JOURNAL OF BACTERIOLOGY, Oct. 1997, p. 6228–6237 Vol. 179, No. 20
0021-9193/97/$04.00+0
Copyright © 1997, American Society for Microbiology

Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type *Escherichia coli*: Application to Open Reading Frame Characterization

ANDREW J. LINK,[1]† DERETH PHILLIPS,[1] AND GEORGE M. CHURCH[1,2]*

Department of Genetics[1] and Howard Hughes Medical Institute,[2] Harvard Medical School, Boston, Massachusetts 02115

Received 6 March 1997/Accepted 8 August 1997

We have developed a new system of chromosomal mutagenesis in order to study the functions of uncharacterized open reading frames (ORFs) in wild-type *Escherichia coli*. Because of the operon structure of this organism, traditional methods such as insertional mutagenesis run the risk of introducing polar effects on downstream genes or creating secondary mutations elsewhere in the genome. Our system uses crossover PCR to create in-frame, tagged deletions in chromosomal DNA. These deletions are placed in the *E. coli* chromosome by using plasmid pKO3, a gene replacement vector that contains a temperature-sensitive origin of replication and markers for positive and negative selection for chromosomal integration and excision. Using kanamycin resistance (Kn^r) insertional alleles of the essential genes *pepM* and *rpsB* cloned into the replacement vector, we calibrated the system for the expected results when essential genes are deleted. Two poorly understood genes, *hdeA* and *yjbJ*, encoding highly abundant proteins were selected as targets for this approach. When the system was used to replace chromosomal *hdeA* with insertional alleles, we observed vastly different results that were dependent on the exact nature of the insertions. When a Kn^r gene was inserted into *hdeA* at two different locations and orientations, both essential and nonessential phenotypes were seen. Using PCR-generated deletions, we were able to make in-frame deletion strains of both *hdeA* and *yjbJ*. The two genes proved to be nonessential in both rich and glucose-minimal media. In competition experiments using isogenic strains, the strain with the insertional allele of *yjbJ* showed growth rates different from those of the strain with the deletion allele of *yjbJ*. These results illustrate that in-frame, unmarked deletions are among the most reliable types of mutations available for wild-type *E. coli*. Because these strains are isogenic with the exception of their deleted ORFs, they may be used in competition with one another to reveal phenotypes not apparent when cultured singly.

With the completion of the *Escherichia coli* K-12 genome sequence (http://www.genetics.wisc.edu/ and http://mol.genes.nig.ac.jp/ecoli/), a variety of tools will be required to determine the functions of the vast array of uncharacterized open reading frames (ORFs) found within the genome. Even in an organism as well studied as *E. coli*, over 58% of the putative coding regions remain without a recognized function, and many others are only partially understood. To study these regions, we devised a system for creating in-frame deletions of any desired sequence in wild-type *E. coli*.

Gene replacements in *E. coli* have generally relied on specific genetic backgrounds as starting strains, such as *polA*, *recD*, *strR*, *sup*$^+$, or F′ (15, 21, 37, 39, 44). After replacement of a wild-type sequence with an in vitro-altered sequence in a mutant background, the altered chromosomal region must then be transduced into a wild-type genetic background. Unfortunately, these methods often require the transduction of a marker along with the mutant allele. This marker can obscure the phenotype of the mutant allele because it may itself cause a mutant phenotype.

Bacterial genes needed in a particular pathway tend to be grouped in cotranscribed clusters or operons (32). Insertional, frameshift, nonsense, or antisense disruption of an ORF within an operon can affect upstream and downstream gene expression in addition to the gene targeted for inactivation. These polar effects could confuse the assignment of a mutant phenotype to the disrupted gene. At the other extreme, point mutants can leave significant parts of the gene intact. To reduce these problems, we developed methods for creating precisely engineered deletions of *E. coli* ORFs by using a procedure known as crossover PCR (18, 19). To integrate these PCR-generated deletions into the genome of wild-type *E. coli*, we constructed a new gene replacement vector, pKO3.

Hamilton et al. have described a method for gene replacement in wild-type *E. coli* that uses homologous recombination between the bacterial chromosome and a plasmid carrying cloned chromosomal sequences whose replication ability is temperature sensitive (16). At the nonpermissive temperature, cells maintain drug resistance only if the plasmid integrates into the chromosome by homologous recombination between the cloned fragment and the bacterial chromosome. Excision of the integrated plasmid is allowed at the permissive temperature. Depending on the position of the second recombination event that excises the plasmid, the chromosome retains either the wild-type sequence or the altered sequence from the plasmid. Although this method can be applied to wild-type strains, there is no selection for loss of the excised plasmid. The *Bacillus subtilis* gene *sacB* encodes levansucrase, an enzyme that catalyzes the hydrolysis of sucrose and levan elongation (12). When expressed in *E. coli* growing on media supplemented

* Corresponding author. Mailing address: Department of Genetics, Warren Alpert Building, Room 513, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 432-7562. Fax: (617) 432-7266. E-mail: church@salt2.med.harvard.edu.

† Present address: Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195.

with sucrose, the *sacB* gene is lethal (14). Blomfield et al. developed a system for using a temperature-sensitive plasmid and a counterselectable *sacB* marker in the chromosome to facilitate allelic exchange (5). We have reduced this system to one component by incorporating the *sacB* gene into a gene replacement plasmid (pKO3) and have developed a protocol for introducing altered alleles into wild-type *E. coli* strains. By combining the crossover PCR and gene replacement methods, we demonstrate a system for creating precise deletions that eliminate gene function without introducing polar affects on expression of distal genes in an operon.

When making a survey of the most abundant proteins in *E. coli*, we found two poorly understood genes, *yjbJ* and *hdeA*, that encode unexpectedly abundant proteins in the cell (25). The *E. coli yjbJ* gene, with sequence similarity to the uncharacterized ORF *ywmH* in *B. subtilis*, encodes a small 69-amino-acid protein that is highly abundant during early stationary phase in rich media. HdeA is a 121-amino-acid protein with a 23-amino-acid signal peptide whose expression is affected by mutations that eliminate the protein HU-1 (45, 46). The HdeA protein is abundant during growth in minimal media and during stationary phase in rich media (25). To determine if mutant alleles of *yjbJ* and *hdeA* have significant phenotypes, we replaced the chromosomal genes with both insertion and deletion alleles by using the pKO3 gene replacement protocol. In the following text, we will discuss the advantages and disadvantages of both insertional and deletion methods. In light of the completion of the genomic sequences of several free-living organisms, the results of these gene replacements are discussed as paradigms for addressing the function of chromosomal sequences.

MATERIALS AND METHODS

Strains. All plasmid constructions were electroporated and propagated in *E. coli* DH5α [F^- λ^- *endA1 hsdR17 hsdM*$^+$ *supE44 thi1 recA1 gyrA96 relA1* Δ(*argF lacZYA*)U169 φ80d Δ(*lacZ*)M15]. The gene replacement experiments used the recombination proficient wild-type K-12 strain EMG2 (F′ λ^+).

Media and growth conditions. All strains were grown in LB medium (1% Bacto Tryptone, 0.5% yeast extract, 0.5% NaCl) with the appropriate selection. For antibiotic selection, the concentrations of antibiotics were 50 mg/ml (ampicillin and kanamycin) and 20 mg/ml (chloramphenicol). For selection against *sacB*, LB medium was supplemented with sucrose to a final sucrose concentration of 5% (wt/vol).

DNA purification. Plasmid DNA was purified by the alkaline lysis method (4). Genomic DNA was purified by previously described methods (10).

Partial digestion. All partial digestion of genomic and plasmid DNA used serial dilution of the restriction enzyme and a constant 1-h incubation time (26). The reactions were stopped by adding 0.25 M EDTA (pH 8) to a final concentration of 50 mM.

Blunt-end reactions. Unless stated otherwise, T4 DNA polymerase and deoxynucleoside triphosphates (dNTPs) were used to create all blunt-ended DNA fragments.

Ligation. Ligations were performed overnight at room temperature, using a DNA concentration of 10 μg/ml and an insert-to-vector molar ratio of 1:1 or an oligonucleotide-to-insert molar ratio of 160:1. The ligation buffer used for the reactions contained 66 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 50 mM dithiothreitol, 1 mM ATP, and 0.05 Weiss units of T4 DNA ligase/μl. The ligated DNA was ethanol precipitated, washed with 70% ethanol, vacuum dried, and resuspended either in 10 mM Tris-HCl (pH 8)–1 mM EDTA or 40 μl of electroporation-competent *E. coli* cells (for immediate transformation).

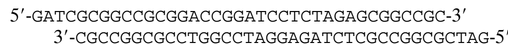
Electroporation. Electroporation-competent cells (40 μl; 10^{11} CFU/ml) were mixed with 1 to 3 μl of DNA solution in an ice-cold microcentrifuge tube and transferred to a 0.2-cm electroporation cuvette (Bio-Rad, Inc.). The cells were electroporated at 2.5 kV with 25 μF and resistance of 200-ohms. Immediately after electroporation, 1 ml of SOC medium (2% Bacto Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added to the cuvette. The cells were transferred to a 17- by 100-mm polypropylene tube and allowed to recover for 1 h at either 30°C (for temperature-sensitive plasmids) or 37°C with shaking at 250 rpm before plating on selective media.

PCR. All PCRs were performed in a Perkin-Elmer 9600 thermal cycler. PCR buffer (28) consisted of 30 mM tricine (pH 8.4), 2 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.01% (wt/vol) gelatin, 0.1% (wt/vol) Thesit, 200 μM each dNTP, 600 μM each primer, and 1 U of *Taq* polymerase (Boehringer Mannheim, Inc.). After addition of template DNA, the PCR mixture was denatured at 94°C for 3 min before addition of the *Taq* polymerase. The thermal cycle profile was 15 s at 94°C, 15 s at 55°C, and 30 s at 72°C. All experiments used 30 cycles and a final 5-min 72°C hold step.

Analysis of PCR products. PCR products were analyzed on 2% high-strength agarose–1% NuSieve agarose gels (FMC, Inc.) or 1% high-strength agarose gels cast in 0.5× Tris-borate-EDTA with ethidium bromide.

pKO3 plasmid construction. The gene replacement vector pKO3 was constructed as follows. First, the 1.6-kb *Eco*NI fragment from pMAK700 (16) containing the temperature-sensitive pSC101 replication origin and the 1.6-kb *Bbv*II-*Bsu*36I fragment of pMAK700 containing the *cat* gene were blunt ended and ligated together to create pKO1.

Second, the 1.35-kb *Not*I-*Nru*I fragment from pBS-TS (2a) containing the *sacB* gene and the 5.6-kb *Sph*I-linearized pMAK705 plasmid (16) were blunt ended and ligated to create pMAK705s. The following *Not*I polylinker was then ligated into the *Bam*HI site of pMAK705s to create pMAK705so:



The 550-bp *Bgl*I-*Bsm*AI fragment from pBluescript II SK− (Stratagene, Inc.) containing the M13 origin of replication and *Hin*dIII-linearized plasmid pMAK705so were blunt ended and ligated to create pMAK705som. The single *Pst*I site in plasmid pMAK705som was deleted by using T4 DNA polymerase and dNTPs to create pMAK705somp.

Finally, the 2.4-kb *Ecl*136II-*Eco*RV fragment from pMAK705somp containing the polylinker, M13 origin of replication, and *sacB* gene was blunt ended and ligated to *Age*I-linearized, blunt-ended plasmid pKO1 to create pKO3.

Crossover PCR deletions and subcloning. Crossover PCR deletion products were constructed in two steps, as illustrated in Fig. 4. In the first step, two different 25-μl asymmetric PCRs were used to generate fragments to the left and right of the sequences targeted for deletion. The PCR conditions were as described above except that the primer pairs were in a 10:1 molar ratio (600 μM outer primer and 60 μM inner primer). In the second step, the left and right fragments were annealed at their overlapping region and amplified by PCR as a single fragment, using the outer primers. Specifically, 1 μl of each of the two asymmetric PCR mixtures and 600 μM each of the two outside primers were mixed together and PCR amplified. The fusion products were phenol-chloroform extracted, ethanol precipitated, washed with 70% ethanol, vacuum dried, resuspended in 50 μl of 1× *Bam*HI restriction buffer containing 40 U of *Bam*HI restriction enzyme, and digested overnight at 37°C. The fusion products were gel purified, ligated into *Bam*HI-digested and phosphatase-treated pKO3 vector, electroporated into *E. coli*, and plated on chloramphenicol plates at 30°C. The recombinant colonies were screened for inserts with PCR, using primers pKO3-L and pKO3-R (described below).

To construct the 286-bp deletion of *yjbJ* by crossover PCR, the following set of oligonucleotide primers was used: yjbJ-No, 5′-CGCGGATCCTCACCTTTAC CGCCTATGCG-3′; yjbJ-Ni, 5′-CCCATCCACTAAACTTAAACACCGTCA CGTTGCGGCAAACC-3′; yjbJ-Co, 5′-CGCGGATCCTTGCGCCTGATGAG TCTGCAGG-3′; and yjbJ-Ci, 5′-TGTTTAAGTTTAGTGGATGGGGTGGAT TGGGAAACCCGC-3′.

To construct the deletion of *hdeA*, the following set of primers was used: hdeA-No, 5′-CGCGGATCCGAAATTATGACTGCGGTTGC-3′; hdeA-Ni, 5′-CCCATCCACTAAACTTAAACAGCCTAATACTTTTTTCATCG-3′; hdeA-Co, 5′-CGCGGATCCTACTCCTTTTTACTTGCACC-3′; and hdeA-Ci, 5′-TG TTTAAGTTTAGTGGATGGGAAAGGCGAATGGGACAAAAT-3′.

DNA sequencing. DNA sequencing was performed as previously described with the Stratagene Cyclist sequencing kit H (27). Sequencing products were labeled with [α-^{32}P]dATP and resolved on a 4.5% wedge-gradient sequencing gel. Sequencing primers used for the pKO3 left and right vector-insert junctions were pKO3-L (5′-AGGGCAGGGTCGTTAAATAGC-3′) and pKO3-R (5′-T TAATGCGCCGCTACAGGGCG-3′). Sequencing primers used to prime from multiplex tag 04 (10) were CP-04 (5′-AGTGTGAGGTTTAAATATTG-3′) and CE-04 (5′-TGTTTAAGTTTAGTGGATGG-3′). Sequencing primers used to prime from multiplex tag 01 (10) were CP-01 (5′-TGATTAGTTGTAATGAA AGG-3′) and CE-01 (5′-TAGTATGATTTTATTGGGGG-3′).

Gene replacement. Mutant alleles cloned into the pKO3 gene replacement vector were electroporated into EMG2 and allowed to recover for 1 h at 30°C. The cells were plated on prewarmed chloramphenicol-LB plates and incubated at 43 and 30°C. The integration frequency was calculated as the ratio of colonies at 43°C to colonies at 30°C. From the 43°C plate, one to five colonies were picked into 1 ml of LB broth, serially diluted, and immediately plated at 30°C on either 5% (wt/vol) sucrose or 5% sucrose-kanamycin plates and at 43°C on chloramphenicol plates. The excision frequency is the ratio of 30°C-grown sucrose-resistant colonies to 43°C-grown chloramphenicol-resistant colonies. The 5% sucrose plates were replica plated to chloramphenicol plates at 30°C to test for loss of the replacement vector. The gene replacement was confirmed by PCR using primers flanking the targeted ORF.

Construction of multiplex interposons. To construct the kanamycin resistance (Kn^r) interposon, the 1.3-kb *Dra*III-*Bam*HI fragment from pNK2859 (23) containing the *kan* gene was blunt ended and ligated to the following *Bst*XI linkers:

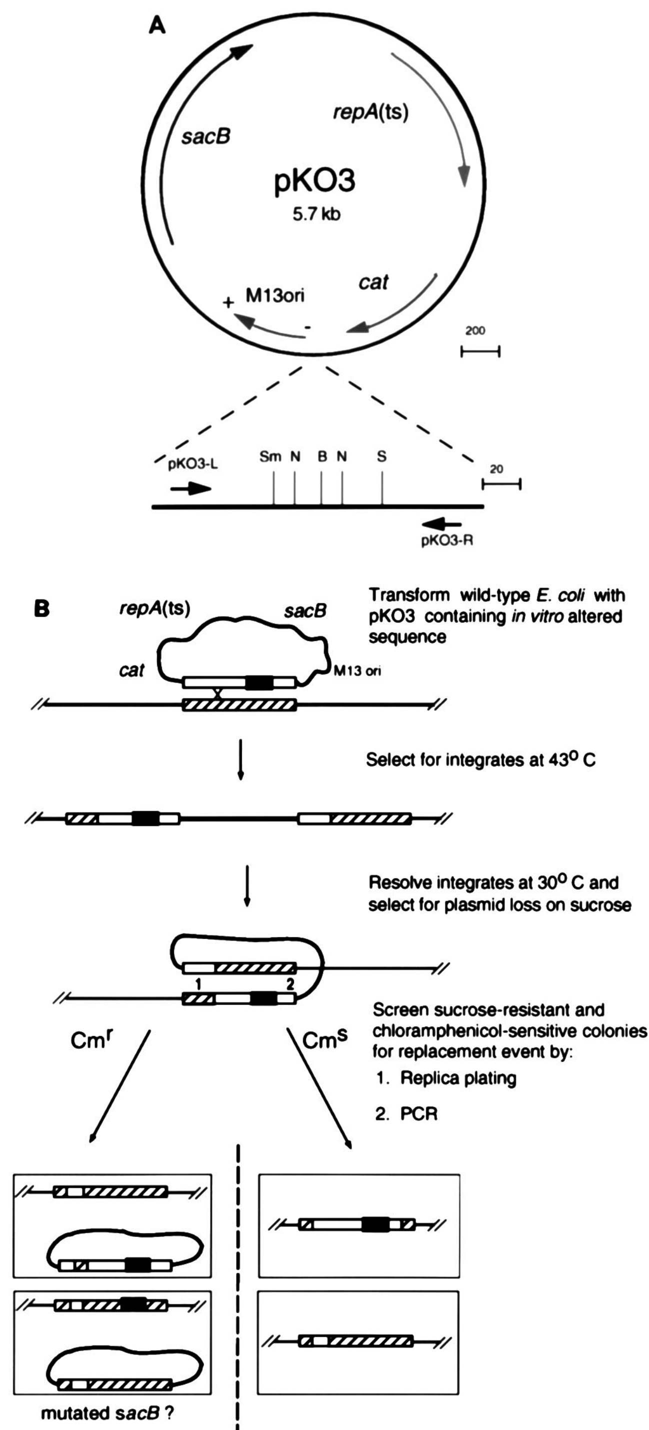


FIG. 1. Gene replacement vector and protocol. (A) The pKO3 vector used in the gene replacement experiments. The cloning region is enlarged. Arrows in the circular plasmid indicate the direction of transcription and the direction of M13 replication. The arrows in the enlarged region are the DNA primer sites. Unique restriction sites are shown (B, *Bam*HI; N, *Not*I; S, *Sal*I; Sm, *Sma*I). ori, origin of replication. (B) Protocol used for replacing wild-type sequences on the chromosome with in vitro-altered sequences. The gene replacement vector carrying in vitro-altered sequences is transformed into *E. coli* and plated at the nonpermissive temperature of the plasmid replicon. An integration event allows replication of plasmid sequences by the chromosomal origin. When shifted to 30°C, the plasmid is excised from the chromosome at either crossover point 1 or 2. The counterselectable *sacB* marker is used to select for loss of plasmid sequences. The sucrose-resistant colonies are screened for loss of vector sequences by replica plating to chloramphenicol plates and then for the gene replacement event by PCR. The "mutated *sacB*?" in the left panels indicates loss of *sacB* gene function by some unknown mechanism. The wavy, thin line represents the gene replacement vector sequences. The straight, thin line represents the *E. coli* chromosome. The boxes represent homologous sequences cloned into the vector (open) and located in the *E. coli* chromosome (striped). The black box within the homologous vector sequence could represent any type of sequence alteration (insertion, deletion, single-base change, etc.).

5′-TCTAGACCACCTGC-3′
3′-AGATCTGGTG-5′

The *kan* fragment with the attached *Bst*XI linkers was ligated to the 2.5-kb *Bst*XI fragment from multiplex vector plex.04B containing multiplex tags 04 (10) to create plasmid pplexkan04B. The Kn^r fragment with the attached linkers was similarly inserted into multiplex vectors 01, 02, 07, 09, 10, 11, 14, 16, 17, 18, and 19 (10) to construct a series of plexkan interposons.

To add a 5′-CG-3′ overhang on the Kn^r interposon, the 1.4-kb *Not*I fragment from plasmid pplexkan.04B was ligated with the following adapters and gel purified:

Dap1 5′-CGCCCCCTGCAGGA-3′
Dap2 3′-GGGGGACGTCCTCCGG-5′

Constructing *pepM*, *rpsB*, *yjbJ*, and *hdeA* insertion mutations. A gel-fractionated genomic library of 3- to 7-kb *Sau*3A inserts prepared from EMG2 genomic DNA was ligated into the phosphatase-treated *Bam*HI site of pKO3. This library was electroporated into *E. coli* DH5α, plated on chloramphenicol plates, and overlaid with nylon membranes to create colony lifts essentially as previously described (26). To identify recombinant plasmids carrying the desired genomic inserts, the colonies were screened by hybridization using oligonucleotides labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase at a probe concentration of 1 nM, overnight at 42°C (10, 26). The probes used were complementary to the 5′ ends of *pepM* (5′-TTCTGTCCATCAGCGTCGGTG-3′), *yjbJ* (5′-GCCGGCTTCATCTTTATTCAT-3′), and *hdeA* (5′-CCACCAAGAAT AACGCCTAAT-3′). The *pepM* oligonucleotide was used to screen for *rpsB* clones simultaneously.

To identify clones with at least 1 kb of genomic DNA flanking each side of the desired genes, positive clones were screened by a combination of restriction mapping and DNA sequencing across the vector-insert junctions. These results were compared to physical maps of the regions. To create lesions in *pepM* and *rpsB*, a positive clone was partially digested with a mixture of five four-base recognition site restriction enzymes that create a 5′-CG-3′ overhang (*Aci*I, *Hpa*II, *Hin*pI, *Mae*II, and *Taq*I). The singly cut, linearized plasmid was gel purified and ligated with the Kn^r interposon plexkan04, using adapters Dap1 and Dap2. For *yjbJ* and *hdeA*, positive clones (p1.7 [*yjbJ*] and p15.3 [*hdeA*]) were linearized by partial digestion with restriction enzymes with unique sites in the ORFs (e.g., *Nae*I for *yjbJ* and *Pst*I or *Pvu*II for *hdeA*). The singly cut, linearized plasmids were gel purified and ligated to a blunt-ended multiplex interposon (e.g., *yjbJ*::plexkan04, *hdeA*::plexkan01 [*Pst*I site], and *hdeA*::plexkan04 [*Pvu*II site]). Before performing the gene replacement, we characterized the *in vitro*-altered insert by DNA sequencing across both the vector-insert junctions (using the primer sites in the vector as primer sites) and the interposon-insert junctions (using the interposon's multiplex tags as primer sites).

Screening for gene replacements. PCR was used to screen for gene replacements of *yjbJ* and *hdeA*. The *yjbJ* gene replacement was confirmed by using the primers yjbJ-Nout and yjbJ-Cout flanking the gene. The *hdeA* gene replacement was confirmed by PCR using primer pair hdeA-Nout1 plus hdeA-Cout1, hdeA-Nout2 plus hdeA-Cout2, or hdeA-Nout3 plus hdeA-Cout3 flanking the gene. Sequences of the primers are as follows: yjbJ-Nout, 5′-AGGTGAAAAAGAA ACCGCGTT-3′; yjbJ-Cout, 5′-TGGTTTGCCGCAACGTGACGG-3′; hdeA-Nout1, 5′-CGCGGATCCCATATACAGAAAACC-3′; hdeA-Cout1, 5′-CGCG GATCCTTTTAAAGAAGATAT-3′; hdeA-Nout2, 5′-CTGATGCATCTGTAA CTCATT-3′; hdeA-Cout2, 5′-AACGCAGATTGTGCGTTCACC-3′; hdeA-Nout3, 5′-GGATGAAGAAATAGCCGATC-3′; and hdeA-Cout3, 5′-CTTCCC ATGCCAATTAATAC-3′.

Competition experiments. Competition experiments were performed by coculturing equal concentrations of two strains in rich media and then sampling the population density of each strain at various time points. Equal optical densities at 600 nm of diluted overnight cultures of the various strains were mixed in the following combinations and sampled at various time points. EMG2 *yjbJ*::plekan04 and EMG2 *hdeA*::plekan01 strains were each cocultured with the wild-type EMG2 strain. In a second competition experiment, EMG2 *yjbJ*::plexkan04 was cocultured with EMG2 Δ*yjbJ*. Each mixed culture was grown aerobically in a 250-ml Erlenmeyer flask containing 50 ml of LB medium at 37°C shaking at 250 rpm (New Brunswick Scientific G2 platform). Since each culture contained both a marked and unmarked strain, the survival ratios could be determined by plating on both LB and kanamycin plates at various time points and counting the colonies surviving on each plate.

RESULTS

Developing an improved gene replacement method. We constructed a gene replacement vector for creating null mutations in the chromosomal sequences of wild-type *E. coli* strains as described in Materials and Methods and illustrated in Fig. 1A. The plasmid is derived from a previously described gene replacement vector and has the *lac* sequence removed to eliminate homologous recombination at the *lac* region in the *E. coli* chromosome (16). The *repA*(Ts) replication origin is derived from pSC101 and has a permissive temperature of 30°C but is inactive at 42 to 44°C. The *cat* gene (encoding chloramphenicol resistance) is used as a marker to select for chromosomal integrates and as a marker for cells harboring vector sequences after plasmid excision. The *sacB* gene is used to counterselect vector sequences by growing cells harboring the plasmid on medium supplemented with 5% sucrose. The M13 replication origin facilitates generation of single-stranded copies of the plasmid by using helper phage (a feature not used in this study). Finally, the primer sites pKO3-L and pKO3-R flanking the cloning site enable screening the vector for inserts by PCR or for DNA sequencing across the vector-insert junctions.

Figure 1B diagrams the protocol that we used to perform gene replacements in *E. coli*. The in vitro-altered sequences carried in the vector pKO3 are transformed into *E. coli*, and the transformed cells are allowed to briefly recover at the permissive temperature. The cells are then plated on chloramphenicol plates at the nonpermissive temperature to select for chromosomal integrates. This was more effective for obtaining the final gene replacement event than plating cells at 30°C and shifting them to 43°C. We found that integrates could also be obtained by serially diluting cells harboring the plasmid at 30°C and plating them at 43°C. To select cells in which the plasmids are excised and lost, we picked and suspended colonies from the 43°C plates, diluted the suspension, and plated the cells on LB plates containing 5% sucrose at 30°C. Only cells that have excised the plasmid sequences and lost *sacB*'s counterselectable function should grow under these conditions. We found this procedure worked better for getting the final gene replacement event than simply replica plating colonies from 43°C to sucrose plates at 30°C. Finally, the sucrose-resistant and chloramphenicol-sensitive colonies are screened for the desired gene replacement event by using PCR and primers to the genomic DNA flanking the altered sequences or by Southern hybridization.

Replacing *yjbJ* with an insertional allele. Suspecting that the null allele of *yjbJ* would be lethal, we decided to disrupt *yjbJ* by inserting a specialized Kn^r selectable marker, or interposon, into the gene (29). A 5.5-kb DNA fragment from a genomic library containing *yjbJ* was cloned into pKO3, and a Kn^r gene (plexkan04) was inserted at the unique *Nae*I site in the gene (see Materials and Methods). Before doing the gene replacement, we sequenced both the vector-insert junctions and the insertion site of the Kn^r gene and showed that the insertional allele had at least 1 kb of chromosomal sequence flanking both sides of the interposon (Fig. 2A). When the *yjbJ* replacement vector was transformed into *E. coli* and plated at 43°C, the integration frequency was 10^{-2} of the plated cells. Several integrates were picked, serially diluted, and plated at 30°C on various selective media to induce the plasmid excision and loss (Fig. 2B). These different master plates were then replica plated to chloramphenicol plates and kanamycin plates to identify colonies that retained the Kn^r gene and not the vector (Fig. 2B). When the integrate cells were plated on kanamycin medium without sucrose at 30°C, most of the sucrose-resistant colonies were still chloramphenicol resistant, indicating that the cells retained the vector sequences (Fig. 2B, row a). When integrate cells were plated on kanamycin–5% sucrose medium, more than 98% of the sucrose-resistant colonies were chloramphenicol sensitive, indicating loss of plasmid sequences and a probable gene replacement event (Fig. 2B, row b). When the integrate cells were plated on rich medium containing 5% sucrose, 48% of the sucrose-resistant colonies were chloramphenicol sensitive and kanamycin resistant, indicating loss of the plasmid and a probable gene replacement event (Fig. 2B, row c). We verified the structure of the initial 43°C integration and the replacement of *yjbJ* with the insertional allele by screening colonies via PCR using primers flanking *yjbJ* (Fig. 2C). These results proved that the pKO3 replacement system worked and showed that *yjbJ* is a nonessential gene under these environmental conditions.

Lethal gene replacement phenotype. To observe the results of the gene replacement protocol when trying to replace an essential *E. coli* gene with an insertional allele, we tested two known essential genes, *pepM* and *rpsB*. The *pepM* (*map*) gene encodes methionine aminopeptidase, and *rpsB* encodes the ribosomal protein S2 (7, 9). Each gene was cloned into the gene replacement vector pKO3, and insertion mutations were constructed by using the Kn^r gene (see Materials and Methods). DNA sequencing and restriction enzyme mapping showed that both inserts had at least 1 kb of genomic DNA flanking each side of the insertion site.

The pKO3 plasmids carrying the insertionally disrupted essential genes were electroporated into *E. coli* and plated at 43°C to select for integration. The integration frequency was approximately 10^{-2} to 10^{-3}, similar to that for *yjbJ*. Ten integrate colonies were picked, suspended in medium, serially diluted, and plated at 30°C on 5% sucrose–kanamycin plates. We found the sucrose resistance frequencies for both the *pepM* and *rpsB* integrates were approximately 10^{-8}, compared to a frequency of 10^{-2} to 10^{-4} for the insertional disrupted nonessential *yjbJ* gene replacement. For both *pepM* and *rpsB*, all of the sucrose-resistant, kanamycin-resistant colonies remained chloramphenicol resistant, indicating that plasmid sequences were still present in the cell. In addition, the colonies had a mucoid phenotype compared to colonies that had lost the plasmid sequences. It is unknown whether *sacB*'s activity had been directly compromised by a mutation in the gene or if a secondary mutation in the genome conferred sucrose resistance. These results showed the phenotype expected when one tries to replace an essential gene with a disrupted allele. Using the pKO3 gene replacement procedure, Brown et al. have shown that an essential gene, *murA*, can be replaced on the *E. coli* chromosome with a deletion allele, as long as the deletion is complemented by another copy of the essential gene (8).

Paradoxical phenotypes of different *hdeA* insertional alleles. Speculating that *hdeA* might be an essential gene, we constructed two different insertional alleles of *hdeA* (Fig. 3). Both were made using the same chromosomal insert cloned into the vector pKO3. In one allele, the inserted Kn^r gene was cloned into the *Pst*I site of the *hdeA* gene; in the second allele, the Kn^r gene was cloned into the *Pvu*II site in the opposite orientation (see Materials and Methods).

In the first step of the gene replacement procedure, the two plasmids transformed and integrated at similar frequencies. However, when resolving the integrates, we found that the *Pvu*II insertional allele had a sucrose resistance frequency of $<10^{-8}$, compared to approximately 10^{-3} for the *Pst*I allele. All of the sucrose-resistant, kanamycin-resistant colonies with the *Pvu*II insertional allele were chloramphenicol resistant, indicating that the plasmid sequences were still present. This finding suggested that the *Pvu*II allele is a lethal mutation. How-

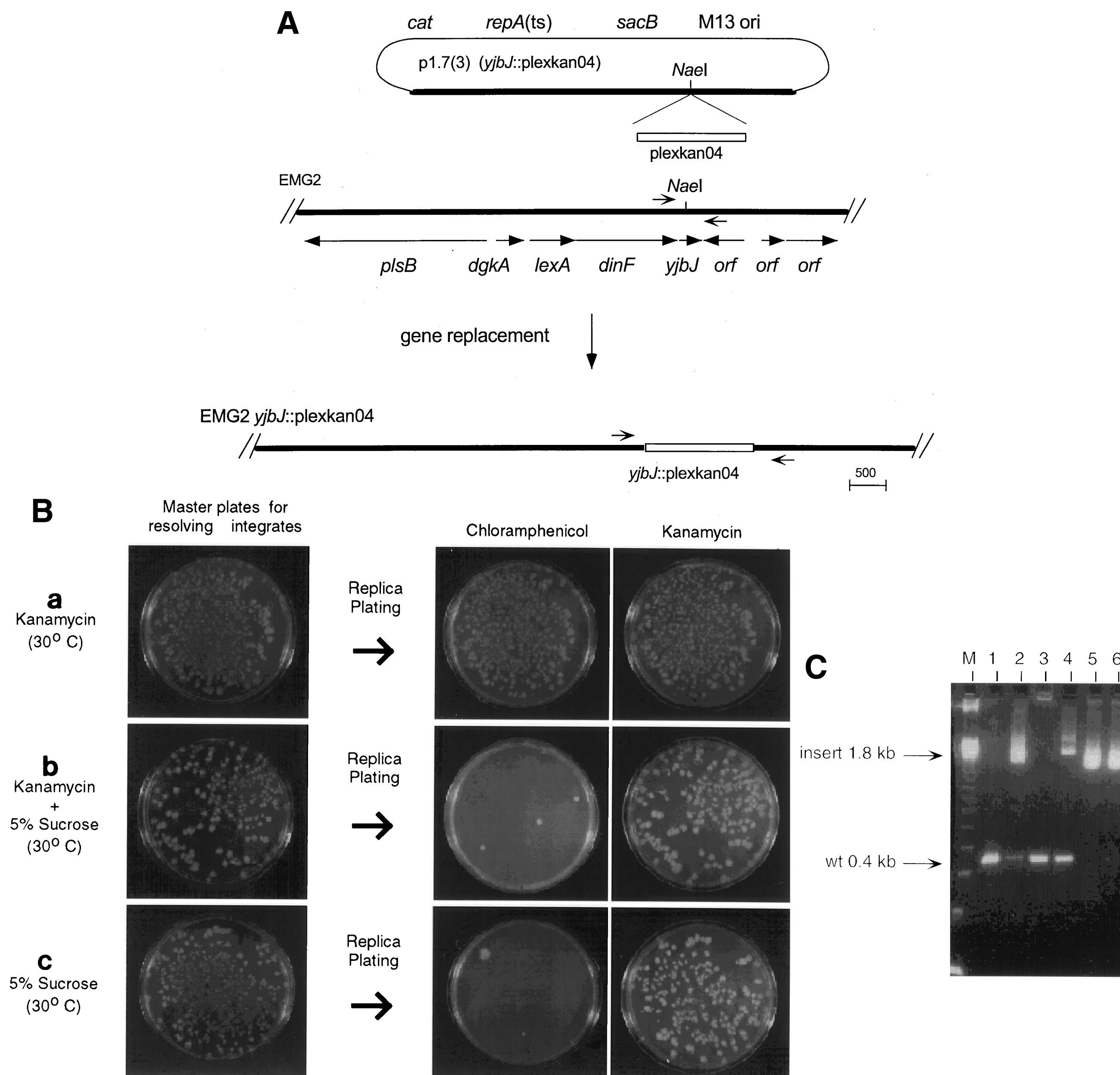


FIG. 2. Replacement of *yjbJ* with an insertional allele. (A) The top panel shows a genomic insert containing *yjbJ* cloned into pKO3 and mutagenized with the plexkan04 interposon. The interposon insertion at the *Nae*I site was confirmed by sequencing across the interposon-genomic insert junction by using primers CE-04 and CP-04, which flank the interposon. The insert was mapped on the *E. coli* chromosome by sequencing across the vector-insert junctions by using primers pKO3-L and pKO3-R. The physical map of the wild-type chromosomal region is shown. The arrows flanking the *Nae*I sites are the PCR primer sites yjbJ-Nout and yjbJ-Cout used to identify the *yjbJ* allele. The expected sizes of the PCR products are 0.4 kb for *yjbJ* and 1.8 kb for *yjbJ*::plexkan04 alleles. The thick lines represent chromosomal fragment sequences. The thin line represents the vector pKO3. The open box represents the interposon. (B) Resolution of *yjbJ*::plexkan04 integrates when plated at 30°C under different selection conditions. Five 43°C integrates were picked, serially diluted, and plated on the master plates shown at the left. These master plates were replica plated first to chloramphenicol plates and then to kanamycin plates at 30°C to detect loss of plasmid sequences (chloramphenicol sensitive) and retention of the interposon (kanamycin resistant). (C) Use of PCR to verify the replacement of *yjbJ* with the insertional allele. The gel shows the products of the PCRs using primers yjbJ-Nout and yjbJ-Cout to detect either the wild-type (wt) or the disrupted *yjbJ* allele. The sources of the template DNA are plasmid p1.7 (*yjbJ*) (lane 1), plasmid p1.7(3) (*yjbJ*::plexkan04) (lane 2), EMG2 genomic DNA (lane 3), genomic DNA from 43°C integrate of p1.7(3) (*yjbJ*::plexkan04) into the EMG2 (lane 4), EMG2 *yjbJ*::plexkan04 genomic DNA from a sucrose-resistant and chloramphenicol-sensitive colony shown in row b (lane 5), and EMG2 *yjbJ*::plexkan04 genomic DNA from a sucrose-resistant and chloramphenicol-sensitive colony shown in row c (lane 6). The size marker is a 123-bp ladder (lane M).

ever, the replacement of the wild-type gene with the *Pst*I insertion was confirmed by PCR screening of colonies with primers flanking *hdeA*. The latter finding suggested that the *Pst*I allele is nonlethal. These *Pst*I and *Pvu*II results together illustrated the difficulty of classifying *hdeA* as either an essential or a nonessential gene because the phenotype varied according to the insertion site of the marker and/or its orientation in the disrupted allele.

Engineering in-frame deletions to minimize polar effects. To avoid problems associated with insertional mutations, we developed a system that replaces ORFs with in-frame deletions. Figure 4 shows how we used crossover PCR to create a dele-

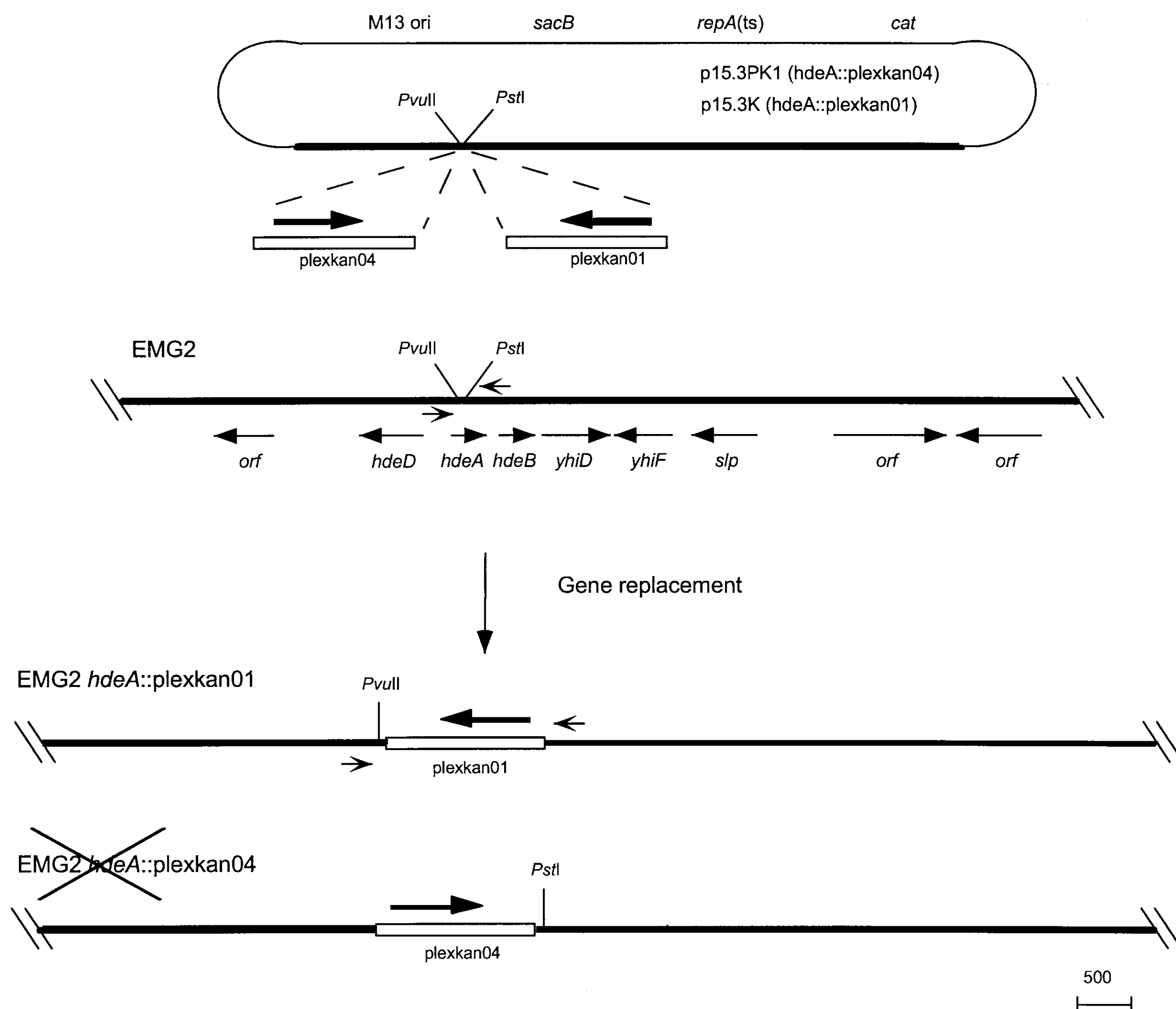


FIG. 3. Paradoxical phenotypes of replacing *hdeA* with insertional alleles. Physical maps of the DNA fragment containing *hdeA* and the chromosomal region before and after gene replacement with the insertional alleles are shown. The unique *Pvu*II and *Pst*I sites in *hdeA* are the insertion sites for the plexkan01 and plexkan04 interposons, respectively. The arrows flanking the *Pst*I and *Pvu*II sites are the PCR primer sites hdeA-Nout1 and hdeA-Cout1 flanking the *hdeA* gene. The construct marked with an "X" could be integrated but not resolve to the replacement allele. Details are as described in the legend to Fig. 2A.

tion of any *E. coli* ORF (18, 19). Complementary oligonucleotide primers and asymmetric PCR are used to generate two DNA fragments having overlapping ends. The two fragments are combined in a fusion reaction in which the overlapping ends annealed and served as primers for 3′ extension of the complementary strand. This fusion molecule is then amplified by PCR using the outer primers.

To construct the deletions, we developed the following rules for designing the oligonucleotides to ensure sufficient homology for recombination during gene replacement and to minimize disruption of flanking sequences. The lengths of the two fragments flanking the deletion are at least 500 bp. The decision to use 500 bp is based on published integration frequencies for various lengths of chromosomal regions cloned into similar gene replacement vectors. The predicted integration frequency should be approximately 10^{-4} (5, 16). The two complementary oligonucleotides (C and B) have at least a 21-base complementary region to allow the products from the asymmetric first and second PCRs to anneal and extend (Fig. 4). The primers were designed so that the deletion maintained the original translational reading frame of the ORF and the added bases provided unique sequences for tracking the deletion in a population of different *E. coli* deletion strains. To minimize potential affects on expression of neighboring genes, we engineered the deletion of the ORF to begin 18 bp downstream of the translation start site and end 36 bp upstream of the stop codon. The oligonucleotides (A and D) have *Bam*HI restriction sites in the 5′ end to allow efficient cloning of the fusion product.

Deletion of *hdeA*. To further investigate the null phenotype of the *hdeA* mutation, we engineered a deletion of the gene (Fig. 5A). Using the above-specified rules, we deleted a 279-bp region, or 84% of the coding region of *hdeA* and replaced it with a 21-bp in-frame sequence tag, using the crossover PCR protocol (see Materials and Methods). Figure 5B shows the two complementary PCR products and the final crossover PCR deletion product. The deletion fragment was cloned into the vector pKO3, and chromosomal deletions were introduced into the chromosome by using our pKO3 gene replacement protocol. The deletion plasmid had an integration frequency of 1.8×10^{-3}. Integrates were serially diluted and plated at 30°C on 5% sucrose plates to select for excision and loss of the

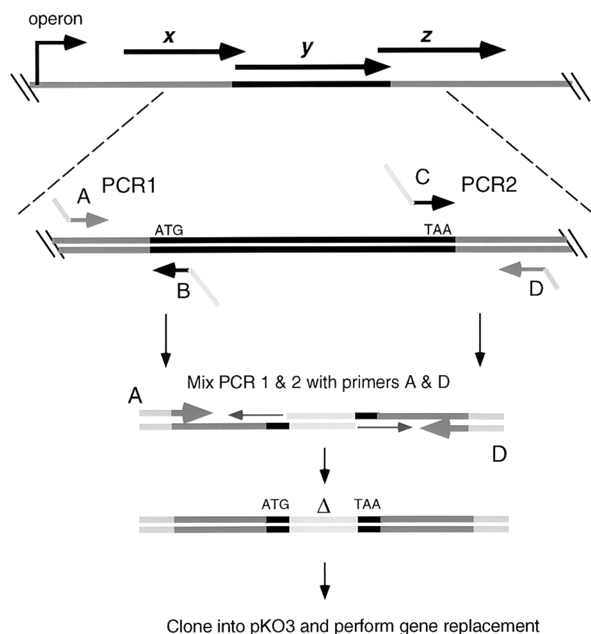


FIG. 4. The creation of in-frame deletion constructs. The top line represents a region of the chromosome where genes *x*, *y*, and *z* form a polycistronic operon. The second line is an expanded view of gene *y* showing the two PCRs used to generate fragments (PCR1 and PCR2) which will form an in-frame deletion of gene *y* when fused. The PCR primers B and C are complementary over 21 nucleotides (represented by the light gray lines) so that when the two PCR products are mixed, the complementary regions anneal and prime at the 3′ overlapping region for a 3′ extension of the complementary strand. In the third line, the fused molecule is amplified by PCR with primers A and D. Primers A and D have *Bam*HI sites incorporated into the 5′ ends of both oligonucleotides (represented by the gray lines) so that the fusion product can be restriction digested and cloned into pKO3.

plasmid sequence. Figure 5C shows the results of screening a fraction of the resolved colonies by PCR with primers flanking *hdeA*. Approximately 7% of the sucrose-resistant and chloramphenicol-sensitive resolved integrate colonies had the deletion replacing the wild-type *hdeA* sequence. Why the resolution frequency for replacing the gene with the deletion was not the expected 50% is unknown. Recovery of the deletion demonstrated that the gene was nonessential under these environmental conditions and suggested that the apparent lethal effect of the Kn^r insertion into *hdeA* was probably due to an effect of the insertion on elements outside of the cloned segment.

Deletion of *yjbJ*. A similar set of experiments was performed to delete 146 bp, or 73%, of *yjbJ*, with the rules previously described. Although the deletion product was successfully amplified, it could not be cloned into pKO3. A PCR assay showed the deletion insert ligated to pKO3, and so we hypothesize that either the protein produced by the deletion mutant was toxic or the insert interfered with plasmid replication in *E. coli*. An analysis of the genomic region identified a potential promoter 107 bp upstream of the *yjbJ* translational start. We designed a second deletion to remove most of the predicted promoter region while leaving the upstream *dinF* gene intact. The second deletion extended from 3 bp downstream of the *dinF* stop codon to 36 bp upstream of the *yjbJ* stop codon, deleting 286 bp of the region, including 82% of coding region of *yjbJ*. This second crossover PCR deletion product was successfully cloned into pKO3. Using the gene replacement protocol, we found that the deletion plasmid had an integration frequency of 7.9×10^{-5}. Similar to *hdeA*, only 3% of the sucrose-resistant

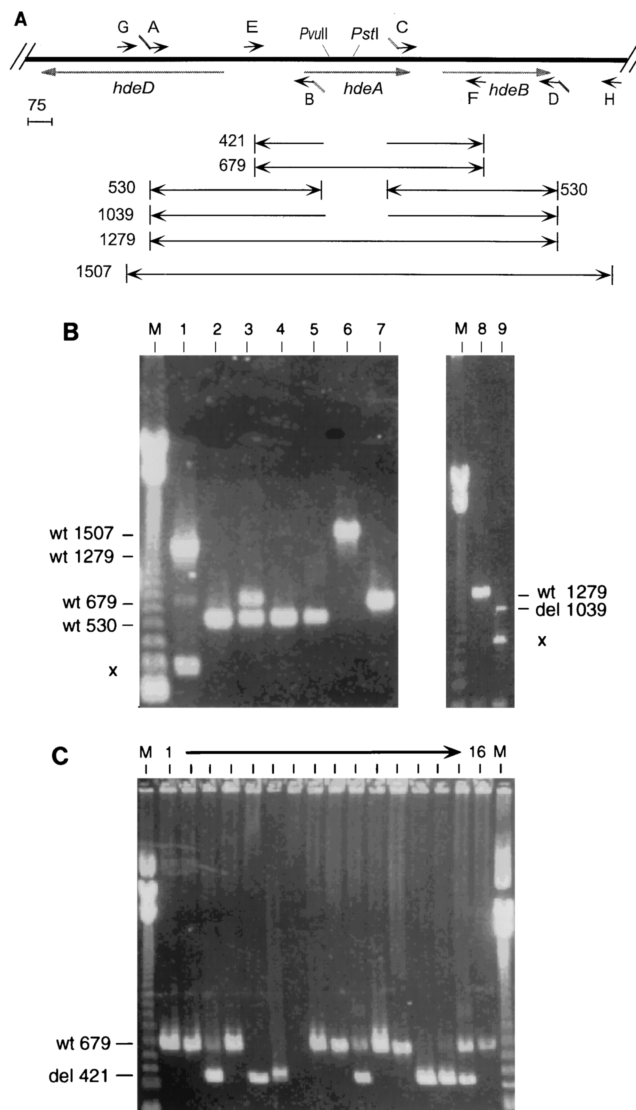


FIG. 5. Constructing and replacing the gene *hdeA* with a precisely engineered deletion. (A) Diagram of the *hdeA* region. The small arrows marked with capital letters are the PCR primer sites used to construct the deletion and to PCR assay either the wild-type or deletion allele of *hdeA*. The predicted sizes of the PCR products are shown below the physical map. Primers: A, hdeA-No; B, hdeA-Ni; C, hdeA-Ci; D, hdeA-Co; E, hdeA-Nout2; F, hdeA-Cout2; G, hdeA-Nout3; H, hdeA-Cout3 (see Materials and Methods). (B) Analysis of PCR products used to construct precise deletion of *hdeA*. The left gel shows the two fragments that will form the deletion product and the PCR products made by using primers flanking *hdeA*. The sizes of the PCR products are shown. EMG2 genomic DNA was used as the template DNA for lanes 1 to 7. The DNA primer pairs used for the PCRs are A-D (lane 1), A-B (1:1) (lane 2), A-B (10:1) (lane 3), D-C (1:1) (lane 4), D-C (10:1) (lane 5), G-H (lane 6), and E-F (lane 7) (see panel A). The right gel shows the amplified deletion (μl) product (lane 9). Using the products of lanes 3 and 5 as templates, the left and right fragments of the deletion were combined, annealed, extended, and PCR amplified by using primers A and D (lane 9). This gel also shows the PCR product made by using the same primer pairs but starting with EMG2 genomic DNA as the template (lane 8). The deletion fusion product was cloned into pKO3 and used in the gene replacement protocol. The "x" indicates an unknown PCR by-product. The size marker is a 123-bp ladder (lane M). wt, wild type. (C) Verification of the replacement of *hdeA* with the crossover PCR deletion product. After integration at 43°C, integrates were plated at 30°C on 5% sucrose plates and replica plated to chloramphenicol plates. The chloramphenicol-sensitive, sucrose-resistant colonies were screened by PCR using primers E and F (see panel A). Those containing the precise deletion give a 421-bp product, while those containing the wild-type allele give a 679-bp product. This gel shows a subset of the colonies screened for the deletion (lanes 1 to 16). The size marker is a 123-bp ladder (lane M).

and chloramphenicol-sensitive resolved-integrate colonies had the deletion replacing the wild-type *yjbJ* sequence. These results prove that the YjbJ protein is nonessential under these environmental conditions and agree with the earlier results obtained by replacing the gene with the Kn^r insertion allele.

Competition experiments to compare insertional and deletion phenotypes. To compare the phenotypes of the various mutant strains, isogenic strains with the insertional *yjbJ* and *hdeA* alleles were compared in a growth and survival competition with wild-type *E. coli*. In a second experiment, the *yjbJ* insertion and deletion strains were compared (see Materials and Methods). Figure 6A shows the survival of the *yjbJ* and the viable *hdeA* insertion mutants in competition with the wild-type strain. Under these conditions, the *hdeA* deletion causes a slight growth defect with respect to wild-type EMG2, while the *yjbJ* insertion strain outcompetes the wild-type strain. Figure 6B shows the competition results for the strain with the *yjbJ* insertional allele versus the strain with the deletion *yjbJ* allele. Surprisingly, two different phenotypes are observed for the different mutant alleles. In this assay, the *yjbJ* insertion strain outcompetes the *yjbJ* deletion strain.

DISCUSSION

We have presented an improved method for performing gene replacements in *E. coli*. The method is similar to the pop-in/pop-out method used for *Saccharomyces cerevisiae* (6, 31, 33) and the hit-and-run procedure used for mouse embryonic stems cells (17). Unlike other methods used for gene replacements in *E. coli* that use ColE1 plasmids in a *polA1* background or transformation of linear DNA into *recBC*, *sbcB*, or *recD* strains, this protocol can be performed directly in wild-type strains (15, 21, 37, 44). Since the system is plasmid based, gene replacements are easily performed in any genetic background that is recombination proficient and supports the replication of pSC101 plasmids. Using this system, we have created another 44 *E. coli* strains with in-frame deletions of other ORFs (27a).

Although not attempted in our lab, the pKO3 gene replacement method can be used for constructing *E. coli* strains with multiple mutations without the need for multiple drug resistance markers or for replacing DNA sequences in the chromosome with precise point mutations. Finally, the method can be used for altering large exogenous fragments of DNA cloned into the single-copy P1 or BAC vectors which use *E. coli* as the host cell (38, 42).

In contrast to the deletion method, the insertion method creates mutations by inserting a Kn^r gene (interposon) into cloned chromosomal DNA segments similar to a previous protocol (29). We designed this method for a gene that is predicted to be essential and uses selection instead of a screen to assess gene replacement. The Kn^r gene was chosen as the marker since the gene has no homology to either the gene replacement vector pKO3 or the *E. coli* chromosome. We engineered the Kn^r interposons with a different multiplex sequencing tag flanking each side of the interposon so that mutagenized clones could be sequenced by either cycle or multiplex sequencing (10, 27).

The two distinct phenotypes resulting from the insertional mutagenesis of *hdeA* highlight the unreliability of insertional mutagenesis. The comparison of the *yjbJ* insertion and deletion strains in the competition experiment also illustrates the phenotypic differences that can occur as a result of the particular type of mutation created. The *yjbJ* insertion strain appears to have an advantage over both the wild type and its respective deletion strain under the selection condition tested. Insertional

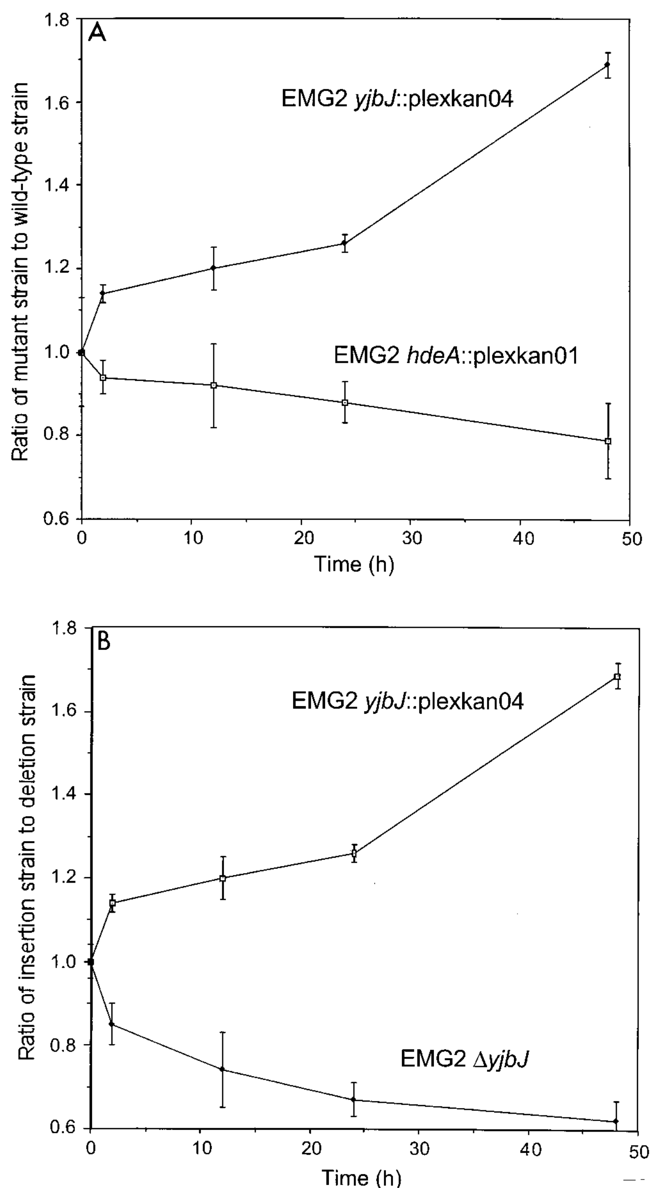


FIG. 6. Survival and growth competition between isogenic strains having either insertion or deletion alleles of *hdeA* and *yjbJ*. Equivalent numbers of cells from each strain were inoculated into rich medium and grown in competition under aerobic conditions at 37°C. At various time points, the cells were plated on rich media with and without kanamycin, and the viable cell density of each strain was assayed. (A) Relative survival of EMG2 *yjbJ*::plekan04 and EMG2 *hdeA*::plekan01 insertion strains when competed against wild-type EMG2. (B) Comparative survival of the EMG2 *yjbJ*::plekan04 insertion strain and the EMG2 Δ*yjbJ* deletion strain when cultured together.

mutagenesis has the potential for several undesired side effects, including polar termination-induced reduction of downstream operon expression (3), fusion products (2, 20), and misregulation of adjacent genes due to the insertion marker's promoter (11, 22, 43). Assigning a phenotype to a mutated gene may be problematic if the phenotype is actually a consequence of both the primary mutation and its effects on the surrounding genes.

This system, of replacing targeted ORFs with in-frame deletions was developed to reduce the inherent problems of insertional mutagenesis. Sensitive to the existence of transcrip-

tional and translational overlap in prokaryotic operons (13), our deletions were designed to retain translational coupling and to minimize the disruption of the regulation of neighboring genes in an operon (13, 24). The first six codons (18 bp) at the 5′ end of the gene were retained to maintain the gene's translation start signals. The last 12 codons (36 bp) at the 3′ end of the gene were retained based on the maximum overlap of coding regions observed in a sequence analysis of *E. coli* and *Salmonella typhimurium* operons (30). The largest observed overlap of a gene's 5′ coding region into a neighboring gene's 3′ coding region was 20 bp (5′-*cbiF-cbiG*-3′ in the *cob* operon of *S. typhimurium*). The 36-bp overlap was chosen to maintain translational coupling in a gene cluster for operons with potentially even greater overlapping regions and for ambiguity in downstream translation start site assignment.

The expected frequency of colonies bearing the deletion allele after resolution of the plasmid integrates is 50%. As expected, the frequency of colonies bearing the Kn^r insertion allele of *yjbJ* after plasmid resolution was approximately 50%. However, the observed resolution frequency of colonies with PCR-generated deletion alleles of both *yjbJ* and *hdeA* was only 3 to 7%. We speculate that this reduction in resolution frequency is caused by the reduced length of homologous sequences combined with possible PCR-generated DNA mismatches that flank one side of the duplications, causing the resolution of the integrate to be asymmetric (1, 34–36). In *E. coli*, a Chi site represented by the octanucleotide sequence 5′-GCTGGTGG-3′ stimulates recombination, depending on the length of the recombination interval and the location of the Chi site with respect to the interval (40, 41). We searched the genomic regions flanking *yjbJ* and *hdeA* and did not find the octanucleotide sequence in the vicinity of the two genes.

This research indicates that in the emerging post-genome sequencing era, when high-throughput evaluation of uncharacterized ORFs becomes a necessity, insertional mutagenesis by traditional methods will not be sufficiently reproducible to assign phenotypes based on subtle strain-by-strain variations. Because the engineering of in-frame deletions enables us to avoid many of the phenotypic artifacts mentioned earlier, we should be able to attach significance to a greater number of the phenotypes that we observe. This method will help investigators to systematically assign functions to the vast number of new ORFs revealed by current microbial sequencing projects.

ACKNOWLEDGMENTS

We thank Claire and Doug Berg for plasmid pBS-TS and the Kushner lab for plasmids pMAK700 and pMAK705. We thank Robert Kolter, Richard Baldarelli, Fritz Roth, and Pete Estep for helpful discussions. We are especially grateful to Elizabeth A. Malone and Martha Bulyk for critical reading of the manuscript.

This work was funded by DOE grant DE-FG02-87ER60565.

REFERENCES

1. **Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller.** 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29:**319–328.
2. **Bassford, P. J., T. J. Silhavy, and J. R. Beckwith.** 1979. Use of gene fusion to study the secretion of maltose-binding protein into *Escherichia coli* periplasm. J. Bacteriol. **139:**19–31.

2a. **Berg, C.** Personal communication.

3. **Berg, C. M., and D. E. Berg.** 1996. Transposable element tools for microbial genetics, p. 2588–2612. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
4. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513.
5. **Blomfield, I. C., V. Vaughn, R. F. Rest, and B. I. Eisenstein.** 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. Mol. Microbiol. **5:**1447–1457.
6. **Boeke, J. D., J. Truehart, G. Natsoulis, and G. R. Fink.** 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. **154:**164–175.
7. **Bollen, A., R. Lathe, A. Herzog, D. Denicourt, J. P. Lecocq, L. Desmarez, and R. Lavalle.** 1979. A conditionally lethal mutation of *Escherichia coli* affecting the gene coding for ribosomal protein S2 (*rpsB*). J. Mol. Biol. **132:**219–233.
8. **Brown, E. D., E. I. Vivas, C. T. Walsh, and R. Kolter.** 1995. MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. J. Bacteriol. **177:**4194–4197.
9. **Chang, S.-Y. P., E. C. McGary, and S. Chang.** 1989. Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. J. Bacteriol. **171:**4071–4072.
10. **Church, G. M., and S. Kieffer-Higgins.** 1988. Multiplex DNA sequencing. Science **240:**185–188.
11. **Ciampi, M. S., M. B. Schmid, and J. R. Roth.** 1982. Transposon Tn*10* provides a promoter for transcription of adjacent sequences. Proc. Natl. Acad. Sci. USA **79:**5016–5020.
12. **Dedonder, R.** 1966. Levansucrase from *Bacillus subtilis*. Methods Enzymol. **8:**500–505.
13. **Draper, D. E.** 1996. Translational initiation, p. 902–908. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
14. **Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado.** 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. **164:**918–921.
15. **Gutterson, N. I., and D. E. Koshland.** 1983. Replacement and amplification of bacterial genes with sequences altered *in vitro*. Proc. Natl. Acad. Sci. USA **80:**4894–4898.
16. **Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner.** 1989. New method for generating deletions and gene replacements in *Escherichia coli*. J. Bacteriol. **171:**4617–4622.
17. **Hasty, P., R. Ramirez-Solis, R. Krumlauf, and A. Bradley.** 1991. Introduction of a subtle mutation into the *Hox*-2,6 locus in embryonic stem cells. Nature **350:**243–246.
18. **Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease.** 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77:**51–59.
19. **Horton, R. M., H. D. Hunt, S. N. Ho, J. K. Pullen, and L. R. Pease.** 1989. Engineering hybrid gene without the use of restriction enzymes; gene splicing by overlap extensions. Gene **77:**61–68.
20. **Ito, K., P. J. Bassford, and J. R. Beckwith.** 1981. Protein localization in *E. coli*: is there a common step in the secretion of periplasmic and outer-membrane proteins? Cell **24:**707–717.
21. **Jasin, M., and P. Schimmel.** 1984. Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. J. Bacteriol. **159:**783–786.
22. **Kendrick, K. E., and W. S. Reznikoff.** 1988. Transposition of IS50L activates downstream genes. J. Bacteriol. **170:**1965–1968.
23. **Kleckner, N., J. Bender, and S. Gottesman.** 1991. Uses of transposons with emphasis on Tn*10*. Methods Enzymol. **204:**139–180.
24. **Landick, R., C. L. Turnbough, and C. Yanofsky.** 1996. Transcription attenuation, p. 1263–1286. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. Washington, D.C.
25. **Link, A. J., K. Robison, and G. M. Church.** 1997. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis **18:**1259–1313.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Murray, V.** 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res. **17:**8889.

27a. **Phillips, D., and G. Church.** Available at http://arep.med.harvard.edu/gmc/ecoko.html.

28. **Ponce, M. R., and J. L. Micol.** 1992. PCR amplification of long DNA fragments. Nucleic Acids Res. **20:**623.
29. **Prentki, P., and H. M. Krisch.** 1984. *In vitro* insertional mutagenesis with a selectable marker. Gene **29:**303–313.
30. **Roth, J. R., J. G. Lawrence, M. Rubenfield, S. Kieffer-Higgins, and G. M. Church.** 1993. Characterization of the cobalamin (vitamin B12) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. **175:**3303–3316.
31. **Rothstein, R.** 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. **194:**281–301.
32. **Rudd, K. E., W. Miller, C. Werner, J. Ostell, C. Tolstoshev, and S. G. Satterfield.** 1991. Mapping sequenced *E. coli* genes by computer: software, strategies and examples. Nucleic Acids Res. **19:**637–647.
33. **Scherer, S., and R. W. Davis.** 1979. Replacement of chromosome segments

with altered DNA sequences constructed *in vitro*. Proc. Natl. Acad. Sci. USA **76:**4951–4955.
34. **Senecoff, J. F., and M. M. Cox.** 1986. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. J. Biol. Chem. **261:**7380–7386.
35. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics **112:**441–457.
36. **Shen, P., and H. V. Huang.** 1989. Effect of base pair mismatches on recombination via the RecBCD pathway. Mol. Gen. Genet. **218:**358–360.
37. **Shevell, D. E., A. M. Abou-Zamzam, B. Demple, and G. C. Walker.** 1988. Construction of an *Escherichia coli* K-12 *ada* deletion by gene replacement in a *recD* strain reveals a second methyltransferase that repairs alkylated DNA. J. Bacteriol. **170:**3294–3296.
38. **Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA **89:**8794–8797.
39. **Slater, S., and R. Maurer.** 1993. Simple phagemid-based system for generating allele replacement in *Escherichia coli*. J. Bacteriol. **175:**4260–4262.
40. **Smith, G. R.** 1987. Mechanism and control of homologous recombination in *Escherichia coli*. Annu. Rev. Genet. **21:**179–201.
41. **Smith, G. R.** 1988. Homologous recombination in procaryotes. Microbiol. Rev. **52:**1–28.
42. **Sternberg, N.** 1990. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. USA **87:**103–107.
43. **Wang, A., and J. R. Roth.** 1988. Activation of silent genes by transposons Tn*5* and Tn*10*. Genet. **120:**875–885.
44. **Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161:**1219–1221.
45. **Yoshida, T., C. Ueguchi, and T. Mizuno.** 1993. Physical map location of a set of *Escherichia coli* genes (*hde*) whose expression is affected by the nucleoid protein H-NS. J. Bacteriol. **175:**7747–7748.
46. **Yoshida, T., C. Ueguchi, H. Yamada, and T. Mizuno.** 1993. Function of the *Escherichia coli* nucleoid protein, H-NS: molecular analysis of a subset of proteins whose expression is enhanced in a *hns* deletion mutant. Mol. Gen. Genet. **237:**113–122.