Xer Site-Specific Recombination: Promoting Vertical and Horizontal Transmission of Genetic Information

CAROLINE MIDONET1 and FRANCOIS-XAVIER BARRE1

1Institute for Integrative Biology of the Cell (I2BC), Université Paris Saclay, CEA, CNRS, Université Paris Sud, 1 avenue de la Terrasse, 91198 Gif sur Yvette, France

ABSTRACT Two related tyrosine recombinases, XerC and XerD, are encoded in the genome of most bacteria where they serve to resolve dimers of circular chromosomes by the addition of a crossover at a specific site, dif. From a structural and biochemical point of view they belong to the Cre resolvase family of tyrosine recombinases. Correspondingly, they are exploited for the resolution of multimers of numerous plasmids. In addition, they are exploited by mobile DNA elements to integrate into the genome of their host. Exploitation of Xer is likely to be advantageous to mobile elements because the conservation of the Xer recombinases and of the sequence of their chromosomal target should permit a quite easy extension of their host range. However, it requires means to overcome the cellular mechanisms that normally restrict recombination to dif sites harbored by a chromosome dimer and, in the case of integrative mobile elements, to convert dedicated tyrosine resolvases into integrases.

INTRODUCTION

It was Barbara McClintock who first described the problems of segregation arising from the circularity of chromosomes during her studies on maize variegation (1). The importance of this observation, which could have passed as a mere oddity at the time because of the linear nature of chromosomes in Eukaryota, was only realized after the demonstration of the circular nature of the Escherichia coli chromosome by François Jacob and Elie Wollman in the 1960s (2). Since then, the wealth of information gained by genomic studies has shown that circular chromosomes are the norm in Bacteria and Archaea.

DNA circularity can result in the formation of two major topological threats for the segregation of genetic information at the time of cell division: sister chromosome catenation and concatenation [Fig. 1(A)]. Replication of the double-stranded DNA helix introduces one catenation link per helical turn (3). In E. coli, catenanes are largely unlinked by the action of a Type II topoisomerase, Topo IV, which transports one double-stranded DNA segment through another after having introduced a double strand break in one of the two segments and then seals back the cleaved segment (Fig. 1(A), [4]). Sister chromosome concatenation, which is more classically referred to as chromosome dimer formation, results from odd numbers of crossovers due to homologous recombination between sister chromatids (Fig. 1(A), [5]). In E. coli, chromosome dimers are resolved by the addition of a crossover at a specific locus, dif, by a pair of chromosomally encoded tyrosine recombinases, XerC and XerD, which re-establishes the parity of strand exchanges.
XerCD-dif are also able to unlink catenanes in a stepwise manner, which can compensate for a partial loss of the activity of Topo IV (Fig. 1A, [10, 11, 12]). In addition to E. coli, chromosome dimer resolution has been studied in several Bacteria, including Bacillus subtilis, Campylobacter jejuni, Caulobacter crescentus, Haemophilus influenzae, Helicobacter pylori, Lactococcus lactis, Staphylococcus aureus, Vibrio cholerae, and Xanthomonas campestris, and in two Archaea, Pyrococcus abyssi and Sulfolobus solfataricus [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. Evidence of its importance for the fitness of cells has been directly obtained in B. subtilis, C. crescentus, E. coli, H. pylori, S. solfataricus, and V. cholerae [13, 14, 15, 16, 17, 27, 28, 29].

Other than its importance for the evolutionary success of organisms harboring circular chromosomes, research on Xer recombination has been motivated by three characteristics. First, two recombinases are most often used, XerC and XerD. Each of them catalyzes the exchange of a specific pair of strands between recombining sites, with a Holliday Junction (HJ) as an essential reaction intermediate. As a result, two pathways of reaction can be defined depending on whether XerC- or XerD-catalysis initiates the reaction [Fig. 1(B)]. This feature makes it an excellent model for studying the molecular interplay among the proteins during site-specific recombination. Second, XerC and XerD are exploited by numerous mobile genetic elements: some plasmids use it to maintain the monomeric state of their genome and some lysogenic phages use it to integrate their genome into the genome of their host. Indeed, the Xer recombinases were initially discovered because of their implication in the stability of the Escherichia coli ColE1 plasmid [7, 30]. Exploitation of Xer recombination by plasmids and integrative mobile

N-terminal domains are omitted for clarity. Tails have been added to indicate the C-terminal interactions of the recombinases. Red and black lines indicate the two strands of the recombining sites. Full and empty circles represent the XerC and XerD cleavage points, respectively. Horizontal and vertical substrates are proficient for XerC and XerD-strand exchanges, respectively. (C) Consensus sequence obtained from the alignment of the dif sites of 715 bacterial chromosomes. The XerC and XerD recognition sites are underlined. Double-stranded DNA sequence of V. cholerae dif1, dif2 and difG, of the core cer and psi plasmid sites and of the three types of attachment sites observed in the genome of integrative mobile elements exploiting Xer are shown below (ET: El Tor CTX; VGJ: VGJ phage; TLC: TLC satellite phage). XerC and XerD process the top and bottom strands, respectively. Bases differing from the consensus are shown in red. Lower case letters indicate the absence of conventional Watson–Crick pairing interactions.

doi:10.1128/microbiolspec.MDNA3-0056-2014.f1

FIGURE 1. Chromosome dimer resolution and sister chromosome decatenation. (A) Topological maintenance of bacterial chromosomes. Violet and pink circles represent sister circular chromosomes. Topo IV resolves catenanes. The Xer machinery resolves both catenanes and dimers. (B) Xer recombination. Light and dark grey ovoid shapes depict the C-terminal domains of the XerC and XerD tyrosine recombinases, respectively. The
elements relies on the extreme versatility of the Xer recombination machinery, which can recombine sites with very different structures and/or sequences (Fig. 1(C), [31]). The third interesting feature of Xer recombination is that despite the diversity of sites that XerC and XerD can recombine, their actions remain under very strict cellular and/or topological control.

**THE XER SITE-SPECIFIC RECOMBINATION MACHINERY**

**The Xer recombinases**

Although they share only 38% of their identity, *E. coli* XerC and XerD are the closest relatives to each other in the tyrosine recombinase family. Orthologs of *E. coli* XerC and XerD are readily identified in the genome of most bacteria that carry circular chromosomes (13, 32). When this is the case, a *dif* sequence can almost always be predicted on each of the circular chromosomes of the bacterium based on its homology to *E. coli* *dif* (Fig. 1(C), [13, 32]). The XerC gene, the XerD gene, and *dif* are invariably encoded in different regions of the genome, which suggests that the Xer machinery appeared early in the evolution of bacteria (13, 32, 33, 34). However, the XerCD-*dif* dimer resolution machinery is not universal: in the streptococci and the lactococci, chromosome dimer resolution depends on a single Xer-like tyrosine recombinase, XerS (25); a subgroup of the *ε*-proteobacteria, including *Helicobacter pylori*, uses a single Xer-like tyrosine recombinase, XerH (33); a single tyrosine recombinase is also used in Archaea, XerA in *P. abyssi* (22) and SSQ0395 in *S. solfataricus* (15). These recombinases do not derive from each other, but can all be grouped with XerC and XerD in the phylogenetic tree of tyrosine recombinases. It is probable that the XerS and XerH machineries recently replaced the conventional XerCD-*dif* machinery in the bacteria that harbor them. Indeed, several *ε*-proteobacteria still possess a conventional XerCD-*dif* machinery (33). In addition, both XerC, XerD, and XerS have been found encoded in the genome of *Lactococcus helveticus* (32). Finally, the genes encoding for XerH, XerS, and XerA are immediately adjacent to their target site on the chromosome, which raises the possibility that they could have been acquired by horizontal gene transfer, via the integration of a mobile element (22, 25, 33).

The use of two recombinases is unusual but not unique among tyrosine recombinases. For instance, two recombinases mediate an inversion gene switch that regulates the expression of Type I fimbriae (35). The Tn554 conjugative transposon also uses two related tyrosine recombinases (36, 37).

Despite high sequence divergence, the crystal structure of XerD is strikingly similar to that of Cre (38). The predicted secondary structure of XerC fits well with the known structure of XerD. It is therefore very likely that the structure of XerC is similar to that of XerD. The XerC and XerD proteins can be divided into a small N-terminal and a large C-terminal domain that make a C-shaped clamp into which the DNA is bound. Like Cre, and in contrast to Flp, a functional catalytic unit is contained within a single recombinase molecule. Indeed, each recombinase can by itself act as a site-specific type I topoisomerase by relaxing DNA containing its binding site (39). However, this activity is prevented when the two recombinases are bound together (39). This is likely due to the cyclic interactions that regulate and coordinate the two pairs of strand exchanges during recombination (Fig. 1(B), [40, 41, 42]).

**The recombination mechanism**

Recombination occurs within a nucleoprotein complex consisting of the two recombining sites and a pair of each of XerC and XerD. By analogy to Cre/loxP recombination, cyclic interactions made via the extension of the extreme C-terminus of the recombinases [the donor region, depicted as little tails in Fig. 1(B)] are thought to hold together the nucleoprotein complex. In these interactions, the donor region of each recombinase contacts the neighboring partner recombinase close to its active site (the acceptor region). The synaptic complex is asymmetric, the donor region of two of the recombinases contacting the acceptor region of the partner recombinases in *cis*. By analogy to Cre/loxP, this is thought to involve a more extended conformation of the donor recombinases, which activates their catalytic tyrosine (depicted in Fig. 1(B) by the addition of an open and filled circle at the position where the XerD and XerC recombinases will cleave their target strands, respectively). Each pair of activated recombinases catalyzes the cleavage of a specific pair of strands immediately 3′ to their binding sites. It generates recombinase/DNA covalent phosphorothysterolyl linkages on one side of the cleaved strands and free 5′-hydroxyl extremities on the other side. Following cleavage, a few bases from each of the two liberated 5′-hydroxyl extremities melt from their complementary strand and attack the recombinase/DNA covalent phosphorothysterol linkage of their recombining partner. Subsequent ligation requires the stabilization of strand invasion by Watson–Crick or Wobble base pairing interactions (43). Small movements of DNA and proteins in the resulting HJ allow a change in conformation that reciprocally activates the second pair of strand exchanges (depicted by a 90° rotation of the XerC
and XerD binding sites in Fig. 1(B) and in the rest of the figures of this chapter. This model is supported by both genetic and biochemical ensemble experiments on various chromosome, plasmid, and synthetic recombination sites, as well as synthetic HJ intermediates (44, 45, 46, 47, 48, 49, 50). It is also supported by the direct observation of the predicted conformational changes during the recombination of two dif sites in single molecule experiments (51).

**CHROMOSOME DIMER RESOLUTION**

**Cell cycle regulation**

Decatenation and chromosome dimer resolution are highly regulated processes. They depend on two main factors: cell division and the organization of the chromosome.

During cell proliferation, the timing of DNA synthesis, chromosome segregation, and cell division must be coordinated to ensure the stable inheritance of the genetic material. In eukaryotes, this is achieved by checkpoint mechanisms that separate these processes in time. No such strict temporal separation exists in bacteria, in which these processes are concomitant. Replication can even be initiated more than once in a single cell, giving rise to siblings that inherit already partially duplicated chromosomes (52). This notably allows for shorter doubling times than the time required for the replication of the entire genome (53). The length of time required between the end of a new round of replication and the next cell scission event, D, is one of the key factors limiting the doubling time of the cell population [Fig. 2(A)]. Cell scission takes place at the end of the D period, however the assembly of the cell division apparatus can start much earlier [Fig. 2(A)]. Two steps can be defined in the process: (i) mid-cell assembly of a ring of early cell division proteins, including the tubulin-like FtsZ protein, and (ii) recruitment of late cell division proteins such as FtsI, which is involved in the remodeling of the rigid cell wall. Using thermosensitive mutants of early and late cell division proteins, it was shown that Xer recombination at dif happens only after the assembly of the complete cell division apparatus (54, 55, 56). Using cephalixin, an antibiotic that inhibits cell constriction without disrupting the cell division machinery, it was further shown that dif recombination occurs shortly before, or commensurate with, septum closure (Fig. 2(A), [54, 55, 56]). Likewise, even though Topo IV removes precatenanes behind the replication forks during the course of replication, the majority of topoisoasmerase IV activity is restricted to the D period of the cell cycle (57, 58).

**FIGURE 2** Spatial and temporal control of chromosome dimer resolution. (A) Temporal restriction of Xer recombination during the bacterial cell cycle. White disk: origin of replication region; Converging arrows: terminus of replication region. The two sister chromatids are depicted as pink and purple tubes. (B) Spatial restriction of Xer recombination along bacterial chromosomes. The dif activity zone corresponds to the region in which dif can still resolve dimers if displaced. (C) FtsK controls Xer recombination. Violet and pink circles represent bacterial sister circular chromosomes. White arrows indicate the KOPS motifs and their orientation. dif sites are shown as red and black lines. The FtsK protein is drawn in blue. doi:10.1128/microbiolspec.MDNA3-0056-2014.f2

Bacterial chromosomes harbor a single origin of bidirectional replication. When they are circular, replication ends in a region opposite the origin of replication, the terminus region, which defines two replichores of approximately equivalent lengths [depicted as arrows in Fig. 2(B)]. There is an asymmetry between the nucleotide compositions of the leading strand and the lagging strand, which results in an inversion in GC-skew, i.e., the richness of G over C and T over A, between the two replichores of bacterial chromosomes. The position at which the two replication forks most frequently
collide on the circular chromosome can be determined by computing the GC-skew (59). Analysis of hundreds of bacterial genomes demonstrated that dif sites were invariably located at a short distance from the termination point determined by GC-skew profiling (32, 60). The dif site was found to be a hot spot of TopoIV activity on the E. coli chromosome, linking decatenation to the replisome organization of bacterial chromosomes (61). In E. coli, chromosomal inversion and dif-displacement studies further showed that dif needs to be located in the terminus region of the chromosome to be effective for chromosome dimer resolution (62, 63, 64). In addition, excision of small DNA cassettes inserted in the chromosome between two directly-repeated dif sites was found to be limited to the terminus region of chromosomes in E. coli and V. cholerae (13, 27, 56, 65). However, there is no direct coupling between the recombination activity of XerCD at dif and replication: cells in which termination is displaced away from dif remain proficient in chromosome dimer resolution as long as dif is kept at the inversion of polarity of the two replisomes (66, 67, 68).

It was the discovery of the function of a cell division protein, FtsK, in chromosome segregation that led to our understanding of the temporal restriction of dif recombination to the time of cell division and of its spatial restriction to the terminus region of the chromosomes. Most of our knowledge on FtsK came from the study of its role in E. coli and from the study of B. subtilis SpoIIIE, a homolog implicated in sporulation. B. subtilis and E. coli each possess nucleoid occlusion machineries (69, 70). Nevertheless, septum formation can initiate over partially segregated chromosomes during vegetative cell division in both species, at least when a chromosome dimer has been created by homologous recombination (65, 71, 72). In addition, it invariably occurs during sporulation in B. subtilis (73). The correct distribution of the genetic material then requires the respective activities of SpoIIIE and FtsK (Fig. 2C). The two proteins share a common structural organization (Fig. 2C, [74]). This includes an integral domain at the amino-terminus (FtsK\textsubscript{N}), which is anchored to the inner membrane by 4 transmembrane segments (75), a linker domain (FtsK\textsubscript{L}), which lacks any evolutionarily conserved features, and a conserved RecA-type ATPase fold at the C-terminus (FtsK\textsubscript{C}, [76]). In E. coli, FtsK\textsubscript{N} and FtsK\textsubscript{L} are implicated in the stabilization of the cell division apparatus, and the presence of at least one of these two domains is essential for the assembly of the cell division apparatus (77). FtsK\textsubscript{C} assembles into hexameric motors that use the energy of ATP to translocate on DNA (78, 79). As FtsK motors are anchored at the site of division by FtsK\textsubscript{N} and FtsK\textsubscript{L}, it results in the mobilization of chromosomal DNA (28, 80).

FtsK translocation on chromosomal DNA is not random. Using a combination of genetics, bioinformatics, and biochemical experiments at the single molecule level, specific DNA motifs that dictate the orientation of translocation of FtsK were identified on the E. coli chromosome, the KOPS (FtsK oriented polar sequence) (81, 82, 83, 84). Identical motifs orient the activity of V. cholerae FtsK (13). KOPS are small, highly repeated, polar motifs that point from the origin of replication of bacterial chromosomes towards their dif site in their terminus region (13, 82, 85). As a result, FtsK motors pump the origin-proximal regions of a chromosome dimer away from the division site and the zone of convergence of the KOPS to mid-cell [Fig. 2(C)]. The sequence of KOPS is not conserved in all bacteria. The motifs that orient the activities of B. subtilis SpoIIIE and L. lactis FtsK differ from the E. coli and V. cholerae KOPS and between each other, which suggests that different polar DNA motifs have been co-opted during the course of evolution to serve as FtsK oriented polar sequences (85, 86). Many studies have focused on the mechanism of translocation of FtsK and on the mode of action of the KOPS (see (87, 88) for a review). The end result of the action of FtsK is that the two dif sites of a dimer are brought together if they are correctly located in the zone of convergence of the KOPS but separated if they are located elsewhere on the chromosome, which explains the dif activity zone [Fig. 2(B) and 2(C)]. FtsK was also shown to stimulate the activity of Topo IV, which might explain in part the increase in Topo IV activity in the D period of the cell cycle (89, 90). Impeding FtsK translocation has varying effects on fitness depending on the bacterial species. In E. coli, FtsK\textsubscript{C} participates in the orderly segregation of the terminus region of sister chromatids, whether monomeric or dimeric (91). However, it is not essential. It is only required when a chromosome dimer has been formed, which under laboratory growth conditions has been estimated to occur in ~15% and 10% of cell divisions in E. coli and V. cholerae, respectively (65, 92). In contrast, FtsK\textsubscript{C} is essential in Caulobacter crescentus, which is attributed to its role in chromosome decatenation (93).

A direct interaction between FtsK and XerD is necessary for the addition of a crossover between dif sites (78, 94). This is the limiting factor in Xer recombination at dif (65, 78, 95, 96). FtsK does not take charge of the segregation of the bulk of the DNA, but only of the terminus regions of chromosomes (56, 91). As a result,
Xer recombination is normally limited to dif sites located in this region, which explains the absence of excision of dif cassettes inserted elsewhere on chromosomes (56, 65, 66, 92) and the low efficiency of excision of dif cassettes on plasmids (65, 97). It also explains why recombination at chromosomal dif sites is limited to the time of division since this is the only period when FtsK needs to take charge of chromosomal DNA. However, hexamers of FtsK and SpoIIIE form before any visible sign of cell constriction (98, 99). Therefore, it remains to be understood why in E. coli and V. cholerae most recombination events only occur after the initiation of constriction and not directly after the recruitment of FtsK to mid-cell (54, 65). In addition, it remains unclear why there is over a 10-fold drop in the frequency of chromosomal tandem dif cassette excision when homologous recombination is abolished in E. coli (54, 65, 92). It was long postulated that monomeric chromosomes were segregated away from mid-cell before the activation of the FtsK translocation, thereby limiting Xer recombination to chromosome dimers. However, microscopic observations suggest that, even though FtsK-dependent DNA translocation is not essential in E. coli, FtsK normally takes charge of the orderly segregation of the terminus region of all the chromosomes, whether concatenated or not (91). In addition, dif cassette excision was found to be independent from chromosome dimer formation in V. cholerae, further suggesting that the influence of RecA might be linked to a specificity of the regulation of the activity and/or production of the E. coli recombinases (56). In favor of this hypothesis, Xer recombination between plasmid-borne dif sites is also slightly stimulated by chromosome dimer formation in E. coli (97, 100).

Control of catalysis
Most of what we know about the control of Xer recombination at dif by FtsK has been gained from the study of the E. coli Xer recombination machinery. By default, the XerD recombinases are inactive and the XerC recombinases promote a low level of HJ formation in vitro and in vivo (Fig. 3(A), [65]). However, these HJs could only be observed in vivo in the presence of a DNA intercalating and crosslinking agent, psoralen (65). Their formation is favored in vitro by high glycerol concentrations and the presence of a DNA intercalating agent, ethidium bromide (65). This recombination pathway is a dead-end. HJs resulting from XerC-catalysis cannot be converted into a crossover by XerD-catalysis (44). In addition, no product can be created by replication across them in vivo, suggesting that they are unstable and very transient (65, 78). Indeed, HJs resulting from XerC-catalysis are rapidly eliminated by reverse reactions (Fig. 3(A), [44, 45, 49, 50, 65, 78]). During chromosome dimer resolution, the XerC-first futile recombination cycle is broken by a direct interaction with FtsK, which triggers the exchange of a first pair of strands by XerD-catalysis (Fig. 3(B), [13, 51, 78, 96, 101]). Structural and functional studies indicate that the 62-amino acid C-terminal fragment of FtsK, FtsKγ, interacts directly with the XerD C-terminus (94, 95). Indeed, FtsKγ can activate XerCD-dif recombination in the absence of the translocase domain, when it is fused to XerCD or added in isolation (102). Based on the structure of the Cre/loxP synaptic complex, it is not possible to switch a synapse in which the XerC recombinases are in a suitable conformation to be active into a configuration suitable for XerD catalysis without breaking the cyclic interactions between the recombinases [Fig. 1(B)]. Single molecule studies demonstrated that XerCD/dif synaptic complexes readily form in vitro, ruling out models in which FtsK activated XerD prior to synapsis and in which FtsK actively remodeled complexes initially poised to undergo XerC catalysis (51, 103). In vitro dissection of the reaction further indicated that the HJs created by XerD catalysis rapidly isomerized into a configuration suitable for XerC-strand exchange (51). Correspondingly, HJs created by XerD catalysis are rapidly converted into product by XerC-catalysis in vitro and in vivo (Fig. 3(B), [13, 51, 78, 96, 101]). As XerD catalysis is impeded by coordinated reciprocal switches in recombinase activity, HJs created by XerD catalysis are readily detectable without chemical treatment when XerC catalysis is inhibited and are stable enough to be converted into product by replication (13, 45, 78).

Work in H. influenzae, V. cholerae and B. subtilis suggest that the mechanism by which FtsK controls Xer recombination at dif is conserved in bacteria harboring two Xer recombinases (13, 72, 95, 104). Even more interestingly, the FtsK control seems to have been kept in L. lactis even though its XerCD machinery was replaced by a single tyrosine recombinase, XerS (105).

Topological control
FtsK translocation is not absolutely required to activate Xer recombination at dif (102, 105). However, the ATPase activity of FtsK considerably enhances the efficiency of the reaction in vitro (51, 78, 81, 96, 101). In addition, products of FtsK-dependent Xer recombination on supercoiled circular substrates containing directly repeated dif sites have a simple fixed topology.
The only products that could be detected were free circles, as if the two dif sites are always brought together by a slithering mechanism (Fig. 3(C), [78]). In contrast, in the absence of translocation, FtsK-dependent Xer recombination of supercoiled circular substrates harboring tandem dif sites produces catenanes, as expected from random collision events (24, 102). FtsK can also simplify the products of Cre/loxP recombination (10). The capacity of chromosome dimer resolution machinery to remove catenation links in a stepwise manner is probably related to this property of FtsK (11, 12). The mechanism by which FtsK might help bring together...
dif sites in a specific topological configuration is not understood yet. Some clues to the answer might reside in the observation that although FtsK and SpoIIIE can strip proteins off the DNA (101, 106, 107), FtsK translocation is stopped by XerCD-dif (107, 108).

**PLASMID DIMER RESOLUTION**

Multimerization of circular plasmids reduces the number of independently segregating units, which leads to failures in their vertical transmission from mother cell to daughter cells (109, 110, 111, 112, 113). Hence, many plasmids and phages encode a tyrosine recombinase that serves to maintain the monomeric state of their genome. This is notably the case for phage P1, which encodes the Cre recombinase. However, several plasmids do not possess their own dedicated dimer resolution machinery and exploit the Xer recombinases of their host (30, 114, 115, 116, 117, 118). These plasmids do not carry a dif site and resolution of plasmid multimers is independent of FtsK. Indeed, addition of a dif site to a plasmid rather leads to its concatenation inside the cell (97).

**Topological regulation**

The ColE1 and pSC101 dimer resolution systems have been characterized in great detail (112, 116). ColE1 and pSC101 carry a complex recombination site composed of a core dif-like sequence and ~150 bp of accessory sequences, cer and psi, respectively (Fig. 1(C) and Fig. 4(A), [119, 120, 121]). In contrast to Xer recombination at dif, Xer recombination at cer and psi senses the topological connectivity between the sites such that they recombine only if they are directly repeated on the same DNA molecule (122). This is due to a “topological filter,” which ensures recombination selectivity, i.e., that multimers are resolved rather than formed by Xer recombination. A direct consequence of the topological filter is that Xer recombination yields products with a specific topology (122): recombination between two psi sites produces four-noded catenanes, in which the two component circles are wrapped around each other exactly four times (Fig. 4(B), four-noded catenane); recombination between two cer sites produces an equivalent structure with an HJ.

In the cer and psi recombination reactions, the two product circles are wrapped around each other in a right-handed fashion and the sites in these circles are in antiparallel orientations (122). It was demonstrated using the mathematical tangle theory that this specific topology could only be produced by a single 3-dimensional model of the recombination reaction (Fig. 4(A), [123, 124]).

**FIGURE 4** Plasmid dimer resolution. (A) Schematic representation of the topological filter. Yellow circles represent accessory proteins. P: PepA; A: ArgR or phosphorylated ArcA; Green tubes: accessory sequences. (B) Topology of the products of Xer recombination at cer and psi multimer four-node catenanes. (C) The topological filter controls Xer catalysis for plasmid dimer resolution. doi:10.1128/microbiolspec.MDNA3-0056-2014.14

The cer and psi accessory sequences are binding sites for two host proteins, PepA and ArgR (125, 126, 127) and PepA and phosphorylated ArcA (128), respectively. All three proteins are known transcription factors with specific DNA binding properties. ArgR binds as a hexamer to a single ArgR box within cer; PepA binds as a
hexamer to two synapsed cer accessory sequences on both sides of the ArgR binding site (Fig. 4(A), [127]). PepA and phosphorylated ArcA bind to the psi accessory sequences at places similar to those bound by PepA and ArgR in cer (Fig. 4(A), [128]). Based on the X-ray structure of PepA and on the analysis of mutants of PepA, a molecular model could be built showing how two hexamers of PepA and one hexamer of ArgR could bind to the accessory sequences of two cer sites (129, 130). Atomic force microscopy now supports this model (131).

In contrast to recombination at cer and psi, recombination between loxP sites requires only the Cre protein and gives products of different topologies depending on the reaction conditions and the degree of supercoiling of the substrate (132, 133). Complex product topologies are created after synapsis by random collision (Fig. 3(C), [134, 135]). However, the addition of PepA was sufficient to alter the topology of Cre recombination at hybrid sites consisting of accessory sequences from cer or psi adjacent to a loxP core (136). These results show that the accessory sequences and the accessory proteins are sufficient to create a topological filter and thus ensure topological selectivity.

**Control of catalysis**

The pathway of recombination followed by the Xer machinery at cer and psi is reversed compared to the pathway of recombination at dif; the two XerC recombinases of the nucleoprotein synaptic complex catalyze the first pair of strand exchanges in the reaction (Fig. 4(A), [7, 137]). It was partly because of these observations that the XerC-first pathway was proposed to be the default pathway of recombination in the early models on the control of Xer recombination at dif (Fig. 3(A) and 3(B), [45, 63]). However, in the same year when the role of FtsK in promoting a first pair of strand exchanges by XerD catalysis at dif was demonstrated, it was shown that the accessory sequences and the accessory proteins determined the order of strand exchanges at psi: a plasmid containing two inverted-core psi sites recombined with a reversed order of strand exchange, but with unchanged product topology (121). This result suggests that the topological constraints imposed by the accessory sequences and the accessory proteins switch the synaptic complex into a configuration suitable for XerC-strand exchanges and that they switch the resulting HJ intermediate into a configuration suitable for XerD-strand exchanges. This result also has structural implications: the recombinase–core nucleoprotein complex must form with the recombinase C-terminal domains facing the accessory proteins and sequences to explain both the order of strand exchanges and the topology of the products observed with normal and inverted psi sites (121).

The recombination reaction doesn’t proceed further than the formation of an HJ at cer (138). This is probably linked to the structure of the cer core site, in which the central region contains 8 bp instead of the canonical 6 bp overlap region of most tyrosine recombinase sites [Fig. 1(C)]. The HJ intermediate is not processed by the host RuvC HJ resolvase (138). Instead, it is proposed that initiation of replication on the HJ-containing circular molecule leads to the production of two plasmid monomers and one plasmid dimer [Fig. 4(C)]. This also seems to occur at psi when the catalytic activity of XerD is inactivated (45, 121). In agreement with this model, the HJ intermediates resulting from XerC-catalysis at cer and psi are stable enough to be detected without any chemical trap, in contrast to the HJs resulting from XerC-catalysis at dif. It is probable that the topological constraints imposed by the accessory proteins and sequences permit the stabilization of the HJs by switching them into a conformation unsuitable for XerC-catalysis (and reciprocally suitable for XerD-catalysis in the case of psi).

**Cell cycle regulation**

An advantage of the mechanism by which plasmids exploit the Xer machinery of their host to maintain their monomeric state is to escape the spatial and temporal control normally exerted on Xer recombination. Moreover, plasmids are able to control the cell cycle of their host to allow more time for their replication and/or for multimer resolution before the cell divides. The RepA protein of plasmid pSC101 controls the cell division of *E. coli* to permit the repopulation of cells with plasmids when copy number falls and thus limit variation in plasmid copy number (139). ColE1 encodes a short, untranslated RNA called Rcd (regulator of cell division) that delays cell division to allow more time for multimer resolution (140, 141, 142). Rcd and its promoter are found within the accessory sequences of cer and plasmid concatenation triggers its expression (143). The exact mechanism by which Rcd inhibits cell division might be linked to the production of indole, a small molecule that affects the transmembrane potential of bacteria (144, 145, 146, 147).

**INTEGRATIVE MOBILE ELEMENTS EXPLOITING XER**

Many mobile genetic elements encode a tyrosine recombinase for the integration/excision of their genome into the genome of their host [Fig. 5(A)]. This is notably the
case for lysogenic phages, such as phage λ. However, some integrative mobile elements do not encode their own recombinase but use the Xer machinery of their host to integrate at dif. These elements are referred to as integrative mobile elements exploiting Xer (IMEXs). IMEXs are found integrated in the genome of many bacteria, including Enterobacteriaceae like E. coli and Yersinia pestis, Xanthomonadaceae like X. campestris and Xylella fastidiosa, Neisseriaceae like Neisseria gonorrhoeae and Neisseria meningitidis, and Vibrionaceae like V. cholerae (31, 148, 149, 150, 151, 152, 153).

Initial work on IMEXs was motivated by their implication in the evolution of human pathogens. In particular, the genes encoding cholera toxin, the principal virulence factor of the diarrhea-causing Gram-negative bacterium Vibrio cholerae, are encoded by one of the first phages to be identified as an IMEX, CTXϕ (152, 153). Several other IMEXs use the Xer machinery of V. cholerae. In each case, a functional dif site is invariably recreated after their integration, at the attR junction [Fig. 5(B)]. The new dif site is composed of the XerD-side of the previous dif site and of the XerC-side of the attP site of the element (Fig. 5(B), [156, 43, 157, 158]). This implies that the IMEX attP sites must carry a conserved XerC-binding site. dif restoration is important for the fitness of the host, and thus for the success of the IMEX. In addition, it permits multiple successive integration events [Fig. 5(B)]. Correspondingly, analysis of more than 150 complete genomes of clinical and environmental V. cholerae isolates revealed that most of them harbored large IMEX arrays (159, 160). Molecular interactions between the different IMEXs that compose these arrays

![FIGURE 5 IMEX integration depends on little homology with dif. (A) Schematic of the integration/excision of mobile elements into the genome of their host. (B) IMEX integration generates a new dif site, which allows for multiple successive integration events. (C) IMEX integration depends on limited homology. The distance separating the two bases of a base pair indicates the quality of the base pair interactions that are formed. N.D.: not determined doi:10.1128/microbiolspec.MDNA3-0056-2014.f5](https://www.asmscience.org/MicrobiolSpectrum)
are implicated in the ecological cycle that permits the constant and rapid acquisition of new cholera toxin gene variants, as observed in the ongoing 7th cholera pandemic (157, 158, 159, 160, 161, 162, 163, 164).

**Xer recombination between nonhomologous sites**

The bacterial attachment site of a given integrative mobile element, attB, must match the attachment site of the element, attP, and should be in a highly conserved genomic region of the host genome (163, 166). In this regard, exploitation of the Xer machinery is advantageous since chromosomal dif sites are highly conserved (13, 32, 33). The simplest solution for an element to become an IMEX would be to harbor a dif site. Indeed, *E. coli* and *V. cholerae* XerC and XerD can promote the efficient integration of plasmids harboring a dif site into their chromosomal counterpart in a FtsK-dependent manner (65, 158). However, such integration events are highly unstable (56, 65). All the IMEXs described to date possess an attachment site with limited homology to dif, as illustrated with the different *V. cholerae* IMEXs [Fig. 1(C) and Fig. 5(C)]. The genome of *V. cholerae* is divided on two circular chromosomes, chromosome I and II. Both chromosomes harbor a unique dif site. *V. cholerae* dif sites are polymorphous. They can be grouped into three classes based on the sequence of their central region, dif1, dif2, and difG (Fig. 1(C) and 5(C), [13, 43]). Chromosome II always carries a dif2 site whereas chromosome I generally harbors a dif1 site. Each of the different *V. cholerae* IMEXs specifically targets a subset of these sites (43, 157, 158). This implies some homology between the central regions of the attP site and the targeted dif sites. However, the amount of homology that is required is limited (Fig. 5(C), [43, 157, 158]): a single Watson–Crick strand on one side of the exchanged overlap regions and a G–T Wobble base pair on the other side are sufficient to stabilize the exchange of a pair of strands [Fig. 5(C)]. From this point of view, the integration/excision reactions of IMEXs are markedly different from plasmid and chromosome dimer resolution reactions. The ability of XerCD to recombine sites with very little homology is further highlighted by the observation of rare dif1–dif2 chromosome fusion events (167).

**Strategies for the control of integration and of excision**

It is critically important for temperate phages to control their timely excision from the host genome, such as when the survival of the host is compromised. In most cases, this is ensured via a virally encoded accessory protein, the excisionase. In the absence of the excisionase, the phage recombinase recombines attP and attB to produce attL and attR. In the presence of the excisionase, it recombines attL and attR to produce attB and attP again. To date no excisionases have been characterized in the genome of IMEXs. This is apparently compensated by smart life cycle strategies.

Three very different types of IMEX attP sites have been identified so far, which correspond to three categories of IMEXs with distinct life cycle strategies [Fig. 1(C) and Fig. 6, [31, 148]]. These three IMEX categories are referred to as the CTX-, VGJ- and TLC-families, according to their most well-studied representative in *V. cholerae*, CTXϕ (152, 153), VGJϕ (168) and TLCϕ (162, 163).

**The CTXϕ life cycle**

CTXϕ is a filamentous phage. Its double-stranded DNA (dsDNA) replicative form carries two dif-like sites in inverted orientation (169). None of them can be recombined with any of the *V. cholerae* dif sites because the central region of the first one has no homology with the central regions of the *V. cholerae* dif sites and because the central region of the other one is too long to permit XerC and XerD cyclic interactions (156). Instead the attachment site of CTXϕ, attPCTX, consists of the stem of a hairpin that is generated by the folding of the region encompassing these two sites in the single-stranded DNA (ssDNA) phage genome: the one which is packaged in the phage particles and which is generated by rolling circle replication [Fig. 1(C)]. Integration is catalyzed by the XerC recombinases, which promote the formation of an HJ intermediate [Fig. 5(C) and 6(A), [43, 156]]. The lack of homology on the XerD-side of the overlap regions of attPCTX and dif prevents any potential XerD-mediated strand exchange. Correspondingly, CTXϕ-integration does not depend on the catalytic activity of XerD, which only plays a structural role in the formation of the synaptic complex (43). The HJ intermediate is probably converted into product by replication [Fig. 6(A)]. The active form of attPCTX is masked in the double-stranded genome of prophage. As a result, the integration of CTXϕ is intrinsically irreversible (156). However, free CTXϕ genome copies can be produced by a process analogous to rolling circle replication after the integration of a second IMEX harboring the same integration/replication machinery, such as the RS1 satellite phage (164). Production of these new extrachromosomal copies of the genome of CTXϕ is under the control of SOS, at least in part because the SOS repressor, LexA, regulates the expression of the phage nickase, RstA (170, 171).
The VGJφ life cycle

Like CTXφ, VGJφ is a filamentous phage. However, its attachment site, attPVGJ, is found in its dsDNA replicative form. As in the case of CTXφ, VGJφ-integration results from a single pair of strand exchanges between this site and its chromosomal target, which is catalyzed by XerC ([Fig. 5(C), [157]]. The resulting HJ is probably converted into product by replication ([Fig. 6(B), [157]].

**FIGURE 6** Diversity of the IMEX integration pathways. (A) Irreversible integration of CTX-type elements. The single-stranded DNA and double-stranded DNA forms of the element are represented in pink. The host genome is shown in purple. The incapacity of XerD to perform strand exchanges is indicated in yellow. Orange hexagons labeled with the letter E are Endoll. (B) Integration/excision of VGJ-type elements. The double-stranded DNA replicative form of the element is shown in pink. The host genome is shown in purple. A yellow explosion indicates the impossibility for XerD to perform strand exchanges. The orange circle labeled with a question mark indicates a putative unknown integration factor. (C) Integration/excision pathway of TLC-type elements. The double-stranded DNA form of the element is shown in pink. The host genome is shown in purple. The Blue circle labeled with a question mark represents an unknown factor that could permit the binding of XerD and its catalytic activation. doi:10.1128/microbiolspec.MDNA3-0056-2014.f6
However, a major difference with the integration of CTXϕ is that a functional phage attachment site is recreated at the attL junction. As a result, a new round of Xer recombination can excise integrated copies of VGJϕ: XerC can promote the exchange of a pair of strands between the phage attachment site that is recreated at attL and the dif site that is recreated at attR; replication of the HJ intermediate leads to a free copy of the dsDNA genome of VGJϕ, one sister chromosome free of the prophage and one sister chromosome in which the prophage is still present (157). Correspondingly, IMEXs of the VGJ-type are only very rarely found integrated in the genome of clinical and environmental strains (159, 160).

How do these cycles of integration and excision participate in the life cycle of VGJϕ? The dsDNA replicative form of the phage, which is normally extremely stable, is easily lost in cells in which Xer recombination is abolished, suggesting that the integration/excision cycles participate in the vertical inheritance of the phage genome and in the steady state production of new viral particles (157). As a consequence, VGJϕ production escapes the SOS control, which is markedly different from most other lysogenic phages (157).

The TLCϕ life cycle

TLCϕ is a satellite phage that exploits phages of the VGJϕ IMEX family for dissemination (162). As in the case of VGJϕ, its attachment site, attP_{TLC}, is found in its dsDNA replicative form. However, the overlap region of attP_{TLC} is fully homologous with the overlaps of its target dif site in contrast to attP_{CTX} and attP_{VGJ} [Fig. 5 (B)]. In addition, it lacks a bona fide XerD-binding DNA arm [Fig. 1(B)]. Correspondingly, attP_{TLC} is a poor XerD binding substrate (158). Nevertheless, a pair of strand exchanges catalyzed by XerD initiates TLCϕ-integration, the resulting HJ intermediate being converted into product by XerC-catalysis (Fig. 6(C), [158]). A functional phage attachment site is recreated at the attL junction so that the reaction is fully reversible: XerD can promote the exchange of a pair of strands between the phage attachment site that is recreated at attL and the dif site that is recreated at attR, the resulting HJ intermediate being converted into product by XerC-catalysis (Fig. 6(C), [158]).

Control of catalysis

The propensity of HJs resulting from XerC-catalysis to be eliminated by reverse reactions should severely compromise the chances for replication to finalize the integration process of CTXϕ and VGJϕ (44, 45, 65, 172). In this case of CTXϕ, Endo III, a ubiquitous base excision repair enzyme, seems to dislodge the Xer recombinases from the HJs once they are formed, which prevents XerC from reversing the reaction (172). This action is crucial to the life cycle of CTXϕ because it increases the chances for success of independent integration events, which facilitates the formation of tandem CTXϕ arrays (172). However, it is not an absolute requirement. attP_{CTX}/dif HJs are stable enough to be processed by replication in the absence of Endo III (172). Likewise, no other host or phage factor seems to be involved in the integration of VGJϕ, suggesting that attP_{VGJ}/dif HJs are stable enough to be processed by replication (172). In contrast, the dif/dif HJs that result from XerC-strand exchanges cannot be converted into product by replication [Fig. 3(A)]. These observations suggest that the unconventional central region of attP_{CTX} and attP_{VGJ} might help stabilize their HJ integration intermediates.

The integration/excision strategy of TLCϕ came as a surprise because all other mobile elements so far characterized, whether plasmids or IMEXs, exploited the XerC-first pathway of recombination (see above). In this respect, the TLCϕ integration pathway is similar to the chromosome dimer resolution pathway. However, TLCϕ integration and excision were both found to be independent of FtsK, whereas chromosome dimer resolution depends on the activation of XerD-catalysis by FtsK (158). A second puzzling aspect of the exploitation of Xer by TLCϕ is that integration seems to be much more efficient than excision, even though they rely on the same Xer reaction pathway (158). The factor(s) that permit the activation of XerD-catalysis in the TLCϕ reactions and help favor its integration remain to be explored.

CONCLUDING REMARKS

Two related tyrosine recombinases, XerC and XerD, are encoded in the genome of most bacteria where they serve to resolve dimers of circular chromosomes by the addition of a crossover at a specific site, dif. The most striking feature of Xer recombination is possibly the variety of the mechanisms that permit their exploitation as chromosome and plasmid dimer resolvases as well as mobile genetic element integrases.

The versatility of exploitation of Xer recombination seems intimately connected to the many possibilities that exist to coordinate and control the strand exchange reactions catalyzed by XerC and XerD. In particular, four recombination pathways can be considered, depending on whether recombination is initiated by XerC- or XerD-catalysis, and whether it ends with a second
pair of strand exchange or with the conversion of the HJ intermediate into product by replication. Each of these different recombination pathways has been observed. The molecular bases of their control are being deciphered. This includes the reciprocal control of catalysis that the two Xer partner recombinases exert over each other, the topological modifications imposed by accessory sequences and accessory factors, and the sequence of the recombination sites.

A difficulty in our progress towards the molecular characterization of the Xer recombination pathways has long been the lack of structural information on the different intermediates of the reactions. It is safe to assume that Xer recombinases more amenable to structural studies will be found with the characterization of Xer machineries from more and more organisms. In addition, the development of new biophysical techniques to track the recombination reaction at the single molecule level promises future great advances in the field.

ACKNOWLEDGMENTS

We would like to acknowledge support from the Agence Nationale pour la Recherche [ANR-11-BLAN-O2401], the Fondation Bettencourt Schueller [2012 Coup d’Elan award], and the European Research Council under the European Community’s Seventh Framework Programme [FP7/2007-2013 Grant Agreement no. 281590].

REFERENCES


Xer Site-Specific Recombination


Double-stranded DNA translocation: structure and mechanism of hexa-


