A direct selection strategy for shotgun cloning and sequencing in the bacteriophage M13

Richard A. Guilfoyle and Lloyd M. Smith
Department of Chemistry, University of Wisconsin, Madison, WI 53706, USA

Received June 15, 1993; Revised and Accepted October 26, 1993

ABSTRACT

A new cloning strategy is described which utilizes direct selection of recombinants for shotgun sequencing in the filamentous bacteriophage M13. Direct selection is accomplished by insertional inactivation of the M13 gene X protein, a powerful inhibitor of phage-specific DNA synthesis when overproduced. An extra copy of gene X was inserted into the intergenic region of M13 and placed under the control of the bacteriophage T7 gene 10 promoter and RBS. Random fragments are cloned into the EcoRV cloning site of the new gene X cistron and recombinants are selected in an E.coli male strain producing T7 RNA polymerase. Cloning efficiencies obtained with M13-100 or phosphatase-treated M13mp19 vector are comparable. The direct selection capability of M13-100 was demonstrated to have the following advantages: (a) consistently achieved high ratios of recombinants to religated vector in the libraries, averaging about 500:1 (0.2% background), and (b) the elimination of the need for phosphatase treatment of the vector to reduce background. The direct selection strategy significantly improves the efficiency of shotgun library construction in M13, and should therefore facilitate the cloning aspects of large scale sequencing projects.

INTRODUCTION

A goal of our research is to automate the ‘front end’ (1) of the DNA sequencing process for large scale sequencing projects. The front end of most large scale DNA sequencing strategies depends on randomly fragmenting the target molecule into small pieces which can be subcloned into a bacteriophage like M13 (2–6). These vectors produce template DNA in a single stranded form, which can be subcloned into a bacteriophage like M13 (2–6). The data obtained from such ‘shotgun cloned’ fragments are combined and overlapped until approximately 80–95% of both strands are covered, after which gap filling techniques may be utilized to complete the sequence (4). The shotgun approach (2,3) may require up to 500 million templates to obtain closure of the entire human genome. Automation and streamlining is clearly essential to the success of such efforts. In this manuscript, a strategy is described which should substantially simplify the cloning aspects of large-scale sequence analysis.

The present method for shotgun cloning is to ligate randomly fragmented DNA into M13mp18/19 type vectors (2,8). The primary advantage of such filamentous bacteriophage is the ease with which highly purified single stranded DNA can be obtained. A disadvantage is the indirect means by which recombinant clones are distinguished from non-recombinants using a marker (β-galactosidase) inactivation assay. Two rather expensive chemicals (X-GAL and IPTG) are prepared and mixed with the growth media. Non-recombinant clones growing on this media are visually distinguished as blue plaques, whereas recombinant clones are colorless. This assay can prove subjectively difficult to interpret. Another problem is the variable amount of non-recombinant vector which can survive, lowering the overall efficiency of gene library construction. This problem may be minimized by introducing yet another step, the removal of the phosphate group from the vector 5'-ends using alkaline phosphatase which lowers the chance of vector religating to itself, thus reducing the background of parental plaques. This step is also variable, but when efficient can eliminate the need for color discrimination in shotgun library constructions. Nevertheless, the multiple steps required generally makes cloning in this particular system difficult to automate. What is needed is a more efficient or ‘one-step’ process which allows direct or ‘positive’ selection of recombinants.

We have developed a novel bacteriophage M13 cloning strategy which utilizes a new vector called M13-100 and permits the direct selection of recombinant clones. With this strategy recombinant phage have the highest probability of replicating efficiently and thereby being detected in the selection host employed. All of the plaques which develop can thus be utilized for DNA sequence analysis without fear of sequencing appreciable amounts of parental vector DNA. In addition, the requirement for phosphatase treatment has been eliminated. The direct selection capability of the M13-100 vector is achieved by taking advantage of the DNA synthesis ‘inhibitory’ or ‘repression’ activity of M13 gene X. The gene X product is one of only three phage proteins needed for phage DNA replication. Genes II and V encode the other two proteins (for a review, see ref. 9). The gene X protein is 144 amino acid residues in length and is identical to the C-terminal third of the gene II protein. Its translation initiates at codon 300, AUG, of gene II and is required for efficient accumulation of viral DNA export (10). The gene X protein is a powerful inhibitor of phage-specific DNA synthesis in vivo and is thought to operate at the earliest stages of DNA replication, thus giving the protein a regulatory function without which the infected cell is incapable of accumulating progeny ssDNA particles. If over-produced, the resulting high level of gene X protein blocks phage-specific DNA synthesis (11).
The vector was designed to allow over-expression of an extra copy of gene X inserted into the M13 intergenic (IG) region near the origin of replication. Placed under control of the bacteriophage T7 gene 10 promoter region (12,13), high levels of replication repressor activity are then obtained in the JM109(DE3) host, a male strain of E.coli producing T7 RNA polymerase. Random foreign DNA fragments can be inserted into the new gene X cistron such that its expression is disrupted. Recombinant phage are then specifically selected by growth on the JM109(DE3) cells.

MATERIALS AND METHODS

Growth of phage and bacterial strains

Wild type M13 phage was obtained from ATCC (Rockville, MD) and served as the source of phage needed for the construction of M13RV and M13SPS1, parent vectors to M13-100. E.coli male strains JM109 and JM109(DE3) were purchased from Promega Corp., Madison, WI. The genotype of JM109 is endA1, recA1, gyrA96, thi, hsdR17 (rK-mK-), relA1, supE44, (lac-proAB), (F', traD36, proA, lacI, ΔM15). JM109(DE3) contains a stable chromosomal integrant of recombinant lambda phage DE3 which contains T7 bacteriophage gene I encoding T7 RNA polymerase under the control of the inducible lacUV5 promoter (12,13). For the preparation of M13 plaques and high-titer supernatants, bacterial infections were done in either LB or 2XYT growth media using established protocols (14).

DNA purification

Single stranded M13 DNA was purified from high-titer phage supernatants using Sephaglas (Pharmacia LKB Biotechnology, Piscataway, NJ) and served as template for both PCR amplification of the M13 gene X region and for DNA sequencing reactions. Double stranded replicative form (RF) DNA midi-preps were prepared from phage infected cell pellets of 100 ml liquid cultures using a plasmid purification kit purchased from Qiagen, Inc. (Chatsworth, CA). The RF DNA served as the cloning vectors after linearization by restriction digestion. M13 phage mini-preps were prepared from cell pellets of 1.5 ml infected cell cultures using the Magic Miniprep DNA Purification System™ (Promega). The pGEMEX-1 (Promega Corp.) was used as the template for PCR amplification of the phage T7 region. Restriction fragments and PCR products were purified after electrophoresis on agarose gels either by electroelution or from low melt agarose gels using the Promega Purification System™ (Promega). The pGEMEX-1 (Promega Corp., Madison, WI) was used for extension of the —35SP oligonucleotide primer (see Fig. 4) that was 5'-end labeled with 32P-ATP (Amersham Corp., Arlington Heights, IL) and T4 polynucleotide kinase (New England Biolabs). Calf intestine alkaline phosphatase (CIAP) (Boehringer Mannheim, (Indianapolis, IN) was used for dephosphorylation of HindII digested M13mp19 RF DNA. Fluorescent sequencing reactions were performed with a PRISM T7 Terminator Single-Stranded DNA Sequencing Kit (Applied Biosystems, Inc. (ABI), Foster City, CA) and analyzed on a ABI 373A automated sequencer.

Construction of M13-100 by SOE

'Splicing by overlap extension' (SOE) (16)) was used to construct the region in M13-100 which confers the direct selection property. This region contains an extra copy of M13 gene X fused to a sequence harboring the phage T7 gene 10 5' flanking region. The fusion was performed using the following oligonucleotide primers (synthesized by the U. WI Biotechnology Core Center): PCR primer XL: 5' AAGGAGATATCATATGAAATTTAT ATACGATTCC. This primer is designed to contain the T7 gene 10 5'-flanking region. The region in M13-100 which confers the direct selection property. This region contains an extra copy of M13 gene X fused to a sequence harboring the phage T7 gene 10 5' flanking region. The fusion was performed using the following oligonucleotide primers (synthesized by the U. WI Biotechnology Core Center):

Enzymes and reactions

Restriction endonucleases were purchased from Promega and New England Biolabs (Beverly, MA). The reaction conditions used for restriction enzymes were those recommended by the manufacturer except when linearizing M13-100 and M13mp19 vector DNAs by EcoRV and HindIII digestion, respectively. In these cases, nominal incubation times (30 min. to 1 hr) and limiting amounts of enzyme (0.1 to 0.5 units/μg DNA) were used to help minimize damage to the cloning site of the vector DNA by potential contaminating exonuclease activity. CviJI was obtained from James Van Eten (Dept. of Plant Pathology, U. Nebraska) and is not yet commercially available. The reaction conditions used were as those previously described (15). AmpliTaq DNA polymerase (Perkin Elmer Cetus, Norwalk, CT) was used for PCR reactions according to the manufacturer's recommended conditions provided in the GeneAmp PCR amplification kit. Optimizations involved varying the primer concentrations between 0.1 and 1.0 μM as well as the MgCl2 concentrations between 0.1 and 0.4 mM. One thermal cycling profile was used for all amplifications: 35 cycles; 94°C, 30 sec.; 55°C, 30 sec., 72°C, 30 sec. All thermal cyclings were performed in the 9600 GeneAmp PCR system (Perkin Elmer Cetus). DNA ligations were incubated overnight at 15°C using T4 DNA ligase (2,000,000 units/ml) purchased from New England Biolabs.

Dideoxy DNA sequencing reactions were carried out by using Bst DNA polymerase (Bio-Rad Laboratories, Richmond, CA) for extension of the —35SP oligonucleotide primer (see Fig. 4) that was 5'-end labeled with 32P-ATP (Amersham Corp., Arlington Heights, IL) and T4 polynucleotide kinase (New England Biolabs). Calf intestine alkaline phosphatase (CIAP) (Boehringer Mannheim, (Indianapolis, IN) was used for dephosphorylation of HindII digested M13mp19 RF DNA. Fluorescent sequencing reactions were performed with a PRISM T7 Terminator Single-Stranded DNA Sequencing Kit (Applied Biosystems, Inc. (ABI), Foster City, CA) and analyzed on a ABI 373A automated sequencer.

The steps in the construction of the M13-100 vector were as follows: (a) M13 gene X (335 bp including TAA) was PCR-amplified from WT M13 template ssDNA using primers XL and XR; (b) the T7 promoter region (100 bp) was PCR-amplified from the pGEMEX-1 plasmid template using primers T7L and T7R; (c) The M13 gene X and T7 promoter region PCR products were purified from agarose gels by electroelution and then fused together in another PCR amplification reaction containing only primers T7L and XR. The final PCR fusion product is called the 'extra gene X cistron'; (d) The PCR fusion product was blunt-end ligated into M13SPSI RF DNA cut with EcoRV. M13SPSI was derived from M13RV (17), a vector previously constructed

Nucleic Acids Research, 1994, Vol. 22, No. 1
by oligo-directed mutagenesis in order to introduce a unique EcoRV restriction between bases 5604 and 5605 within the M13 IG region (18). The M13SPS1 vector was then constructed by ligating a 57 bp oligonucleotide into the EcoRV which contains a new EcoRV site flanked by EcoRI and BamHI sites; (e) electrocompetent JM109 cells were transfected with the ligation reaction by electroporation (according to the protocol described below). Plaques were screened by double digesting mini-prep RF DNAs with EcoRI and BamHI to look for the released fusion product by electrophoresis on an agarose gel. One positive isolate, now called M13-100, was sequenced and shown to contain the extra gene X cistron cloned in the opposite orientation of the endogenous M13 gene X, as depicted in Figure 1. Figure 2 shows the DNA sequence with pertinent controlling elements as well as the flanking BamHI and EcoRI sites originally derived from the M13SPS1 parent vector. The T7 gene 10 promoter sequence is highlighted in italics.

Figure 1. The M13-100 cloning vector. The circular map shows the position of the new gene X cistron relative to the origin of replication and flanking M13 genes: I, II, IV, V, and the natural gene X. (For a complete map, see ref. 9). The linear map shows a schematic blow up of the new gene X cistron with pertinent elements: the bacteriophage T7 gene 10 promoter (T7Pr); the -35 primer used for dideoxy sequencing reactions; the T7 gene 10 ribosome binding site (RBS); The restriction endonuclease cleavage site used for shotgun cloning (EcoRV), and the gene X translation initiation codon (AUG).

Figure 2. The DNA sequence of the M13-100 new gene X cistron. The 451 bp DNA sequence was determined by using the oligonucleotides T7L and XR as primers in dideoxy sequencing reactions. Sequences in bold: The XR, T7L, -35SP primers and the T7 gene 10 ribosome binding site (RBS); Sequences underlined: the gene X start codon (ATG) and stop codon (TAA), the EcoRV shotgun cloning site (derived from XL and T7R PCR primers, see Materials and Methods), the Bgl II and XbaI restrictions sites brought in with the T7 region, and the Bam HI and Eco R1 site originally derived from the M13SPS1 parent vector. The T7 gene 10 promoter sequence is highlighted in italics.

RESULTS

Repression of M13-100 in E.coli expressing T7 RNA polymerase

Table I shows the results of an experiment testing the replication of M13-100 phage on the selection host, JM109(DE3), a lambda lysogen which expresses T7 RNA polymerase under IPTG-inducible lacUV5 control (Promega Corp.). Tenfold serial dilutions of phage from two different high-titer supernatant M13-100 phage stocks (A,B in Table I) were titered in the presence and absence of the inducer (IPTG) on JM109(D3) cells as well as on JM109 cells which do not express T7 RNA polymerase. The results indicate that even without induction, M13-100 replicates very poorly in JM109(DE3) cells compared...
JM109(DE3) cells must be enough to overproduce the gene X repressor (19). Evidently, synthesis of the lac repressor is manifested during the course of infection. The JM109(DE3) cells observed on the JM109(DE3) cells relative to growth in JM109 to 15 thousand fold (Expt. A) reduction in phage titer was 

(0.5/2.0/5.0 ng vector DNA equivalents were plated) A,B: two different hts stocks of Ml3-100 of high titer supernatant (hts) phage stocks.

1. Repression of M13-100 phage replication in JM109(DE3) cells

Table

Shotgun library construction using the M13-100 vector

The experiments discussed above indicated that the M13-100 phage can produce a ‘self-inhibitory’ infection in the JM109(DE3) cells producing T7 RNA polymerase. This strongly suggested that the polymerase was responsible for the overproduction of the gene X protein (pX) and subsequent inhibition of phage-specific DNA synthesis, as predicted by the experiments of Fulford and Model (11). We reasoned that pX overexpression would be ‘derepressed’ and replicate efficiently in JM109(DE3) cells, whereas phage containing only religated vector would remain self-inhibitory. The simplest strategy to test M13-100 as a cloning vehicle consisted of placing a unique restriction site between the Shine–Dalgarno (RBS) sequence and the gene X start codon. Insertion of foreign fragments in this position might sufficiently disrupt gene X overexpression by distanc ing the gene from its controlling elements. In fact, this approach was originally proposed for cloning into M13mp1, the parent to the mp vector series (20). An EcoRV site (see Figs. 1,2) was introduced to permit the ligation of blunt-end fragments.

Two approaches were taken in order to assess the efficacy of M13-100 as a shotgun cloning vehicle: (a) comparison with cloning in the widely used M13mp19 system, and (b) analysis of several independently constructed M13-100 libraries. In this way, the following important parameters could be evaluated:

(a) cloning efficiency, defined here as the total number of plaque forming units (pfu) obtained per 50 ng of vector DNA used to transfect the selection host. Cloning efficiency can be affected by numerous variables, particularly efficiencies of ligation and transformation (2, 4,14,21).

(b) the ratio of recombinant (R) to background (B) clones in a library, where recombinants are defined as clones containing target inserts, and the background clones consists of vector devoid of target DNA. Hereafter, this ratio will be referred to as the R/B ratio. A ‘direct’ selection vector should consistently yield high R/B ratios. The number of recombinants should be high enough to assure a ‘representative gene library,’ or 5—20× coverage of the target DNA sequence, assuming the ability to acquire 350—500 bp of raw sequence per clone (22). The number of background clones should be low enough to render the effort of sequencing religated vector DNA negligible.

(c) the degree of ‘false negatives,’ in the foreground population. These are recombinants which go undetected in the library because they either replicate poorly or contain inserts which are either selected against or actually lethal to the cell. There may be a strong bias toward certain inserts, which cannot be directly measured, but it will be manifested as gaps or regions of thin coverage in the overall target DNA sequence. The fraction of false negatives can contribute to lowering R/B ratios.

to JM109 cells. In these experiments, a 7 thousand (Expt. B) to 15 thousand fold (Expt. A) reduction in phage titer was observed on the JM109(DE3) cells relative to growth in JM109 cells. The basal level of expression of T7 RNA polymerase in JM109(DE3) cells must be enough to overproduce the gene X protein so that its inhibitory function can be dramatically manifested during the course of infection. The JM109(DE3) cells harbor the I9 mutation which results in overproduction of the lac repressor (19). Evidently, synthesis of the lac repressor is somewhat leaky in these cells such that the strong lacUV5 promoter is not completely inactivated. This inhibition of phage replication was further increased several hundred fold after inducing an even higher level of T7 RNA polymerase production by the addition of IPTG. That is, the drop in titers were augmented to 6 million and 4 million fold in experiments A and B, respectively. These observed levels of repression represent the ideal limits of background when using M13-100 as a cloning vector. As a control, the level of growth of the M13SPS1 parent phage is shown to be unaffected by either the presence of T7 RNA polymerase or its induction with IPTG.

### Table 1. Repression of M13-100 phage replication in JM109(DE3) cells

<table>
<thead>
<tr>
<th></th>
<th>JM109</th>
<th>JM109(DE3)</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>- IPTG</td>
<td>+ IPTG</td>
</tr>
<tr>
<td>M13SPS1</td>
<td>1.5x10^13</td>
<td>1.4x10^13</td>
<td>1.5x10^13</td>
</tr>
<tr>
<td>M13-100 A</td>
<td>2x10^13</td>
<td>2x10^6</td>
<td>5x10^6</td>
</tr>
<tr>
<td>M13-100 B</td>
<td>7x10^12</td>
<td>1x10^6</td>
<td>1.6x10^6</td>
</tr>
</tbody>
</table>

*All pfu shown are averages of values obtained from three 10-fold serial dilutions of high titer supernatant (hts) phage stocks.

A,B: two different hts stocks of M13-100 (-), (+) = minus, plus IPTG

### Table 2. Comparison of cloning efficiencies in M13-100 and M13mp19 shotgun libraries (pfu per 50 ng vector DNA)

**A. M13-100 Libraries**

<table>
<thead>
<tr>
<th></th>
<th>Expt</th>
<th>JM109</th>
<th>JM109 (DE3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13-100 uncut</td>
<td>6,000,000</td>
<td>1,100</td>
<td></td>
</tr>
<tr>
<td>M13-100 + Eco RV</td>
<td>1) 1,200</td>
<td>0 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2) 1,600</td>
<td>0 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3) 1,100</td>
<td>0 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4) 1,800</td>
<td>0 ***</td>
<td></td>
</tr>
<tr>
<td>M13-100 + Eco RV + Ligase</td>
<td>36,000</td>
<td>20 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2) 80,000</td>
<td>50***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3) 50,500</td>
<td>30 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4) 47,800</td>
<td>0 ***</td>
<td></td>
</tr>
<tr>
<td>M13-100 + Eco RV + Ligase + Insert</td>
<td>1) LAMBDA</td>
<td>18,000</td>
<td>6,000</td>
</tr>
<tr>
<td></td>
<td>1) GABA</td>
<td>28,000</td>
<td>5,700</td>
</tr>
<tr>
<td></td>
<td>2) GABA</td>
<td>64,700</td>
<td>12,800</td>
</tr>
<tr>
<td></td>
<td>3) GABA</td>
<td>64,700</td>
<td>13,400</td>
</tr>
<tr>
<td></td>
<td>4) GABA</td>
<td>50,500</td>
<td>27,200</td>
</tr>
</tbody>
</table>

**B. M13mp19 Libraries**

<table>
<thead>
<tr>
<th></th>
<th>(JM100)</th>
<th>Blue</th>
<th>Clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13mp19 uncut</td>
<td>5,500,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP</td>
<td>650</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP</td>
<td>125</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase</td>
<td>5,200</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase</td>
<td>75</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase + GABA(1)</td>
<td>6,800</td>
<td>4,050</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase + GABA(1)</td>
<td>275</td>
<td>2,575</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase + LAMBDA</td>
<td>7,625</td>
<td>5,750</td>
<td></td>
</tr>
<tr>
<td>M13mp19 + Hinc II - CIAP + Ligase + LAMBDA</td>
<td>825</td>
<td>4,625</td>
<td></td>
</tr>
</tbody>
</table>

*/***/100×25×10× extrapolations of actual plaque counts obtained (0.5/2/0.5 ng vector DNA equivalents were plated)

CIAP = calf intestine alkaline phosphatase

GABA, LAMBDA = random Cvi11 target inserts (GABA = Tri21/ws plasmid)
were constructed irrespective of the difference in the two target gene libraries. Therefore, representative M13-100 and M13mpl9 gene libraries were obtained from M13-100 (experiment 1) and M13mpl9 (+CLAP): GABA—2575 clear; lambda^J625 clear).

Each library yielded enough recombinants to obtain a significantly high R/B ratio (eg. 255:1, 446:1, and 544:0 for libraries 2, 3, and 4). The R/B ratios for the M13-100 GABA and lambda libraries were 285:1 (5700/20) and 300:1 (6000/20), respectively. It is important that the background should be considered 'apparent background' since it is determined on plates separate from those containing the recombinants where the 'true' background cannot be directly estimated. Given the 10,000 fold repression of M13-100 phage growth on the selection host, JM109(DE3), it is expected that growth of the background clones derived from intact religated vector should be selected against. However, sequencing a large number of clones will be needed to verify this assumption. False positives, on the other hand, would survive selection and may represent a significant fraction of the background plaques found on the religated vector control plates. To determine if, in fact, any of the M13-100 background clones were truly false positives arising from alteration of the vector DNA, ten clones were sequenced using the T7L oligonucleotide as primer (highlighted in bold in Fig. 2).

In this sampling, one appeared to be a vector dimer, three were found to be intact religated vector, and six clones contained deletions (up to 150 bp) extending in either direction from the cloning site (data not shown). These deletions were presumably created as a result of exonuclease activity contaminating the EcoRV, T4 DNA ligase, and/or CviJI enzyme. After passaging the M13-100 phage several times, no evidence of sequence instability within the extra gene X cistron has been found which could account for the deletions observed.

A useful direct selection vector is one with which high R/B ratios can consistently be achieved, and without the need for the variable phasate treatment step. A good statistical R/B average for a number of libraries should range between 100:1 and 1000:1 (eg. 0.1% and 1.0% background). For any given library, variability within this range will be caused by differences associated with the quality of the DNA modifying enzymes used during library constructions. To gain preliminary insight into a value for the average R/B ratio for the M13-100 system, cloning efficiencies were obtained from three additional independently constructed M13-100 libraries using the Tri/21ws plasmid (GABA) as target DNA. The fragmentation and ligation conditions employed were the same as those used for the library construction experiments described for experiment 1 of Table 2. The cloning efficiencies obtained per 50 ng vector DNA for these three GABA libraries are also indicated in Table 2 (experiments 2, 3, and 4). The R/B ratio of plaques obtained on JM109(DE3) were 255:1, 446:1, and 544:0 for libraries 2, 3, and 4, respectively. Including the GABA and lambda libraries of expt. 1, an average R/B ratio of 542:1 was calculated from the five trials.

Because M13-100 and M13mpl9 utilize different cloning sites, caution must be taken in comparing their R/B ratios due to potential differences in contaminating nuclease activities between the EcoRV and HincII/phosphatase preparations, respectively.
Nevertheless we reasoned that a comparative study would be informative when maintaining constant all of the other variables including: fragmentation, ligated vector and target DNA concentrations, and transfections. M13mp19 libraries can be analyzed with or without the color discrimination provided by the B-galactosidase marker inactivation assay. To be fair to a direct comparison with the 'apparent' M13-100 background, an analysis without color selection is appropriate since M13-100 does not provide this capability to distinguish between parentals and false positives within the background population. Without the addition of X-GAL and IPTG to the media, the analogous M13mp19 background due to intact religated vector can be considered 85 plaques (75 blue + 10 clear) after CIAP treatment in this experiment. In this case, the R/B ratios can be considered 33:1 (2850/2575 blue plaques) and 64:1 (5450/825 blue)/85 for the GABA and lambda libraries, respectively. These ratios are respectable in terms of approaching the M13-100 direct selection values, but still relied on the CIAP treatment to achieve them. Clearly when CIAP treatment is employed, the (inducible) color selection scheme is not necessarily a requirement for the generation of M13mp type libraries, and the use of IPTG and X-GAL can be avoided. Instead, β-galactosidase interruption merely serves as a quality control measure to estimate the level of background clones devoid of target inserts. In this particular experiment the CIAP treatment was not optimal, resulting in 1.4% religated vector (75/5200 blue plaques) in the absence of insert DNA and about 11% (825/7625) in the presence of insert DNA. The augmentation of the blue plaque background in the presence of target DNA fragments is probably related to the fact that elevated DNA concentrations in ligation reactions will increase blunt-end ligation efficiency (26). The 'true' (vs. apparent) M13mp19 R/B ratios can then be considered 9:1 (2575/2575 + 10) and 6:1 (4625/825 + 10) for the GABA and lambda libraries, respectively. An analogous rise in parental pfu would not be expected to be detectable in a M13-100 library since they should be repressed in the selection host. However, much lower M13mp19 background levels comparable to those obtained by M13-100 direct selection should be achievable with a more efficient CIAP treatment. It can be seen that without phosphatase treatment, the percentage of clones in a M13mp19 shotgun library which arise due to religated vector can be very high, or in this experiment about 60% of the total plaques obtained (GABA: 6800 blue/6800 blue + 4050 clear; lambda: 7625/7625 + 5750). It is generally recognized that a wide degree of variability in blue plaque levels can be experienced among different M13mp19 library constructions, with or without the phosphatase treatment.

In the M13mp19 libraries, 'false positives' are indicated by clear plaques arising from the religated vector control transfection. As a fraction of the GABA and lambda recombinants obtained, these were approximately 0.4% (10/2575) and 0.2% (10/4625) after CIAP treatment, respectively. These false positives were presumably the result of point or deletion mutations that altered the Lac Z gene reading frame. These are usually produced by nuclease impurities present in the phosphatase, ligase, or more commonly, restriction enzyme preparations (eg. HindII). Alternatively, false positives can arise due to the insertion of contaminating bacterial chromosome or other episomal fragments co-purifying with the vector RF DNA, but this can be greatly minimized by taking the extra precaution of gel purifying the DNA after restriction digestion. False positives derived from EcoRV linearized M13-100 cannot be visualized without sequencing, since no marker exists for plaques arising due to religation of the vector. Therefore, the percentage of false positives cannot be estimated for M13-100 libraries as in the M13mp19 system, but its upper limit is assumed to be derived from the number of pfu on JM109(DE3) cells arising from the religated vector control reactions. Based on the average R/B ratio of about 500:1 described above, the false positive element could therefore average up to 0.2%, a percentage comparable to that obtained above for cloning into Hinc II digested M13mp19 after phosphatase treatment.

The results presented have only addressed the performance of M13-100 as a shotgun cloning vehicle in the presence of basal expression of T7 RNA polymerase in the selection host, JM109(DE3). However, it was shown in Table 1 that up to a 500 fold augmented level of repression due to pX over-expression could be elicited by the addition of IPTG since the polymerase is under inducible lac control in these cells. Assuming that IPTG has no influence on the growth of library recombinants, the inducer's overall effect should therefore be to lower the background of clones arising from religated parental vector, thereby increasing the R/B ratio. However, preliminary experiments demonstrated that both recombinant and background levels decreased with increasing concentrations of IPTG present in the growth media (data not shown). Because we were more interested in developing a cloning system which did not require the utilization of expensive reagents such as IPTG, we have not yet studied this effect thoroughly. We speculate that this effect on recombinants in the induced state is the result of increased phage replication repression through augmented expression of the gene X protein.

Sizing and sequencing analysis of M13-100 shotgun inserts
It was necessary to establish if an M13-100 library could yield templates with random insert sizes useful for DNA sequence analysis consistent with current sequencing paradigms (350–500 bp per clone). Therefore twelve individual plaques were picked from library #4 of Table 2 and used to inoculate 1.5 ml JM109 cultures from which double- and single-stranded DNAs were prepared for restriction analysis and DNA sequencing, respectively. To estimate insert sizes, the RF DNAs were double digested with BamHI and EcoRI and then electrophoresed on an agarose gel. The recognition sites for these restriction enzymes flank the extra gene X cistron and its inserts, as depicted in Figure 2. The restriction analysis is shown in Figure 3 and indicates that a random distribution of insert sizes is obtained. After subtracting the size of the extra gene X cistron (450 bp), these insert sizes were estimated to range from about 200 bp (lane 7) to 3500 bp (lane 3), with the majority falling between 500 bp and 2000 bp in length. With regard to the stability of inserts in M13-100, the same twelve clones in addition to 38 other recombinants were passaged 2–3 times in liquid culture to high titers, and their EcoRI/BamHI restriction profiles were shown to remain unaltered (data not shown).

Next, the M13-100 vector was tested as a template for enzymatic sequencing using single stranded DNA prepared from the same set of clones. High yields of plus strand DNA were obtained from the 1.5 ml cultures, ranging from 6.5 μg to 13 μg of ssDNA. It is interesting to note that because M13-100 has two copies of gene X in opposite orientations, three secondary structural forms of the ssDNA can be seen on a 1% agarose gel (data not shown) which we believe are: A) the major relaxed, open circular form, B) an 'hourglass' structure resulting from...
intramolecular annealing of the two gene X sequences, and C) a dimer form resulting from intermolecular annealing of the gene X sequences. Forms B and C can be melted out at 90°C such that mostly form A prevails. However, these alternate forms do not interfere with dideoxy sequencing reactions performed at 70°C with the thermostable Bst DNA polymerase. This enzyme was used to extend the 32\(^\text{P}\)-labelled 3'-35SP\(^\text{P}\) primer (Fig. 2) which anneals to the T7 region of the ssDNA templates. On a sequencing gel, the first base of each insert was demonstrated to be a cytosine as expected for CvUI generated fragments (data not shown). The sequencing ladders extended very long distances from the primer, and the sequences of the ten isolates were non-overlapping as expected from this small sampling size.

Interestingly, the secondary structural forms of M13-100 ssDNA could markedly affect reactions performed with T7 DNA polymerase. The effect is manifested by the inability to determine vector sequence beyond the 3' cloned insert junction in the region where the two gene X sequences reanneal (data not shown). This effect is particularly pronounced in sequencing reactions performed on DNA templates prepared from M13-100 shotgun clones containing smaller inserts (e.g. 200–500 bp). The symptom observed on an ABI 373A automated sequencer using fluorescent dye-conjugated dideoxy terminators (27) is a sudden drop in peak intensity near the 3' cloning junction followed by inaccurate base-calling in the gene X region of the vector sequence. Importantly, the ability to accurately read the entirety of the insert sequences remains unaffected. This effect is probably attributed to predominance of the hourglass form created during the stringent primer pre-annealing step utilized for the T7 DNA polymerase reactions. A potential benefit of this result is that with visual or computer-aided inspection of the analyzed sequence data, the 3' insert-vector junction can rapidly be delineated. Perhaps this will have utility in shotgun sequencing projects as a way of discarding unwanted vector sequence during fragment assembly.

**DISCUSSION**

The direct selection capability of M13-100 is enabled by insertional inactivation of M13 gene X, a repressor of phage-specific replication when overproduced in E.coli. High level expression was accomplished by inserting an extra copy of gene X into the IG region and placing it under phage T7 gene 10 promoter/RBS control. The gene X selection function is only activated in cells expressing T7 RNA polymerase, such as JM109(DE3) cells. To our knowledge, this is the only DE3-lysogenized F' strain currently available commercially, but any male strain permissive for F1-specific phage growth could easily be lysogenized (e.g. a lambda DE3 lysogenization kit is available from Novagen, Madison, WI.). Our results indicate that although JM109(DE3) is an F' strain, the strong lacUV5 promoter still permits leaky expression of T7 RNA polymerase in these cells since a 10,000 fold gene X-mediated inhibition of M13-100 growth can be elicited without IPTG induction (Table 1). Whether this effect can be reproduced by lysogenizing a different strain of E.coli with DE3 or the like remains to be determined.

In order to maximize the ratio of recombinants (R) to background (B) clones, shotgun library constructions using the widely employed mp type vectors require either phosphatase treatment of the vector DNA or ligation of sticky-ended adaptors to target fragments for forced orientation cloning. When efficient, these steps do not require the blue/clear color discrimination assay. However, high efficiency cannot be routinely guaranteed, and therefore the \(\beta\)-galactosidase inactivation assay is often invoked as a convenient method to estimate the fraction of the two major subgroups of clones within the background population, that is, those arising due to religated vector and those representing false positives. Another disadvantage of this assay is the high expense associated with the inducer, IPTG, and indicator, X-GAL. The M13-100 direct selection strategy is expected to facilitate a more efficient front-end system by removing these uneconomical reagents and the unreliability associated with attempting to consistently obtain high R/B ratios.

The R/B ratio was considered of paramount importance in evaluating direct selection. Indeed, high R/B ratios were consistently achieved in five independent M13-100 shotgun libraries (Table 2), averaging 542:1 for the five constructions tested. By comparison, the R/B ratios achieved in two M13mp19 (+CIAP) libraries plated from ligations sharing the same fragment preparations were significantly lower, averaging about 50:1. The difference in these ratios reflects the nonoptimal phosphatase treatment of the M13mp19 DNA in this experiment. By contrast, cloning in M13-100 has so far consistently averaged up to 10-fold lower backgrounds without the added phosphatase treatment. However, either vector system should be equally vulnerable to other variables known to influence cloning efficiency, including (a) target DNA complexity, (b) random fragment size distribution, (c) quality of DNA modifying enzymes, (d) ligation efficiency, and (e) transfection efficiency.

The cloning efficiencies of the M13-100 and M13mp19 systems were shown to be similar and capable of generating essentially comparable representative gene libraries (Table 2). This suggested that the M13-100 cloning strategy did not strongly favor the generation of false negatives. For any vector system, their contribution cannot be accurately determined without completely sequencing the target DNA. False negative production could be favored by (a) overexpression of the selection gene, (b) a metabolic overload that slows the rate of cell division and/or phage replication, and/or (c) the accumulation of toxic or unstable
cloned sequences. Clearly, the degree of false negative accumulation will affect the levels of random coverage of a particular target DNA. Furthermore, the method of DNA fragmentation can also influence the degree of randomness of fragment distribution within a library. For this study, the target DNAs were fragmented with the two cutters, CviHI. The use of this enzyme has several advantages over the more commonly employed methods of fragmentation (15), but its influence on the generation of nonrandomness and false negatives remains to be more rigorously analyzed.

False positives, on the other hand, often arise by inactivation of the marker or selection gene due to a nuclease component present in the DNA modifying enzymes. For instance, deletions disrupting the lacZ reading frame in alpha-complementation vectors have been reported as a problem for a number of restriction endonucleases (28). We found no significant difference in the level of false positives between M13-100/EcoRV and M13mp19/HincII libraries (< 1%). On the other hand, M13mp19/SmaI exhibited up to 12% non-recombinant clear plaques in M13 libraries that we have constructed. Fortunately, a number of restriction enzymes are now commercially available (Promega) in which some of these troublesome nuclease components has been removed.

Other filamentous phage vectors have been described (29, and see 30,31,32 for reviews). In the f1 vector series (30), the cloning site is at position 5614, between the packaging signal and the origin of (−) strand synthesis (located only 10 bp downstream of the of the insertion site for the extra gene X cistron in M13-100). In these vectors no color discrimination or positive selection is provided. However, an advantage is that some larger inserts in this region that were unstable further downstream in the mp series were reportedly more stable in this region of f1. After analyzing the inserts of 50 M13-100 shotgun clones, our results indicate a similar trend, although insert stability as a function of size and sequence awaits an exhaustive analysis.

Another positive selection system for shotgun cloning in M13 has been described (29) which relies on the EcoK restriction/modification system. These authors describe a M13 vector which utilized insertional inactivation of an EcoK cassette used for blunt-end cloning. Recombinants can be directly selected by growth on a ++ strain of E. coli. However, only a 20 fold selection against religated vector could be accomplished when shotgun cloning in this vector. By contrast, a 10,000 fold selection against religated M13-100 vector in JM109(DE3) cells is observed. This corresponds with the survival rates also observed by infection (Table 1), transfection with untreated RF DNA (Table 2), and transfection with plus strand DNA (data not shown). This strong inhibition of phage growth no matter what form the infectious agent takes is consistent with Fulford and Model’s (1988) observation that unphysiologically high levels of X protein blocks all phage-specific DNA synthesis at the earliest stages of replication. Given also that filamentous phage maturation and secretion occurs within only 20 minutes post-infection, it seems unlikely that a cellular 'lethality' function in lieu of gene X could as efficiently block or slow down the M13 phage replication cycle.

Further development of a more generalized vector containing a multiple cloning region and reverse primer is in progress for purposes of forced orientation cloning and bidirectional sequencing. However, until an appropriately modified M13-100 vector is available, the prototype described here should have immediate application to both small and large scale DNA sequencing projects utilizing the shotgun cloning strategy.

ACKNOWLEDGEMENTS

We gratefully acknowledge Jim Uzgiris and Dr Dave Mead for providing the M13SPS1 parent vector, and Dr Richard French-Constant for providing the purified Tri/21ws plasmid. We thank Dr James Van Etten for his generosity in furnishing the CviI restriction enzyme. We also thank Karen Whelpley for her excellent technical assistance. This work was supported by DOE Human Genome Program grant # DE-FG02-90ER61026.

REFERENCES