MINI-REVIEW

Bacteriophage recombination systems and biotechnical applications

Nafiseh Nafissi · Roderick Slavcev

Abstract Bacteriophage recombination systems have been widely used in biotechnology for modifying prokaryotic species, for creating transgenic animals and plants, and more recently, for human cell gene manipulation. In contrast to homologous recombination, which benefits from the endogenous recombination machinery of the cell, site-specific recombination requires an exogenous source of recombinase in mammalian cells. The mechanism of bacteriophage evolution and their coexistence with bacterial cells has become a point of interest ever since bacterial viruses' life cycles were first explored. Phage recombinases have already been exploited as valuable genetic tools and new phage enzymes, and their potential application to genetic engineering and genome manipulation, vectorology, and generation of new transgene delivery vectors, and cell therapy are attractive areas of research that continue to be investigated. The significance and role of phage recombination systems in biotechnology is reviewed in this paper, with specific focus on homologous and site-specific recombination conferred by the coli phages, λ, and N15, the integrase from the *Streptomyces* phage, ΦC31, the recombination system of phage P1, and the recently characterized recombination functions of *Yersinia* phage, PY54. Key steps of the molecular mechanisms involving phage recombination functions and their application to molecular engineering, our novel exploitations of the PY54-derived recombination system, and its application to the development of new DNA vectors are discussed.

N. Nafissi
School of Pharmacy, Faculty of Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: nnafissi@uwaterloo.ca

R. Slavcev (✉)
School of Pharmacy, University of Waterloo,
200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada
e-mail: slavcev@uwaterloo.ca

Introduction

Bacteriophages (phages) are viruses that infect bacterial cells and use the host cell machinery to generate progeny. Phages are divided into two major categories: "virulent" and "temperate". After infection of a sensitive host cell, lytic phages immediately propagate themselves by using the host cell molecular machinery, releasing thousands of new phage particles into the environment upon killing of the infected cells. In contrast, temperate bacteriophages "make a decision" between lytic growth and lysogeny, a latent state, whereby the genome is passively replicated until the prophage is induced (Groth and Calos 2004). Temperate bacteriophages that integrate their genome into the host chromosome in the establishment of lysogeny facilitate integration by encoding an integrase enzyme that mediates unidirectional site-specific recombination between two DNA recognition sites, the bacteriophage attachment site, *attP*, and the bacterial attachment site, *attB*. The lysogen in this case is flanked by two hybrid sites, *attL* and *attR*, which consist of half *attP* and half *attB* sequences that become the substrates for extreme recombination by the excisionase. Several unfavorable environmental factors affect the excisionase reaction, which is followed by lytic phage release and the initiation of a new infection cycle. The ability of phage integrases to unidirectionally recombine two short DNA sequences offers utility toward many genetic engineering applications (Groth and Calos 2004). As such, phage integrases are of growing importance in the genetic manipulation of living eukaryotic cells, particularly mammals that possess large genomes, for which there are few tools for

precise genetic manipulation. Integrases of different phages have been shown to work efficiently in mammalian cells, carrying out efficient integration at introduced native *att* sites or pseudo sites available on the host genome with partial identity to *att* sequence. This reaction may be exploited in biotechnical applications such as manipulation of prokaryotic and eukaryotic cells, bacterial engineering, construction of transgenic higher eukaryotes, or cell/gene therapy. Each phage-encoded integrase recognizes specific individual sequences. Some integrases act autonomously, while others act with the help of other phage proteins and/or bacterially encoded host factors (Groth et al. 2000).

Homologous recombination

The process of homologous recombination involves the alignment of similar DNA sequences to form mobile forks between four strands of DNA termed Holliday junctions that are highly conserved in both prokaryotic and eukaryotic cells. Homologous recombination requires energy and cofactors from host cells (Thyagarajan et al. 2000). The frequency of integration by homologous recombination appears to generally occur at a frequency of about 10^{-6} for most mammalian cells, although it was reported that this frequency may be increased by up to 20-fold using completely isogenic DNA or up to 1,000-fold following the introduction of a double-strand break at the target site; however, the resources necessary to generate such a break into endogenous sequences remains a challenge (Porteus 2012).

The highly efficient genetic engineering system, "recombineering", is based on a homologous recombination, but exploits a phage-encoded recombination system (Thomason et al. 2001). Recombineering and homologous recombination now represent the methods of choice for knock-in and knock-out strategies to either manipulate bacterial and mammalian genome or as an alternative to traditional restriction enzyme-mediated cloning with PCR products (Thomason et al. 2001; Clokie et al. 2009; Sharan et al. 2009). Homologous recombination systems occur via either the host RecA-dependent or RecA-independent pathways or *Escherichia coli* RecA-dependent recombination pathway.

In prokaryotes, homologous recombination plays an important role in the generation of genetic diversity and in DNA repair processes. One of the best studied mechanisms is the RecA-dependent or the RecBCD pathway that is useful in the repair of chromosomal DNA double-strand breaks (Dillingham and Kowalczykowski 2008). The RecBCD recombination complex recognizes specific sequences within the bacterial chromosome called "*chi* sites" that prevent DNA degradation. Foreign DNA segments that lack *chi* sites are thus degraded to protect the *E. coli* cell from invading nucleic acids. RecBCD, also known as Exonuclease V

(Exo V), is an enzyme that recognizes a double-stranded (ds) DNA break and initiates recombination and repair. The enzyme complex is composed of three different subunits called RecB, RecC, and RecD, hence the complex name, RecBCD. Both the *alpha* (RecD) and *beta* (RecB) subunits are energy-dependent helicases that unwind and separate the strands of DNA, where the *beta* subunit is also a nuclease that introduces single-stranded (ss) nicks into the DNA. The *gamma* subunit (RecC) recognizes the 5′-GCTGGTGG-3′ (*chi*) sequence to initiate recombination (Wigley 2013).

The multifunctional protein, RecA, is a significant contributor to recombinational repair of damaged DNA in *E. coli*, with a functional homolog present in every species. RecA activity is energy-dependent, where the protein strongly binds to and coats the single-stranded DNA generated by the exonuclease activity of Exo V to form a nucleoprotein filament. Since it has more than one DNA-binding site, RecA can hold single- and double-stranded DNA together, a feature that makes it possible to catalyze a DNA synapse reaction between a DNA double helix and a homologous region of single strand (ssDNA). This reaction initiates the exchange of strands between two recombining DNA double helices, which progresses down the strand by DNA branch migration (Dillingham and Kowalczykowski 2008).

We previously reported on the application of RecA$^+$ *E. coli* K-12 (W3110) cells to process homologous recombination-mediated insertions of bacteriophage-derived recombinase genes into the *lac* operon of *E. coli* using the pBRINT-*cat* integrating plasmid (Nafissi and Slavcev 2012). The pBRINT family of integrating vectors facilitates the homologous recombination-mediated chromosomal integration of cloned sequence of interest into the *lacZ* gene of *E. coli* in a RecA-dependent manner (Le Borgne et al. 1998). In contrast, a *recA* mutant host is often employed for cloning and plasmid maintenance to ensure the stability of extrachromosomal DNA.

The RecA-independent bacteriophage λ Red recombination system

The recombination genes of the bacteriophage λ Red system were among the earliest genes to be described. Following the isolation of *recA*, *recB*, *recC*, or *recD* recombination-deficient mutants, it was found that phage λ could effectively initiate homologous recombination independent of any of the bacterial recombination pathways due to the encoding of its own recombination enzymes. The λ Red recombination was thus considered to be a separate pathway (Christensen 2001). The *red* system is comprised of the *exo* (α), *bet* (β), and *gam* (γ) genes, all clustered in the *pL*-governed operon of the λ genome and is regulated by the CI repressor (Stahl 1998). The *exo* gene encodes the 24-kDa 5′-3′ exonuclease (Exo) that targets dsDNA ends; *bet* encodes the 28-kDa ssDNA-

binding protein (Bet) capable of annealing complementary ssDNA strands; and *gam* encodes the 16-kDa polypeptide (Gam) that inhibits Exo V activity and confers protection against nuclease attack and digestion of the λ linear dsDNA genome. During the λ lytic cycle, λ Red recombination is stimulated by endonuclease activity of the λ terminase on the phage's terminal cohesive ends (*cos*) sites, thereby introducing the dsDNA break following the 5′-3′ exonuclease activity of Exo that provides the 3′ overhang to initiate recombination (Fig. 1). In this case, if an intact dsDNA with no free ends is the only available partner, recombination proceeds through the *E. coli* RecA-dependent mechanism by strand invasion. In contrast, if the available partner is a damaged/nicked dsDNA or a replicating chromosome with a free double-strand end, then recombination can proceed in a RecA-independent manner through Bet-mediated annealing. The Red recombination system is facilitated by the λ Gam function that inhibits *E. coli* RecBCD exonuclease activity to prevent further digestion of λ phage DNA by host cell, thus promoting Exo and Bet access to DNA ends to promote Red-mediated recombination (Poteete 2001).

The Red recombination system was applied for the first time in 1998 for efficient and simplified genetic engineering of bacterial cells (Murphy 1998). The system was exploited by expressing the λ *red* genes, either via plasmid or chromosomally inserted, to promote recombination between the bacterial chromosome and a nonreplicative linear dsDNA molecules introduced into the cell. Most of the recombination systems that had been developed previous to this study required long

Exo binds to linear DNA and generates a 3' ssDNA 5' to 3' exonuclease activity.

3'   3'

Exo

Beta binds to ssDNA

Beta

Target DNA

Recombination occurs between two homology arms

Homologous recombination generates a new recombinant DNA

Fig. 1 The molecular mechanism of Red-mediated recombination

homology arms (>500 bp), whereas the plasmid-encoded *red* system facilitated insertion and recombination events between PCR products with short homology arms (>50 bp) into the bacterial chromosome with high efficiency. This technology has since been applied for genome editing in different prokaryotic cells, inserting genes for long term and stable production of recombinant proteins and enzymes (Murphy et al. 2012).

Site-specific recombination

Permanent modification of the target genome is of great utility in genetic engineering and is generally carried out by three different approaches: (1) random integration (illegitimate recombination), (2) homologous recombination, and (3) site-specific recombination. Although random integration can be used to place the introduced gene into the genome, the lack of control over the position of introduced DNA can result in undesirable side effects, including unpredictable expression of the introduced gene and potential mutagenesis of neighboring genes. As such, it is a method best suited for insertional mutagenesis. A strategy that can exclusively yield efficient site-specific integration into safe locations in the target genome would be ideal. However, while homologous recombination provides excellent specificity in integration sites, it occurs at too low frequency to be optimal for genetic engineering in multicellular organisms (Kuhlman and Cox 2010).

Site-specific DNA recombination systems are derived from prokaryotes and unicellular yeasts, and among them, bacterial viruses provide a repertoire of recombinational systems, a number of which have been exploited to facilitate efficient DNA exchange in human cells. As previously mentioned, temperate bacteriophages are often integrated into the host chromosome at a specific site by recombination, a process that requires specialized "attachment sites", *attP* and *attB* in the phage and the host chromosomes, respectively. In the prophage, the enzymes generate two junction sites, *attL* and *attR*, consisting of hybrid sequences of *attP* and *attB*, and the integration reaction strictly conserves the original sequences in the recombined product sequences. Lysogeny is reversible, and a minority of cells within a population of lysogens will spontaneously lyse, releasing progeny phage. The integrated prophage is flanked by *attL* and *attR*, which are the substrates for excisionase recombination, a process that reforms *attP* and *attB* and releases the prophage. During the initial stage of the conversion to the lytic phase, a phage-encoded excisionase catalyzes the excision of the prophage genome by cooperating with the phage integrase, and the original *attP* and *attB* sequences are regenerated via site-specific recombination between *attL* and *attR* (Turan et al. 2011).

Site-specific recombination facilitates integration, excision, and inversion of defined DNA segments that usually requires no DNA synthesis or high-energy cofactor (Groth and Calos
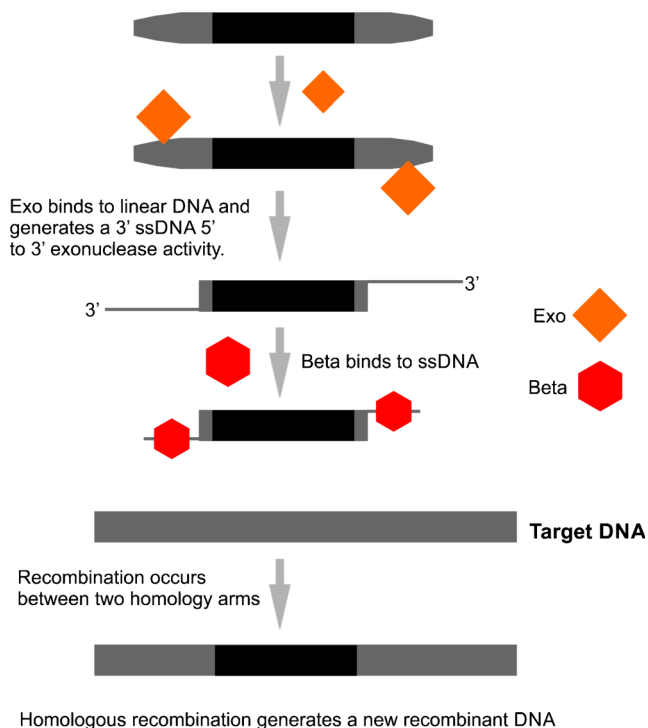
2004). Site-specific recombinases are classified into two major families of tyrosine or serine, which were named based on the amino acid residue that forms a covalent protein-DNA linkage in the intermediate reaction. Although enzymes of the site-specific recombinase families share high specificity and efficiency, their recombination mechanisms are distinctly different. In fact, all recombinases typically mediate efficient "cut and paste"-type DNA exchange between recognition sites in the range of 30–40 bp or longer (Groth and Calos 2004). Recombination is conservative and is characterized by the following events: (1) recombination occurs at a specific site on the interacting DNA molecules, (2) expression and synthesis of the recombinase enzyme, (3) strand exchange occurs at small regions of DNA homology within the recombining sites, (4) pairing of the interacting recombination sites followed by strand exchange resulted in structural intermediates, and finally (5) resolution of intermediates followed by strand migration (Groth and Calos 2004). However, tyrosine recombinases form "Holliday intermediates" by cleaving one strand of each duplex at a time; each crossover site is nicked and must be joined to its partner before the second strand can be cut. In contrast, serine recombinases cut all strands in advance at both crossover sites before any exchange, following 180° rotation and rejoining of cleaved substrate (Sauer and Henderson 1988). The tyrosine-type recombinases are subdivided into two families based on the directionality of site-specific recombination: (1) unidirectional tyrosine-type phage integrases and (2) bidirectional tyrosine-type simple recombinases. Generally, the tyrosine family of integrases tends to recognize longer *attP* sequences and requires helper proteins encoded by the phage or the host bacteria for their activity. In contrast, serine recombinases recognize shorter *attP* sequences that are still long enough to

be specific on a genomic scale and do not require host cofactors. Serine-type recombinases are subdivided into two families based on protein size: (1) small serine-type resolvase/invertases and (2) large serine-type phage integrases (Hirano et al. 2011).

Some integrases function with no requisite for cofactors or helper proteins and therefore offer improved activity in heterologous environments. From this group, bacteriophage P1-derived Cre (cyclization recombinase), bacteriophage lambda-derived Int (integrase), and *Saccharomyces cerevisiae*-derived Flp have been applied widely in gene manipulation of higher eukaryotic organisms and in the production of transgenic species (Groth et al. 2000). Recombinases such as Cre, Flp, and Int perform both integration and excision at the same target sites, and thus, their net integration frequency in mammalian cells is generally low. An ideal application would see that the transgene remain inexcisable following integration (Groth and Calos 2004); a property shown by some phage-derived recombinases such as the *Streptomyces* phage-encoded ΦC31 integrase (Chavez and Calos 2011), *Escherichia* phage-encoded N15 TelN protelomerase (Deneke et al. 2000), and *Yersinia* phage-encoded PY54 Tel protelomerase (Hammerl et al. 2007) (Table 1).

Here, we examine the application of phage-derived enzymes toward (1) human genome engineering strategies and regenerative medicine and (2) vectorology and in engineering of superior DNA transgene delivery vectors.

Phage lambda integrase

Phage λ-derived integrase (Int) is a conservative site-specific recombinase that belongs to the tyrosine-type integrase family

Table 1 Characteristics of site-specific recombinases active in mammalian cells

	1 Catalytic residue	2 Recombination mechanism	3 Recombinase family	4 Protein size (a.a.)	5 Recombination directionality	Minimal recognition site
Phage λ Int	Tyrosine	Formation and resolution of Holliday junction	Tyrosine-type phage integrases	356	Unidirectional	Nonidentical 250-bp λ *attP* and 25-bp *attB*
Phage P1 Cre	Tyrosine	Formation and resolution of Holliday junction	Simple recombinase	343	Bidirectional	Identical 34-bp *loxP*
Saccharomyces cerevisiae FLP	Tyrosine	Formation and resolution of Holliday junction	Simple recombinase	423	Bidirectional	Identical 48-bp *FRT*
Phage ΦC31 integrase	Serine	180° rotation and rejoining of cleaved DNAs	Phage integrase (large recombinase)	613	Unidirectional	Nonidentical ~40-bp *attP* and *attB*
Phage N15 TelN	Tyrosine	Cleaving joining activity at perfect palindrome target sequence to generate DNA intermediate with hairpin ends	Protelomerase	630	Unidirectional	Perfect palindrome *telRL* sequence 46 bp
Phage PY54 Tel	TelN-like	Cleaving joining activity at perfect palindrome target sequence to generate DNA intermediate with hairpin ends	Protelomerase-like	617	Unidirectional	Perfect palindrome 42-bp *pal* sequence

of recombinases. The natural function of Int is to carry out integration of the λ genome into the *E. coli* genome. In combination with the Xis (excisionase) protein, Int is also involved in excision of λ upon prophage induction. Although the Int-*att* system was considered as a potential candidate for human cell genome manipulation, it was not selected due to certain obstacles encountered in clinical studies. Int requires the host-encoded proteins (integration host factor, IHF) for efficient integration, which limits its integrative recombination activity in heterologous cells. However, several attempts to overcome the problems associated with the use of these integrases in heterologous cells have been made (Sanger et al. 1982). For example, IHF-independent mutant λ integrases have been generated that catalyze DNA intermolecular integration and excision reactions at low levels in the absence of accessory proteins in mammalian cells (Lorbach et al. 2000). We previously used the conditional replication, integration, and modular (CRIM) integrating plasmids (Haldimann and Wanner 2001) possessing the λ *attP* site-specific sequence to insert a foreign operon into a Rec+ *E. coli* K12 (W3110) genome (Nafissi and Slavcev 2012). These modified "CRIM" plasmids were integrated into the host bacterial attachment (*attB*) site by supplying phage λ integrase (Int) from a helper plasmid. This system proved very efficient for site-specific integration of the gene of interest (GOI) into the host genome. However, λ integrase usually catalyzes bidirectional site-specific recombination in cells other than the natural *E. coli* host, which limits its application in clinical studies where the strict unidirectional integration of foreign DNA into a heterologous genome is desirable (Lorbach et al. 2000; Hirano et al. 2011). As such, λ Int is not an ideal recombinase for stable and long-term human genomic modifications.

Phage P1 Cre recombinase and *S. cerevisiae* Flp recombinase

Phage P1-derived Cre ("causes recombination") and *S. cerevisiae*-derived Flp ("flipase") recombinases were discovered by Austin et al. (1981) and Broach et al. (1982), respectively. These recombinases catalyze reversible site-specific recombination events between two short, identical sequences: a 34-bp *loxP* target site for Cre and the 48-bp *FRT* site for Flp. Both Cre and Flp belong to the simple/bidirectional tyrosine-type recombinase family and, as such, do not require host-encoded accessory proteins or specific DNA structures (Hirano et al. 2011). The simplicity of the Cre/*loxP* and Flp/*FRT* systems has led to their extensive use as tools for in vivo genomic engineering such as DNA rearrangements and gene knock-in and knock-outs in a wide variety of heterologous environments, including human cells. Both recombinases have been extensively used to activate or switch

gene expression in clinical applications, and integration of foreign DNA sequences into chromosomally integrated *loxP* or *FRT* sites is reversible since wild-type Cre and Flp recombinases predominantly catalyze re-excision of the integrated DNA molecule via the reverse reaction. This feature has been widely applied in knock-out studies when excision of a particular gene is required.

Cre recombinase differs from typical tyrosine-type temperate phage integrases in that it is not involved in prophage establishment, but rather resolution. Cre is essential for P1 lysogeny as the prophage does not integrate into the host chromosome, but rather replicates as an episomal replicon in its host. The P1 lysogenic cycle begins with circularization of the injected linear P1 DNA into infected cells by site-specific recombination at *loxP* sites in the phage genome. P1 plasmids replicate independently from the host chromosome, and Cre/*loxP* activity resolves dimeric P1 plasmids into monomers, thereby facilitating stable partitioning of P1 into daughter cells during cell division (Austin et al. 1981). Cre activity in cointegrate resolution limits the utility of the Cre/*loxP* system for applications involving stable genomic integration. The Flp recombinase from yeast with 2-μm circle, similar to Cre, also differs from typical temperate phage integrases in its physiological function. Initially, the Flp recombinase leads to amplification of the 2-μm circle by inverting the DNA segment flanked by the two *FRT* sites; this rearrangement initiates rolling-circle replication by changing the direction of the replication fork. Next, the enzyme resolves the multimeric array generated by rolling-circle replication into the 2-μm circle monomers (Broach et al. 1982). Thus, the physiological function of Flp recombinase is the inversion and excision of a DNA segment flanked by two *FRT* sites. As such, Flp/*FRT*, like Cre/*loxP*, is not ideal for applications requiring stable integrative recombination.

Several strategies have been employed to overcome the issues associated with application of these recombination systems to unidirectional integrative recombination. For instance, the recombination-mediated cassette exchange (RMCE) and the "LE/RE mutant system" have led to the development of two types of mutant *loxP* sequences to facilitate unidirectional site-specific recombination by Cre. The Cre recombinase recognizes 34-bp *loxP* sites, each of which consists of two 13-bp inverted repeats that bind to Cre and a central 8-bp spacer region where strand exchange occurs. If one nucleotide in the central 8-bp spacer region is modified, Cre recombinase is still able to recognize the modified target site and carry out the integration event, but cannot execute its excision function. This was shown in heterologous genomes, where a pre-inserted DNA sequence flanked by two mutant *loxP* sites containing modified spacer sequences, resulted in the loss of Cre endonuclease capacity (Bouhassira et al. 1997). Modified *loxP* sites have been used to integrate foreign DNAs into a wide variety of heterologous host genomes including

those of plants (Albert et al. 1995), mammals (Araki et al. 1997), and bacteria (Suzuki et al. 2007).

RMCE-type recombination based on the Flp/FRT system has also been utilized to facilitate unidirectional integrative recombination since FRT target sequences are longer and more complex than loxP sites (Turan et al. 2011). Each 48-bp FRT site consists of three 13-bp repeats and an 8-bp spacer region, in which strand exchange occurs. Two nearly identical 13-bp inverted repeats bind the Flp recombinase and flank the 8-bp spacer, while the additional 13-bp repeat follows a 1-bp gap forming a perfect direct repeat with the repeat at one side of the spacer. While the inversion and excision reactions occur with a 34-bp core site consisting of the two 13-bp inverted repeats and 8-bp spacer region, Flp-mediated recombination is more efficient with the intact FRT site that contains the third repeat serving as an extra Flp-binding sequence (Hirano et al. 2011).

Both λ and P1 phage-derived site-specific recombination systems have been exploited toward the production of modern plasmid-derived DNA transgene delivery vectors (Table 2).

The first system described in 1997 for bacterial sequence-free DNA "minicircle" production, employed the λ Int/attP site-specific recombination system. In this study, recombinant Int was expressed endogenously from E. coli, and two attP target sequences were inserted into a plasmid DNA vector flanking the eukaryotic expression cistron. Gene int expression

resulted in site-specific excision of att sites and generation of two mini circular constructs, one of which contained the bacterial sequence to form the "miniplasmid" and the other encoding the gene of interest (GOI) cistron to form the desired "minicircle" (Darquet et al. 1997). Despite the obvious advantage, the system was limited by the toxicity of phage λ Int recombinase in bacterial cells, thus necessitating tight control over the recombinase induction process in order to produce the minicircle in culture, which led to inefficient recombination (40–70 %) (Darquet et al. 1997, 1999). As such, the project instead turned to an alternative strategy of using the Cre-loxP recombination system. Similar to the original design, Cre was expressed either endogenously or episomally (Bigger et al. 2001). Cre would then catalyze site-specific recombination between loxP sites available on standard plasmid DNA vector (pDNA) leaving the bacterial sequence-free minicircle for further application (Bigger et al. 2001). Until recently, the Cre-loxP recombination system was the method of choice to generate DNA minicircle vectors, despite its limitation in processing pure DNA minicircles with no conventional plasmid residual contamination. The in vivo efficiency of this system was <100 %, and the undigested parent pDNA was one of the most challenging steps in purification. Other groups have since applied different recombination systems and tactics to improve DNA minicircle production leading to the achievement of

Table 2 Systems developed for production of bacterial sequence-depleted DNA vectors

6Type of DNA vector	7General features	8Generation Systems	9Improvements	10Limitations
Minicircle	Nonreplicating; CCC; provides sustained and high expression of GOI; no immunostimulatory effect; site-specific recombination dependent; high-yield production is available	Phage λ attB-attP recombination	–	Size exclusion/sequence-specific purification; low recovery due to bidirectional recombination reaction
		Cre/mutated loxP recombination	Unidirectional recombination reaction, tight control of recombinase gene expression	Size exclusion/sequence-specific purification
		ΦC31 Integrase	Unidirectional recombination reaction, tight control of recombinase gene expression, high-yield production of minicircle	
Ministring	Nonintegrating; nonreplicating; in vitro and in vivo production available; LCC topology; provides sustained and high expression of transgene; no immunostimulatory effect; TelN/telRL or Tel/pal recombination dependent for in vivo production; no limit on high-yield production; no isoform generation	"MIDGE"	Enhanced targeting	Enzymatic production process, high cost, limited high-yield production
		Purified N15 phage protelomerase TelN	No need to ligase to generate covalently closed ends	Size exclusion/sequence-specific purification, limited high-yield production, needs purified TelN
		PCR amplified	Enhanced targeting	Synthetic production process, high cost, limited high-yield production
		TelN/telRL and Tel/pal recombination	Unidirectional recombination reaction, tight control of recombines expression, high-yield production of ministrings	

highly pure minicircles with very low parent plasmid or miniplasmid residual contamination (Kay et al. 2010) (Fig. 2).

Phage ΦC31 integrase

In the 1990s, actinophage ΦC31 integrase was discovered and found to be homologous to resolvase/invertases (Kuhstoss and Rao 1991). This enzyme belongs to the large integrase serine-type recombinase family and catalyzes unidirectional site-specific recombination between *attP* and *attB* sites, both of which are normally ~40 bp in size and do not require any supplementary host factors or specific DNA structures. In general, both *att* sites have an identical 2-bp core sequence flanked by 20–25-bp imperfect, inverted repeats (P and P′ in *attP*; B and B′ in *attB*), where there is little sequence similarity between *attP* and *attB* sequences. This integrase requires the slightly longer *attP* sequence for efficient integration, and *attP* encodes more perfect inverted repeats than *attB*. The integrase possesses highly homologous N-terminal catalytic domains (140 amino acids (a.a.)) and diverse C-terminal DNA-binding domains (300–500 a.a.), and the DNA-binding domains confer sequence-specific DNA binding to distinct *att*-integrase pairs (Kuhstoss and Rao 1991). These integrases function on a recombination mechanism similar to that of the resolvase/ invertase, serine-type recombinases, involving the 180° rotation and rejoining of cleaved *att* DNAs, except that they recognize two different sites and catalyze a tightly regulated unidirectional recombination. The unidirectional recombination is mediated by the larger C-terminal DNA-binding domains of the integrase that possess a motif controlling recombinational direction. The excision reaction requires the phage-encoded excisionase, Xis, which binds tightly to the *attP*- and

attB-bound dimers, inhibiting the integration reaction, and binds less strongly to the *attL*- and *attR*-bound dimers to facilitate the excision reaction. It is this characteristic in particular that makes integrases the most ideal recombination enzyme of choice in clinical applications where there is a need for insertion of a specific gene into a heterologous genome and for provision of a long-term and stable expression of the protein product (Keravala et al. 2011).

The relative simplicity of serine-type phage integrase-mediated recombination has spurred the development of heterologous gene integration systems based on these enzymes. Genomic integration systems based on ΦC31 integrase have been developed for human genome editing (Brown et al. 2011). This integrase efficiently catalyzes the site-specific integration of an *attP*-containing DNA vector into preexisting (engineered) *attB* or *attP* sites in the mammalian genome. The second strategy (*attP*) imparted with better outcome owing to the complexity and lower accessibility of the chromosomal DNA compared to that of the pDNA structure (Rice et al. 2010). In addition, "pseudo *attP*" sites with only 40 % identical sequence to wild-type *attP* have been identified that can bind to ΦC31 integrase and mediate ΦC31 integrase-dependent integration at 10-fold higher rates than random sites. Interestingly, these "pseudo *attP*" sites are often present in transcriptionally active regions of genome and euchromatin chromosomal DNA, and thus, the insertion of a vector into pseudo *attP* sites often provides higher expression levels than those inserted randomly into the human genome. These findings indicate that multiple distinct site-specific recombination systems, based on various serine-type phage integrases, may be employed to simultaneously or sequentially introduce multiple foreign DNAs into *att* sites pre-inserted into heterologous genomes.

The integration frequency mediated by the ΦC31 integrase in mammalian cells is approximately 10- to 100-fold higher than Cre-mediated integration at an inserted wild-type *loxP* site, higher than FLP-mediated integration at an inserted FRT site, and more efficient than phage integrases of the tyrosine-catalyzed site-specific recombinase family, such as λ Int (Thyagarajan et al. 2001). Furthermore, recent advances in recombinase engineering showed promising results to target the GOI into a specific site in the mammalian genome, for example, by fusing a DNA binding domain into the serine family recombinase, serin-type-mediated recombination rates were dramatically improved in mammalian cells (Gaj et al.).
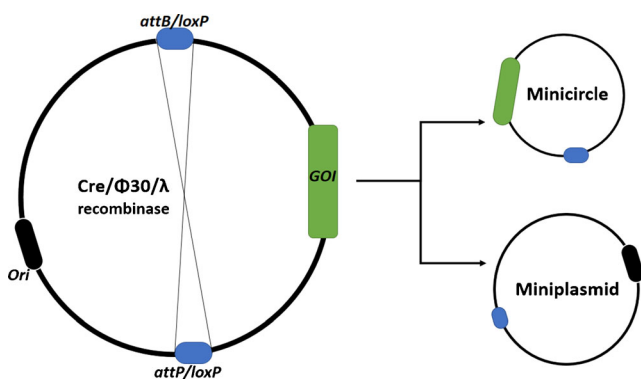
Fig 2 *The in vivo production of minicircle DNA vectors.* Vectors are generated by site-specific recombination via two excision sequences (*blue rounded rectangle*), which flank the mammalian gene of interest expression cassette (*green rounded rectangle*) and the bacterial amplification unit (*black rounded rectangle*) in a recombinant plasmid vector. *Arrows* indicate the direction of the recombination reaction. System employing the phage P1 Cre, or ΦC31 and λ Int, recombinase supplied by the bacterial host strain and the corresponding target sequences *loxP* or *attB* and *attP*, respectively

Phage N15 TelN protelomerase

The *E. coli* bacteriophage, N15, discovered by Ravin in 1964 in Moscow, has some unique features compared to other temperate bacteriophages with respect to its genetic composition and organization (Ravin et al. 2000). N15 not only represents the first

example of a linear prophage genome with covalently closed ends, but still remains the only known naturally occurring example of a linear plasmid in *E. coli* since its discovery (Valentin and Rybchin 1999). While N15 virions are very similar to phage lambda in morphology, burst size, plaque morphology, and length of genome, N15 prophage is very unusual in that it exists and autonomously replicates as a linear covalently closed (LCC) DNA plasmid in the lysogenic state. Like λ, the linear N15 mature packaged phage genome also possesses ssDNA cohesive ends, *cos*L and *cos*R. The linear plasmid prophage termini, *tel*R and *tel*L (*telRL*), are located near the center at position 24.8 kb on the mature N15 DNA map. The *telRL* site is located in a similar site that is occupied by the attachment site, *attP*, in λ and defines the boundary between the left and right arms (Ravin et al. 2000). In the mature N15 genome, the genes responsible for formation and maintenance of the linear plasmid in the host cell (protelomerase, anti-repressor, replication, and partitioning genes) are located on the left (0–10 kb) arm of the genome. This cluster has some overlap with genes involved in the lytic growth of phage although most lytic genes are located on the right arm. Upon deletion of the sequences between the left and right arms, forming a 5.2-kb mini plasmid product, the derivative is still capable of replication (Ravin 2011).

During the establishment of N15 lysogeny, the genome circularizes by annealing and ligating its terminal cohesive (*cos*) ends. The telomere-forming *telRL* site on the circular plasmid prophage becomes the target for N15-encoded protelomerase cleaving-joining enzyme (Deneke et al. 2002). The *telRL* sequence contains a 56-bp palindrome with two different base pairs at the 12th and 14th positions on both sides of the palindrome center (*telO*), which forms a 14-bp perfect palindrome recognized and cleaved by TelN protelomerase to create ssDNA termini. After cleavage, the self-complementary ssDNA ends reanneal and create perfect hairpin structures at each end (Valentin and Rybchin 1999).

TelN is a slightly acidic 630 a.a. protein that is about 72.2 kDa in size, and while essential for the generation and replication of a linear plasmid prophage, it also plays a significant role in the lytic growth of N15 phage (Deneke et al. 2002).

TelN and λ Int play a similar role in the creation of lysogeny. As previously mentioned, the C-terminal segments of integrases are highly conserved, particularly their catalytically active domains. Analytical studies on TelN structure revealed a region (residues 390–427) that is paralogous to the C-termini of integrases that belong to the tyrosine recombinase family. TelN possesses a tyrosine at the same position as integrases, which suggests that protelomerases and tyrosine recombinases possess similar catalytic mechanisms for DNA cleavage and ligation and likely evolved from a common ancestor. Both generate a 3′-phosphotyrosine DNA intermediate that enables the covalent rejoining of cleaved DNA strands without the use of a high-energy cofactors (Valentin and Rybchin 1999). TelN cleaving-joining activity is functional in the absence of any other N15-encoded factors, and purified TelN is the only protein required to convert circular DNA substrates possessing the *telRL* sequence to the linear molecule with covalently closed ends (Heinrich et al. 2002). In vitro studies on the functionality of purified TelN demonstrate that in the presence of the 56-bp *telRL* target site, and in particular the 22-bp *telO* target site, purified TelN processes circular and linear pDNA in vitro with the same covalently closed end DNA topology as was observed in vivo (Deneke et al. 2000). In addition, it was shown that the central telomerase occupancy site (*tos*), within the *telO* perfect palindrome, is occupied by two TelN molecules to create a stable TelN-*telRL* complex. The *telO* palindrome alone is not sufficient for specific stable complex formation and requires the additional sequences of *telRL* to stabilize the TelN-target complexes that process the binding reaction of TelN homodimers to the 3′-phosphoryl end of the cleaved strands and the generation of LCC DNA (Deneke et al. 2002).

TelN protelomerase represents a significant addition to the repertoire of exploitable enzymes in biotechnology with respect to its unique property in linearizing the plasmids carrying its target *telRL* site. The "pJAZZ" series of transcription-free, linear cloning vectors and "linear miniplasmids" are N15-based linear cloning systems that facilitate cloning of up to 30-kb DNA sequences, including AT-rich inserts or inverted repeats, which are very difficult to clone by standard methods into plasmid
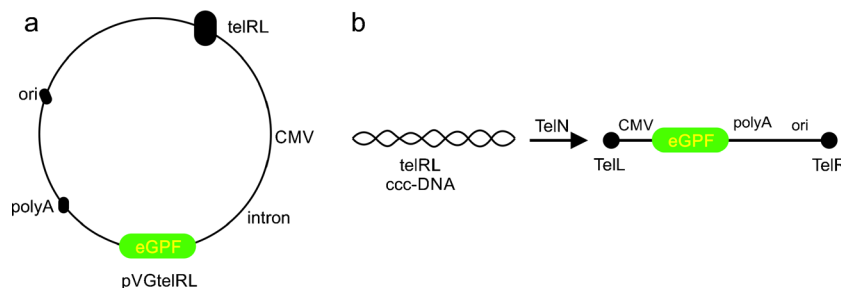

Fig. 3 *The in vitro application of TelN/telRL recombination system.* **a** Schematic presentation of eGFP expressing plasmid DNA carrying the telRL target site. **b** Schematic representation of the processing of the *telRL* substrate by recombinant TelN protein. TelN cuts the *telRL* site and produces a linear covalently closed (LCC) plasmid DNA (modified from Heinrich et al. 2002)

Fig. 4 *The phage PY54 pal sequence and the mechanism of conversion into closed ends.* The 42-bp palindrome is flanking, inverted repeat and the putative start codon are *boxed* (modified from Hertwig et al. 2003)

vectors (Ravin et al. 2000). TelN/*telRL* cleaving-joining activity has also been applied to enhance in vitro production of LCC DNA vectors (Fig. 3), where LCC DNA vectors demonstrate heightened and sustained expression of a gene of interest both in vitro and in vivo (Heinrich et al. 2002).

Phage PY54 Tel protelomerase

Bacteriophage PY54 was first isolated in 2003 from *Yersinia enterocolitica* as a temperate phage, belonging to the *Siphoviridae* family with a 46-kb dsDNA genome (Stefan et al. 2003). Like N15, PY54 also exhibits a λ-like morphology with similar genome size and is a temperate phage that coexists with the infected host as a linear plasmid with similar telomere-like hairpin ends. However, despite these similarities, these two phages are evolutionary quite distant, where N15 is far more closely related to λ than to PY54 (Stefan et al. 2003).

A paralog of N15 *telN*, the PY54 *tel* gene encodes a 77-kDa TelN-like protelomerase with observably identical function and is able to process recombinant plasmids containing the 42-bp *telRL*-like palindrome (*pal*) (Stefan et al. 2003). The *pal* site (Fig. 4) is a 42-bp perfect palindrome, that unlike N15 *telRL*, possesses only partial function in vivo in the absence of adjacent sequences (Stefan et al. 2003). Sequencing and nucleotide BLAST of *telRL* and *pal* indicates that the 10-nucleotide sequence, 5′-TACGCGCGTA-3′, in the center of the paralogous palindromes are 100 % identical. Furthermore, the protein BLAST alignment of TelN and Tel also reveals 60 % similarity (Fig. 5).

To date, there are no in vivo applications of the Tel-*pal* system of PY54 such as the construction of Tel-expressing cells in genome engineering, or in the production of new pDNA vectors that exploit this system. While the generated LCC and mini LCC DNA vectors have been employing the TelN-*telRL* system, enzyme production was limited to in vitro use and LCC DNA processing. We are currently applying the

Fig. 5 *Protein sequence alignment of the PY54 and N15 protelomerases.* The amino acid alignment of N15 TelN and PY54 Tel proteins showing ~60 % a.a. identity (*black box*) and similarity in *gray box*

N15 and particularly the PY54 recombination systems to develop a novel in vivo technology for production of superior bacterial sequence-free mini DNA vectors for transgene delivery. LCC DNA vectors may be superior to conventional CCC DNA vectors in that they are more efficient and likely safer. The obvious drawback to LCC DNA vectors is the multistep approach to production that adds expense and makes them less desirable in clinical trials due to the fact that pDNA needs to be purified from bacterial cells, digested, chemically modified to produce the LCC DNA, and then finally purified as LCC vector. We have developed an in vivo, cost-effective, one-step, and robust technology for the production of LCC transgene expression cassettes, called "DNA ministring" vectors. This technology exploits the phage PY54 recombination system in comparison to N15 (Nafissi and Slavcev 2012). The PY54-derived Tel-pal system relies only upon proteolomerase activity and the palindromic target site, offering the major advantage of a single enzyme and a single-step cleaving-joining enzymatic reaction that can be manipulated for unlimited in vivo production of mini LCC DNA vectors.

Conclusion and perspective

In recent decades, several site-specific recombinases have been discovered from various microorganisms and applied to in vivo genome engineering for a wide variety of heterologous cells. These enzymes have been classified into tyrosine-type and serine-type recombinases according to their catalytic mechanisms, where both families are further subdivided into two subgroups identified (see Table 1). Tyrosine-type phage integrases, such as λ integrase, catalyze tightly regulated unidirectional recombination, but due to their characteristic bidirectional recombination in heterologous cells, their application to chromosomal insertions in eukaryotes remains limited. In contrast, the tyrosine-type simple recombinases, such as Cre and Flp, catalyze bidirectional recombination and have been widely used as tools for gene integration into heterologous genomes by development of RMCE and LE/RE mutant site systems. Serine-type phage integrases, such as ΦC31 integrase catalyze unidirectional recombination between short attachment sites with no need for additional host-encoded proteins, cofactors, or specific DNA structures, which, due to simplicity, favors their use in the development of human gene integration systems. Homologous recombination is generally favored for site-specific introduction of foreign DNAs into a wide variety of heterologous genomes, but the efficiency of homologous recombination is dependent on the host cell features and DNA structure of targeting region. In addition to the abovementioned methods, the "restriction-modification (R-M) and CRISPR-Cas systems" could be a method of choice in the future for more policed recombination in mammalian cells. The R-M and CRISPR-Cas systems is a

defending mechanism for bacterial cells against invading phages (Dupuis et al. 2013).

References

Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7:649–659

Araki K, Araki M, Yamamura K-i (1997) Targeted integration of DNA using mutant lox sites in embryonic stem cells. Nucleic Acids Res 25:868–872

Austin S, Ziese M, Sternberg N (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. Cell 25:729–736

Bigger BW, Tolmachov O, Collombet J-M, Fragkos M, Palaszewski I, Coutelle C (2001) An araC-controlled bacterial cre expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy. J Biol Chem 276:23018–23027

Bouhassira EE, Westerman K, Leboulch P (1997) Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange. Blood 90:3332–3344

Broach JR, Guarascio VR, Jayaram M (1982) Recombination within the yeast plasmid 2 µ circle is site-specific. Cell 29:227–234

Brown WRA, Lee NCO, Xu Z, Smith M (2011) Serine recombinases as tools for genome engineering. Methods 53:372–379

Chavez LC, Calos PM (2011) Therapeutic applications of the PhiC31 integrase system. Cur Gene Ther 11:375–381

Christensen A (2001) Bacteriophage lambda-based expression vectors. Mol Biotechnol 17:219–224

Clokie MJ, Kropinski A, Thomason L, Oppenheim A and Court D (2009) Modifying bacteriophage lambda with recombineering. In: Bacteriophages. pp. 239–251: Humana, Totowa

Darquet AM, Cameron B, Wils P, Scherman D, Crouzet J (1997) A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. Gene Ther 4:1341

Darquet AM, Kreiss P, Schwartz B, Naimi S, Delaere P, Crouzet J, Scherman D (1999) Minicircle: an improved DNA molecule for in vitro and in vivo gene transfer. Nature 6:209–218

Deneke J, Ziegelin GN, Lurz R, Lanka E (2000) The proteolomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. Proc Natl Acad Sci U S A 97:7721–7726

Deneke J, Ziegelin G, Lurz R, Lanka E (2002) Phage N15 telomere resolution target requirements for recognition and processing by the proteolomerase. J Biol Chem 277:10410–10419

Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev 72:642–671

Dupuis M-ÀV, Villion M, Magadán AH, Moineau S (2013) CRISPR-Cas and restriction modification systems are compatible and increase phage resistance. Nat Commun 4:2087

Gaj T, Mercer AC, Gersbach CA, Gordley RM, Barbas CF (2011) Structure-guided reprogramming of serine recombinase DNA sequence specificity. Proc Natl Acad Sci U S A 108:498–503

Groth AC, Calos MP (2004) Phage integrases: biology and applications. J Mol Biol 335:667–678

Groth AC, Olivares EC, Thyagarajan B, Calos MP (2000) A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci U S A 97:5995–6000

Haldimann A, Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol 183:6384–6393

Hammerl JA, Klein I, Appel B, Hertwig S (2007) Interplay between the temperate phages PY54 and N15, linear plasmid prophages with covalently closed ends. J Bacteriol 189:8366–8370

Heinrich J, Schultz J, Bosse M, Ziegelin G, Lanka E, Moelling K (2002) Linear closed mini DNA generated by the prokaryotic cleaving-joining enzyme TelN is functional in mammalian cells. J Mol Med 80:648–654

Hertwig S, Klein I, Schmidt V, Beck S, Hammerl JA, Appel B (2003) Sequence analysis of the genome of the temperate *Yersinia enterocolitica* Phage PY54. J Mol Biol 331:605–622

Hirano N, Muroi T, Takahashi H, Haruki M (2011) Site-specific recombinases as tools for heterologous gene integration. Appl Microbiol Biotechnol 92:227–239

Kay MA, He C-Y, Chen Z-Y (2010) A robust system for production of minicircle DNA vectors. Nat Biotechnol 28:1287–1289

Keravala A, Chavez CL, Hu G, Woodard LE, Monahan PE, Calos MP (2011) Long-term phenotypic correction in factor IX knockout mice by using phiC31 integrase-mediated, gene therapy. Gene Ther 18: 842–848

Kuhlman TE, Cox EC (2010) Site-specific chromosomal integration of large synthetic constructs. Nucleic Acids Res 38:e92–e92

Kuhstoss S, Rao RN (1991) Analysis of the integration function of the streptomycete bacteriophage φC31. J Mol Biol 222:897–908

Le Borgne S, Palmeros BI, Valle F, Bolivar F, Gosset G (1998) pBRINT-Ts: a plasmid family with a temperature-sensitive replicon, designed for chromosomal integration into the *lacZ* gene of *Escherichia coli*. Gene 223:213–219

Lorbach E, Christ N, Schwikardi M, Dröge P (2000) Site-specific recombination in human cells catalyzed by phage λ integrase mutants. J Mol Biol 296:1175–1181

Murphy KC (1998) Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. J Bacteriol 180: 2063–2071

Murphy KC, Magorzata A, and Wacaw S (2012) Phage recombinases and their applications. In: Advances in Virus Research pp. 367–414: Academic, San Diego

Nafissi N, Slavcev R (2012) Construction and characterization of an in-vivo linear covalently closed DNA vector production system. Microb Cell Factories 11:154

Porteus M (2012) Homologous recombination-based gene therapy for the primary immunodeficiencies. Ann NY Acad Sci 1246:131–140

Poteete AR (2001) What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol Lett 201:9–14

Ravin NV (2011) N15: the linear phage-plasmid. Plasmid 65:102–109

Ravin V, Ravin N, Casjens S, Ford ME, Hatfull GF, Hendrix RW (2000) Genomic sequence and analysis of the atypical temperate bacteriophage N15. J Mol Biol 299:53–73

Rice P, Mouw KW, Montaño SP, Boocock M, Rowland S-J, Stark WM (2010) Orchestrating serine resolvases. Biochem Soc Trans 38:384

Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB (1982) Nucleotide sequence of bacteriophage λ DNA. J Mol Biol 162:729–773

Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci U S A 85:5166–5170

Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc 4:206–223

Stahl FW (1998) Recombination in phage lambda: one geneticist's historical perspective. Gene 223:95–102

Stefan H, Iris K, Rudi L, Erich L, Bernd A (2003) PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. Mol Microbiol 48:989–1003

Suzuki N, Inui M, Yukawa H (2007) Site-directed integration system using a combination of mutant lox sites for *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 77:871–878

Thomason L, Court DL, Bubunenko M, Costantino N, Wilson H, Datta S, Oppenheim A (2001) Recombineering: genetic engineering in bacteria using homologous recombination. In Current Protocols in Molecular Biology. Wiley, New York

Thyagarajan B, Guimaraes MJ, Groth AC, Calos MP (2000) Mammalian genomes contain active recombinase recognition sites. Gene 244: 47–54

Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP (2001) Site-specific genomic integration in mammalian cells mediated by phage [phgr] C31 integrase. Mol Cell Biol 21:3926–3934

Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. J Mol Biol 407:193–221

Valentin N, Rybchin ANS (1999) The plasmid prophage N15: a linear DNA with covalently closed ends. Mol Microbiol 33:895–903

Wigley DB (2013) Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat Rev Microbiol 11:9–13