An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective

MAKKUNI JAYARAM,1 CHIEN-HUI MA,1 AASHIQ H KACHROO,1 PAUL A ROWLEY,1 PIOTR GUGA,2 HSUI-FANG FAN,3 and YURI VOZIYANOV4

1Department of Molecular Biosciences, UT Austin, Austin, TX 78712; 2Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Department of Bio-organic Chemistry, Lodz, Poland; 3Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 112, Taiwan; 4School of Biosciences, Louisiana Tech University, Ruston, LA 71272

ABSTRACT Tyrosine site-specific recombinases (YRs) are widely distributed among prokaryotes and their viruses, and were thought to be confined to the budding yeast lineage among eukaryotes. However, YR-harboring retrotransposons (the DIRS and PAT families) and DNA transposons (Cryptons) have been identified in a variety of eukaryotes. The YRs utilize a common chemical mechanism, analogous to that of type IB topoisomerases, to bring about a plethora of genetic rearrangements with important physiological consequences in their respective biological contexts. A subset of the tyrosine recombinases has provided model systems for analyzing the chemical mechanisms and conformational features of the recombination reaction using chemical, biochemical, topological, structural, and single molecule-biophysical approaches. YRs with simple reaction requirements have been utilized to bring about programmed DNA rearrangements for addressing fundamental questions in developmental biology. They have also been employed to trace the topological features of DNA within high-order DNA interactions established by protein machines. The directed evolution of altered specificity YRs, combined with their spatially and temporally regulated expression, heralds their emergence as vital tools in genome engineering projects with wide-ranging biotechnological and medical applications.

INTRODUCTION

Tyrosine family site-specific recombinases (YRs), named after the active site tyrosine nucleophile they utilize for DNA strand breakage, are widely distributed among prokaryotes. They were thought to be nearly absent among eukaryotes, the budding yeast lineage (Saccharomyces) being an exception in that a subset of its members houses nuclear plasmids that code for YRs (1, 2). However, YR-harboring DIRS and PAT families of retrotransposons and presumed DNA transposons classified as Cryptons have now been identified in a large number of eukaryotes (3, 4). The presence of functional YRs encoded in Archaeal genomes has been established by a combination of comparative genomics and modeling complemented by biochemical and structural analyses (5, 6). Over 1300 YR sequences mined from bacterial genome databases have been organized into families and subfamilies, providing a better understanding of the evolutionary relationships among them (7). These classifications also encourage investigations into the potential functional significance of YRs whose
genes are present as pairs or trios in bacterial and plasmid genomes.

YRs are remarkable enzymes that utilize a common chemical mechanism to bring about a wide array of biological consequences. They range from the choice of lysogenic or lytic developmental pathways in phage λ and related phage, equal segregation of phage, plasmid and bacterial chromosomes by resolving genome dimers or multimers formed by homologous recombination into monomers, to the resolution of hairpin telomeres that mark the termini of certain bacterial and phage genomes (8, 9, 10, 11, 12). In addition, YRs promote the transposition of conjugative mobile elements, the resolution of cointegrate intermediates formed by the Tn3-related toluene catabolic transposon Tn4651, the unidirectional insertion of the Vibrio cholerae phage CTXφ into the host chromosomes and the copy number control of budding yeast plasmids (13, 14, 15, 16, 17). A subset of YRs has been utilized as tools for directed genome manipulations with potentially important biotechnological and medical applications (18). In this overview of tyrosine site-specific recombination, we present our current understanding of the mechanism of the reaction from biochemical, chemical, structural and topological perspectives, and highlight the utility of this knowledge in addressing problems of fundamental importance in biology and in developing new technologies for biomedical engineering (see also chapters by M. Boocock, A. Landy, A. Segall, G. van Duyne, D. Mazel, F-X Barre, J. Gardner, and G. Chaconas).

THE RECOMBINATION REACTION: SYNAPTIC ORGANIZATION OF DNA PARTNERS AND STRAND EXCHANGE MECHANISM

The biochemically and structurally most well characterized YRs are phage λ integrase (λ Int), phage P1 coded Cre, Flp coded for by the Saccharomyces cerevisiae plasmid 2 micron circle and XerCD of Escherichia coli (10, 19, 20, 21, 22, 23). They have provided the templates for the chemical and conformational attributes of the strand cleavage and strand exchange steps during tyrosine recombination. The reaction is executed in the context of two core DNA target sites, each bound by two recombinase monomers, brought together in a synaptic complex by protein–protein interactions [Fig. 1(A)].

The association of a recombinase, a monomer in solution, with its binding element (a little over one turn of DNA) activates the scissile phosphate adjacent to it. The amino- and carboxyl-terminal domains of the recombinase cradle the DNA between them through a small number of base-specific and many more phosphate contacts. Recognition specificity is imparted to a significant degree through indirect readout. The 13 bp Flp binding element, for example, contains an A/T-rich segment with a characteristically narrow minor groove, and A to T changes within it are well tolerated with respect to binding (24). At the same time, C to G changes are detrimental to binding, and replacement of guanosine with inosine alleviates this negative effect by eliminating the obstructive 2-amino group from the minor groove. Within a DNA substrate, the scissile phosphates are positioned 6 to 8 bp apart (depending on individual systems) on opposite strands, specifying the extent of the strand exchange region. In general, two identical monomers of a recombinase occupy the two binding elements flanking the strand exchange region in a head-to-head (inverted) fashion. In rare instances, as with XerCD, a target site is bound by one monomer each of XerC and XerD.

Strand cleavage by the active site tyrosine nucleophile utilizes the type IB topoisomerase mechanism, yielding a 3’-phosphotyrosyl intermediate and an adjacent 5’-hydroxyl group (20). Strand exchange involves the nucleophilic attack by the 5’-hydroxyl group on the phosphotyrosyl bond across DNA partners to reseal the strand breaks in the recombinant configuration. The reaction is completed in two temporally separated cleavage-exchange steps, the first yielding a Holliday junction intermediate and the second resolving it into reciprocal recombinants.

INHIBITION OF TYROSINE RECOMBINATION BY AGENTS THAT TARGET HOLLIDAY JUNCTIONS

Short synthetic hexapeptides rich in aromatic amino acids inhibit tyrosine recombination by trapping the Holliday junction intermediate (25, 26, 27) (also chapter by A. Segall). The current model for peptide action, based on gel mobility shift and fluorescence quenching results, together with crystal structure data for a Cre recombinase-Holliday junction-peptide ternary complex, posits that the binding of a peptide dimer across the junction core stabilizes the junction in a nearly square-planar (but nonfunctional) conformation (27, 28, 29). More recent analyses of peptide–junction interactions by a combination of single molecule FRET (fluorescence resonance energy transfer), SAXS (small angle X-ray scattering) and gel mobility shifts suggest that peptide binding yields an ensemble of highly dynamic junction
conformations that do not fit the canonical square-planar and stacked X-conformations (unpublished observation). The induced conformational heterogeneity likely results from multiple stacking arrangements of aromatic amino acids with the bases surrounding the junction core, perhaps reflecting an intrinsic property of positively charged hydrophobic peptides. Peptide association with a protein–Holliday junction complex may inhibit subsequent reaction steps by inducing global changes in the junction conformation or local changes in the conformation of the bases surrounding the junction core.

**FIGURE 1** Tyrosine family site-specific recombination. (A) The two target sites, each bound by two recombinase monomers across the strand exchange region, are arranged within the recombination synapse in an almost perfectly planar, antiparallel fashion. The left and right arms of the sites are marked as L1, L2 and R1, R2, respectively. The reaction proceeds by the cleavage/exchange of one pair of strands to form a Holliday junction intermediate, isomerization of the junction, and exchange of the second pair of strands to give the recombinant products (L1R1 + L2R2 → L1R2 + L2R1). The scissile phosphates engaged by the “active” active sites at distinct stages of the reaction are indicated by the filled circles. (B) The “half-of-the-sites” activity, responsible for the two-step strand exchange mechanism, is revealed by the crystal structure of the Flp-DNA complex (34, 36). Within each recombination partner (left), the green Flp monomer (bound at R1 or R2) is poised to promote the cleavage of the scissile phosphate adjacent to it (red circle). The tyrosine nucleophile for cleavage is donated in trans by the neighboring Flp monomer (bound at L1 or L2; magenta). Following isomerization of the Holliday junction intermediate (right), there is a switch between the active and inactive Flp pairs, signifying the imminent cleavage of the scissile phosphates adjacent to Flp monomers bound at L1 and L2. The tyrosine nucleophiles are donated across DNA partners, in the R1 to L2 and R2 to L1 configuration. doi:10.1128/microbiolspec.MDNA3-0021-2014.f1
in the active site environment. Alternatively, peptide binding could accelerate protein dissociation from the junction, and then inhibit further reaction by inducing unfavorable junction conformations. The concept of inhibiting biologically important nucleic acid transactions by enhancing, rather than constraining, conformational freedom may be broadly applicable to peptide and nonpeptide ligands that recognize specific nucleic acid structures. In addition to inhibiting tyrosine recombination, hexapeptides also impede the unwinding of branched DNA structures by the RuvG helicase of E. coli, and interfere with Holliday junction resolution by the RuvABC complex \(^{(28)}\). Consistent with these properties, the inhibitory peptides appear to hold promise as potential antimicrobial agents \(^{(30)}\).

### SIMPLE AND COMPLEX YRs: CONTROLLING THE DIRECTIONALITY OF RECOMBINATION

Simple YRs such as Cre and Flp are not particular about DNA topology or target site orientation. They can act on supercoiled or nicked circles as well as linear molecules, and promote intra- and intermolecular reactions. They bring about DNA inversion between a pair of sites in head-to-head (inverted) orientation and DNA deletion between sites in head-to-tail (direct) orientation. More complex YRs (\(\lambda\) Int and XerCD, for example), depending on the reaction context, may require DNA supercoiling, and may utilize the interaction between accessory factors and their cognate sites to regulate the reaction and/or impart directionality to it. The crystal structures of \(\lambda\) Int tetramers bound to synapsed DNA partners and the Holliday junction intermediate, together with biochemical data, suggest how interactions of the amino-terminal domains of Int with the ‘arm-type’ sequences (which also include multiple binding sites for the accessory proteins: IHF, Xis and Fis) can stabilize 2-fold symmetric configurations of the recombination complex \(^{(21, 31)}\). The cumulative DNA–protein and protein–protein interactions thus coordinate Int activity at the core recombination sites as well as bias strand exchange towards a particular outcome. Consistent with this model, when the amino-terminal domain of \(\lambda\) Int is fused to the normally unregulated and bias-free Cre, the latter acquires the regulatory features and directionality of Int \(^{(32, 33)}\). The action of the Int-Cre chimera on attenuated core target sites containing appropriate embellishments with the arm-type sequences is controlled by the accessory factors as if recombination were being performed by native Int. The conversion of a simple recombinase, whose origin likely traces back to an ancestral type IB topoisomerase, into a complex one by just the addition of a peptide domain suggests a possible “self-promoting” evolutionary scheme for the emergence of the latter class of recombinases. The relevant gene fusion may be performed by the recombinase itself via low frequency crossover events between suitably positioned secondary target sites, which are fortuitously scattered within a genome and may be harbored by the recombinase gene. Alternatively, the “complexity” domain may be acquired via the action of the host’s recombination machinery. \(\lambda\) Int may thus be a representative of the evolutionary trajectory from topoisomerase to simple and then complex recombinases (see chapter by A. Landy).

### HALF-OF-THE-SITES ACTIVITY OF THE YR ACTIVE SITE

Within the recombination synapse, which has a 2-fold symmetry, only two of the four potential active sites are active at any one time \(^{(34, 35)}\) \[Fig. 1(B)\]. The two active sites responsible for the first cleavage-exchange step become inactive following the isomerization of the Holliday junction intermediate. The other two active sites, which now become activated, resolve the junction into recombinant products. This “half-of-the-sites” activity accounts for the two-step, single-strand exchange mechanism of recombination. Consistent with this mechanism, three Cre or Flp recombinase monomers bound to a three-armed DNA substrate (Y-junction) can yield two functional active sites capable of resolving the junction into a linear and a hairpin product \(^{(36, 37, 38)}\).

### THE LOCAL GEOMETRY OF PARTNER SITES WITHIN THE RECOMBINATION SYNAPSE

The DNA target sites are almost entirely planar in their paired state, and are arranged in an antiparallel fashion \[Fig. 1(A)\]. Topological analyses and crystal structure data support this synapse geometry, with strand cleavage and exchange occurring in a diagonal fashion \(^{(31, 34, 39, 40, 41, 42, 43, 44)}\). However, a few experiments based on electron microscopy, proximity of DNA ends reported by ligation, and a combination of atomic force microscopy, tangle analysis (see below under “Topological and chiral features of tyrosine recombination”) and modeling suggest parallel arrangement of sites or their nonplanar configuration with a potential tetrahedral geometry for the Holliday junction intermediate \(^{(45, 46, 47)}\). These could represent transient or intermediate states that precede the functional synapse or...
isomerization of the Holliday junction. Or, they could be comprised of aberrant complexes. The reactive orientation of key catalytic residues with respect to the scissile phosphates in the crystal structures strongly imply that the antiparallel disposition of the partner sites revealed by them represents the functional geometry of the recombination synapse (19, 31, 39).

THE YR ACTIVE SITE: KEY CATALYTIC RESIDUES

The signature active site motif of YRs consists of a tyrosine nucleophile assisted by an invariant or highly conserved catalytic pentad: Arg, Lys, His, Arg, and His/Trp (23). In addition, a sixth conserved residue, Asp/Glu, appears to contribute indirectly to transition state stabilization by hydrogen bonding to catalytic residues and promoting the integrity of the active site (39). In Flp, the catalytic hexad is comprised of Arg-191, Asp-194, Lys-223, His-305, Arg-308, and Trp-330 with Tyr-343 providing the cleavage nucleophile (Fig. 2). In Cre (see chapter by G. van Duyne), the corresponding residues are Arg-173, Glu-176, Lys-201, His-289, Arg-292, Trp-315, and Tyr-324. In λ Int, the second and sixth conserved positions are aspartic acid (Asp-215) and histidine (His-333), respectively. Mutational analyses combined with structural data have provided insights into the mechanistic roles of several of these residues.

As would be expected from chemical principles, the two invariant arginine residues balance the negative charges on the nonbridging oxygen atoms of the scissile phosphate group (31, 34, 39). In Cre, Lys-201 appears to function as the general acid that stabilizes the leaving 5′-hydroxyl group during strand cleavage. The absence of this residue can be rescued when the leaving group pKₐ is decreased by substituting the 5′-oxygen by sulfur (5′-thiolate) (48). The potential general acid function of Lys-223 of Flp has not been similarly tested. In the type IB vaccinia topoisomerase, Lys-167 (which corresponds to Lys-201 of Cre) collaborates with Arg-130 (corresponding to Arg-173 of Cre and Arg-191 of Flp) to facilitate leaving group departure, perhaps by a proton relay mechanism (49, 50). His-305 of Flp may serve as the general base that abstracts the proton from Tyr-343 to activate it as a nucleophile (51). A tyrosine mimic with reduced pKₐ (3-fluorotyrosine, pKₐ = 8.2 as opposed to tyrosine, pKₐ = 10.0), when supplied in the context of a short native peptide, can restore the cleavage potential of Flp(H305Q) to a large extent. The predominant majority of YRs contain a histidine rather than tryptophan at the final hexad position. The hydrogen bonding between the indole-nitrogen of Trp-315 and the scissile phosphate observed in the Cre crystal structure (52) seemed to suggest that either histidine or tryptophan at this position helps catalysis through their hydrogen bonding potential. However, this is not the case for Cre or Flp, as replacement of Trp-315 or Trp-330, respectively, by histidine results in lower recombination activity compared to replacement by phenylalanine or tyrosine in Cre and phenylalanine in Flp (39, 53). Consistent with these biochemical results, structural data for Flp suggest that hydrophobic/van der Waals contacts made by Trp-330, over a hydrophobic surface of ~380 Å², help position Tyr-343 in its active orientation (53) (Fig. 3). In addition, Trp-330 appears to play a secondary role in stabilizing the 5′-hydroxyl leaving group, as suggested by the stimulation in the cleavage activities of Flp mutants lacking this residue on 5′-thiolate substrates (54). The vanadate transition state mimic structure of Cre-DNA reveals Trp-315 to be located on a turn between two helices (αL and αM), providing a sizable hydrophobic surface on to which the αM helix carrying the Tyr-324 nucleophile docks through van der Waals contacts (39) (see chapter by G. van Duyne).

HIDDEN RNA CLEAVAGE ACTIVITIES OF THE YR ACTIVE SITE

When a ribonucleotide is substituted at the cleavage position or immediately 3′ to it within the Flp target (FRT) site, two latent RNase activities of Flp can be unveiled (55, 56, 57). The 2′-hydroxyl, when present as part of the cleavage position nucleotide, can attack the phosphotyrosyl intermediate formed by strand cleavage to give a 2′,3′ cyclic phosphate. Subsequent attack of the cyclic bond by a water nucleophile gives the 3′-phosphate as the end product. This activity, which closely follows the recombination mechanism, has been termed type I RNase. When the 2′-OH is placed on the nucleotide adjacent to the cleavage position, it directly attacks the scissile phosphate to yield the cyclic phosphate intermediate, which is then hydrolyzed to the 3′-phosphate. This activity, which resembles the classical pancreatic RNase mechanism, has been termed type II RNase. Perhaps the latent RNase activities are the relics of the evolutionary progression of Flp from an elementary nuclease to a recombinase, likely via a topoisomerase. When the interaction between two Flp monomers bound to an FRT site is weakened by increasing the length of the strand exchange region, a weak topoisomerase activity of Flp can be unmasked. The type I, but not the type II, RNase activity has been detected in Cre as well.
FIGURE 2 Organization of conserved catalytic residues within the recombinase active site. The arrangements of the catalytic hexad (Arg-Asp/Glu-Lys-His-Arg-His/Trp) and the tyrosine nucleophile in Cre, Flp and λ Int active sites are shown (31, 34, 36, 52, 94, 147). The states of the active site with the scissile phosphate uncleaved and cleaved are shown at the left and right, respectively. The role of the conserved Asp/Glu of the hexad in transition state stabilization is likely indirect, by promoting the structural integrity of the active site. doi:10.1128/microbiolspec.MDNA3-0021-2014.f2
The ability of 2′-hydroxyl groups to compete effectively with the tyrosine nucleophile (in the case of Flp) and with the 5′-hydroxyl group (in the case of Cre and Flp), as suggested by their latent RNase activities, speaks to the considerable catalytic flexibility of the tyrosine recombinase active site. These activities also expose the potential threat to recombination by errant nucleophiles that might gain entry into the active site.

**ASSEMBLY OF THE YR ACTIVE SITE IN CIS OR IN TRANS**

In general, the active site of a tyrosine recombinase is assembled entirely within a monomer, although its strand cleavage activity may be stimulated by allosteric contact with an adjacent monomer. Flp and the related subfamily of YRs coded for by 2 micron-like plasmids of budding yeasts are unusual in that they assemble an active site at the interface of two neighboring monomers (54, 59, 60) [Fig. 1(B); Fig. 3]. Biochemical and structural evidence suggests that the integrase of SSV1, a virus that infects the extremely thermophilic archaeon *Sulfolobus shibate*, may also harbor a shared active site (61, 62). A cis active site (the Cre active site, for example) is responsible for the activation and cleavage of the scissile phosphate engaged by it. A trans active site, exemplified by that of Flp, activates the scissile phosphate but relies on the tyrosine nucleophile donated to it for strand cleavage. There are two trans modes of DNA cleavage. For the first strand exchange step and formation of the Holliday junction, the tyrosine nucleophile performs cleavage across the strand exchange region within a DNA substrate [Fig. 1(B), left]. For the resolution of the Holliday junction and formation of the recombinant products, the cleavage by tyrosine occurs across partner substrates [Fig. 1(B), right].

Comparison of the crystal structures of Cre and Flp synaptic structures suggests how a simple switch in the connectivity of two helices can switch a cis active site into a trans active site or vice versa. (23, 34). In the shared active site, the Tyr-343 from an Flp monomer gains entry into the proactive site of the second monomer as part of the protruding “M” helix. Trp-330 of the hexad in the proactive site of Flp may play a particularly important role in helping to dock the M helix by packing against it through contacts with Ser-336, Ala-339, and Tyr-343 (53) [Fig. 3(A)]. These interactions are further augmented by the nearly perfect stacking of His-309 from the recipient Flp over Tyr-343, with likely assistance from His-345 (donor) and His305 (recipient) (34, 63). The extremely weak recombination activity of Flp...
(W330A) can be restored to nearly wild type level by a second mutation that changes Ala-339 to methionine (54). The longer side-chain of Met-339 located in the M helix can compensate for the lack of Trp-330 in the proactive site, further highlighting the importance of interprotomer hydrophobic interactions in the assembly of the trans active site [Fig. 3(B)].

From a purely mechanistic point of view, there is apparently little advantage of a cis active site over a trans active site or vice versa. By the “Cheshire cat” paradigm, if one were to erase all the amino acid residues from Cre and Flp structures, except for the tyrosine nucleophile and its principal catalytic cohort, the ‘catalytic grins’ of the two recombinases would look almost identical (64, 65) (Fig. 2). However, the trans active site offers considerable advantages in the analysis of recombination mechanism. For example, strand cleavage can be performed by providing exogenous nucleophiles such as hydrogen peroxide or tyramine or by supplying chemically modified tyrosines embedded in a short native Flp peptide (51, 66, 67). As already noted, the potential general base/acid role of His305 of Flp has been inferred from the ability of tyrosine analogs with lowered pKas to confer cleavage competence on Flp(H305Q).

From a functional perspective, each type of active site may have its strengths and weaknesses. The trans active site will delay DNA cleavage until a target site has been occupied by two monomers of the recombinase. However, the time delay between the binding of the two monomers may allow rogue nucleophiles, the abundant water nucleophile for example, to attack the activated phosphodiester bond (68). Since the cis active site positions all the catalytic residues, including the tyrosine nucleophile, in concert around the scissile phosphate, the chances for aberrant strand cleavage are minimized. However, cleavage could potentially occur even before the full occupancy of a DNA site by the recombinase. Premature strand breaks may be minimized if the cleavage potential of the cis active site is activated by allosteric interactions between recombinase monomers within a DNA site or between partner sites within a recombination synapse. For Flp, whose physiological function is to trigger plasmid DNA amplification by a replication-coupled recombination event (14, 17), DNA cleavage in trans may hold special significance. Should an Flp monomer covalently linked to the cleaved phosphate be dislodged from its binding element by the replication machinery, and be unfolded or partially degraded by a protease as a consequence, the Flp monomer bound to the other binding element will be able to promote healing of the DNA break via ligation.

**REQUIREMENT OF HOMOLOGY BETWEEN THE STRAND EXCHANGE REGIONS OF RECOMBINATION PARTNERS**

According to the generally accepted paradigm, based on evidence from the λ Int, Cre and Flp systems, successful recombination requires perfect homology in the strand exchange regions of the DNA partners, even though the sequence per se of the exchanged segment can be altered in a number of ways without affecting reaction efficiency. The original notion that homology promotes end-to-end branch migration of the Holliday junction through the strand exchange region (69, 70) has been discounted by biochemical and structural evidence (31, 34, 52, 71, 72, 73, 74). The cleaved strands are swapped in a segmental fashion, perhaps as triplets during the first and strand exchange steps (74). Nonhomology would disfavor stable strand exchange, as mismatches in DNA hinder the ligation reaction (73). In this model, recombination is blocked by nonhomology either because the Holliday junction is not formed or because the junction with mismatched DNA is quickly resolved in the parental mode. For Flp, which mediates the exchange of 8 bp rather than 6 bp (the extent of exchange in the Cre system), the strand swaps at the initiation and termination steps of recombination may be separated by an intermediate step of limited branch migration through the central base pairs (75).

The strict requirement of homology in strand exchange has been called into question as the mechanisms of more YRs have been brought to light, in particular, the integrases of integrative conjugative elements (ICEs), also referred to as conjugative transposons (16) (see chapter by J. Gardner). These enzymes mediate strand exchange across overlap regions that include 5 to 6 bp nonhomologies. IntDOT, the integrase of the *Bacteroides* conjugative transposon CTnDOT, utilizes a 2 bp homology within a 7 bp overlap region for the first exchange step, and carries out the second exchange in a homology-independent manner [Fig. 4(A)]. The resulting heteroduplex DNA is resolved by replication.

The CTXφ phage of *Vibrio cholera* manipulates the XerCD recombinase of its host for its integration in a rather unusual reaction that also challenges the homology rule in its conventional sense (13) [Fig. 4(B)] (see chapter by F-X Barre). The + strand of the phage DNA folds itself into a forked hairpin structure to generate a pseudo XerCD target site within it. The first strand exchange between the phage and chromosome target sites is mediated by XerC, and utilizes 3 bp of homology adjacent to its cleavage site. The nonhomology adjacent to XerD stops the reaction at this pseudo-Holliday...
junction stage. Resolution of this structure by replication generates a chromosome with an integrated copy of the phage. As the lysogen is not flanked by functional recombination sites, the integration reaction is irreversible, proscribing the excision of the replicative form of the phage. However, tandem copies of the lysogen generated by successive integration events permits the production of + single-stranded phage genome by a rolling circle type of replication. An analogous strategy of integration via single-strand exchange by a tyrosine recombinase, followed by replication-mediated resolution of the resulting Holliday junction, is also devised during the capture of exogenous gene cassettes by integrons (76), which are important not only for their role in the spread of antibiotic resistance but also for their potential relevance to bacterial genome evolution. As in the case of the *V. cholerae* phage integration, the recombination target site on the cassette is assembled by the folding of a single-stranded DNA region. The phage and integron systems, rather than breaching the homology rule outright, seem to bend it by cleverly manipulating the strand exchange reaction in their favor.

Even the archetypal YRs may violate the homology rule in its strictest sense. Analysis of the Cre reaction between a wild type LoxP (the Cre target site) and mutant LoxPs containing single bp substitutions of the

**FIGURE 4** Challenges to the homology rule during tyrosine site-specific recombination. (A) The integrase (IntDOT) of the conjugative transposon CTnDOT catalyzes exchange of both strands between target sites that contain five consecutive nonhomologous positions within the 7 bp segments swapped between them (16). (B) The folded form of the “+” strand of the CTXφ phage contains an imposter target site for XerCD recombinase of its host bacterium, *V. cholera*. Single-strand exchange mediated by the XerC active site between the phage DNA and the bacterial chromosome results in phage integration (13). The heteroduplex integrant in (A) and the pseudo-Holliday junction in (B) are likely resolved via replication. The flat horizontal arrows indicate recombinase binding sites. The short vertical arrows denote points of strand cleavage. doi:10.1128/microbiolspec.MDNA3-0021-2014.f4
6 bp strand exchange region reveals several types of outcomes: significant amounts of the Holliday junction intermediate without detectable recombinant product, small amounts of the product with higher amounts of the Holliday junction, small amounts of the Holliday junction with higher amounts of the product and small amounts of the product with no detectable Holliday junction (77). The overall strand exchange, the sum of the Holliday junction and recombinant yields, is most diminished for substitutions adjacent to the scissile phosphate, whose cleavage initiates the first strand exchange step, with one exception. The T to A transversion at position 2 from the initiation end gives a modest amount of the product with much smaller amounts of the Holliday junction. The accumulation of the Holliday junction as nonhomology shifts from the initiation-proximal positions to the central and termination-proximal positions of the strand exchange region is consistent with the normal execution of the first strand exchange step, while the second exchange step is blocked or severely impeded by nonhomology. Stable single-strand exchange by Cre, dictated by the location of nonhomology, is thus somewhat analogous to the formation of the pseudo-Holliday junction by XerCD during CTXφ integration.

NONHOMOLOGY INDUCED KNOTTING OF SUPERCOILED PLASMIDS BY A YR

An even more flagrant violation of the homology rule is brought to light by Flp reactions between two FRT sites, nonhomologous at the central two positions of their 8 bp strand exchange regions and located within negatively supercoiled plasmids (44, 78). While no stable recombinant products result, the reaction ties the plasmids into knots of wide ranging complexity, but all in their parental configuration. When the two sites are in head-to-head orientation (with respect to the six homologous bp), the knots are even numbered; when the sites are in head-to-tail orientation, the knots are odd numbered. The knotting may be explained by two or repeated even rounds of recombination, giving rise initially to unstable recombinants containing mismatched base pairs that then rapidly recombine back to the parental form.

An obvious question is whether the observed knot complexity is due to DNA crossings added during an iterated series of recombination events occurring within a given synapse. A similar increase in topological complexity has been described for the serine recombinases when they attempt to recombine sites that harbor nonhomology in the overlap region (79, 80). However, members of this recombinase family follow a completely different mechanism (see chapter by W. M. Stark). They make concerted double-strand cuts within target sites arranged in a parallel fashion, promote right-handed rotation of the broken DNA through 180 degrees, and reseal the strands in the recombinant configuration. Mismatches between partner strands impede or reverse joining and encourage a second 180 degree rotation to restore complementarity, which favors joining. A repetition of these dual half-rotation steps will progressively increase the complexity of products. Can nonhomology alter the normal synapse geometry of tyrosine recombination and processively generate products of increasing complexity?

An alternative explanation for the knotting reaction by Flp is that the complexity of the knots reflects the topological complexity of the plasmid substrate from which they are generated. Since the pairing of FRT sites occurs by random collision, a range of supercoils (crossings between the two DNA domains bordered by the sites; blue × red in Fig. 5) can be trapped by the synapse. The antiparallel geometry of the FRT sites requires an odd number of such crossings between head-to-head sites and even number of crossings between head-to-tail sites (Fig. 5). Depending on the number of trapped interdomainal crossings, the first recombination event will generate an unknotted inversion circle plus 3-, 5-, 7- etc. crossing knots from the head-to-head sites [Fig. 5(A)]. Similarly, the products from the head-to-tail sites will be a pair of unlinked deletion circles plus 2-, 4-, 6- etc. crossing catenanes (linked circles) [Fig. 5(B)]. When the parental FRT sites are nonhomologous, these products contain mismatches, and are prone to a second round of recombination after dissociation of the original synapse. The addition of one more crossing during this step will convert the knots with odd number crossings from the inversion reaction to knots with even number crossings (4, 6, 8 etc.) [Fig. 5(A)]. Similarly, the catenanes from the deletion reaction will be converted to fusion knots with odd number crossings (3, 5, 7 etc.) [Fig. 5(B)]. The prediction then is that when the synapse topology is simplified and made unique, the product topology must be correspondingly simple and unique. This indeed is the case (44). When Flp reaction is carried out after assembling the Tn3 resolvase synapse [which traps precisely three interdomainal negative supercoils, as in Fig. 5(A)], and taking care to minimize random entrapment of supercoils, the product yielded by the head-to-head sites is predominantly a 4-noded knot; that yielded by the head-to-tail sites is
predominantly a 5-noded knot (44). The topologies of the products from corresponding reactions between two native (homologous) FRT sites are a 3-noded knot and a 4-noded catenane. The difference of one in the crossing numbers between the knot and the catenane is consistent with the need to arrange the sites in the

**FIGURE 5** Flp-mediated knotting of supercoiled plasmids by recombination between two FRT sites harboring nonhomology within the strand exchange region. (A) The first recombination event between two head-to-head (inverted) FRT sites from a synapse containing an odd number of interdomainal (blue × red) supercoil crossings will generate a torus knot with the same number of crossings. The product from a synapse with one blue × red crossing will be an unknotted inversion circle, as it takes a minimum of three crossings to form the simplest knot. In the example shown, a 3-noded knot is formed from a 3-crossing synapse. A second recombination event after dissociation of the first synapse, and the assembly of a de novo synapse, can give rise to a twist knot with four crossings. (B) For FRT sites in head-to-tail (direct) orientation, the first recombination event from a synapse with an even number of interdomainal crossings yields a catenane with the same number of crossings. The product from a synapse with no crossings will be two unlinked deletion circles. The diagram illustrates the formation of a 4-noded catenane from a 4-crossing synapse. A second round of dissociative recombination can convert the 4-noded catenane into a 5-noded knot. In the reactions shown in (A) and (B), intradomainal supercoils (blue × blue or red × red crossings) are omitted for clarity, as they do not contribute to knot or catenane crossings. The products from the second rounds of recombination revert to the parental configuration. The noncomplementarity in the product formed by recombination between FRT sites containing nonhomology in their strand exchange regions encourages a second recombination event that restores base pairing and parental DNA configuration (44). doi:10.1128/microbiolspec.MDNA3-0021-2014.f5
same functional geometry, antiparallel, for them to recombine [Fig. 5(B)]. In addition to the three crossings anchored by resolvase, a fourth crossing must be trapped from the supercoiled plasmid substrate to keep the head-to-tail sites antiparallel (see also the section on “Difference topology”). Thus, nonhomology does not block recombination by Flp; nor does it induce processive recombination by altering the normal reaction mechanism. The unstable (mismatched) recombinants resulting from the first recombination event are restored to the more stable parental state by a second dissociative recombination event.

PROBING ACTIVE SITE MECHANISM USING CHEMICAL MODIFICATIONS OF THE SCISSILE PHOSPHATE GROUP

Mechanistic analysis of strand breakage and joining reactions in nucleic acids has greatly benefited from chemical modifications of the phosphate group in the nonbridging oxygen atoms to alter its electronegativity and/or stereochemistry, and in the bridging oxygen atoms to manipulate leaving group properties. The potential general acid role for Lys-201 and a subsidiary role for Trp-330 in leaving group stabilization during strand cleavage by Cre and Flp, respectively, have been revealed with the help of 5′-thiolate substrates (48, 54). Shuman and colleagues have successfully exploited phosphorothioate (replacement of a nonbridging oxygen by sulfur), methyl phosphonate (MeP; replacement of a nonbridging oxygen by the methyl group) and 5′-thiolate substrates to investigate the active site mechanisms of vaccinia topoisomerase (81, 82, 83, 84).

Recent studies employing MeP-substrates [Fig. 6(A)] have revealed active site attributes of Cre and Flp that could not have been deduced from reactions of native phosphate containing substrates. These analyses have been performed predominantly using half-site substrates containing a single scissile phosphate or a modified scissile phosphate [Fig. 6(B)], together with recombinase variants harboring specific active site mutations. The chemical synthesis of MeP-half-sites is considerably

FIGURE 6 Reactions of half-sites containing methylphosphonate substitution at the scissile phosphate position. (A) The structures of methylphosphonate (MeP) are compared to that of the native phosphate in DNA. There are two possible stereoisomers of MeP (R<sub>P</sub> or S<sub>P</sub>). (B) The possible reactions of a half-site containing MeP at the scissile phosphate position are illustrated. The 5′-hydroxyl group on the bottom strand of the half-site is blocked by phosphorylation to prevent it from taking part in a pseudo-joining reaction. Attack of the MeP bond by the active site tyrosine will give the MeP-tyrosyl intermediate, which may undergo slow hydrolysis. The hydrolysis product may also be formed by direct water attack on the MeP bond. The two-step (type I) and single-step (type II) reaction pathways are mechanistically analogous to the type I and type II RNA cleavage activities of Flp (see text). doi:10.1128/microbiolspec.MDNA3-0021-2014.f6

A

Phosphodiester

O 3′ DNA

O 5′ DNA

R<sub>P</sub> methylphosphonate

O 3′ DNA

O 5′ DNA

S<sub>P</sub> methylphosphonate

B

5′ HO-TTT3′

5′

Mep

AAAGATCT(p) 5′

3′

MepTTT 3′

AAAGATCT(p) 5′

5′ HO-TTT3′
An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective

easier than that of full-sites. Although the half-site reaction involves the breakage of one scissile phosphate within a substrate molecule, it faithfully preserves the chemical mechanism of the normal reaction. Associations of a recombinase-bound half-site can give rise to dimers, trimers and tetramers (85), so that the shared active site assembly and the trans mode of DNA cleavage are obeyed during Flp half-site reactions.

The reactivity of Flp variants on MeP-substrates demonstrates that neutralization of the phosphate negative charge in its ground state permits transition state stabilization in the absence of one of the two conserved arginines (either Arg-191 or Arg-308) (68, 86). Flp (R191A) and Flp(R308A) are active in the MeP reaction [Fig. 7(A)], while both these variants are almost completely inactive on phosphate containing DNA substrates. The electrostatic suppression of the lack of a positive charge in the recombinase active site by a compensatory charge substitution in the scissile phosphate of the DNA substrate has been demonstrated for

**FIGURE 7** Distinct activities of Flp(R191A) and Flp(R308A) on an MeP-half-site. (A) Flp (R191A) cleaves the MeP-half-site (S) using Tyr-343 to form the protein–DNA adduct (revealed by SDS-PAGE; top) (86). This intermediate is converted to the hydrolysis product (HP) (revealed by denaturing PAGE; bottom) in a subsequent slow reaction. Flp(R308A), by contrast, yields the hydrolysis product directly, without going through the MeP-tyrosyl intermediate (68). (B) The binding of an Flp monomer to FRT activates the scissile phosphate, leaving it exposed until the binding of a second Flp monomer delivers Tyr-343 to the active site in trans. (C) Concomitant with the binding of a Cre monomer to the LoxP site, Tyr-324 engages the scissile phosphate in cis, thus protecting it against direct water attack. (D) As vaccinia topoisomerase, like Cre, assembles its active site in cis, the scissile phosphate is protected at the strand cleavage step during DNA relaxation. However, the protein’s grip on DNA is loosened during the strand rotation step, leaving the phosphotyrosyl bond vulnerable to attack by water. doi:10.1128/microbiolspec.MDNA3-0021-2014.f7
the Cre recombinase as well (87, 88). Not only do Cre (R173A) and Cre(R292A) yield strand cleavage in an MeP-half-site, the double mutant Cre(R173A, R292A) also mediates this reaction. Presumably, the overall electrophilic character of the Cre active site is sufficient to neutralize the diminished negative charge present in the MeP, compared to the phosphate, transition state.

As the methyl substitution of one of the nonbridging oxygen atoms turns the normally symmetric phosphate group into an asymmetric center [Fig. 6(A)], an additional utility of the MeP substrates is in unveiling the stereochemical course of the recombination reaction. Stereochemically pure $R_P$ and $S_P$ forms of the MeP substrates are currently being used to dissect the individual stereochemical contributions of Arg-191 and Arg-308 in Flp, and to prove how other members of the catalytic hexad might influence these contributions.

**DISTINCT ACTIVITIES OF FLP(R191A) AND FLP(R308A) IN THE MEP REACTION**

The absence of Arg-191 or Arg-308 has strikingly different effects on the activity of Flp on an MeP-half-site [Fig. 7(A)] (68, 86). Flp(R308A) does not utilize the Tyr-343 nucleophile, but promotes direct hydrolysis of the MeP bond. Consistent with this mechanism, the double mutant Flp(R308A,Y343F) also yields the hydrolysis product with similar kinetics and $V_{\text{max}}$ (maximal velocity) as Flp(R308A). Apparently, the lack of Arg-308 permits the abundant water nucleophile to access the reaction center, where it out competes Tyr-343 to give a dead-end product. However, the possibility that Arg-308 is required for the positioning of Tyr-343 or its nucleophilic activation cannot be ruled out. The corresponding arginines, Arg-292 of Cre and Arg-410 of Leishmania topoisomerase I, are hydrogen bonded to the catalytic tyrosine in their respective vanadate-transition state structures (39, 89). Flp(R191A), by contrast, utilizes Tyr-343 as the cleavage nucleophile to yield the tyrosyl intermediate. Direct hydrolysis in the Flp(R191A) reaction is only a minor component. Cre(R173A) and Cre(R292A) are mechanistically similar to Flp(R191A) in that they promote Tyr-324-mediated cleavage of MeP (87, 88).

**POTENTIAL ROLES FOR ACTIVE SITE AND PHOSPHATE ELECTROSTATICS IN PREVENTING FUTILE PHOSPHORYL TRANSFER**

As suggested by the MeP reactions, in addition to balancing the phosphate negative charge, Arg-308 of Flp appears to protect the normal reaction course from abortive hydrolysis, perhaps by electrostatically mis-orienting water nucleophile (which is a dipole) from the activated phosphate. The phosphotyrosyl bond formed by vaccinia topoisomerase during DNA relaxation is apparently protected from hydrolysis by an analogous mechanism, utilizing the negative charge on the scissile phosphate (84). Furthermore, the potential role of the Arg-308 side-chain in orienting or activating Tyr-343 (see above under “Distinct activities of Flp(R191A) and Flp(R308A) in the MeP reaction”) suggests an alternative or collaborative mechanism for preventing futile break- age of the DNA backbone by increasing the local concentration of the tyrosine nucleophile. As noted earlier, the need to shield the scissile phosphate from extraneous nucleophiles would be more critical for Flp because of its $trans$ active site. The scissile phosphate, activated by the proactive site of a bound Flp monomer, stays exposed until Tyr-343 is provided in $trans$ [Fig. 7(B)]. Binding by a Cre or topoisomerase monomer to DNA and the alignment of the tyrosine nucleophile with respect to the scissile phosphate would be nearly concomitant events because of their $cis$ active sites [Fig. 7(C)]. In the case of the topoisomerase, which acts as a monomer, the strand rotation step may open the phosphotyrosyl bond to attack by water [Fig. 7(D)], which is prevented by phosphate electrostatics. Such a protective mechanism is likely unnecessary for the recombinases, as the tight organization of the recombinase tetramer-DNA complex (Fig. 1) and the dynamics of strand exchange within it preclude water from accessing the phosphotyrosyl bond. The extrusion of the cleaved strand into the center of the “strand exchange cavity” seen in the structure of the Cre-recombination synapse (52) would be consistent with strand swap being nearly concomitant with strand cleavage.

**TYROSINE RECOMBINATION STEP-BY-STEP FROM START TO FINISH: SINGLE MOLECULE ANALYSIS**

Single molecule analysis of tyrosine recombination using real-time tethered particle motion (TPM), tethered fluorophore motion (TFM) and fluorescence energy transfer (FRET) have provided deeper insights into the pre-chemical, chemical, and conformational steps of the reaction pathway by revealing transient states as well as long- and short-range movements of DNA (90, 91, 92, 93). The results of these studies reveal interesting similarities and contrasts among Cre, Flp and $\lambda$ Int. The kinetics of recombinase binding to target sites and the...
pairing of bound sites are quite fast in all three cases, ruling out intrinsic barriers to synapsis, at least in vitro. There is a strong commitment to recombination following the association of Flp with the FRT sites. The formation of nonproductive complexes (those that do not synapse) and wayward complexes (those that do not form the Holliday junction intermediate or complete recombination after synapsis) constitute only minor detractions from the productive pathway. The stability of the synapse is enhanced by strand cleavage in the case of Flp and λ Int. However, Cre forms stable synapse even in the absence of strand cleavage. Recombination by Flp is efficient, and the frequency of occurrence of the Holliday junction intermediate is quite low. λ Int exhibits a strong and early commitment to a directed reaction path, likely assisted by its accessory factors bound to their cognate sites. Unidirectionality of an initiated recombination event would be a desirable attribute in vivo in bringing about the desired DNA rearrangement, without reversing course midway through a reaction. The Holliday junction formed during Cre recombination, however, is long lived, thanks to a rate-limiting step that follows its isomerization. This kinetic barrier affords the opportunity for the reaction to be interrupted and to go backwards, at least in vitro. It is possible that the in vitro Cre reaction fails to recapitulate the native regulatory features of recombination occurring within the P1 phage genome organized into a nucleoprotein complex.

**SINGLE MOLECULE TPM AND FRET ANALYSES AS PROBES FOR THE GEOMETRY OF SITE-PAIRING AND ORDER OF STRAND EXCHANGE**

TPM analysis is based on the rationale that the Brownian motion (BM) amplitude of a small polystyrene bead (~200 nm in diameter) attached to one end of a DNA molecule, whose other end is held in place, will be determined by the length of the DNA (Fig. 8). Since individual steps of recombination (binding of recombinase to target sites and their bending, synapsis of two bound sites by DNA looping, formation of the Holliday junction, and excision of DNA between two head-to-tail sites) are accompanied by characteristic changes in DNA length, TPM is well suited for the stepwise analysis of recombination. The pre- and postchemical changes in DNA length can be distinguished by challenging the reaction with SDS. Upon protein dissociation, the prechemical states and completed inversion reactions (from head-to-head sites) will return to the length of the substrate DNA molecules. The Holliday intermediates from the inversion or the excision reaction and the linear product from the latter will retain their reduced “tether” lengths. By performing the excision and inversion reactions in parallel using DNA substrates identical in length and in the location and spacing of the recombination sites, a complete analysis of the reaction path is possible. TPM is also useful for verifying the geometry of a pair of target sites within the recombination synapse, as described below.

For sites in head-to-head orientation, the entry and exit points of the DNA will be at the same end of the synapse if the sites are aligned in a parallel fashion. If the sites are in antiparallel alignment, the DNA will enter and exit the synapse from opposite ends.
ends. These situations will be reversed for a pair of sites in the head-to-tail orientation [Fig. 9(B)]. The proximal disposition of the entry and exit points imposes a stronger constraint on the DNA than their distal configuration, and makes it effectively shorter by a small amount. A significant difference in the BM amplitudes of two DNA molecules of identical length harboring a pair of equally spaced recombination sites, head-to-head in one case and head-to-tail in the other, would indicate a preferred geometry of the synapse. doi:10.1128/MicrobiolSpectrum.MDNA3-0021-2014.f9

**FIGURE 9** Effect of synapse geometry on the BM amplitude of DNA. (A) The DNA contours for a pair of synapsed head-to-head sites are outlined for their alignment in parallel (left) or antiparallel (right) geometry. (B) Similar diagrams as in (A) represent the antiparallel (left) and parallel (right) synaptic configurations for head-to-tail sites. The effective length of DNA is slightly larger when its entry and exit points are at opposite ends of the synapse than when they are at the same end. For two DNA substrates that differ only in the relative orientation of the recombination sites, a difference in the BM amplitudes of synapsed head-to-head versus head-to-tail sites signifies a preferred geometry of the synapse.

Geometry is restricted to being either parallel or antiparallel by a short single-stranded tether joining them (44) (Fig. 10). A change in the FRET state, in the expected direction, upon binding of Flp(Y343F) is observed only for the pair of FRT sites constrained in the antiparallel sense.

Given the approximate 2-fold symmetry of the core recombination sites, one might have imagined that they would synapse in a parallel or antiparallel fashion, even if only one of the two arrangements was productive for recombination. Topological and FRET results argue for preferred antiparallel synapsis of FRT sites even in the absence of the chemical steps of recombination (41, 44). Perhaps an asymmetric DNA bend within the strand exchange region of an Flp bound FRT site may preclude two similarly bent sites from occupying the synapse in a parallel fashion. A sharp bend located at a single bp step at one end of the strand exchange region has been observed in the structures of LoxP complexed with cleavage-incompetent mutants of Cre (94).

The synapse geometry raises the question of “order of strand exchange” during recombination. Depending on the location of the asymmetric bend with respect to the scissile phosphate, there are two geometrically equivalent and chemically competent configurations of the antiparallel synapse. One would correspond to “top strand” cleavage, and the other to “bottom strand” cleavage [Fig. 11(A)]. If one of the two synaptic configurations is preferred, the order of strand cleavage/exchange will reflect this preference. Current evidence suggests that Flp performs strand exchange without obvious bias (95, 96), indicating that the two modes of antiparallel synapsis are equally likely. Cre, by contrast, performs ordered strand exchange. FRET analysis with donor–acceptor dye pairs suitably positioned with respect to the strand exchange region demonstrates a preferred synapse, the DNA bend within which is consistent with the biochemically mapped preference in strand cleavage [Fig. 11(B), (C)] (48).

Ordered strand exchange is the norm in the λ Int and XerCD systems as well. The constraints imposed by high-order protein assemblies and DNA topology on the synapsis of the Int-bound core sites can dictate which pair of scissile phosphates is primed for initial cleavage (31, 70, 97). In the case of XerCD, depending on the reaction context, cleavage susceptibility may be determined by the synapse topology organized by accessory factors or may be altered by the presence or absence of an interacting regulatory protein (98, 99).
TOPOLOGICAL AND CHIRAL FEATURES OF TYROSINE RECOMBINATION

Tyrosine recombinases (Cre, Flp and λ Int) in general assemble the synapse by random collision of their target sites (see chapter by M. Boocock). In the case of Int, this randomness appears to be superposed over an intrinsic topological specificity (see below). As noted in discussing the role of homology in Flp recombination, it is the interdomainal crossings trapped during synapsis (blue × red in Fig. 5) that appear as knot or catenane crossings in the recombination products. As pointed out earlier, the inversion reaction results in an unknotted circle (with the blue domain inverted with respect to the red) together with a range of increasingly complex knots; the deletion reaction produces unlinked circles as well as a range of increasingly complex catenanes.

The topology of Cre recombination is sensitive to reaction conditions. Relatively high pH tends to increase the complexity of the products, while lower pH has the opposite effect (43). Computer simulations, combined with DNA cyclization assays, suggest that the topological difference between Cre and Flp can be accounted for by the difference in the presynaptic bends that they induce in their target sites (~35° for LoxP and ~78° for FRT) (100). The larger bend tends to localize two presynaptic FRT sites within separate branches of a plectonemically supercoiled circle [Fig. 12(A)], while the smaller bend tends to place two presynaptic LoxP sites in the same branch [Fig. 12(B)]. Interbranch recombination results in topologically complex products; intrabranch recombination gives simple products. Thus, protein-induced local changes in the statistical properties of large DNA molecules can strongly influence their global topology, and dictate the outcomes of the reactions they partake in.

The topological outcomes of XerCD recombination are dictated by the contexts in which the reaction occurs. Recombination between dif sites (utilized for the resolution of E. coli chromosome dimers) requires the ATP driven DNA translocase FtsK (99). The reaction yields topologically simple products from negatively supercoiled substrates. When dif-recombination is activated by the carboxyl-terminal γ-domain of FtsK, which lacks the ATPase function, the products are topologically complex. The topology of recombination between two cer sites or two psi sites (utilized for the resolution of plasmid dimers) is dictated by accessory protein factors (101). The reaction normally occurs between sites in head-to-tail orientation, and requires negative supercoiling. The unique right-handed 4-crossing catenane produced from psi × psi recombination, as well as the structure of the Holliday junction formed by cer × cer

FIGURE 10 Preferred antiparallel synapsis of a pair of tethered FRT sites. (A) The two FRT sites, whose orientation is indicated by the arrowheads, are constrained by a single-stranded tether (wavy line) to align only in the antiparallel geometry. The positions of the donor (Cy3) and acceptor (Cy5) fluorophores are indicated by the green and red circles, respectively. The shift towards lower FRET upon Flp(Y343F) binding is consistent with the synapsis of the FRT sites as schematically diagrammed (44). (B) In the tethered DNA substrate, analogous to that diagrammed in (A), the FRT sites are constrained to pair only in the parallel geometry. Flp(Y343F) binding produces no change in FRET, suggesting the absence of parallel synapsis. doi:10.1128/microbiolspec.MDNA3-0021-2014.f10
strand exchange, conforms to a three-crossing synapse topology. According to tangle analysis, the recombination synapse fits a unique three-dimensional model, with three solutions that correspond to three distinct views obtained by rigid body movements of the synapse and projection on to a planar surface (102).

The FtsK-dependent topology simplification by XerCD recombination is also manifested in the unlinking of catenanes harboring \textit{dif} sites, either in the parallel or anti-parallel sense (103). Catenanes with parallel \textit{dif} sites are topologically analogous to catenanes resulting from the replication of circular plasmids and chromosomes, which are unlinked by the type II topoisomerase Topo IV (104). Rather surprisingly, XerCD-FtsK can support the resolution of chromosome catenanes \textit{in vivo} when Topo IV activity is compromised (105). The recombination mechanism would suggest that unlinking by XerCD-FtsK proceeds by removing one crossing at a time (Fig. 13). This intuitive model, based on product distributions observed in \textit{in vitro} reactions with plasmid substrates, has been validated mathematically by a combination of tangle analysis and knot theory under the assumption that each recombination event reduces the topological complexity of the substrate (106, 107).

**FIGURE 11** The preferred assembly of one of two possible types of antiparallel synapse can specify the order in which strands are cleaved and exchanged during recombination. (A) A LoxP site bound by Cre is bent asymmetrically, the bend center being located close to one end of the strand exchange region. The two possible asymmetric bends would specify the cleavage of the bottom (blue) or the top strand (red). The scissile phosphates primed for cleavage are indicated by the filled circles; the quiescent ones are shown as open circles. For convenience of orienting the sites, the DNA arms are labeled as L (left) and R (right) as in Fig. 1. (B) Based on the structure of the Cre-LoxP complex, fluorophores can be so positioned as to minimize donor (green)–acceptor (red) distance, and induce efficient FRET when the synapse favoring bottom strand cleavage (shown at the left) is assembled by Cre. In this fluorophore configuration, the FRET efficiency will be low for the synapse favoring top strand cleavage (shown at the right). (C) By reversing the left–right orientation of the fluorophores with respect to the strand exchange region, while maintaining their relative positioning, the synapsis favoring top strand cleavage (right) can be made to acquire the high FRET state. Experimental results indicate a clear preference for the synapse shown at the left in (B) suggesting that recombination is initiated by bottom strand cleavage and exchange (48). doi:10.1128/microbiolspec.MDNA3-0021-2014.f11
An intriguing aspect of tyrosine recombination, brought to light primarily from the analysis of λ Int reactions is the apparent chirality of the reaction (108). The chirality of knots and catenanes formed from inversion and deletion reactions, respectively, in negatively supercoiled substrates follows from the right-handed chirality of plectonemic negative supercoils. However, quite unexpectedly, even reactions of nicked substrates turn out to be chiral. In the reactions between attP and attB sites in nicked substrates, two inter-domainal right-handed crossings are trapped by the DNA Interactions of Int and its accessory factors; in the reactions between attL and attR sites, the corresponding number is one. It would take only a single additional right handed DNA crossing randomly trapped in the substrate to generate a chiral three-noded knot as the product of an attP-attB inversion reaction. During attP-attB reactions in negatively supercoiled molecules, Int (in conjunction with accessory factors) traps three right handed DNA crossings, one more than that deduced for the same reactions in nicked circular molecules. However, for attL-attR reactions, this number is still one, unchanged between nicked and negatively supercoiled substrates.

The topological and chiral features of a recombination reaction are conveniently and succinctly summarized by tangle diagrams such as those illustrated in Fig. 14. A tangle may be perceived as a three-dimensional ball within which strings representing double stranded DNA may cross in a variety of ways. In the two dimensional projection of a tangle, the entry and exit points of DNA are placed at the NE, NW, SE and SW corners (in a geographical sense). The Ob and Of tangles harbor DNA constrained by protein binding and ‘free’ DNA, respectively. The P and R tangles contain the DNA segments that engage in crossover in their parental and recombined states, respectively. Ob for attL-attR reactions, containing one right-handed crossing, is a +1 tangle for nicked as well as negatively supercoiled substrates (as shown In Fig. 14 A, B). For attP-attB reactions in nicked substrates, the Ob tangle is +2; for the same reactions in negatively supercoiled substrates, the Ob tangle is +3.

Most surprisingly, recombination reactions from nicked substrates by Cre and Flp also appear to be chiral, trapping one right-handed inter-domainal crossing in the synapse. This crossing is proposed to predispose the reaction towards a chiral product via a right-handed Holliday junction intermediate. The near perfect planarity of the DNA arms in the crystal structures of Cre and Flp (34, 52) challenges this postulate. Nevertheless, the slight out-of-plane disposition of the Holliday junction arms in the Flp crystal structure is consistent with the proposed right-handed chirality.

An irksome aspect of chirality is the difficulty in accommodating the experimental observation that the linking number change (ΔLk) associated with Flp- or λ Int-mediated inversion reactions between FRT sites and...
attL-attR sites, respectively, is either +2 or −2 (41, 109), and the two outcomes are equally likely for a nearly perfectly relaxed substrate. The right-handed chirality would predict a ΔLk of exclusively +2. For example, the right-handed crossing (a − node) trapped by Int would change its sign (a + node) as a result of DNA inversion (ΔLk = +1 − (−1) = +2). The tangle diagram depicting this change in the node sign in O₈ is shown in Fig. 14A. The two suggested tangle solutions to resolve this paradox are shown in Fig. 14B, C. In Fig. 14B, the substrate DNA enclosed by the O₈ tangle harbors two + crossings, one to compensate for the − crossing trapped by Int (in the O₈ tangle) and an additional one to arrange the recombination sites with the antiparallel geometry in the P (parental) tangle. The inversion of each of these crossings would give ΔLk = −2, [(+1 − (−1)] + [(−2) − (+2)]. The problem, though, is that the energetic cost of introducing additional O₈ crossings should make ΔLk = −2 less likely than ΔLk = +2, in violation of the experimental result. In Fig. 14C, the P tangle is switched from an ∞ tangle to a 0 tangle, so as to preserve the tangle notation in O₈ and still produce a ΔLk = −2, [−1 − (+1)]. This is also unsatisfactory, as it accommodates ΔLk = −2 by a sleight of hand(edness). If one follows the contour of the DNA circle, it is obvious that the crossing in O₈ is left-handed, not right-handed. Chirality of tyrosine recombination and the ΔLk paradox arising from it remain an enigmatic curiosity that calls for further exploration.

CONTRIBUTIONS OF YRs TO BASIC BIOLOGY AND BIOENGINEERING APPLICATIONS

Biochemical, biophysical, and structural studies of YRs have been seminal to unveiling the mechanisms of an important class of phosphoryl transfer reactions in nucleic acids and to understanding conformational dynamics associated with strand exchange between two DNA partners (20, 21, 23, 90, 91, 92, 93). The simple requirements of Cre and Flp have been exploited to carry out specific genetic rearrangements in bacteria, fungi, plants, and animals. By combining the DNA delivery properties of mobile group II introns in bacteria and the DNA exchange potential of tyrosine recombination, a new platform for genome editing via targetrons and recombinases (GETR) has been developed (110). In general, prokaryotic and eukaryotic cells engineered to express a recombinase and harboring its target sites in the genome or housed in an extrachromosomal vector carry out the expected reaction with high efficiency. Directed insertion

FIGURE 13 Unlinking of replication catenanes by XerCD-FtsK. (A) The unlinking of replication catenanes in E. coli is normally carried out by the type II topoisomerase Topo IV. For a 4-noded replication catenane containing parallel dif sites, unlinking by Topo IV will be completed in two steps (the straight path), removing two crossings at each step. Unlinking of the same catenane by FtsK-XerCD-mediated recombination at the dif sites requires four steps (the zigzag path), by removal of one crossing at a time. (B) The mechanisms for topology simplification by Topo IV and FtsK-XerCD are illustrated. doi:10.1128/microbiolspec.MDNA3-0021-2014.f13
of a desired foreign DNA into a genome as well as inversions, deletions, or translocations of selected genomic segments can thus be accomplished with reasonable ease. These manipulations have been particularly helpful in addressing fundamental problems in cell and developmental biology. The utilization of controlled and efficient site-specific recombination between homologous chromosomes to generate mosaic flies has provided a technical breakthrough for tracking cell lineages in Drosophila (111). Analogous strategies, in conjunction with multicolored reporter genes and live-cell imaging, have expanded the power and range of lineage tracking to higher organisms and facilitated its integration with the monitoring of intracellular signaling pathways (112). Methodologies for tissue-, cell type- and stage-specific induction of recombination activity make it possible to analyze spatial and temporal controls of developmental programs in intricate detail (113, 114). Another, perhaps less widely publicized, utility of Cre and Flp in basic biology is exemplified by “difference topology,”

**FIGURE 14** Tangle diagrams of attL-attR recombination performed by λ Int; ΔLk associated with DNA inversion. The λ Int mediated inversion reaction between attL and attR sites in three relaxed circular substrate molecules is represented by tangle diagrams (A–C). The O_b tangle contains inter-domainal DNA crossings trapped by Int (likely assisted by the accessory factors). O_f contains randomly trapped crossings in the ‘free’ DNA. The core recombination sites reside in the P tangle in anti-parallel geometry. The R tangle represents the post-recombination state of the sites. The tangle notations are shown at the top in bold; the corresponding DNA crossing (node) signs are given at the bottom in parentheses. The convention for the crossing signs (+1 or −1) is illustrated at the right, with the arrow heads denoting the direction (arbitrarily assigned) for the circular DNA axis. The simplest tangles (0, +1, −1, ∞) are diagrammed at the far right. A. In the DNA molecule shown here, one right-handed crossing is trapped in O_b, and none are contained in O_f. In tangle parlance, recombination changes P(∞) tangle to the R(0) tangle, yielding an unknotted inversion circle. Note that a right-handed crossing in O_b in the substrate is +1 by the tangle convention, but −1 by the sign convention. In the recombinant product, the crossing sign in O_b becomes +1 because of DNA inversion. The linking number change (ΔLk) accompanying the attL-attR reaction is +2. B. The same reaction as in B is shown for a molecule with two left-handed crossings present in O_f. The ΔLk for the reaction is −2. C. A molecule performing the same reaction as in A and B is represented with P(∞) and R(0) switched to P(0) and R(∞), respectively. The ΔLk associated with recombination is −2 in this case as well. The Lk changes are explained in more detail in the text. doi:10.1128/Microbiolspec.MDNA3-0021-2014.f14
an analytical method for tracing the topological path of DNA within high-order DNA–protein complexes (115, 116). Finally, Cre, Flp, and to a limited extent, λ Int have been put to practical use in a number of biotechnology-related applications. A brief description of the principles and practice of difference topology and of the potential impact of site-specific recombination on biotechnology is given below.

DIFFERENCE TOPOLOGY: DECIPHERING DNA TOPOLOGY WITHIN DNA–PROTEIN MACHINES

The elimination of topological randomness from Cre and Flp reactions by assembling a unique synapse with the assistance of protein factors (40, 41, 42, 43) is the basis for the analytical tool called difference topology. The method is useful for determining the number of supercoils sequestered by two DNA sites when they functionally interact with each other. As we have seen already, when three negative supercoils are trapped adjacent to Cre or Flp synapse, say by utilizing the resolvase synapse, the inversion and deletion reactions yield a 3-noded knot and a 4-noded catenane, respectively (Fig. 5). The inversion knot faithfully preserves the number of DNA crossings in the external synapse, as three (or an odd number) crossings would bring head-to-head LoxP or FRT sites in the antiparallel geometry that promotes recombination. A fourth crossing, easily provided by the negatively supercoiled substrate, is necessary for the deletion reaction, as it takes an even number of crossings to confer antiparallel geometry on head-to-tail sites. By similar arguments, two external negative supercoil crossings would be revealed in the difference topology analysis as a 2-noded catenane for the deletion reaction and a 3-noded knot for the inversion reaction. Random entrapment of supercoils in the hybrid synapse can be avoided by suitable placement of the recombination sites with respect to the external synapse. The crossings in the inversion knot and the deletion catenane, analyzed by gel electrophoresis and electron microscopy, would thus accurately report on the DNA topology of the external synapse. A simplified description of the concepts and experimental applications of difference topology can be found in a recent review (2). Using this analysis, the topology of the interactions among the left and right ends of phage Mu and its transposition enhancer element within the transposition complex organized by the MuA protein has been mapped as a three-branched, five-crossing plectoneme (117) (see chapter by R. M. Harshey).

ENGINEERING OF EUKARYOTIC GENOMES USING YRs

Before the practical utilities of site-specific recombination came into prominence, genome manipulations in higher eukaryotes relied either on nonhomologous or homologous recombination. Nonhomologous recombination promotes the efficient integration of linear DNA molecules into most genomes, but does so randomly. Homologous recombination permits modification of genomic loci with high specificity, but suffers from very low efficiency. Site-specific recombinases circumvent the drawbacks of both inefficiency and promiscuity, but require the prior integration of their cognate target sites into the genome to be modified. Furthermore, targeting multiple loci within a genome is limited by the number of available recombinases with the desired properties. This problem can at least be partly circumvented by taking advantage of the homology rule that dictates successful recombination. Mutually incompatible, but individually functional, target sites may be designed by introducing nonhomologies within their strand exchange regions. An even better solution, at least in principle, is the directed evolution of recombinases with altered target specificities (118, 119, 120, 121). Among YRs, Cre and Flp have been, by far, the enzymes of choice for applications in biotechnology. Variants of λ Int that are not functionally limited by cofactor requirements (122) have so far been running a rather distant third. The integrase of the Streptomyces phage φC31 and chimeras derived from an activated form of the Tn3 resolvase represent serine recombinases that have shown promise as tools in applied genetics (123, 124). Zinc finger nucleases (ZFNs), transcription activator effector-like nucleases (TALENs) and clustered regulatory interspersed short repeats (CRISPR)-Cas based RNA guided nucleases have complemented and augmented site-specific recombinases in the bioengineer’s arsenal for analyzing and reshaping genomes (125, 126, 127).

The optimal performance of a recombinase and the tight regulation of its activity in a given biological context often require amino acid substitutions in the native protein sequence and/or the introduction of allosteric control regions. For example, the preferred growth temperature of budding yeast (∼30 °C), in which Flp normally functions, differs from that of mammalian cells (37 °C). A thermo-tolerant variant of Flp (Flpe), with higher activity at 37 °C than Flp, was obtained by mutagenesis coupled to selection (128). Flpe harbors four amino acid changes from Flp: P2S, L33S, Y108N, and S294P. Flpe, which still underperforms Cre in mammalian cells, has been further improved by adding a nu-
An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective

clear localization signal and by ‘humanizing’ its codons (129). The new variant Flpo is comparable to Cre in its activity in mammalian cells. Continuous expression of a recombinase from a constitutively active promoter could be counterproductive because of the likelihood of the intended reaction being reversed and potential toxic effects arising from rare off-target recombination events. These impediments can be overcome or minimized by conditional recombinase expression from a regulatable promoter (Tet-on or Tet-off, for example) (130). Small molecule effectors are also useful for controlling recombination activity. The recombination potential of Flp or Cre fused to the ligand binding domain of the steroid hormone receptor is activated only in the presence of natural estrogens or synthetic estrogen receptor antagonists (131, 132, 133, 134).

Attempts to increase the target repertoire of a recombinase by generating variants with nonoverlapping specificities have been moderately successful. The basic strategy involves the screening of a large pool of the mutagenized recombinase gene for those that code for “shifted” or “switched” specificities. Simple bacterial genetic assays or more rapid high-throughput cell sorting screens, based on chromogenic reporter genes, have been effectively employed for identifying recombinases with the desired recombination potential (119, 121). Substrate designs that place the target sites and the recombinase genes in cis so as to link them by the act of recombination (substrate linked protein evolution; SLiPE) can accelerate screening by simple PCR-based protocols (135). In a distinct cell-free approach, in vitro compartmentalization (IVC) has been used to obtain altered specificity variants of λ Int (120). The IVC method relies on compartments of an oil-in-water emulsion in which in vitro expressed Int variants and the target sites are encapsulated.

In the Flp system, the strategy of mutagenesis followed by a bacterial dual reporter screen yielded recombination capability initially towards FRT sites containing single point mutations in the Flp binding element and subsequently towards sites containing combinations of these mutations (121, 136). It was further demonstrated that hybrid FRT sites, harboring distinct specificities in their two binding elements, can be recombined by a binary combination of Flp variants, each with the appropriate monospecific recognition potential (118). A step-wise directed evolution scheme with intermediate DNA shuffling steps is necessary to progressively coax Flp into accepting multiple changes within the FRT site. Consistent with the array of DNA contacts made by Flp, and the rather complex mode of substrate recognition, mutations involved in the acquisition of new specificities are distributed among amino acids that directly contact DNA as well as those that are located at monomer–monomer interfaces or in the proximity of catalytic residues.

Altered specificity variants of Cre have been evolved by structure-based substitution of base pairs recognized by Cre and randomization of selected amino acid positions in close proximity to them (119). Structural analysis of a subset of these variants suggests that two target sequences can be functionally recognized by a Cre variant through similar backbone contacts in conjunction with distinct base-specific contacts (137). These alternative modes of recognition are facilitated by a network of water-mediated contacts and an unexpected shift in the DNA backbone configuration. The contributions of water networks and macromolecular plasticity to DNA–protein interactions may thus complicate efforts to evolve new target specificities based on predictive schemes.

Directed evolution of recombinases that can act on naturally occurring sequences in their native biological context would signify a giant step forward in site-specific genome remodeling. Search algorithms such as TargetFinder and TargetSiteAnalyzer have been developed to identify genomic sequences that match the size of a given recombination target site, and rank them according to the degree of their resemblance in organization and sequence to the chosen site (138, 139). There are >600,000 potential FRT-like sequences in the human genome (roughly one such sequence per 5,000 bp). Their distributions are inversely correlated to the average G/C content of individual chromosomes, in agreement with the A/T richness of the FRT site. The highest density (one FRT-like sequence per ~4 kb) is in chromosomes 4 and 13 with an average G/C content of 38%, and the lowest (one FRT-like sequence per ~8 kb) is in chromosomes 19 and 22 with an average G/C content of 48%. The majority of duplicate FRT-like sequences are located in the copies of LINE1, while others form part of the LTRs (long terminal repeats) of endogenous retroviruses, Alu repeats and other repetitive DNA sequences. The potential genomic target sites located by search algorithms not only facilitate the manipulation of genetic loci of interest but also promote stringent specificity by providing sequences for counterselection during the steps of directed evolution of novel specificity recombinases.

Once an appropriately placed “high-ranking” site has been identified, the procedures of progressive directed evolution of the recombinase, aided by structural information in at least some of the cases, can be employed to turn it into an authentic recombination target. This strategy has produced Flp variants that utilize an FRT-
like sequence located upstream of the human IL-10 gene and analogous sequences found in the human β globin locus (138, 139). Members of the latter set of sites perform well in mammalian cells when they are present on episomal vectors (139). However, further optimization of specificity, recombinase expression, and activity will be required before the system operates efficiently in the native chromosomal context. The step-wise evolutionary approach has proven to be powerful enough to yield a Cre variant capable of deleting a proviral DNA of HIV-1 pseudotype by recombination between LoxP-like sequences located in the LTRs (140).

An important and frequently employed genome engineering reaction is recombinase-mediated cassette exchange (RMCE), a replacement reaction that exchanges two DNA fragments by a double recombination event between sites flanking them at either end (141, 142) (Fig. 15). The two sites harbored by each DNA partner are designed to be incompatible (heterotypic) for intramolecular recombination but compatible (homotypic) for intermolecular recombination in one configuration of the partners. In its early formats, RMCE reactions utilized a single recombinase such as Cre or Flp to perform exchange at both DNA ends (141, 143) [Fig. 15 (A)]. In more recent versions of RMCE, referred to as dual RMCE (144), recombination at each end is mediated by a separate recombinase, Flp and Cre or Flp and the integrase of the λ related HK022 phage [Fig. 15(B)]. To obtain the best replacement results by dual RMCE, the relative expressions of the two recombinases have to be carefully optimized (145, 146). The Cre-Flp pair can yield RMCE in up to 35 to 45% of the transfected cells, while the corresponding yield for the Flp-HK022 Int pair is ∼12%.

The Cre-Flp based dual RMCE system is being successfully employed by several mouse genome engineering programs for systematically knocking out protein coding regions, expressing reporter cassettes from cellular promoters, and for amino-terminal protein tagging of gene trap clones in situ. The organizations leading these efforts are the International Mouse Phenotyping Consortium (IMPC; www.mousephenotype.org), the European Conditional Mouse Mutagenesis Program (EUCOMM; www.eucomm.org), the KnockOut Mouse Project (KOMP; www.knockoutmouse.org) and the German Gene Trap Consortium (www.genetrap.de). The beneficiaries from these endeavors will be high-throughput genomics and proteomics related to molecular medicine.

**EPILOGUE**

The intellectual seed for the advances in our understanding of site-specific recombination was sown more than fifty years ago by a simple and elegant model proposed by Allan Campbell for the integration of the phage λ genome into the *E. coli* chromosome. Over these five decades, the study of recombination has been transformed from a geneticists’ sanctuary to the playing fields.

**FIGURE 15** Recombination-mediated cassette exchange. (A) In the classical RMCE, the replacement of a native locus by a donor DNA fragment is mediated by the same recombinase acting on two pairs of target sites (RTa and RTa*) that are compatible only in one configuration of the DNA partners. The reaction may be followed by the replacement of one fluorescent reporter (RFP; red fluorescent protein) by another (GFP; green fluorescent protein). (B) In dual RMCE, the same reaction as in (A) is mediated by two separate recombinases acting on their respective cognate sites (RTa and RTb). doi:10.1128/microbiolspec.MDNA3-0021-2014.f15
of biochemists and to the roaming grounds of crystallographers and biophysicists. Their collective contributions have unveiled the chemical simplicity, mechanistic elegance, and structural sophistication of the reaction. Genome engineers, biotechnologists, and system biologists have now almost completely taken over the field and seem poised to lead it in new directions.

ACKNOWLEDGMENTS
I thank Phoebe Rice, Ian Grainge, F.-X. Barre and Jeff Gardner for helpful criticisms and suggestions. The work in the Jayaram laboratory on the mechanism of site-specific recombination has been supported by the National Institutes of Health, the National Science Foundation, the Human Frontiers in Science Program, and the Texas Higher Education Coordinating Board. Funding for our recent work on the DNA topology and single molecule analysis of recombination has been provided by the National Science Foundation (MCR-1049925) and the Robert F Welch Foundation (F-1274).

REFERENCES


An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective


