to make the insert fragments. (1) There is no need to have the appropriate restriction sites at both ends of the fragment in order to maintain the correct open reading frame. (2) Even the most carefully selected restriction fragment may not necessarily be the best one to use to isolate a stable fusion protein in bacteria. (3) The use of randomly cut DNA fragments bypasses the confusion and frustration caused by frameshift errors in the DNA sequence determination of the starting DNA. (4) If a variety of fusion proteins are desired from a single gene, randomly cut fragments from all regions of the substrate DNA will be ligated to pMR100 and the resultant lac+ clones selected by their production of easily detectable amounts of stable fusion protein.

**Preparation of DNA Fragments**

DNA used to prepare small fragments for the detection of open reading frames should be free of contaminating DNA, such as that from the bacterial chromosome or lambda phage and plasmid vectors; these contaminating DNAs are rich in open reading frame sequences. Since most cloned DNAs are prepared by gel electrophoresis, it is important to digest the DNA under conditions that minimize degradation, to avoid vector fragments.

To follow each of the subsequent preparative steps for the insert fragments, start with a large excess (e.g., 20 μg) of the source DNA. If the

---

DNA is in a lambda phage vector, or a relatively low copy number plasmid, it is worth the small investment of time to reclone it in a high copy number plasmid such as those of the pUC and pEMBL series.\(^6\)\(^7\) In this way, 0.5–1.5 mg of plasmid DNA containing the target DNA can be easily prepared from a single 500 ml culture of transformed cells, providing more than enough DNA to prepare small fragments for ligation into pMR vectors.

**Purification of Starting DNA Fragments.** Any reliable method can be used to purify the starting DNA; we have found that electrophoresis on small agarose gels to be the simplest and fastest. A DNA fragment can be easily removed from the gel by first trapping the stained band on a small piece of DEAE membrane (NA-45, Schleicher and Schuell) by electrophoresis, and then eluting the DNA fragment from the membrane in 400 µl of 50 mM arginine plus 1.0 M NaCl (two incubations at 65°, 200 µl each). We have observed that DNA fragments purified in this way are not resistant to further enzyme treatments, unlike those purified by electrolution from agarose gel slices in dialysis bags; the DEAE method is much faster and less cumbersome as well. The purified fragments can be stored in any volume until sonication, or precipitated in ethanol; in either case, it is useful to check for recovery by electrophoresis of a small fraction of the purified target DNA on minigels. Throughout all of the fragment preparation steps, it is important to avoid irradiation and damage of DNA by short-wave ultraviolet light; photograph the ethidium bromide-stained preparative gels using only long-wave (greater than 300 nm) ultraviolet light sources.

**Sonication.** The purified DNA fragments are sonicated to a size range of 100–1000 bp, with most of the fragments 400–500 bp, as follows. Suspend the DNA in TE\(_1\) (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) plus 0.2 M NaCl, in a tube suitable for sonication; plastic tubes are less likely than glass tubes to shatter accidentally. Carefully clamp the tube so that the sonicator tip does not touch the wall of the tube. Surround the tube with ice water so that the DNA solution does not overheat during sonication. Set the power output to a level just below the maximum recommended for the tip. Position the sonicator tip so that when the power is on, the DNA solution does not splash and foam; when the tip is not fully submerged in the liquid during sonication, the DNA is not fragmented well and the tip may become damaged. Sonicate for 30 sec at a time, with at least 30 sec in between bursts to allow time for cooling. To reduce most of the fragments to the length range 100–1000 bp, with most of the DNA around 450 bp,

sonicate for a total of 3 min. Small (<1.5 kb in length) target DNA fragments occasionally require longer times of sonication than larger (~4 kb) molecules.

After sonication, concentrated (at least 20 μg/ml) DNA fragments may be recovered by ethanol precipitation. More dilute DNA solutions should be passed over a DEAE–cellulose column (DE-52; Whatman); alternatively, disposable columns that attach to syringes (Elutips; Schleicher and Schuell) can be used. After binding on DEAE–cellulose in low salt buffer (TE plus 0.2 M NaCl or less), the DNA fragments are eluted in a small (300–400 μl) volume of high salt buffer (TE plus 1 M NaCl). Precipitate the DNA fragments in a microfuge tube by adding 2 volumes of 95% ethanol, leaving the tube on dry ice for 10 min, then centrifuging for 5 min; wash the pellet in 70% ethanol, recentrifuge, and dry the pellet. Dissolve the DNA fragments in a small (10–50 μl) volume of the buffer used for the end repair steps to follow.

Repair of Termini to Blunt Ends. After sonication, the ends of the DNA fragments are virtually unligatable since they are most probably a random mixture of 5' and 3' overhangs of variable lengths. There are many enzymes that can be used to repair the ends of the fragments to blunt ends. Nucleases that remove single-stranded DNA ends or polymerases that fill in by using the single-strand overhangs as templates work well. Preparations of each enzyme vary over a wide range in their efficiency of making blunt ends; it is best to evaluate several different approaches with some abundant source of sonicated DNA before selecting a repair strategy for the purified target DNA fragments of interest.

We have achieved good results in making blunt ends by a short digestion of the fragments with S1 nuclease (to remove single-stranded termini), followed by a fill-in treatment using T4 polymerase; treatment of the sonicated fragments with either of these enzymes alone has been less successful. In this approach, the precipitated sonicated DNA (2–10 μg) is resuspended in 50 μl of S1 buffer, placed at 37°, and digested with S1 nuclease (final concentration 500 units/ml) for 15 min. Stop the reaction by adding EDTA to 25 mM; dilute in 200 μl of TE plus 0.2 M NaCl, extract with 250 μl of buffered phenol/chloroform (1:1), and precipitate the S1-digested fragments with ethanol. Dissolve the lyophilized pellet in 20 μl of TA buffer (33 mM Tris-acetate, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, pH 7.9). Place the DNA at 37° and add one unit of T4 polymerase for each microgram of DNA fragments. In the absence of added nucleotides, the 3' to 5' exonuclease activity of T4 polymerase converts all single-stranded termini to 5' overhangs. After 3 min, add all four nucleotides to 0.1 mM for the fill-in reaction (at least 30 min at 37°). The reaction is stopped by phenol/chloroform extraction and
reprecipitation in ethanol of the repaired fragments. The success of the end repair reactions should be tested by self-ligating a small aliquot of the repaired and unrepaired fragments; compare ligated and unligated samples by electrophoresis on minigels.

**Size Fractionation of Sonicated Fragments.** To avoid cloning very small random ORFs from genomic sequences that do not encode proteins, it is necessary to select by size the DNA fragments desired for ligation into pMR100. The probability that a fragment of random sequence of length $x$ codons is the correct size $(3n + 2)$ and entirely open reading frame is $1/3 \times (61/64)^x$ (without correcting for GC composition differences). In this calculation, there is a $1/3$ probability that the fragment is the proper length, and a $61/64$ chance that any codon will not be a stop codon. If the fragments (of random sequence) are in the range of 200 bp (or 66 codons), then the probability of the fragment being suitable for correcting the frameshift in pMR100 is $4.6 \times 10^{-3}$. If the fragments are 300 bp long, then the probability is $9.1 \times 10^{-4}$, and for 500 bp, $3.5 \times 10^{-5}$. The probability that a fragment of DNA from a large open reading frame sequence will give rise to a pMR100 lac$^+$ clone is considerably higher. It is the product of the probability of having the correct length ($1/3$) times that of starting in the correct frame ($1/3$) times the probability of being cloned in the proper orientation ($1/2$). For completely open reading frame DNA, the probability is $5.5 \times 10^{-2}$ and is independent of the length of the fragment. For a 500-bp fragment, the difference in the probabilities for the detection of an ORF fragment in random sequence versus open reading frame is large enough to strongly suggest that it is derived from a bona fide open reading frame from the protein coding region of a gene. Alternatively, for a 150-bp fragment, the difference is very small.

The very small sonicated fragments can be removed by size selection after electrophoresis. Electrophorese the DNA on a 10% acrylamide/TBE gel (0.6 mm thick), along with appropriate size markers, at 200 V until the bromphenol blue has traveled 10–15 cm (about 2 hr). Up to 10 µg of DNA can be loaded into a well 1.6 cm wide. Stain the gel in ethidium bromide, destain in distilled water, and photograph using a long-wave ultraviolet light source. Electrottransfer the DNA fragments from the gel to DEAE–cellulose paper (Whatman DE-81) as follows. Assemble a sandwich consisting of the following: a Scotchbrite pad (3M Corporation), two layers of filter paper (Whatman 3MM), the gel, a piece of DEAE–cellulose paper, two more layers of filter paper, and a second Scotchbrite pad. Submerge the sandwich between the two rigid supports of the electrottransfer apparatus after it has been filled with enough transfer buffer (20 mM Tris–HCl,

---

1 mM EDTA, pH 8.0) to cover the gel. The sandwich must be oriented so that the DEAE-cellulose paper is between the gel and the positive electrode. Electrophoresis at 150 mA (about 8–15 V) for 2 hr is sufficient to transfer completely all fragments under 800 bp in length. For larger fragments, increase the current (up to 600 mA at room temperature) or increase the time of electrophoresis.

Remove the very fragile wet DEAE paper by covering the gel and DEAE paper with a piece of plastic wrap, inverting the gel so that the plastic wrap is on the bottom, and then lifting the gel away from the DEAE paper. Cover the DEAE paper with another layer of plastic wrap if it is not possible to elute the DNA immediately; if the DEAE paper is allowed to dry, the DNA fragments do not elute easily. Photograph the DEAE paper using a long-wave UV light source; compare the photograph of the gel to that of the DEAE paper in order to check for recovery. Mark on the plastic wrap the region that contains the DNA fragments of the desired size, using the size standards as a guide. We usually discard any fragments under 200 bp, and save size fractions of 200–400 bp and 400–700 bp. Cut out the DEAE paper that contains the fragments and pack it into a 0.5 ml microfuge tube. Cover the paper with 100 μl of TE plus 0.2 M NaCl to prevent drying. Make a small hole in the tip of the tube using a 26-gauge needle and place this tube into a 1.5-ml microfuge tube. This arrangement is used to wash the DEAE paper and elute the DNA by centrifugation in a microfuge. Wash the paper 3 times with 100 μl of TE plus 0.2 M NaCl. After all of the wash buffer is spun out and discarded, elute the DNA with three 100 μl washes of TE plus 1 M NaCl. Pool the 300 μl of DNA solution and precipitate in ethanol as described above. Resuspend each size fraction of DNA in 10–20 μl of TE; electrophorese 1 μl of each DNA sample on a minigel along with appropriate size and concentration standards in order to estimate recovery. It is sufficient to have 100 ng of each size fraction for ligation with pMR100.

Ligation of Sonicated Fragments to pMR100

For each sample of sonicated DNA to be ligated to pMR100, 0.5 μg of SmaI-digested pMR100 is used; it is convenient to prepare 10–20 μg at one time for many ligations. In order to prevent recircularization of the linearized vector molecules, digest the SmaI-cut DNA with calf intestinal phosphatase (CIP; Boehringer Mannheim) in order to remove the terminal 5' phosphate groups. This treatment reduces transformation (by greater than 99%) by plasmid molecules which lack inserts and makes it possible to transform with much more ligated DNA on each plate. In our experience, phosphatasing the vector does not affect the efficiency of ligation of
DNA fragments to vector molecules, nor does it increase the frequency of false positives due to frameshifts at the Smal site.

After digestion of pMR100 with Smal, check for complete digestion by electrophoresis of a small portion of the DNA. If there are no supercoiled or open circular bands present, increase the volume of the DNA to 200 μl in 10 mM Tris–HCl (pH 8.0) and add approximately one unit of CIP for each microgram of vector DNA. Digest at 37°C for 30 min, and then phenol/chloroform extract and precipitate the phosphatased vector in ethanol; resuspend the DNA to 0.5 μg/μl in TE.

Mix 0.5 μg pMR100 vector DNA, 50–100 ng of insert DNA fragments, and no more than 1 unit of T4 DNA ligase (Boehringer Mannheim) in a 10 μl volume of ligase buffer. Addition of more than 1 unit of ligase results in fewer transformants and a large increase in false positives (discussed below). When ligating pMR100 with sonicated fragments, always set up an identical ligation that is missing the insert fragments. Incubate for 12–16 hr at 15°C before transformation. It is informative to compare the ligated DNA samples by electrophoresis of very small aliquots (0.5 μl of the 10 μl total) on minigels. The intensity of the ligated open-circle vector band above the unligated linear band provides a reliable prediction of the expected number of transformants.

**Transformation of Host Bacteria**

Any transformation method that will consistently yield 10⁶ or more transformants from 1 μg of supercoiled plasmid DNA is sufficient. We have used the method of Dagert and Ehrlich, with some minor modifications, for transforming LG90 cells with ligated pMR100 DNA.

Prepare competent LG90 cells by diluting a saturated culture (less than 7 days old) 1 : 100 in 100 ml of L broth in a 500-ml culture flask. Grow the cells to early log phase (about 90 min at 37°C) with vigorous shaking for maximal aeration; if the cells are grown too long, the transformation efficiency is decreased by 2- to 10-fold. Put the cells on ice for 10 min, and then centrifuge at 7,000 rpm (4°C) for 5 min. Discard all of the L broth, and resuspend the bacteria in one-half of the original culture volume of ice-cold 100 mM CaCl₂. Leave the cells on ice for 20 min, and then centrifuge again; resuspend the pellet in 1% of the original culture volume of ice-cold 100 mM CaCl₂. For optimal efficiency, store the cells at 0°C for 12–20 hr before transformation.

In our hands, phenol extraction and ethanol precipitation of ligated DNA increases the transformation efficiency 5- to 10-fold. Add 200 μl of

---

TE plus 0.2 M NaCl and 10 μg of tRNA carrier to each ligated DNA sample before phenol/chloroform extraction and precipitation. Resuspend the dry pellet in TE at 10 ng of vector DNA/μl.

Pipet the DNA into sterile tubes on ice. For each plate, use 50 ng of ligated vector DNA (or 10 ng if the vector was not phosphatased); it is best to have about 500 colonies on each plate. After 5 min, add 50 μl of competent LG90 cells to each tube; mix and leave the tubes on ice for another 5 min. Temperature shock the cells by putting the tubes in a 37°C water bath for 5 min. Add 1 ml of L broth (with no ampicillin) to each tube, and shake at 37°C for 1 hr. Transfer each sample of transformed cells to 1.5-ml microfuge tubes, and centrifuge for 10 sec. Discard all of the L broth and resuspend the pellet of transformed cells in one drop of sterile distilled water. Spread the cells on a MacConkey agar (Difco Labs) plate (50 μg ampicillin/ml) until the surface of the plate appears dry. When the plates have lower (25–40 μg/ml) concentrations of ampicillin, often more transformants are recovered, but satellite nontransformed colonies frequently appear after overnight incubation; if a larger number of transformants is desired, a lower concentration of ampicillin should be tested.

Incubate the plates at 37°C upside down for at least 24 hr. The transformed colonies will be visible at 12 hr, but the red color of the lac+ transformants may take as long as 40 hr to develop. There is a wide range of expression of β-galactosidase in open reading frame clones, depending on the DNA sequence of the insert; this results in lac phenotypes ranging from a slowly developing (40–48 hr) slightly red to a very rapidly developing (less than 12 hr) dark red. After incubation at 37°C, mark the positions of the lac+ colonies on the back of the plate with a felt-tip pen. The plates can be stored at room temperature for up to 5 days without any change in the red color phenotype; when the plates are stored at 4°C, all colonies, including lac− ones, will turn red.

The number of transformants resulting from each ligated DNA sample may vary over a wide range and is dependent on the efficiency of each of the preparative steps preceding transformation. Typical results for different types of experiments are shown in Table II. For ligation of sonicated DNA inserts, the efficiency of the insert fragment end-repair steps is the most important variable.

Selection of lac+ Transformants

Some of the red lac+ colonies are not always the result of the insertion of an open reading frame fragment in the frame-shifted pMR100 fusion gene. The proportion of artifact red colonies (discussed below) is much higher if the number of transformants is low (when there is little or no
<table>
<thead>
<tr>
<th>Vector</th>
<th>Inserts</th>
<th>Transformants/μg vector</th>
<th>% lac⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMR100 uncut</td>
<td>None</td>
<td>1.0–1.5 × 10⁶</td>
<td>0</td>
</tr>
<tr>
<td>pMR200 uncut</td>
<td>None</td>
<td>1.0–1.5 × 10⁶</td>
<td>100</td>
</tr>
<tr>
<td>pMR100 Smal-cut</td>
<td>None</td>
<td>0.5–1.0 × 10⁶</td>
<td>0.5–1.0</td>
</tr>
<tr>
<td>pMR200 Smal-cut</td>
<td>None</td>
<td>0.5–1.0 × 10⁶</td>
<td>99.0–99.5</td>
</tr>
<tr>
<td>pMR100 Smal-cut</td>
<td>Fragments with HaeIII</td>
<td>0.5 × 10⁶</td>
<td>0.5–5.0⁺</td>
</tr>
<tr>
<td>pMR200 Smal-cut</td>
<td>Fragments with HaeIII</td>
<td>0.5 × 10⁶</td>
<td>Up to 75</td>
</tr>
<tr>
<td>pMR100 Smal/CIP</td>
<td>None</td>
<td>5–10 × 10³</td>
<td>0.5–1.0</td>
</tr>
<tr>
<td>pMR200 Smal/CIP</td>
<td>None</td>
<td>5–10 × 10³</td>
<td>99.0–99.5</td>
</tr>
<tr>
<td>pMR100 Smal/CIP</td>
<td>Fragments with HaeIII</td>
<td>0.25–0.5 × 10⁶</td>
<td>0.5–5.0⁺</td>
</tr>
<tr>
<td>pMR200 Smal/CIP</td>
<td>Fragments with HaeIII</td>
<td>0.25–0.5 × 10⁶</td>
<td>0.5–5.0⁺</td>
</tr>
<tr>
<td>pMR100 Smal/CIP</td>
<td>Sonicated with repaired ends</td>
<td>0.5–2.5 × 10⁴</td>
<td>0.5–5.0⁺</td>
</tr>
<tr>
<td>pMR200 Smal/CIP</td>
<td>Sonicated with repaired ends</td>
<td>0.5–2.5 × 10⁴</td>
<td>1.0–75⁺</td>
</tr>
</tbody>
</table>

* Yield depends on the insert frequency, the length of the insert fragments, the presence of open reading frame sequence, and the presence of sequences that can cause initiation of translation in bacteria.

increase in the number of transformants obtained with the ligation of phosphatased pMR100 and insert fragments over that of vector alone). For this reason, it is best to check each lac⁺ transformant for the presence of the correct inserts by Grunstein–Hogness colony screening, using a probe made from the target DNA.¹⁰ Since it is about as much work to screen 50 transformants as it is to check 500, it is useful to transform LG90 cells with all of the ligated pMR100 DNA on many plates (e.g., 10 plates from one ligation of 0.5 μg of vector DNA) before the colony hybridization experiments.

Inoculate a fresh MacConkey agar plate containing ampicillin (50 μg/ml) with small (1–2 mm in diameter) patches of cells from each lac⁺ transformant in a numbered grid pattern; use a paper template taped to

the underside of a plate. At the same time, inoculate 2–3 known positive and 2–3 negative control transformants in some asymmetric pattern among the 80–100 unknown transformants on each plate. Grow the cells at 37° for at least 16 hr, and record the intensity of the red color phenotypes of the transformants. The range of the intensity of the red color is often a reflection of the total amount of β-galactosidase proteins produced by the transformant line; this information is helpful in the electrophoretic analysis of the fusion proteins (discussed below). Replica plate the transformants to several nitrocellulose filters and one new plate. Grow the transformants on the nitrocellulose for at least 12 hr and then lyse the cells with alkali and fix the DNA. Replicate the filters with probes made by nick translation of the purified fragments used to make the open reading frame clones. Probes made from whole recombinant lambda phage vectors and most plasmids will hybridize with pMR100, because of shared lac operon, lambda cI, or pBR322 DNA sequences. Any purified fragment greater than 400 bp in length can be labeled by nick translation; small fragments can be nick translated easily by making concatemers of them by self-ligation. Expose the hybridized and washed filters to X-ray film with an intensifying screen at −70° for 3–4 hr; positive signals should be clearly distinguishable from the negative controls.

The proportion of the lac+ transformants that have inserts derived from the correct DNA source is highest when the total insert frequency is high and there is some open reading frame sequence present in the target DNA. The lac+ clones that do not hybridize to the target DNA probe are caused by the following artifacts: (1) A single base deletion at the Smal cloning site of pMR100 that results in correction of the frameshift in the cI::lacI fusion gene. Red transformants of the control plasmid pMR200 are an example of these plasmids; these transformants are distinctive in that they always produce high levels (higher than most open reading frame clones) of β-galactosidase and have a very dark red MacConkey agar phenotype. These colonies are almost certainly caused by exonuclease contaminants either in the Smal restriction enzyme or the T4 DNA ligase. The proportion of these clones is usually less than 1% in transformations with the ligated vector without inserts. (2) Rare plasmid deletions that put the fusion gene in frame. (3) Correction of the frameshift by the insertion of an open reading frame fragment derived from the wrong DNA. Common sources of contaminating DNA are the largely open reading frame vector sequences and bacterial chromosomal DNA.

An alternative approach which is occasionally available is screening

---

the lac$^+$ transformants by using polyclonal antibody preparations directed against the protein sequences that are expected to be expressed as fusion proteins in the open reading frame clones.$^{12}$

**Analysis of Fusion Proteins**

Some of the insert-containing lac$^+$ pMRI00 transformants do not produce full-length (cI::inserted polypeptide sequence::lacI) fusion proteins, because of translation initiation within the inserted RNA sequence (discussed below). For this reason, it is important to examine the β-galactosidase polypeptides by electrophoresis on Laemmli SDS gels.$^{13}$

Prepare protein samples from 2 ml overnight cultures (L broth plus 50 μg/ml ampicillin) of each transformant. Transfer the cells to 10 × 75 mm Pyrex glass tubes and centrifuge at 7,000 rpm for 5 min. Discard all of the L broth and completely resuspend the cells in 100 μl of 1.2× Laemmli sample buffer. Lyse the cells by incubation at 100° for 3 min. If the lysate is not clear, the cells are not completely lysed; large amounts of cell debris interfere with resolution of large proteins on the SDS gels. Often, the cause of incomplete lysis is decomposition of the 5× Laemmli sample buffer. The clear protein lysates are extremely viscous and hard to pipette because of bacterial chromosomal DNA. The high viscosity can be reduced either by aspiration of the lysate through a long 22 gauge needle several times, or by brief sonication. The protein lysates can be stored for several months at −20°.

Prepare a 0.6-mm-thick 7.5% acrylamide SDS gel with a 4.5% acrylamide stacking gel, using a comb with wells 4 mm wide. Electrophorese 4 μl of each protein sample (along with samples prepared from both pMR100 and pMR200 transformants) as follows: (1) Run the tracking dye all the way through the stacking gel at 100 V (about 30 min). (2) Increase the voltage to 200 V for 90 min. (3) Increase again to 250 V; let the dye front run off and electrophorese for 30 min more (if the running gel is about 11 cm long). Stain the gel for at least 1 hr in 0.25% Coomassie blue in 50% methanol plus 10% acetic acid. Destain in 50% methanol plus 10% acetic acid until the background is clear (several hours). Fix the gel in 10% acetic acid for long-term storage or drying.

An example of a protein gel prepared as described above is shown in Fig. 2. The β-galactosidase band is visible as the only high molecular weight protein that varies both in size and abundance among the samples.


FIG. 2. Gel electrophoresis of fusion proteins. Samples (4 µl) of protein minilysates prepared from various transformants were electrophoresed on a 7.5% acrylamide/SDS gel and stained with Coomassie blue as described. Lanes: 1, LG90; 2, 17, pMR100; 3, 20, pMR200; 4-11, Ubx ORF clones UBX7, UBX17, UBX35, UBX27, UBX20, UBX10, UBX12, and UBX13, respectively (Table I; Fig. 3); 12-14, CPV ORF clones CPV3, CPV39, and CPV170, respectively; 15, UBX102; 16, UBX103; 18, CPV6; 19, CPV27. Arrows: lower, the position of wild-type β-galactosidase; upper, the pMR200 cl::lacI fusion protein (lanes 3, 20).

The staining intensity of the β-galactosidase proteins often correlates with the degree of red color of the transformant colony on MacConkey plates.

The β-galactosidase proteins produced by transformants with weak lac⁺ phenotypes are occasionally difficult to see by Coomassie blue staining of the gel. Protein blots (also termed Western blots) are useful for the analysis of clones of this type. In Western blot analysis there is no confusion caused by co-migrating proteins; long exposure times allow visualization of very small amounts of β-galactosidase fusion proteins. We have used anti-β-galactosidase and anti-λ cl as probes to visualize proteins encoded by the pMR100 fusion gene.

Any protein blot procedure is sufficient; we have used the procedure of Towbin with consistent success. SDS gels are prepared as described

above, and then electroblotted to nitrocellulose after electrophoresis. The protein blots are incubated with anti-β-galactosidase, anti-cI, or antibodies against the polypeptide sequence encoded by the inserted DNA sequence. Finally, the blots are incubated with ¹²⁵I-labeled protein A, washed, and exposed to X-ray film.

**Analysis of the Open Reading Frame DNA Insert Fragments**

For several reasons, it is useful to make plasmid DNA from 1-ml cultures of each open reading frame clone. This DNA can be used to regenerate the clone even after years of storage. The length of the ORF insert fragment can be determined easily by digestion of the plasmid with BamHI. The SmaI cloning site is between the only two BamHI sites in pMR100; the small BamHI fragment from each recombinant BamHI contains only 10 bp of vector sequence. In this way, the blunt-ended, randomly cut fragments now have BamHI ends. (If the inserted ORF fragment has a BamHI site, then two small fragments will be found after digestion with BamHI.) The insert fragment can be easily sequenced by recloning the BamHI fragments into sequencing vectors such as those of the M13 or pEMBL series. Alternatively, the inserts can be sequenced in pMR100, without recloning, by using double-stranded DNA sequencing methods. Primers homologous to sequences in pMR100 on either side of the SmaI site can be used. The miniprep DNA can also be used to make nick-translated probes for Southern or Northern blots.

A reliable plasmid DNA miniprep protocol is the alkaline lysis method. Resuspend the DNA pellet (about 1 µg of DNA) from a 1 ml saturated culture of transformed cells in 25 µl of TE. In order to visualize the DNA insert easily, digest 5 µl of the plasmid DNA with 2 units of BamHI in a total volume of 10 µl for 60 min at 37°C in the presence of a small amount (0.5 µl of a 0.1 mg/ml stock) of RNAse A. Electrophorese the digested DNA samples along with the appropriate size markers on a 10% acrylamide/TBE gel.

**Preparation of Antibodies Specific for the Inserted Polypeptide Sequence of pMR100 Fusion Proteins**

The β-galactosidase fusion proteins produced by pMR100 ORF clones can be purified easily and used as immunogens for the production of antibody probes. Many different protein purification strategies can be

---


used, taking advantage of the large molecular weight of β-galactosidase or its substrate specificity. Described below is a fusion protein and immunization protocol; it is certainly not the only strategy or necessarily the best one for raising antibodies directed against the inserted polypeptide sequence of pMR100 fusion proteins, but it has worked well in our hands.

*Preparation of the Bacterial Total Protein Lysate.* We have prepared total protein lysates from bacteria using a modification of a standard protocol. Inoculate 1-liter cultures of each LG90 transformant in L broth with 40 μg/ml ampicillin; shake at 37° until log phase (A₅₆₀ = 0.6). Measure the β-galactosidase enzyme activity using the protocol described by Miller. In order to follow the fusion protein during subsequent purification steps, activity assays should be performed on small aliquots; for this reason, it is useful to know the amount of β-galactosidase initially present in the culture. After growing the cells, all subsequent steps should be performed at 4°.

Centrifuge the cells at 7,000 rpm for 5 min; discard the supernatant medium and resuspend the cells in 400 ml of buffer B [200 mM Tris, 250 mM NaCl, 10 mM MgAc₂, 10 mM 2-ME (2-mercaptoethanol), 5% glycerol, pH 7.6]. Centrifuge and resuspend the washed pellet of cells in 20 ml buffer B (without glycerol); transfer the cells to 50-ml plastic centrifuge tubes. Add lysozyme to a final concentration of 2 mg/ml; mix and incubate on ice for 30 min. Add phenylmethylsulfonyl fluoride (PMSF; a protease inhibitor) to a final concentration of 1 mM. Quick freeze on dry ice for 10 min and thaw at 25°; repeat the freeze–thaw procedure 3 times to lyse the cells. In order to reduce the viscosity caused by chromosomal DNA, sonicate the lysate 3 times for 30 sec; keep the lysate on ice during sonication. Centrifuge the lysate in 10 ml plastic centrifuge tubes at 35,000 rpm for 30 min at 4° using a fixed-angle ultracentrifuge rotor.

To remove nucleic acid, add to the supernatant 30% streptomycin sulfate equivalent to 10% of the total volume; stir for 30 min on ice. Centrifuge at 7,000 rpm for 10 min; remove the supernatant and mix in an equal volume of 80% ammonium sulfate. Stir for 1 hr on ice; centrifuge at 10,000 rpm for 20 min. Discard the supernatant and resuspend the ammonium sulfate pellet in 8 ml of buffer D (10 mM Tris, 250 mM NaCl, 10 mM MgCl₂, 1 mM EDTA, 10 mM 2-ME, pH 7.6) plus 0.1% Triton X-100; dialyze the resuspended pellet in buffer D plus 0.1% Triton X-100 as follows. Dialyze over a period of 16 hr at 4° using at least 3 changes of 2 liters each. After dialysis, centrifuge the crude protein mixture at 10,000 rpm for 20 min, and save the clear protein supernatant.

---

Purification of $\beta$-Galactosidase by Affinity Chromatography. Proteins with $\beta$-galactosidase activity can be purified using the substrate analog $p$-aminophenyl-$\beta$-d-thiogalactopyranoside agarose (Sigma, A 8648)\textsuperscript{17}. Prepare a 15-ml column of the substrate analog and equilibrate with buffer D. Pass the crude protein mixture over the column 3 times. Wash the column with 40 ml of buffer D plus 0.1% Triton X-100. Wash the column again with 15 ml of buffer D without Triton X-100 until no more protein comes off the column ($A_{280} < 0.005$). Elute the $\beta$-galactosidase with 100 mM sodium borate (pH 10.0); collect 30 1-ml fractions. Assay $\beta$-galactosidase activity in 0.1 $\mu$l of each column fraction and 100 $\mu$l of the flowthrough fraction in buffer D; pool all of the fractions that include $\beta$-galactosidase and adjust the pH carefully to 7.5 with 0.1 N HCl. Most of the $\beta$-galactosidase activity is usually found in fractions 10–16. Precipitate the protein by mixing in an equal volume of 80% ammonium sulfate; stir on ice for 2 hr and pellet by spinning at 10,000 g for 20 min.

Usually, about 80–90% of the precipitated protein is the $\beta$-galactosidase fusion protein. Resuspend the pellet in PBS (10 mM sodium phosphate, 0.8% NaCl, pH 7.5) and dialyze overnight in PBS as described above, before final preparation for injection. If a higher level of homogeneity is desired, it is useful to purify further the fusion protein by gel electrophoresis.

Preparative SDS Gel Electrophoresis of Fusion Proteins. Resuspend the ammonium sulfate precipitate (after the affinity column) in 20 ml of PBS plus 0.1% Triton X-100 plus 10 mM 2-ME. Dialyze overnight in PBS plus Triton plus 2-ME as described above. Add 125 $\mu$l of Laemmli 5× sample buffer to 0.5 ml of the dialyzed protein and boil at 100° for 3 min. Load the sample across the full width of a 5% acrylamide SDS gel that is 1.5 mm thick, 16 cm wide, and 18 cm long.\textsuperscript{13} Electrophorese at 200 V for 4–5 hr until the tracking dye has reached the bottom. Usually 0.5 ml of the resuspended pellet contains 150–200 $\mu$g of protein; if the fusion protein is $\sim$100 $\mu$g, then the final recovery from the gel is usually about 50–70 $\mu$g. We have used 300 $\mu$g of fusion protein for immunization; four or five preparative gels should be run. After electrophoresis, cut a small vertical slice of the gel and stain it in 0.25% Coomassie blue as described above. Align the stained gel slice with the unstained gel in order to locate the region that contains the fusion protein band. Cut out the fusion protein band and crush the gel slice using a glass homogenizer; allow the protein to elute by diffusion overnight into 2 ml of PBS. In order to estimate the final recovery and check for homogeneity, electrophorese 4–5 $\mu$l of the protein on an analytical SDS gel as described above. The denatured protein eluted from the gel does not have any $\beta$-galactosidase activity.

Remove the acrylamide fragments by passing the eluted protein mixture over a small glass-wool column; wash the column with 1 ml of PBS.
Precipitate the fusion protein by adding 4 volumes of cold (−20°C) acetone and leaving the mixture at −20°C overnight. Centrifuge at 18,000 rpm for 30 min; discard the supernatant and resuspend the dry protein pellet in a volume of PBS suitable for immunization.

Immunization of Rabbits. We have immunized rabbits with pMR100 fusion proteins, using the protocol below. Monoclonal antibodies specific for determinants of the inserted polypeptide sequence have also been successfully isolated using pMR100 fusion proteins as antigens.19

Before the initial injection, collect a pre-immune blood sample. For the initial injection, mix 200 μg (in 1–2 ml PBS) of the fusion protein with an equal volume of complete Freund's adjuvant (Difco Labs, 0639-60-6); inject the protein subcutaneously into several regions of the back and neck of the rabbit. Six weeks after the initial immunization, inject 50–70 μg of fusion protein in 1 ml PBS mixed with an equal volume of incomplete Freund's adjuvant (Difco Labs, 0638-60-7). Later booster injections can be given 3–4 weeks apart.

Rabbits are bled (40 ml) 10 days after each booster immunization; immunoreactivity of the serum is estimated by ELISA assay using purified β-galactosidase (Sigma) and/or purified fusion protein as antigen.20 Immunoreactivity is usually observed after the first booster injection; maximal immune response levels are usually achieved after the third boost.

Purification and Characterization of the Antibodies. Allow the blood samples to clot at room temperature for 2 hr and then at 4°C for 12–16 hr. Spin down the clot at 10,000 rpm for 10 min, and remove the serum. Precipitate the proteins from 1 ml of serum by mixing in 1 ml of saturated ammonium sulfate; incubate at 4°C for 1–2 hr with constant stirring. Centrifuge at 10,000 rpm for 20 min; resuspend the pellet in 20 ml of phosphate buffer (17.5 mM sodium phosphate, pH 6.5) and dialyze in phosphate buffer as described above.

The serum proteins can be further enriched for IgG molecules by passing the dialyzed sample over a DEAE–cellulose (Whatman DE-52) column; many non-IgG proteins bind to DEAE. Prepare a 1-ml column; equilibrate with phosphate buffer. Check the pH and conductivity of the phosphate buffer before and after the column in order make sure that the column is properly equilibrated. Add the dialyzed serum proteins to the column and collect 0.5-ml fractions; assay the fractions for proteins by optical absorbance (A280). Pool the protein fractions and dialyze overnight in BSB (150 mM NaCl, 17.5 mM boric acid, pH 7.5) as described above.

Antibody molecules specific for the pMR100 recombinant fusion protein are purified by their ability to bind to the fusion protein used as the immunogen. Prepare an Affigel-10 (Bio-Rad Chemicals, #153-6046) column with the fusion protein coupled to the gel. Couple 500 μg of substrate analog-purified fusion protein to 0.5 ml of Affigel-10 following the manufacturer’s directions. Wash the column with 200 mM glycine–HCl plus 500 mM NaCl, pH 2.5, and then neutralize with BSB. Treat the column with 1 ml of BSB plus 0.1% SDS (SDS treatment is discussed below) and equilibrate the column with BSB. Pass the rabbit serum proteins (in BSB) over the column 3 times at 4°C; only antibody molecules specific for any portion of the fusion protein will bind to the column. Wash the column with 10 ml of BSB; wash the column again with 2 ml of BSB plus 500 mM NaCl in order to remove nonspecifically bound or low-avidity antibodies. Elute the bound antibody molecules with 5 ml of 200 mM glycine–HCl, pH 2.5 plus 500 mM NaCl; collect 12 0.5-ml fractions; neutralize (pH 7.0) each of the antibody fractions immediately by the addition of 200 mM Tris base. Add bovine serum albumin (BSA) to a final concentration of 0.5 mg/ml to each of the fractions; dialyze the first 6 fractions separately in BSB as described above. Assay all fractions and the flowthrough sample using the ELISA assay with both β-galactosidase and recombinant fusion proteins as antigens. Usually, fractions 2–5 have most of the anti-fusion protein antibody; pool all the antibody-containing fractions after dialysis.

To remove antibody molecules specific for the λ cI repressor and E. coli β-galactosidase antigens and recover only those specific for the inserted polypeptide sequence, the dialyzed antibody preparation above is allowed to adsorb to Affigel-10 coupled with the pMR200 fusion protein. The pMR200 fusion protein contains only the cI and lacIZ portions of pMR100 recombinant fusion proteins (Fig. 1). Prepare the pMR200 Affigel-10 according to the manufacturer’s directions; equilibrate with BSB. Mix 250 μl of pMR200–Affigel-10 with the antibody preparation; shake gently for 1 hr at 4°C. Centrifuge at 2,000 rpm for 5 min; repeat the adsorption of the supernatant with a second aliquot of pMR200–Affigel-10. Antibodies specific for cI and lacIZ antigens will bind to the pMR200–Affigel-10 and those specific for the inserted polypeptide will be left in the supernatant. Assay the final antibody preparation using the ELISA method with the recombinant fusion protein, pMR200 fusion protein, and β-galactosidase as antigens. The specificity of the antibody can be tested using Western blots, as discussed above,14 or Western dot blots21 (Fig.

Fig. 3. (A) Preparation of antibodies specific for the inserted polypeptide sequence of pMR100 fusion proteins. Duplicate bacterial cell lysate samples (each containing approximately 1 ng of fusion protein) treated with 0.1% SDS were blotted onto nitrocellulose. The blot on the left was incubated with affinity-purified antibodies to the inserted polypeptide sequence. The blot on the right was incubated with pre-immune serum subjected to the same purification procedure as that of the immune serum. Both blots were then incubated with 125I-labeled protein A, washed, and exposed to X-ray film with an intensifying screen at -70°. (B) Effect of SDS on the antigenicity of pMR200 fusion proteins. Duplicate fusion protein samples (as above) in either PBS or PBS plus 0.1% SDS were blotted onto nitrocellulose; samples of PBS and PBS plus 0.1% SDS with no proteins were also blotted. The blot on the left was incubated with a 1:500 dilution of rabbit immune serum; the blot on the right was incubated with pre-immune serum diluted 1:500. Both blots were then incubated with 125I-labeled protein A, washed, and exposed to X-ray film.

The dot blots are prepared using a minifold apparatus (Millipore). The final volume of purified antibody is usually 1–2 ml; dilute to 1:300 for Western blots.

It has been reported that proteins purified by SDS–gel electrophoresis are strong immunogens and can be used to generate high-titer antisera.22 In one well-studied case in our experiments, when the immunogen used was pMR100 fusion protein purified on SDS gels, the resultant polyclonal rabbit antiserum recognized SDS-treated antigens much better than untreated antigens. When the antigen is mixed with 0.1% SDS, the antibody–antigen reaction is 5-fold higher using either ELISA assays or Western type dot blots (Fig. 3B).

Examples

Location of Open Reading Frames in Uncharacterized DNA. Selection of open reading frame fragments using pMR100 has been used to characterize DNA from the bithorax locus in Drosophila melanogaster.23

Over 95 kb of cloned DNA from the *Ubx* portion of the locus was sonicated and fragments longer than 200 bp ligated to pMR100 as described above. Lac⁺ transformants were screened for the correct DNA inserts by colony hybridization using the appropriate *Ubx* probes. From 60,000 transformants, 1200 were lac⁺ (2% of the total). Approximately one-half of these had *Ubx* DNA inserts, and the proportion of these that made full-length fusion proteins varied from 0 to 65%, depending on the region of the *Ubx* DNA from which the insert originated. All clones with inserts longer than 300 bp, as well as some smaller ones, were selected for fusion protein analysis; 61 of these produced full-length fusion proteins that include the inserted *bithorax* polypeptide sequence. Each of these ORF DNA inserts was mapped to restriction fragments in *Ubx* DNA by Southern blot analysis; the results are summarized in Fig. 4.

The only known protein-coding regions of *Ubx* are found within 4 exons spread over 75 kb (Fig. 4).24,25 The 5' exon contains the first 840 bp of the long ORF found in *Ubx* cDNA clones. The two central miniexons are each 51 bp in length; the 3' exon contains only 200 bp of open reading frame sequence. There are some sequence data from other regions within the *Ubx* transcription unit; no additional open reading frames greater than 350 bp in length have been found.26

There are two substantial clusters of ORF fragments found within *Ubx*. One cluster is in the restriction fragment that includes the 840-bp ORF at the 5' exon; the longest ORF fragments (including one of 569 bp) map at this location. The other cluster of fragments maps in a region 15 kb downstream from the 5' exon within the second intron; the longest of these clones is 335 bp in length. There is no other molecular evidence of which we are aware that suggests that a protein-coding exon is present at this position. Most of the remaining fragments map outside of the DNA known to be part of the *Ubx* transcription unit; those within *Ubx* are all less than 300 bp in length and are scattered throughout the *Ubx* region. These results predicted that a long open reading frame exists at the genomic position shown by cDNA mapping to contain the main part of the protein-coding region of the gene. Few or no significant ORF sequences exist at other positions in *Ubx*, consistent with current molecular data indicating a lack of other long protein-coding regions in *Ubx*.25 These results suggest that the approach of "shotgun" cloning into pMR100 to locate ORF sequences in a large region of uncharacterized DNA can provide a reasonably accurate picture of the open reading frame sequence

---

25 M. O'Connor and W. Bender, personal communication.
Fig. 4. Position of ORF fragments on the molecular map of the Ubx region of the bithorax complex in Drosophila melanogaster. The map coordinates are in kilobases; the short vertical lines on the map denote EcoRI sites. The positions of the four exons found in Ubx cDNA clones are shown below the restriction map; the diagonal lines represent the intron regions. The positions of the Ubx, abx, and bx regions are summarized above the map. The numbers indicate the positions and lengths of the ORF fragments.
distribution. This approach for mapping of ORF regions has also been useful in the analysis of the serendipity locus of Drosophila.27,28

Expression of Portions of a Well-Characterized Open Reading Frame Sequence. In order to see if many (or all) regions of an entirely open reading frame DNA fragment can be expressed in pMR100 as described above, a 1423-bp fragment that encodes part of the virus coat protein of canine parvovirus (CPV) was used to prepare insert fragments.29 All lac+ transformants with CPV inserts (300 bp or longer) were analyzed further. Of 192 lac+ transformants with CPV inserts, 44 made full-length fusion proteins. These were mapped within the 1423-bp CPV fragment by restriction site mapping or by DNA sequencing. The open reading frame fragments mapped to all regions of the sequence, suggesting that no part of CPV sequence necessarily gives rise to an unstable fusion protein (Fig. 5).

Analysis of lac+ Transformants That Produce Small Fusion Proteins. In both the Ubx and CPV experiments, most of the lac+ clones that were shown to have the appropriate inserts by colony hybridization produced β-galactosidase proteins that were shorter than expected—frequently the size of β-galactosidase without any additional fused polypeptide sequences, including the λ cI portion of the pMR200 fusion protein (Fig. 2). Three of the Ubx and 7 of the CPV lac+ clones that made only small proteins were studied further, by recloning the inserts in pEMBL9 and sequencing the inserts. All 10 fragments contained the wrong number of bases needed to correct the pMR100 frameshift and lacked a continuous open reading frame in either orientation. All 10 fragments carried an internal ATG codon followed by an open reading frame in phase with that of the downstream lacIZ portion of the pMR100 fusion gene. Upstream from these ATG codons were short sequences that matched or closely resembled bacterial Shine–Delgarno sequences.30 This suggested that β-galactosidase proteins lacking the cI part of the protein sequence and much of the insert sequence might be the result of the initiation of translation within the insert. Indeed, the sizes of the small proteins suggested that this was the case; the distance between the ATG codon and the 3' end of the insert correlated well with the increase in the size (relative to that of the wild-type protein) of the β-galactosidase protein (Fig. 6). The data suggest that if the DNA fragment has a sequence that can cause the initiation of translation in bacteria at a nearby downstream ATG codon, it
will give rise to a lac⁺ pMR100 transformant that makes a small β-galactosidase protein. If an open reading frame fragment of the correct size (with $3n + 2$ bp) and in the correct frame at both ends has a Shine–Delgarno sequence upstream from a nearby ATG codon which is in the correct reading frame, then two sizes of β-galactosidase proteins might be present—a large full-length protein and a small protein with an amino terminus corresponding to the ATG codon within the insert. The open reading frame insert of one of the CPV clones (CPV170) that produced both large and small proteins was sequenced and shown to fulfill these sequence requirements (Figs. 2 and 6).

The only requirements for fragments to cause the initiation of translation in the insert sequence of the pMR100 fusion gene mRNA are the insertion in the right orientation and in the proper reading frame at the junction of the 3' end with the lacIZ portion of the fusion gene. The variability in the fraction of the total lac⁺ clones that make small proteins with the DNA source is most likely a reflection of the relative abundance of the Shine–Delgarno-like sequences and open reading frames in the source DNA.

The clones that produce only small β-galactosidase proteins often are weakly lac⁺ on MacConkey plates and produce small amounts of β-galactosidase; the CPV example discussed above is exceptional (Table 1; Fig. 

---

**FIG. 5.** Molecular map positions in the 1423-bp CPV coat protein gene fragment of ORF fragments expressed in pMR100.
FIG. 6. (A) Molecular map positions of CPV fragments that cause initiation of translation within the inserted sequence of the pMR100 fusion gene mRNA. The numbers above each fragment identify the clone; the numbers at the left end of each fragment represent the number of bases needed to join correctly the CPV ORF with that of the cI portion of the pMR100 fusion gene. The designation “OK” indicates that the CPV and pMR100 fusion gene sequences are joined in the correct frame. (B) DNA sequences that might act as translation start sites in pMR100 recombinant plasmids. The CPV sequence is the same as that in (A); the Ubx sequences are from three different cloned fragments, as indicated. It has been proposed that the translational start sequences, termed Shine–Delgarno sequences, form hybrids with the 3‘ end of 16 S rRNA during translation; the sequence of the E. coli 16 S rRNA thought to participate in mRNA binding is shown.
2). The ratio of the intensity of the β-galactosidase bands in the CPV170 sample suggests that the CPV translation start sequence is as efficient as that of the pMR100 fusion gene. The variability in the amount of the small β-galactosidase protein made by a lac⁺ transformant might be a reflection of how efficiently the sequence in the insert can initiate translation in bacteria. The exact sequence of the Shine–Delgarno sequence and the neighboring DNA, as well as the distance from a downstream ATG codon, affect the rate of translation to a large extent. Alternatively, the in vivo stability of the small and large proteins may vary significantly.

To date, all of the clones we have analyzed that make small proteins have inserts with sequences that match those found associated with the initiation of translation. Thus, it is likely that this is the most prevalent cause for the small proteins in pMR100 lac⁺ transformants, rather than protein degradation in vivo, as suggested previously. The incidence of an ATG codon with a nearby upstream Shine–Delgarno sequence should vary with each DNA source used; if these sequences are present in a larger region of DNA that includes an open reading frame, then the frequency of small protein clones should be at least 3 times higher than that of the large protein clones, because the latter is dependent on the proper reading frame at both the 5' and 3' ends of the insert fragment. The translation-initiating fragments need to have the correct configuration only at the 3' end. Because of the large variety of very short sequences that have been identified as Shine–Delgarno sequences in bacteria, it is not surprising that eukaryotic DNA has many sequences that can act as translation initiators when placed in the appropriate context in bacterial genes.

Comments

As discussed above, "shotgun" cloning of randomly cut fragments from uncharacterized DNA into pMR100 can be used for the mapping and expression of many and in some cases all of the open reading frames in the initial DNA. The only limitation of which we are aware is that with some DNA sources many of the lac⁺ pMR100 clones synthesize β-galactosidase without expressing an open reading frame DNA insert, due to internal translation starts. In most of our experiments, only about 25% of the insert-containing lac⁺ pMR100 clones make full-length fusion proteins (because of the presence of translation-initiating sequences); this proportion may vary from 0 to 100% depending on the relative abundance of open reading frame and translation-initiating sequences in the substrate DNA. In some cases, the transformants that synthesize the shortest of the small β-galactosidase proteins can be avoided by screening the lac⁺ colo-
The β-galactosidase fusion proteins produced by recombinant pMR100 transformants can be used as antigens for the production of antibodies; this has been successful for *Drosophila* proteins encoded by the *Ubx* and *per* loci,19,31 trypanosome surface glycoproteins,32 and canine parvovirus coat protein.33 The strategy of shotgun cloning in pMR100 has also been used to quickly express parts of genes from uncharacterized viral genomes, such as that of HTLV III.34

Acknowledgments

We would like to thank our colleagues Hildur Coiot, Linda Gritz, Julia Sue, and Tobie Tishman for their advice and assistance, and Gabrielle Peterson for performing ELISA assays. This work was supported by grants from the National Institutes of Health (GM33205 and HD08887).

31 P. Reddy and M. Rosbash, unpublished experiments.
33 G. Mazzara, unpublished experiments.

[9] Use of Open Reading Frame Expression Vectors

*By George M. Weinstock*

Introduction

Producing foreign polypeptides in *Escherichia coli* poses a variety of problems. These include the proper joining of bacterial transcription and translation initiation signals to the foreign coding sequence and assaying for expression. Open reading frame (ORF) expression vectors provide a general solution to these problems. ORF vectors do not express an intact gene; rather, only a part of the coding sequence, lacking translation termination codons, is expressed to produce a polypeptide representing a part of the complete foreign protein. Such partial proteins can be used to produce antibodies which can then be used to detect the complete protein in its natural host. It is not necessary to know the DNA sequence of the foreign coding region in order to use ORF vectors to express it in *E. coli*.