Genomics **73,** 56–65 (2001)
doi:10.1006/geno.2000.6451, available online at http://www.idealibrary.com on IDEAL®

A Highly Efficient *Escherichia coli*-Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA

E-Chiang Lee,* Daiguan Yu,† J. Martinez de Velasco,* Lino Tessarollo,* Deborah A. Swing,* Donald L. Court,† Nancy A. Jenkins,* and Neal G. Copeland*,[1]

*Mouse Cancer Genetics Program and †Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute–Frederick, Frederick, Maryland 21702

Received November 1, 2000; accepted November 16, 2000

Recently, a highly efficient recombination system for chromosome engineering in *Escherichia coli* was described that uses a defective λ prophage to supply functions that protect and recombine a linear DNA targeting cassette with its substrate sequence (Yu *et al.,* 2000, *Proc. Natl. Acad. Sci. USA* 97, 5978–5983). Importantly, the recombination is proficient with DNA homologies as short as 30–50 bp, making it possible to use PCR-amplified fragments as the targeting cassette. Here, we adapt this prophage system for use in bacterial artificial chromosome (BAC) engineering by transferring it to DH10B cells, a BAC host strain. In addition, arabinose inducible *cre* and *flpe* genes are introduced into these cells to facilitate BAC modification using *lox*P and *FRT* sites. Next, we demonstrate the utility of this recombination system by using it to target *cre* to the 3′ end of the mouse neuron-specific enolase (*Eno2*) gene carried on a 250-kb BAC, which made it possible to generate BAC transgenic mice that specifically express Cre in all mature neurons. In addition, we show that fragments as large as 80 kb can be subcloned from BACs by gap repair using this recombination system, obviating the need for restriction enzymes or DNA ligases. Finally, we show that BACs can be modified with this recombination system in the absence of drug selection. The ability to modify or subclone large fragments of genomic DNA with precision should facilitate many kinds of genomic experiments that were difficult or impossible to perform previously and aid in studies of gene function in the postgenomic era. © 2001 Academic Press

INTRODUCTION

Bacterial artificial chromosomes (BACs) (Shizuya *et al.,* 1992) have become the tool of choice for generating long-range physical maps, positionally cloning disease genes, and whole-genome sequencing due to their ability to be stably propagated. However, for them to become an equally powerful tool for functional genomics, easy methods for precisely manipulating BAC DNA are required. A significant advance in our ability to manipulate DNA was provided by Zhang *et al.* (1998), who showed that it is possible to modify DNA in *Escherichia coli* by homologous recombination using the RecET proteins of *E. coli.* This methodology makes it possible to target PCR-amplified linear DNA fragments with short regions of homology (<60 bp) at their ends to virtually any target DNA such as a high-copy plasmid, the *E. coli* chromosome, or a BAC (Muyrers *et al.,* 1999; Zhang *et al.,* 1998).

Recently, an alternative approach for chromosome engineering in *E. coli* was described that makes use of a defective prophage to supply functions that protect and recombine the electroporated linear DNA (Yu *et al.,* 2000). This defective prophage carries a deletion between *cro* and *bio*A (Yu *et al.,* 2000). The *P*L operon encoding *gam* and the *red* recombination genes, *exo* and *bet,* is under the tight control of the temperature-sensitive λ repressor (allele *cI857*). Recombination functions can thus be transiently supplied by shifting the cultures to 42°C for 15 min. This recombination system does not require RecA function and depends primarily on the expression of Exo, Beta, and Gam. Gam inhibits the *E. coli* RecBCD nuclease from attacking the electroporated linear DNA, while Exo and Beta generate recombination activity.

λ recombination functions expressed from a plasmid have been used previously to modify a BAC (Muyrers *et al.,* 1999). However, the λ prophage system appears to be at least 50- to 100-fold more efficient than this, or the RecET system, where recombination functions are expressed from plasmids (Muyrers *et al.,* 1999; Narayanan *et al.,* 1999; our unpublished results). Expressing the λ recombination functions from plasmids also creates other problems not presented by the prophage.

[1] To whom correspondence should be addressed at the Mouse Cancer Genetics Program, Building 539, Room 229, National Cancer Institute–Frederick, Frederick, MD 21702. Telephone: (301) 846-1260. Fax: (301) 846-6666. E-mail: copeland@ncifcrf.gov.

0888-7543/01 $35.00

For example, drug resistance is used to maintain plasmids in the cell. For experiments using gene replacement, drug-resistant markers often become rate-limiting. A plasmid-based system also precludes using that particular plasmid type for recombination cloning studies because of incompatibility. Importantly, plasmid expression systems are leaky; thus, Gam and Red functions are always present at some level. The presence of Gam causes a RecBCD defect, a condition that results in plasmid instability and loss of cell viability (Feiss *et al.,* 1982; K. Sergueev, unpublished results). Gam and Red can also cause BAC instability.

Here, this prophage system is modified for use in BAC engineering, and the utility of this system for recombinogenic targeting and subcloning of BAC DNA is demonstrated. The ability to modify large fragments of genomic DNA with precision should facilitate studies of gene function in the postgenomic era.

MATERIALS AND METHODS

Bacterial strains. All of the strains used in this study, except DH10B, were maintained at 32°C because of the temperature-inducible prophage. DY303 was constructed by infecting DH10B cells (Gibco) with a A phage carrying *recA* (A*cI857 recA* +) (a gift from F. W. Stahl), and lysogens were selected. Strain EL11 was constructed by replacing the *tet*gene of DY380 with a cassette containing the *cat* and *sac*B genes. EL11 cells are Tet^S, Cm^R, and sensitive to 2% sucrose. Strain EL250 was constructed by replacing the *cat–sac*B cassette of EL11 cells with *ara*C and the arabinose promoter-driven *flpe* recombinase gene (P_{BAD}*flpe*). EL250 cells are resistant to 2% sucrose. Strain EL350 was constructed in a similar manner except that *cre* replaced *flpe.*

Construction of plasmids. The *IRES– eGFPcre–FRT– kanamycin* (*kan*)*–F RT* targeting cassette was PCR-amplified from pICGN21, which was constructed by subcloning a 1.9-kb *Hin*dIII/*Acc*I-digested and filled-in *FRT– kan–FRT* fragment from pFRTneo into the *Not*I/*Bcl*I-digested and filled-in cloning site of pIRESeGC. The FRTneo was constructed by amplifying the *kan* gene along with the [3-lactamase promoter from pEGFP-C1 (Clontech) with primers 5'CTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCGTCAGGTGGCACTTTCGGG and 5'CTCAGAAGAACTCGTCAAGAAGG. The amplified fragment was then targeted between the *FRT* sites in pNeo[3-gal (Stratagene). The pIRESeGC was generated by inserting the 2-kb *Nhe*I/*Mlu*I-digested and filled-in eGFPcre fragment from pEGC into the 3.5-kb *Bam*HI-digested and filled-in cloning site of pNTRlacZPGKneoloxP (Arango *et al.,* 1999). The pEGC was generated by subcloning a 1.05-kb *Eco*RI/*Kpn*I PCR fragment containing the *cre* gene from pGKmncre (a gift from P. Soriano) into the *Eco*RI/*Kpn*I site of pEGFP-C1. This PCR fragment was generated by amplifying the *cre* gene from pGKmncre with primers 5'GTAGGTACCTCGAGAATCGCCATCTTCCAGCAGGC and 5''TCGAATTTTCTGCATCCAATTTACTGACCGTACACC, which contain *Eco*RI and *Kpn*I cleavage sites, respectively, at their 5' ends.

To construct the pTamp vector, the *amp*-targeted pBeloBAC11 was first generated by replacing the *lox*P site in pBeloBAC11 (Shizuya *et al.,* 1992) with the PCR-amplified *amp* gene from pEGFP (Clontech). The primers used for amplification are 5'*GCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTG* TCTTAGACGTCAGGTGGCAC and 5'*CATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGAAC* CTCACGTTAAGGGATTTTGGTC, which are homologous to the *amp* gene of pEGFP (in roman type) and to sequences flanking the *loxP* site in pBeloBAC11 (in italics). A 2.4-kb PCR fragment amplified from *amp*-targeted pBeloBac11 with primers 5'GCAGGATCCAGTTTGCTCCTGGAGCGACA and 5'TGCAGGTCGACTCTAGAGGATC was then cloned into the *Xho*I/*Xba*I and filled-in site of pCS (Stratagene) to create the pTamp vector. The 2.4-kb *amp* cassette containing an *amp* gene along with 920 bp of 5', and 370 bp of 3', pBel0BAC11 vector sequence flanking the *lox*P site can be released by *Bam*HI digestion and used directly to replace the *loxP* site in any pBeloBAC11-derived BACs with *amp.*

The pKO4 vector containing the *cat–sac*B targeting cassette is a derivative of pKO3 (Link *et al.,* 1997) in which 605 bp between *cat* and *sac*B had been deleted.

The *ara*C–P_{BAD}*flpe* targeting cassette was amplified from pBADflpe, which was constructed by subcloning a 1.4-kb *Pst*I/*Kp n* I fragment from pOGFlpe (Buchholz *et al.,* 1998) into pBAD/MycHis-A (Invitrogen). The *ara*C–P_{BAD}*cre* targeting cassette was amplified from pBADcre, which was constructed by introducing a 1.2-kb *Hin*dIII/*Nco*I fragment from pGKmncre into pBAD/His-C (Invitrogen).

Amplification primers for targeting or GAP repair cassette DNAs. For all primers listed below, nucleotides in italics are homologous to the targeted sequence, while those in roman type are homologous to amplification cassettes. The Tet^R cassette used for targeting *cro– bio* in DY330 was amplified from Tn10 with primers 5'*TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATG* CTCTTGGGTTATCAAGAGGG and 5'*GGCGCTGCAAAAATTCTTTGTCGAACAGGGTGTCTGGATC* ACTCGACATCTTGGTTACCG. The *cat–sac*B cassette used for replacing the *tet* gene in DY363 was amplified from pKO4 with primers 5'*TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATG* CTGTGACGGAAGATCACTTCG and 5'*GGCGCTGCAAAAATTCTTTGTCGAACAGGGTGTCTGGATC* CTGAGGTTCTTATGGCTCTTG. The *ara*C–P_{BAD}*flpe* and *ara*C–P_{BAD}*cre* cassettes used for replacing the *cat–sac*B in EL11 were amplified from pBADflpe and pBADcre with primers 5'*TGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTGTATG* AAGCGGCATGCATAATGTGC and 5'*GGCGCTGCAAAAATTCTTTGTCGAACAGGGTGTCTGGATC* CTGTGTCCTACTCAGGAGAGCGTTC. The *IRES– eGFPcre–FRT– kan–FRT* cassette used for targeting the *Eno2* locus was amplified from pIGCN21 with primers 5'*CGCTTCGCGGGACATAATTTCCGAAATCCCAGTGTGCTGTGA*GCCAAGCTATCGAATTCCGCC and 5'*GAGGCTCCAGGAGAATGAGATGTTCCCGCGTTCAGGCAAGCG* CTATTCCAGAAGTAGTGAGGA. The oligonucleotides used to target the flag cassette into the 5' end of the *Sox4* gene were annealed and polymerase-extended using primers 5'*GCGAGCGTGTGAGCGCGCGTGGGCGCCCGGCAAGCCGGGGCCATG* GATTACAAGGATGACGACGATAAGGTACAACAGA and 5'*GGCCAGCAGAGCCTCAGTGTTCTCCGCGTTGTTGGTCTGTTGTAC* CTTATCGTCGTCATCCTTGTAATCCATGGCCCCC. The linear pBR322 derivative used to subclone the 25-kb fragment from the modified *Eno2* locus was amplified with primers 5'*CTCTCCATGCCTGTCTGGGTGAGGGTGGCCCAGGGGCGATGGCTATGAGAGAG* GTCGACTTCTTAGACGTCAGGTGGCAC (*Eno2*-C-L1) and *GCAATGCAGAGAAGCCTTGTACTGGGA TGACAGAGACGGAGGGGAAGAGGAG* GCGGCCGCGATACGCGAGCGAACGTGA (*Eno2*-C-R1/2). The amplification primers for the other experiments were as follows: 48-kb modified fragment, 5'*GACTTCTATGACCTGTACGGAGGGGAGAAGTTTGCGACGTGACAGAGCTGGTC*GTCGACTTCTTAGACGTCAGGTGGCAC (*Eno2*-C-L2/3/4) and *Eno2*-C-R1/2; 60-kb modified fragment, *Eno2*-C-L2/3/4 and 5'*GCCCCATACACGTAAATGTACATAGAATCACACAGCATCACTTCTATGGAT* GCGGCGGCCGCGATACGCGAGCGAACGTGA (*Eno2*-C-R3); 80-kb modified fragment, *Eno2*-C-L2/3/4 and 5'*CATCCAGTAGAACTTGGGAGTGAAGCTAGAGCCAAGGCCATCTAAGTGACAG*GCGGCCGCGATACGCGAGCGAACGTGA (*Eno2* -C-R4). These primers contained 5' regions homologous to the target sequence and 3' regions homologous to pBR322. PCR products were purified using a Qiaex II gel extraction kit (Qiagen) and digested with *Dpn*I to remove contaminated template.

Preparation of electrocompetent cells and generation of recombinants. For BAC modification, overnight cultures containing the BAC were grown from single colonies, diluted 50-fold in LB medium, and grown to an OD_{600} = 0.5–0.7. Ten-milliliter cultures were then induced for Beta, Exo, and Gam expression by shifting the cells to 42°C for 15 min followed by chilling on ice for 20 min. Cells were then centrifuged for 5 min at 5500*g* at 4°C and washed with 1.5 ml of ice-cold sterile water three times. Cells were then resuspended in 50

TABLE 1

Bacterial Strains Constructed in This Work

Strains	Genotype
DH10B	F⁻ *mcr*A (*mrr–hsd*RMS–*mcr*BC) cp80d*lac*Z M15 *lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 (*ara, leu*)7649 *gal*U *gal*K *rsp*L *nup* G
DY303	DH10B [A*cl*857 *rec*A⁺]
DY330	W3110 *lac*U169 *gal*490 [A*cl*857 (*cro–bio*A)] W3110
DY363	*lac*U169 *gal*490 [A*cl*857 (*cro–bio*A) <> *tet*][a] DH10B
DY380	[A*cl*857 (*cro–bio*A) <> *tet*]
EL11	DH10B [A*cl*857 (*cro–bio*A) <> *cat–sac*B]
EL250	DH10B [A*cl*857 (*cro–bio*A) <> *ara*C–P$_{BAD}$*flpe*][b]
EL350	DH10B [A*cl*857 (*cro–bio*A) <> *ara*C–P$_{BAD}$*cre*]

[a] (*cro–bio*A) <> *tet* indicates substitution of *cro–bio*A with *tet.*
[b] P$_{BAD}$ represents the promoter of *ara*BAD.

μl of ice-cold sterile water and electroporated. For BAC transformation, the induction step was omitted.

Cell transformation was performed by electroporation of 100–300 ng linear DNA into 50 μl of ice-cold competent cells in cuvettes (0.1 cm) using a Bio-Rad gene pulser set at 1.75 kV, 25 μF with a pulse controller set at 200 ohms. One milliliter of LB medium was added after electroporation. Cells were incubated at 32°C for 1.5 h with shaking and spread on appropriate selective or nonselective agar media.

Production of transgenic mice. Modified BAC and the p25-kb subclone DNAs were purified using cesium chloride gradients as described (Antoch *et al.,* 1997). The 25-kb subclone DNA was linearized by *Not*I digestion before microinjection. BAC DNA (1 μg/ml) and 25-kb subclone DNA (2 μg/ml) were microinjected into the pronucleus of (C3H/HeN-Mtv⁻ 3 C57BL/6Ncr)F_2 zygotes. Transgenic founders were subsequently identified by Southern analysis using a *cre* probe or by PCR using primers 5'CTGCTGGAAGATGGCGATTCTCG and 5'AACAGCAGGAGCGGTGAGTC that flank the 3' insertional junction.

Histochemical analysis of [3-*galactosidase expression.* Mice at 4 to 5 weeks of age were sacrificed in CO_2 and perfused with 4% paraformaldehyde in PBS (pH 7.3). The brains, spinal cords, and eyes were removed and postfixed for 3 h. Vibratome sections (20 μm) of brains were mounted on slides and used directly for X-gal staining or for immunocytochemistry. For spinal cords and eyes, cryostat sections (20 μm) were used that were made by cryoprotecting tissues in 30% sucrose in PBS overnight and embedding the tissues in freezing compound (OCT, Sakura). Before X-gal staining, samples on slides

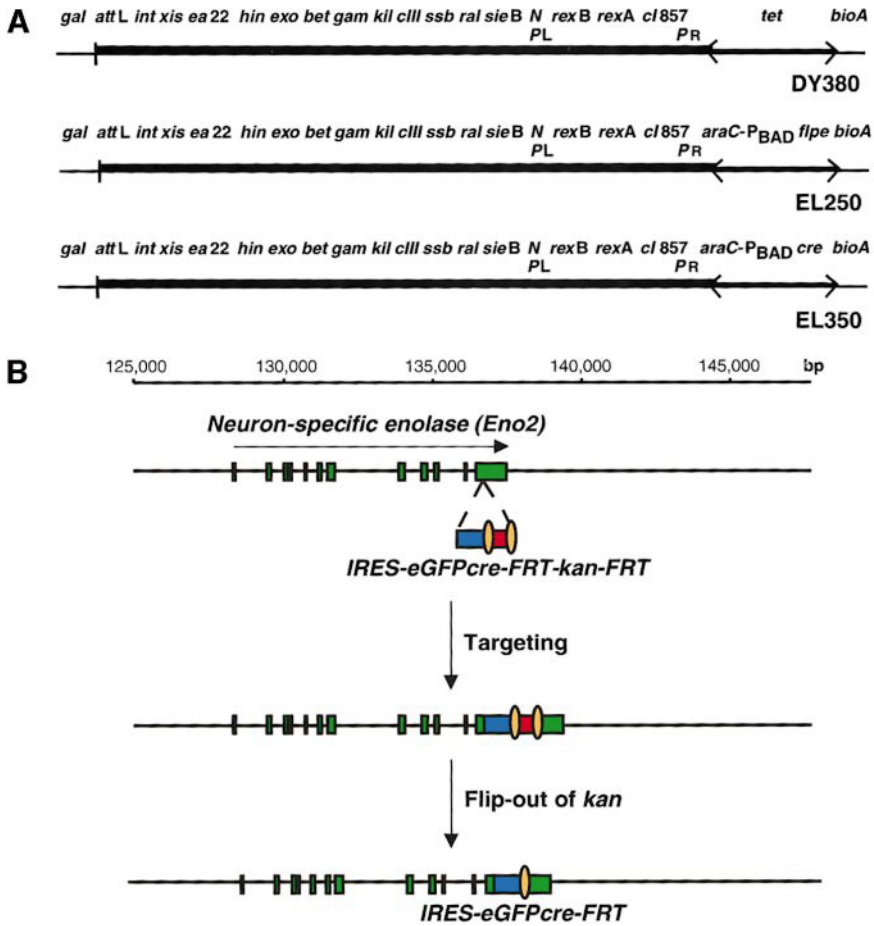

FIG. 1. General strategy for BAC engineering. (**A**) The defective prophages used for BAC engineering contain the A genes from *cI*857 to *int. P*L and *P*R denote the left and right promoters, respectively. The *gam* and *red* genes, *exo* and *bet,* are under the control of *P*L, which is repressed by the temperature-sensitive repressor cI857 at 32°C and derepressed at 42°C. *tet* replaces the segment from *cro–bio*A in DY380 cells. The *ara*C–P$_{BAD}$*flpe* cassette or the *ara*C–P$_{BAD}$*cre* cassette replaces the segment from *cro–bio*A in EL250 or EL350 cells, respectively. The promoter of the *ara*BAD operon (P$_{BAD}$), which can be induced by L-arabinose, controls the expression of the *flpe* or *cre* genes. Thick black lines designate the prophage while thin lines represent *E. coli* sequence. Angle brackets define the ends of the *cro–bioA* region that was replaced with *tet, araC*–P$_{BAD}$ *flpe,* or *araC*–P$_{BAD}$ *cre.* (**B**) The relative position of the *Eno2* gene in the fully sequenced 250-kb BAC, 284H12, and the different steps used to introduce Cre into the last exon of *Eno2.* In the targeting cassette, *FRT* sites are denoted by orange ellipses, the *kan* gene is represented by a red rectangle, and the *GFPcre* fusion gene is represented by a blue rectangle. The green boxes represent *Eno2* exons.

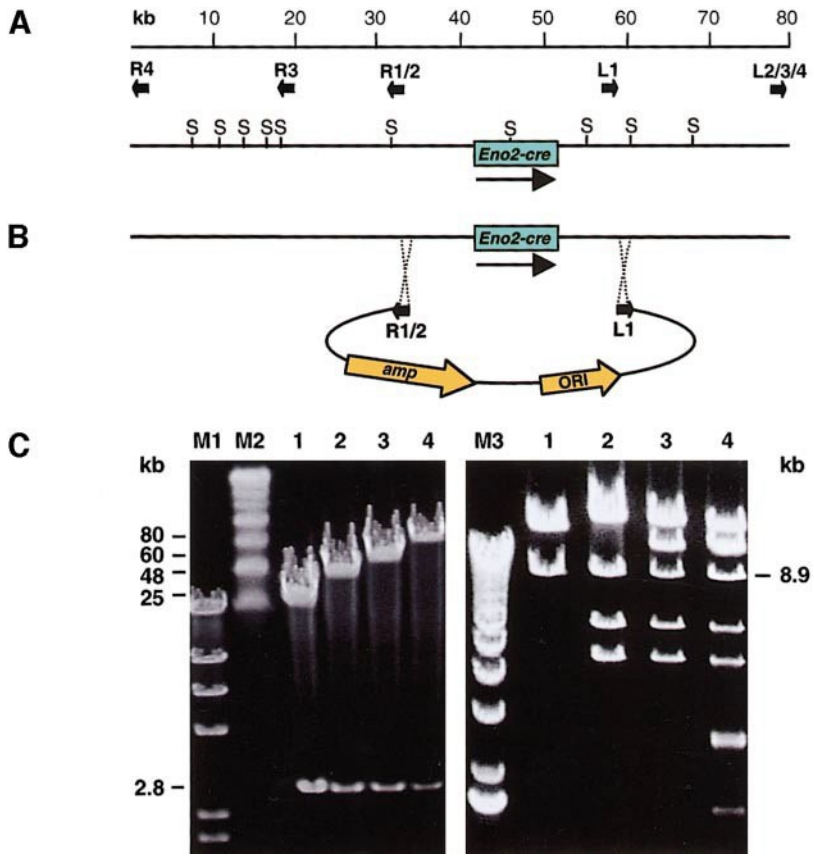

FIG. 2. Fragments as large as 80 kb can be subcloned from BACs by gap repair. (**A**) The location of the 5' homologies on the amplification primers used to amplify pBR322 for subcloning by gap repair are shown as thick black arrows. Each primer also contains 20-nt segments at its 3' end to prime pBR322. *Not*I and *Sal*I cleavage sites were included in these primers to facilitate release of the subcloned fragments from the plasmid backbone. The location of *Spe*I restriction sites near *Eno2* is also shown. *Spe*I restriction sites are not present on the linear amplified pBR322 vector. (**B**) Gap repair intermediate showing pairing between a typical amplified pBR322 targeting cassette and the modified *Eno2* BAC. amp, *amp*-resistance gene; ORI, origin of replication. (**C**) (**Left**) Representative CHEF gel results of *Not*I/*Sal*I-double digested BAC DNA from the different subcloning experiments. A 2.8-kb vector fragment and a rescued fragment can be seen in each lane. **M1,** A/*Hin*dIII marker. **M2,** midrange marker II (New England Biolabs). The *Spe*I-digested restriction pattern of each BAC is shown in the right panel. As expected, an 8.9-kb band containing the 3' end of the *Eno2* gene is seen in all the lanes. **M3,** 1-kb ladder (Gibco).

were postfixed with 0.25% glutaraldehyde in PBS and briefly washed with rinse solution (0.1 M phosphate buffer, pH 7.3, 0.1% deoxycholic acid, 0.2% NP-40, and 2 mM $MgCl_2$). X-gal staining was performed by incubating samples in staining buffer (2.5 mg/ml X-gal, 5 mM potassium ferricyanide, and 5 mM potassium ferrocyanide in staining buffer) for 2 h at 37°C followed by counterstaining with 0.25% eosin (Fisher).

Immunocytochemistry. Immunostaining was carried out using the ABC Vectastain kit (Vector Labs) on 20-μm vibratome sections. Sections were blocked with PBS (pH 7.3; containing 0.2% Triton X-100, 1.5% bovine serum albumin, and 5% normal goat serum) at room temperature for 2 h and incubated with primary *Eno2* antibody, a polyclonal rabbit anti-*Eno2* antiserum (Chemicon) at 1:100 dilution in PBS solution. After incubation with a secondary biotinylated antibody and the ABC reagent, peroxidase was reacted with 0.05% diaminobenzidine tetrahydrochloride and 0.003% hydrogen peroxide.

RESULTS

Transfer of the Defective A *Prophage into DH10B Cells*

None of the *E. coli* strains carrying the defective A prophage generated in previous experiments can be efficiently transformed with BAC DNA (Yu *et al.,* 2000), precluding their use for BAC engineering. To facilitate its use for BAC engineering, the prophage was introduced into DH10B cells, a BAC host strain. The prophage could not, however, be transferred by standard genetic methods because DH10B is *recA* $^-$. To circumvent this problem, a *tet* selectable marker was introduced between *cro* and *bio*A in the defective A prophage carried in strain DY330 (Table 1). A P1 lysate was then prepared on this new strain (DY363; Table 1) and used to infect DY303 cells, a DH10B strain that carries a A *cI857recA* + lysogen (Table 1). By selecting cells that are *recA* $^-$ based upon their UV sensitivity, it was possible to obtain cells in which *tet* replaced the *recA* $^+$ gene in the A *cI857recA* + lysogen. This new strain, DY380 (Table 1), can be transformed with BAC DNA at efficiencies of 10^{-6} to 10^{-4} (data not shown). The targeting efficiency of DY380 cells is also similar to that of the original *recA* $^-$ DY strain (Yu *et al.,* 2000).

Creation of DY380 Derivatives Containing Arabinose-Inducible Cre or Flpe Genes

BAC targeting often makes use of a selectable marker to introduce the targeting cassette into the targeted locus. The selectable marker can, however, interfere with the subsequent function of the targeted locus. If the selectable marker is flanked with *FRT* or *lox*P sites, it can be removed from the targeted locus by Flp or Cre recombinases, thus eliminating this problem. To facilitate this operation in *E. coli,* two new strains, EL250 and EL350, were constructed, by replacing the *tet* gene in the prophage carried in DY380 cells with *araC* and the arabinose-inducible *flpe* and *cre* genes, respectively (Fig. 1A, Table 1). *tet* is located between *cI*857 and *bioA* in the DY380 prophage. *flpe* is a genetically engineered *flp* that has a higher recombination efficiency than the original *flp* gene (Buchholz *et al.,* 1998). Thus, both strains have homologous recombination (the A *red* genes) and site-specific recombination (*flpe* or *cre*) functions, with the former controlled by temperature and the latter by arabinose. This dual regulation allows both selective targeting by recombination and the subsequent removal of the selection marker from the targeted locus by site-specific recombination.

The General Strategy for BAC Engineering

To test the efficiency of this prophage system for BAC engineering, an *IRES– eGFPcre–FRT– kan–FRT* cassette was targeted to the *Eno2* locus carried on a 250-kb BAC (284H12, Research Genetics) (Fig. 1B). The *Eno2* gene is located in the middle of this fully sequenced BAC (Ansari-Lari *et al.,* 1998). The *Eno2* gene was targeted because it is neural-specific and expressed in most mature neurons (Marangos and Schmechel, 1987). The goal was to create a BAC transgenic line that expresses Cre in all mature neurons for use in conditional knockout studies. A BAC approach was used since conventional transgenes often lack important regulatory sequences required for proper gene expression.

The *Eno2* BAC was modified using previously described methods (Yu *et al.,* 2000). First, the *Eno2* BAC was electroporated into EL250 cells, and six chloramphenicol-resistant (Cm^R) colonies were selected. Digestion of BAC DNA from six Cm^R colonies with *Eco*RI or *Hin*dIII showed that one had an abnormal digestion pattern. However, in other BAC electroporation experiments involving the analysis of more than 76 additional colonies, no abnormal BACs were identified (our unpublished results). These results indicate that BAC rearrangements during electroporation are rare. The *IRES– eGFPcre–FRT– kan–FRT* cassette was then PCR-amplified from a template plasmid, pIGCN21, using chimeric 63-nt primers. The 3' 21 nt of each primer was homologous to the targeting cassette used for amplification, while the 5' 42 nt was homologous to the last exon of *Eno2* where the cassette was to be targeted by recombination. The primers were designed to target precisely the cassette downstream of the *Eno2* stop codon and upstream of its poly(A) site. EL250 cells carrying the BAC were then shifted to 42°C for 15 min to induce Exo, Beta, and Gam expression. The cells were then electroporated with 300 ng of the amplified cassette, and kanamycin-resistant (Km^R) colonies were selected. Approximately, 5200 Km^R colonies were obtained from 10^8 electroporated cells for a targeting efficiency of ~10^{-5}. No colonies were obtained from uninduced cells.

Whole-cell PCR analysis of 24 selected colonies using primers that flanked the targeted locus indicated that all were correctly targeted. Sequencing of the targeted region from six colonies, however, showed that three carried point mutations. To determine whether these point mutations were introduced during PCR amplification or during homologous recombination, the targeting was repeated. This time, however, the PCR-amplified *IRES– eGFPcre–FRT– kan–FRT* cassette was subcloned into the *Sma* I site of pBluescript by blunt-end ligation before targeting, and plasmids carrying wildtype amplified cassettes were identified by DNA sequencing. These cassettes were then released from the plasmid by *Bam*HI digestion and used for targeting. Using this two-step method, all 12 targeted BACs that were subsequently sequenced contained wildtype *IRES– eGFPcre–FRT– kan–FRT* cassettes. These results indicate that the point mutations were introduced during PCR amplification of the targeting cassette rather than during targeting. Surprisingly, the targeting efficiency using this two-step approach was similar to direct targeting even though the *Bam*HI fragment contains 5'-protruding ends and extra bases from the pBluescript polylinker that bear no homology to the targeted region. This does not mean that subcloning and sequencing should be performed routinely as the targeted BACs must still be sequenced to verify that they carry wildtype targeted cassettes.

Next, the *kan* selectable marker was removed to prevent its possible interference with Cre expression. Overnight cultures from single Km^R colonies were diluted 50-fold in LB medium and grown till $OD_{600} = 0.5$. Flpe expression from the EL250 cells was then induced by incubating the cultures with 0.1% L-arabinose for 1 h. The bacterial cells were subsequently diluted 10-fold in LB medium, grown for an additional hour, and spread on chloramphenicol plates (12.5 μg/ml). The next day, 100 Cm^R colonies were picked and replated on *kan* plates (25 μg/ml) to test for loss of kanamycin resistance. All colonies were Km^s and contained a single *FRT* site at the targeted locus. The high recombination efficiency likely reflects the tight control of Flpe expression afforded by the P_{BAD} promoter and the fact that the *FRT* sites are located *in cis* rather than *in trans* to one another.

Finally, the *lox*P site contained in the BAC vector backbone, pBeloBAC11 (Shizuya *et al.,* 1992), was removed by a final round of gene targeting. This was

performed to prevent any concatemerized BAC transgenes from being excised from the mouse germline by Cre recombinase. To facilitate the removal of this *lox*P site, a new plasmid, pTamp, was constructed that contains an *amp* gene flanked by 920 bp of pBeloBAC11 sequence located 5' of the *lox*P site and 370 bp of pBeloBAC11 sequence located 3' of the *lox*P site (Shizuya *et al.,* 1992). This *amp* insert can be released from pTamp by *Bam*HI digestion and used to replace the *loxP* site in the BAC transgene by gene targeting (see Materials and Methods). This targeting reaction is very efficient due to the large amount of homology between the *amp* cassette and the pBeloBAC11 vector (56,200 colonies per 10^8 electroporated cells).

Subcloning by GAP Repair

This recombination system can also be used to subclone fragments from BACs without the use of restriction enzymes or DNA ligases. In this case, subcloning relies on gap repair to recombine the free ends of a linear plasmid vector with homologous sequences carried on the BAC (Fig. 2B). The linear plasmid vector with an *amp* selectable marker and an origin of replication carries the recombinogenic ends. The vector is generated by PCR amplification using two chimeric primers. The 5' 45–52 nt of each primer is homologous to the two ends of the BAC sequence to be subcloned while the 3' 20 nt is homologous to plasmid DNA. Recombination generates a circular plasmid in which the DNA insert was retrieved from the BAC DNA via gap repair. Circular plasmids are selected by their Amp^R.

To determine the maximum-sized fragment that can be subcloned from BACs using this method, several different pairs of primers were generated in which the homology segments were located 25, 48, 60, or 80 kb apart in the *Eno2* BAC DNA (Fig. 2A). Rare-cutter *Not*I and *Sal*I restriction sites were also incorporated into these primers so that the subcloned fragments could be released from the recombinant clones intact. Using pBluescript as the cloning vector, it was possible to subclone the 25-kb fragment, but it was impossible to subclone larger fragments (data not shown). Perhaps subclones containing larger fragments on a high-copy vector are toxic to the cell.

To determine whether this is the case, pBR322 was used as the cloning vector with its copy number control element intact. As shown in Fig. 2C (left panel), fragments as large as 80 kb could be subcloned with this lower copy number vector. Unlike targeting, however, not all subclones had the correct inserts as determined by restriction enzyme pattern analysis (Fig. 2C, right panel). Some subclones lacked inserts while others contained inserts with aberrant restriction patterns (Table 2). It is important therefore to fingerprint enough subclones so that subclones with wildtype inserts are identified.

TABLE 2

Subcloning Fragments from the Modified BAC by Gap Repair

Rescued fragment size (kb)	No. Amp^R colonies[a]	No. Correct recombinants/ No. colonies examined[b]
25	555	8/12 (1, 3)[c]
48	504	4/12 (0, 8)
60	514	3/24 (2,19)
80	487	2/24 (1,21)

[a] Induced EL250 competent cells carrying the modified BAC were electroporated with 300 ng of the different recombinogenic amplified pBR322 cloning vectors. The total number of Amp^R colonies from each experiment is shown. The number of surviving colonies without drug selection was ~10^8 for each experiment.

[b] Twelve or 24 Amp^R colonies were randomly picked and analyzed by Southern analysis following *Not*I/*Sal*I digestion. Correct recombinants contained a 2.8-kb cloning vector and the expected size insert.

[c] The first number in parentheses indicates the number of colonies in which the amplified cloning vector ends were joined without undergoing recombination. The second number in parentheses depicts colonies with inserts that are aberrant in size.

The ability to subclone large fragments of genomic DNA by gap repair should facilitate many studies in genome research that were difficult or impossible to perform previously.

Production of Transgenic Mice

To determine whether the modified BAC contains all of the regulatory sequences needed for neural-specific Cre expression, it was injected into (C3H/HeN-Mtv^- X C57BL/6Ncr)F_2 zygotes. A BAC transgenic line carrying approximately two copies of the transgene was then established. While eGFP was included in the original Cre targeting cassette to provide a visual marker for Cre expression, none of the subsequent Cre-expressing transgenic lines expressed detectable levels of eGFP (data not shown). The reason for this is unclear as cultured embryonic fibroblast cells transfected with this same targeting cassette expressed readily detectable levels of eGFP (data not shown). As a control, two transgenic lines carrying the 25-kb subclone were also established (Fig. 2). The 25-kb subclone contains the entire modified *Eno2* coding region as well as 10 kb of 5' flanking sequence and 5 kb of 3' flanking sequences (Fig. 2A). One transgenic line, 25kbp-1, carries approximately four copies of the transgene, while the second, 25kbp-2, carries approximately five copies of the transgene.

To assess Cre activity, the transgenic mice were crossed to ROSA26 reporter mice, which contain a lacZ reporter that can be activated by Cre recombinase (Soriano, 1999). Double heterozygotes were subsequently analyzed by X-gal staining at 4 weeks of age. Several different tissues were examined for X-gal expression including the brain, spinal cord, eye, lung, heart, intestine, muscle, liver, spleen, and kidney. Blue-stained cells were found only in neural tissue in the three transgenic lines, indicating that both the BAC and the

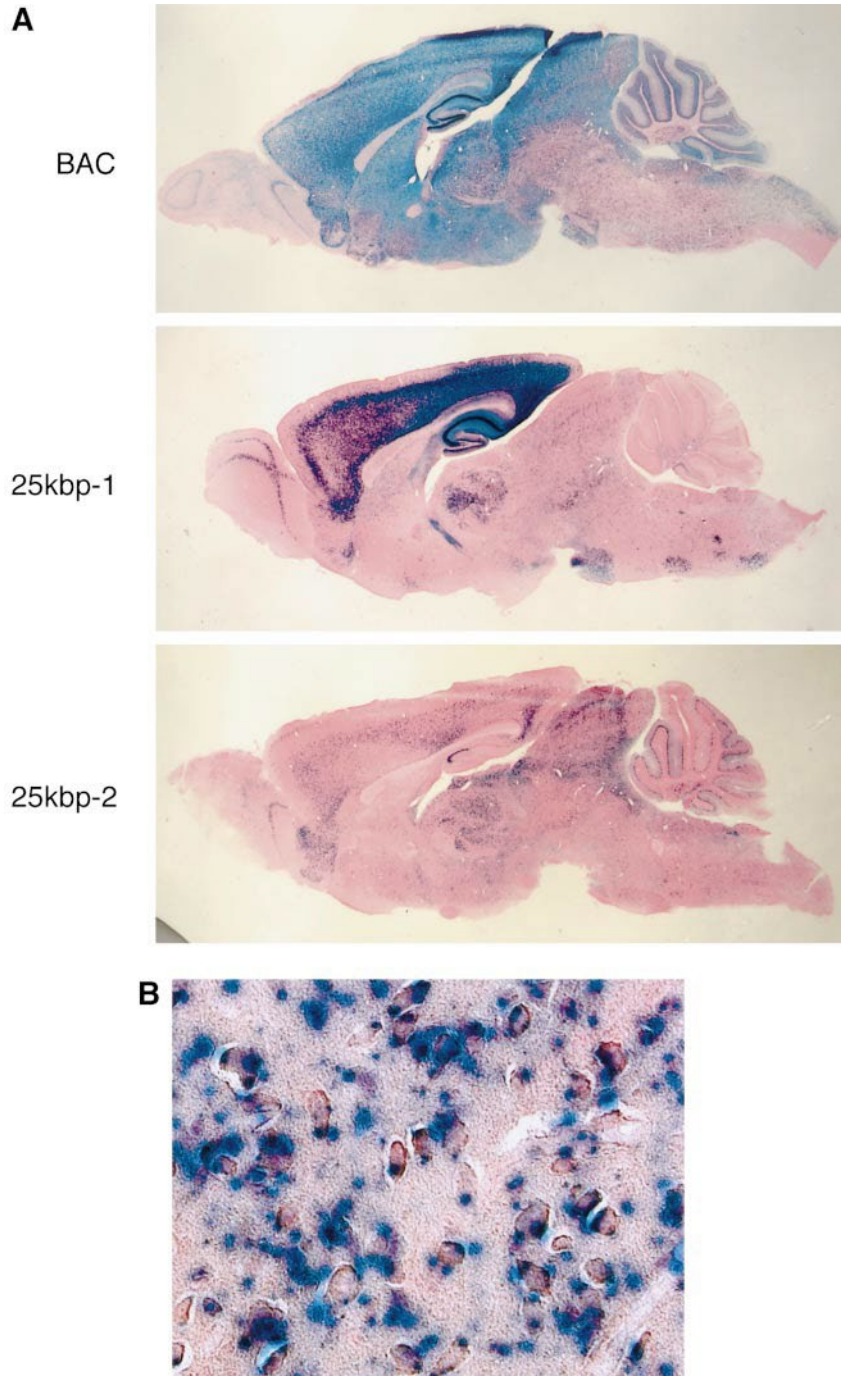

FIG. 3. Cre activity in the brains of transgenic mice carrying the Rosa26R reporter gene. (**A**) X-gal-stained sections from the brains of 4-week-old animals carrying the Rosa26R reporter gene in addition to the BAC transgene, the 25kb-1 transgene, or the 25kb-2 transgene. X-gal staining is indicative of Cre activity. (**B**) A section from the superior colliculus region of the brain of a BAC transgenic animal was immunostained with an anti-Eno2 antibody (Chemicon) followed by X-gal staining for Cre activity. Blue staining is indicative of Cre activity, while brown staining indicates Eno2 protein-positive cells.

25-kb subclone contain the regulatory elements needed for neural-specific expression. The pattern of Cre activity was, however, different in the three lines. Vibratome sections of the brain from the BAC transgenic mice showed blue-stained cells throughout the gray matter but not in the white matter, indicative of Cre activity in most neurons but not in glial cells (Fig. 3A). In contrast, X-gal staining in the 25kb-1 and 25kb-2 transgenic mice was present in only a subset of neurons, and expression was variable between the two different lines (Fig. 3A).

Higher power magnification of the cerebellum of the BAC transgenic mice showed that Cre was expressed in virtually all neuronal cells. This included Purkinje cells in the Purkinje cell layer, granule and Golgi cells in the granular layer, basket cells and

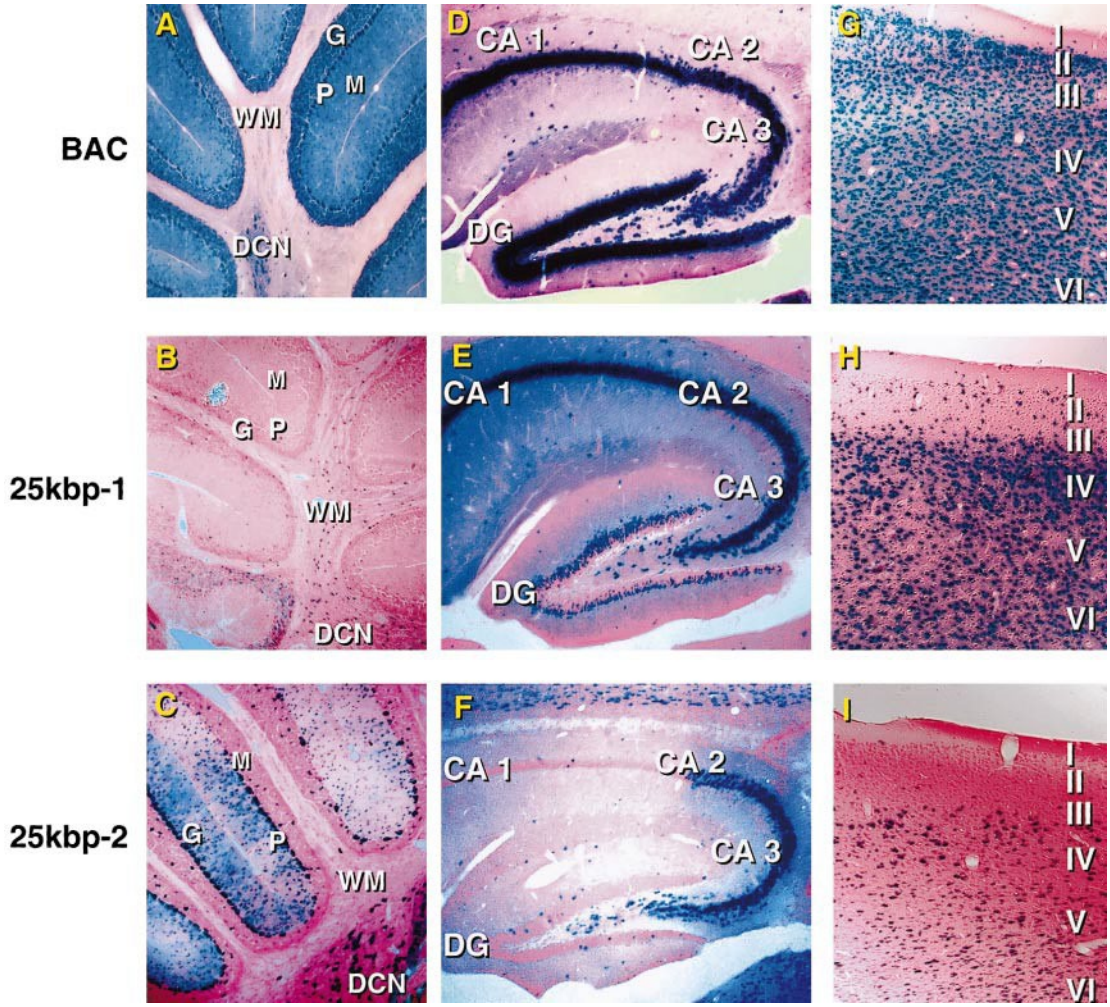

FIG. 4. Cre activity in the cerebellum, hippocampus, and cortex of transgenic mice. Higher power magnification of X-gal-stained sections from the cerebellum (**A, B,** and **C**), hippocampus (**D, E,** and **F**), and cortex (**G, H,** and **I**) of a BAC transgenic animal (**A, D,** and **G**), a 25kbp-1 transgenic animal (**B, E,** and **H**), and a 25kbp-2 transgenic animal (**C, F,** and **I**). The Purkinje cell layer (P), granular layer (G), molecular layer (M), deep cerebellar nuclei (DCN), and white matter (WM) of the cerebellum, the three subregions of the CA (CA1, CA2, and CA3), the dentate gyrus (DG) of the hippocampus, and the six layers (I to VI) of the cortex are labeled.

stellate cells in the molecular layer, and neurons of the deep cerebellar nuclei (Fig. 4A). In contrast, in the 25kbp-1 line, Cre was expressed in only a subset of Golgi cells in addition to a few cells in the granule and Purkinje cell layers (Fig. 4B). Glial cells of white matter also expressed Cre, indicative of leaky expression. In the 25kbp-2 line, Cre expression was limited to the gray matter and included a variety of neuronal cell types, including most basket cells, stellate cells, Purkinje cells, and neurons of the deep cerebellar nuclei (Fig. 4C). In contrast, few granule cells and Golgi cells in the granule layer expressed Cre.

Higher power magnification of the hippocampus and cortex showed similar results. In the hippocampus of BAC transgenic mice, virtually all neurons in the cornu Ammonis (CA) region and the dentate gyrus (DG) expressed Cre (Fig. 4D). The same was true in the cortex, where all five layers of the cortex that contained neurons (layers II–VI) expressed Cre. In contrast, the hippocampus of 25kbp-1 transgenic mice showed reduced Cre expression in the DG (Fig. 4E) and layers II and III of cortex (Fig. 4H). The 25kbp-2 transgenic mice showed even lower levels of Cre expression in the DG (Fig. 4F). The CA1 and CA2 regions of the CA also failed to express Cre (Fig. 4F). Cre expression was also greatly reduced in the cortex, with layers II and III showing the most reduction (Fig. 4I).

Cre activity in the spinal cord, dorsal root ganglion, and retina of the transgenic mice was also examined to determine whether Cre was expressed in mature neurons within the peripheral nervous system. Similar to what was observed for the central nervous system, Cre was expressed in most mature peripheral neurons in the BAC transgenic mice, while fewer peripheral neurons expressed Cre in the two 25-kb transgenic lines (data not shown).

To determine whether Cre was expressed in all Eno2 protein-positive neurons, a section from the brain of a BAC transgenic animal was immunostained with an anti-Eno2 antibody followed by X-gal

staining for Cre activity. As shown in Fig. 3B, virtually all Eno2-positive neurons were active for Cre.

Targeting without Selection

The high level of efficiency of recombination obtained with this prophage system suggested that it might be possible to perform targeting without drug selection. Direct targeting would facilitate genomic experiments in which the presence of a selectable marker, or even a *FRT* or *lox*P site, is undesirable. To determine whether targeting can be achieved without drug selection, a 24-bp flag tag was targeted to the 5' end of the SRY-box containing gene 4 (*Sox4*) carried on a 125-kb BAC. For these experiments, a 114-bp targeting cassette was generated in which two 45-bp arms homologous to the *Sox4* gene flanked the 24-bp flag sequence. This DNA fragment was created by synthesizing two 79-bp oligonucleotides that overlapped at their 3' ends by 44 bp. These overlaps were annealed and filled in by *Taq* polymerase. Following electroporation of the flag-tagged cassette into induced DY380 cells carrying the *Sox4* BAC, the cells were spread on LB plates to a density of ~2000 cells per plate. Colonies containing the flag tag were subsequently identified by colony hybridization using a 30-bp flag-specific oligonucleotide probe (24-bp flag tag and 3 bp on each side that was homologous to the *Sox4* targeted site). Among 3800 colonies screened from uninduced cells, no flag-positive colonies were identified. In contrast, 7 flag-positive colonies were identified in 4210 colonies obtained from induced cells for an overall targeting frequency of 1.7×10^{-3}. PCR amplification and direct sequencing showed that each of the seven flag-positive colonies was correctly targeted.

DISCUSSION

In the studies presented here, we describe a highly efficient recombination system for manipulating BAC DNA in *E. coli* that uses a defective λ prophage to supply functions that protect and recombine the electroporated linear DNA targeting cassette with the BAC sequence. Because the recombination functions are expressed from a defective prophage rather than a plasmid, the recombination functions are not lost during cell growth as often happens with plasmid-based systems. Another advantage of this prophage system is that the λ *gam* and *red* recombination genes are under the control of the temperature-sensitive λ repressor that provides a much tighter control of *gam* and *red* expression than can be obtained on plasmids. This tight regulation, combined with the strong λ *P*L promoter, which drives *gam* and *red* expression to very high levels, makes it possible to achieve recombination frequencies that are at least 50- to 100-fold higher than those obtained with plasmid-based systems (Narayanan *et al.,* 1999; Muyrers *et al.,* 1999; our unpublished results).

The ability to manipulate large fragments of genomic DNA precisely, independent of the location of appropriate restriction enzyme sites, has many applications for functional genomics, both in the mouse and in other organisms. As shown here, Cre can be introduced into the coding regions of genes carried on BACs facilitating the generation of Cre-expressing transgenic lines for use in conditional knockout studies or for use in conditional gene expression studies. Genes can also be epitope-tagged and microinjected into the germline of mice carrying a mutation in the gene. If the epitope-tagged transgene rescues the mutant phenotype, the epitope-tagged protein is functional, and the epitope tag can serve as a marker for expression of the gene. Likewise, a gene carried on a BAC can be replaced with another gene, and the function of the "knock-in" mutation can be assayed in transgenic mice.

This recombination system also facilitates the generation of complicated conditional targeting vectors. While the generation of such vectors often used to take several months, it can now be performed in a only few weeks. The ability to express reversibly Cre or Flpe recombinases in *E. coli* speeds this process even further. A selectable marker flanked with *lox*P or *FRT* sites can now be introduced into an intron of a gene and then removed by transient Cre or Flpe expression, leaving behind a solo *lox*P or *FRT* site in the intron. A limitation of this approach at the present time is the lack of a BAC-based mouse physical map and the paucity of mouse genome sequence information. This should all dramatically change, however, next year, as the draft sequence of the mouse comes on-line and the BAC physical map is completed.

The high recombination efficiency offered by this recombination system also makes it possible to manipulate BAC DNA without drug selection. Point mutations, deletions, or insertions can now be engineered into any gene on a BAC in the absence of a confounding linked drug selection marker or a *lox*P or *FRT* site. In cases where the gene is mutated in human disease, the exact disease-causing mutations can be engineered on the BAC, and the effect of these mutations can be analyzed in transgenic mice.

This recombination system also makes it possible to subclone fragments as large as 80 kb from BACs by gap repair. Targeting vectors or transgenic contructs generated by BAC engineering can now be subcloned with ease, and virtually any region of the engineered BAC can be included in the final subclone. Subcloning by gap repair should also facilitate the identification of regulatory elements or locus control regions that may be located at some distance from a gene. Many such potential regulatory elements will be identified over the next few years by comparative genome sequencing. The ability to modify precisely these regulatory sequences on BACs, combined with the ability to include or exclude them during the subcloning process, should make it possible to dissect the function of these se-

quences in the whole animal at a level not previously possible.

ACKNOWLEDGMENTS

We thank P. Soriano for the Rosa26 cre reporter line and the pPGKmncre vector and R. Behringer for the pNTRlacZPGKneoloxP vector. The National Cancer Institute, DHHS, supported this work. All vectors and strains described in this article are freely available for basic research purposes. To request these vectors and strains, please send an e-mail to N. G. Copeland at copeland@ncifcrf.gov.

REFERENCES

Ansari-Lari, M. A., Oeltjen, J. C., Schwartz, S., Zhang, Z., Muzny, D. M., Lu, J., Gorrell, J. H., Chinault, A. C., Belmont, J. W., Miller, W., and Gibbs, R. A. (1998). Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res.* **8:** 29–40.

Antoch, M. P., Song, E. J., Chang, A. M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H., and Takahashi, J. S. (1997). Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell* **89:** 655–667.

Arango, N. A., Lovell-Badge, R., and Behringer, R. R. (1999). Targeted mutagenesis of the endogenous mouse Mis gene promoter: In vivo definition of genetic pathways of vertebrate sexual development. *Cell* **99:** 409–419.

Buchholz, F., Angrand, P. O., and Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **16:** 657–662.

Feiss, M., Siegele, D. A., Rudolph, C. F., and Frackman, S. (1982). Cosmid DNA packaging in vivo. *Gene* **17:** 123–130.

Link, A. J., Phillips, D., and Church, G. M. (1997). Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli:* Application to open reading frame characterization. *J. Bacteriol.* **179:** 6228–6237.

Marangos, P. J., and Schmechel, D. E. (1987). Neuron specific enolase, a clinically useful marker for neurons and neuroendocrine cells. *Annu. Rev. Neurosci.* **10:** 269–295.

Muyrers, J. P., Zhang, Y., Testa, G., and Stewart, A. F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27:** 1555–1557.

Narayanan, K., Williamson, R., Zhang, Y., Stewart, A. F., and Ioannou, P. A. (1999). Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in *E. coli* DH10B using an inducible homologous recombination system. *Gene Ther.* **6:** 442–447.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (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.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21:** 70–71.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **97:** 5978–5983.

Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli. Nat. Genet.* **20:** 123–128.