RECOMBINEERING IN PROKARYOTES

Lynn C. Thomason, Richard S. Myers, Amos Oppenheim, Nina Costantino, James A. Sawitzke, Simanti Datta, Mikhail Bubunenko, and Donald L. Court

19

A new in vivo genetic engineering technology, recombineering, has been developed for use in *Escherichia coli*. Due to its precision, efficiency, and simplicity, recombineering may replace standard in vitro cloning techniques. Recombineering is enabled by generalized recombination functions encoded by lambdoid bacteriophages; these proteins efficiently catalyze homologous recombination in vivo between DNA sequences with homologies as short as 35 bases. These recombinases work in the absence of most other bacteriophage genes, and in certain cases, a single phage protein can catalyze the reaction. Recombineering is carried out with PCR products or synthetic oligonucleotides (oligos) and provides methods to clone and modify genes carried on plasmids, bacterial artificial chromosomes (BACs), the chromosomes of enteric bacteria, and bacteriophages without the necessity of restriction enzymes or DNA ligase. This new technology allows rapid and precise in vivo manipulations of DNA and shows great promise for use in functional genomic analyses. This chapter describes the present state of recombineering in *E. coli* and details the essential elements of the system. The success of this technology has encouraged its development for use in other organisms. We discuss that goal and some obstacles that must be overcome in order to fully utilize recombineering in other prokaryotes.

In the life of nearly all organisms, homologous recombination is recognized as a key process for DNA replication and repair as well as for the generation of genetic diversity. Recombination has also been an important tool for studies of bacterial gene function. Molecular biology and modern biology have expanded in the >30 years since the development of genetic engineering. In vitro genetic engineering, although extremely useful, depends on and is limited by the presence of unique and appropriately positioned restriction sites. Recombineering is a recently developed method of genetic engineering that does not rely on restriction sites for cloning; instead, it uses short

Lynn C. Thomason, Nina Costantino, James A. Sawitzke, Simanti Datta, Mikhail Bubunenko, and Donald L. Court, Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute-Frederick, Frederick, MD 21702–1225. *Richard S. Myers*, Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL 33101–6129. *Amos Oppenheim*, Department of Molecular Genetics and Biotechnology, The Hebrew University-Hadassah Medical School, Jerusalem, Israel 91120.

Phages: Their Role in Bacterial Pathogenesis and Biotechnology
Edited by Matthew K. Waldor, David I. Friedman, and Sankar L. Adhya

regions of homology and bacteriophage proteins capable of catalyzing recombination between these homologies to link novel combinations of genes and other genetic elements in vivo. Genetic engineering by phage-promoted in vivo homologous recombination is gaining widespread acceptance as scientists discover its ease of use, high efficiency, and versatility. It is now possible to make "ideal" DNA constructs for experimental needs: an exact desired DNA molecule can nearly always be generated by recombineering. Another powerful use of recombineering is the retrieval of chromosomal DNA onto a plasmid, in a "gap-repair" reaction. This cloning by retrieval is done without amplifying the desired chromosomal clone by PCR; thus, PCR-generated mutations are avoided. The advantages of recombineering for functional genomic analyses have been clearly demonstrated, and it is already being used to modify eukaryotic clones carried on BACs and P1 artificial chromosomes (PACs) in *E. coli* (16, 61). Recombinants can be made in a single step, only short homologies are required to achieve efficient recombination, and in some instances, the frequencies of recombination are high enough to permit screening for desired constructs in the absence of selection (17).

For the creation of new genetic constructs by recombineering, a bacterial strain expressing a bacteriophage recombination system is required. The proteins used to catalyze recombineering are from lambdoid phages, and *E. coli* is the standard host. Bacteriophage λ carries three genes that are critical for recombineering. The *exo* and *bet* genes encode a 5′ to 3′ double-strand-dependent exonuclease, Exo, and a single-stranded DNA (ssDNA) annealing protein, Beta, respectively (82). Together, these two proteins can catalyze recombination of a double-stranded PCR product with flanking homologies to the desired genetic target (60). The λ *gam* gene encodes a protein, Gam, which inhibits the *E. coli* RecBCD enzyme (82) and thus prevents degradation of the linear double-stranded PCR product, allowing it to recombine. These three functions comprise the λ Red recombination system. Single-stranded oligonucleotides (ss-oligos or ssDNAs) can also be targeted to a desired location in a reaction requiring only the Beta protein (25). The λ Red proteins catalyze efficient recombination in the absence of the *E. coli* RecA function (9, 25, 79, 98, 100).

A cryptic lambdoid prophage called Rac is found in some strains of *E. coli* (40). This prophage contains the *recE* and *recT* genes, which encode homologous recombination functions analogous to those of *exo* and *bet*, respectively (31, 38, 39). In fact, RecE and RecT can substitute for Exo and Beta in λ (30) and for recombineering (100). Mutants of the Rac prophage have been selected in which the expression of these recombination genes has been activated. These *sbcA* (suppressor of *recBC*) mutations result in increased recombination in the absence of RecBCD (7).

The recombineering enzymes can be expressed from a truncated lambda prophage (25, 98), from synthetic operons in the *E. coli* genome (58), or from a plasmid-based system (21, 100). The defective prophage lacks genes for DNA replication and lytic functions and is incapable of producing a functional virus particle. Of these three systems, the prophage system has the advantage of tight gene regulation combined with high coordinate expression, which results in higher recombination frequencies (see below). This is an important advantage, since in many cases high recombination frequencies are essential for obtaining the desired recombinant. In addition, having the recombination system on the bacterial chromosome eliminates plasmid incompatibility issues when the recombineering reaction is targeted to a plasmid. On the other hand, plasmid-based systems have the advantage of mobility: they can be transferred even to non-*E. coli* strains (10, 22, 34, 92). At the end of this chapter, we describe new plasmid systems which utilize the phage control regions, thereby providing ease of mobility between species and highly coordinated gene regulation.

For the creation of different chromosomal constructs by recombineering, the modifying DNA, either a double-stranded PCR product or a synthetic ss-oligo, is introduced into the Red recombination-expressing strain by electropo-

ration. Recombinants are obtained either by selection or by screening of the population of cells surviving electroporation. After the desired construct is obtained, the now unwanted prophage or plasmid may be removed in a subsequent step. Alternatively, engineered alleles on the chromosome can be moved into a different host by bacteriophage P1 transduction. Unlike some other in vivo genetic engineering methods (74; for a review, see reference 19), recombineering is largely independent of the host *recA* function (98). In many cases, a recombineering host strain that has a defective *recA* gene is preferable, since unwanted molecular rearrangements are less likely to occur if the strain is recombination proficient only during a short time when the phage functions are induced (16). A detailed discussion of the molecular mechanisms of recombination and recombineering can be found in the work of Court et al. (19). In this chapter, we emphasize the biochemistry of the proteins used for recombineering, the genetic manipulations possible, and the relative efficiencies of each. We also discuss potential issues in developing recombineering for prokaryotes other than *E. coli*.

BIOCHEMISTRY OF THE Red SYSTEM

λ Exo: a 5′ to 3′ dsDNA-Dependent Exonuclease

λ Exo degrades linear dsDNA in a 5′ to 3′ direction from both ends. Estimates of the degradation rate from bulk measurements and from single-molecule studies hover around 10 to 30 bases per s in vitro (11, 51, 70, 85, 94), although in a study by Matsuura et al. (53), the remarkably high rate of 1,000 bases per s was reported, with single-molecule kinetics. When dsDNA has 5′-phosphoryl ends, single nucleotides are removed processively to leave long 3′ ssDNA overhangs (11, 36, 51, 85). This reaction initiates less efficiently in the absence of a 5′ P. λ Exo does not initiate at nicks or gaps in the DNA (12, 13); the enzyme requires a dsDNA end to initiate digestion and remains bound to the dsDNA as it degrades one strand. The active form of the protein is a trimeric toroid with a tunnel through the protein (45, 95). The entrance hole to the tunnel accommodates a dsDNA, but the exit hole diameter is the size of ssDNA. Thus, Exo binds a dsDNA end and slides along the 3′-ended single strand that it generates while cleaving mononucleotides from the 5′ strand, resulting in a 3′ overhang (45, 85).

λ Beta: a ssDNA Binding Protein That Anneals Complementary ssDNA Strands

The λ Beta protein binds stably to ssDNAs (72) of >35 nucleotides (65, 98), protects DNA from digestion by single-strand nucleases (41, 56), and promotes pairing or annealing between complementary ssDNAs (41, 44, 56). RecT (and thus Beta by analogy) has been reported to have a limited strand invasion activity (66), but the spectrum of recombination events that the Beta and RecT proteins catalyze in vivo is consistent with a predominantly single-strand annealing activity (28, 83).

Combined Action of Exo and Beta

Based on genetic and biochemical studies, Exo and Beta are thought to form a complex in vivo (13, 78, 96, 99). As Exo degrades a dsDNA from the 5′ end, it generates a ssDNA overhang to which Beta binds (Color Plate 13) and ultimately pairs with a complementary ssDNA target. The Beta-Exo complex has been termed SynExo, for synaptase (Beta) exonuclease (Exo) (96). A compelling model for this complex based on in vitro data (12) received impressive support in vivo (83). The model provides for mechanistic coupling between the processing nuclease and the homologous pairing activity. A similar model has been inferred from studies of RecBCD and RecA actions (2). The proposed coupling of a specific exonuclease with its partner annealing protein may be essential for regulating the nuclease activity, limiting DNA digestion to the amount needed for annealing and recombination. Heterospecific pairs of single-strand annealing protein and exonuclease components fail to produce active SynExo recombinases (62). In other words, RecE-RecT and Red Exo-Beta initiate recombination at

dsDNA breaks by specifically interacting with their respective partners, but the RecE-Beta and Exo-RecT hybrid combinations have not been demonstrated to catalyze recombination (62), probably because of their protein-protein interaction specificities.

λ Gam: a Modifier of RecBCD and SbcCD

Although RecBCD is present at some 10 molecules per *E. coli* cell (86), it rapidly destroys any linear dsDNA within the cell. The λ Gam protein has been shown to bind stoichiometrically to the RecBCD enzyme, forming a Gam-RecBCD complex (42, 57) and inhibiting RecBCD activities, including the potent nuclease (42, 57, 91). Genetic studies indicate that Gam also inhibits a second *E. coli* enzyme, the SbcCD endonuclease (14, 27, 46). In vivo, SbcCD has been shown to repair double-strand breaks on the bacterial chromosome through recombination with a sister chromosome (20). Thus, Gam inhibits two nucleases which are involved in double-strand-break-dependent recombination. If RecBCD is inactivated by mutation, Exo and Beta can catalyze the recombination of linear DNA in the absence of the Gam protein (58). As yet, it is unclear whether the SbcCD function is involved in linear DNA recombination.

RecBCD destroys rolling-circle replication products. By inhibiting RecBCD, Gam action can lead to abnormal rolling-circle replication of plasmids present in the cell (15). This aberrant rolling-circle replication, which depends on an active recombination pathway (80, 81), can cause plasmid instability and cell toxicity (47). This abnormal replication also complicates the targeting of some plasmids by recombineering (25, 98). These issues are discussed more fully in the section on targeting recombineering to plasmids.

GENETIC MANIPULATIONS POSSIBLE WITH RECOMBINEERING

Several different types of genetic construction can be engineered by recombineering, and the desired genetic product dictates the type of DNA substrate used in the recombineering reaction. The following four types of substrates have been used successfully in our laboratory and other laboratories: PCR products; short, partially dsDNA created by annealing ss-oligos; gapped linear plasmid DNA; and ssDNA oligonucleotides. Table 1 gives typical efficiencies for each of these reactions.

Gene or Chromosome Replacement with dsDNA

With wild-type *E. coli*, recombineering using a dsDNA substrate requires the three phage proteins Exo, Beta, and Gam (98). Exo and Beta are needed to process the PCR product prior to its incorporation into the chromosome, and Gam is needed to prevent the degradation of the linear dsDNA by the RecBCD nuclease and possibly by the SbcCD nuclease (46).

PCR products with flanking homology to the target DNA can be used to create heterologous insertions, deletions, and gene replacements (Color Plate 14). Such PCR products are created by using chimeric primers, with 35

TABLE 1 Typical recombination efficiencies for standard recombineering reactions

Reaction conditions	No. of recombinants/10^8 viable cells
galK < > *cat* dsDNA targeted to *E. coli*[a]	2.5×10^4
70-base oligos overlapping by 34 bp with 5′ overhangs; selection for Gal$^+$[b]	1.5×10^5
Retrieval by gap repair of linear plasmid from *E. coli* chromosome[c]	1.6×10^2
70-base *galK*$^+$ ss-oligo targeted to *E. coli*[d]	3.5×10^5

[a]DY330, Δ(*argF-lac*)*U169 gal490* (λ*cI857-bioA*) (98).
[b]DY411, W3110 (λ*cI857-bioA*) *galK* < > *kan*, with 34 bp of *galK* deleted (99).
[c]DY378, W3110 (λ*cI857-bioA*) (89).
[d]HME6, W3110 Δ(*argF-lac*)*U169* $galK_{TYR145UAG}$ (λ*cI857-bioA*). The oligo creates a T-C mismatch that is well repaired by the MMR system (17).

to 50 nucleotides (nt) homologous to the desired target at the 5′ end followed by an adequate priming sequence (~20 nt) at the 3′ end, to amplify the DNA to be inserted (such as an antibiotic resistance gene or cassette) (98). We have successfully deleted a 70-kb region by this method (16). A dsDNA substrate is especially useful when the desired construct has a precise deletion of a gene and its concomitant replacement with a drug marker. If a desired change cannot be selected, e.g., inserting a green fluorescent protein (GFP) fusion, a two-step system can be used (Color Plates 14 and 15). The first step is the direct selection of a dual-function (selectable and counterselectable) cassette at the desired location; the second step is the removal of that marker by its replacement with the final desired sequence by counterselection. In the first reaction, for example, a set of two linked genes, a *cat sacB* cassette, may be inserted, and chloramphenicol resistance is selected. The *sacB* gene confers sensitivity to sucrose (26). This *cat sacB* cassette is removed in a second round of recombineering by replacing the cassette with a desired DNA sequence and selecting on minimal plates containing sucrose. Recombinants in which the desired DNA has replaced the *cat sacB* cassette will grow. This general strategy is useful for making gene and operon fusions, for the production of specific hybrid proteins, and for constructing reporter gene fusions (48, 60, 63). As with all counterselections, adequate outgrowth must be provided to ensure complete chromosome segregation so that the recombinant phenotypes can be expressed, and in this example, all of the *cat sacB* gene copies are segregated. We have shown (unpublished data) that a complete loss of *sacB* function can take at least 3 h in Luria-Bertani medium. For a detailed protocol, see the work of Thomason et al. (88).

Molecular tags such as GFP, His tags, and antigen tags can be inserted by dsDNA (or ss-oligo) recombineering. Counterselection can be used to insert a DNA segment for which there is no direct positive selection by two rounds of recombineering, as discussed above. Alternatively, a DNA inserted without selection or counterselection can be identified directly by colony hybridization to a labeled DNA probe. This method of recombinant identification requires very efficient recombineering, with ≥1 recombinant/10,000 colonies, so that the desired recombinant can be found by screening manageable numbers of colonies (48). Since cells are growing in exponential phase in rich broth when they are made competent for electroporation, they contain multiple copies of their chromosome (77). The recombining DNA often goes to only one chromosome copy per cell (17), and hence the cell will be heterozygous for the introduced phenotype, and when plated nonselectively, recombinant colonies will be sectored. Colonies testing positive by hybridization should be purified, and new isolates should be retested. Alternatively, adequate outgrowth time for complete chromosome segregation can be allowed before plating the cells. However, longer outgrowths will result in lower apparent frequencies and will make screening more laborious.

dsDNA recombineering can also be done with overlapping ss-oligos that anneal when they are coelectroporated, resulting in short dsDNAs with single-stranded overhangs (99). It was shown previously that if the annealed oligos have 5′ ssDNA overhangs, then recombination will be high, reaching an efficiency of >0.1%. The oligos need only overlap by 10 bases. This technique can be efficiently used to mutagenize a region of a gene. First, a small deletion (<70 bp) in the gene of interest is made. Two complementary oligos are then used to repair the deletion and to insert desired codons in the gene. Multiple overlapping oligos were also shown to yield recombinants (99), raising the possibility of using recombineering with overlapping oligos to generate entire genes, in an in vivo reaction analogous to PCR assembly (84).

Retrieval onto a Plasmid

A second useful manipulation which can be done by recombineering is cloning by the retrieval of DNA onto a linearized plasmid vector. The following two types of reaction are feasible: (i) the insertion of a PCR product or

other linear dsDNA fragment directly into a gapped plasmid and (ii) the rescue of a DNA segment from the chromosome or an episome into a gapped plasmid (Color Plate 16). For this gap repair method, PCR can be used to amplify all or part of a plasmid, creating a linear product with homology to the target at the ends. The linear plasmid DNA, containing its replication origin and a selectable marker, is introduced by electroporation into cells which are competent for recombineering. Recombination with the target DNA regenerates a circular plasmid and confers drug resistance. The plasmid DNA is isolated and screened for the retrieved insertion. Retrieval onto a plasmid works well, with the following caveat: the Red system will catalyze "recombination" between directly repeated sequences as small as 6 bp (43, 99, 101), and thus, if the linear ends of the transformed plasmid contain even relatively short repeats, a high background of end-joined vector lacking an insert is possible. Such end-joining events are more likely when a vector is linearized within a multiple cloning site, as these regions often have many repeat sequences. A judicious choice of vector, careful primer design, and analysis of the sequence within a few hundred bases of the gapped linear DNA ends for the presence of such repeats can avoid this problem. In a slight modification of the retrieval procedure, a gene of interest can first be tagged by the insertion of a nearby drug marker, and then a PCR-amplified plasmid origin lacking a drug marker can be used to retrieve both the gene of interest and the adjacent antibiotic resistance gene (see Color Plate 19). This modification eliminates any background of religated vector.

The retrieval method has been used successfully with pBR322 and pACYC184 (48, 101) derivatives, but not with pSC101-based plasmids. The failure of pSC101 plasmids to retrieve DNA has at least two possible explanations. The replication of pSC101 requires the plasmid-encoded Rep protein (5), and when a linear pSC101 molecule is introduced into the cell by transformation, it may be unable to establish itself in the absence of Rep function. Another possibility is that rolling-circle replication is actually required for successful gap repair. Whereas colE1 and p15A origin-based plasmids undergo a rapid transition to rolling-circle replication, the pSC101 plasmid enters rolling-circle replication much more slowly (8).

If the insertion of a PCR product or oligo into a plasmid is desired, the substrate can be either cotransformed with the linear plasmid, as described above, or targeted to a plasmid already resident in cells that are proficient for recombineering, depending on the *ori*. Direct targeting of a PCR product to a resident plasmid works with all plasmid derivatives (25). However, complex multimer topologies can be generated with colE1 and p15A origins, often with only one copy in the multimer having the desired change. Multimer formation may be a result of rolling-circle replication stimulated by Gam. In general, if one desires to alter a colE1 or p15A vector, coelectroporating the plasmid and the linear fragment gives more satisfactory results. pSC101-based plasmids, on the other hand, are targeted well, and multimer formation is less of a problem.

Recombineering with ssDNA

With ssDNA oligonucleotides of about 70 bases in length, several types of genetic rearrangements can be created by recombineering. For example, it is possible to alter a single base or several bases, to create a nonsense or missense mutation, to delete a few bases, or alternatively, to remove >9 kb of DNA (25). ssDNA has been used to simultaneously delete a large DNA segment and replace it with up to 34 bp (99). This makes ssDNA recombineering especially useful for site-directed mutagenesis. There are several advantages to recombineering with ss-oligos. The efficiency of ss-oligo recombination can be higher than that of dsDNA and is increased substantially in a mismatch repair mutant or by using sequences that form poorly corrected DNA mismatches in the heteroduplex DNA intermediate (17). Recombineering with ss-oligos requires only the single-strand annealing protein, Beta (25). ss-oligos do not require processing by λ Exo, and in the presence of Beta they are not subject to degradation by

RecBCD exonuclease (ExoV). Thus, the Gam protein, which inhibits RecBCD, is not needed for recombineering with ss-oligos. Sometimes the recombinant change can be selected for by a loss or gain of function; if not, the frequency of ss-oligo recombination is so high that the desired recombinant can often be found by screening the colony phenotypes (17), by a changed restriction pattern, by colony hybridization, or even by direct sequencing of unselected colonies (our unpublished experiments).

High-efficiency targeting of oligos to a plasmid is possible by use of a Beta-only expression system operating in the absence of mismatch repair (see below). For this method, the plasmid and the oligo are coelectroporated, and frequencies of several percent can be achieved (unpublished data). The expression of Beta function in the absence of Gam may also minimize the complex topology of the plasmid products.

EFFICIENCY OF Red RECOMBINATION SYSTEM

A recombinant frequency of $\geq 10^4/10^8$ viable cells is routinely obtained with PCR products containing drug resistance cassettes (Table 1). The efficiency of gap repair with a plasmid is much lower than that of gene modification by the use of PCR products. A recombination frequency of 10^2 to $10^3/10^8$ viable cells is common for gap repair (48, 89).

The efficiency of recombineering with ss-oligos is higher and depends on which of the two possible complementary strands is used. One strand consistently displays a higher efficiency than the other, and that is the one corresponding to the lagging strand during DNA replication (25, 102). For this reason, we believe that the Beta-bound ss-oligo anneals to single-strand gaps of the lagging strand at the DNA replication fork, effectively mimicking an Okazaki fragment (Color Plate 17). With lagging-strand oligos, the efficiency of repairing a point mutation is approximately $3 \times 10^5/10^8$ viable cells. This efficiency can be increased a hundredfold if a recombineering strain defective in the methyl-directed mismatch repair system (MMR system) is used (17, 49). In the absence of MMR, recombinant frequencies of 1×10^7 to $3 \times 10^7/10^8$ viable cells (up to 30%) are routinely observed for lagging-strand oligos. These extremely high recombination frequencies can also be achieved in MMR-proficient cells by using oligos that, when paired to the target, produce C-C mismatches in the heteroduplex recombination intermediate (Table 2) (17). This is because a C-C mismatch is not corrected by the MMR system in *E. coli* or in other bacterial species in which it has been studied (55). What this means is that a G at any chromosomal position on the lagging strand can be changed to a C with an expectation of high recombination. A C-C mismatch can also prevent the repair of a nearby non-C-C mismatch: recombination frequencies for other mismatches are enhanced if they are within ~6 bp of a C-C mismatch, on either side (N. Costantino, unpublished data). If a leading-strand oligo is used for recombineering, the C-C mismatch efficiencies are 20-fold lower than those obtained with the lagging-strand oligo (17). Since the MMR system acts efficiently only on point mutations or insertions and deletions of 3 nt or less (68), the absence of MMR does not increase the efficiency of recombination with dsPCR substrates containing drug cassettes or other heterologous insertions.

ACCURACY OF THE RECOMBINATION SYSTEM

Accuracy of Targeting with dsDNA

The use of the λ Red system for the replacement of a gene with a drug marker proceeds efficiently when adequate targeting homology is provided at the ends of the drug cassette. Almost without exception, PCR cassettes go to the targeted regions. Yu et al. (98) replaced the *E. coli galK* gene with a *cat* cassette containing 50-bp flanking *galK* homologies and found that of 50 Cm^r colonies, all had become Gal^-, illustrating the efficient insertion of recombining DNA at the desired location. Screening more chloramphenicol-resistant colonies directly on MacConkey galactose indicator agar demonstrated that Cm^r Gal^+ colonies were present at

TABLE 2 Recombination efficiencies of various recombineering systems with ss-oligos and dsDNA

Source of Red	Plasmid origin	Approximate copy no./cell	No. of recombinants/10^8 viable cells			
			galK$^+$ ss-oligos targeted to *E. coli*[a] (creates a C-C mismatch)	*galK*$^+$ ss-oligo background (uninduced) recombination targeted to *E. coli*[a] (creates a C-C mismatch)	*galK* <> *amp* dsDNA targeted to *E. coli*[a]	*galK* <> *amp* dsDNA targeted to *S. enterica* serovar Typhimurium[b]
Standard defective prophage (*N-kil*$^+$)	NA[c]	4[d]	1.5×10^7	1.8×10^3	2.5×10^4	NA[c]
Minimal defective prophage (Δ*N-kil*)	NA[c]	4[d]	9.0×10^6	2.7×10^3	1.2×10^4	NA[c]
pSIM2 (Cm^r),[e] pSIM4 (Amp^r)	pBR322 lacking copy number control	220[f]	1.7×10^7	1.1×10^6	1.5×10^4	ND[c]
pSIM5 (Cm^r),[e] pSIM6 (Amp^r)	pSC101*ts*	16[h]	1.8×10^7	1.5×10^3	2.8×10^4	9.0×10^3
pSIM7 (Cm^r),[e] pSIM8 (Amp^r)	pBBR1	30–40[i]	2.7×10^7	1.5×10^4	2.3×10^4	9.6×10^3
pKD119[j]	pSC101*ts*	16[h]	4.9×10^6	7.3×10^2	2.6×10^3	8.5×10^2
pKD119[k]	pSC101*ts*	16[h]	2.8×10^5	7.3×10^2	1.3×10^1	1.0×10^1

[a]W3110, *E. coli* K-12 IN(*rrnD-rrnE*).
[b]TS616, *S. enterica* serovar Typhimurium LT2 *his-6165 ilv-452 metA22 metE551 trpC malE616*::Tn*10 galE496 xyl-404 rpsL120 fla-66 hsdL6 hsdSA29* from *Salmonella* Genetic Stock Center.
[c]NA, not applicable.
[d]From reference 77.
[e]Minimal prophage on plasmid (see text and Color Plate 19). The data in the table were generated with the Cm^r plasmid.
[f]From reference 50.
[g]ND, not determined.
[h]From reference 33.
[i]From reference 3.
[j]10 mM arabinose.
[k]1 mM arabinose.

a frequency of <0.1%. These rare colonies proved to be diploid for the *gal* operon. They carried a *galK*$^+$ gene as well as a perfectly targeted *galK cat* knockout (98; M. Bubunenko, unpublished data). Duplications of the bacterial chromosome naturally exist in culture (for a review, see reference 52), and because recombineering is so efficient, these diploids are also targeted.

Accuracy of ss-oligo Recombination

Oppenheim et al. (67) performed a careful study of the possible mutagenicity of ss-oligo recombineering, looking at phage λ mutants with an easily scored phenotype, namely, plaque morphology. An ss-oligo was used to convert an allele of the λ *cI* repressor gene that confers a conditionally clear plaque morphology, *cI857*, to the wild-type sequence, thus conferring an unconditionally turbid plaque phenotype. Oppenheim et al. (67) found that some 10% of the yield was recombinant, as ascertained by turbid plaque morphology, but unexpectedly, 0.5 to 2% of the plaques had become unconditionally clear. The *cI* region was sequenced for several of the clear mutant phages; the sequencing revealed that the correction to the wild-type sequence had occurred but that secondary mutations causing the unconditional clear plaque phenotype had also occurred. These mutations were primarily frameshifts in the region included in the oligo. Purification of the oligo by polyacrylamide gel electrophoresis reduced this background of unconditionally clear mutants. This work demonstrated that aberrant oligos in the stock were to blame for most, if not all, of the unsolicited mutations arising during recombineering with ss-oligos.

In a separate control study, an oligo containing a *galK* amber mutation was recombined into a bacterial strain already containing the same *galK* amber mutation. When Gal$^+$ colonies were selected, Gal$^+$ recombinants were found at >1,000-fold above background levels. An oligo that annealed nearby but did not cover the amber mutation did not give rise to Gal$^+$ recombinants. These results are consistent with errors in chemical synthesis of the oligo at the amber codon (Costantino, unpublished data). This type of error is less frequent than the frameshift described above.

EXPRESSION OF Red FUNCTIONS FROM A DEFECTIVE λ PROPHAGE

We now focus on the use of a defective λ prophage for the expression of Red functions. As mentioned earlier, the use of the prophage with its endogenous regulatory elements achieves controlled, coordinate expression of the required genes. The prophage contains the phage immunity region and the main leftward operon under the control of the P_L promoter but is missing the major rightward operon encoding the DNA replication genes, the lysis cassette, and the structural genes. Operationally, this means that after prophage induction, the prophage chromosome cannot excise and the cells will not lyse, nor will phage particles be produced. The *exo*, *bet*, and *gam* genes are clustered in the P_L operon (Color Plate 18) and are expressed after induction of the prophage. The CI repressor directly controls the P_L promoter. A temperature-sensitive repressor mutation, *cI857*, is used so that cells transferred to 42°C are rapidly induced as the repressor is inactivated. This mutant repressor rapidly regains activity upon transfer of the cells to a lower temperature 15 min after heating, so that recombination functions are expressed transiently and then shut off completely. Following removal of the repressor by heat induction, the expression of the *exo*, *bet*, and *gam* genes from P_L is initially prevented by transcriptional terminators. Ultimately, the λ N function, encoded by the first gene in the P_L operon to be expressed after induction, modifies RNA polymerase to prevent transcription termination and to allow expression of the recombination functions (18), thereby coordinately activating all of the genes in the P_L operon. A caveat of the prophage system is that, if the strain to be engineered has an endogenous prophage with a CI-type repressor, this repressor may prevent induction of the recombineering functions. In this situation, the use of an alternative regulatory system (21, 62) is preferable.

The expression of the recombinase functions in the context of the prophage regulatory machinery is preferable to expression from standard cloning vector promoters (Table 2), since the phage regulatory elements are already optimized for the expression of Red. Additionally, having the recombination functions "on" for only a short time and when they are needed prevents undesired DNA recombination and potential rearrangements (59). Under many circumstances, however, one may prefer a plasmid for the recombineering system. Plasmid-based systems have the advantage of mobility and ease of introduction into different bacterial strains and species. S. Datta (unpublished data) recently developed several Red expression vectors by combining a defective λ prophage encoding Red (under the control of an intact *cI* repression system) with various plasmid origins (Color Plates 18 and 19). In this minimal prophage, most of the nonessential region of the P_L operon has been removed, including the toxic *kil* gene, transcription terminators, and the antitermination gene *N* (Color Plate 18). The *rex* genes downstream of the *cI* repressor gene have been replaced by drug cassettes, allowing selection for either chloramphenicol (*cat*) or ampicillin (*bla*). The P_L promoter on these constructs is still regulated by the temperature-sensitive CI857 repressor, with O_L and O_R operators present to ensure tight control (23). With this minimal system, raising the temperature transiently to inactivate the repressor directly induces the Red functions without the intermediate step of N antitermination. A comparison of the prophage systems described here and the pBAD plasmid expression system of Datsenko and Wanner (21) is presented in Table 2.

One of the most promising of the plasmid-prophage systems has a temperature-sensitive pSC101 replication origin (5) (pSIM5 is Cm^r, and pSIM6 is Amp^r). This plasmid catalyzes Red recombination as efficiently as the defective prophage in *E. coli*, and the level of unwanted background recombination is also as low as that of the prophage (Table 2). The same minimal prophage has also been combined with a temperature-sensitive RK2 origin of replication (93) to make a broad-host-range, low-copy-number vector (pSIM9, which is Cm^r) (Table 2). A low copy number is critical for reducing undesirable background recombination: a pUC-based plasmid origin lacking copy number control and carrying the same minimal prophage gives an unacceptably large number of recombinants in the absence of heat induction (Table 2) and places an undesirable metabolic load on the host, as evidenced by the slow growth of cultures carrying these high-copy-number plasmids (pSIM1 and pSIM2). Another broad-host-range plasmid, pSIM7/8, has a pBBR1 origin (3) and an intermediate copy number (Table 2). Modification of the broad-host-range plasmids by the addition of mobilization functions would enable mating between more distantly related species.

POSSIBILITY OF USING RECOMBINEERING WITH OTHER PROKARYOTES

In order to adapt recombineering for use in other prokaryotic systems, we need to consider several issues, as follows. (i) For the use of recombineering in other biological systems, either the Red or RecET system must be adapted for these organisms, or new Red-like recombination systems will have to be developed. (ii) The means of expressing the recombination system must be determined. The system must be well expressed in the organism in which it is to be utilized, yet this activity must exist transiently in the cell to minimize unwelcome genomic plasticity catalyzed by the recombinases. (iii) The recombination functions must be moved into the new host. (iv) The organism to be engineered must have been (at least partially) sequenced so that targeting oligonucleotides can be designed. (v) The DNA substrate of the recombineering reaction must be introduced at high levels into those cells. (vi) Methods for selecting or screening recombinants must be available. Several of these issues are discussed further below.

Choice of Recombination System

Whether λ Red will be able to catalyze recombineering in an organism of interest may depend on the organism's relatedness to *E. coli*.

For Red to act, the recombination functions must be expressed and, following expression, must interact appropriately with necessary host functions. The more distantly related an organism is to *E. coli*, the more likely that differences in housekeeping functions will lead to poor utilization of the λ Red system. For example, the recombination enzymes must interact appropriately with the DNA replication machinery. DNA modifications such as methylation may also differ from those present in *E. coli*.

Recombineering is now routinely used in several laboratories, and the functionality of the Red system has already been demonstrated for *Salmonella* (for examples, see references 10 and 92), *Yersinia* (22), and pathogenic *E. coli* (59). We expect that the Red or RecET system will be operative in other bacteria related to *E. coli*, including *Shigella* and *Vibrio*, as well as in *Pseudomonas* and *Streptomyces*. For distantly related bacteria, the use of a different recombination system may be preferable. Many dsDNA viruses other than λ encode Red-like two-component recombinases (54, 64, 73, 96) that are likely to catalyze efficient and accurate recombination in their particular host. Genomic analysis has demonstrated that the recombination functions are organized into gene modules, with the Red-like genes adjacent to each other and their general location conserved among phages (35). Phage functions that have coevolved with host bacteria are optimized to maintain protein-specific interactions with the host and are most likely to facilitate high-efficiency and high-fidelity recombination. As stated previously, the efficient coordination of exonuclease activity with single-strand annealing activity appears to be critical, and for optimal recombineering the nuclease and single-strand annealing protein pairs should probably be evolutionary partners (62).

Currently, there are multiple groups of single-strand annealing proteins and exonucleases, which have been identified by BLAST searches, structural analyses, and enzymology (37, 64, 69). The single-strand annealing proteins include Beta and RecT from λ and Rac, respectively, the Erf proteins of P22 and HK97 (37, 71), ICP8 from herpes simplex virus type 1 (73), and LEF-3 from baculovirus (54). The exonucleases include λ Exo and RecE as well as the alkaline nucleases of herpes simplex virus type 1 (73) and baculovirus (54). Single-strand annealing proteins like Beta and RecT are found in many bacteria, including *Borrelia*, *Listeria*, and *Streptococcus*; the genes for these proteins appear to have originally evolved in DNA bacteriophages (37). Exonucleases similar to RecE are present in many bacteria, including *Bacillus*, *Streptococcus*, *Lactococcus*, and *Listeria* and their phages. Employing endogenous functions, e.g., using the *Bacillus* phage SPP1 Chu/gp35 Red-like recombinases (6, 96) to engineer *Bacillus* bacterial cells, will undoubtedly provide advantages over the use of Red Exo-Beta or RecET.

Just as protein-specific interactions between recombinase and replisome components may be required for efficient recombination, specific interactions between anti-host nuclease proteins such as λ Gam and the host nucleases may also be important. The λ Gam protein specifically inhibits the *E. coli* RecBCD and SbcCD nucleases, and a BLAST analysis suggested that Gam is not as well conserved as the Exo and Beta proteins, as it is present only in the prophages of some pathogenic strains of *E. coli*, *Shigella*, and *Salmonella*. While the RecBCD-like ExoV enzymes are represented in many bacterial species, other bacteria often contain a two-subunit form of this enzyme, e.g., the AddAB complex of *Bacillus subtilis* (1) and the RexAB complexes of *Lactococcus lactis* (24) and *Streptococcus pneumoniae* (32). Given the widespread distribution of RecBCD-like activity in bacteria, there may be functional analogs of Gam in other phages that are different enough in sequence that BLAST fails to identify them. As suggested earlier, a Gam gene is likely to be located adjacent to the Exo- and Beta-like genes.

Some phages have proteins that protect linear DNA from degradation (75), and some of these may, like λ Gam, protect dsDNA while still allowing its participation in recombination reactions. Not all phage-encoded nuclease inhibitors will be useful in this context, since some (Mu Gam and T4 gp2) act by binding DNA

ends and protecting them from degradation but also make that DNA unavailable to participate in recombination (87) or cause nonviability of the host (4). Other potentially useful inhibitors of RecBCD-like enzymes are still uncharacterized.

Since recombineering with linear dsDNA has the potential problem of nuclease-mediated DNA degradation, engineering with ssDNA holds the most promise for the initial development of recombineering in any organism. Genetic engineering using ss-oligos requires only a Beta-like single-strand annealing protein and is less likely to be inhibited by cellular nucleases. The use of a Beta-like protein that is closely related to the host will optimize the chances for success since the phage synaptase presumably interacts with other cellular proteins. A phage single-strand annealing protein expressed in a distantly related organism may fail to catalyze recombination only because it does not fit properly with the endogenous cellular machinery.

Choice of Expression System

Expression of the recombination genes at high levels for a brief period of time gives optimal efficiencies (98) and is a preferred method (59). The temperature-inducible prophage allows rapid, efficient induction of Red functions simply by raising the temperature for time periods as short as 10 min. For other expression systems, exogenous inducers such as IPTG (isopropyl-β-D-thiogalactopyranoside) (58, 59, 60) and arabinose (21, 59, 101) can be added to activate gene expression; these agents will be removed when the cells are washed during preparation for electroporation. For such inducers, the optimal concentration and length of induction will need to be determined individually. Note that, as shown in Table 1, for an Ara^+ host, the induction of Red with 10 mM arabinose gives considerably higher recombination levels, although not further optimized, than the 1 mM concentration used by Datsenko and Wanner (21) with their Ara^- strain. Many plasmid expression systems are leaky, and recombination functions may catalyze undesired DNA rearrangements when expressed for extended periods. The choice of expression system will largely be determined by which regulatory elements operate in the host under consideration.

Moving the Recombination System into Other Strains

After the recombination and expression functions are selected, the complete system, including the genes for recombination functions and their regulatory elements, must be introduced into the desired strain, either on the host chromosome or on a stable episome. The functional limitations imposed by transcriptional and translational regulatory elements will operate. For example, the λ Red genes are unlikely to function optimally, if at all, in gram-positive bacteria, while *Bacillus* phage SPP1 Chu/gp35, which functions under the control of its own phage regulatory elements, or a tightly regulated *Bacillus* expression system is more likely to work well.

There are advantages to both chromosomal and plasmid locations for the recombination functions. A chromosomal construct yields recombineering strains that are genetically stable, do not require drug selection, and are less likely to cause toxicity upon induction. Locating the recombination genes on the chromosome simplifies recombineering of a plasmid and avoids the problems of plasmid incompatibility.

Although the bacterial chromosome may be the ideal location for recombinase genes, plasmid-carried alternatives will work well in many cases. If a plasmid expression system is chosen, broad-host-range vectors that are able to shuttle between *E. coli* and other bacteria can be used to move Red into closely related species. Bear in mind that some plasmids containing a *gam*-like gene may become unstable unless they are tightly regulated since plasmids may undergo recombination-dependent rolling-circle replication. Such replication may result in possible gene rearrangements and plasmid loss, which may lead to toxicity if drug selection is applied. Mobile fertility factors are also viable candidates for the introduction of the recombination functions. Such factors have more in

common with the chromosome than with small plasmids in terms of their stability and copy number.

A theoretical alternative, which is still undeveloped for recombineering, is to coelectroporate purified recombinase proteins and the DNA substrate, thus eliminating the need for any genetic construct. Initial attempts with purified Beta and ssDNA complexes have not been successful in *E. coli* (H. Ellis, D. Court, and K. Murphy, unpublished experiments). This technology has, however, been developed for insertional transposon mutagenesis in several bacterial species as well as in *Saccharomyces cerevisiae* (29).

Introduction of Substrate DNA into Bacterial Host

Electroporation is the method of choice for introducing substrate DNAs (PCR products and ss-oligos) into bacterial cells. For *E. coli*, electroporation is extremely efficient: we routinely introduce ss-oligos and short dsDNAs into at least 25% of viable cells, as evidenced by recombinant frequencies (17). We have found that, for *E. coli* with Red functions induced from the prophage, ~100 ng of PCR-generated dsDNA is saturating (98). Electroporation of *Salmonella* is reported to be as efficient as that of *E. coli* (76), and our recombineering results with TS616, a strain lacking endogenous restriction systems, largely support this observation (Table 1); similar numbers of recombinants were obtained in a wild-type LT2 strain with all of its restriction systems intact (S. Datta, unpublished data). For some bacteria, however, electroporation will continue to be a rate-limiting step, for conventional cloning methods as well as for use in recombineering. Perhaps this step will eventually be circumvented by the development of alternative methods of introducing DNA substrates, such as mating and transduction with phages. Bacteria that are naturally competent at some stages of their life cycle may take up DNA efficiently enough for recombineering.

Once the donor DNA is introduced, it may be subject to restriction by endogenous nucleases, so if mutations in host restriction-modification systems are available, they can be utilized and may increase the yield of successfully transformed cells. Interestingly, we have not seen any detrimental effects from active restriction systems in our experiments, even though the chloramphenicol acetyltransferase drug cassette that we use as a PCR-generated dsDNA substrate contains an EcoK restriction site (98). Possibly, the dsDNAs that we introduce are too short to be good substrates for restriction; alternatively, the concentration of introduced DNA may be such that the restriction enzymes are titrated.

Some organisms may never yield to recombineering technology. For these organisms, the relevant segment of DNA can be cloned into BACs or PACs and then moved into *E. coli* for manipulation, as has been done for eukaryotic systems (16). Viruses from distantly related organisms can also be cloned intact into BACs in *E. coli*, modified, and then returned to their native organism (97). The development of recombineering for other organisms is likely to proceed by trial and error as various schemes are attempted. Some things may not work on the first try, and the reasons may not be obvious, but the benefits of successful recombineering systems will make efforts toward optimization time well spent.

CONCLUSIONS

New genomic technologies are allowing the characterization of the microbial world on a level that has never before been possible. As of August 2004, more than 100 complete bacterial genome sequences were in the Entrez Genome NCBI database; the number is over 1,000 for viral genomes, including more than 100 bacteriophages. Analyses of these organisms, with their anticipated similarities and diversities in their habitats and behaviors, will provide a wealth of new information about the microbial world. It is anticipated that many useful applications for humans will arise from the study of these recently sequenced microorganisms. As discussed by Toussaint et al. (90), the experimental testing of predictions arising from bioinformatics is likely to be the rate-limiting step in these analyses for the foreseeable future.

Recombineering, or in vivo genetic engineering using homologous recombination, has made it possible to efficiently, rapidly, and precisely generate genetic constructs in the *E. coli* chromosome, its plasmids, and its phages. Recombineering also holds great promise as a tool for analyses of the many less-well-characterized bacteria. Catalyzed by the same lambdoid prophage functions, recombineering is now being utilized in gram-negative bacteria that are closely related to *E. coli*. Recombineering in more distantly related bacteria will be facilitated by the identification and characterization of Red-like functions in phages from those organisms. Other requirements for the ready use of recombineering are those in common with other modes of genetic analysis, such as the development of systems permitting regulated gene expression and means of easily introducing DNA into the organism under study.

REFERENCES

1. **Alonso, J. C., G. Luder, and T. A. Trautner.** 1992. Intramolecular homologous recombination in *Bacillus subtilis* 168. *Mol. Gen. Genet.* **236:**60–64.
2. **Anderson, D. G., and S. C. Kowalczykowski.** 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell* **90:**77–86.
3. **Antoine, R., and C. Locht.** 1992. Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from gram-positive organisms. *Mol. Microbiol.* **6:**1785–1799.
4. **Appasani, K., D. S. Thaler, and E. B. Goldberg.** 1999. Bacteriophage T4 gp2 interferes with cell viability and with bacteriophage lambda Red recombination. *J. Bacteriol.* **181:**1352–1355.
5. **Armstrong, K. A., R. Acosta, E. Ledner, Y. Machida, M. Pancotto, M. McCormick, H. Ohtsubo, and E. A. Ohtsubo.** 1984. A 37 × 10^3 molecular weight plasmid-encoded protein is required for replication and copy number control in the plasmid pSC101 and its temperature-sensitive derivative pHS1. *J. Mol. Biol.* **175:**331–348.
6. **Ayora, S., R. Missich, P. Mesa, R. Lurz, S. Yang, E. H. Egelman, and J. C. Alonso.** 2002. Homologous-pairing activity of the *Bacillus subtilis* bacteriophage SPP1 replication protein G35P. *J. Biol. Chem.* **277:**35969–35979.
7. **Barbour, S. D., H. Nagaishi, A. Templin, and A. J. Clark.** 1970. Biochemical and genetic studies of recombination proficiency in *Escherichia coli*. II. Rec+ revertants caused by indirect suppression of *rec*⁻ mutations. *Proc. Natl. Acad. Sci. USA* **67:**128–135.
8. **Biek, D. P., and S. N. Cohen.** 1986. Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *J. Bacteriol.* **167:**594–603.
9. **Brooks, K., and A. J. Clark.** 1967. Behavior of λ bacteriophage in a recombination-deficient strain of *Escherichia coli*. *J. Virol.* **1:**283–293.
10. **Bunny, K., J. Liu, and J. Roth.** 2002. Phenotypes of *lexA* mutations in *Salmonella enterica*: evidence for a lethal *lexA* null phenotype due to the Fels-2 prophage. *J. Bacteriol.* **184:**6235–6249.
11. **Carter, D. M., and C. M. Radding.** 1971. The role of exonuclease and β protein of phage λ in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *J. Biol. Chem.* **246:**2502–2512.
12. **Cassuto, E., T. Lash, K. S. Sriprakash, and C. M. Radding.** 1971. Role of exonuclease and β protein of phage λ in genetic recombination. V. Recombination of λ DNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **68:**1639–1643.
13. **Cassuto, E., and C. M. Radding.** 1971. Mechanism for the action of λ exonuclease in genetic recombination. *Nat. New Biol.* **229:**13–16.
14. **Chalker, A. F., D. R. Leach, and R. G. Lloyd.** 1988. *Escherichia coli sbcC* mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* **71:**201–205.
15. **Cohen, A., and A. J. Clark.** 1986. Synthesis of linear plasmid multimers in *Escherichia coli* K-12. *J. Bacteriol.* **167:**327–335.
16. **Copeland, N. G., N. A. Jenkins, and D. L. Court.** 2001. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2:**769–779.
17. **Costantino, N., and D. L. Court.** 2003. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **100:**15748–15753.
18. **Court, D. L., and A. B. Oppenheim.** 1983. Phage lambda's accessory genes, p. 251–277. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), *Lambda II*. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
19. **Court, D. L., J. A. Sawitzke, and L. C. Thomason.** 2002. Genetic engineering using homologous recombination. *Ann. Rev. Genet.* **36:**361–388.
20. **Cromie, G. A., C. B. Millar, K. H. Schmidt, and D. R. Leach.** 2000. Palindromes as substrates for multiple pathways of recombination in *Escherichia coli*. *Genetics* **154:**513–522.
21. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Esche-*