Review

Genetic recombination in bacteriophage lambda

Christopher R. T. Hillyar*

School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK.

* Corresponding author: Cancer Research UK/MRC Gray Institute for Radiation Oncology and Biology, Churchill Hospital, University of Oxford, Old Road Campus Research Building, Off Roosevelt Drive, Oxford OX3 7DQ, UK. Tel: +44 07500978173. Email: christopher.hillyar@oncology.ox.ac.uk

Supervisor: Dr Gary J. Sharples, School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK.

Bacteriophage λ (lambda) infects Escherichia coli and induces a dramatically increased rate and altered profile of genetic recombination as part of the cycle of infection. The genetic recombination processes augment the host-encoded recombination proteins with phage-encoded recombination proteins, promoting particular recombination pathways. This review characterizes the protein machinery involved in the most important processes underlying λ-mediated genetic recombination.

Key words: recombination, lambda, orf, Exo, beta, Rap

Submitted on October 2011; accepted on January 2012

Introduction

Genetic recombination is defined as the rearrangement of genes or parts of genes. Throughout evolution, the acquisition of ‘new’ genes and the rearrangement of ‘old’ ones has been driven by genetic recombination between bacteriophage and bacterial genomes. In Escherichia coli, genetic recombination is carried out ‘cautiously’ through carefully regulated mechanisms. Although genetic recombination may be highly active in repairing DNA damage during periods of stress or in promoting genetic diversity, uncontrolled recombination causes genomic instability. This relatively ‘static’ recombination mechanism shifts dramatically when E. coli is infected by the temperate bacteriophage λ (lambda) (Sawitzke and Stahl, 1994). Each and every λ progeny undergoes at least one recombination event, despite possessing a tiny chromosome of only 50 kb in length. Thus, the bacterium is rendered a hotbed of genetic exchange in its last few minutes of existence during the natural history of the cycle of infection (Poteete, 2001). The highly efficient rate of recombination induced by λ is called ‘hyper rec’ and requires just 40 bp of homology (Morimatsu and Kowalczykowski, 2003). The λ-encoded genetic recombination protein machinery termed the Red system has been shown by Zagursky and Hays (1983) to be harmless to E. coli if constitutively expressed from a multicopy plasmid. The Red system eliminates the dependence of λ-mediated recombination on many of the host E. coli recombination proteins. In contrast to the E. coli RecA protein, the Red system does not mediate unwanted recombination in bacterial artificial chromosomes. For this reason, Red-mediated processes complement conventional cloning technologies, such as restriction enzymes and DNA ligase. Examination of the Red system has enhanced the versatility of genetic engineering (Court, Sawitzke and Thomason, 2002). Here, the ‘double-strand invasion’ and ‘single-strand annealing’ pathways underpinning λ-mediated genetic recombination are discussed. These Red-mediated pathways proceed in a manner that is RecA-dependent and RecA-independent, respectively. Other λ-encoded proteins, viz. Orf and Rap, contribute to the early and end stages of these reactions. A critical feature of the λ recombination proteins is their ability to mediate genetic recombination efficiently even in the absence of the E. coli RecA protein (Court, Sawitzke and Thomason, 2002). Another important point that should be noted is that recombinant progeny may be generated from <50 bp of identity, analogous to the more complex recombination systems in eukaryotic cells (Court, Sawitzke and Thomason, 2002).

Genetic recombination in bacteriophage λ

During rolling circle replication, each λ progeny acquires a double-strand break at cos (the cohesive site). If a cut at cos is
the only double-strand-break available, λ recombination is generally RecA-dependent (Tarkowski et al., 2002). RecA-dependent genetic recombination with a linear λ chromosome is a ‘break-join’ mechanism. This mechanism involves the λ exonuclease Exo, which degrades the 5′-ending strand of duplex DNA, while the 3′-ending strand is preserved, generating a 3′-ssDNA (single-stranded DNA) overhang. Subsequently, the E. coli RecA protein binds to the processed 5′-ssDNA-overhang generated by Exo. This step allows the assembly of multiple RecA monomers that form a presynaptic filament on ssDNA. Invasion of the filament into a homologous duplex results in synapse formation via a replication-dependent double-strand invasion recombination pathway (Fig. 1A) (Stahl et al., 1997). In the presence of a homologous partner ssDNA, however, λ may mediate recombination with Exo-processed substrates via a RecA-independent single-strand annealing pathway. This requires the λ-encoded ssDNA-binding protein Bet, which anneals partner ssDNA (Fig. 1B) (Court, Sawitzke and Thomason, 2002). Thus, the single-strand annealing and double-strand invasion pathways are distinct paradigms for generating recombinant progeny. The λ-encoded proteins involved in these reactions are analogous to the E. coli RecBCD enzyme (Red system), the RecFOR enzyme (Orf) and the RuvC resolvase (Rap) (Webb, Cox and Inman, 1997; Court, Sawitzke and Thomason, 2002). The absence of RecFOR, Orf is required for early events of recombinational exchange (Sawitzke and Stahl, 1992, 1994; Poteete, 2004, 2008; Rybalchenko et al., 2004). The Rap resolvase targets recombination intermediates such as Holliday junctions, which arise as a consequence of host restriction, terminase cleavage, DNA damage, rolling circle replication or replication fork collapse (Court, Sawitzke and Thomason, 2002; Sharples et al., 2004).

**The Red system**

In E. coli, a mutation in the recA gene eliminates the host RecBCD and RecFOR pathways. Nevertheless, λ recombination is as efficient in recA mutants as in wild-type E. coli. In contrast, Red (recombination-deficient) gene knockout λ phage is partially recombination-deficient in wild-type E. coli, but grossly defective in recA mutants (Gingery and Echols, 1968; Singer and Weil, 1968). This finding suggests that the red genes play a critical homologous recombination role. The Red genes are clustered in the PL operon of the λ genome. Expression is controlled by the CI repressor. The red genes include exo (or redA), bet (or redB) or, simply, β and gam (or γ) (Poteete, 2001). The exo, bet and gam genes encode the 24 kDa exonuclease Exo, the 28 kDa ssDNA-binding protein Bet and the 16 kDa polypeptide Gam, respectively (Muniyappa and Radding, 1986). Because Exo and Bet were originally found to co-purify together and defects in Bet cause defects in Exo, these proteins are thought to form a functional complex. Gam confers protection against nuclease attack of linear dsDNA (double-stranded DNA) by inhibiting the E. coli SbcCD and RecBCD enzymes, perhaps via a conserved binding site (Court, Sawitzke and Thomason, 2002).

**The Exo nuclease loads Bet onto processed ssDNA**

The exonuclease Exo is a toroidal ring-shaped homotrimer with a 3-fold symmetry. Each monomer contains an αβ fold and a metal ion-binding active site (Zhang, McCabe and Bell, 2011). The central channel ranges from 15 to 30 Å in diameter (Kovall and Matthers, 1997) and accommodates dsDNA at the wide end and processed 3′-ending ssDNA at the narrow end (Court, Sawitzke and Thomason, 2002). Although, like DNA, the ‘outside’ of the protein is negatively charged, this charge is distributed away from the central channel, facilitating ssDNA binding through the channel (Kovall and Matthers, 1997). When bound to DNA, Exo is asymmetrically tilted such that the 5′-ending strand binds to an active site in one of the three subunits. DNA binding is non-sequence-specific, involves hydrogen bonding with the DNA backbone and causes no significant conformational changes in three-dimensional structure of the protein (Zhang, McCabe and Bell, 2011). Exo is able to resect the 5′-end of dsDNA and forms a 3′-ending overhang in a similar manner to the E. coli Exonucleases I–III (Little, 1967; Court, Sawitzke and Thomason, 2002). Arg45 has been proposed to act as the ‘rudder’ keeping the enzyme on track. A hydrophobic wedge comprising Val73, Ala75, Ala77 and Leu78 splits apart the base pairs as the enzyme processes along the DNA. This mechanism requires two Mg2+ metal ions, spaced 4 Å apart. Arg28 forms a positively charged pocket that binds to the 5′-phosphate of the DNA, pulling the DNA into the active site. The active site comprises Glu85, Asp119 and Lys131. Mg2+-dependent digestion at 37°C occurs at a rate of ~30
Bet is required for RecA-independent single-strand annealing

The ssDNA-binding protein Bet has Mg²⁺-dependent and adenosine triphosphate (ATP)-independent ssDNA annealing properties. Bet loads onto the 3′-end of Exo-processed ssDNA due to 3′-5′ polar cooperativity. This activity allows Bet to spontaneously form ring structures or helical filaments on >35 nucleotides long ssDNA in vitro (Court, Sawitzke and Thomason, 2002). The half-time for Bet-mediated renaturation, like RecA protein, is independent of ssDNA concentration. This feature contrasts with the renaturation reactions mediated by the E. coli single-stranded binding protein (SSB; Muniyappa and Radding, 1986) and is a second-order reaction that allows Bet to form filaments on dsDNA analogous to the eukaryotic Rad52 protein (Court, Sawitzke and Thomason, 2002). Thus, Bet promotes renaturation of homologous ssDNA partner sequences that have been processed by Exo preferentially at sites rich in AT base pairs, while the RecA protein prefers GC-rich sequences (Muniyappa and Radding, 1986; Rybalchenko et al., 2004; Maxwell et al., 2005). The strand exchange activities of Bet may extend to a homologous duplex with a single-strand gap, allowing displacement of the complementary strand with the free energy of the annealing reaction. Whether these reactions require RecA is unclear (Court, Sawitzke and Thomason, 2002). Bet remains bound to recombinant dsDNA after ssDNA assimilation, although Bet does not bind to ssDNA directly. This has been suggested to confer resistance against nuclease attack (Muniyappa and Radding, 1986). Because Bet forms a complex with Exo in vitro, Court, Sawitzke and Thomason (2002) suggested that Bet modulates the nucleolytic and recombinational activities of Exo through a direct mechanism. This interaction may allow Bet to protect recombination intermediates from degradation by inhibiting Exo. Thus, the main function of Bet is to promote 3′-end assimilation, whereas preferentially presenting 5′-ends to Exo for degradation (Court, Sawitzke and Thomason, 2002; Poteete, 2004; Rybalchenko et al., 2004).

Gam protects recombination intermediates from nuclease attack

Gam is a 276-amino acid homodimer that binds to and inhibits the nuclease activity of E. coli-encoded nucleases (Poteete, 2001). Gam inhibits all the known activities of the E. coli RecBCD enzyme, including its helicase, nuclease and ATPase activity. Expression of Gam confers the types of defects associated with RecBCD mutants. The inactive RecBCD-Gam complex is thought to exist in an equilibrium between free and complex forms, allowing some portion of the enzyme to exist in active state (Murphey, 2007). Nevertheless, Gam prevents RecBCD-mediated degradation of linear duplexes in vitro (Court, Sawitzke and Thomason, 2002; Mosberg, Lajoie and Church, 2010). Indeed, RecBCD mutants and RecBCD-Gam recombination both produce DNA concatemers during λ replication in vitro (Poteete, Fenton and Murphey, 1998). Inhibition of RecBCD likely occurs by Gam preventing RecBCD from binding to dsDNA ends. The mode of Gam binding to RecBCD is thought to involve disruption of amino acid residues in RecB that are important for DNA binding. Addition of Gam to RecBCD–dsDNA complexes has been shown to lead to complex dissociation, suggesting that Gam interferes with the RecBCD DNA-binding site even when that site is occupied (Murphey, 2007). Interestingly, Gam has been reported to confer radiotolerance to X-rays, but sensitivity to ultraviolet (UV). The reason for this paradox is likely to centre on the types of DNA lesions produced by these modes of ionizing radiation. UV-induced lesions tend to be ssDNA daughter-strand gaps caused by blockage of the replisome at pyrimidine dimers or abasic sites. Only occasionally do these lesions generate dsDNA ends that are processed by RecBCD. In contrast, the proportion of dsDNA breaks generated by X-rays is thought to be higher, because X-rays produce a higher proportion of clustered single-strand breaks that become ‘frank’ double-strand breaks that are processed by RecBCD. Therefore, Gam prevents the RecBCD-mediated digestion of dsDNA that may be repaired by genetic recombination (Murphey, 2007). RecBCD-mediated reactions are usually destructive, until the activity of the enzyme is modified by the χ recombination hotspot.

Red-mediated genetic recombination mechanisms

Bacteriophage λ induces the Red system to engage in efficient homologous recombination in E. coli recABC mutants (Murphey, 1998). Red-mediated recombination produces ssDNA 3′-overhangs from double-strand breaks through Bet-modulated Exo activity (Poteete, 2001). According to the
Court mechanism, this leads either to strand invasion dependent on RecA or Bet-mediated strand annealing with a suitable ssDNA substrate independent of RecA. In the absence of RecA, the Red system is unable to recombine a linear DNA duplex with a non-replicating DNA circle, but can recombine linear DNA with the bacterial chromosome (Court, Sawitzke and Thomason, 2002). The dependence on RecA, RecFOR and RecG also varies in different E. coli strains, suggesting that the Red system does not simply substitute for RecBCD in promoting λ and bacterial recombination (Poteete, 2008). By replacing the recC–ptr–recB–recD gene cluster in E. coli with the red genes, it has been demonstrated that most host recombination genes can be knocked out while maintaining host viability (Poteete, 2001).

Furthermore, a number of researchers have suggested that the Red system can mediate strand invasion reactions that are RecA-independent if the target chromosome is replicating (Luksi-Deluca and Kolodner, 1992; Kussano et al., 1994; Silberstein, Tzifati and Cohen, 1995; Ellis et al., 2001; Rybalchenko et al., 2004). Knocking out the recf and recG genes increases the efficiency of the Red system in recBCD mutants of E. coli (Poteete, 2001). This led Tarkowski et al. (2002) to propose a model of Red-mediated recombination in which RecJ degrades Red-processed recombination intermediates in the absence of Rap. In contrast, Poteete and colleagues implicated the Red system in ‘template switching’ during replication. The Poteete mechanism involves Exo-mediated processing of dsDNA. Subsequently, Bet binds to the 3′-ssDNA end and mediates annealing with a homologous partner sequence in a lagging template strand, displacing the original template and inducing a template switch (Poteete, 2008).

More recently, Mosberg, Lajoie and Church (2010) proposed a non-mutually exclusive mechanism which utilizes a single-strand intermediate to repair single-strand gaps in a replicating lagging strand. This mechanism may proceed in parallel with the Court and Poteete mechanisms. Mosberg proposed that the dsDNA molecule is completely processed by one Exo enzyme before a second can bind, in contrast to preceding pathways in which both ends of the dsDNA molecule were processed by two Exo enzymes. It is thought that the Mosberg mechanism is feasible due to the high processivity of the Exo endonuclease. The viability of a single-strand being a recombinogenic intermediate with lagging-strand bias was validated in vitro. It should be noted that the viability of this reaction in vivo is not readily calculable (Mosberg, Lajoie and Church, 2010).

Orf plays an analogous role in genetic recombination to the E. coli RecFOR enzyme

The λ orf gene was originally discovered by Sawitzke and Stahl (1992) as an open-reading frame mapping to ORF146 (ninB) that conferred a gain of genetic recombination function of the RecFOR pathway. Subsequently, Sawitzke and Stahl (1992) named the gene product Orf for recQ, recE-like functions, notwithstanding minimal sequence similarity to the recFOR genes. Further experimental observations showed Orf expression in recBC, sbcB, sbcC and recO mutants from a multicopy plasmid reduced viability 3- to 8-fold, while a lac repressor supplied in trans from a compatible plasmid conferred non-lethal expression (Sawitzke and Stahl, 1994). In E. coli strains dependent on the Red system for recombination, Tarkowski et al. (2002) noticed a 58% reduction in recombinants in strains crossed with λ lacking the nin5 region, which contains the orf and rap genes. This suggested that orf contributed in some way to Red recombination. Sawitzke and Stahl (1997) noticed that in crosses where E. coli was dependent on the Red system and the infecting λ phage was deficient for the red genes, recombination was dependent on the RecFOR pathway and required RecA and RecJ, but not RecFOR, RecQ, RuvA nor RuvB. This indicated that orf expressed a function that could substitute for at least some components of the RecFOR pathway. Other observations showed a trans-acting plasmid-borne orf gene conferred partial RecFOR independence in the absence of the Red system in recBC and sbcBC mutants (Sawitzke and Stahl, 1992). Poteete and Fenton (2000) showed constitutive expression of Orf-reduced RecFOR dependence in E. coli with the RecC–ptr–RecB–RecD gene cluster replaced with the Red system, while the same background not expressing Orf showed a growth retardation effect. Moreover, Sawitzke and Stahl (1997) suggested Orf influenced the initial phase of recombination by substituting for RecFOR in the presence of the Red system to mediate RecFOR-independent Red recombination. Thus, Orf is partially required for Red recombination and functions as a RecFOR analogue in Red-mediated recombination.

Ninety-eight percent of λ recombination events occur near the cos when RecFOR and orf is expressed, while the absence of orf promotes increased ‘unfocused’ recombination events. Thus, Orf appears to promote ‘focused’ genetic recombination at the cos (Sawitzke and Stahl, 1994, 1997; Poteete, 2004). Although Orf is not fully required for Red-mediated recombination, it nevertheless participates in recBC and sbcBC backgrounds (possibly in cooperation with RecFOR), promoting strand invasion reactions to attenuate the production of potentially lethal linear multimers that inhibit recombination by competing for short-supply recombination machinery (Poteete, 2004). Orf restored UV resistance equally in RecFOR, RuvC and RuvAB mutant strains, suggesting Orf partly substitutes for RuvC and RuvAB (Poteete, 2004). Poteete (2004) suggested that Orf may modify RecFOR-like and RuvABC-like proteins or modify some other protein complexes rendering them RecFOR- and RuvABC-like. Furthermore, these authors suggest that Orf may also modulate SSB activity.

Orf’s crystal structure and function

Orf is a basic (pI 8.94) 146-amino acid protein (16.6 kDa), which belongs to the α+β protein family (Maxwell et al.,...
The 2.5 Å resolution crystal structure shows a 33 kDa asymmetric ring including residues 1–141 (monomer A) and 1–133 (monomer B) (Maxwell et al., 2005). Each monomer contains 19 lysine residues (comprising the central channel and binding cleft) and 7 semi-conserved tryptophans (Curtis et al., 2010). Maxwell et al. (2005) showed two interacting regions determined dimerization: residues 3–9 (β1), 14–22 (αA) and 26–38 (β2 loop and β2), and 67–80 (αC), 83–89 (β3) and 95–97 (β4). Comprising 20% of the surface of the A and B monomers (1407 Å2), this interface was enriched with nine hydrophobic residues (Ile8, Val21, Val33, Ile36, Val75, Val86, Leu89, Val95, Ile97) (Maxwell et al., 2005). A twist at residues Asn40 to Ser42 in the backbone created asymmetry. In monomer A, there was a fifth C-terminal α-helix after the fourth. In contrast, there was a random coil C-terminus until residue 133 after the fourth α-helix in monomer B (Maxwell et al., 2005). The flexibility of monomer B's C-terminus, coupled to the importance of Arg140 in binding to ssDNA, suggested that this C-terminus might act as a clamp when forming an Orf-ssDNA complex (Curtis et al., 2010).

Furthermore, Lys3A/B, Arg41A, Lys73A/B and Lys81A of the dimer render the interior of the central channel positively charged, possibly for ssDNA binding. This central channel was thought too narrow (20–8 Å) to accommodate ssDNA but no change in the ratio of frequencies of patch polarities compared with wild-type crosses (Gumbiner-Russo and Rosenberg, 2007). Therefore, Orf may equally promote either 5′–3′ or 3′–5′ (inverse) RecA loading.

**Orf’s mode of DNA binding**

Orf has been shown to bind biphasically to ssDNA in preference to dsDNA by electrophoretic mobility shift assay and stopped flow fluorescence spectroscopy. The initial 10 s rapid binding phase was followed by a slower 20 s binding phase. Orf showed no preference in binding DNA substrates carrying a 5′- or 3′-single-strand overhang or the intersection between ss- and dsDNA (Maxwell et al., 2005). To determine whether the C-terminal domain of the Orf monomer was involved in ssDNA binding, Curtis et al. (2010) used versions of the Orf protein carrying the W141F point mutation and ΔC6 and ΔC19 C-terminal truncations in mobility shift assays. The ΔC19 truncation completely abolished dsDNA-binding ability, suggesting the C-terminus conferred DNA binding. Two conserved arginine residues (R132 and R140) are probably important for binding to the phosphodiester backbone of ssDNA (Curtis et al., 2010).

**Orf’s interactions with other proteins**

Because *E. coli* RecFOR enzyme binds SSB, facilitating RecA loading onto SSB-coated ssDNA, and Orf is thought of as analogous to RecFOR, this raised the possibility that Orf may also bind SSB. Curtis et al. (2010) showed that Orf interacts with SSB and observed no reduction in binding using two C-terminal point mutations (SSB133, P176S) and a version of SSB lacking the last 10 residues completely (ΔC10). This Orf–SSB interaction, therefore, is independent of SSB's C-terminus. This contrasts with Exol, PrIA, PolIII, PolV, RecG, RecO, RecQ, Udg, topoisomerase III and the γ subunit of PolII, that target the conserved acidic tail of SSB (Curtis et al., 2010). Nevertheless, taken with Orf’s role as a substitute for RecFOR, this suggests that Orf may interact with SSB to overcome SSB’s inhibitory effect on RecA-dependent Red-mediated recombination. This interaction with SSB raised further possibilities that Orf may mediate RecA loading. In *E. coli* strains dependent on the Red system, expressing orf increased Red recombination compared with similar strains lacking the orf transgene, while *E. coli* mutants dependent on Red recombination, lacking orf, but expressing recA803—which confers constitutive RecA-mediated displacement of SSB from ssDNA—were unable to facilitate RecFOR-independent recombination (Potete, 2004). This suggested that Orf promoted RecA loading during Red recombination by displacing Bet, not SSB. Thus, Potete (2004) suggested the sequential ‘Bet-RecA’ hypothesis to account for this RecA-dependent salvage pathway that was active only when Red-mediated strand annealing was blocked due to the absence of ssDNA partner sequences. These extra capabilities, which orf confers to wild-type *E. coli* strains, do not influence patch polarity. λ crosses lacking orf and Rap showed reduced amounts of 5′- and 3′-recombinant patches, but no change in the ratio of frequencies of patch polarities compared with wild-type crosses (Gumbiner-Russo and Rosenberg, 2007). Therefore, Orf may equally promote either 5′–3′ or 3′–5′ (inverse) RecA loading.

**Rap**

The rap (recombination adept with plasmid or ninG) gene resides in the ninR region and encodes the Rap protein, a homodimeric ion-dependent endonuclease promoting the targeting of recombination intermediates (Sharples, Corbet and Graham, 1998). Both RecBCD- and Red-dependent recombination are reduced in rap mutants, a defect which is partially suppressed in recJ mutant backgrounds. In ruvC mutants, Rap partially substitutes for RuvC (Potete, 2004). Rap cleaves 5′-·C·C dinucleotides in D-loop and Holliday junction structures (Sharples et al., 2004), increasing λ-plasmid recombination catalysed by the Red system or the host RecBCD enzyme (Tarkowski et al., 2002).

**Conclusion**

In conclusion, the temperate bacteriophage λ, a pre-existing cellular structure or metabolic processes, has ‘learned’ partial independence from its prokaryotic host, *E. coli* (Hendrix et al., 2000). The absolute majority of genomes on Earth (10e6), these dsDNA-containing tailed phages outnumber bacteria 10-fold. Approximately 4500 pervasively mosaic phages infect a huge diversity of bacteria (Hendrix, 2002, 2003). Bacteriophage-bacterial transduction of *Shigella* Stx toxin has generated the enterohaemorrhagic *E. coli* strain O157:H7, a novel human pathogen associated with haemolytic-uremic syndrome (Zhou et al., 2010). Thus, genomic diversity and
virulence generated by promiscuous recombination and extensive horizontal genetic exchange at legitimate and illegitimate sites between bacteriophages and bacteria has important implications for animal and human health and disease (Fishers, Hofreuter and Haas, 2001; Hendrix, 2003). Genetic recombination in bacteriophage λ should be further studied, as this model system continues to inform our understanding of the mechanisms underpinning these complex genetic exchanges that drive ecological evolution (Friedman and Court, 2001).

Author biography

Christopher Hillyar completed his undergraduate degree in BSc (Hons) Biomedical Sciences with Placement Year at Durham University. Graduating top of his class, Christopher was awarded the 2011 Institute of Biomedical Sciences Northern Region Symposium Prize and received a nomination for a European Science, Engineering and Technology Award. In sandwich, he worked in industry at Reckitt and still doing its thing.

References


