

# DNA Cloning Using In Vitro Site-Specific Recombination

James L. Hartley, Gary F. Temple, and Michael A. Brasch<sup>1</sup>

Life Technologies, Inc., Rockville, Maryland 20850, USA

As a result of numerous genome sequencing projects, large numbers of candidate open reading frames are being identified, many of which have no known function. Analysis of these genes typically involves the transfer of DNA segments into a variety of vector backgrounds for protein expression and functional analysis. We describe a method called recombinational cloning that uses in vitro site-specific recombination to accomplish the directional cloning of PCR products and the subsequent automatic subcloning of the DNA segment into new vector backbones at high efficiency. Numerous DNA segments can be transferred in parallel into many different vector backgrounds, providing an approach to high-throughput, in-depth functional analysis of genes and rapid optimization of protein expression. The resulting subclones maintain orientation and reading frame register, allowing amino- and carboxy-terminal translation fusions to be generated. In this paper, we outline the concepts of this approach and provide several examples that highlight some of its potential.

The functional analysis of genes and their coding sequences (open reading frames [ORFs]) typically requires that each ORF be expressed, the encoded protein purified, antibodies produced, phenotypes examined, intracellular localization determined, and interactions with other proteins sought. Each step of characterization requires subcloning into one or more specialized vectors that impart particular functional properties to the cloned segment. The same is true for medical application of genes, for example, in gene therapy or genetic immunization. When characterizing multiple genes, such as those encoding members of intracellular pathways, or candidates from functional screens (e.g., two-hybrid, phage display, expression cloning), the subcloning required can present a significant barrier to progress. The need for more efficient cloning/subcloning methods is keenly apparent when considered in the context of the hundreds of thousands of genes predicted from ongoing genome projects.

Several approaches have been described that facilitate the cloning process. Examples that take advantage of homologous recombination in *Escherichia coli* (Bubeck et al. 1993; Oliner et al. 1993; Degryse 1996; Zhang et al. 1998) or yeast (Lafontaine and Tollervey 1996; Storck et al. 1996), site-specific transposition (Luckow et al. 1993), or site-specific recombination (Peakman et al. 1992; Boyd 1993; Liu et al. 1998) have been published. These have significant value for particular applications but are limited in scope by requirements for specific hosts, by selection schemes, or by the vector attributes they can effectively contribute. We therefore sought to develop a flexible approach

that could provide high-efficiency, high-fidelity cloning and subcloning reactions, independent of vector function or host background. Here, we describe such a system and summarize its application to a number of genomics projects.

## RESULTS

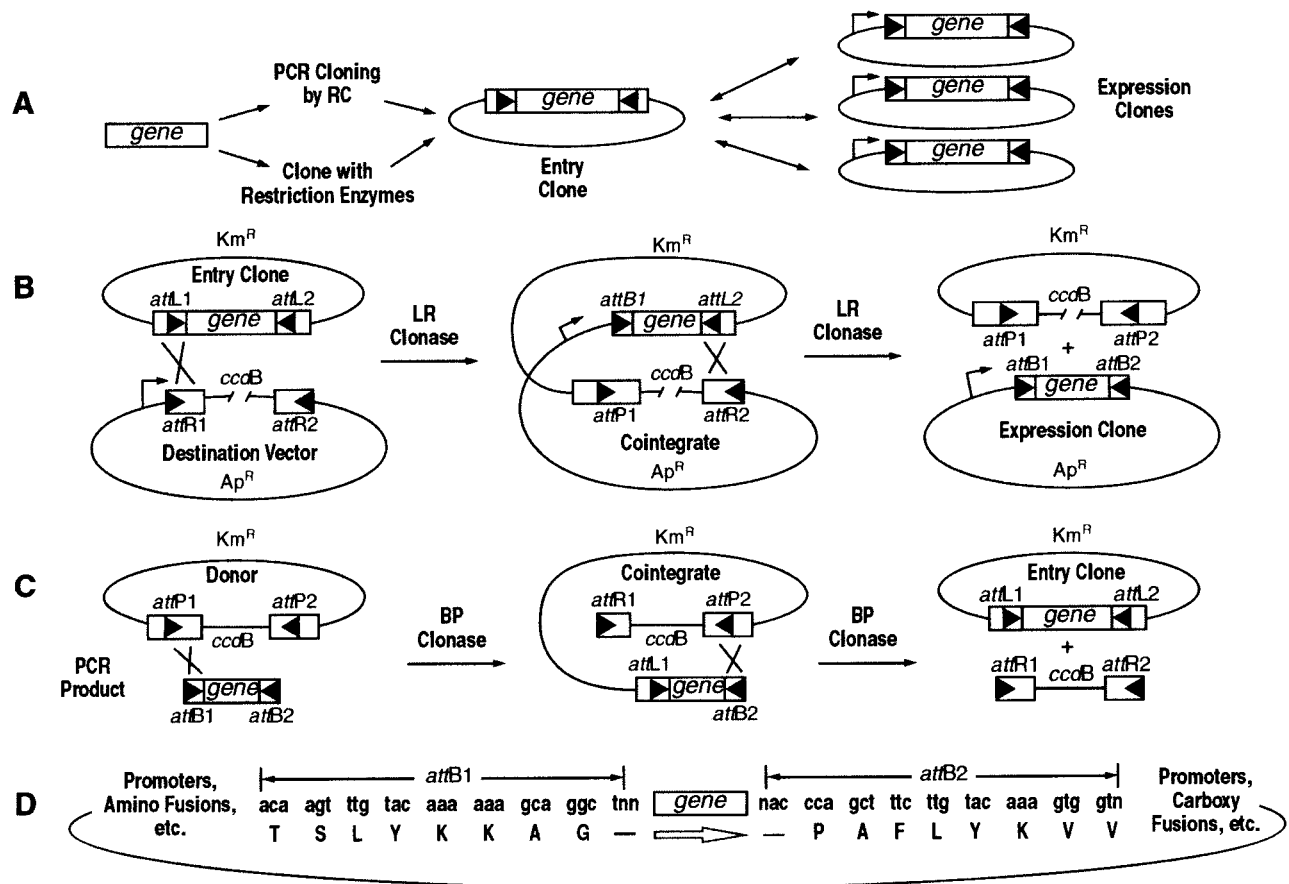
### Recombinational Cloning

The site-specific recombination reactions mediated by the  $\lambda$  integrase family of recombinases are conservative (no net gain or loss of nucleotides) and highly specific (Landy 1989). We designed an approach (Fig. 1A–C) whereby DNA segments (e.g., genes) flanked by recombination sites can be mixed in vitro with a new vector also containing recombination sites and incubated with bacteriophage  $\lambda$  integrase recombination proteins to accomplish the transfer of the gene into the new vector. We refer to this process as recombinational cloning (RC). Our experiments initially used both the bacteriophage  $\lambda$  system and the Cre/loxP system (Abremski and Hoess 1984; Gopaul et al. 1999). However, the bacteriophage  $\lambda$  system proved more suitable. This system carries out two reactions: (1)  $attB \times attP \rightarrow attL + attR$  mediated by the integrase (Int) and integration host factor (IHF) proteins and (2)  $attL \times attR \rightarrow attB + attP$  mediated by Int, IHF, and excisionase (Xis). Thus, the direction of the reactions is controlled by providing different combinations of proteins and sites. The mutant  $attB$  recombination sites ( $attB1$  and  $attB2$ ; 25 bp) we developed are shown in Figure 1D. Note that  $attB1$  will recombine with  $attP1$  but not  $attP2$ , thereby maintaining orientation of the DNA segment during recombination. Aberrant recombination events have not been identified in hundreds

<sup>1</sup>Corresponding author.

E-MAIL [jhartley@lifetech.com](mailto:jhartley@lifetech.com); FAX (301) 610-8371.

Article and publication are at [www.genome.org/cgi/doi/10.1101/gr.143000](http://www.genome.org/cgi/doi/10.1101/gr.143000).



**Figure 1** (A) Schematic of gene cloning and transfer by recombinational cloning. The triangles represent recombination sites. Genes are cloned into Entry Vectors by in vitro recombination of PCR products or with restriction enzymes and ligase. Thereafter, genes are moved entirely by recombination. (B) Subcloning of a gene from the Entry Vector into a Destination Vector. (C) Cloning attB-PCR products by in vitro recombination. (D) Sequences of the attB1 and attB2 sites that flank the gene either in a PCR product made with primers containing attB sites or in an Expression Clone.

of sequenced RC clones (J. LaBaer, pers. comm.; M. Vidal, pers. comm.; S. Wiemann, pers. comm.).

The in vitro recombination reaction initially contains two starting DNAs: an Entry Clone (*attL1*-gene-*attL2*), which carries the DNA segment to be transferred, and a Destination Vector (*attR1*-*ccdB*-*attR2*), the vector into which the DNA will be subcloned (Fig. 1B). Incubating these DNAs with recombination proteins (Int + Xis + IHF) results in Int-mediated recombination, first generating a cointegrate molecule and then resolving it, to accomplish transfer of the cloned DNA segment into the Destination Vector. To obtain only the Expression Clone following introduction of the mixture into *E. coli* by transformation, we imposed two selection schemes. First, the Entry Clone (kanamycin resistant [Km<sup>R</sup>]) and the Destination Vector (ampicillin resistant [Ap<sup>R</sup>]) contain different antibiotic resistance genes. Second, the Destination Vector contains a selection marker, the F-plasmid-encoded *ccdB* (Bernard and Couturier 1992; Miki et al. 1992) gene, which inhibits growth of *E. coli*. Hence, as shown in Figure 1B, trans-

formants selected for Ap<sup>R</sup> will contain only the Expression Clone (the Destination Vector in which the *ccdB* gene is replaced by the DNA segment of interest flanked by the small attB sites). Background colonies contain inactive or deleted *ccdB* genes.

#### Parallel Subcloning of a Test Gene into Multiple Destination Vectors

As an initial demonstration of RC, 12 diverse cloning vectors were converted to Destination Vectors by insertion of a blunt-end cassette comprising *attR1*-chloramphenicol resistance gene (Cm<sup>R</sup>)-*ccdB*-*attR2* (Table 1). Potentially any vector can be similarly converted. To propagate vectors that contain the *ccdB* gene, we isolated an *E. coli* strain (DB3.1) containing a *gyrA462* mutation that provides resistance to the effects of *ccdB*. An Entry Clone (~50 ng) containing the chloramphenicol acetyl transferase (CAT) gene flanked by attL sites was separately mixed with each Destination Vector (~50ng) and LR Clonase (a mixture of Int, IHF, and Xis) and incubated for 30 min. An aliquot

**Table 1.** Parallel Transfer of the Chloramphenicol Acetyl Transferase (CAT) Gene into 12 Destination Vectors

Function of Destination Vector <sup>1</sup>	Colonies <sup>2</sup>	Control <sup>3</sup>	Correct <sup>4</sup>
Expression of native protein in <i>E. coli</i>	15,000	0	6/6
His6-fusion protein in <i>E. coli</i>	10,650	0	6/6
GST-fusion protein in <i>E. coli</i>	9200	0	6/6
Thioredoxin-fusion protein in <i>E. coli</i>	11,000	0	6/6
Sequencing, probe synthesis, (+) strand	13,950	0	6/6
Sequencing, probe synthesis, (–) strand	8950	0	6/6
Transient expression of native protein in mammalian cells	7950	0	6/6
Stable expression of native protein in mammalian cells	10,500	1100	6/6
Expression of native protein in insect cells	7800	15	6/6
Native protein in mammalian cells. Semliki Forest Virus vector	4150	0	6/6
His6-fusion protein in insect cells	6350	30	6/6
Tetracycline-regulated expression of native protein in mammalian cells	11,650	0	6/6

<sup>1</sup>Standard vectors were converted to Destination Vectors by inserting a cassette containing the motifs *attR1*–*Cm<sup>R</sup>–ccdB–attR2*.

<sup>2</sup>Aliquots (20 or 200  $\mu$ L) of the 1 ml expression mixture were plated on ampicillin. Colonies are expressed as the number calculated for mL. Cells were competent at  $1 \times 10^8/\mu$ g pUC.

<sup>3</sup>Identical reactions except the CAT Entry Clone DNA was omitted.

<sup>4</sup>Based on miniprep DNAs (four colonies) and restriction digests (two colonies) from random colonies.

(1/10 reaction) was then introduced into *E. coli* DH5 $\alpha$  by transformation with selection for colonies containing Ap<sup>R</sup> Expression Clones (Fig. 1B). Each reaction generated thousands of transformants. Negative control reactions that lacked CAT Entry Clone DNA gave 200- to 15,000-fold fewer colonies (except CMVneo), indicating that the *ccdB* gene was effectively inhibiting transformation by unreacted Destination Vectors and, therefore, that most Ap<sup>R</sup> colonies contained the desired Expression Clone. The CMVneo Destination Vector was unstable in *E. coli* and gave a background of ~10%; recently constructed versions show backgrounds similar to other Destination Vectors. All colonies tested contained a plasmid of expected size (four tested per reaction; 48/48), and all had the predicted restriction pattern (two minipreps per reaction, 24/24; data not shown). These results demonstrate parallel transfer of a test gene into multiple Destination Vectors in an efficient and orientation-specific manner with minimal elapsed and hands-on time.

### RC-Mediated Cloning of PCR Products

PCR products flanked by *attB* sites can be generated by incorporating *attB* sites (25 base + 4 G residues) at the 5' end of PCR primers: *attB1* in the forward primer and *attB2* in the reverse primer. These *attB*-flanked PCR products can be cloned by incubating with *attP*-containing vectors in the presence of Int and IHF (BP Clonase) to generate Entry Clones (*attL1*-PCR product-*attL2*; Fig. 1C,D). Such clones retain the orientation and reading frame specified by their starting *attB* sites.

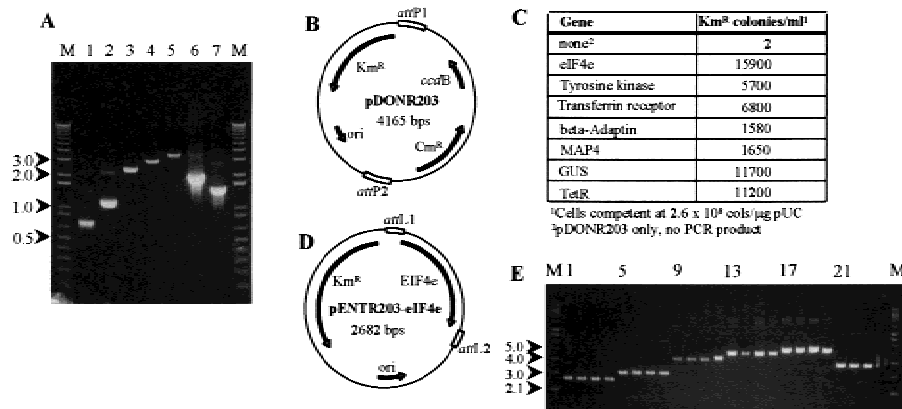
Five human genes (Table 2) were amplified from first-strand cDNA synthesized using total HeLa RNA as a template (Fig. 2A). In addition, the *E. coli gus* ( $\beta$ -glucuronidase) and tetracycline resistance (*tetR*) genes

**Table 2.** Sources of Genes

Gene	Accession No.	Open Reading Frame size	Protein size
Hu eIF4e	M15353	0.65 kb	25 kD
Hu tyrosine kinase	U02680	1.1 kb	39 kD
Hu transferrin receptor	X01060	2.3 kb	84 kD
Hu $\beta$ -adaptin	M34175	2.3 kb	105 kD
Hu MAP4	M64571	3.3 kb	121 kD
<i>E. coli gus</i>	P05804	1.8 kb	68 kD
<i>E. coli tetR</i>	J01749	1.4 kb	NA

were amplified from plasmid clones. All primers contained *attB* sites. The *tetR* amplicon contained the natural promoter, ribosome binding site, and stop codon, and thus conferred tetracycline resistance. Amplification of the other six genes began at their methionine start codons and extended to the carboxy-terminal stop codons. Both prokaryotic and eukaryotic translation sequences (Methods) were added between the *attB1* site and ATG of the forward PCR primers (except *tetR*) to allow expression as native protein or amino-terminal fusions in *E. coli* or eukaryotic cells following transfer from an Entry Clone into the appropriate Destination Vector.

Equal volumes (2  $\mu$ L) of each PCR reaction (40–1000 ng product), 300 ng of *attP* vector (pDONR203, Fig. 2B), and BP Clonase were mixed and incubated for 1 hr, after which an aliquot (1/10 reaction) was introduced into *E. coli* DH5 $\alpha$  cells by transformation, with selection for Km<sup>R</sup> (Fig. 2C). The map of one representative clone, pENTR203-eIF4e, is shown in Figure 2D. Transformants resulting from RC-mediated cloning of the *tetR* PCR product were found to be tetracycline re-



**Figure 2** (A) Agarose gel (1%) of PCR amplification products (2.5  $\mu$ L of 35  $\mu$ L) using *attB*-primers (see Methods) encoding *elf4e* (lane 1), tyrosine kinase (lane 2), transferrin receptor (lane 3),  $\beta$ -adaptin (lane 4), MAP4 (lane 5), glucuronidase (*Gus*, lane 6), or the tetracycline-resistance gene (*tetR*; lane 7). (B) *attP* cloning vector pDONR203. PCR products cloned by in vitro recombination replace the chloramphenicol-resistance and *ccdB* genes; the recombination reactions convert the *attP* sites to *attL* sites. (C) Colonies resulting from transformation of *E. coli* with reactions (2  $\mu$ L of 22  $\mu$ L) containing the PCR products, pDONR203, and BP Clonase. The negative control contained all components except PCR product. (D) Entry Clone pENTR203-*elf4e*, the product of recombinational cloning of the *elf4e* PCR product into pDONR203. (E) Miniprep DNA (Entry Clones) from the colonies in C. Lanes 1–4, *elf4e*; lanes 5–8, tyrosine kinase; lanes 9–12, transferrin receptor; lanes 13–16,  $\beta$ -adaptin; lanes 17–20, MAP4; lanes 21–24, *Gus*; M, supercoiled DNA ladder.

sistant at high efficiency (195 of 197). Similarly, four colonies from each of the other six cloning reactions (Fig. 2E) showed all 24 of the plasmids with the expected sizes. This approach therefore provides high efficiency cloning of numerous PCR products in parallel with only a few hours of laboratory time. The resulting plasmids contain the amplified DNA in an Entry Clone capable of transferring the insert into any number of Destination Vectors. (We have also constructed vectors that contain multiple cloning sites flanked by *attL1* and *attL2* sites, allowing Entry Clones to be constructed by standard restriction enzyme cloning.)

### RC Allows Rapid Optimization of Protein Expression

Numerous systems exist for expressing proteins in a range of organisms (e.g., *E. coli*, yeast, insect, or mammalian cells), from various promoters, or as native proteins or proteins linked to fusion tags. At the outset, for any given gene it is typically unknown which system will provide sufficient expression levels to allow purification. RC can be used to generate constructs and simultaneously to test multiple fusion tags and/or expression in several hosts. As an example, DNA from one colony of each Entry Clone constructed above (Fig. 2E, lanes 1,5,9,13,17) was used to transfer the five cloned human genes into Destination Vectors for expression of amino-terminal His6 fusion proteins in *E. coli* (pDEST17, which contains a phage T7 promoter) and expression of native proteins in insect cells (pDEST8). The *gus* gene (Fig. 2E, lane 21) was also transferred into these vectors and into baculovirus vectors

that expressed amino-terminal His6 (pDEST10) or GST (pDEST20) fusion proteins in insect cells.

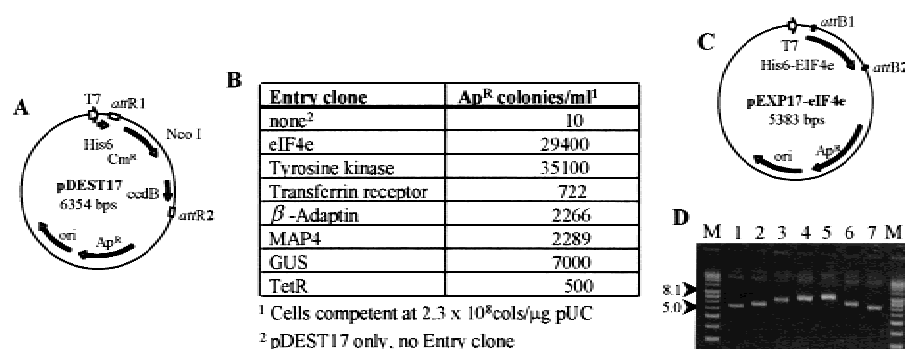
Approximately 200 ng of miniprep Entry Clone DNA (except the transferrin receptor gene, 60 ng, and the *tetR* gene, 40 ng) was mixed with ~300 ng of Destination Vector and LR Clonase and then incubated 1 hr. An aliquot (1/10) of each reaction was introduced into *E. coli* DH5 $\alpha$  by transformation with selection for transformants containing Ap<sup>R</sup> Expression Clones. The data for each cloning into pDEST17 are shown in Figure 3. Transformants from the *tetR* Entry Clone reaction showed 96 of 102 resistant to tetracycline. Based on this high percentage of desired clones, and on the low number of background

colonies, one random colony was examined from each cloning and was found to contain an Expression Clone of the expected size (Fig. 3D). In total, 16 clonings using seven different genes and five different Destination Vectors generated the desired Expression Clones with minimal effort.

Protein expression for each of the pDEST17 derivatives (except for *tetR*) was examined in *E. coli* strain BL21SI, which expresses T7 RNA polymerase under the control of a salt-inducible promoter (Bhandari and Gowrishankar 1997). Cultures were induced by addition of NaCl, and total cell extracts were applied to an SDS-PAGE gel (Fig. 4A). Fusion protein was observed for His6-*elf4e*, His6- tyrosine kinase, and His6-GUS (lanes 1,2,6), showing that fusions across the *attB1* site expressed efficiently. His6-MAP4 gave relatively weak expression (lane 5), whereas no expression was observed for  $\beta$ -adaptin (lane 4) or transferrin receptor (lane 3; cells grew very slowly). The transferrin receptor,  $\beta$ -adaptin, MAP4, and *gus* genes were also transferred into Destination Vectors for amino-terminal fusions with GST (pDEST15) or thioredoxin (pDEST16) in *E. coli*. Although the *gus* gene showed expression levels in both systems similar to that observed for His6-GUS, expression of the human genes was not observed (data not shown).

Each of the baculovirus Expression Clones generated above was introduced into SF9 insect cells, and extracts were examined for protein expression. In contrast to the results in *E. coli*, fair to good expression was observed for tyrosine kinase,  $\beta$ -adaptin, and MAP4 pro-





**Figure 3** (A) Destination Vector pDEST17, for expressing His6 fusions in *Escherichia coli*. (B) Colonies resulting from transformation of DH5 $\alpha$  cells with 2  $\mu$ L (of 22  $\mu$ L) of LR reactions. The negative control contained all components except the Entry Clone DNA. pDEST17 was linearized at the unique NcoI site prior to mixing with Entry Clone DNA. (C) The Expression (attB) clone pEXP17-eIF4e, which resulted from the LR reaction with pENTR203-eIF4e and pDEST17. The subcloned eIF4e gene has replaced the chloramphenicol resistance-*ccdB* segment, with the amino end of the gene downstream of the T7 promoter and the His6 tag in frame with the eIF4e open reading frame. The recombination reactions convert the *attR* sites (25 bp). (D) Miniprep DNA of single colonies (Expression Clones) from each reaction. Lane 1, eIF4e; lane 2, tyrosine kinase; lane 3, transferrin receptor; lane 4,  $\beta$ -adaptin; lane 5, MAP4; lane 6, Gus; lane 7; TetR. M, supercoiled DNA ladder.

teins, but not for eIF4e (Fig. 4B). A small amount of protein of the expected size of the transferrin receptor was also seen. Expression of the *gus* gene as native His6- and GST-amino fusions was extensive (lanes 6–8), further demonstrating translation across the *attB1* site (and, for native protein, the embedded ribosome binding sites; Table 1). We further transferred the *gus* gene into a Destination Vector containing a CMV promoter (pDEST12.2) and assayed glucuronidase activity in COS7 and CHO cells. Cells from both transfections stained intensely, indicating expression of functional GUS protein (data not shown).

## DISCUSSION

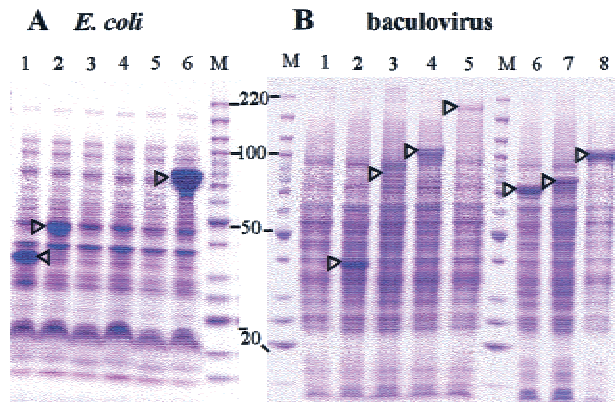
RC uses in vitro site-specific recombination to exchange DNA segments flanked by recombination sites between two parental molecules. A first, intermolecular recombination forms a cointegrate molecule, which then resolves into two daughter molecules by a second intramolecular reaction. The desired clone is obtained from this mixture of molecules by imposing antibiotic resistance selection for the desired construct and a selection (encoded by the *ccdB* gene) against starting molecules and intermediates (Fig. 1). Performing these reactions in vitro eliminates problems of plasmid segregation inherent with in vivo recombination schemes. The use of the  $\lambda$  site-specific recombination system allows engineering of recombination sites to provide high specificity (*attB1* reacts with *attP1* but not *attP2*, etc.) and activity, thereby maintaining orientation of the transferred DNA segment and yielding a high proportion of desired clones. Use of the  $\lambda$  system also provides control over reaction directionality, because different combinations of proteins and binding

sites mediate the *attB*  $\times$  *attP* reaction and the *attL*  $\times$  *attR* reaction. This feature helps to maximize the amount of starting molecules that can be driven to product (without competing reverse reactions that regenerate starting molecules), a critical issue when attempting to transfer complex mixtures of clones (e.g., cDNA libraries) between vectors. In contrast to *loxP* sites (Liu et al. 1998), *attB* sites have no secondary structure to interfere with protein expression or DNA sequencing. Because no net synthesis or loss of DNA occurs during DNA segment transfer, reading frame register is always maintained. This allows the faithful transfer of

ORFs from Entry Clones into Destination Vectors that provide amino-terminal and carboxy-terminal translation fusions. Moreover, because transfer by RC does not rely on a replicative step (e.g., PCR-based strategies), sequence alterations to the subcloned DNA segment are not expected. Collectively, these design considerations provide RC with substantial flexibility.

Cryptic *attL*, *attR*, and *attP* sites are unlikely to occur even in large genomes because of the number and arrangement of binding sites for Int, Xis, and IHF required (Landy 1989) for functionality. However, cryptic *attB* sites may be encountered. Not only are *attB* sites small (25 bp), but mutations can be introduced into the 25-bp *attB* sites while maintaining fully active recombination. For example, *attB2* (Fig. 1D) differs in sequence from the wild-type *attB* site at eight positions. To participate in cloning of a PCR product, a cryptic *attB* site must both be recognized by the intasome (the complex of the *attP* site with Int and IHF) and share sufficient homology with the *attP* site (*attP1* or *attP2* in the Donor plasmid; Fig. 1C) to allow productive formation and resolution of the reaction intermediates (Landy 1989). We speculate that for purposes of calculating the frequency of aberrant cloning events the effective size of an *attB* site is in the range of 12–16 bp. Thus, if PCR products totaling millions of base pairs are cloned by the RC BP Clonase reaction, it would not be surprising if a functional *attB* site was encountered in genomic DNA sequence.

The efficiency of the in vitro RC reactions decreases with increasing size of the DNAs involved, as judged by the number of colonies produced (Figs. 2C and 3B). This effect can be minimized by using equal



**Figure 4** (A) SDS-PAGE gel of proteins expressed from Expression Clones of the genes in pDEST17 (His6 fusion) vector in *Escherichia coli* strain BL21 SL. (Lane 1), eIF4e; (lane 2), tyrosine kinase; (lane 3), transferrin receptor; (lane 4),  $\beta$ -adaptin; (lane 5), MAP4 (MAP4 is known to migrate aberrantly on SDS-PAGE gels; Chapin et al. 1995); (lane 6), Gus. (B) SDS-PAGE gel of expression of genes subcloned from the same Entry Clones as in (A) into baculovirus Destination Vectors. (Lanes 1–5) are pDest8 (native expression; (lane 1), eIF4e; (lane 2), tyrosine kinase; (lane 3), transferrin receptor; (lane 4),  $\beta$ -adaptin; (lane 5), MAP4). (Lanes 6–8) the *gus* gene, from the same Entry Clone as in (A), expressed in the baculovirus vectors pDest8 (lane 6, native, 68 kD), pDest10 (lane 7, His6 fusion, 74 kD), and pDest20 (lane 8, GST fusion, 97 kD). As in *E. coli* expression (above), MAP4 protein migrated as a ~200-kD protein instead of as a 121-kD protein (Chapin et al. 1995).

moles of DNA and by incubating for longer times (G. Temple, unpubl.). A 10-kb PCR product has been cloned with the BP reaction (G. Temple, unpubl.). A Destination Vector >100 kb in size has been used to clone and express genes in insect cells (K. Franke, pers. comm.).

We provide several examples of the feasibility and utility of RC. Flanking the CAT gene with *attL* recombination sites in an Entry Clone allowed parallel transfer of this gene into 12 diverse Destination Vectors. These reactions required little bench time. They also yielded a high percentage of desired clones, reducing the need to screen numerous colonies for the construct of interest. We further applied this approach to optimizing protein expression. The successful expression and purification of proteins often requires assessment of multiple systems (e.g., different promoters or fusion tags) and host backgrounds (e.g., *E. coli*, yeast, insect or mammalian cells). Each of five cloned human genes and two *E. coli* genes were transferred into collections of Destination Vectors that allowed production of native His6-, GST-, or thioredoxin-fusion proteins in *E. coli* or insect or mammalian cells (Figs. 3 and 4). Expression of the *E. coli gus* gene was high in all systems tested. In contrast, expression of the human ORFs was system dependent. These results highlight the value of generating multiple systems in parallel to optimize expression. In several direct comparisons of expression constructs containing or lacking *attB* recombination

sites, no detectable differences have thus far been observed in protein yield or activity (data not shown).

RC can be used for directional cloning of PCR products by incorporating an *attB1* site in the forward primer and *attB2* in the reverse primer and recombining the *attB*-flanked PCR product with a vector containing *attP1* and *attP2* sites. The product of this reaction contains the cloned amplified DNA flanked by *attL* sites (an Entry Clone) that can subsequently be transferred into Destination Vectors. We demonstrate here the amplification and cloning of five human genes from HeLa RNA template and two *E. coli* genes from plasmid templates. Primers containing *attB* sites have also been used with genomic DNA and cDNA libraries as templates with a variety of polymerases. Altogether, hundreds of genes have now been amplified and cloned in this way (data not shown).

So long as genes and ORFs could not be manipulated in a uniform manner—independent of size, sequence, or restriction sites—manipulating large numbers of them was not possible. Because RC is nearly independent of these constraints and is highly efficient, genes may be cloned, subcloned, screened for phenotypes, and retrieved from screening protocols with high-throughput procedures. For example, Walhout et al. (2000) have begun a genome-wide survey of all the protein–protein interactions of *Caenorhabditis elegans*. The 19,000 ORFs predicted from the genomic sequence will be amplified and the PCR products cloned by RC, then subcloned into two-hybrid bait or prey vectors. As a starting point, an RC-compatible cDNA library was screened using 27 bait clones, constructed with RC, that contained genes implicated in a developmental pathway. The screen yielded 124 unique interactors, 109 of which were previously unknown. Simpson et al. (2000) have used RC to clone >100 unknown human ORFs and express them as fusions to green fluorescent protein in mammalian cells. The observed intracellular locations of the proteins usually, but not always, supported inferences about function gained from bioinformatics methods.

It is important to note that RC-compatible clones can be transferred easily into other RC-compatible vectors. ORFs from *C. elegans* can be transferred to vectors that express them as GFP fusions and tested for intracellular location. Human ORFs can be transferred to yeast two-hybrid vectors to screen for protein–protein interactions. These transfers can be accomplished robotically if desired, because of the efficiency and uniformity of the RC reactions. Thus, one of the most important advantages of RC-based genomics is the potential to widely distribute sequence-verified ORFs, of known or unknown function, for use in a variety of technology platforms. The Harvard Institute of Proteomics (<http://www.hip.harvard.edu/>) has been established to use RC to clone all known human ORFs for

this purpose. RC-compatible cDNA libraries from human tissues are available from the National Institutes of Health Cancer Genome Anatomy Project (in the vector pCMVSPORT6; <http://www.ncbi.nlm.nih.gov/ncicgap/>). The availability of RC-compatible clones of genes from diverse model organisms will assist the rapid evaluation of hypotheses from other high-throughput methods and can thus contribute to the solution of important biological problems.

## METHODS

All materials, including PCR primers, were obtained from Life Technologies, Inc., unless otherwise noted, and were used according to the manufacturers' instructions. RC-related materials are available as the Gateway Cloning Technology. Sequences of DNAs are available at [www.lifetech.com](http://www.lifetech.com). The bacteriophage  $\lambda$  recombination sites used here can be derived from the sequences of the *attB1* and *attB2* sites (Fig. 1D) and the sequence of the bacteriophage (GenBank accession no. J02459) according to the mechanism of  $\lambda$  recombination (Weisberg and Landy 1983). The *attP* sites in pDONR203 correspond to  $\lambda$  coordinates 27586 through 27818. The *attR* sites in the Destination Vectors lack the bases between 27586 and 27618 (deletion of the P1 and H1 domains improves the excision reaction [Bushman et al. 1985]), and base 27630 has been changed to a G to remove an *NdeI* site. Miniprep DNAs were prepared by alkaline lysis of overnight cultures and dissolved in TE (10 mM Tris HCl pH 7.5, 1 mM EDTA) containing RNase A (Sigma; 1  $\mu$ g per mL cell culture). Midi and maxi scale DNA preparations used the Concert purification system. Destination Vectors for use in RC may be constructed by inserting the appropriate blunt reading frame cassette (of the general form, *attR1* – Cm<sup>R</sup> – *ccdB* gene – *attR2*) into an available restriction site of any plasmid and propagated in *E. coli* strain DB3.1 (RR1 *endA recA gyrA462*), which is resistant to the toxic *ccdB* gene product (Bernard and Couturier 1992).

## PCR

Sources of templates are shown in Table 2. Primer sequences were, for human eIF4e, [A]atggcgactgtcgaaaccgaaa (fwd) and [B]actaaacaacacatttttagtggtgga (rev); for human tyrosine kinase, [A]atgtccaccagaccggcatc (fwd) and [B]tcagtagtagcttcagttccgct (rev); for human transferrin receptor, [A]atgatgatcaagtagatcagca (fwd) and [B]actaaaactcattgtcaatgtccc aaacg (rev); for human  $\beta$ -adaplin, [A]atgactgactcaaatatttcac cac (fwd) and [B]actagttcttgaggatgtctctgtagg (rev); for human MAP4, [A]atggctgacctcagctctgcag (fwd) and [B]actagatgctt gtctcctg-gatctggc (rev); for *E. coli gus*, [A]atggtccgtctgtagaaacc (fwd) and [B]actattgttgcctccct-gctgcgg (rev); and for *E. coli tetR*, [A]aattctcatgttgacagcttacc (fwd) and [B]cgatggatgttct gccaag (rev). [A] = ggggacaagttgtacaaaaagcaggctcattta-actttaagaaggagatatatacc, in which sequences in bold type correspond to *attB1* and those following or in italics are translation signals for *E. coli* or mammalian cells, respectively. [B] = ggggaccactttgtacaagaagctgggt, in which sequences in bold types correspond to *attB2*. Primers were desalted but otherwise unpurified. First-strand cDNA (made with ThermoScript reverse transcriptase) equivalent to 200 ng of total HeLa RNA was used as template in the amplification of the human genes; 1 ng of the appropriate *E. coli* plasmid DNA was used for amplification of *gus* and *tetR* genes. Following amplification with Pfx polymerase, amplification products were pre-

cipitated with polyethylene glycol + Mg to remove primer dimers (which clone efficiently) and dissolved in TE. Aliquots (2.5  $\mu$ L) of each product were applied to a 1% agarose/ethidium bromide gel (Fig. 2A).

## Clonases

BP Clonase, containing Int and IHF, catalyzes the integrative (BP) reaction (*attB*  $\times$  *attP*  $\rightarrow$  *attL* + *attR*). LR Clonase, containing Xis, Int, and IHF, catalyzes the excisive (LR) reaction (*attL*  $\times$  *attR*  $\rightarrow$  *attB* + *attP*). IHF (heterodimer; GenBank accession no. X04864 and V00291), Int, and Xis, (bacteriophage  $\lambda$ ; GenBank accession no. J02459) were purified from *E. coli* strains containing the cloned, overexpressed genes. Clonases and other RC materials are available from Life Technologies, Inc., as part of the Gateway Cloning System.

## BP Reactions for Cloning PCR Products

The *attP* plasmid pDONR203 (300 ng; Fig. 2B) was mixed with 2  $\mu$ L of each purified PCR product in reactions (20  $\mu$ L) that contained 4  $\mu$ L BP Clonase in 25 mM Tris HCl pH 7.5, 22 mM NaCl, 5 mM EDTA, 5 mM spermidine HCl, 1 mg/mL BSA. After incubation for 60 min at 25°C, proteinase K (4  $\mu$ g in 2  $\mu$ L) was added, and each reaction was incubated at 37°C for 20 min. Aliquots (2  $\mu$ L) of each reaction were transformed into *E. coli* DH5 $\alpha$  (Library Efficiency) and plated on kanamycin plates (100  $\mu$ g/mL) Fig. 2C) incubated at 37°C. Miniprep DNA was prepared from four colonies from each reaction, and 2  $\mu$ L aliquots were applied to a 1% agarose/ethidium bromide gel (Entry Clones; Fig. 2E).

## LR (Subcloning) Reactions

The LR reactions described in Table 1 used earlier versions of DNAs, proteins, and reaction conditions. Only the conditions used for the seven genes amplified in Figure 2 will be described. One miniprep of each cloned gene (Fig. 2E, lanes 1,5,9,13,17,21), and *tetR* was chosen as the Entry Clone. Aliquots containing ~200 ng of each miniprep DNA (except ~60 ng and 40 ng of the transferrin receptor and *tetR* Entry Clones, respectively) were incubated with 300–400 ng of the appropriate Destination Vector (linearized within the chloramphenicol resistance-*ccdB* region; Fig. 3A) in 20  $\mu$ L reactions containing 4  $\mu$ L of LR Clonase, 50 mM of Tris HCl pH 7.5, 50 mM of NaCl, 0.25 mM of EDTA, 2.5 mM of spermidine HCl, and 0.2 mg/mL of BSA. Then proteinase K (4  $\mu$ g in 2  $\mu$ L) was added, and reactions were incubated at 37°C for 20 min. Aliquots (2  $\mu$ L) of each reaction were transformed into *E. coli* DH5 $\alpha$  and plated on ampicillin (100  $\mu$ g/mL) plates (Fig. 2C) incubated at 37°C. Miniprep DNAs of the resulting Expression Clones were prepared from one colony from each reaction, and aliquots were applied to a 1% agarose/ethidium bromide gel (Fig. 3D).

## Protein Expression

For expression in *E. coli*, miniprep DNAs of Expression Clones in pDEST17 (His6 amino fusion) or other Destination Vectors (pDEST15 for GST amino fusions or pDEST16 for thioredoxin amino fusions) containing the T7 promoter were transformed into competent BL21 SI cells (Bhandari and Gowrishankar 1997) and plated at 37°C on LB ampicillin plates lacking NaCl. Colonies were picked into LB broth lacking NaCl. After growth at 30°C to an A590 ~0.3, NaCl was added to 0.3 M, and expression was continued at 25°C for 2 h. Cells were lysed in SDS/ $\beta$ -mercaptoethanol, and aliquots (0.05 A590 units) were

electrophoresed on a 4%–20% polyacrylamide Tris-glycine SDS gel (Novex) and stained with Coomassie blue. For expression in insect cells, miniprep DNAs of subclones in pDEST8 (for native protein) or other baculovirus Destination Vectors (pDEST10 for His6 amino fusions, pDEST20 for GST amino fusions) were transformed into *E. coli* DH10Bac cells, in which transfer of the expression region into bacmid DNA occurred in vivo (Luckow et al. 1993). Minipreps of bacmid DNAs were transfected into Sf9 cells using Cellfectin. Viral supernatants were harvested after 72 h at 27°C, clarified, and used to infect fresh cultures at a multiplicity of infection of five, assuming a titer of  $1 \times 10^7$  viral particles/mL. Cells were harvested 48 h after infection. Proteins from  $1\text{--}2 \times 10^5$  cells were applied to lanes of a 4%–20% polyacrylamide Tris-glycine SDS gel. For expression in mammalian cells, the *gus* gene was subcloned into pDEST12.2 (CMV promoter, neomycin resistant), and the resulting Expression Clone DNA was purified using the Concert system and transfected into COS-7 and CHO-K1 cells using Lipofectamine2000. Cells were stained for Gus activity using X-glucuronide.

## ACKNOWLEDGMENTS

We thank Dr. Howard Nash for generously providing purified  $\lambda$  recombination proteins in early stages of this work and to our colleagues at Life Technologies, Inc., for their many essential contributions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Abremski, K. and Hoess, R. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259**: 1509–1514.
- Bernard, P. and Couturier, M. 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226**: 735–745.
- Bhandari, P. and Gowrishankar, J. 1997. An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. *J. Bacteriol.* **179**: 4403–4406.
- Boyd, A.C. 1993. Turbo cloning: A fast, efficient method for cloning PCR products and other blunt-ended DNA fragments into plasmids. *Nucleic Acids Res.* **21**: 817–821.
- Bubeck, P., Winkler, M., and Bautsch, W. 1993. Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res.* **21**: 3601–3602.
- Bushman, W., Thompson, J.F., Vargas, L., and Landy, A. 1985. Control of directionality in  $\lambda$  site specific recombination. *Science* **230**: 906–911.
- Chapin, S.J., Lue, C.M., Yu, M.T., and Bulinski, J.C. 1995. Differential expression of alternatively spliced forms of MAP4: A repertoire of structurally different microtubule-binding domains. *Biochemistry* **34**: 2289–2301.
- Degryse, E. 1996. In vivo intermolecular recombination in *Escherichia coli*: Application to plasmid constructions. *Gene* **170**: 45–50.
- Gopaul, D.N., Van Duyne, G.D., and Guo, F. 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl. Acad. Sci.* **96**: 7143–7148.
- Lafontaine, D. and Tollervey, D. 1996. One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res.* **24**: 3469–3471.
- Landy, A. 1989. Dynamic, structural, and regulatory aspects of  $\lambda$  site-specific recombination. *Ann. Rev. Biochem.* **58**: 913–949.
- Liu, Q., Li, M.Z., Leibham, D., Cortez, D., and Elledge, S.J. 1998. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**: 1300–1309.
- Luckow, V.A., Lee, S.C., Barry, G.F., and Olins, P.O. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J. Virol.* **67**: 4566–4579.
- Miki, T., Park, J.A., Nagao, K., Murayama, N., and Horiuchi, T. 1992. Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. Mutants of DNA gyrase subunit A suppress *letD* (*ccdB*) product growth inhibition. *J. Mol. Biol.* **225**: 39–52.
- Oliner, J.D., Kinzler, K.W., and Vogelstein, B. 1993. In vivo cloning of PCR products in *E. coli*. *Nucleic Acids Res.* **21**: 5192–5197.
- Peakman, T.C., Harris, R.A., and Gewert, D.R. 1992. Highly efficient generation of recombinant baculoviruses by enzymatically mediated site-specific in vitro recombination. *Nucleic Acids Res.* **20**: 495–500.
- Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. 2000. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Reports* **1**: 287–292.
- Storck, T., Kruth, U., Kolhekar, R., Sprengel, R., and Seeburg, P.H. 1996. Rapid construction in yeast of complex targeting vectors for gene manipulation in the mouse. *Nucleic Acids Res.* **24**: 4594–4596.
- Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116–122.
- Weisberg, R.A. and Landy, A. 1983. In *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Zhang, Y., Buchholz, F., Muirers, J.P., and Stewart, A.F. 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**: 123–128.

Received April 16, 2000; accepted in revised form September 18, 2000.