ROLE OF THE CORE IN CTNDOT INTEGRATION

BY

JENNIFER LAUREN LAPRISE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Microbiology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Commitee:

Professor Jeffrey F. Gardner, Chair
Professor James M. Slauch
Associate Professor Andrei Kuzminov
Associate Professor Richard I. Tapping

Abstract

IntDOT is the tyrosine recombinase encoded by the *Bacteroides* conjugative transposon CTnDOT. It catalyzes integration and excision into and out of the bacterial host chromosome. Although it is a member of the well studied tyrosine recombinase family, IntDOT is of particular interest because it can catalyze recombination between substrates containing heterology in the overlap region. This tolerance for heterology suggests that IntDOT uses a different mechanism for recombination.

IntDOT performs initial cleavage on the top strand adjacent to the D and B core sites. The first two bases within the overlap region are a conserved GC dinucleotide that provide the only sequence identity between the two substrates. I used complementary DNA oligonucleotides to invert the overlap region 180 degrees to relocate the GC dinucleotide to the bottom strand adjacent to the B' core site of *attB*. I tested the inverted overlap *attB* site in an *in vitro* integration assay with wild-type *attDOT* and showed that integration still occurs, albeit in the opposite orientation relative to the wild type reaction. I used nicked *attB* substrates to show that initial cleavage of the inverted overlap *attB* substrate takes place on the bottom strand adjacent to the B' core site, suggesting that the location of sequence identity within the overlap determines the site of initial cleavage. I confirmed this hypothesis by using an *attB* site containing a symmetric overlap sequence.

Acknowledgements

There are many people who made this thesis possible, and I am eternally grateful for their help.

I would like to thank Dr. Jeff Gardner for being an outstanding advisor. He has been the biggest contributor to my scientific achievements and I am grateful for his patience, care and understanding.

I would like to thank Dr. Abigail Salyers for her encouragement and efforts to make me a stronger person.

I would like to thank my committee members, Dr. Jim Slauch, Dr. Andrei Kuzminov, and Dr. Richard Tapping for their guidance and advice.

I am thankful to the members of the Gardner and Salyers lab for their friendship and entertainment over the years. I am fortunate to be constantly surrounded by such an amazing group of people on a daily basis.

I especially want to thank my family and close friends. I am blessed to have the most incredible mother any person could hope to have, as without her love and support I would never have been able to accomplish the things I have. I am also grateful for the love, tolerance, and support of Steve, who has always been by my side. Last, but definitely not least, I must thank Sofie for her tireless dedication, companionship and love.

To my best buddy Lyle, this is for you.

Table of Contents

CHAPTER 1: GENERAL INTRODUCTION..1

CHAPTER 2: MECHANISM OF ORDERED STRAND EXCHANGE BY INTDOT...41

CHAPTER 3: INTDOT INTERACTIONS WITH THE CORE-TYPE SITES DURING INTEGRATION...81

CHAPTER 4: ATTEMPTS TO TRAP HOLLIDAY JUNCTION INTERMEDIATES...109

CHAPTER 5: SUMMARY AND FUTURE DIRECTIONS..................................120

Chapter 1

General Introduction

Clinical Significance of *Bacteroides* spp.

Bacteroides spp. are Gram negative obligate anaerobes that comprise over 40% of the human gastrointestinal microbiota (29). Some species are known to be opportunistic pathogens and can cause abscesses and septicemia if they are released from the gut. These infections are becoming increasingly difficult to treat due to the prevalence of antibiotic resistance genes carried by *Bacteroides.* While naturally resistant to aminoglycosides, certain species are also resistant to β-lactams, tetracycline, and MLS (macrolide-lincosamide-streptogramin) drugs (33, 67, 75, 76, 80, 84, 93). Recent studies have shown increasing resistance to a variety of antibiotics such as clindamycin and cefoxitin (33). An older study demonstrated that resistance to tetracycline increased from 30% to 80% between 1970 and 2000 (80). Growing evidence suggests that *Bacteroides* spp. can act as a reservoir for antibiotic resistance genes found on transmissible elements such as conjugative transposons. These genes are transferred via conjugation to other pathogens passing through the gut.

Conjugative Transposons

Conjugative transposons (CTns) are also called integrative conjugative elements (ICEs). They are self-transmissible elements that combine features of plasmids, transposons and bacteriophages (75). Normally, CTns are found integrated into the host chromosome, but under appropriate conditions they will excise to form a closed circular intermediate (Figure 1.1). This intermediate can either re-integrate into a different location in the host genome, or can be nicked at the *oriT,* and a single strand transferred to a recipient cell via conjugation. The complementary strand is synthesized in the new host, and the double stranded element can integrate into the recipient chromosome (10, 24, 25). In addition to self-mobilization, CTns can also mobilize coresident plasmids and mobilizable transposons (MTns) by providing the necessary cellular machinery for transfer via conjugation or by inserting into a plasmid and transferring as a cointegrate (10, 74, 79)

The elements mobilized by CTns can also carry antibiotic resistance genes. For example, NBU2 (*B. fragilis*) carries lincomycin resistance (92), Tn*4555* (*B. vulgatus*) carries cefoxitin resistance (82); and the plasmid pIP419 (*B. thetaiotaomicron*) carries 5-nitroimidazole resistance (68). The CTns and the elements they mobilize are stably maintained in the absence of antibiotic selection and are not exclusive of each other, meaning that a strain can acquire multiple resistances. Even more striking is the broad host range of *Bacteroides*

elements which is only limited by maintenance of the elements and expression of transfer proteins in the recipient (73, 75, 96).

The *Bacteroides* Conjugative Transposon CTnDOT

CTnDOT is the best studied CTn in *Bacteroides.* It carries the *tetQ* gene that provides resistance to tetracycline via a ribosome protection mechanism. A very high sequence conservation of the *tetQ* gene across different species indicates that horizontal gene transfer is responsible for the proliferation of tetracycline resistance. In addition to the *tetQ* gene, CTnDOT also carries an erythromycin resistance gene called *ermF* (Figure 1.2). *ermF* is part of the MLS family of resistance genes that also confer resistance to clindamycin (66). CTnDOT encodes genes for mobilization, transfer, excision and integration. It also contains a central regulatory operon that includes a two-component regulatory system (Figure 1.2). In addition to regulating CTnDOT excision and transfer, the central regulatory proteins can also mobilize unlinked mobilizeable elements such as NBU1 (81, 86, 95).

Regulation of CTnDOT excision and transfer is tetracycline-dependent and is controlled by a translational attenuation mechanism (21, 22, 57). In the presence of tetracycline, the regulatory operon is induced 100 to 1000 fold to produce the RteA and RteB proteins which have high similarity to members of two component regulatory systems. RteA is the sensor and RteB is the response regulator (86). RteB promotes expression of the major regulatory gene, *rteC*,

which is essential for control of both excision and transfer of CTnDOT (Figure 1.3).

CTnDOT encodes its own integrase, a tyrosine recombinase called IntDOT, which is constitutively expressed and directly catalyzes the integration and excision reactions. Although it is a member of the tyrosine recombinase family, IntDOT appears to have a different mechanism for recombination than that proposed for lambda Int and other tyrosine recombinase family members (21, 22, 31, 51). IntDOT contains three important domains: an N-terminal arm binding domain (N), a core binding domain (CB) and a catalytic domain (CAT). The CB domain binds to a pair of inverted repeats called core-type sites that directly flank the region of strand exchange. The CAT domain introduces sequential 7 bp staggered cuts within the region of strand exchange, or overlap regions of the DNA substrates. During integration, initial cleavage takes place on the top strand and the first strand exchange is homology-dependent. The second strand exchange is homology independent (52). This homology-independent step is what differentiates IntDOT from the other tyrosine recombinases.

During excision, IntDOT cleaves the bacterial host DNA adjacent to the ends of the integrated element. This newly excised element forms a closed circular intermediate containing a 5 bp region of heterology within the overlap region that is referred to as a coupling sequence (Figure 1.4). During conjugation, a single strand is transferred into a recipient cell. The heterology is resolved via replication in both the donor and the recipient cells.

The joined ends of the closed circular form of the excised CTnDOT are referred to as *attDOT.* The core-type sites of *attDOT* are called D and D' (Figure 1.5). Integration occurs into one of several known *attB* sites on the *Bacteroides* chromosome. The *attB* site is a simple site that consists only of two inverted repeat core-type sequences called B and B' that directly flank the overlap region. The D core site of *attDOT* contains a conserved sequence shared by the B site in all known *attBs*. We believe this sequence is important for IntDOT recognition and binding.

Tyrosine Recombinases

There are two types of site-specific recombinases - tyrosine and serine recombinases. Tyrosine recombinases use the amino acid tyrosine as the catalytic nucleophile to cleave DNA by forming a 3' phosphotyrosyl intermediate. They catalyze sequential cleavage, strand exchange and ligation steps that form a Holliday junction intermediate before repeating the process on the second pair of DNA strands (8, 91). In contrast, serine recombinases cleave and exchange all four DNA strands simultaneously in a process that includes formation of a 5' phosphoserine linkage. Serine recombinases do not form Holliday junction intermediates (83).

Tyrosine recombinases mediate a variety of important biological reactions such as resolution of multimeric plasmids or chromosomes to ensure proper partitioning during cell division, integration and excision of viral genomes, and

regulation of gene expression by inversion of DNA segments (For review see (8)). The best studied tyrosine recombinases are lambda Int, Cre recombinase from bacteriophage P1, Flp recombinase from *Saccharomyces cerevisae,* and the bacterial XerC/D.

Simple and factor-assisted recombination systems

There are two main subclasses of tyrosine recombinases - simple and factor-assisted. Cre and Flp are examples of simple recombinases. They contain two domains: a core binding (CB) domain, and a catalytic (CAT) domain. In these systems, the enzyme binds to a pair of inverted repeats called core-type sites, which surround the region of strand exchange or crossover region. Cleavage occurs at the border between the core-type site and the 5' end crossover region. Cre and Flp require only the core-type binding sites within their DNA substrates, *loxP* and FRT.

Additional protein factors and DNA sequences are required for the more complex factor-assisted recombinase systems. Lambda Int is the best studied example of a factor-assisted recombinase. In addition to the CB and CAT domains, Int also has a large N-terminal arm binding domain (N) that binds to DNA sites distal to the core-type sites, called arm-type sites. The distal arm regions of the lambda *attP* site also contain binding sites for IHF, Fis, and Xis. The binding of the recombinase and the additional factors to the DNA forms a higher order nucleoprotein structure called an intasome. Depending on which

arm sites are bound and which factors are present, the intasome can be either in the integrative or excisive conformation.

Lambda integration involves formation of the integrative intasome on the phage particle *attP*. Lambda Int binds to both core-type DNA sites (C and C') and to arm-type DNA sites. The *E. coli* architectural protein IHF bends the *attP* DNA to allow interactions between both arm- and core-type sites by Int (55). The integrative intasome synapses with a naked bacterial attachment site *attB* in the host chromosome (69). The *attB* site consists only of core-type DNA sites and the region of strand exchange, or overlap region.

Tyrosine recombinase reaction mechanism

Tyrosine recombinases utilize topoisomerase I chemistry to cleave and ligate strands of DNA. They recognize two partner substrates that contain a pair of 9-13 bp inverted repeats directly flanking a central 6-8 bp overlap region. The recombinase binds to the inverted repeats and makes staggered cuts at each end of the overlap region. Recombination requires complete homology within the overlap region and two pairs of recombinase monomers. The first pair of monomers cleaves the DNA strands to form a covalent 3'-phosphotyrosyl linkage between the catalytic tyrosine and the phosphate backbone of the DNA (Figure 1.6). The free hydroxyl groups generated by this reaction attack the phosphotyrosyl bond on the partner strand, generating a Holliday junction (HJ) intermediate. It has been suggested that the homology requirement in the

overlap region is to allow proper Watson-Crick base pairing between the exchanged strands in order for ligation to occur (15, 46, 59, 62). Isomerization of the Holliday intermediate induces activation of the second pair of monomers while inactivating the first pair. These monomers resolve the HJ by cleavage, exchange and ligation of the second pair of DNA strands (30).

Structure of the recombinases

All tyrosine recombinases share a common catalytic domain fold containing the signature RKHRHY residues (4, 7, 32, 60) (Figure 1.7). The tyrosine is the catalytic nucleophile for which the family is named. The two arginines coordinate the non-bridging oxygens of the phosphodiester bond and stabilize the pentavalent intermediate. These three residues are required for catalysis. The two histidine residues are not completely conserved. In some recombinases, the first His is replaced by Arg, Lys or Tyr, while the second His is often replaced by Trp. The histidines are predicted to act as a general base and a general acid to remove the proton from the catalytic tyrosine and protonate the leaving group during the next step of the reaction. The His is substituted with Lys or Arg in some recombinases (87, 91). Finally, the lysine makes contact with bases adjacent to the scissile phosphate to coordinate the reaction.

Tyrosine recombinases share similar structures in the core binding and catalytic domains despite having very little primary sequence similarity (20, 32, 35-37, 60). The CB domain is located adjacent to the C-terminal CAT domain (91), and both domains work together to form a C-shaped clamp with the CAT

domain interacting with the minor groove on one face of the DNA ,while the CB domain interacting with the major groove on the opposite face. These two domains comprise the functional units of simple recombinases like Cre and Flp (Figure 1.8). Factor assisted recombinases like lambda Int and IntDOT contain an additional N-terminal arm binding domain. The arm binding domain binds to arm-type DNA sites distal to the region of strand exchange and is ultimately responsible for the directionality of the reaction (16, 42, 88).

Role of homology within the overlap region

Tyrosine recombinases have a strong requirement for homology, or sequence identity, between the overlap regions of DNA substrates. When a mutation was made in the overlap of one of the lambda *att* sites, recombination levels become severely depressed. However when the same mutation was made on the partner DNA site, recombination levels were restored, suggesting that homology, rather than sequence, is important in the overlap region (9). Similar results were found in Cre and Flp recombinases (40, 78).

Early models predicted that the homology was required for branch migration of the Holliday junction intermediate, and that mismatches would introduce energetic barriers that could not be overcome. The branch migration model proposed that the first strand exchange occurs at one end of the overlap, and that the HJ intermediate "branch migrates" toward the opposite end of the overlap, where it is resolved by the second pair of strand exchanges (94). To study this further, synthetic HJs were made by blocking the second strand

exchange via an *att* site with a single stranded nick (61) or with a bridging phosphorothioate linkage (44) at the bottom strand cleavage site. More recently, synthetic peptides were developed that can trap HJ intermediates by intercalating into the central, protein free region and inhibiting HJ resolution (14, 18, 19).

Observations that Int and Flp did not require branch migration to resolve synthetic HJs raised doubts about the branch migration model. Synthetic HJs that limited branch mobility within the center of the overlap region could be efficiently resolved by Int (59). Resolution of junctions with a fixed branch point was best when the crossover point was fixed at 2-3 bp within the overlap, whereas resolution was poor at the ends of the overlap. These experiments suggest that the branch point is at the center of the overlap and not at the ends as previously thought. This idea led to a new model, the strand swapping isomerization model proposed by Nunes-Duby et al. The strand swapping isomerization model involves two sequential, symmetrical swaps of three bases between each partner strand. The annealing step would test for homology before ligation occurred. Isomerization of the HJ around the central 1 to 2 bp would promote the second strand exchange (59).

Lee and Jayaram also proposed this model for the Flp system (46) by using half-*frt* sites to measure cleavage and ligation, and by doing crosses with half sites and full sites to measure the amount of strand transfer product. They showed that the ligation step requires homology between the first two to three bases adjacent to the site of ligation. Beyond these bases, ligation was not affected by the presence of heterology. The homology appears to promote proper

base pairing to position the free 5’ hydroxyl group for the ligation reaction. Other data showing that Flp requires 2 bp of homology at either end of the overlap region for efficient strand exchange supports this idea (5, 46, 64, 99). Nunes-Duby et al. further showed that homology is sensed prior to ligation and that strand transfer occurs even when ligation is blocked by a 5’ phosphate (62, 65). They also found that complementary base pairing at one site stimulated strand transfer at the other mismatched site.

The idea that homology is required for the ligation step was further supported by earlier experiments showing that synapsis and initial strand exchange were not affected by the presence of heterology within the overlap region (43, 61). Cleavage assays using heteroduplex *attB* sites and *attP-saf* variants that were not homologous to either strand of the heteroduplex *attB* validate the conclusion that homology is not required for the capture or cleavage of *attB* (69).

Tyrosine Recombinase Family Diversity

The tyrosine recombinase family contains over one hundred known members (60) that catalyze recombination events including integration and excision of phage genomes, maintenance of plasmid copy number, and conjugative transposition. Type IB topoisomerases and eukaryotic type I topoisomerases are also included in this family. The chemistry of tyrosine recombinases is defined by two features: the ability to cleave and ligate DNA

without high energy cofactors, such as ATP; and the presence of highly conserved residues in the catalytic domain (8).

The tyrosine recombinase family can be broken down into four groups on the basis of their reaction complexity. The members of the simplest family branch function as monomers whose primary role is the cleavage and ligation of a single DNA strand. This branch includes the vaccinia virus topoisomerase and type IB topoisomerases. In contrast, all other family members function as tetramers - one monomer for each strand of DNA to be cleaved. The second family branch includes recombinases that do not require accessory proteins. These are sometimes referred to as simple recombinases, and include Cre from bacteriophage P1 and Flp from *Saccharomyces cerivisiae*. The third family branch contains recombinases that bind to core-type DNA sites and require accessory proteins for regulation or topological specificity. The best studied member of this group is *E. coli's* XerCD system which is involved in chromosomal segregation. The fourth family branch is comprised of the most complex systems. Recombinases in the fourth branch are heterobivalent, meaning they bind to both core- and arm-type DNA sites, and require accessory proteins. The integrases of temperate bacteriophages such as lambda, HP1, and L5 are examples of this branch. Other members of this group are found in conjugative elements such as Tn916 and CTnDOT. Some of the best studied tyrosine recombinases are described in further detail below.

Lambda Integrase

The integrase of bacteriophage lambda is one of the best studied members of the tyrosine recombinase family. In fact, the family is often called the lambda Int family of recombinases. Int catalyzes integration and excision of the phage genome into and out of the host chromosome (for a review see (8)). It is a 40 kDa heterobivalent protein with three domains that bind to two different types of DNA sequences (70). The N-terminal (N) binds to arm-type DNA sites that are responsible for ordered strand exchange and the directionality of the reaction (42). The central core-binding domain (CB) binds to core-type DNA sites that directly flank the region of strand exchange. Finally, the C-terminal domain (CAT) also binds to core-type sites and is the catalytically active region of the enzyme (56).

Integration occurs at a single *attB* site in the *E. coli* chromosome located between the *gal* and *bio* operons. The *attB* site is comprised of 21 bp, 15 of which are homologous to the core sequence in the phage *attP* site. This common core region is duplicated during phage integration. Int catalyzes integration by making 7 bp staggered cuts with the initial cleavage taking place on the top strand between bases -3 and -2. The Holliday junction formed after the first strand exchange is resolved when Int makes a second round of cleavage on the bottom strands between bases +4 and +5. The 7 bp overlap region is directly flanked by the core-type binding sites, which are a pair of imperfect inverted repeats bound by the CB and CAT domains of Int. For recombination to occur, the overlap region must be identical between the two *att* sites.

The phage *attP* is much larger than the 21 bp *attB* site. It consists of the core region flanked by the P and P' arms that contain binding sites for Int and other proteins that determine the efficiency and directionality of recombination (41) (Figure 1.9). In addition to the two core sites C and C', there are five arm-type sites - P1 and P2 on the P arm, and P'1, P'2, and P'3 on the P' arm (70, 71). Lambda integration requires the *E. coli* host factor IHF (integration host factor) to stimulate Int binding to both core- and arm-type sites by bending the *attP* DNA. IHF binds to three sites on *attP* - H1 and H2 on the P arm, and H' on the P' arm. Together with Int, IHF forms a nucleoprotein complex, called an intasome, on *attP*. The intasome synapses with a naked *attB* site for integration (69). Integration requires IHF bound to all three sites, and Int bound to P1, P'2 and P'3 (34, 58) (Figure 1.9).

The excision reaction requires an additional phage encoded protein called Xis. It is a small, basic protein that has an architectural role in forming the excisive intasome, and also engages in protein-protein interactions with Int (2, 23). Xis binds to two sites on *attR* - X1 and X2. A third monomer can bind to X1.5, located between X1 and X2 that stabilizes the Xis filament on the DNA (1). Cooperative interactions occur between Xis bound to X1/X2 and Int bound to P2. Int, Xis and IHF assemble on *attR* to form an excisive intasome complex with *attL* for excision of the phage (Figure 1.9). Xis is an important factor in determining directionality of recombination by inhibiting integration. Similarly, IHF will bind preferentially to H1 during integration, and IHF bound to H1 inhibits excision.

Fis (factor for inversion stimulation) is another host factor that bends DNA to promote recombination. The Fis binding site (F) overlaps with the X2 site (Figure 1.9). When Xis concentrations are decreased, Fis can stimulate excision by binding to X2 (90). Because of this, Fis and Xis were originally thought to compete for binding to F/X2 but recent studies have shown that they actually bind cooperatively and that Fis enhances binding of three Xis monomers (63).

Cre recombinase

Cre recombinase is encoded by bacteriophage P1 to resolve phage chromosome dimers and to assist in cyclization of the linear genome after infection (69). Cre contains two domains - CB and CAT. It recognizes the *loxP* site, which contains an asymmetric 6 bp overlap region flanked by two 14 bp inverted repeats. Cre binds to the inverted repeats and makes 6 bp staggered cuts on either side of the overlap region (38, 39) (Figure 1.8). Recombination requires only the *loxP* DNA substrates and Cre recombinase, but no accessory proteins (3, 85).

Flp recombinase

Flp recombinase is encoded by the *Saccharomyces cerevisiae* 2 μm plasmid. It is responsible for maintaining control of the high copy number of the plasmid during cell division. The 2 μm plasmid contains two copies of a 599 bp region that contains the minimal *FRT* site, which is recognized by Flp. The minimal *FRT* consists of two 13 bp inverted repeats separated by an 8 bp overlap

region. Like Cre, Flp does not require any accessory proteins or sequences (77). Flp recombinase cleaves in *trans*, meaning that the catalytic tyrosine of one monomer reaches across the overlap region and cleaves adjacent to the partner monomer (47).

XerC/D

The Xer system is conserved in most organisms containing circular chromosomes. Unlike the other recombinases discussed earlier in this chapter, the Xer system requires two tyrosine recombinases - XerC and XerD. These enzymes are the closest relatives in the tyrosine recombinase family, yet share only 37% identity (11). XerC/D ensure proper chromosome segregation at cell division by converting plasmid and chromosome dimers to monomers (13). They recognize the *cer* site of ColE1 plasmids (27), the *psi* site on pSC101 plasmids (26), and the *dif* site at the chromosomal replication terminus (13). These sites contain 11 bp core-type sites as inverted repeats surrounding a 6 bp (*dif* and *psi*) or 8 bp (*cer*) overlap region. Initial cleavage and strand exchange is performed by XerC, while the second round of cleavage and strand exchange is performed by XerD (6, 12, 26). Interestingly, depending on the substrate, XerC/D can function autonomously or require additional factors. For example, only XerC/D are required for recombination at the *dif* site (48), whereas recombination at *cer* and *psi* requires accessory proteins (PepA and ArgR at *cer*, PepA and ArcA at *psi*) to direct formation of a productive synapse. These proteins are transcription

factors that bind to accessory sequences adjacent to the XerC binding site (26, 54).

Tn*916*

Tn*916* from *Enterococcus faecalis* confers tetracycline resistance through *tetM*. It is highly promiscuous and can transfer between many Gram positive and Gram negative bacterial species (25). Integration of Tn*916* only requires the Tn*916* Integrase, which is a heterobivalent protein similar to lambda Int (50). The C-terminal domain binds to core sites at the ends of the transposon and flanking host DNA. It contains the catalytic residues responsible for DNA cleavage. The N-terminal domain binds to direct repeats located 150 bases from the left end and 90 bases from the right end (24). Excision from the chromosome requires the integrase and Xis (72). The host factor HU enhances the excision reaction but is not strictly required (28). Tn*916* excises from the host chromosome when the integrase makes 6 bp staggered cuts at the ends of the element, generating short overhangs that can join to form a circular intermediate (53, 89). A feature of this intermediate that distinguishes it from the intermediates formed by the recombinases discussed above is that the ends of the excised Tn*916* element are separated by a heteroduplex called a coupling sequence (17). This heterology is normally a barrier for other tyrosine recombinases, but Tn*916* Int can catalyze integration between non-homologous overlaps. The mechanism for this is still unknown.

CTnDOT Integrase, IntDOT

IntDOT is the tyrosine recombinase encoded by the *Bacteroides* conjugative transposon CTnDOT. It catalyzes the integration and excision of CTnDOT into and out of the host chromosome. During excision, IntDOT makes 7 bp staggered cuts at each end of CTnDOT, and the two ends form a covalently closed circular intermediate (21, 49, 51). Like the excised form of Tn*916*, the excised CTnDOT contains heterology in 5 out of the 7 bases, giving it a 5 bp coupling sequence (21) (Figure 1.4).

Integration of CTnDOT occurs into one of several known *attB* sites. All of these sites contain a 10 bp sequence that is nearly identical to a 10 bp sequence found at one end of CTnDOT. IntDOT makes 7 bp staggered cuts on *attDOT* and the target *attB* site. The first strand exchange is homology-dependent, and requires two bases of homology at the site of initial cleavage (45, 52). The second strand is homology-independent. Integration requires only IntDOT and a *Bacteroides* host factor recently identified as BHFa (K. Ringwald, unpublished), although *E. coli* IHF can substitute in this reaction.

Excision of CTnDOT is much more complex than integration. In addition to IntDOT and BHFa, the excision reaction requires Xis2C, Xis2D, and Exc which are all encoded on the excision operon of CTnDOT (22, 57). Xis2C and Xis2D are small basic proteins that bind to the ends of the integrated element and act as excisionases that help to bend the DNA to facilitate intasome formation (C. Keeton, et. al., unpublished). Exc has topoisomerase IA activity, but mutation of the catalytic tyrosine has no effect on excision. It appears to enhance excision

efficiency but the mechanism of how it does so is unclear (C. Keeton, et. al., submitted).

Thesis Summary

The first chapter of my thesis is an introduction to *Bacteroides,* conjugative transposons, and tyrosine recombinases. I explain in detail the mechanism of recombination, giving examples of transposons and bacteriophages that use tyrosine recombinases for integration and excision. The lifecycle of the *Bacteroides* conjugative transposon CTnDOT and the proteins involved in integration and excision are both described.

The second chapter is a description of the role of homology in ordered strand exchange during IntDOT mediated recombination. Most tyrosine recombinases require complete sequence homology between the overlap sequences of the substrate DNA. IntDOT is of particular interest because it only requires two bases of homology at the site of initial cleavage. There is a conserved GC dinucleotide at the top strand adjacent to the B and D core sites where the initial cleavage and strand exchange take place. I used complementary DNA oligonucleotides to move the GC dinucleotide to the bottom strand adjacent to the B' core site where the second cleavage and strand exchange occurs. IntDOT was able to catalyze integration by changing the orientation of the integrative intasome relative to the naked *attB* substrate such that the initial cleavage still took place adjacent to the GC dinucleotide. To

confirm that the bottom strand was the site of initial cleavage when the overlap region was inverted, I synthesized nicked DNA substrates using oligonucleotides. The nick was located at either the top or the bottom strand cleavage site. Cleavage and strand exchange can only occur with the intact DNA strand; the nick prevents these events from taking place. The results of the nicked substrate experiment revealed that the order of strand exchange had switched from initial top strand cleavage to initial bottom strand cleavage, concurrent with the location of the GC dinucleotide. These results were surprising because in the closely related lambda system, the order of strand exchange is dictated by the arm-type sites, not the overlap region.

In addition to the change in the initial strand exchange, the orientation of the integrating *attDOT* had also changed. I made an *attB* site containing a symmetric overlap region with the GC dinucleotide located on both ends. After the integration reaction, I performed an SspI restriction digest of the recombinants and separated the fragments via agarose gel electrophoresis. The fragmentation pattern was consistent with integration in both orientations at equal frequency. This showed that not only does synapsis occur in both orientations, but that successful recombination requires homology at the site of initial cleavage.

The third chapter describes a detailed analysis of the core sites involved in the integration reaction. The B and D sites share a conserved sequence that we presume to be important for IntDOT recognition and binding. The sequence is GTANNTTT. In contrast to the B and D sites, the sequences of the B' and D' sites

are poorly conserved. I started making triple mutations along the B site and tested their effect on integration. Not surprisingly, triple mutations containing the conserved bases nearly abolished recombination. Mutations in the D site also severely reduced recombination. Triple mutations in the B' site had a less significant effect, with recombination only reduced by half. Mutations in the D' site were not made, but would presumably be similar to the results seen with the B' site mutations.

Because the mutations in the conserved bases had the greatest effect on recombination, I made single base mutations along the entire core site and tested each mutant in the *in vitro* recombination assay. The single base mutations with the greatest effect on integration were the middle T of TTT (position -5), and the last two bases of the GTA (positions -9 and -10). These three positions were selected for further analysis by substituting all 4 bases at each site. The bases that were the best and least tolerated gave us an indication of which functional groups might be important for IntDOT contact. I used the base analogs 2-aminopurine (2AP) and diaminopurine (DAP) to show that the presence of an amino group in the major groove is important for IntDOT recognition of the core sequence.

The fourth chapter will describe future studies that could further our understanding of how IntDOT is able to tolerate heterology between its substrates. Data from synthetic Holliday junctions shows that the presence of the arm-type DNA sites is important for resolution of heterologous substrates. This suggests that the intasome itself is somehow different than intasomes formed by

other Y-recombinases and this difference is responsible for the ability to tolerate heterology. Currently, we are working on making crystals of IntDOT in complex with DNA, which should give us a better understanding of the integrative intasome structure. The crystallography results should also confirm the base contact positions identified in my biochemical experiments.

Figures

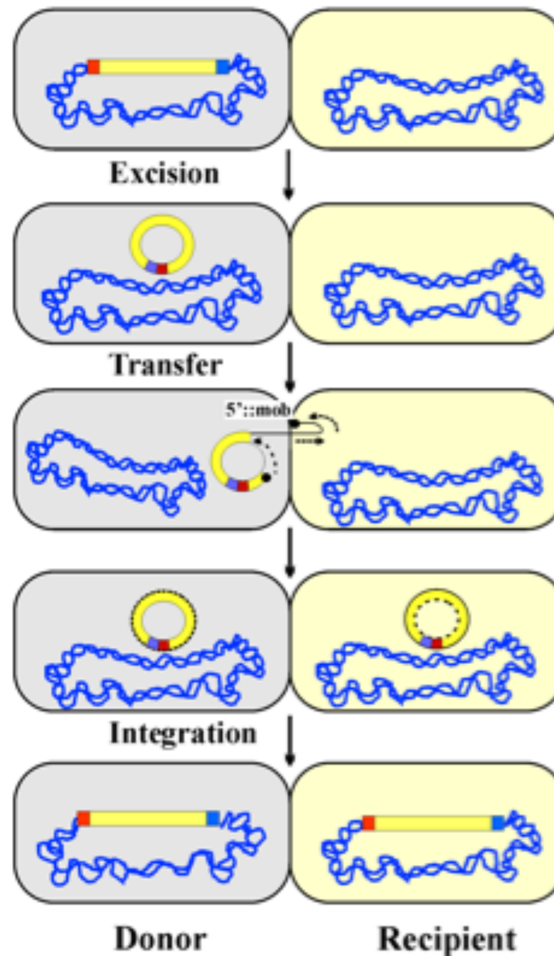


Figure 1.1: Lifecycle of a conjugative transposon. Normally CTns are found integrated in the host chromosome. Under appropriate conditions, the element will excise to form a covalently closed circular intermediate. This intermediate is nicked at the *oriT* and a single strand is transferred to a recipient cell by conjugation. The complementary strands are synthesized in both the donor and the recipient, and the CTn integrates into the chromosome. Figure taken from (96).

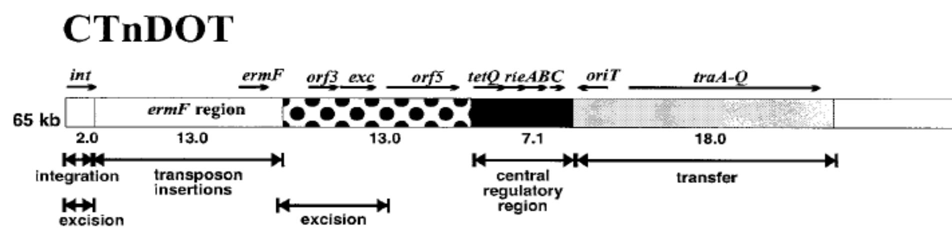


Figure 1.2: Genetic map of CTnDOT. CTnDOT is approximately 65 kb in size. It carries a tetracycline resistance gene *tetQ,* and an erythromycin resistance gene *ermF.* The different fills indicate functional regions, while the arrows above the map indicate promoters and their transcriptional direction. Figure adapted from (86).

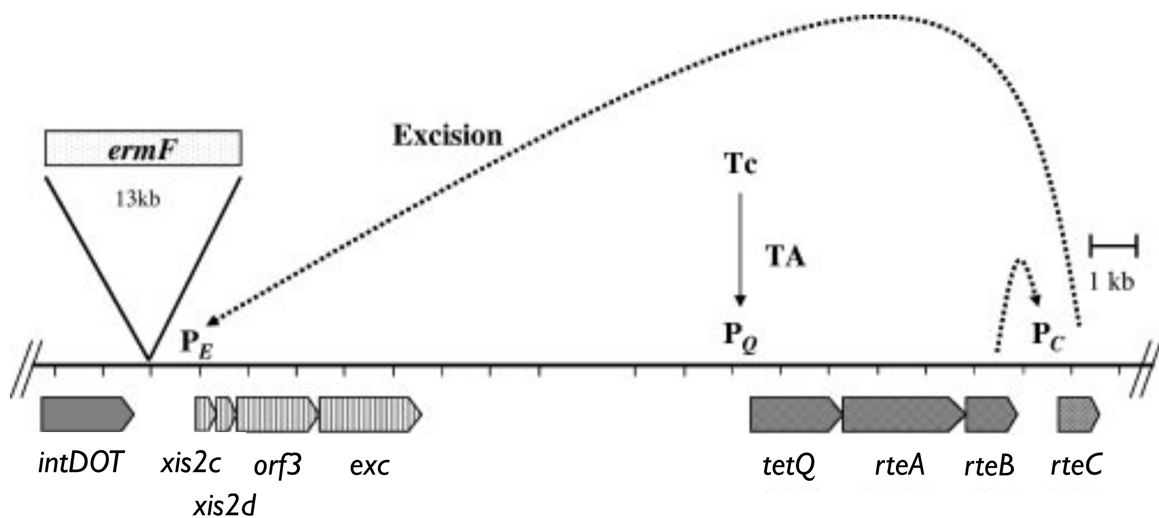


Figure 1.3: Regulation of CTnDOT excision. The presence of tetracycline stimulates excision 100 to 1000 fold. Expression of the *tetQ* operon is controlled by translational attenuation (TA). RteB activates transcription of *rteC*, which is a transcriptional regulator of the excision operon containing *xis2c, xis2d* and *exc.* Figure adapted from (57).

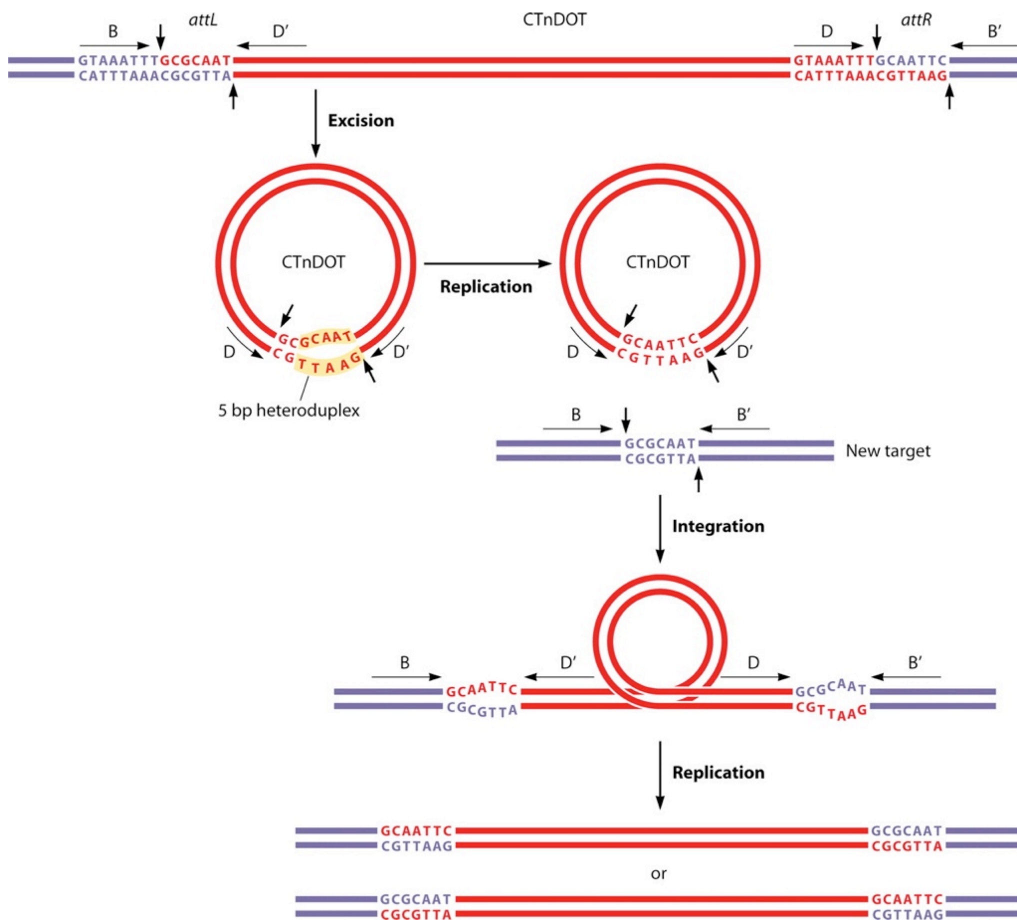
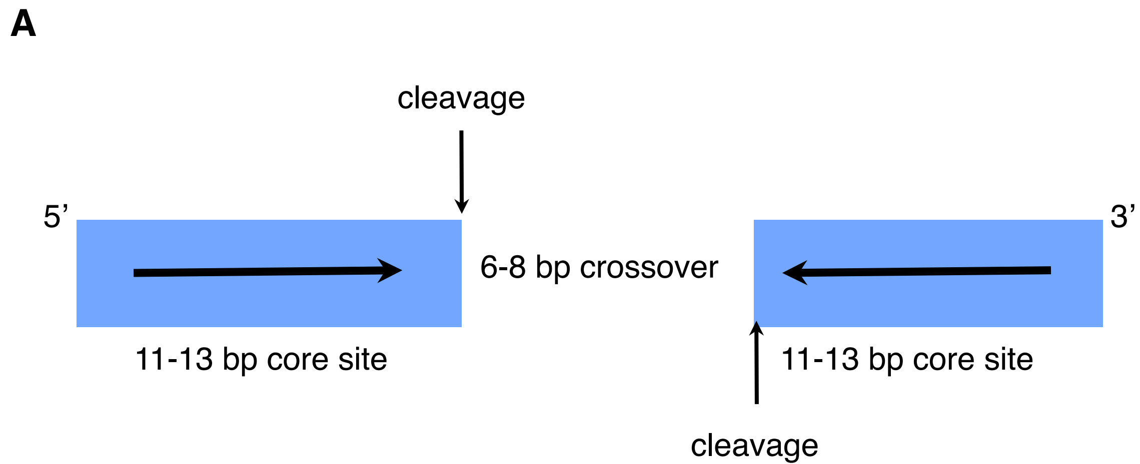


Figure 1.4: Excision and integration of CTnDOT. During excision, IntDOT makes 7 bp staggered cuts at *attL* and *attR* (denoted by the vertical arrows). The element circularizes to form a covalently closed circular intermediate with a 5 bp heteroduplex. The heterology is resolved after conjugation and complementary strand synthesis in both the donor and recipient cell. The double stranded CTnDOT integrates into an *attB* site, introducing temporary mismatches that are resolved by mismatch repair or by replication. Figure taken from (65).

A



B

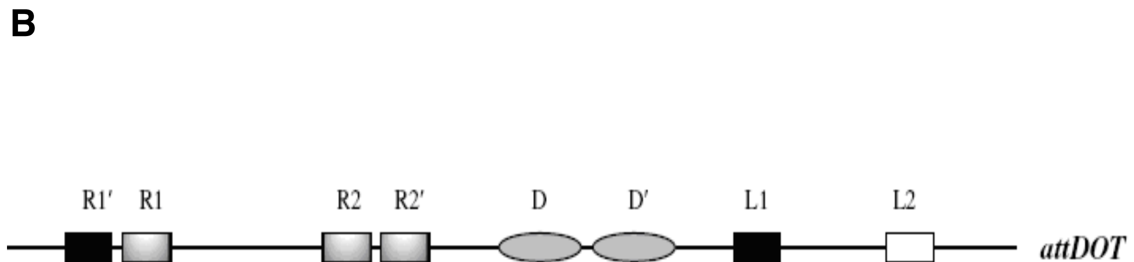


Figure 1.5: Binding sites recognized by tyrosine recombinases. (A) All tyrosine recombinases recognize core site DNA. The site consists of a pair of 11-13 bp inverted repeats that directly flank a central 6-8 bp crossover region. The crossover region is also referred to as the overlap or region of strand exchange. The recombinase binds to the inverted repeats and makes staggered cuts as shown by the vertical arrows. (B) IntDOT recognizes both core and arm-type DNA sites. The core sites D and D' are flanked by arm sites on both sides. The R1' and L1 sites (black boxes) are required for integration. R1, R2, and R2' (grey boxes) are important for intramolecular interactions while L2 (white box) is not required for integration or excision. Figure taken from (97).

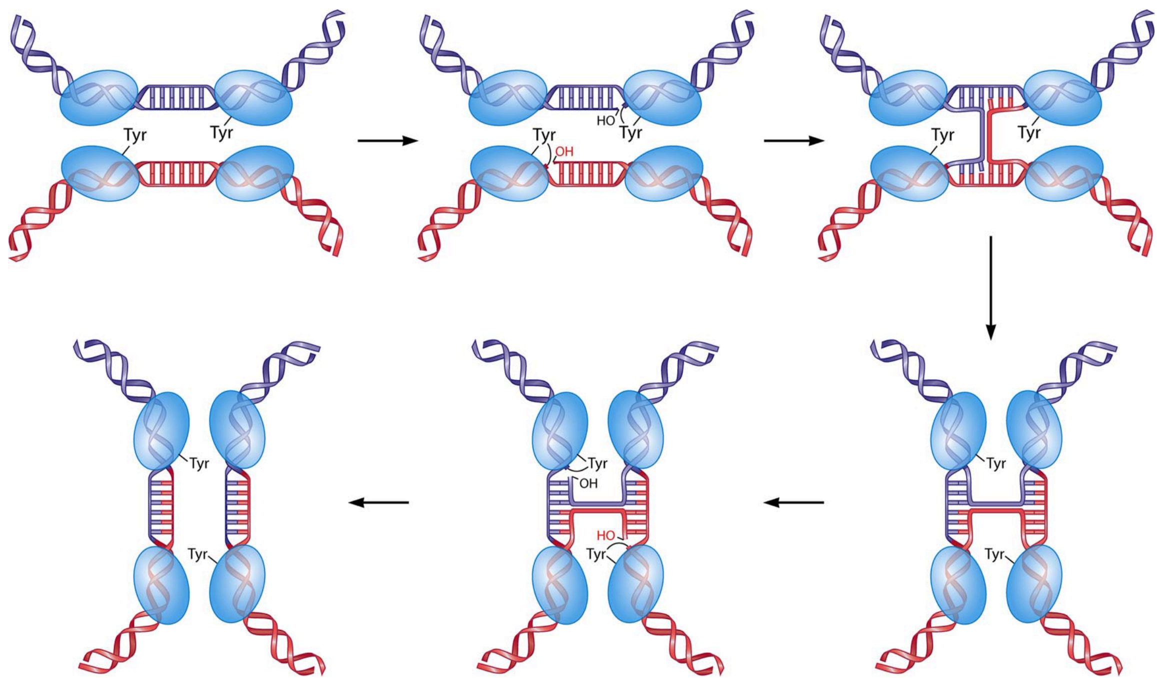


Figure 1.6: Recombination by tyrosine recombinases. The DNA substrates are shown in red and blue. Four recombinase monomers (blue ovals) bind to the DNA. Only two monomers are active at any given time. The active monomers cleave and exchange the first pair of DNA strands to form a Holliday junction intermediate. The isomerization of the HJ results in the inactivation of the first pair of monomers and the activation of the second pair of monomers. The now active pair of monomers carries out the second round of cleavage and strand exchange. Figure taken from (65).

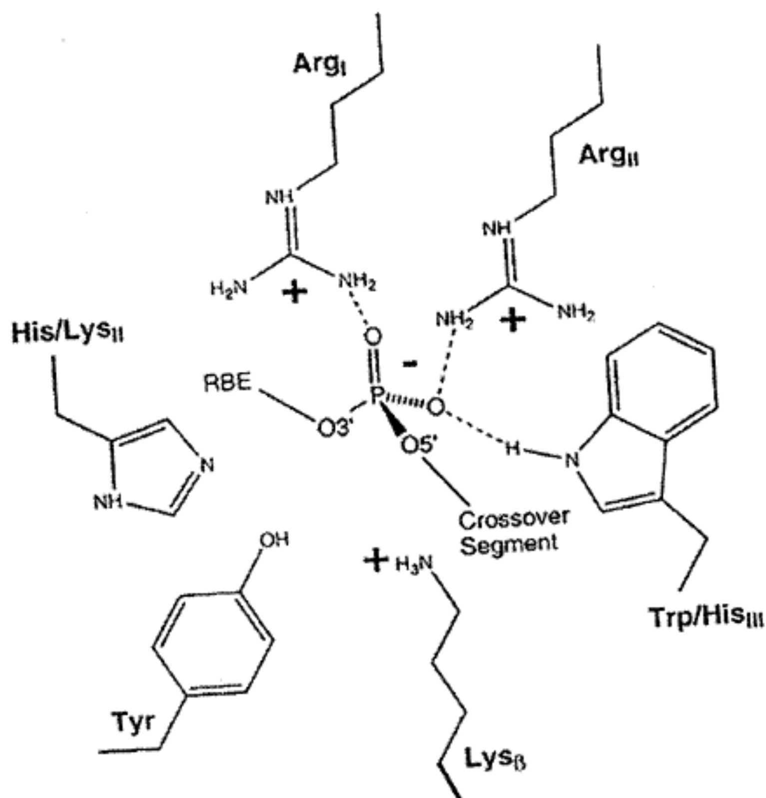


Figure 1.7: Active site of tyrosine recombinases. The scissile phosphate is shown in the center. Six catalytically important residues interact with the phosphate. The two arginines interact with non-bridging oxygens. The hydroxyl group on tyrosine is the catalytic nucleophile. The lysine probably acts as a general acid while the histidines may act as a general base or interact with the scissile phosphate. Figure taken from (91).

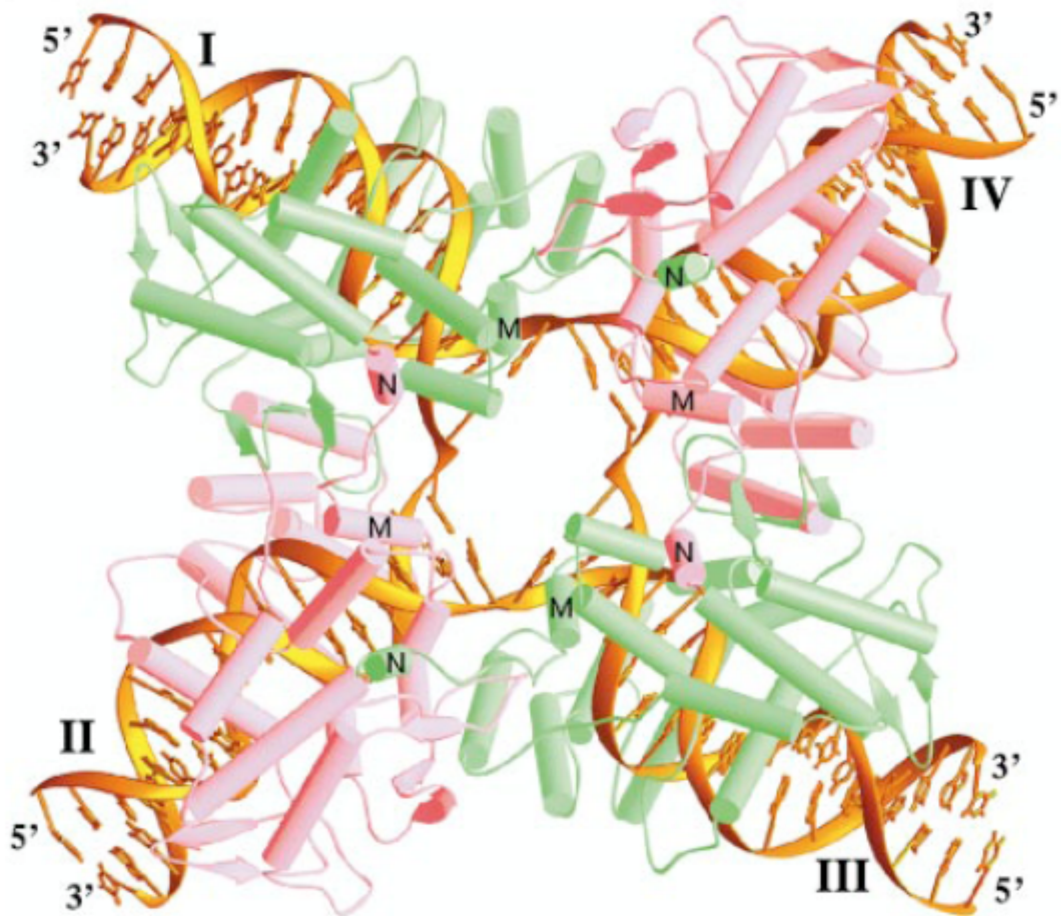


Figure 1.8: Crystal structure of the Cre-Holliday junction complex. The DNA is shown in gold, and the two pairs of monomers are shown in pink and green. The N helix of one monomer interacts with the adjacent monomer. These intermolecular interactions help to stabilize the complex. Figure taken from (35).

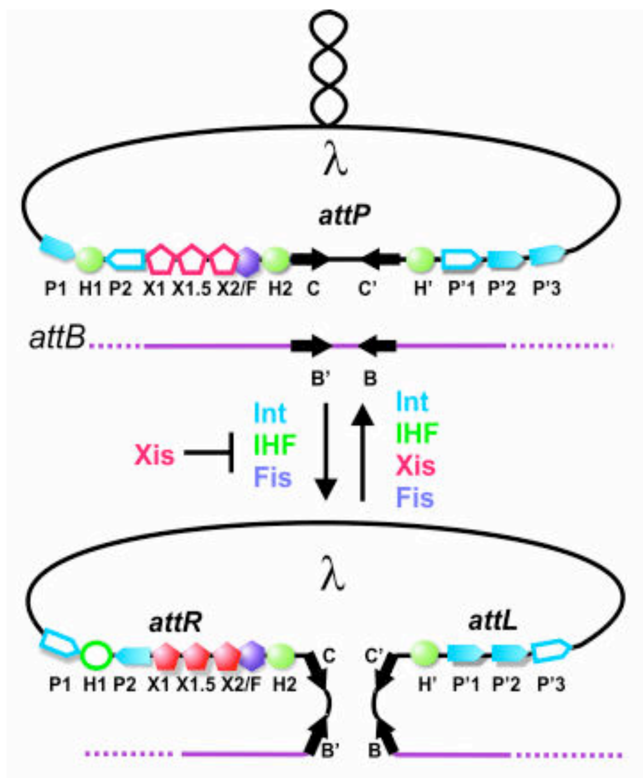


Figure 1.9: Integration and excision of bacteriophage lambda. Integration occurs when the supercoiled *attP* inserts into the *E. coli attB* site to generate *attL* and *attR*. The proteins required for integration are Int, IHF, and Fis. Xis is required for excision and inhibits integration. Filled symbols represent binding sites used. Figure taken from (98).

References

1. **Abbani, M. A., C. V. Papagiannis, M. D. Sam, D. Cascio, R. C. Johnson, and R. T. Clubb.** 2007. Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. Proc Natl Acad Sci U S A **104:**2109-2114.

2. **Abremski, K., and S. Gottesman.** 1982. Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. J Biol Chem **257:**9658-9662.

3. **Abremski, K., and R. Hoess.** 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J Biol Chem **259:**1509-1514.

4. **Abremski, K. E., and R. H. Hoess.** 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. Protein Eng **5:**87-91.

5. **Amin, A., H. Roca, K. Luetke, and P. D. Sadowski.** 1991. Synapsis, strand scission, and strand exchange induced by the FLP recombinase: analysis with half-FRT sites. Mol Cell Biol **11:**4497-4508.

6. **Arciszewska, L. K., and D. J. Sherratt.** 1995. Xer site-specific recombination in vitro. EMBO J **14:**2112-2120.

7. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. Narayana, L. S. Pierson, 3rd, and et al.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J **5:**433-440.

8. **Azaro, M. A., and A. Landy.** 2002. Lambda Integrase and the lambda Int family, p. 119-148. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.

9. **Bauer, C. E., J. F. Gardner, and R. I. Gumport.** 1985. Extent of sequence homology required for bacteriophage lambda site-specific recombination. J Mol Biol **181:**187-197.

10. **Bedzyk, L. A., N. B. Shoemaker, K. E. Young, and A. A. Salyers.** 1992. Insertion and excision of *Bacteroides* conjugative chromosomal elements. J Bacteriol **174:**166-172.

11. **Blakely, G., G. May, R. McCulloch, L. K. Arciszewska, M. Burke, S. T. Lovett, and D. J. Sherratt.** 1993. Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. Cell **75:**351-361.

12. **Blakely, G. W., A. O. Davidson, and D. J. Sherratt.** 1997. Binding and cleavage of nicked substrates by site-specific recombinases XerC and XerD. J Mol Biol **265:**30-39.

13. **Blakely, G. W., and D. J. Sherratt.** 1994. Interactions of the site-specific recombinases XerC and XerD with the recombination site *dif*. Nucleic Acids Res **22:**5613-5620.

14. **Boldt, J. L., C. Pinilla, and A. M. Segall.** 2004. Reversible inhibitors of lambda integrase-mediated recombination efficiently trap Holliday junction intermediates and form the basis of a novel assay for junction resolution. J Biol Chem **279:**3472-3483.

15. **Burgin, A. B., Jr., and H. A. Nash.** 1995. Suicide substrates reveal properties of the homology-dependent steps during integrative recombination of bacteriophage lambda. Curr Biol **5:**1312-1321.

16. **Bushman, W., J. F. Thompson, L. Vargas, and A. Landy.** 1985. Control of directionality in lambda site specific recombination. Science **230:**906-911.

17. **Caparon, M. G., and J. R. Scott.** 1989. Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. Cell **59:**1027-1034.

18. **Cassell, G., M. Klemm, C. Pinilla, and A. Segall.** 2000. Dissection of bacteriophage lambda site-specific recombination using synthetic peptide combinatorial libraries. J Mol Biol **299:**1193-1202.

19. **Cassell, G. D., and A. M. Segall.** 2003. Mechanism of inhibition of site-specific recombination by the Holliday junction-trapping peptide WKHYNY: Insights into phage lambda integrase-mediated strand exchange. J Mol Biol **327:**413-429.

20. **Chen, Y., U. Narendra, L. E. Iype, M. M. Cox, and P. A. Rice.** 2000. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol Cell **6:**885-897.

21. **Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2000. Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. J Bacteriol **182:**4035-4043.

22. **Cheng, Q., Y. Sutanto, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2001. Identification of genes required for excision of CTnDOT, a *Bacteroides* conjugative transposon. Mol Microbiol **41:**625-632.

23. **Cho, E. H., R. I. Gumport, and J. F. Gardner.** 2002. Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. J Bacteriol **184:**5200-5203.

24. **Churchward, G.** 2002. Conjugative Transposons and Related Mobile Elements, p. 177-191. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press.

25. **Clewell, D. B., S. E. Flannagan, and D. D. Jaworski.** 1995. Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons. Trends Microbiol **3:**229-236.

26. **Colloms, S. D., R. McCulloch, K. Grant, L. Neilson, and D. J. Sherratt.** 1996. Xer-mediated site-specific recombination in vitro. EMBO J **15:**1172-1181.

27. **Colloms, S. D., P. Sykora, G. Szatmari, and D. J. Sherratt.** 1990. Recombination at Cole1 *cer* Requires the *Escherichia coli xerC* Gene-Product, a Member of the Lambda-Integrase Family of Site-Specific Recombinases. J Bacteriol **172:**6973-6980.

28. **Connolly, K. M., M. Iwahara, and R. T. Clubb.** 2002. Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. J Bacteriol **184:**2088-2099.

29. **Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight.** 2009. Bacterial community variation in human body habitats across space and time. Science **326:**1694-1697.

30. **Craig, N. L.** 1988. The mechanism of conservative site-specific recombination. Annu Rev Genet **22:**77-105.

31. **Craig, N. L., and H. A. Nash.** 1983. The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell **35:**795-803.

32. **Esposito, D., and J. J. Scocca.** 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. Nucleic Acids Res **25:**3605-3614.

33. **Falagas, M. E., and E. Siakavellas.** 2000. *Bacteroides*, *Prevotella*, and *Porphyromonas* species: a review of antibiotic resistance and therapeutic options. Int J Antimicrob Agents **15:**1-9.

34. **Gardner, J. F., and H. A. Nash.** 1986. Role of *Escherichia coli* IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. J Mol Biol **191:**181-189.

35. **Gopaul, D. N., F. Guo, and G. D. Van Duyne.** 1998. Structure of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. EMBO J **17:**4175-4187.

36. **Guo, F., D. N. Gopaul, and G. D. Van Duyne.** 1999. Asymmetric DNA bending in the Cre-*loxP* site-specific recombination synapse. Proc Natl Acad Sci U S A **96:**7143-7148.

37. **Guo, F., D. N. Gopaul, and G. D. van Duyne.** 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature **389:**40-46.

38. **Hoess, R. H., and K. Abremski.** 1984. Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. Proc Natl Acad Sci U S A **81:**1026-1029.

39. **Hoess, R. H., and K. Abremski.** 1985. Mechanism of strand cleavage and exchange in the Cre-*lox* site-specific recombination system. J Mol Biol **181:**351-362.

40. **Hoess, R. H., A. Wierzbicki, and K. Abremski.** 1986. The role of the *loxP* spacer region in PI site-specific recombination. Nucleic Acids Res **14:**2287-2300.

41. **Hsu, P. L., W. Ross, and A. Landy.** 1980. The lambda phage att site: functional limits and interaction with Int protein. Nature **285:**85-91.

42. **Kitts, P. A., and H. A. Nash.** 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. J Mol Biol **204:**95-107.

43. **Kitts, P. A., and H. A. Nash.** 1987. Homology-dependent interactions in phage lambda site-specific recombination. Nature **329:**346-348.

44. **Kitts, P. A., and H. A. Nash.** 1988. An Intermediate in the Phage-Lambda Site-Specific Recombination Reaction Is Revealed by Phosphorothioate Substitution in DNA. Nucleic Acids Res **16:**6839-6856.

45. **Laprise, J., S. Yoneji, and J. F. Gardner.** 2010. Homology-dependent interactions determine the order of strand exchange by IntDOT recombinase. Nucleic Acids Res **38:**958-969.

46. **Lee, J., and M. Jayaram.** 1995. Role of partner homology in DNA recombination. Complementary base pairing orients the 5'-hydroxyl for strand joining during Flp site-specific recombination. J Biol Chem **270:**4042-4052.

47. **Lee, J., M. Jayaram, and I. Grainge.** 1999. Wild-type Flp recombinase cleaves DNA in trans. EMBO J **18:**784-791.

48. **Leslie, N. R., and D. J. Sherratt.** 1995. Site-specific recombination in the replication terminus region of Escherichia coli: functional replacement of dif. EMBO J **14:**1561-1570.

49. **Li, L. Y., N. B. Shoemaker, and A. A. Salyers.** 1995. Location and characteristics of the transfer region of a Bacteroides conjugative transposon and regulation of transfer genes. J Bacteriol **177:**4992-4999.

50. **Lu, F., and G. Churchward.** 1994. Conjugative transposition: Tn916 integrase contains two independent DNA binding domains that recognize different DNA sequences. EMBO J **13:**1541-1548.

51. **Malanowska, K., A. A. Salyers, and J. F. Gardner.** 2006. Characterization of a conjugative transposon integrase, IntDOT. Mol Microbiol **60:**1228-1240.

52. **Malanowska, K., S. Yoneji, A. A. Salyers, and J. F. Gardner.** 2007. CTnDOT integrase performs ordered homology-dependent and homology-independent strand exchanges. Nucleic Acids Res **35:**5861-5873.

53. **Manganelli, R., S. Ricci, and G. Pozzi.** 1996. Conjugative transposon Tn916: evidence for excision with formation of 5'-protruding termini. J Bacteriol **178:**5813-5816.

54. **Mcculloch, R., L. W. Coggins, S. D. Colloms, and D. J. Sherratt.** 1994. Xer-Mediated Site-Specific Recombination at Cer Generates Holliday Junctions in-Vivo. Embo Journal **13:**1844-1855.

55. **Moitoso de Vargas, L., S. Kim, and A. Landy.** 1989. DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. Science **244:**1457-1461.

56. **Moitoso de Vargas, L., C. A. Pargellis, N. M. Hasan, E. W. Bushman, and A. Landy.** 1988. Autonomous DNA binding domains of lambda integrase recognize two different sequence families. Cell **54:**923-929.

57. **Moon, K., N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2005. Regulation of excision genes of the Bacteroides conjugative transposon CTnDOT. J Bacteriol **187:**5732-5741.

58. **Numrych, T. E., R. I. Gumport, and J. F. Gardner.** 1990. A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage lambda. Nucleic Acids Res **18:**3953-3959.

59. **Nunes-Duby, S. E., M. A. Azaro, and A. Landy.** 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. Curr Biol **5:**139-148.

60. **Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy.** 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res **26:**391-406.

61. **Nunes-Duby, S. E., L. Matsumoto, and A. Landy.** 1987. Site-specific recombination intermediates trapped with suicide substrates. Cell **50:**779-788.

62. **Nunes-Duby, S. E., D. Yu, and A. Landy.** 1997. Sensing homology at the strand-swapping step in lambda excisive recombination. J Mol Biol **272:**493-508.

63. **Papagiannis, C. V., M. D. Sam, M. A. Abbani, D. Yoo, D. Cascio, R. T. Clubb, and R. C. Johnson.** 2007. Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. Journal of molecular biology **367:**328-343.

64. **Qian, X. H., R. B. Inman, and M. M. Cox.** 1992. Reactions between Half-Flp and Full-Flp Recombination Target Sites - a Model System for Analyzing Early Steps in Flp Protein-Mediated Site-Specific Recombination. J Biol Chem **267:**7794-7805.

65. **Rajeev, L., K. Malanowska, and J. F. Gardner.** 2009. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. Microbiol Mol Biol Rev **73:**300-309.

66. **Rasmussen, J. L., D. A. Odelson, and F. L. Macrina.** 1986. Complete nucleotide sequence and transcription of ermF, a macrolide-lincosamide-streptogramin B resistance determinant from Bacteroides fragilis. J Bacteriol **168:**523-533.

67. **Reysset, G., A. Haggoud, and M. Sebald.** 1993. Genetics of resistance of Bacteroides species to 5-nitroimidazole. Clin Infect Dis **16 Suppl 4:**S401-403.

68. **Reysset, G., A. Haggoud, W. J. Su, and M. Sebald.** 1992. Genetic and molecular analysis of pIP417 and pIP419: Bacteroides plasmids encoding 5-nitroimidazole resistance. Plasmid **27:**181-190.

69. **Richet, E., P. Abcarian, and H. A. Nash.** 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell **52:**9-17.

70. **Ross, W., and A. Landy.** 1982. Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. Proc Natl Acad Sci U S A **79:**7724-7728.

71. **Ross, W., A. Landy, Y. Kikuchi, and H. Nash.** 1979. Interaction of int protein with specific sites on lambda att DNA. Cell **18:**297-307.

72. **Rudy, C., K. L. Taylor, D. Hinerfeld, J. R. Scott, and G. Churchward.** 1997. Excision of a conjugative transposon in vitro by the Int and Xis proteins of Tn916. Nucleic Acids Res **25:**4061-4066.

73. **Salyers, A. A., and N. B. Shoemaker.** 1997. Conjugative transposons. Genet Eng (N Y) **19:**89-100.

74. **Salyers, A. A., N. B. Shoemaker, and L. Y. Li.** 1995. In the driver's seat: the Bacteroides conjugative transposons and the elements they mobilize. J. Bacteriol **177:**5727.

75. **Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L. Y. Li.** 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol Rev **59:**579-590.

76. **Sebald, M.** 1994. Genetic basis for antibiotic resistance in anaerobes. Clin Infect Dis **18 Suppl 4:**S297-304.

77. **Senecoff, J. F., R. C. Bruckner, and M. M. Cox.** 1985. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proc Natl Acad Sci U S A **82:**7270-7274.

78. **Senecoff, J. F., and M. M. Cox.** 1986. Directionality in Flp Protein-Promoted Site-Specific Recombination Is Mediated by DNA-DNA Pairing. Journal of Biological Chemistry **261:**7380-7386.

79. **Shoemaker, N. B., and A. A. Salyers.** 1990. A cryptic 65-kilobase-pair transposonlike element isolated from Bacteroides uniformis has homology with Bacteroides conjugal tetracycline resistance elements. J Bacteriol **172:**1694-1702.

80. **Shoemaker, N. B., H. Vlamakis, K. Hayes, and A. A. Salyers.** 2001. Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. Appl Environ Microb **67:**561-568.

81. **Shoemaker, N. B., G. R. Wang, and A. A. Salyers.** 2000. Multiple gene products and sequences required for excision of the mobilizable integrated Bacteroides element NBU1. J Bacteriol **182:**928-936.

82. **Smith, C. J., and A. C. Parker.** 1993. Identification of a circular intermediate in the transfer and transposition of Tn4555, a mobilizable transposon from Bacteroides spp. J Bacteriol **175:**2682-2691.

83. **Smith, M. C., and H. M. Thorpe.** 2002. Diversity in the serine recombinases. Mol Microbiol **44:**299-307.

84. **Speer, B. S., N. B. Shoemaker, and A. A. Salyers.** 1992. Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. Clin Microbiol Rev **5:**387-399.

85. **Sternberg, N., and D. Hamilton.** 1981. Bacteriophage P1 site-specific recombination. I. Recombination between *loxP* sites. J Mol Biol **150:**467-486.

86. **Stevens, A. M., N. B. Shoemaker, L. Y. Li, and A. A. Salyers.** 1993. Tetracycline regulation of genes on Bacteroides conjugative transposons. J Bacteriol **175:**6134-6141.

87. **Subramanya, H. S., L. K. Arciszewska, R. A. Baker, L. E. Bird, D. J. Sherratt, and D. B. Wigley.** 1997. Crystal structure of the site-specific recombinase, XerD. EMBO J **16:**5178-5187.

88. **Swalla, B. M., E. H. Cho, R. I. Gumport, and J. F. Gardner.** 2003. The molecular basis of co-operative DNA binding between lambda integrase and excisionase. Mol Microbiol **50:**89-99.

89. **Taylor, K. L., and G. Churchward.** 1997. Specific DNA cleavage mediated by the integrase of conjugative transposon Tn916. J Bacteriol **179:**1117-1125.

90. **Thompson, J. F., L. Moitoso de Vargas, C. Koch, R. Kahmann, and A. Landy.** 1987. Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. Cell **50:**901-908.

91. **van Duyne, G.** 2002. A structural view of tyrosine recombinase site-specific recombination, p. 2257. *In* N. L. Craig, R. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. American Society of Microbiology press, Washington DC.

92. **Wang, J., N. B. Shoemaker, G. R. Wang, and A. A. Salyers.** 2000. Characterization of a Bacteroides mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. J Bacteriol **182:**3559-3571.

93. **Waters, V. L.** 1999. Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. Front Biosci **4:**D433-456.

94. **Weisberg, R. A., L. W. Enquist, C. Foeller, and A. Landy.** 1983. Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. J Mol Biol **170:**319-342.

95. **Whittle, G., N. B. Shoemaker, and A. A. Salyers.** 2002. Characterization of genes involved in modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT. J Bacteriol **184:**3839-3847.

96. **Whittle, G., N. B. Shoemaker, and A. A. Salyers.** 2002. The role of Bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. Cell Mol Life Sci **59:**2044-2054.

97. **Wood, M. M., J. M. Dichiara, S. Yoneji, and J. F. Gardner.** 2010. CTnDOT integrase interactions with attachment site DNA and control of directionality of the recombination reaction. J Bacteriol **192:**3934-3943.

98. **Yin, S., W. Bushman, and A. Landy.** 1985. Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. Proc Natl Acad Sci U S A **82:**1040-1044.

99. **Zhu, X. D., G. Pan, K. Luetke, and P. D. Sadowski.** 1995. Homology requirements for ligation and strand exchange by the FLP recombinase. J Biol Chem **270:**11646-11653.

Chapter 2

Mechanism of Ordered Strand Exchange by IntDOT

Introduction

Bacteroides spp. are Gram-negative anaerobes that colonize the human gastrointestinal tract. They are host to a number of integrated mobile genetic elements that have been implicated in the transfer of antibiotic resistance genes (24, 41, 42, 47). These elements are called conjugative transposons, or more recently, integrative and conjugative elements (ICEs) (8). CTnDOT is the best studied *Bacteroides* ICE. It carries genes encoding tetracycline resistance, *tetQ*, and erythromycin resistance, *ermF* (6, 39, 40). Excision and integration of CTnDOT are catalyzed by its integrase, IntDOT, which is a member of the large tyrosine recombinase family (15, 25, 31, 34). Tyrosine recombinases use a topoisomerase I-type mechanism for recombination (4, 9, 13), which involves an initial strand exchange to form a Holliday junction intermediate. In most systems this is followed by a homology-dependent strand swapping isomerization step, and then a second strand exchange to form the product (30). In some cases, recombination stops at the Holliday junction step, and the structure is resolved by host processes (7, 12, 28). Sequence identity between overlap regions is a requirement for recombination of most tyrosine recombinases. It is assumed that

Watson-Crick base pairing must take place after the strand exchange in order for ligation to occur (32).

Site-specific recombinases can be sub-classified into two groups depending on whether or not their recombination reaction is regulated. There are autonomous systems in which the recombinase binds only core-type DNA sites. These enzymes contain core-binding (CB) and catalytic (CAT) domains that interact directly with core-type sites. In the more complex systems the recombinase has an additional N-terminal (N) domain that binds to arm-type sites that flank the core-type sites. These systems require accessory proteins (4, 44). There does not appear to be any correlation between the complexity of the system and the regulation of the order of strand exchanges. The order of strand exchange can be influenced by a variety of different factors including additional DNA binding domains, accessory proteins, and DNA structure.

IntDOT contains an N domain and requires accessory proteins for integration and excision, so it was initially assumed that these would influence the order of strand exchange in the integration reaction. In this paper we show that for IntDOT, DNA sequence identity in the overlap region dictates the order and position of strand exchange and that a single base of sequence identity is required for the first strand exchange. Interestingly, unlike most tyrosine recombinases, IntDOT does not require sequence identity for the strand swapping-isomerization step or for the second strand exchange. This unique mechanism for strand exchanges may place IntDOT into a new subclass of tyrosine recombinase.

Materials and Methods

Preparation of Radiolabeled *attB* Substrates

One strand of the *attB* DNA oligonucleotide (IDT) was 5' end-labeled with γ-^{32}P-ATP (Perkin-Elmer) using T4 polynucleotide kinase (Fermentas) and free γ-^{32}P removed using G-25 spin columns (Amersham Biosciences). The labeled DNA was mixed with the unlabeled complementary strand at a 1:5 molar ratio and annealed in an annealing buffer (0.1M KCl, 10mM Tris-HCl pH 8, 5mM EDTA) by heating to 90° C for two minutes followed by slow cooling to 25°C. The *attB* DNA substrates containing a nick in either the top or bottom strand were prepared by mixing a labeled intact strand with two unlabeled strands complementary to the labeled strand at ratios of 1:5:5 respectively. These were annealed as described above. All nicked *attB* substrates were phosphorylated at the 5' end of the oligo at the cleavage site. A list of oligonucleotides used in this study is shown in Table 2.1.

In Vitro Recombination Assay

The *attDOT* and *attB* substrates were incubated in a 20μl reaction volume containing 0.17 μM *E. coli* IHF, 1 unit of IntDOT, 30 mM Tris-HCl (pH 7.4), 5 mM DTT, 0.1 mg/ml tRNA, 0.07 mg/ml BSA, 2.6% glycerol and 50 mM KCl. The final concentrations of *attDOT* and *attB* were 2 nM as determined by O.D.$_{260}$. A unit of IntDOT is defined as the minimum amount of IntDOT needed to produce maximum recombination between *attDOT* and *attB* (29). The reaction was shown

to proceed over 16 hours so samples were incubated overnight at 37^oC and the reaction quenched with the addition of 5 μl stop solution (30% glycerol, 10% SDS, 0.25% xylene cyanol, and 0.25% bromophenol blue). All samples were subjected to electrophoresis on a 1% agarose gel. Gels were dried on a vacuum slab drier then exposed on phosphorimager screens and the recombination efficiency quantified using a Fujifilm FLA-3000 phosphorimager and Fujifilm Image Gauge software (Macintosh v.3.4). Efficiency was calculated by subtracting background photostimulated luminescence (psl) from the recombinant and unreacted *attB* bands, then dividing the recombinant psl by the total psl.

Restriction Digest Analysis of the Recombinant Products

Some experiments were done where products were cleaved by SspI endonuclease. A double volume of the standard recombination assay described above was performed but was terminated by heating for 20 minutes at 60^oC instead of by the addition of stop solution. Magnesium chloride was added to a final concentration of 10 mM. Half the reaction volume was transferred to a fresh microcentrifuge tube and digested with SspI endonuclease (Fermentas). The digest was stopped by heating at 60^oC for 20 minutes and the sample was subjected to electrophoresis on a 1% agarose gel at 100 V for 2 hours and analyzed as described above.

Site-Directed Mutagenesis of *attDOT*/pUC19

Mutations were made in the 7 bp core overlap region of *attDOT* using the Stratagene Quikchange Mutagenesis Kit as described by the supplier (Stratagene). Primers carrying the specific mutations are shown in Table 3. The *attDOT* regions of the plasmid were sequenced to confirm the presence of the desired mutation and to ensure no other mutations were present.

Two-Dimensional Gel Electrophoresis

A double volume recombination assay was performed and terminated with stop solution. Two 15 μl samples were subjected to electrophoresis on a 1% agarose gel in 1X TBE for 2 hours at 100 mA. One lane was sliced out of the gel and prepared for the second dimension; the other lane was dried and exposed on a phosphorimager screen. The gel slice for the second dimension was soaked in alkaline buffer (50 mM NaOH, 5 mM EDTA) for 1 hour before being set in a gel tray with 1% agarose dissolved in alkaline buffer. The gel was electrophoresed in alkaline running buffer (50 mM NaOH, 5 mM EDTA) for 16 hr at 30 V then dried and exposed on a phosphorimager screen.

Detection of Covalent DNA/Protein Complexes

A double volume recombination assay was performed but was stopped with the addition of 0.1 vol of 10% SDS. A standard SDS/KCl precipitation was performed as described previously (43). The fractions were separated on a 1% agarose gel for 2 hours at 100 V.

Results

In this study we used a gel based *in vitro* integration assay developed previously (11). The *in vitro* integration reaction utilizes two DNA substrates carrying a linear radiolabeled *attB* site and a supercoiled plasmid containing the *attDOT* site. The reaction requires IntDOT and *E. coli* integration host factor (IHF). (The *Bacteroides* host factor had not yet been identified at that point, but *E. coli* IHF substitutes (11)). Because IntDOT mediated recombination requires an accessory factor, all experiments were performed with IntDOT and *E. coli* IHF. Recombination between the *attDOT* and *attB* sites produce a 3.6 kb linear recombinant that can be separated from unreacted *attB* DNA on an agarose gel (See Materials and Methods). Several *attB* sites have been identified (10), but most experiments described here use the *attB1* site.

The overlap region defines the sites of cleavage and strand exchange in the top and bottom strands of the *attDOT* and *attB* sites (26). It consists of a 5 base pair coupling sequence that varies in sequence from site to site, and a conserved GC dinucleotide that is found at the left side of the overlap regions in *attDOT* and all known naturally occurring *attB* sites (10) (Figure 2.1). (The term “GC dinucleotide” is used to describe the complementary 5’ GC/3’CG sequence at the left end of the overlap region). It has been shown previously that the integration reaction catalyzed by IntDOT proceeds via a three-step mechanism: a pair of top strand exchanges on the left side of the overlap region that are dependent upon sequence identity, followed by a sequence-independent

isomerization step, and then a pair of bottom strand exchanges on the right side of the overlap region that are independent of sequence identity (26). It appeared that the only sequence identity between the sites that was required for integration was the GC dinucleotide base pairs on the left side of the overlap regions in both *attDOT* and *attB*. For example, previous work showed that only the base pairs at the position of the GC dinucleotide need to be identical in both substrates for efficient recombination - *attDOT* and *attB* sites where both sites contained CG or AT substituted for the GC dinucleotide were shown to be active (26). Thus productive recombination occurs when both sites have complementary dinucleotide sequences, while the sequence itself is not important.

The overlap regions are flanked by IntDOT core-type binding sites referred to as B and B' in *attB,* or D and D' in *attDOT*. The D and B sites contain a conserved sequence ending in TTTGC (Figure 2.1). The first strand exchanges occur after initial cleavage between the T and G in *attDOT* and *attB* sites. The second strand exchange occurs by cleavage between the A and C residues on the bottom strand of *attDOT* and the two A residues on the bottom strand of *attB1*. Note that ligation of the top strands involves complementary base pairs at the site of ligation while ligation of the bottom strand occurs in the presence of mismatches in the DNA.

An *attB* site with an inverted overlap region functions as an efficient recombination substrate

We wanted to distinguish between three possible general models of the mechanism of ordered strand exchange in IntDOT mediated recombination. The first model proposes that sequences outside of the overlap region (the boxed sequences in Figure 2.1A) play a defining role in determining the order of strand exchange. For example, the arm-type sites could be responsible for determining the order of strand exchange, much like the Lambda Int system (20). A variation of this model is that the core-type sites such as the D and B sites determine the order of strand exchange. If true, the location of the GC dinucleotide within the overlap region should not matter, and initial cleavage and strand exchange should always take place at the same site relative to the arm and core-type sites. The second model proposes that the D and B core-type sites must be contiguous with the GC dinucleotide located in the overlap region so that the entire conserved sequence shared by both sites is responsible for determining the order of strand exchange. If the position of the GC dinucleotide relative to the D and B sites is important, then moving the GC dinucleotide to the opposite end of the overlap region will result in loss of recombination. The third model proposes that only the GC sequence identity within the overlap regions of the recombining *att* sites is important, and the position (either at the left or right side of the overlap region) of these bases in both *attDOT* and *attB* should not affect recombination as long as the sequences are identical in both sites. In addition, initial cleavage and strand exchange will take place at the location of the sequence homology

independent of the flanking sequences. In summary, Model 1 predicts that the arm and/or core-type sites determine the order of strand exchange; Model 2 predicts that the GC dinucleotide must be adjacent to the conserved sequence of the B and D sites; and Model 3 predicts that only the GC dinucleotide within the overlap region determines the order of strand exchange.

In an attempt to distinguish between these three possibilities, we designed an *attB1* site with an inverted overlap region (Figure 2.1B, 2.1C). In this substrate, the entire seven base overlap region is rotated 180° so that the GC dinucleotide is relocated to the 5' end of the bottom strand, contiguous with the B' site (Figure 2.2A, 2.2B). Synapsis of the wild type *attDOT* site and the inverted *attB1* site in orientations shown in Figure 2.2B generates 2 mismatched base pairs (GC in *attDOT* and AT in the inverted *attB1* site) at the position of the original GC dinucleotide. If the flanking core- or arm-type sites determine the order of strand exchange, then the reaction will proceed with initial cleavage and strand exchange taking place adjacent to the D arm, but recombination will be substantially decreased because a homology-dependent strand exchange cannot take place (Figure 2.2C). A substrate with the inverted *attB* site should also not be functional in integrative recombination if the GC dinucleotide must be adjacent to the B site for productive recombination, as proposed in the second model.

We found that the *attB1* site with an inverted overlap region recombined nearly as efficiently as the wild type *attB1* with a partner site containing the wild type *attDOT*. These results suggest a less defining role for the arm-type sites, but do not yet distinguish between the other models. As shown in Figure 2.2D, a

180° rotation of the entire *attB1* site would generate an alignment that contained a GC dinucleotide in the correct position in the top strand adjacent to the B' core site. This could indicate that the position of the GC dinucleotide within the core overlap region dictates the orientation of the *att* sites in a synaptic complex that undergoes recombination. If correct, neither the core-type sites, nor the arm-type sites control the orientation of sites in a synaptic complex. In addition, since the top strands containing the GC dinucleotides are exchanged first in recombination with wild type sites (25, 26), it is possible that the strands containing the GC dinucleotides in reactions between *attDOT* and the inverted overlap *attB1* site are also exchanged first, supporting the third model. Alternatively, the first strand exchange could be taking place between non-identical base pairs, and the location of sequence identity within the overlap region is not as important as the orientation of the B and D sites relative to each other, thereby supporting the first model.

Synaptic complexes containing the *attB1* and inverted overlap *attB1* sites

SspI digestion of the linear recombinant product from wild type *att* sites produces two fragments, 1.1 kb and 2.5 kb long, due to a single asymmetric restriction site within the plasmid containing *attDOT* (Figure 2.3A). During synapsis, the *attDOT* site presumably forms an intasome with four IntDOT monomers and an unknown number of IHF heterodimers bound to the DNA (19, 36). There are two potential orientations of the *attB1* site relative to the intasome in a synaptic complex. If the second model is correct, then the restriction pattern

will be the same as that seen with the wild type *att* sites, meaning the intasome will form a productive synapse with the inverted overlap *attB* site in the same orientation as it does with the wild type *attB* site. On the other hand, if the third model is correct, then an *attB1* site with the inverted overlap will interact with the intasome in the reverse orientation, where the GC dinucleotides are exchanged, and the orientation of the integrated *attDOT* plasmid relative to the ends of the recombinant product will be inverted. When the top strand of a wild type *attB1* is radiolabeled, digestion with SspI of the resulting recombinant yields a 1.1 kb fragment. A 2.5 kb fragment appears if the bottom strand is radiolabeled (Figure 2.3B). After recombination with the *attB1* site containing the inverted overlap region, the opposite fragment pattern appears. With a label on the top strand, the 2.5 kb fragment is produced and the 1.1 kb fragment appears with a bottom strand label. This confirms that the orientation of *attDOT* relative to the *attB* sites in the two sites synaptic complexes are different and that the first strand exchange involves exchange of the strands containing the GC dinucleotide. We conclude from this experiment that model 2 is incorrect.

An *attB* site with an inverted overlap sequence shows a strong bias for bottom strand exchange

Since the orientation between *attDOT* and *attB* during productive synapsis appeared to be changed with the *attB* carrying the inverted overlap region, we wanted to determine if the order of strand exchange was also altered. We previously showed that the first strand exchanges between the *attDOT* and *attB1*

sites occurred at the top strands at sites adjacent to the GC dinucleotide (26). We used complementary oligonucleotides that introduced a nick at either the top strand or bottom strand cleavage site of *attB1*. Nicked substrates allow the intact strand to be cleaved and exchanged to form a Holliday junction containing a nick. However a nick at the second cleavage site prevents the exchange of the second strand, thereby trapping a Holliday intermediate. Alternatively, if the nick is at the site of the first strand exchange, no intermediate is formed (Figure 2.4A). In the integration reaction between wild type *attDOT* and wild type *attB1*, the top strand is cleaved and exchanged first, so a nick in the top strand inhibits the first step of the reaction and no product is formed. A nick on the bottom strand allows the top strand to be cleaved and exchanged forming the Holliday junction intermediate product that is detectable on an agarose gel. We constructed nicked inverted overlap *attB1* sites and tested them with wild type *attDOT*. If the bottom strand is cleaved and exchanged first, then the Holliday junction intermediate would be formed in the reaction with the inverted overlap *attB1* site containing a nick in the top strand but not from the *attB1* site containing a nick in the bottom strand.

The results of the experiments with the inverted overlap *attB1* sites are shown in Figure 2.4B and Figure 2.4C. Products were detected from the inverted *attB1* site with a nick at the top strand cleavage site, but to a much lesser extent from the inverted *attB1* site with a nick at the bottom strand cleavage site. This result indicates that the frequency of first strand exchange between the *attDOT* top strand and the bottom strand of the inverted *attB1* site containing the GC

dinucleotide occurs approximately 9 times more frequently than the exchange with the top strand.

An *attB1* site with a symmetric overlap produces recombinants in both orientations

If the orientation of the integrating *attDOT* site relative to the *attB* site is dependent on sequence identity at the position of the first exchange within the overlap region, then presenting an *attDOT* intasome with a site containing a symmetric overlap region with a GC dinucleotide in both strands should produce recombinants where the *attDOT* site recombined with the symmetric core in both orientations with equal frequency (Figure 2.1D). Figure 2.5A shows the results of experiments where the symmetric overlap *attB1* site was labeled on the top or bottom strand. Both the 1.1 and the 2.5 kb products are formed regardless of which strand is labeled (Figure 2.5). These results suggest that the symmetric overlap *attB1* site can synapse with *attDOT* in both orientations at equal frequencies. These results also support the argument that synapsis occurs at equal frequency with wild type *attB1* in both orientations but only one orientation produces recombinants.

These results are confirmed by experiments done using *attB* sites with a symmetric overlap region containing a nick at either the top or bottom strand cleavage site. Products were detected regardless of which strand contained the nick, indicating that synapsis and cleavage takes place with the substrates in either orientation (Figure 2.5B). The higher amount of product detected with the

attB site containing the nick at the bottom strand cleavage site may suggest a slight preference for a particular orientation of the D and B sites relative to each other in the synaptic complex.

Other known *attB* sites produce recombinants similar to *attB1*

There are six *Bacteroides attB* sites that have been identified and sequenced (10). All six sites contain a conserved sequence on the left side, ending in TTTGC (Table S2). We wanted to determine whether other *attB* sites behave similarly to *attB1* sites containing different overlap and B' sequences. Accordingly, we constructed *attB2-6* sites using complementary oligonucleotides and tested them in the *in vitro* integration assay with a wild type *attDOT*. We also constructed *attB3* and *attB5* sites with inverted overlap regions to confirm that the location of the GC dinucleotide within the overlap region dictates the orientation of the integrating *attDOT.* The *attB3* and *attB5* sites were chosen for experiments because their sequences are the least similar to the *attDOT* sequence. The results of the *in vitro* integration assay show that all six *attB* sites behave in a similar fashion and have relatively similar recombination efficiencies, with *attB4* being the least efficient of the *attB* sites (Table 2.2). The site of initial cleavage takes place between the T and G, making the conserved GC dinucleotide part of the overlap region. The recombinants produced from *attB3* and *attB5* sites containing an inverted overlap region show the same restriction pattern as recombinants produced from the *attB1* site with an inverted overlap region. We

conclude from these results that the known *attB* sites appear to act similarly with respect to the mechanism for ordered strand exchange.

The overlap region dictates orientation of integrating *attDOT*

If the B core-type site of *attB* plays a role in orienting the integrative intasome for productive recombination, then having the B site flanking both sides of the overlap region should also induce integration in both orientations. On the other hand, if sequence identity in the overlap region is key, then recombination will depend on the location of sequence identity in the overlap region. To further investigate the role of the conserved B region, an *attB1* site where both core sites contained B sequences were constructed using complementary oligonucleotides. The core overlap regions were varied to be wild type, symmetric or inverted (Figure 2.6A). Analysis of BoB derived recombinant products containing a wild type, inverted or symmetric overlap region shows similar frequencies as *attB1* derived recombinants with the same overlap regions (Figure 2.6B). Taken together, the data strongly suggests that the core overlap region is responsible for determining how the integrating *attDOT* is oriented relative to the *attB* during synapsis that leads to recombination. It also determines which strand is cleaved and exchanged first.

An *attDOT* site with an inverted overlap recombines with *attB1* in the opposite orientation compared to wild type *attDOT*

If the hypothesis that the orientation of the overlap region within an *attB1* site dictates the order of strand exchange is correct, then a substrate containing an inverted overlap *attDOT* site should integrate in the opposite orientation as the wild type *attDOT* when recombined with either wild type or inverted overlap *attB*s. An *attDOT* derivative containing an inverted overlap region between the D and D' sites was produced via site-directed mutagenesis to test this proposition (Figure 2.7). The results are consistent with this model. The restriction pattern from the wild type *attDOT* x wild type *attB* product showed the 1.1 kb product when the top strand was labeled, the 2.5 kb product when the bottom strand was labeled, and the opposite pattern when the *attB*1 site contained the inverted overlap region (Figure 2.3). The restriction pattern from the inverted overlap *attDOT* x wild type *attB1* product showed the predicted pattern, but recombination levels were depressed nearly 15 fold (Figure 2.7 B). Thus it is likely that the identity of the core sites of *attDOT* plays a role in the efficiency of the integration reaction.

A single base sequence identity adjacent to the site of cleavage and ligation is sufficient for recombination

We established above that the location of the GC dinucleotide in the overlap region determines the order of strand exchange and provides enough sequence identity for recombination to occur. We next asked if a single base pair

of sequence identity between both *att* sites is sufficient for recombination and whether that single base pair needs to be at the site of cleavage. Presenting a wild type *attDOT* site with an *attB* site containing a CC in place of the GC dinucleotide should decrease recombination due to a lack of sequence identity between the bases at either cleavage site. As expected, recombination was abolished. This indicates that sequence identity is required at the base adjacent to cleavage in the overlap region and that the presence of the C of the GC dinucleotide is not sufficient for recombination to occur.

We then constructed an *attDOT* mutant containing a terminal G substitution in the overlap region (TermG *attDOT*), making the overlap sequence GCTTAGG (Figure 2.8). We combined this *attDOT* site with both wild type *attB* and an *attB* site containing the overlap sequence ATGCAAA (AT *attB*). Not surprisingly, the *attDOT* containing the terminal G substitution recombined well with the wild type *attB* site since both sites share a G at the cleavage site (Figure 2.8B, 2.8C). Also as expected, the TermG *attDOT* site recombined almost negligibly with the AT *attB* site (< 0.5%). The wild type *attDOT* site, however, was able to recombine with the AT *attB* site, albeit poorly (approximately 3%). This is explained by the fact that the two sites share a single base of sequence identity at the cleavage sites – the T at the 3' end of the *attDOT* overlap region and the T at both ends of the overlap region in an inverted AT *attB* site (Figure 2.8B). Taken together, this data suggests that only one base of sequence identity is sufficient for low levels of recombination, but that two bases of sequence identity are

required for maximum recombination. The sequence identity must be located at the site of cleavage and ligation.

Discussion

There are two main subclasses of tyrosine recombinases. The first is the autonomous class that only contains two domains – CB and CAT domains. These recombinases recognize core-type DNA sites and do not require accessory proteins. The second subclass is characterized by the presence of an N-terminal domain that binds to arm-type DNA sites, and a requirement for additional proteins to determine directionality.

Some tyrosine recombinases also have a preferred order of strand exchanges while others do not (23). There does not appear to be a connection between complexity of the recombination systems and the mechanism of regulating the order of strand exchanges. For example, Flp is an autonomous recombinase that does not have a defined order of strand exchange – it will cleave and exchange both strands with equal frequency (1). Cre recombinase also falls under the autonomous class, yet when it binds to its *loxP* substrate, it induces an asymmetric bend that determines which strand will be exchanged first (16, 22, 23, 27). The Xer system involves two tyrosine recombinases, XerC and XerD that work in tandem to ensure chromosome and plasmid segregation during cell division. XerC/D can work as autonomous or regulated recombinases depending upon the substrate. For example, the chromosomal *dif* site only

consists of core binding sites for XerC and XerD, while the *psi* site on the plasmid pSC101 requires accessory proteins bound to additional DNA sites. XerC makes the initial cleavage at *psi* sites resulting in ordered strand exchange, while XerD can initiate cleavage at *dif* sites (2, 3, 12, 17, 23, 35).

Lambda Int is a factor-assisted recombinase that contains three domains. It requires accessory proteins to form a protein-DNA complex called an intasome that catalyzes excision and integration (33, 36). The arm-type DNA sites contain both Int and accessory protein binding sites, and are arranged asymmetrically around the core-type sites where cleavage and strand exchange takes place (37). Kitts and Nash used inverted *attP* sites to determine whether the core or arm-type sites dictate the order of strand exchange. They found that regardless of the orientation of the core, initial cleavage takes place at the site of the overlap region closest to the P arm. Their results established that the order of strand exchange is determined by the asymmetry of the arm-type sites (20). Interestingly, when the N-terminal arm-binding domain of Lambda Int is attached to Cre, the chimera becomes dependent on IHF, Xis and Fis to regulate directionality in the same fashion as Lambda Int (45). This suggests that the factor-assisted recombinases are particularly reliant upon the contribution of the accessory proteins and the spatial arrangement of the additional DNA sites around the core to determine the mechanism and order of strand exchange.

Like Lambda Int, IntDOT possesses an arm-binding N-terminal domain. The *attDOT* site contains five IntDOT binding sites arranged asymmetrically around the core site, and intasome formation requires additional accessory

proteins (10, 14). Despite the asymmetric arrangement of the arm-type sites relative to the core sites, the order of strand exchange appears to be predominantly dictated by the core sites. In a reaction with wild-type sites, IntDOT preferentially cleaves and exchanges the top strand first. The first strand exchange requires a sequence identity between the recombining *att* sites, but the second strand exchange is independent of sequence identity. It has also been shown that the identity of the base pairs at the site of cleavage is not important, so long as they are identical between the two sites (26). We have shown here that the location of sequence identity within the overlap region determines the order of strand exchange in IntDOT-catalyzed integration, and that the arm-type sites are less important than in other systems.

Tyrosine recombinases catalyze recombination by performing a pair of strand exchanges at opposite ends of the overlap region. Most tyrosine recombinases have an absolute requirement for sequence identity within the overlap region of recombining substrates, and often the initial strand exchange is dependent upon sequence identity between recombining sites. It has been suggested that Watson-Crick base pairing must take place between the partner strands after strand swapping in order for the ligation reaction to occur (30). If no base pairing can occur, then the ligation step is not likely to take place. This idea is supported by work done on Flp showing that the ligation step is the sequence identity-dependent step (21). Likewise, mutations in the overlap region of *loxP* nearly abolishes recombination between wild type and mutant *loxP* sites (18).

Sites containing heterology between the overlap regions are very poor recombination substrates (46). For example, Lambda *site affinity* mutants containing mutations within the overlap region of *attP* or *attB* reduce recombination, but when the same mutations are made in both *att* sites, recombination is restored. Mutations made directly to the left of the cleavage sites, outside of the overlap region, had no effect on recombination (5). It has also been shown that sequence identity in the left side of the overlap region is strongly conserved in Lambda secondary attachment sites, supporting a critical role for sequence identity between recombining substrates (38). Perhaps the sequence identity on the left side promotes initial strand exchange to form the Holliday junction intermediate. IntDOT can resolve the Holliday junction into products due to its tolerance of heterology, whereas the Lambda Holliday junction requires host processes to resolve it into products in the presence of heterology.

Studies done on IntDOT using synthetic ligation substrates containing *p*-nitrophenol at the site of cleavage and ligation have shown that IntDOT can perform the ligation reaction in the presence of heterology (26). The same study also used bridging phosphorothioate substrates to show that cleavage can also take place in the presence of a heteroduplex in the overlap region (26). These results make the role of sequence identity in IntDOT-mediated recombination unclear and raises the question as to why sequence identity is so important within the DOT overlap region for one strand exchange while heterology is tolerated extremely well for the second strand exchange.

We assume that IntDOT, like other tyrosine recombinases, forms an intasome complex with *attDOT* during integration, and that the intasome synapses with a naked *attB* site (36). Synapsis of the intasome with *attB* can occur in two orientations: the intasome can synapse with the *attB* site from the "top", or from the "bottom". Theoretically, cleavage and initial strand exchange could take place in either orientation, but the reaction does not proceed unless Watson-Crick base pairing can take place between the newly-exchanged strands. If a complementary base is not available, the reaction is reversed. The *attDOT* and *attB* sites are formed, and the *attB* site is released. If a complementary base is available in the partner site, the ligation reactions occur and the first strand exchange is executed. A second round of cleavage and ligation reactions occurs at the other border of the overlap region, which does not require Watson-Crick base pairs at the ligation sites, and recombinants are formed. This is consistent with the results of our symmetric overlap recombination experiments. Because the conserved GC dinucleotide is present on both the top and bottom strands of *attB* at the cleavage sites, the *attDOT*/ intasome utilizes both sites of sequence identity with equal frequency.

Warren, et al. suggests that two evolutionary steps would be required for an autonomous recombinase like Cre to evolve into the more factor-assisted recombinase like Lambda Int. These steps are the addition of another DNA-binding domain and a "detuning" of the recombinase, which involves a decrease in the affinity for particular DNA sites and/or reduced protein-protein interactions (45). IntDOT appears to be an evolutionary intermediate that has gained the

second DNA binding domain but has not yet lost its relatively high affinity for the core-type DNA sites. Combined with a relaxed requirement for sequence identity within the overlap region, these characteristics suggest that IntDOT may be a member of a new subclass of tyrosine recombinase that could potentially include the integrases of other conjugative transposons such as Tn*916* and Tn*1545* that recombine *att* sites containing non-identical overlap regions.

Figures and Tables

Table 2.1: Oligonucleotide sequence list

<u>WT *attB1* top</u>
5' AGC TTG CAA CGC TTG CTC GGA AAT TTG CAG TAA ATT TGC GCA ATT AAA ATA CTA AAC AGT AAT TAT ATC ATG GCA GCA AG 3'
<u>WT *attB1* bott</u>
5' CTT GCT GCC ATG ATA TAA TTA CTG TTT AGT ATT TTA ATT GCG CAA ATT TAC TGC AAA TTT CCG AGC AAG CGT TGC AAG CT 3'
<u>Inv. *attB1* top</u>
5' AGC TTG CAA CGC TTG CTC GGA AAT TTG CAG TAA ATT TAT TGC GCT AAA ATA CTA AAC AGT AAT TAT ATC ATG GCA GCA AG 3'
<u>Inv. *attB1* bott</u>
5' CTT GCT GCC ATG ATA TAA TTA CTG TTT AGT ATT TTA GCG CAA TAA ATT TAC TGC AAA TTT CCG AGC AAG CGT TGC AAG CT 3'
<u>Inv. top nick1 *attB1*</u>
5' CGT TGC TCG GAA ATT TGC AGT AAA TTT 3'
<u>Inv. top nick 2 *attB1* (5' p)</u>
5' ATT GCG CGA AAA TAC TAA ACA GTA ATT ATA TCA TGG 3'
<u>Inv. bott nick 1 *attB1*</u>
5' CCA TGA TAT AAT TAC TGT TTA GTA TTT TA 3'
<u>Inv. bott nick 2 *attB1* (5' p)</u>
5' GCG CAA TAA ATT TAC TGC AAA TTT CCG AGC AAC G 3'
<u>Sym. *attB1* top</u>
5' AGC TTG CAA CGC TTG CTC GGA AAT TTG CAG TAA ATT TGC AAT GCT AAA ATA CTA AAC AGT AAT TAT ATC ATG GCA GCA AG 3'
<u>Sym. *attB1* bott</u>
5' CTT GCT GCC ATG ATA TAA TTA CTG TTT AGT ATT TTA GCA TTG CAA ATT TAC TGC AAA TTT CCG AGC AAG CGT TGC AAG CT 3'
<u>Sym. top nick 1 *attB1*</u>
5' CGC TTG CTC GGA AAT TTG CAG TAA ATT T 3'
<u>Sym. top nick 2 *attB1* (5' p)</u>
5' GCA ATG CTA AAA TAC TAA ACA GTA ATT ATA TCA 3'
<u>Sym. bott nick 1 *attB1*</u>
5' CCA TGA TAT AAT TAC TGT TTA GTA TTT TA 3'
<u>Sym. bott nick 2 *attB1* (5' p)</u>
5' GCA TTG CAA ATT TAC TGC AAA TTT CCG AGC AAC G 3'
<u>Sym. BoB top</u>
5' CGT TGC TCG GAA ATT TGC AGT AAA TTT GCA ATG CAA ATT TAC TAA ACA GTA ATT ATA TCA 3'
<u>Sym. BoB bott</u>
5' TGA TAT AAT TAC TGT TTA GTA AAT TTG CAT TGC AAA TTT ACT GCA AAT TTC CGA GCA ACG 3'

Table 2.1 (cont.): Oligonucleotide sequence list

Inv. BoB top
5' CGT TGC TCG GAA ATT TGC AGT AAA TTT ATT GCG CAA ATT TAC TAA ACA GTA ATT ATA TCA 3'
Inv. BoB bott
5' TGA TAT AAT TAC TGT TTA GTA AAT TTG CGC AAT AAA TTT ACT GCA AAT TTC CGA GCA ACG 3'
AT *attB*1 top
5' CGT TGC TCG GAA ATT TGC AGT AAA TTT ATG CAA ATA AAA TAC TAA ACA GTA ATT ATA TCA 3'
AT *attB1* bott
5' TGA TAT AAT TAC TGT TTA GTA TTT TAT TTG CAT AAA TTT ACT GCA AAT TTC CGA GCA ACG 3'
CC *attB1* top
5' CGT TGC TCG GAA ATT TGC AGT AAA TTT CCG CAA TTA AAA TAC TAA ACA GTA ATT ATA TCA 3'
CC *attB1* bott
5' TGA TAT AAT TAC TGT TTA GTA TTT TAA TTG CGG AAA TTT ACT GCA AAT TTC CGA GCA ACG 3'
***attB2* top**
5' ATC CAA TAC ATA TTT AAT AGG TAT TTT TGC ATA ATA AAC ATT GCT GTT CGG TTT GCT AAG 3'
***attB2* bott**
5' CTT AGT AAA CCG AAC AGC AAT GTT TAT TAT GCA AAA ATA CCT ATT AAA TAT GTA TTG GAT 3'
***attB3* top**
5' TCA ATC CCG GTT CTT TTT TCG TAC TTT TGC GGT GAA ATT TTA TAG CTA TGA CAA TAA CTG 3'
***attB3* bott**
5' CAG TTA TTG TCA TAG CTA TAA AAT TTC ACC GCA AAA GTA CGA AAA AAG AAC CGG GAT TGA 3'
Inv. *attB3* top
5' TCA ATC CCG GTT CTT TTT TCG TAC TTT TTC ACC GCA ATT TTA TAG CTA TGA CAA TAA CTG 3'
Inv. *attB3* bott
5' CAG TTA TTG TCA TAG CTA TAA AAT TGC GGT GAA AAA GTA CGA AAA AAG AAC CGG GAT TGA 3'
***attB4* top**
5' ATA AGG TAG TGC AGT ACA TTG TAG ATT TGC TGG ACG AAT GTC TTG AAG ATC TTC CGA GAG 3'
***attB4* bott**
5' CTC TCG GAA GAT CTT CCA GAC ATT CGT CCA GCA AAT CTA CAA TGT ACT GCA CTA CCT TAT 3'

Table 2.1 (cont.): Oligonucleotide sequence list

***attB5* top**
5’ CTC ATT TTG TTT TAC TTT ATG TTG TTT TGC ACT ACA AAG ATA ACA TAT TTC GCA ACG ATT 3’
***attB5* bott**
5’ AAT CGT TGC GAA ATA TGT TAT CTT TGT AGT GCA AAA CAA CAT AAA GTA AAA CAA AAT GAG 3’
Inv. *attB5* top
5’ CTC ATT TTG TTT TAC TTT ATG TTG TTT TGT AGT GCA AAG ATA ACA TAT TTC GCA ACG ATT 3’
Inv. *attB5* bott
5’ AAT CGT TGC GAA ATA TGT TAT CTT TGC ACT ACA AAA CAA CAT AAA GTA AAA CAA AAT GAG 3’
***attB6* top**
5’ ATG CCA GTT TTT CCA ATA TTG TAT ATT TGC TTT ATA AAA ATG AAG GAA GAA TCT GGT TCT 3’
***attB6* bott**
5’ AGA ACC AGA TTC TTC CTT CAT TTT TAT AAA GCA AAT ATA CAA TAT TGG AAA AAC TGG CAT 3’

	SUBSTRATE			AVG. % RECOMBINATION
attDOT	**D:** TA<u>**GTAACTTT**</u>	**gc**ttagt	**D':** TCGTAACTTC	
attB1	**B:** CA<u>**GTAAATTT**</u>	**gc**gcaat	**B':** TAGTATTTTA	31%
attB2	**B:** AG<u>**GTATTTTT**</u>	**gc**ataat	**B':** AGCAATGTTT	24%
attB3	**B:** TC<u>**GTACTTTT**</u>	**gc**ggtga	**B':** CTATAAAATT	18%
attB4	**B:** TT<u>**GTAGATTT**</u>	**gc**tggac	**B':** CCAGACATTC	11%
attB5	**B:** AT<u>**GTTGTTTT**</u>	**gc**actac	**B':** TGTTATCTTT	15%
attB6	**B:** TT<u>**GTATATTT**</u>	**gc**tttat	**B':** CTTCATTTTT	27%
Consensus	**GTANWTTT**	**GCdbdrh**	**HNNNAHNWAW**	

Table 2.2: **Alignment of *attDOT* with known *attB*s**. The conserved D and B regions in *attDOT* and *attB* respectively, are underlined. A consensus sequence is shown below the aligned sequences (B represents C or G or T; D represents A or G or T; H represents A or C or T; N represents any base; R represents A or G; W represents A or T).

Figure 2.1: Core region of *attDOT* and *attB*. A) Wild type sequences of the core binding sites *attDOT* and *attB*. The boxed region indicates the overlap region (o). The conserved GC dinucleotide is shown at the left sides of the overlap regions. The small arrow represents orientation of the overlap region relative to flanking arm sites. The longer arrows indicate the core-type sites. The vertical arrows show the sites of cleavage. B) Simplified schematic diagram of the same core regions. The bubbled area denotes the GC dinucleotide homology between the *attDOT* and *attB1* overlap regions. C) Schematic diagram of the *attB* site with an inverted overlap region. D) Schematic diagram of the *attB* site containing a symmetric overlap region. E) Sequences of the wild type, inverted and symmetric overlap regions.

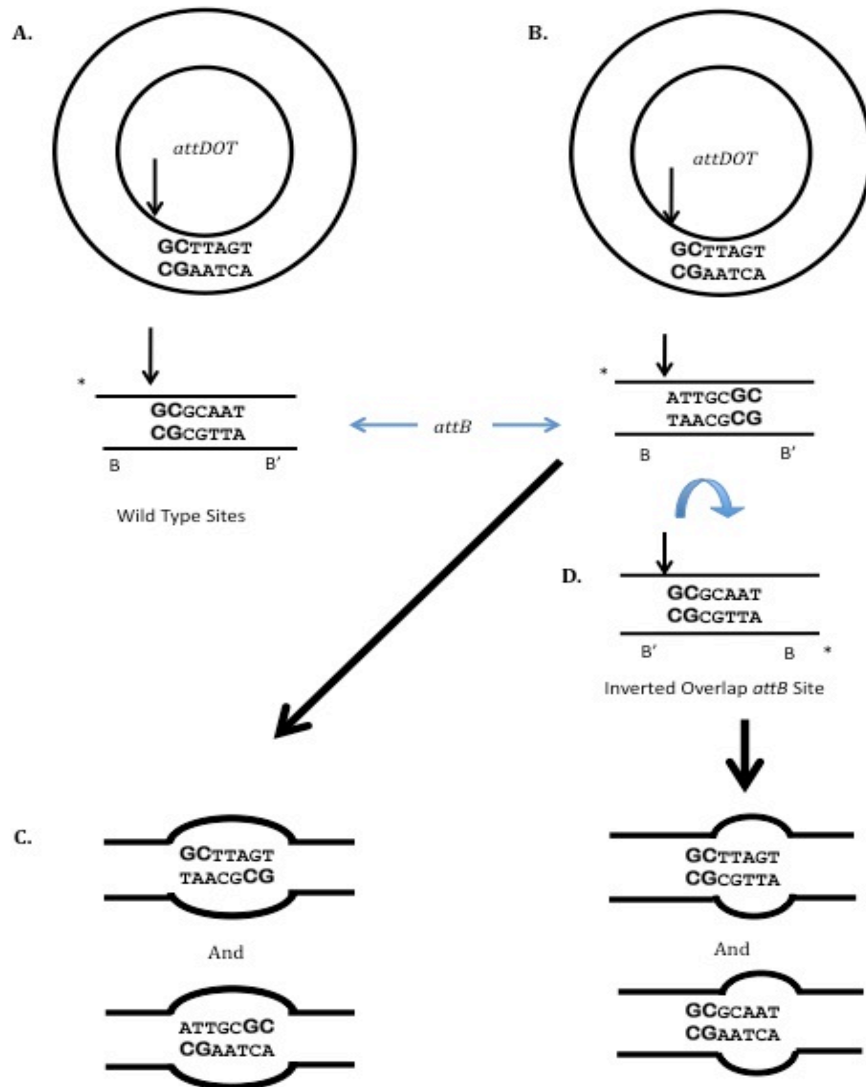


Figure 2.2: How location of sequence identity effects integration. Schematic model of how the plasmid containing the *attDOT* site orients itself relative to the *attB* to maintain alignment of the GC dinucleotide.

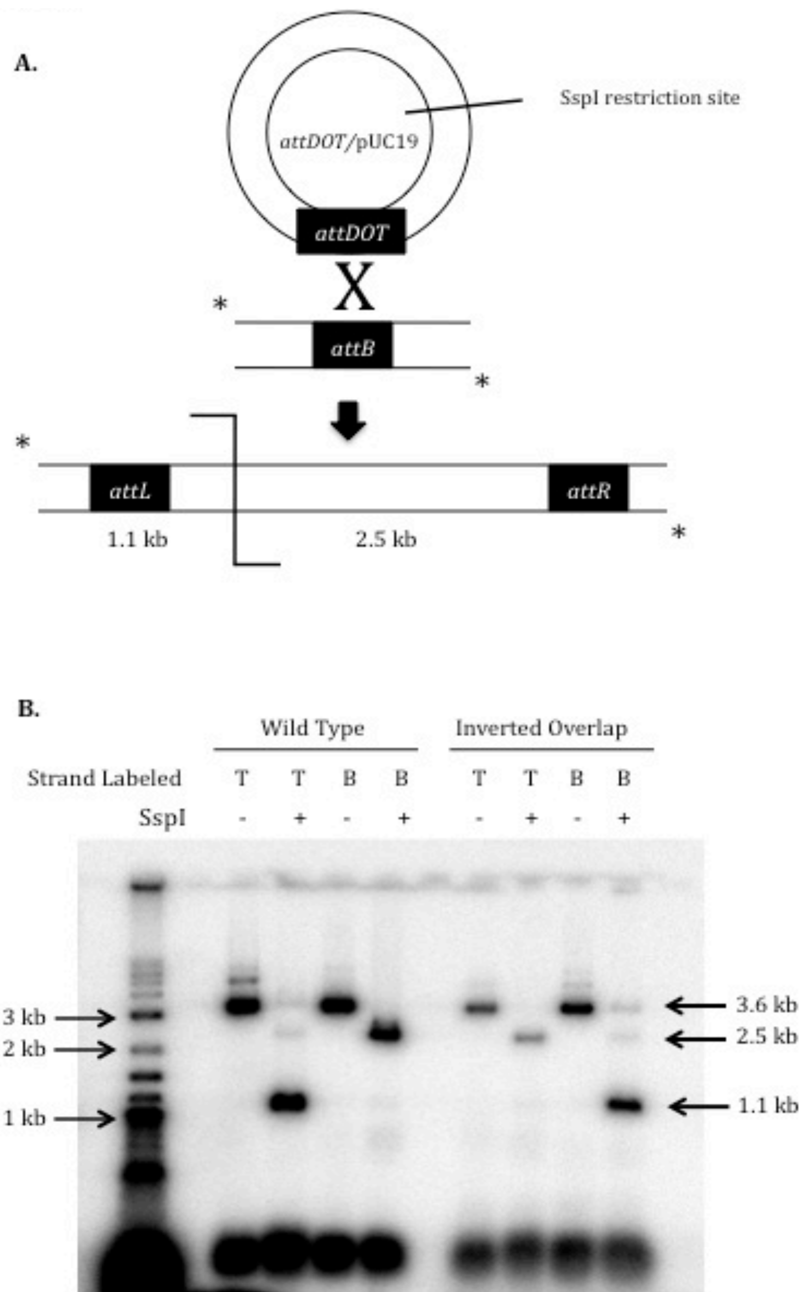
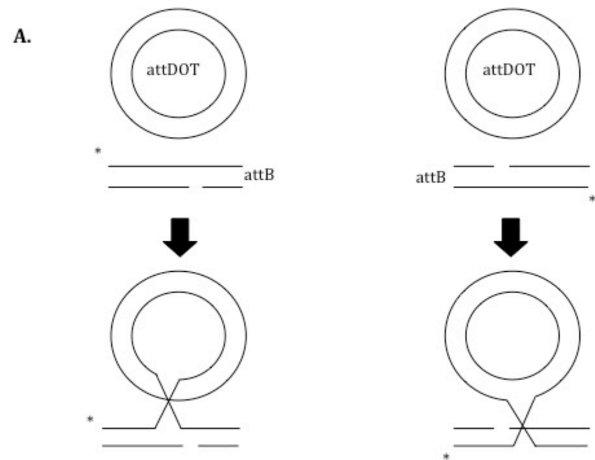
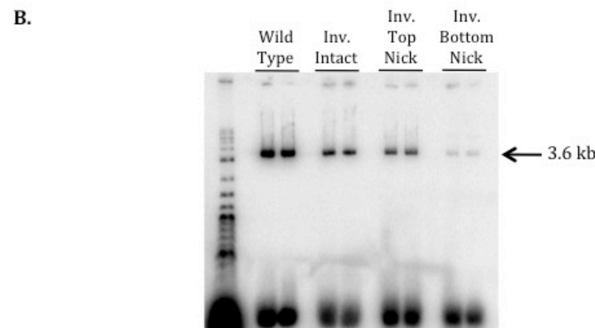


Figure 2.3: *in vitro* integration assay. A) Schematic representation of the *in vitro* integration assay and the single asymmetric SspI site that produces 1.1kb and 2.5 kb fragments from the linear recombinant. B) Results from a SspI digest of recombinants produced from wild type *attDOT* and either wild type or inverted overlap *attB* sites.





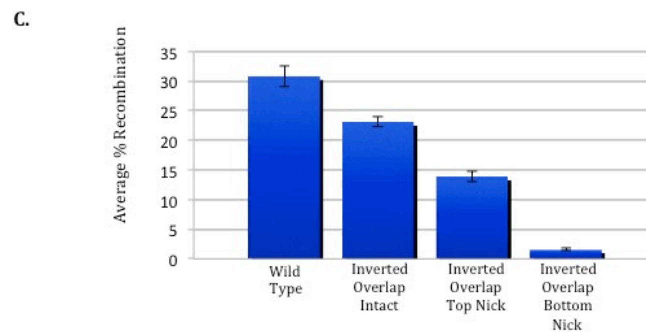
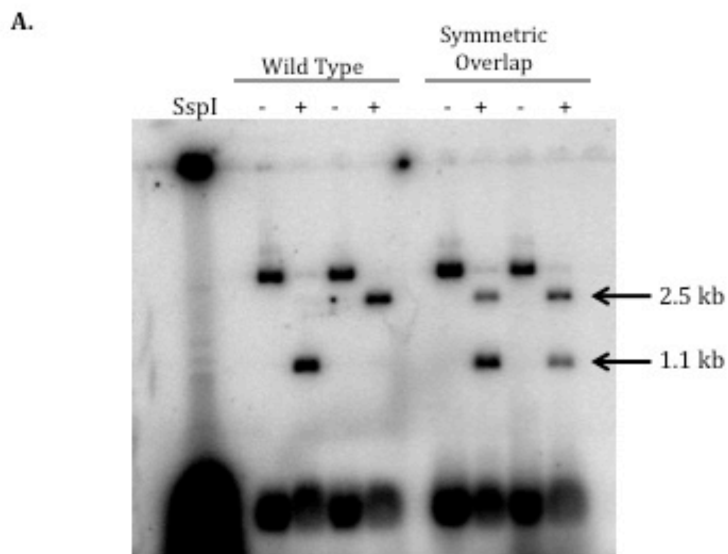


Figure 2.4: Determining the order of strand exchange with nicked substrates. A) Recombination reactions with *attB* suicide substrates nicked on either the top or bottom strand. Initial strand exchange can occur between the intact strand of *attB* and the corresponding strand of *attDOT*. The strand containing a nick cannot undergo recombination and so traps a Holliday intermediate. B) Integration assay results using wild type *attDOT* and either wild type *attB*; inverted overlap intact *attB*; inverted overlap *attB* containing a nick in the top strand; or inverted overlap *attB* containing a nick in the bottom strand. C) Graph showing the average percent recombination of the intact and nicked *attB*s over a minimum of four experiments. The error bars represent standard error.



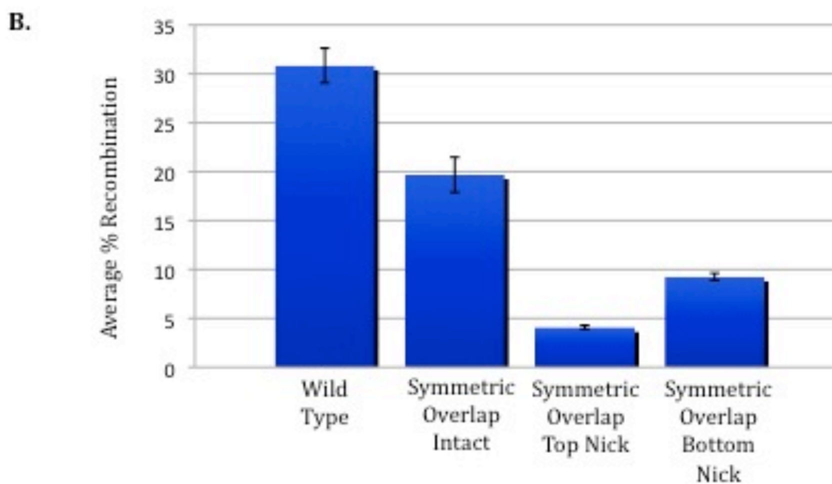
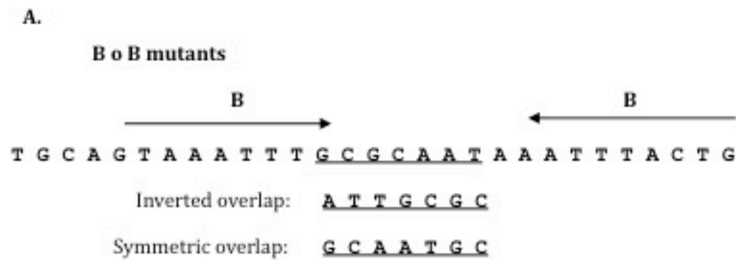


Figure 2.5: Recombination with an *attB* site containing a symmetric overlap region. A) SspI restriction digest of recombinants produced from either wild type *attB1* (WT) or *attB* containing a symmetric overlap sequence (SO). B) Graph comparing the average percent recombination of wild type *attB1*, an *attB site* containing a symmetric overlap region, and *attB* sites with a symmetric overlap region nicked in either the top or bottom strand at the cleavage site. All *attB* sites were combined with a wild type *attDOT* and averaged over a minimum of four experiments.



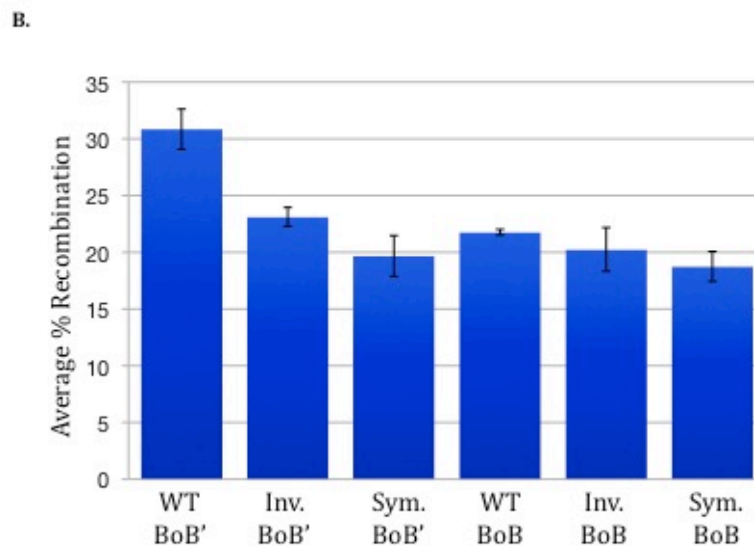


Figure 2.6: Substrates with modified core sequences. A) Sequence of BoB *attB* mutants with each core overlap sequence. The B region of the core is normally only on the 5' side. In the BoB mutant, it is present on both sides. In this study, they are combined with wild type *attDOT* containing both the D and D' core-type sites. B) Comparison of the average percent recombination of wild type *attDOT* combined with wild type (WT BoB'), inverted overlap (Inv. BoB'), or symmetric overlap *attB1* (Sym BoB'), and wild type *attDOT* combined with BoB *attB* sites containing either wild type (WT BoB), inverted overlap (Inv. BoB) or symmetric overlap (Sym. BoB) regions. The average recombination was taken over a minimum of four experiments.

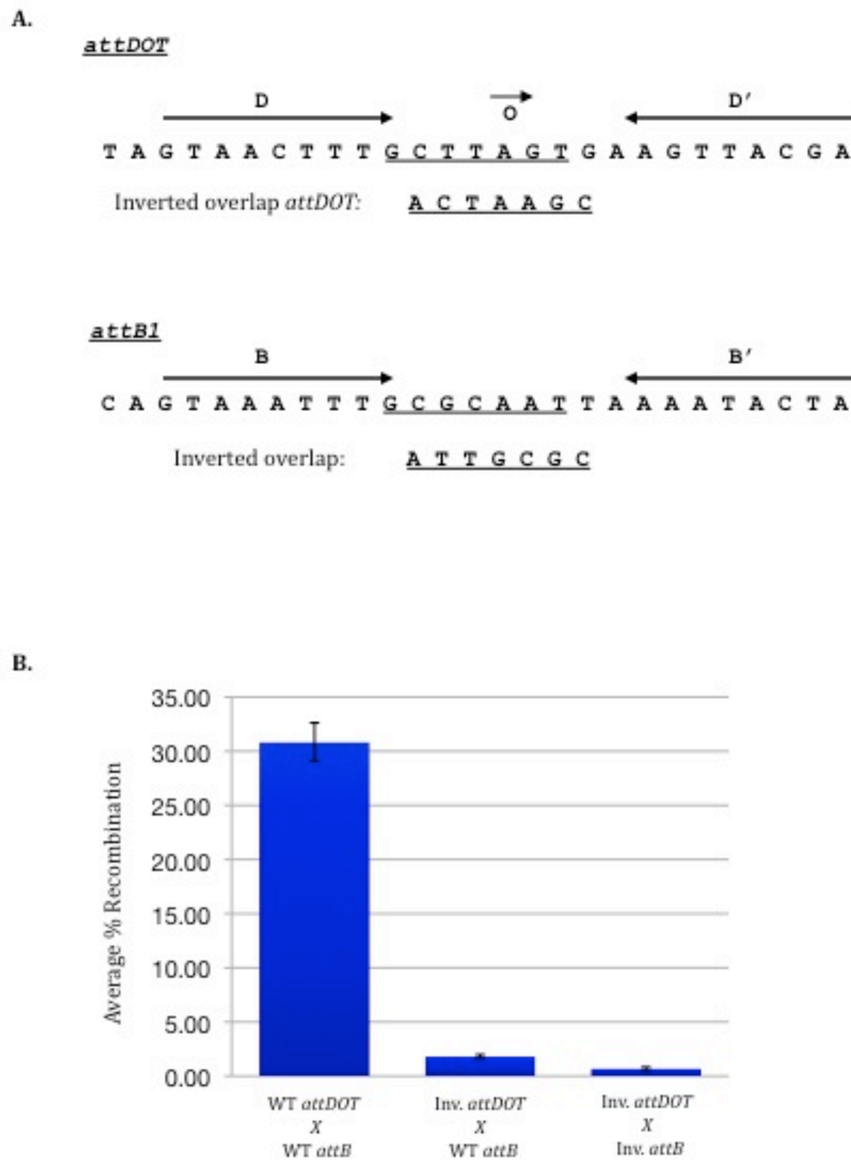


Figure 2.7: Recombination with substrates containing an inverted overlap region. A) Sequences of *attDOT* and *attB1* with either wild type or inverted overlap regions. B) Integration assay results from combinations of wild type (WT) and inverted overlap (I.O.), *attDOT* and *attBs*. The average recombination was taken over a minimum of four experiments.

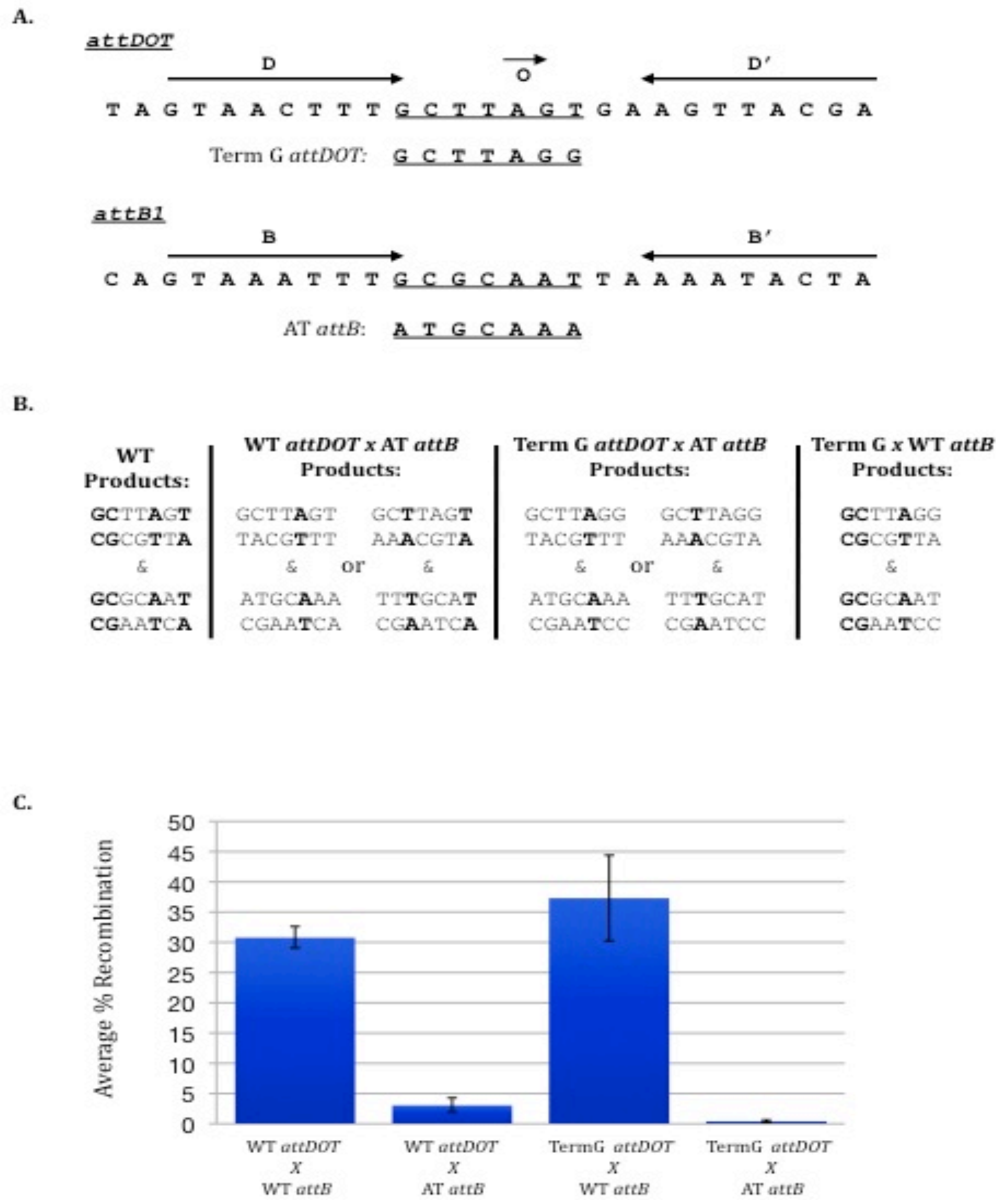


Figure 2.8: Extent of sequence identity required for integration. A) Sequence comparison of wild type *attDOT* and an *attDOT* site with a terminal G mutation in the overlap region, and wild type *attB* with an *attB* site containing substitutions at both ends of the overlap region. B) Sequences of products formed from the recombination of wild type or TermG attDOT with either wild type or AT attB. The second pair of sequences in each combination is the products from an inverted attB site. C) Integration assay results from combinations of wild type or termG *attDOT* sites with either wild type or AT *attB* sites. The average recombination was taken over a minimum of four experiments.

References

1. **Amin, A. A., L. G. Beatty, and P. D. Sadowski.** 1990. Synaptic intermediates promoted by the FLP recombinase. J Mol Biol **214:**55-72.

2. **Arciszewska, L. K., and D. J. Sherratt.** 1995. Xer site-specific recombination in vitro. EMBO J **14:**2112-2120.

3. **Aussel, L., F. X. Barre, M. Aroyo, A. Stasiak, A. Z. Stasiak, and D. Sherratt.** 2002. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. Cell **108:**195-205.

4. **Azaro, M., and A. Landy.** 2002. Lambda Integrase and the lambda Int family, p. 119-148. *In* N. L. Craig, Craigie, R., Gellert, M. and Lambowitz, A. M. (ed.), Mobile DNA II. ASM Press, Washington, D.C.

5. **Bauer, C. E., J. F. Gardner, and R. I. Gumport.** 1985. Extent of sequence homology required for bacteriophage lambda site- specific recombination. J Mol Biol **181:**187-197.

6. **Bedzyk, L. A., N. B. Shoemaker, K. E. Young, and A. A. Salyers.** 1992. Insertion and excision of Bacteroides conjugative chromosomal elements. J. Bacteriol. **174:**166-172.

7. **Blakely, G., S. Colloms, G. May, M. Burke, and D. Sherratt.** 1991. Escherichia-Coli Xerc Recombinase Is Required for Chromosomal Segregation at Cell-Division. New Biologist **3:**789-798.

8. **Burrus, V., G. Pavlovic, B. Decaris, and G. Guedon.** 2002. Conjugative transposons: the tip of the iceberg. Mol Microbiol **46:**601-610.

9. **Cheng, C., P. Kussie, N. Pavletich, and S. Shuman.** 1998. Conservation of Structure and Mechanism between Eukaryotic Topoisomerase I and Site-Specific Recombinases. Cell **92:**841-850.

10. **Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2000. Integration and excision of a Bacteroides conjugative transposon, CTnDOT. J Bacteriol **182:**4035-4043.

11. **Cheng, Q., N. Wesslund, N. B. Shoemaker, A. A. Salyers, and J. F. Gardner.** 2002. Development of an in vitro integration assay for the Bacteroides conjugative transposon CTnDOT. J Bacteriol **184:**4829-4837.

12. **Colloms, S. D., R. McCulloch, K. Grant, L. Neilson, and D. J. Sherratt.** 1996. Xer-mediated site-specific recombination in vitro. Embo J **15:**1172-1181.

13. **Craig, N. L., and H. A. Nash.** 1983. The mechanism of phage lambda site-specific recombination: site- specific breakage of DNA by Int topoisomerase. Cell **35:**795-803.

14. **Dichiara, J. M., A. N. Mattis, and J. F. Gardner.** 2007. IntDOT interactions with core- and arm-type sites of the conjugative transposon CTnDOT. J Bacteriol **189:**2692-2701.

15. **Esposito, D., and J. J. Scocca.** 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. Nucleic Acids Res **25:**3605-3614.

16. **Guo, F., D. N. Gopaul, and G. D. Van Duyne.** 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. Proc Natl Acad Sci U S A **96:**7143-7148.

17. **Hallet, B., L. K. Arciszewska, and D. J. Sherratt.** 1999. Reciprocal Control of Catalysis by the Tyrosine Recombinases XerC and XerD: An Enzymatic Switch in Site-Specific Recombination. Molecular Cell **4:**949-959.

18. **Hoess, R. H., A. Wierzbicki, and K. Abremski.** 1986. The role of the loxP spacer region in PI site-specific recombination. Nucl. Acids Res. **14:**2287-2300.

19. **Hsu, P. L., W. Ross, and A. Landy.** 1980. The lambda phage att site: functional limits and interaction with Int protein. Nature **285:**85-91.

20. **Kitts, P. A., and H. A. Nash.** 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. J Mol Biol **204:**95-107.

21. **Lee, J., and M. Jayaram.** 1995. Role of Partner Homology in DNA Recombination. J. Biol. Chem. **270:**4042-4052.

22. **Lee, L., and P. D. Sadowski.** 2003. Sequence of the loxP Site Determines the Order of Strand Exchange by the Cre Recombinase. Journal of Molecular Biology **326:**397-412.

23. **Lee, L., and P. D. Sadowski.** 2005. Strand Selection by the Tyrosine Recombinases, p. 1-42, Progress in Nucleic Acid Research and Molecular Biology, vol. 80. Academic Press, New York.

24. **Macy, J. M., and I. Probst.** 1979. The biology of gastrointestinal bacteroides. Annu Rev Microbiol **33:**561-594.

25. **Malanowska, K., A. A. Salyers, and J. F. Gardner.** 2006. Characterization of a conjugative transposon integrase, IntDOT. Mol Microbiol **60:**1228-1240.

26. **Malanowska, K., S. Yoneji, A. A. Salyers, and J. F. Gardner.** 2007. CTnDOT integrase performs ordered homology-dependent and homology-independent strand exchanges. Nucleic Acids Res **35:**5861-5873.

27. **Martin, S. S., E. Pulido, V. C. Chu, T. S. Lechner, and E. P. Baldwin.** 2002. The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP Holliday junction complex. J Mol Biol **319:**107-127.

28. **McCulloch, R., L. W. Coggins, S. D. Colloms, and D. J. Sherratt.** 1994. Xer-mediated site-specific recombination at cer generates Holliday junctions in vivo. Embo J **13:**1844-1855.

29. **Nash, H. A.** 1983. Purification and properties of the bacteriophage lambda Int protein. Methods Enzymol **100:**210-216.

30. **Nunes-Duby, S. E., M. A. Azaro, and A. Landy.** 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. Curr Biol **5:**139-148.

31. **Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy.** 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res **26:**391-406.

32. **Nunes-Duby, S. E., D. Yu, and A. Landy.** 1997. Sensing homology at the strand-swapping step in lambda excisive recombination. J Mol Biol **272:**493-508.

33. **Patsey, R. L., and M. F. Bruist.** 1995. Characterization of the interaction between the lambda intasome and attB. J Mol Biol **252:**47-58.

34. **Rajeev, L., K. Malanowska, and J. F. Gardner.** 2009. Challenging a Paradigm: the Role of DNA Homology in Tyrosine Recombinase Reactions. Microbiology and Molecular Biology Reviews **73:**300-309.

35. **Recchia, G. D., M. Aroyo, D. Wolf, G. Blakely, and D. J. Sherratt.** 1999. FtsK-dependent and -independent pathways of Xer site-specific recombination. EMBO J **18:**5724-5734.

36. **Richet, E., P. Abcarian, and H. A. Nash.** 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell **52:**9-17.

37. **Ross, W., and A. Landy.** 1982. Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. Proc Natl Acad Sci U S A **79:**7724-7728.

38. **Ross, W., and A. Landy.** 1983. Patterns of lambda Int recognition in the regions of strand exchange. Cell **33:**261-272.

39. **Shoemaker, N. B., R. D. Barber, and A. A. Salyers.** 1989. Cloning and Characterization of a Bacteroides Conjugal Tetracycline-Erythromycin Resistance Element by Using a Shuttle Cosmid Vector. J Bacteriol **171:**1294-1302.

40. **Shoemaker, N. B., and A. A. Salyers.** 1990. A cryptic 65-kilobase-pair transposonlike element isolated from *Bacteroides uniformis* has homology with Bacteroides conjugal tetracycline resistance elements. J. Bacteriol. **172:**1694-1702.

41. **Shoemaker, N. B., H. Vlamakis, K. Hayes, and A. A. Salyers.** 2001. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. Appl Environ Microb **67:**561-568.

42. **Song, B., N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2007. Integration site selection by the Bacteroides conjugative transposon CTnBST. J Bacteriol **189:**6594-6601.

43. **Trask, D. K., J. A. DiDonato, and M. T. Muller.** 1984. Rapid detection and isolation of covalent DNA/protein complexes: application to topoisomerase I and II. The EMBO Journal **3:**671-676.

44. **Van Duyne, G. D.** 2002. A Structural View of Tyrosine Recombinase Site-specific Recombination. ASM Press, Washington, D.C.

45. **Warren, D., G. Laxmikanthan, and A. Landy.** 2008. A chimeric Cre recombinase with regulated directionality. PNAS **105:**18278-18283.

46. **Weisberg, R. A., L. W. Enquist, C. Foeller, and A. Landy.** 1983. Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. J Mol Biol **170:**319-342.

47. **Whittle, G., N. B. Shoemaker, and A. A. Salyers.** 2002. The role of Bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. Cell Mol Life Sci **59:**2044-2054.

Chapter 3

IntDOT Interactions with the Core-Type Sites During Integration

Introduction

Bacteroides spp. are Gram negative obligate anaerobes that comprise over 40% of the human gut microbiota (6). Although *Bacteroides* normally function as commensals, they can cause serious infections including abscesses and septicemia if released from their normal gut environment. These infections are becoming increasingly difficult to treat due to the high incidence of antibiotic resistance (8, 21, 23). By 2001 over 80% of *Bacteroides* strains were resistant to tetracycline (23). There is evidence that *Bacteroides* is involved in horizontal gene transfer with Gram positive pathogens such as *Clostridium perfringens, Streptococcus pneumoniae* and *Enterococcus faecalis,* which suggests that *Bacteroides* is acting as a reservoir for antibiotic resistance genes (23). Resistance genes are carried on mobile genetic elements such as conjugative transposons, which are often referred to as integrative and conjugative elements (ICEs). CTnDOT is an ICE found in *Bacteroides* spp. that carries resistance genes for both tetracycline (*tetQ*) and erythromycin (*ermF*). The presence of tetracycline actually induces CTnDOT excision and transfer by 1,000-10,000 fold (20).

The excision and integration of CTnDOT are catalyzed by the element-encoded integrase called IntDOT. IntDOT is a member of the tyrosine recombinase family, which includes lambda Int, XerC/D, Cre and Flp. There are two types of systems within this family: autonomous and factor-assisted. Cre and Flp are examples of autonomous recombinases, which do not require additional proteins to catalyze recombination (1, 20, 24). These proteins have core binding (CB) and catalytic (CAT) domains and only recognize one type of DNA site called core-type sites. Core-type sites directly flank the overlap region where cleavage, strand exchange and ligation occur. IntDOT and lambda Int are examples of factor-assisted recombinases which require additional proteins to form a higher order nucleoprotein complex called an intasome. Factor-assisted recombinases have three domains: the CB, CAT, and an N-terminal arm-binding domain (N) that binds to DNA sequences called arm-type sites which flank both sides of the core region. Formation of the intasome requires interactions between the recombinase and both core- and arm-type DNA sites. Additional host factors help facilitate these interactions (7, 17, 25-27).

The core-type sites recognized by IntDOT are B and B' in the bacterial attachment site *attB,* and D and D' in the *attDOT* site. There are multiple *attB* sites in the *Bacteroides* chromosome that have been identified (5). All of these sites share a conserved sequence in the B site - GTANNTTT. This sequence is also present in the D site of *attDOT*, which may suggest that this sequence is important for IntDOT recognition of the core-type sites. Of the six completely conserved base pairs, we wished to determine which are important for IntDOT

interactions with the DNA and ultimately make predictions that complement future crystal structures of IntDOT complexed with DNA.

Materials and Methods

Preparation of Radiolabeled DNA Substrates

Single stranded DNA oligonucleotides were obtained through Integrated DNA Technologies (IDT). One strand of each complementary pair was 5'-end labeled with [$\gamma-^{32}$P]-ATP (Perkin-Elmer) using T4 polynucleotide kinase (Fermentas). Excess [$\gamma-^{32}$P] was removed with Illustra Microspin G-25 columns (GE Healthcare). The oligonucleotides were annealed by combining the labeled strand and the complementary unlabeled strand at a 1:5 molar ratio in an annealing buffer consisting of 0.1 M KCl, 10 mM Tris-HCl (pH8), 5 mM EDTA. The annealing mixture was heated to 90°C for two minutes then slowly cooled to 25°C.

In vitro Integration Assay

The DNA substrates *attDOT* and *attB* were combined at a final concentration of 2 nM each in a 40 µl reaction volume that also contained 0.17 µM *E. coli* IHF, 1 unit of IntDOT, 30 mM Tris-HCl (pH 7.4), 5 mM DTT, 0.1 mg/ml tRNA, 0.07 mg/ml BSA, 2.6% glycerol and 50 mM KCl. A unit of IntDOT is defined as the minimal amount of IntDOT needed to produce maximum recombination between *attDOT* and *attB* (12, 15, 27). The reaction is known to be

very slow, so samples were incubated overnight at 37°C and quenched the following morning by adding 10 µl of a stop solution containing 30% glycerol, 10% SDS, 0.25% xylene cyanol, and 0.25% bromophenol blue. A 10 µl aliquot of each sample was electrophoresed on a 1% agarose gel at 100 V for 2 hours, then dried on a vacuum slab drier and exposed onto a phosphorimager screen. Recombination was quantified using a Fujifilm FLA-3000 phosphorimager and Fujifilm Image Gauge software (Macintosh v.3.4). To account for day-to-day variation in IntDOT activity, a standard reaction containing wild type *attDOT* and *attB1* was included in each experimental group.

Restriction Digest of the Recombinants

A standard *in vitro* integration assay was performed but was terminated by heating at 60°C for 20 minutes instead of by the addition of stop solution. $MgCl_2$ was added to a final concentration of 10 mM. A volume of 18 µl of the reaction was transferred to a new microcentrifuge tube, and 20 units of SspI (Fermentas) were added. The sample was incubated at 37°C for two hours then stopped by heating at 60°C for 20 minutes. The digested and non-digested samples were electrophoresed alongside each other on a 1% agarose gel at 100 V for 2 hours. Gels were analyzed as described above.

Results

Core Site Mutagenesis

The core sites are a pair of inverted repeats that directly flank the overlap region in both the *attDOT* and *attB* sites. In *attDOT*, the core sites are D and D', and in *attB* the core sites are B and B'. During integrative recombination, IntDOT is predicted to bind to these sites with the CB and CAT domains. By analogy to bacteriophage lambda and other tyrosine recombinase systems, it is likely that each of the four sites are bound by a single IntDOT monomer assembled in a nucleoprotein complex on *attDOT* called an intasome. The intasome undergoes synapsis with a naked *attB* site where strand exchange occurs. It is notable that all of the *attB* sites that have been identified to date share a nearly identical B core site sequence (GTANNTTT). This sequence is also conserved in the D core site of *attDOT* (GTANNTTT) (Figure 3.1) (5, 12). This conserved sequence includes the first two base pairs of the overlap region - the GC dinucleotide that provides the only region of required sequence identity within the overlap (Figure 3.1). Based on this highly conserved sequence, we predict that specific base pairs should be important for making direct or indirect contact with the protein, which would be required for the initiation and completion of recombination.

In order to test our hypothesis that specific base pairs in the core sequences are important for IntDOT binding, we first made mutations in the strongly conserved B sequence of *attB1* and measured their effects on integration. If we disrupt important contact positions, we should see a resulting

decrease in recombination levels. To start, we designed complementary *attB* oligonucleotides containing contiguous three base pair substitutions in the B core site and measured the effect of these changes in the *in vitro* integration assay. Recombination with wild type *attDOT* and *attB* substrates averages between 35-40% due to day-to-day variation in IntDOT activity. When base pairs at positions -6 to -4 of *attB* were changed to the complementary sequence (TTT to AAA), the recombination levels dropped to about 2%, suggesting important interactions had been disrupted. Mutation of the non-conserved base pairs at positions -9 to -7 resulted in a reduction of recombination by 10% but with a relatively large margin of error that suggests these bases are likely not important for binding. When base pairs at positions -11 to -9 are changed, recombination plummets to < 1%, which indicates that one or more of these positions is important for IntDOT interactions (Figure 3.2).

Similar sets of 3 bp substitutions were made in the B' site to determine if this less conserved core site has important sequence requirements. As shown in Figure 3.2, when positions +4 to +6 are changed to the complementary sequence (TAA to ATT), recombination levels average about 9%, which is roughly a tfourfold reduction. As shown in Figure 3.1, when positions +7 to +9 are changed, recombination levels average 15-20%. Finally, when positions +10 to +12 are changed, recombination averages 10%, again a threefold reduction relative to the wild type control. The reduction in recombination levels seen with the B' mutations is much less dramatic than that seen when mutations are in the B site. These results may suggest that the B site is key for IntDOT binding to *attB* during

synapsis with the integrative intasome, and may be important for initiating recombination at the site adjacent to the homology in the overlap region.

Analysis of the sequence of *attB1* showed an imperfect direct repeat of the B core site sequence just upstream of the B site. This is similar to the FRT site recognized by Flp recombinase. The FRT site contains a direct repeat upstream of the left core site that is not required for recombination (21). This repeat could potentially be acting as an alternative core site and might increase or decrease IntDOT binding for integration events. In order to determine what role, if any, this potential site may play in integration we disrupted three bases at a time and measured the effect in the *in vitro* integration assay (Figure 3.3). None of the mutations had any significant effect on recombination levels, indicating that the site is not required for integration.

Effects of Core Site Single Base Pair Mutations on Recombination

The results of the earlier mutagenesis experiments suggest that at least some of the conserved base pairs in the B and D sites are important for IntDOT recognition and binding. We wanted to identify the specific bases within this sequence that are likely to be important for IntDOT recognition of the sequence. To identify important bases, we made single base pair changes along the 8 bases of the 10 bp conserved sequence that are not part of the overlap region. Each base was changed to its complement, and the mutated substrate tested for recombination in the *in vitro* integration assay. The results showed that mutations at positions -5, -9 and -10 had the greatest effect on recombination (Figure 3.4).

Positions -7 and -8 are not conserved amongst the core sites, and mutations at these positions had no effect. We chose to analyze more substitutions at positions -5, -9 and -10 by changing the remaining two base pairs at those positions so that all four base pairs would be tested at each location.

We substituted adenine, cytosine, and guanine in place of the thymine that is normally present at position -10. As shown in Figure 3.5A, recombination levels with a T-A base pair at position -10 average 35%. When a G-C base pair is substituted at this position, recombination decreases by 10%. A-T or C-G base pairs at this position are not well tolerated and reduce recombination levels to 5-10%. Variants with substitutions at position -9 had minimal effects. The least effective base pair, C-G, only decreased recombination by 5% (Figure 3.5B). This leads us to conclude that *attB* sites with changes at position -9 do not disrupt favorable interactions with IntDOT, possibly suggesting interactions with the phosphate backbone instead of direct contact with the bases. In contrast, position -5 appears to be the most important base of the conserved sequence. All substitutions at this position result in a significant loss of recombination (Figure 3.5C) with the greatest effect seen when a G-C base pair is present (<1%) followed by a C-G base pair (2%). The dramatic results seen with substitutions at position -5 strongly suggest direct interactions between IntDOT and the T-A base pair at that position.

We confirmed the importance of the T-A base pair at position -5 by making mutations on *attDOT* at positions -4, -5, and -6. Single base mutations to the complementary sequence at positions -4 and -6 did not affect recombination

levels but the single base mutation (-5A) at position -5 reduced recombination threefold (Figure 3.6). Double (-4A, -5A and -5A, -6A) and triple (-4A, -5A, -6A) base mutations that include position -5 completely abolished recombination. This supports our earlier hypothesis that base pairs at position -5 in both *attDOT* and *attB* are interacting directly with IntDOT.

Effect of Core Site Mutations on Orientation of Integration

If the base pair at position -5 is important for IntDOT interactions on the B site, it is possible that there would be similarly important position interactions at the equivalent base pair in the B' site as well. Sequence alignment of the six known *attB* sites reveals that position +5 contains a conserved T-A base pair one base away from the bottom strand cleavage site. This location is spatially and chemically equivalent to the T-A base pair at position -5, yet triple mutations of the B' site containing a change at position +5 do not have the same effect as triple mutations containing a change at position -5. This could suggest that the primary recognition and binding sequence is contained in the B site. What role then does the B' site, specifically the conserved position +5 play in integration? We propose that the B' site, in conjunction with the location of homology within the overlap, may be important for a productive first strand exchange and allow recombination to proceed. Synapsis between the intasome and the *attB* can occur in one of two orientations, but the GC dinucleotides in the overlap regions must be properly aligned for a productive initial strand exchange. We have shown previously that the location of the conserved GC dinucleotide within the

overlap region dictates the orientation of the integrating *attDOT* by providing the homology required for the first strand exchange (12). In the same paper, we showed that an *attB* site containing a symmetric overlap, where the GC dinucleotide is present adjacent to both the top and bottom strand cleavage sites, produced recombinants in both orientations.

Using this symmetric overlap *attB* site as a template, we changed the conserved T-A base pair to G-C at positions -5, +5, and a double mutant containing G-C at both positions. We used SspI restriction digest of the recombinant product to determine the orientation of the integrated element. In a normal reaction between wild type *attDOT* and top strand labeled *attB* substrates, the restriction digest produces a 1.1 kb fragment. In a reaction between a wild type *attDOT* and an *attB* site containing a symmetric overlap region, there are two restriction fragments: 1.1 kb and 2.5 kb. The 2.5 kb fragment corresponds to an integration event that occurred in the inverted orientation. Figure 3.7 shows that reactions containing a wild type *attDOT* site and an *attB* site with a symmetric overlap containing a mutation at position -5 produce only a 2.5 kb restriction fragment. The same reaction containing a mutation at position +5 instead of position -5 produces only a 1.1 kb restriction fragment. When both positions -5 and +5 are mutated, no recombination occurs. These results suggest that the conserved T-A base pair on both the B and B' sites plays a role in orienting the intasome for productive recombination, possibly by interacting directly with IntDOT to position the catalytic tyrosine at the defined

cleavage sites. This explains how initial cleavage and strand exchange can occur in both orientations.

Base Analog Substitutions in *attB*

2-aminopurine

Base analogs are modified versions of the canonical bases. They can be used to introduce or remove functional groups into the major or minor groove of the DNA, which can potentially identify important contact positions of DNA binding proteins. We used base analogs to help determine whether IntDOT is interacting with the major groove or the minor groove to identify core binding sites. The base analog 2-aminopurine (2AP) is structurally similar to adenine and forms base pairs with thymine (Figure 3.8). The difference between adenine and 2AP is the location of the amino group in the DNA. In adenine, the amino group is located in the major groove, while in 2AP the amino group is located in the minor groove. We inserted 2AP where A is the normal base at position -5 in the bottom strand, at position -10 on the bottom strand, and at position -7 on the top strand to act as a control. Position -7 is not conserved, and single base mutations at that site have no effect on recombination. In all cases, the 2AP was paired to a T in the complementary strand.

Having a 2AP at position -10 did not significantly reduce recombination levels (25% compared to 25-30% in the wild type reaction) (Figure 3.8). In contrast, the 2AP at position -7 actually increased recombination levels by twofold to 55%. Sequence comparisons show that most *attB* sites contain either

an A or a T at position -7. Perhaps the presence of the base analog produced more favorable interactions between IntDOT at that location than a canonical A-T base pair. The 2AP at position -5 also had a significant effect on recombination levels, reducing them threefold to less than 9% (Figure 3.8). These results by themselves do not show whether the loss in recombination is due to the loss of the amino group from the major groove or the addition of the amino group to the minor groove.

2,6-diaminopurine

To further investigate whether the decrease in recombination seen with 2AP at position -5 was possibly due to the loss of a functional group from the major groove or the gain of a functional group in the minor groove, we used another base analog: 2,6-diaminopurine (DAP). This base analog contains amino groups in both the major and the minor grooves and pairs with thymine. The T/DAP base pairing also resembles a G/C base pair with three sets of hydrogen bonds instead of two (Figure 3.9). Together with the 2AP results, the DAP should help distinguish between the major and minor groove interactions because it structurally resembles adenine but has an additional amino group in the minor groove. We designed *attB* sites with DAP on the bottom strand of position -5, the top strand at position -7, and at the bottom strand of position -10.

Having a DAP at position -5 kept recombination near wild type levels (Figure 3.9), with recombination at 15-20%. Recombination levels with an *attB* containing a DAP at positions -7 and -10 ranged from 20-25%. This suggests that

the presence of the amino group in the major groove is important for binding of IntDOT and implies that there is direct contact between the protein and the DNA at this position.

Discussion

Interactions with both core- and arm-type sites are necessary for formation of the catalytically active intasome complex (14, 17-19). We are particularly interested in the intasome formed by IntDOT, because it is able to tolerate extensive heterology between the overlap regions of the *attDOT* and *attB* DNA substrates. How IntDOT interacts with the core sites may provide some insight as to the differences between the DOT intasome and the intasomes formed by other tyrosine recombinases. The first notable difference is that IntDOT has a relatively high affinity for the core sites and will shift them independently of the arm-type sites, although affinity is enhanced in the presence of the arm-type sites (27). This is in contrast to lambda Int, which has a higher affinity for the arm-type sites and a low affinity for the core-type sites (3, 11, 13). It is also notable that CTnDOT integrates site-selectively into multiple sites in the *Bacteroides* chromosome. Each of these sites contains a conserved sequence on the B core site that is shared on the D core site of *attDOT.* We suspect that the conserved sequence is important for IntDOT recognition of, and binding to, target *attB* sites. This idea is supported by the poor conservation of the B' sequences.

Many sequence-specific binding proteins recognize their target sequences by inserting an α-helix into the major groove of DNA. In doing so, specific residues on the helix can interact with exposed hydrogen bonding groups in the major groove (16). One of the goals of this paper was to identify the specific bases important for contact with IntDOT. We did this using site-directed mutational analysis to identify the important bases, followed by incorporation of the base analogs 2-aminopurine and 2,6-diaminopurine to identify the functional groups in the major groove that are likely recognized by IntDOT.

Triple mutations that disrupt positions -9, -10, and -11 or positions -4, -5, and -6 abolish recombination, supporting our hypothesis that the conserved sequence is important for IntDOT recognition and binding. Having other base pairs at those positions produces unfavorable interactions between the DNA and IntDOT that prevents recombination from occurring at normal levels.

At position -10, there is a conserved thymine present in the wild type sequence. When the three other bases are substituted at this location, only the guanine is tolerated. Since thymine and guanine both contain a carboxyl group in the major groove, this functional group may provide favorable interactions with IntDOT that allows for productive binding and recombination. At position -9, there is a conserved adenine, but the other three bases are also well-tolerated at this position. Since the substitutions neither disrupt, nor improve interactions between IntDOT and the DNA at this position, we predict that IntDOT is either making contact with the phosphate backbone or making water-mediated interactions with the base pair at this position.

The T-A base pair normally present at position -5 is the only base pair that works, as substitutions at this position reduce recombination from undetectable to about 50% of wild type levels depending on the base pair (Figure 3.5C). This appears to be the single most important position of the entire conserved sequence; with base analog experiments suggesting an important role for the amino group of adenine in the major groove. The experiments with the *attB* site containing a symmetric overlap region support the idea that the T-A base pair at position -5 is important for integration. Mutations changing this position to a G-C base pair abolish recombination in the normal orientation, but still allow recombination in the inverted orientation, suggesting that the T-A base pair located at position +5 on the B' site can compensate, as long as homology exists within the overlap region adjacent to the B' cleavage site. Similarly, mutation at position +5 prevented integration in the inverted orientation. This suggests that strong interactions between IntDOT and positions +5 and -5 occur during synapsis and help to guide the intasome into the proper orientation for productive recombination.

This analysis of IntDOT interactions with core-type site DNA will complement future crystallization studies. We predict at least one direct contact on each core site consistent with the presence of conserved base pairs. The crystal structure will show which bases are in contact with IntDOT and whether these interactions are direct, water mediated, or protein interactions with the phosphate backbone. Crystallization studies done with Flp showed that most of the contacts made between the protein and the DNA are through the phosphate

backbone rather than the bases (4). The direct contacts that Flp does make with the FRT DNA site appear to be with the T-A base pair at position -13 and the C-G base pair at position -11 (4). Both of these base pairs are conserved across both core sites. When position -11 was changed, recombination levels were severely affected (22). Guo et al. showed that most interactions between Cre and *loxP* were water-mediated (9). Studies with Cre showed that there is only one direct contacts between the protein and the major groove of the DNA – a C-G base pair at position 10 which is conserved in both core sites (10). The same study also showed that Cre makes some contacts with bases via the minor groove at positions -16 and -17 (10). Crystallographic studies done with lambda Int showed numerous contacts with the phosphate backbone, but few direct interactions with the DNA bases (2). There are some van der Waals interactions between the C5 methyl group of 4 thymines and two residues directly contacting base pairs that are conserved among the four core-type sites (2). These limited interactions may help explain why lambda Int has such a low affinity for its core-type sites. The stronger interactions seen with IntDOT and its core-type sites more closely resembles the interactions between Flp and it's target FRT, and Cre and it's target *loxP*. These results may help us to understand the role of the core-type sites in intasome formation and perhaps even how *attB* sites are recognized in host cells, since we presume *attB* sites likely consist of only core-type IntDOT binding sites.

Figures

D: **GTAACTTT**

B1: **GTAAATTT**

B2: **GTATTTTT**

B3: **GTACTTTT**

B4: **GTAGATTT**

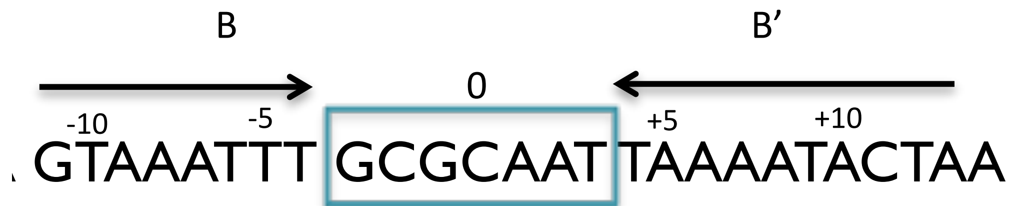
B5: **GTTGTTTT**

B6: **GTATATTT**

Consensus:

GTANWTTT

Figure 3.1: Alignment of the D site in *attDOT* and the B sites of all the known *attBs* the consensus sequence is shown underlined at the bottom.



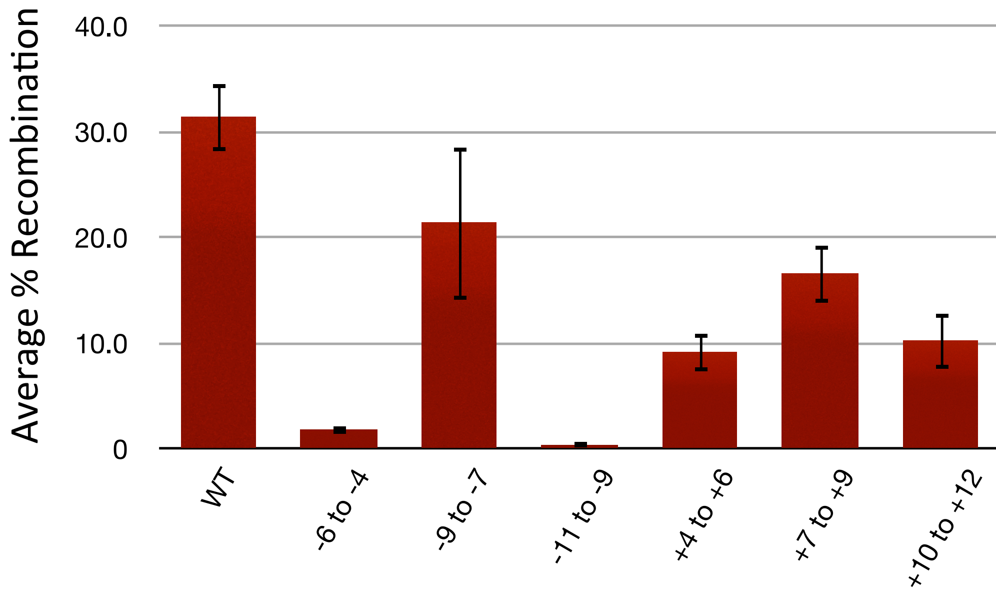
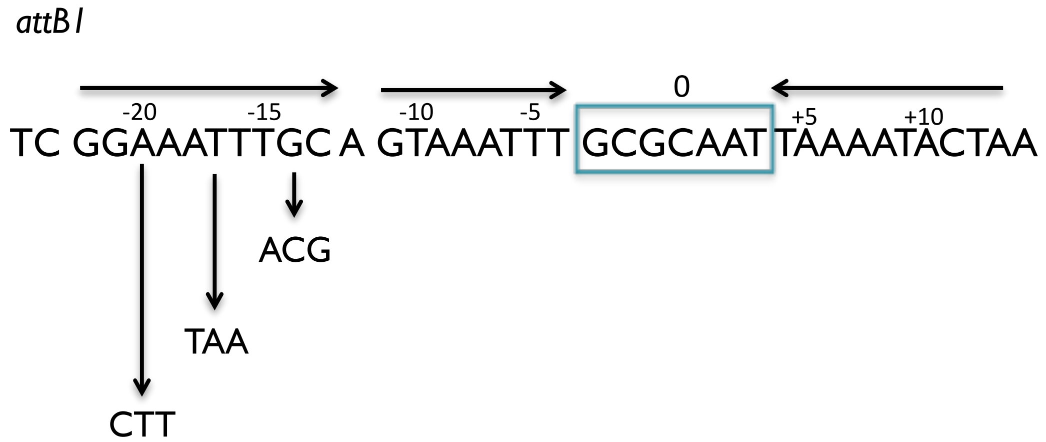


Figure 3.2: Triple mutations in the B and B' core type site changing each position to its complement. Mutations involving the six conserved bases nearly abolish detectable recombination.

A.



B.

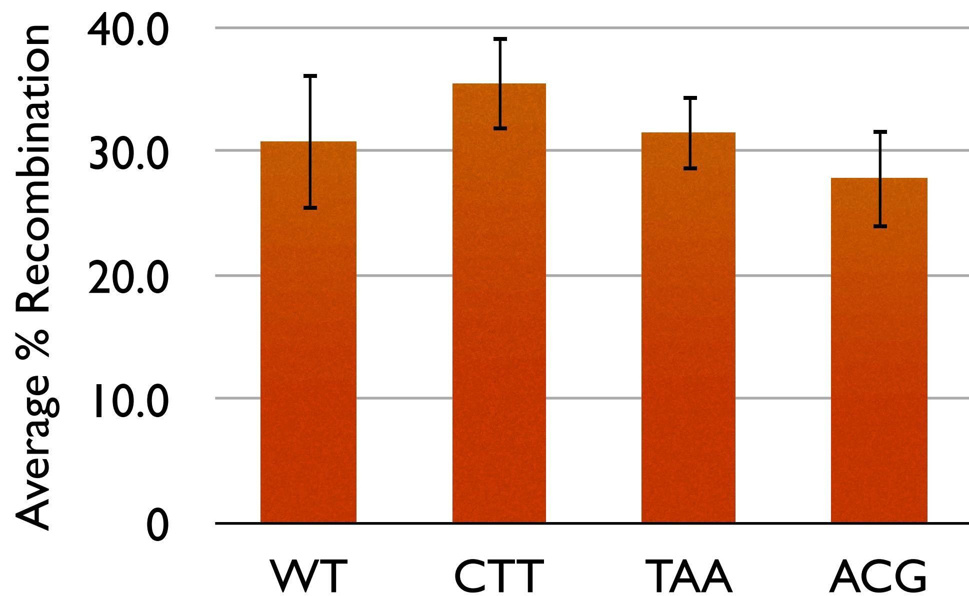


Figure 3.3: A) There is a direct repeat of the 10 bp conserved sequence just upstream of the B site that could potentially act as an IntDOT binding site. We made triple mutations in the sequence and measured the effect on integration. B) Results from integration assays done using substrates containing the upstream mutations.

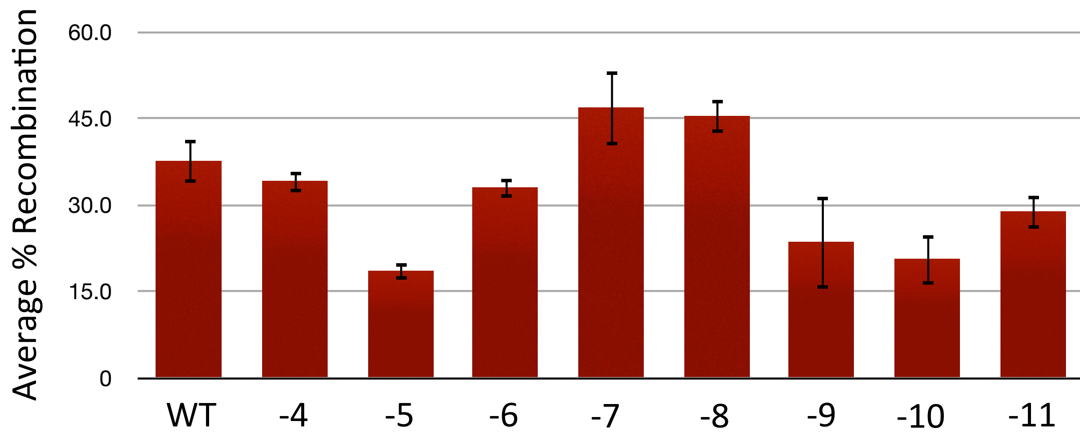


Figure 3.4: Integration assay results of single base mutations along the B core type site. Each base was changed to its complement. Similar experiments done on the D site of *attDOT* showed a comparable recombination profile.

-11 -10 -9 -8 -7 -6 -5 -4

G T A N N T T T

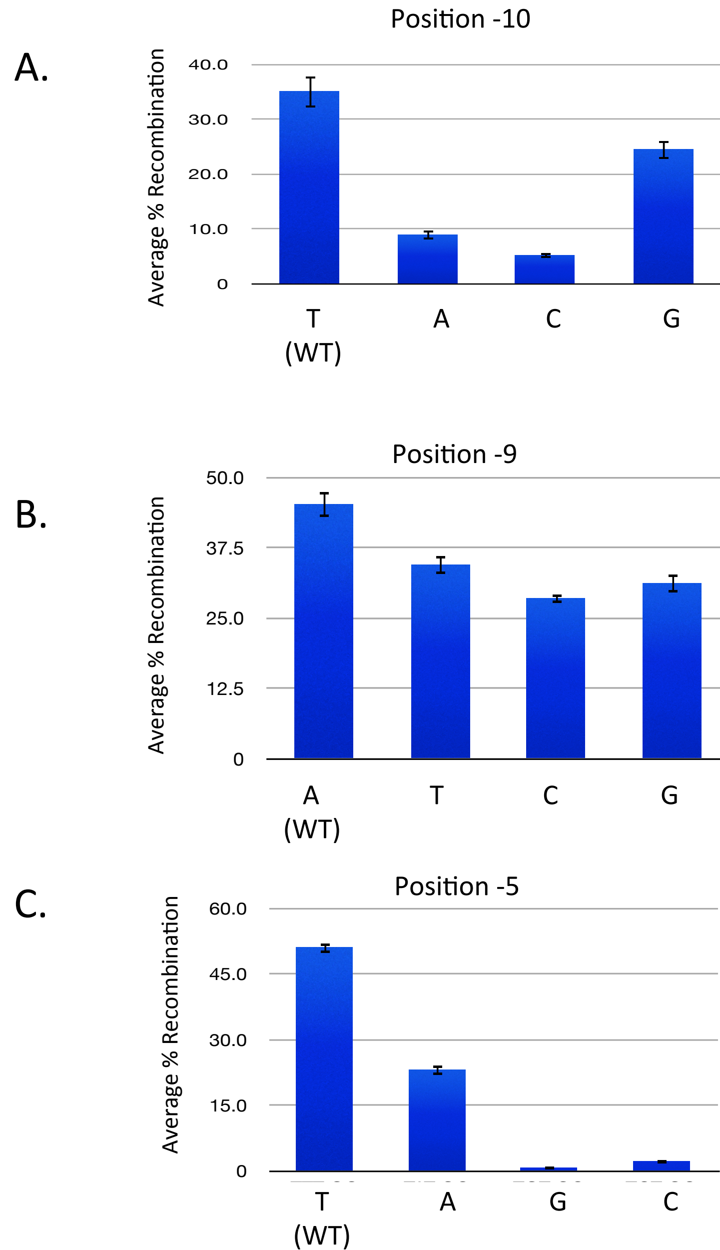


Figure 3.5: Integration assay results of single base substitutions at positions -10, -9 and -5.

-11 -10 -9 -8 -7 -6 -5 -4

GTANNTTT

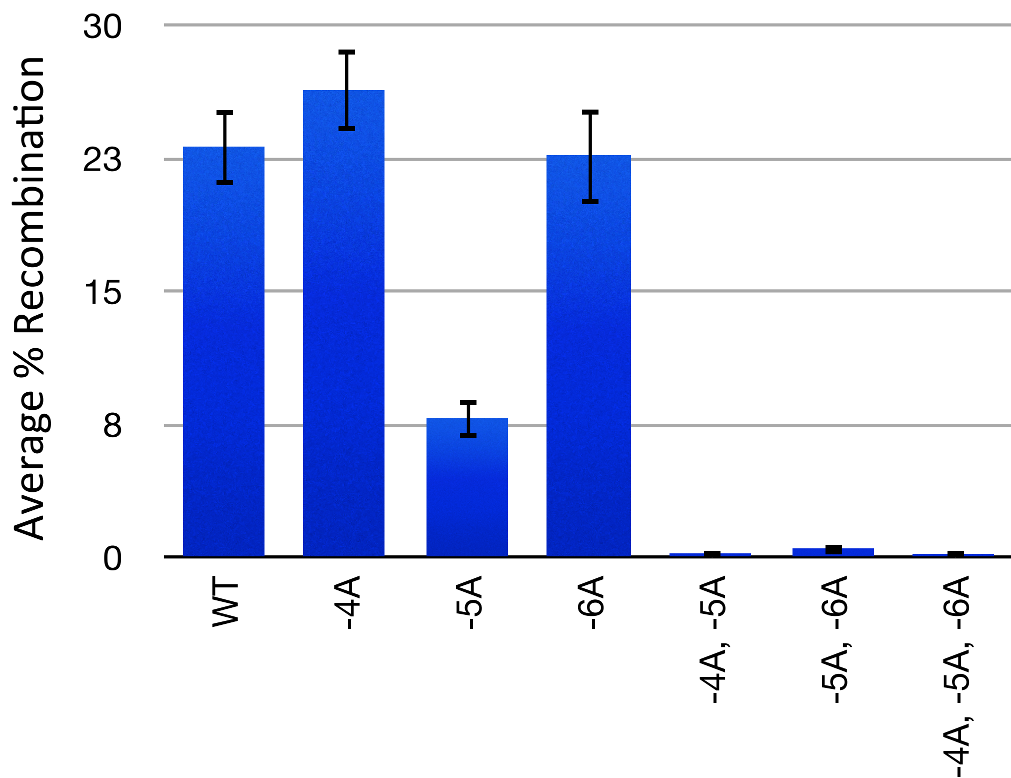
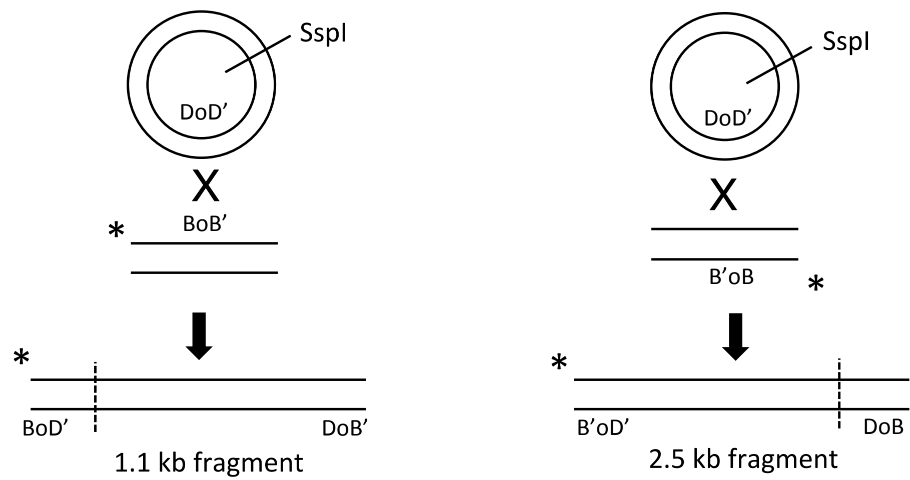


Figure 3.6: Integration assay results of mutations in the D core site of *attDOT.* We made a triple mutation changing each base to its complement (-4A, -5A, -6A), double mutations containing positions -4 and -5, or -5 and -6, and finally single mutations at each of the three positions.



attB	Wild Type		Symmetric Overlap		T-5G		T+5G		T-5G, T+5G	
SspI	-	+	-	+	-	+	-	+	-	+

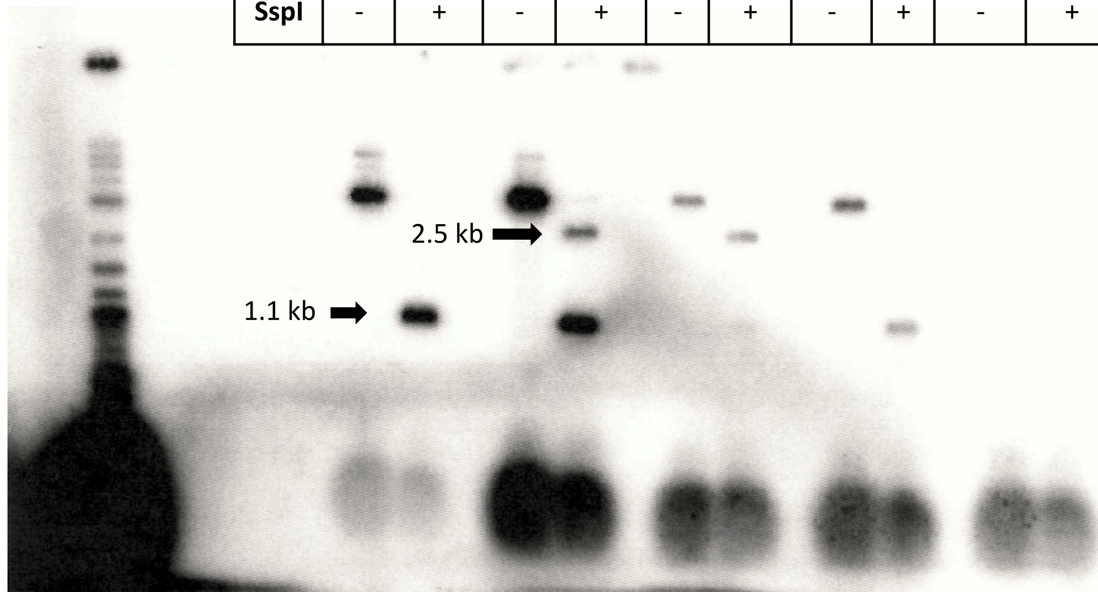


Figure 3.7: Integration assay and restriction digest results of substitutions at positions -5 and +5 of an *attB* site containing a symmetric overlap region. A symmetric overlap region allows integration to occur in both orientations by having a GC dinucleotide at both ends of the overlap. The integrative intasome can synapse with the *attB* site in one of two orientations. This is demonstrated by the different restriction fragments produced from an SspI digestions. Substitutions at positions -5 or +5 inhibit recombination events from a specific synaptic orientation.

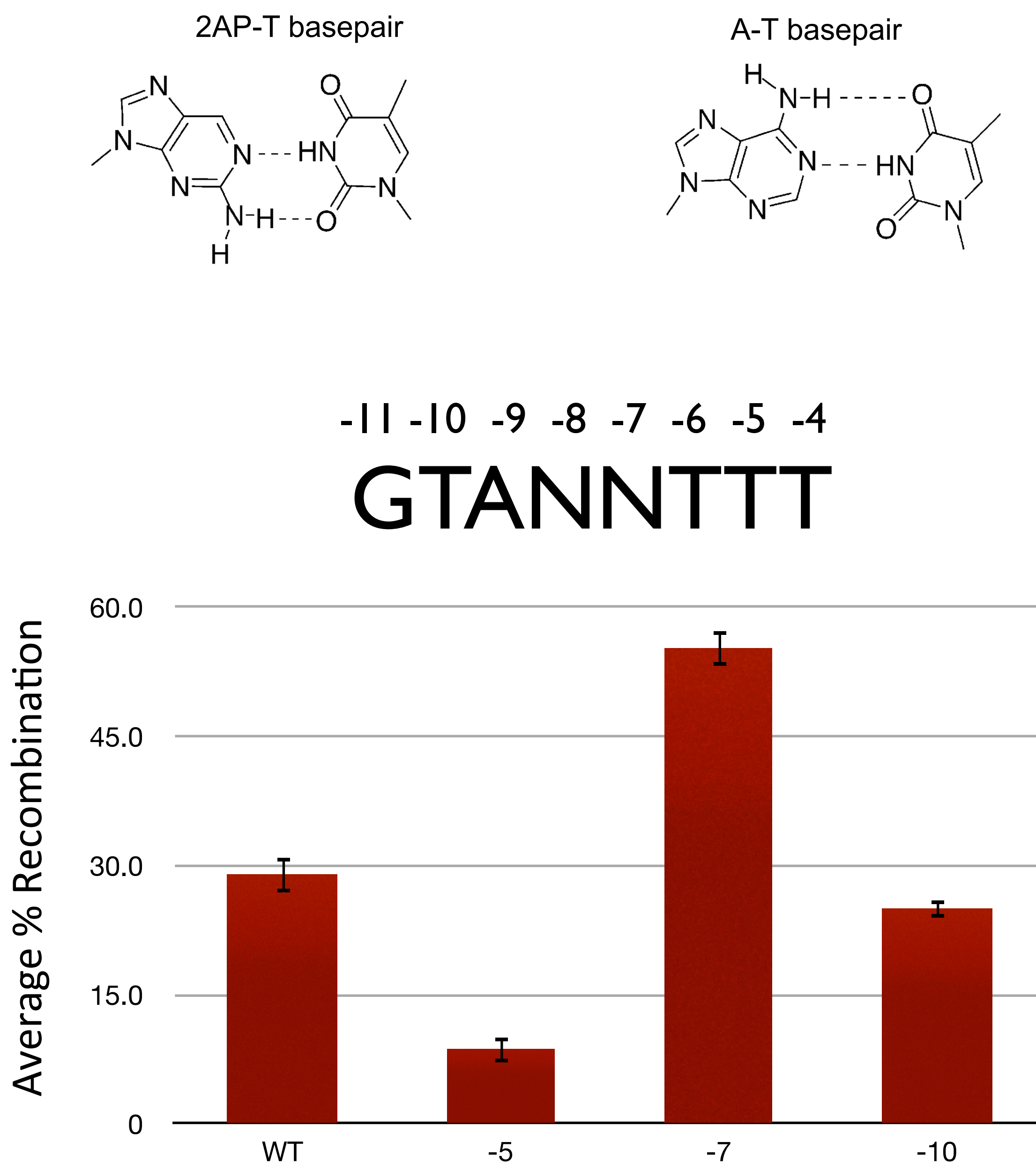
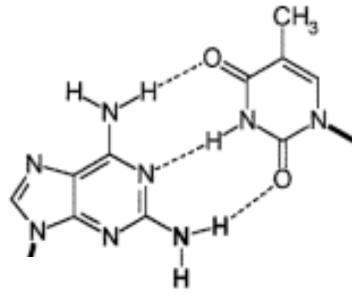
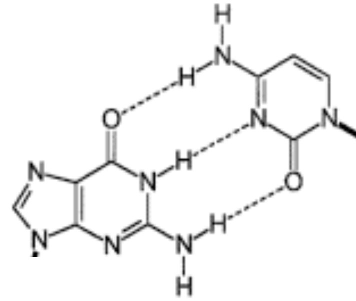


Figure 3.8: Integration assay results of 2-aminopurine substitutions at position -5, -7, and -10. The 2-aminopurine is on the bottom strand at positions -5 and -10 and on the top strand at position -7 so that there is always a 2AP-T base pairing.

DAP-T basepair



G-C basepair



-11 -10 -9 -8 -7 -6 -5 -4

GTANNTTT

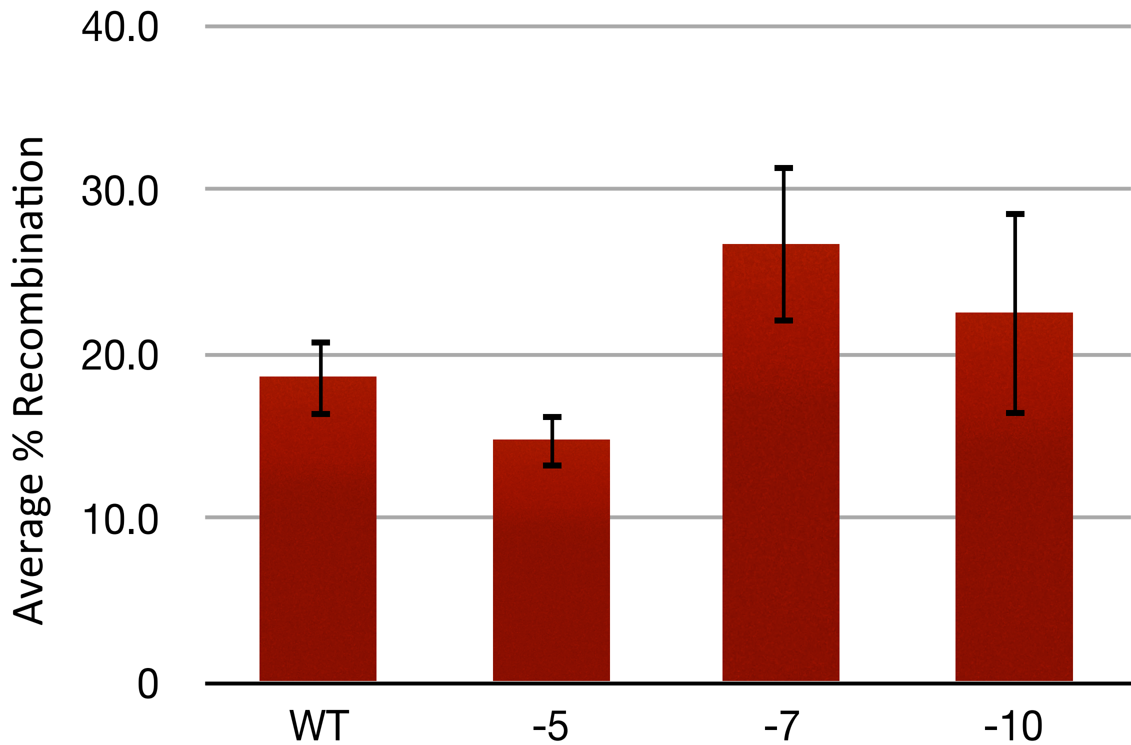


Figure 3.9: Integration assay results of substrates containing diaminopurine (DAP) substitutions at positions -5, -7, -10. Substitutions at positions -5 and -10 are on the bottom strand, while substitution at -7 is on the top strand. In all cases, DAP is paired to a thymine.

References

1. **Abremski, K., and R. Hoess.** 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J Biol Chem **259:**1509-1514.

2. **Aihara, H., H. J. Kwon, S. E. Nunes-Duby, A. Landy, and T. Ellenberger.** 2003. A conformational switch controls the DNA cleavage activity of lambda integrase. Mol Cell **12:**187-198.

3. **Azaro, M. A., and A. Landy.** 1997. The isomeric preference of Holliday junctions influences resolution bias by lambda integrase. EMBO J **16:**3744-3755.

4. **Chen, Y., U. Narendra, L. E. Iype, M. M. Cox, and P. A. Rice.** 2000. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol Cell **6:**885-897.

5. **Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2000. Integration and excision of a Bacteroides conjugative transposon, CTnDOT. J Bacteriol **182:**4035-4043.

6. **Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight.** 2009. Bacterial community variation in human body habitats across space and time. Science **326:**1694-1697.

7. **Dichiara, J. M., A. N. Mattis, and J. F. Gardner.** 2007. IntDOT interactions with core- and arm-type sites of the conjugative transposon CTnDOT. J Bacteriol **189:**2692-2701.

8. **Falagas, M. E., and E. Siakavellas.** 2000. *Bacteroides*, *Prevotella*, and *Porphyromonas* species: a review of antibiotic resistance and therapeutic options. Int J Antimicrob Agents **15:**1-9.

9. **Guo, F., D. N. Gopaul, and G. D. Van Duyne.** 1999. Asymmetric DNA bending in the Cre-*loxP* site-specific recombination synapse. Proc Natl Acad Sci U S A **96:**7143-7148.

10. **Guo, F., D. N. Gopaul, and G. D. van Duyne.** 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature **389:**40-46.

11. **Kim, S., and A. Landy.** 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci attL and attR. Science **256:**198-203.

12. **Laprise, J., S. Yoneji, and J. F. Gardner.** 2010. Homology-dependent interactions determine the order of strand exchange by IntDOT recombinase. Nucleic Acids Res **38:**958-969.

13. **MacWilliams, M., R. I. Gumport, and J. F. Gardner.** 1997. Mutational analysis of protein binding sites involved in formation of the bacteriophage lambda attL complex. J Bacteriol **179:**1059-1067.

14. **Moitoso de Vargas, L., C. A. Pargellis, N. M. Hasan, E. W. Bushman, and A. Landy.** 1988. Autonomous DNA binding domains of lambda integrase recognize two different sequence families. Cell **54:**923-929.

15. **Nash, H. A.** 1983. Purification and properties of the bacteriophage lambda Int protein. Methods Enzymol **100:**210-216.

16. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu Rev Biochem **53:**293-321.

17. **Ross, W., and A. Landy.** 1982. Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. Proc Natl Acad Sci U S A **79:**7724-7728.

18. **Ross, W., and A. Landy.** 1983. Patterns of lambda Int recognition in the regions of strand exchange. Cell **33:**261-272.

19. **Ross, W., A. Landy, Y. Kikuchi, and H. Nash.** 1979. Interaction of Int protein with specific sites on lambda att DNA. Cell **18:**297-307.

20. **Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L. Y. Li.** 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol Rev **59:**579-590.

21. **Senecoff, J. F., R. C. Bruckner, and M. M. Cox.** 1985. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proc Natl Acad Sci U S A **82:**7270-7274.

22. **Senecoff, J. F., P. J. Rossmeissl, and M. M. Cox.** 1988. DNA recognition by the FLP recombinase of the yeast 2 mu plasmid. A mutational analysis of the FLP binding site. Journal of molecular biology **201:**405-421.

23. **Shoemaker, N. B., H. Vlamakis, K. Hayes, and A. A. Salyers.** 2001. Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. Appl Environ Microb **67:**561-568.

24. **Sternberg, N., and D. Hamilton.** 1981. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol **150:**467-486.

25. **Tirumalai, R. S., H. J. Kwon, E. H. Cardente, T. Ellenberger, and A. Landy.** 1998. Recognition of core-type DNA sites by lambda integrase. J Mol Biol **279:**513-527.

26. **Wojciak, J. M., D. Sarkar, A. Landy, and R. T. Clubb.** 2002. Arm-site binding by lambda -integrase: solution structure and functional characterization of its amino-terminal domain. Proc Natl Acad Sci U S A **99:**3434-3439.

27. **Wood, M. M., J. M. Dichiara, S. Yoneji, and J. F. Gardner.** 2010. CTnDOT integrase interactions with attachment site DNA and control of directionality of the recombination reaction. J Bacteriol **192:**3934-3943.

Chapter 4

Attempts to Trap Holliday Junction Intermediates

Introduction

Tyrosine recombinases catalyze recombination by a series of sequential strand exchange and ligation steps, forming a Holliday junction (HJ) intermediate between the first and second strand exchanges. In most systems, the region of strand exchange must have identical sequences between the two recombining DNA sites. In the case of lambda Int, the presence of a single base of heterology is sufficient to block recombination (1). IntDOT differs from other known tyrosine recombinases in that it only requires two bases of sequence identity between the two regions of strand exchange, or overlap regions (Figure 4.1A). Recombination by IntDOT occurs via an initial homology-dependent strand exchange followed by a second strand exchange that is homology-independent (10). How IntDOT is able to resolve a HJ intermediate containing five bases of heterology is still not understood. However, IntDOT must be able to perform ligation reactions where the base to be ligated forms a mismatch with the base of the complementary strand.

We are interested in isolating the HJ intermediate formed during IntDOT mediated recombination in order to establish how IntDOT is able to resolve these structures despite the presence of heterology. We used short peptides developed by the Segall laboratory that have been shown to bind HJs and prevent

resolution in either direction. These peptides have been used successfully to trap HJ intermediates formed by several tyrosine recombinases, including lambda and XerC/D (5). We also used *attB* sites containing a non-bridging phosphorothioate at the top or bottom strand cleavage sites in an attempt to trap the HJ intermediate by preventing IntDOT from performing ligation.

Materials and Methods

Peptides

The peptides were generously provided by Anca Segall at San Diego State University. They arrived dissolved in 100% DMSO with various final concentrations. I made 1 mM dilution stocks from the originals in 10% DMSO and water. The peptides were kept at 4° to avoid multiple thaws. Each of the peptides was used in the *in vitro* integration assay at concentrations ranging from 0.1 μM to 2.5 μM. The four peptides are as follows: D-form dodecamer (dodec) WRWYRGGRYWRW, L-form non-palindromic repeat (non-pal) WRWYRGWRWYRG, L-form W1A substitution (W1A) ARWYRGGRYWRW, and L-form W12A substitution (W12A) WRWYRGGRYWRA.

Phosphorothioates

Oligonucleotides of the *attB* site containing non-bridging phosphorothioate substitutions at either the top strand or bottom strand cleavage sites were obtained through Thermo-Fisher. They were reconstituted according to the

manufacturer and annealed to the partner strand that did not contain a phosphorothioate. These phosphorothioate *attB* substrates were radiolabeled and used in the *in vitro* integration assay.

Preparation of Radiolabeled *attB* substrates

One strand of the *attB* DNA oligonucleotide (IDT) was 5' end-labeled with γ-^{32}P-ATP (Perkin-Elmer) using T4 polynucleotide kinase (Fermentas), and the free γ-^{32}P was removed using G-25 spin columns (Amersham Biosciences). The labeled DNA was mixed with the unlabeled complementary strand at a 1:5 molar ratio and annealed in an annealing buffer (0.1M KCl, 10mM Tris-HCl pH 8, 5mM EDTA) by heating to 90°C for two minutes followed by slow cooling to 25°C.

***In Vitro* Recombination Assay**

The *attDOT* and *attB* substrates were incubated in a 20 μl reaction volume containing 0.17 μM *E. coli* IHF, 1 unit of IntDOT, 30 mM Tris-HCl (pH 7.4), 5 mM DTT, 0.1 mg/ml tRNA, 0.07 mg/ml BSA, 2.6% glycerol and 50 mM KCl. The final concentrations of *attDOT* and *attB* were 2 nM as determined by O.D.$_{260}$. A unit of IntDOT is defined as the minimum amount of IntDOT needed to produce maximum recombination between *attDOT* and *attB* (11). The reaction was shown to proceed over 16 hours, so samples were incubated overnight at 37°C and the reaction quenched with the addition of 5 μl stop solution (30% glycerol, 10% SDS, 0.25% xylene cyanol, and 0.25% bromophenol blue). All samples were subjected to electrophoresis on a 1% agarose gel. Gels were dried on a vacuum

slab drier, then exposed to phosphorimager screens and the recombination efficiency quantified using a Fujifilm FLA-3000 phosphorimager and Fujifilm Image Gauge software (Macintosh v.3.4).

Restriction Digest Analysis of the Recombinant Products

Some experiments were done where products were cleaved by SspI endonuclease. A double volume of the standard recombination assay described above was performed, but the reaction was terminated by heating for 20 minutes at 60°C instead of by the addition of stop solution in order to perform the restriction digest. Magnesium chloride was added to a final concentration of 10 mM. Half the reaction volume was transferred to a fresh microcentrifuge tube and digested with SspI endonuclease (Fermentas). The digest was stopped by heating at 60°C for 20 minutes and the sample was subjected to electrophoresis on a 1% agarose gel at 100 V for 2 hours and analyzed as described above.

Results

Peptides Do Not Trap Holliday Junctions Made by IntDOT

We tested these peptides in the *in vitro* integration assay at concentrations ranging from 0.1-2.5 uM and did not detect HJ accumulation. In addition, the peptides did not reduce the integration efficiency of the wild type reaction (Figure 4.1B). We also tested the peptides in the *in vitro* integration assay with one of the substrates containing an inverted overlap to see if perhaps this reaction adopted

a different conformation that would allow the peptides to be effective. The inverted overlap *attB* substrate contains an overlap region that has been rotated 180 degrees such that the initial cleavage takes place on the bottom strand adjacent to the B' site instead of on the top strand adjacent to the B site (8). These experiments were also unsuccessful.

We then considered that the systems in which the peptides had been shown to accumulate HJs all require complete sequence identity within the region of strand exchange, or overlap region. IntDOT is different from these systems in that only two bases of sequence identity are required and five bases of heterology exist between the wild type substrates. It's possible that, by creating substrates with homologous overlap regions, our system would more closely resemble the other systems in which the peptides are effective. We created an *attB* site containing the same overlap sequence of *attDOT*, thus creating homologous overlap sequences. We used these substrates to test the four peptides again in the *in vitro* integration assay. Unfortunately, no HJs were detected. The plasmid containing the *attDOT* site also contains a single, asymmetric SspI restriction site that will produce fragments of 1.1 kb and 2.5 kb from a 3.6 kb recombinant. If an HJ were present, we would expect to see a different sized fragment corresponding to the χ–structure. Restriction digest analysis with SspI confirmed that no HJ intermediates were present in the assay.

Non-Bridging Phosphorothioates May Trap HJs Formed by IntDOT

A second method we used to try and trap HJ intermediates used non-bridging phosphorothioates at the sites of cleavage. Bridging and non-bridging phosphorothioates have previously been used to trap HJs formed by lambda Int (2, 3, 7). Non-bridging phosphorothioates contain a sulfur in place of a non-bridging oxygen. We designed *attB* oligonucleotides containing a phosphorothioate at either the top strand cleavage site or at the bottom strand cleavage site. Each strand was annealed to the wild type complementary strand so that only one phosphorothioate was present in each *attB* site. We used these phosphorothioate *attB*s as substrates in the *in vitro* integration assay with wild type *attDOT*. We found that recombination levels from the phosphorothioate-containing substrates were roughly half that of the wild type substrates (Figure 4.2). We attribute this to the fact that the phosphorothioates are a racemic mix of R and S isomers and it is likely that IntDOT can only utilize one isomer. Additional analysis of the phosphorothioate recombinants by SspI restriction digest and two dimensional gel electrophoresis showed that no HJs had been accumulated.

Discussion

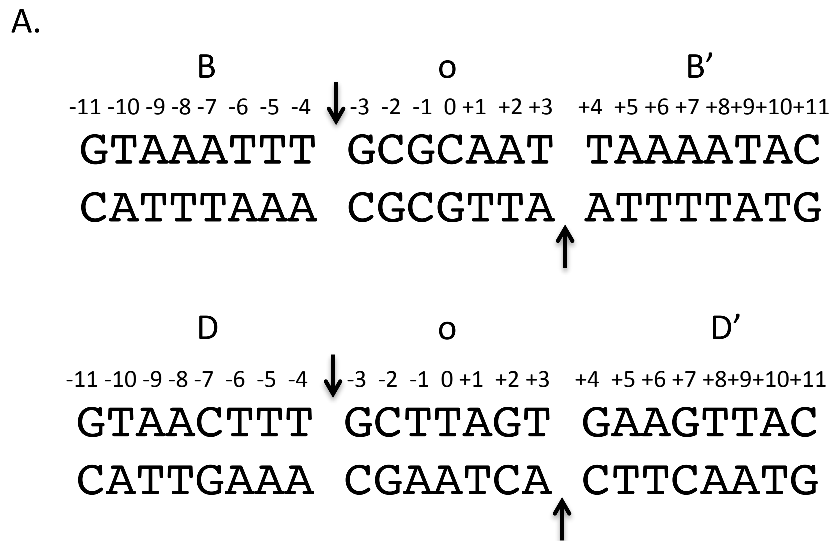
The peptides have been used by the Segall lab at San Diego State University to trap the HJs formed by other tyrosine recombinases including lambda Int, XerC/D and Cre (4, 5). A key conserved feature of the aforementioned recombinases is that they all require sequence homology between the overlap regions of their respective DNA substrates. The homology is

required for efficient strand joining and ligation during the formation of the HJ intermediate (9, 12). In the lambda system, only one base of heterology in the overlap region is enough to inhibit recombination (1). IntDOT is unusual among the tyrosine recombinases because it requires only two bases of homology at the site of initial cleavage (8, 10). Work done with IntDOT by Kim, et. al. showed that the arm-type sites and a host factor were required to resolve a synthetic HJ containing the wild type *attDOT* and *attB* substrates. In the absence of intasome formation, IntDOT would resolve the HJ to substrates rather than to the *attL* and *attR* products (6). This suggests that there is something inherently different about the structure of the IntDOT intasome and the HJ formed by it that allows resolution despite the presence of heterology. The results from the peptide assays and the phosphorothioates support this idea.

The results from the phosphorothioate experiments are somewhat more encouraging. The 50% reduction in recombination is consistent with one isomer of a racemic mixture being able to block ligation by IntDOT. Unfortunately at this time single isomer phosphorothioates are not commercially available, and we don't know which of the two isomers is responsible for the blocking activity. Until single isomer phosphorothioates are commercially available, or until there is an efficient method of producing them ourselves, further attempts to isolate HJs produced by IntDOT will have to be postponed.

Figures

A.



B.

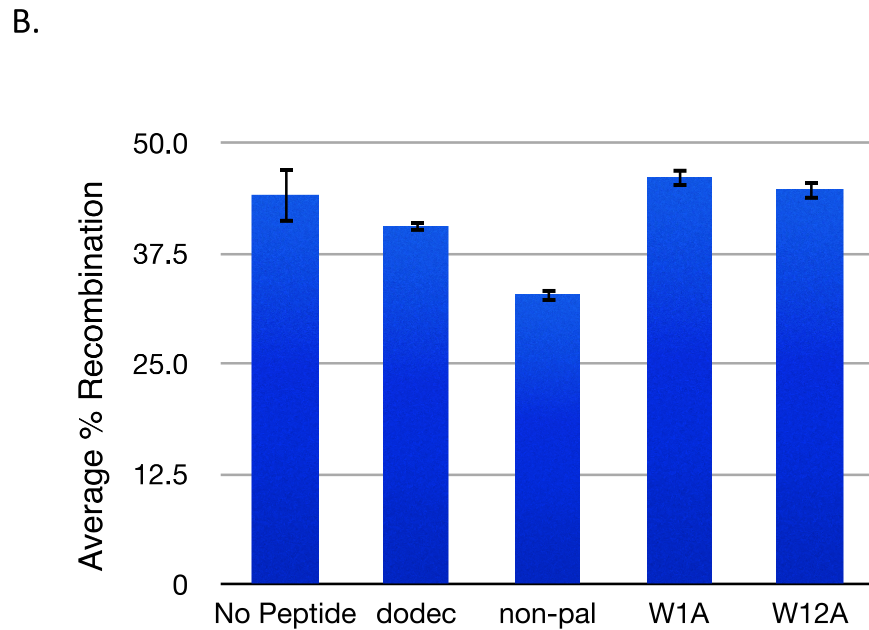


Figure 4.1: A) Core site sequence of the bacterial attachment site *attB1* and of *attDOT*. The vertical arrows indicate sites of cleavage by IntDOT. The B core site sequence is conserved throughout all known *attB* sites as well as within the D core site of *attDOT*. B) Results from an integration assay containing wild type *attDOT* and *attB* substrates and one of four peptides designed to trap the Holliday junction intermediate. The sequence of the peptides are WRWYRGGRYWRW (dodec), WRWYRGWRWYRG (non-pal), ARWYRGGRYWRW (W1A), and WRWYRGGRYWRA (W12A).

Reaction	Wild Type	Wild Type		PTO	PTO
Strand Labeled	Top	Bottom		Top	Bottom
Avg % Recombination	26%	26%		10%	9%

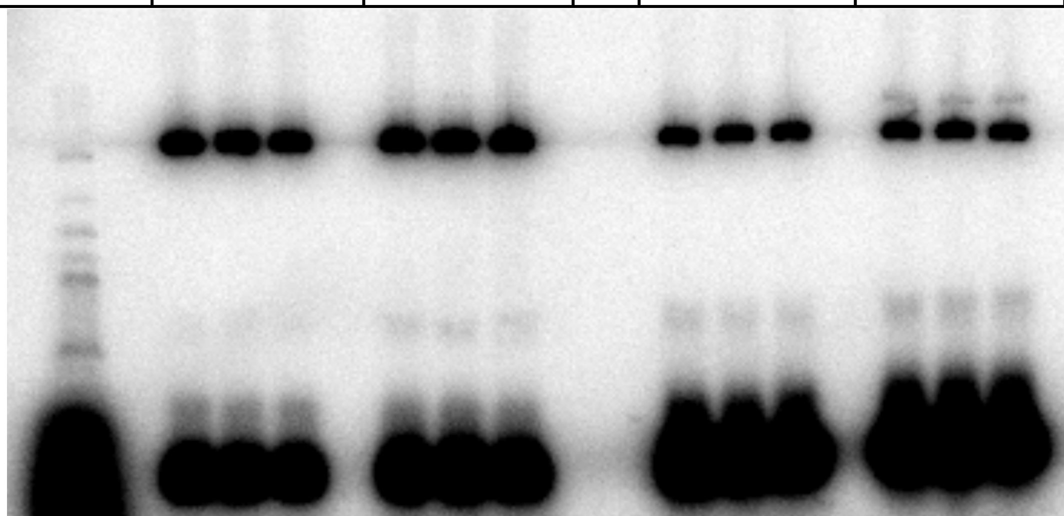


Figure 4.2: Results from an integration assay containing either wild type substrates or substrates containing a non-bridging phosphorothioate (PTO) at either the top strand cleavage site at position -4, or at the bottom strand cleavage site at position +4. The substrates containing the PTO were a racemic mix and we believe IntDOT can only utilize one isomer, resulting in about half the recombination level of the wild type reaction.

References

1. **Bauer, C. E., J. F. Gardner, and R. I. Gumport.** 1985. Extent of sequence homology required for bacteriophage lambda site-specific recombination. J Mol Biol **181:**187-197.

2. **Burgin, A. B., Jr., and H. A. Nash.** 1995. Suicide substrates reveal properties of the homology-dependent steps during integrative recombination of bacteriophage lambda. Curr Biol **5:**1312-1321.

3. **Burgin, A. B. J., B. N. Huizenga, and H. A. Nash.** 1995. A novel suicide substrate for DNA topoisomerases and site-specific recombinases. Nucleic acids research **23:**2973-2979.

4. **Ghosh, K., C. K. Lau, F. Guo, A. M. Segall, and G. D. van Duyne.** 2005. Peptide trapping of the Holliday junction intermediate in Cre-loxP site-specific recombination. Journal of Biological Chemistry **280:**8290-8299.

5. **Gunderson, C. W., J. L. Boldt, R. N. Authement, and A. M. Segall.** 2009. Peptide wrwycr Inhibits the Excision of Several Prophages and Traps Holliday Junctions inside Bacteria. J Bacteriol **191:**2169-2176.

6. **Kim, S., and J. F. Gardner.** 2011. Resolution of Holliday junction recombination intermediates by wild-type and mutant IntDOT proteins. J Bacteriol **193:**1351-1358.

7. **Kitts, P. A., and H. A. Nash.** 1988. An Intermediate in the Phage-Lambda Site-Specific Recombination Reaction Is Revealed by Phosphorothioate Substitution in DNA. Nucleic Acids Res **16:**6839-6856.

8. **Laprise, J., S. Yoneji, and J. F. Gardner.** 2010. Homology-dependent interactions determine the order of strand exchange by IntDOT recombinase. Nucleic Acids Res **38:**958-969.

9. **Lee, J., and M. Jayaram.** 1995. Role of partner homology in DNA recombination. Complementary base pairing orients the 5'-hydroxyl for strand joining during Flp site-specific recombination. J Biol Chem **270:**4042-4052.

10. **Malanowska, K., S. Yoneji, A. A. Salyers, and J. F. Gardner.** 2007. CTnDOT integrase performs ordered homology-dependent and homology-independent strand exchanges. Nucleic Acids Res **35:**5861-5873.

11. **Nash, H. A.** 1983. Purification and properties of the bacteriophage lambda Int protein. Methods Enzymol **100:**210-216.

12. **Nunes-Duby, S. E., D. Yu, and A. Landy.** 1997. Sensing homology at the strand-swapping step in lambda excisive recombination. J Mol Biol **272:**493-508.

Chapter 5

Summary and Future Directions

Summary

The CTnDOT integrase IntDOT is required for excision and integration of CTnDOT into and out of the bacterial host chromosome (1, 2). It has been characterized as a member of the tyrosine recombinase family based on the presence of conserved catalytic residues, but it appears to have a different mechanism of recombination than other family members (3, 4). IntDOT makes a homology-dependent initial strand exchange, followed by a homology-independent second strand exchange (5). I first addressed the question of whether the location of the homology in the overlap is important.

Using complementary DNA oligonucleotides, I designed an *attB* site that contained an inverted overlap sequence. In the normal overlap sequence, there is a conserved GC dinucleotide adjacent to the B core site that provides the homology required for the initial strand exchange. In the inverted overlap *attB* site, the entire overlap sequence was rotated 180° so that the GC dinucleotide was relocated to the bottom strand adjacent to the B' site. When the inverted overlap *attB* site was used in the *in vitro* integration assay with wild type *attDOT*, the recombinant products showed that the orientation of the integrated *attDOT* site was the opposite of the orientation of a wild type recombinant. I used nicked

attB substrates to confirm that relocating the GC dinucleotide had also changed the site of initial cleavage to the bottom strand adjacent to the B' core site.

The results from the *attB* site containing an inverted overlap region suggested that the integrative intasome was synapsing and possibly making initial cleavage and strand exchange, testing for homology, then either ligating and proceeding towards recombinant products or reversing the reaction and releasing the substrates. I showed that this is true by using an *attB* site with a symmetric overlap region, where the GC dinucleotide is present on the top strand adjacent to the B site and also on the bottom strand adjacent to the B' site. Results from the *in vitro* integration assay showed recombinants in both orientations.

Next, I addressed the role of the core-type sites in CTnDOT integration. The core-type sites are D and D' in *attDOT*, and B and B' in *attB*. The B and D core sites share a conserved sequence that extends into the overlap region - GTANNTTTGC, with cleavage taking place between the T and G such that the GC dinucleotide is part of the overlap region. The obvious question is whether the conserved B and D sequence important for IntDOT recognition and binding. I did a mutational analysis on the B and D sites to identify positions important for IntDOT interactions. I identified positions -10, -9, and -5 as particularly important. I substituted all four bases at each of these three positions and noted the effect on recombination. Position -10 requires either a T or a G, possibly for IntDOT interactions with a carboxyl group in the major groove. All four bases were

tolerated at position -9, which may suggest indirect interactions or interactions with the phosphate backbone at this position.

Position -5 appears to be the single most important position of the core sequence, and only a T-A base pair is tolerated. I used the base analogs 2-aminopurine (2AP) and 2,6-diaminopurine (DAP) to identify the important amino group that IntDOT interacts with in the major groove. I further showed that this position is conserved among the poorly conserved B‘ core sites and that interactions between positions -5 and +5 are important for positioning the intasome for productive recombination.

Future Studies

Crystallization of IntDOT

I have used genetic and biochemical approaches to identify positions within the core-type DNA sites that are important for interactions with IntDOT. The next step would be to make crystals of IntDOT in complex with DNA for X-ray crystallographic studies. A crystal structure will test my predictions that IntDOT makes direct interactions at positions -5 and +5, and possibly at position -10. There may also be water-mediated interactions or phosphate backbone interactions with other bases within the core. These results may help us to understand how IntDOT identifies *attB* sites within the *Bacteroides* chromosome, and also define the role of core-type sites in intasome formation.

References

1. **Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2000. Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. J Bacteriol **182:**4035-4043.

2. **Cheng, Q., Y. Sutanto, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers.** 2001. Identification of genes required for excision of CTnDOT, a *Bacteroides* conjugative transposon. Mol Microbiol **41:**625-632.

3. **Craig, N. L., and H. A. Nash.** 1983. The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell **35:**795-803.

4. **Malanowska, K., A. A. Salyers, and J. F. Gardner.** 2006. Characterization of a conjugative transposon integrase, IntDOT. Mol Microbiol **60:**1228-1240.

5. **Malanowska, K., S. Yoneji, A. A. Salyers, and J. F. Gardner.** 2007. CTnDOT integrase performs ordered homology-dependent and homology-independent strand exchanges. Nucleic Acids Res **35:**5861-5873.