STRUCTURE-FUNCTION ANALYSIS OF IntDOT, 
THE CTnDOT INTEGRASE

BY

SEYEUN KIM

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, 2010

Urbana, Illinois

Doctoral Committee:

Professor Jeffrey F. Gardner, Chair
Professor James M. Slauch
Associate Professor Steven R. Blanke
Assistant Professor Carin K. Vanderpool
Abstract

IntDOT is an enzyme required for the recombination reactions that promote movement of CTnDOT that is an integrative conjugative element (ICE) found in Bacteroides spp. IntDOT has been classified in the tyrosine recombinase family. However, unlike most other tyrosine recombinases that require sequence identity in the region where strand exchanges occur, IntDOT recombines overlap sequences that contain mismatched base pairs. Therefore, IntDOT uses different mechanisms from those of other tyrosine recombinases.

IntDOT has three DNA binding domains, the N terminus arm (N) domain, the core binding (CB) domain, the catalytic (CAT) domain. The CAT and CB domain bind to the core type sites and the N domain binds to the arm-type sites that are distantly located from the core-type sites. The CAT domain is the most conserved domain among members of the tyrosine recombinase family and contains six catalytic signature residues, R₁KH₁R₁H₁Y, IntDOT, however, has serine residue in the place of the first catalytic arginine, R₁.

I performed a mutational analysis of IntDOT protein and isolated mutants with substitutions in all three domains of IntDOT proteins, I constructed an E.coli indicator strain that employs the lacI gene as a reporter. The in vivo integration assay using the E.coli indicator strain enables screening for IntDOT mutants from a large population containing random mutations. I isolated twenty five IntDOT mutants that are defective in integration. Four substitutions contained substitutions int eh N domain, two contained substitutions in the CB domain, and the nineteen contained substitutions in the CAT
domain. Biochemical assays such as DNA binding, cleavage, and ligation were used to characterize IntDOT functions to determine which steps in the recombination pathway were defective. The results show that the IntDOT N domain is not required for cleavage and ligation but is important for binding to the arm-type sites. In the CAT domain, residue A352 is likely to be important for positioning the catalytic tyrosine, Y381, in the active site. Residue R285 could provide the missing function of the R₁ residue in catalysis.

During recombination, four IntDOT monomers perform two sets of sequential strand exchanges. The first set of strand exchanges forms the Holliday Junction (HJ) intermediate and the second set of strand exchanges resolves the HJ to complete the recombination reaction. During formation and resolution of the HJ, the four IntDOT monomers communicate with each other and coordinate the resolution of the HJ intermediate. To study protein-protein interactions formed between IntDOT monomers, I developed an HJ in vitro resolution assay using synthetic HJs. I constructed two different types of HJs by designing oligos lacking the arm-type sites but containing the core-type sites. One HJ contains identical overlap sequences as found in most tyrosine recombinase systems. The other HJ contains mismatched overlap sequences as found in natural attDOT and attB sites. I analyzed resolution of two types of HJs by wild type IntDOT and mutant IntDOT proteins. Wild type IntDOT resolves the HJ with identical overlap sequences into both substrates and recombinants but resolves the HJ with mismatched overlap sequences only into substrates. The results indicate that IntDOT and a host factor must form a nucleoprotein complex on attDOT to promote the processing of HJ intermediates through the region of heterology.
Charaterization of HJ resolution by IntDOT mutant proteins showed that three residues, V95, R295, and S368 are important residues for protein-protein interactions in HJ resolution. V95 is a residue in the coupler region between the N terminal and CB domain. The R295 and S368 residues in the CAT domain are on the surface of protein that can likely make protein-protein interactions. I hypothesized that the IntDOT coupler and the N domain interact using charged residues as shown in the interaction between the lambda Int coupler and N domain. To test my hypothesis, I performed a site directed mutational analysis of charged residues in the IntDOT coupler region and N domain. The results indicate that IntDOT coupler is important for HJ resolution, but charged residues in the IntDOT N domain are not involved in protein-protein interactions. It is possible that hydrophobic residues including V95 residue in the IntDOT coupler makes protein-protein interactions.
Acknowledgements

I would like to thank sincerely my advisor Dr. Jeffrey Gardner for his support and guidance. Jeff has given me opportunities to further my understanding of microbiology. His encouragement and understanding made it possible for me to develop my potential during all years of my PhD program. I also like to thank Dr. Abigail Salyers for her understanding. I greatly appreciate their consideration and patience.

I would like to my committee members, Dr. Slauch, Dr. Blanke, Dr. Vanderpool, and Dr. Orlean for their advice and guidance.

I also like to thank members of Gardner’s lab and Salyers’ lab, Carolyn, Margaret, Jennifer, Ken, Sumi, James, Najda, Gui, Jiyoun, and Jill for their help and friendship. I really enjoyed working with them. I particularly like to mention Sumi for her friendship and cheerful company of Korean lunch every week. I also like to special thank to Dr. Brian Swalla who provided me his IntDOT 3D-model.

Lastly, the love, faith, and encouragement of all members of my family: my husband Jaeseob, my son Younseo, my parents, and Jaeseob’s parents made me possible this long journey.
# Table of Contents

Chapter 1. Introduction .............................................................................................................. 1  
1.1 Bacterial resistance to antibiotics .................................................................................. 1  
1.2 Horizontal gene transfer by Conjugative Transposons ............................................ 2  
1.3 Movement of CTnDOT ..................................................................................................... 3  
1.4 Site-specific recombination by tyrosine recombinases .............................................. 6  
1.5 The structures of tyrosine recombinases ........................................................................ 10  
1.6 Diversity in tyrosine recombinases ................................................................................ 17  
Thesis Outline ........................................................................................................................... 22  
Figures and Tables ..................................................................................................................... 25  
References ................................................................................................................................ 38  

Chapter 2. Structure – function analysis of IntDOT ............................................................... 45  
2.1 Introduction ...................................................................................................................... 45  
2.2 Materials and Methods .................................................................................................... 46  
2.3 Results and Discussion .................................................................................................... 54  
Figures and Tables ..................................................................................................................... 71  
References ................................................................................................................................ 81  

Chapter 3. Analysis of HJ resolution by IntDOT ................................................................. 85  
3.1 Introduction ...................................................................................................................... 85  
3.2 Materials and Methods .................................................................................................... 87  
3.3 Results and Discussion .................................................................................................... 89  
3.4 Conclusion ...................................................................................................................... 101  
Figures and Tables ..................................................................................................................... 102  
References ................................................................................................................................ 112  
Summary .................................................................................................................................. 114  

Chapter 4. Future directions ................................................................................................. 117  
4.1 Crystallization of IntDOT ................................................................................................. 117  
4.2 Analysis of the role of IntDOT in excision ...................................................................... 118  
4.3 Roles of the arm-type sites of CTnDOT in HJ resolution ........................................... 119  
References ................................................................................................................................ 120  

Curriculum Vitae ..................................................................................................................... 121
Chapter 1. Introduction

1.1 Bacterial resistance to antibiotics

Mammalian intestines are habitats for a variety of obligate anaerobes. One of the most abundant anaerobes found in the human colon is the gram-negative rod *Bacteroides* spp. that accounts for 25-30% of the normal microflora (62). Under normal conditions *Bacteroides* are not harmful to the host. However, some *Bacteroides* spp. can become pathogenic when they escape from the colon during trauma or surgery causing death.

Most *Bacteroides* strains are resistant to antimicrobial agents and *Bacteroides fragilis* is a commonly isolated anaerobe from clinical specimens (24). In the 1950s and 1960s, *Bacteroides* infections were treated with tetracycline. It is less toxic than other antibiotics and is effective against a broad spectrum of bacteria (61). However, the incidence of *Bacteroides* strains resistant to tetracycline has increased drastically in the past three decades (62).

Three tetracycline resistance mechanisms have been discovered; tetracycline efflux, ribosome protection, and tetracycline inactivation (69). Among them, the ribosome protection mechanism seems to be the most widespread (24). Several different classes of ribosomal protection resistance genes, *tet K*, *tetL*, *tetO*, *tetQ*, and *tetM*, have been identified in a wide range of gram positive and gram negative bacteria (68). In one survey, before the 1970’s, only 20% of the clinical isolates from *Bacteroides* infections carried the *tetQ* gene but from 1980 to 1995 more than 80% of isolates were resistant to tetracycline (62). Currently, tetracycline is only used for a limited range of bacterial infections and is usually less effective as a primary treatment for clinical infections. In
addition to the concern about the increase of tetracycline resistance, resistances to other antibiotics like erythromycin and clindamycin have also increased (66). It is possible that virulence and toxin genes might be spread among the intestinal bacteria population using a horizontal gene transport similar to one for the *tetQ* (57).

### 1.2 Horizontal gene transfer by Conjugative Transposons

Conjugative transposons [CTns; also known as Integrative and Conjugative Elements (ICEs)] are self-transmissible elements that combine the features of transposons, bacteriophage, and plasmids (63). Figure 1.1 shows that CTns are excised from the chromosome and form a double stranded circular DNA intermediate. This intermediate is nicked at the transfer origin (*oriT*) by Mob and one strand is transferred to the recipient by rolling circle replication. The complementary strand is circularized and replicated in (63) and the double stranded circular form of the CTns is integrated into the chromosomes of the donor and recipient (70). CTns are different from transposons because they do not duplicate target sites on integration and also move between cells exclusively by conjugation. They are different from plasmids in that the circular intermediate does not replicate. They are also different from temperate bacteriophages in that CTns don’t exclude other CTns and do not lyse the host after excision (61).

CTns can transfer between the same and different genera such as from gram positive cocci to gram negative *E. coli* (63). Some CTns like CTnDOT also enhance transfer of other co-resident mobile elements such as NBU1 (Non-replicating Bacteroides Unit).
Bacteroides conjugative transposons are known to be responsible for the rapid spread of tetracycline resistance genes among Bacteroides spp. Bacteroides CTns are between 65 to 150 kilobases in size (63). The best studied Bacteroides CTn is CTnDOT (Figure 1.2). CTnDOT carries a tetracycline resistance gene, tetQ, that encodes a ribosomal protection protein gene and a gene encoding erythromycin resistance, ermF (80). The ErmF protein has a ribosomal methyltransferse activity. Since the tetQ alleles isolated from various Bacteroides spp. have 96-100% identity at the DNA sequence level, it suggests that the tetQ gene is transferred horizontally and is maintained stably in the host chromosome in the absence of tetracycline (61).

1.3 Movement of CTnDOT

Excision of CTnDOT

CTnDOT excision is a complicated process and is stimulated 100 to 1000 fold by tetracycline (12). It requires at least four CTnDOT encoded proteins, Orf2c, Orf2d, Exc, IntDOT, and a host factor (70, 72, 73). The attL and attR sites are attachment sites (att) that are required for excisive recombination reaction. Excision reforms the attDOT site which is the joined end of the excised element and the attB site in the intact chromosome (Figure 1.3). The attDOT site made by excision contains mismatched sequences called ‘the coupling sequences’ because the sequences of the ends of the element are not identical. These mismatches are repaired by the host process such as mismatch repair system or by replication of the complementary strand in the donor and recipient (Figure 1.1).
Excision and transfer of CTnDOT are regulated by genes encoded by three operons in CTnDOT (13); the central regulatory region (tetQ-rteA-rteB, and rteC), the excision operon (orf2c-orf2d-exc), and the transfer region (traA to traQ).

In the central regulatory region, the tetQ operon encodes a ribosome binding protein, TetQ, a sensor protein, RteA, and a regulator protein, RteB (Figre 1. 4). A transcriptional regulator RteC is expressed from its own promoter. Expression of tetQ is regulated by translational attenuation at the 5’ leader region of the tetQ operon when the donor is exposed to tetracycline (77, 78). Ribosomes bound with tetracycline release the ribosomal binding site (RBS) in the 5’ leader region of tetQ mRNA operon that was masked by hairpin structures. Translation of the polycistronic mRNA leads to expression of the TetQ, RteA, and RteB proteins. The TetQ protein makes the donor resistant to tetracycline and RteA and RteB proteins activate expression of the rteC gene (49). The RteC protein acts as a transcriptional activator and turns on the expression of excision operon containing the orf2c, orf2d, and exc genes (69). Orf2c and Orf2d are small basic proteins that bind to attL and attR. Their functions are under study. Exc is 78 kDa in molecular weight (73) and has a type 1A topoisomerase activity that can relax negatively supercoiled DNA. However, Exc is not required for in vitro excision (72) in an intermolecular reaction where the attL and attR sites are located on two different plasmids. In vitro excision of CTnDOT with attL and attR sites on the same DNA molecule is enhanced about 5 fold in the presence of Exc compared to the reaction with no Exc (C. Keeton unpublished ). This indicates that Exc may have an architectural role with the natural intramolecular substrates attL and attR in vivo (73).
The transfer operon encodes 17 proteins, TraA to TraQ, that are involved in conjugative transfer of CTnDOT. Recent work on the transfer region (34) shows that the amount of mRNA encoding the transfer genes is enhanced by expression of the excision proteins indicating that the tra region is also controlled at the transcriptional level. It appears that the excision and transfer of CTnDOT are performed by cascade reactions involving these three regulatory regions of CTnDOT.

**Integration of CTnDOT**

Unlike excision of CTnDOT, integration of CTnDOT does not seem to be regulated because IntDOT is constitutively expressed from its own promoter. Integration requires IntDOT and a *Bacteroides* host factor (72). The host factor can be substituted by *E. coli* Integration Host Factor (IHF) *in vitro* (15). The integration reaction occurs between the attB site in the *Bacteroides* chromosome and the attDOT site on the circular intermediate form of CTnDOT.

The attDOT site is ~250 bp in size and the attB site is ~80 bp in size. The core sites of the attDOT and attB sites contain two imperfect inverted repeats called D and D’ in attDOT, and B and B’ in attB. A spacer of 7 bp located between the two inverted repeats is called the overlap region (O) or crossover region. The overlap region contains the site where the strands are exchanged. The attDOT site contains not only the core-type sites but also the arm-type sites that flank the core; R1’, R1, R2, R2’ on the left side and L1, L2 on the right side of the attDOT core site (20, 48) (Figure 1.5).

The *Bacteroides* chromosome has at least 8 preferred attB sites for CTnDOT integration. The B core site of all known attB sites contains the sequence GTA--TTTGC (12) and this sequence is also conserved in the D site of attDOT. The cleavage site is
between T and G. The first two base pairs of the overlap region are called “the GC dinucleotide” and they are invariant and identical in \textit{attB} and \textit{attDOT} sites. The sequences of the other five base pairs in the overlap vary and not identical in \textit{attB} and \textit{attDOT} (Figure 1.6). During CTnDOT integration, two sequential strand exchanges take place in the overlap region which contains non-complementary sequences. This differs from the organization of the overlap regions of other tyrosine recombinase systems. Most proteins in the tyrosine recombinase family such as lambda Int contain sequence identity in the overlap region.

1.4 Site-specific recombination by tyrosine recombinases

The first site specific recombination (SSR) reactions discovered and studied in detail were the integration and excision reactions of phage lambda. Integration and excision of lambda are mediated by the phage encoded integrase protein (Int), which is a member of tyrosine recombinase family (9, 51). In subsequent years, SSRs have been discovered in many temperate bacteriophage systems. In addition, tyrosine recombinases mediate other various biological processes such as resolution of replicon dimers, chromosome segregation, transposition, and regulation of gene expression such as those involved in nitrogen fixation (For a review see (76)).

Most tyrosine recombinases require strict homology in the overlap region for efficient recombination. In the case of lambda integration, a single mismatch in the overlap abolishes recombination. If sequence identity is restored in the partner site, recombination occurs at the wild type level (4). However, some recombinases such as
integrate of Tn916 and CTnDOT perform reactions between sites that contain non-complementary bases in the overlap region.

The intasome and synaptic complex

Tyrosine recombinases initiate the SSR reaction by recognizing specific att sites of their respective elements. Recombinases bind to the att sites and form higher-ordered protein-DNA complexes called intasomes. Some proteins such as lambda Int and IntDOT need a host factor and their arm-type sites to form intasomes. In lambda system, the host factor, IHF bends DNA and promotes formation of the integrative intasome along with Int (30, 75). For lambda excision, two different intasome complexes form on attL and attR. Lambda Int and IHF form a nucleoprotein complex on attL that interacts with the intasome formed on attR (5). The intasome made on attR contains lambda Int, IHF, the phage-encoded Excisionase (Xis), and the host-encoded Factor for inversion stimulation (Fis). The two recombining sites in an intasome form a synaptic complex where the att sites become aligned in a configuration that is favorable for the formation of recombinant products.

Cleavage and ligation reactions

Cleavage and ligation are catalytic reactions that can occur in the absence of intasome formation and can be executed by recombinase monomers. This is consistent with the fact that the CAT domain of lambda Int alone is sufficient to perform cleavage and ligation reactions with appropriate substrates. Cleavage takes place at precisely defined sites. The hydroxyl group of the catalytic tyrosine attacks a phosphodiester bond at the cleavage site. The protein is covalently attached to the 3’ end of the DNA by a phosphotyrosyl bond and a free 5’-hydroxyl group is released. The released 5’-hydroxyl
group is activated as a nucleophile to reform a phosphodiester bond either with the original strand or with a phosphotyrosyl bond linked to a protein on the partner strand. Ligation reforms a phosphodiester bond and proteins leave the DNA backbone.

Cleavage assays have been designed to test the ability of a tyrosine recombinase to cleave DNA. Two types of suicide substrates are used: a phosphorothioate-substituted substrate or a nicked substrate. Recombination reactions with these substrates stop after the cleavage step but before the ligation step to make the reaction irreversible. A phosphorothioate-substituted substrate contains a sulfur in the place of a bridging oxygen of a phosphodiester bond at the cleavage site (83). The cleavage reaction leaves a sulfhydryl group instead of hydroxyl group at the 5’ end. The sulfhydryl group is a poor nucleophile so the reaction cannot proceed forward or backward. A covalently linked protein-DNA complex is trapped and detectable on a gradient gel.

In a nicked suicide substrate, the top strand contains a nick usually created 2-3 bases away from the 3’ end of the cleavage site. The cleavage reaction leaves a protein-DNA complex at the 3’ end and the 2-3 bases between the cleavage site and the nick diffuse away (55). Loss of 5’ hydroxyl group makes this reaction irreversible.

Ligation assays use a substrate containing an activated para-nitrophenol group (pNP) attached to the 3’ end of a cleavage site (8). This substrate mimics the phosphotyrosyl intermediate. In the presence of an active recombinase monomer, the adjacent 5’-hydroxyl attacks the linkage between the pNP group and DNA and reforms a phosphodiester bond. The product formed containing a phosphodiester bond is detected as a ligated single stranded DNA on a denaturing gel.
Holliday Junction (HJ) intermediate

Formation of the HJ intermediate is a hallmark of reactions mediated by tyrosine recombinases. HJs are formed by the first pair of strand exchanges by two active monomers of the four recombinase monomers in a synaptic complex (Figure 1.7). One strands of each site are cleaved and exchanged first to form a HJ. Thus, a HJ intermediate contains four junctions with a recombinase monomer bound to each junction. A HJ can be resolved in one of two directions. The four recombinase monomers communicate by protein-protein interactions to resolve a HJ intermediate into recombinants. If the HJ undergoes a conformational change or isomerization step, the two previously inactive monomers become active. The newly activated monomers cut and exchange the intact strands of DNA at sites 6-8 bases away from the first strand exchanges and the HJ is resolved to produce recombinants. However, if the original pairs of recombinase monomers continue to be activated, this pair of recombinase monomers cuts and exchanges the strands at the same site where the first strand exchanges occurred. This recombination reaction reforms the original substrates. Thus, isomerization is a key step of recombination to determine the direction of the recombination reaction to generate recombinants or substrates.

The lambda recombination reaction requires all 7 bp of sequence identity in the overlap region. A branch migration model was originally proposed to explain the requirement for the homology in the overlap of the lambda system (19, 37). According to the branch migration model, after the first strand exchanges, the junction moves 7 bp toward the site where the second strand exchange occurs. However, the model was
challenged by experiments testing synthetic HJs incorporating heterologies to make a junction constrained at various locations within the overlap region in lambda system (53). This study showed that lambda Int resolves a HJ efficiently when the junction is frozen near the center of the overlap region not the either side of the overlap region. In addition, one base shift of the junction in the middle of the overlap region changes drastically a bias of direction of resolution. One type of product is exclusively by top strand exchanges and the other type of product is exclusively by bottom strand exchanges. This suggested that resolution is not dependent on branch migration across identical sequences and HJs can have two distinct conformations depending on the location of the junction, a HJ favorable to top strand exchanges and a HJ favorable to bottom strand exchanges.

1.5 The structures of tyrosine recombinases

Despite the variety of roles of tyrosine recombinases in cells, most recombinases show a common arrangement of secondary and tertiary structures. The tyrosine recombinase family has two subgroups of proteins depending on whether or not they contain an arm (N) domain at the N terminus of the protein. One group, called “autonomous tyrosine recombinases,” contains two DNA binding domains, the core binding (CB) and catalytic (CAT) domains. The Cre, Flp, and XerC/D proteins are in this group. They perform recombination reactions that do not show regulated directionality. Proteins in the other group, called the “factor assisted tyrosine recombinases,” contain three DNA binding domains: the N domain, the CB domain, and the CAT domain. Integrase of lambda, Tn916, and CTnDOT are in this group. The CB
and CAT domains recognize core-type sites and N domain recognize arm-type sites. The site-specific recombination reactions by these proteins show regulated directionality.

**The N domain**

The N domain is thought to be an accessory domain that has attached to the CB and CAT domain of factor assisted recombinase proteins. The CB and CAT domains that perform the catalytic reaction have been studied in depth, whereas the arm domains are the least studied due to lack of sequence similarity. In spite of the lack of sequence similarity among members of the tyrosine recombinase family, the N domains of three different proteins, lambda Int, Tn916 integrase, and IntDOT show a conserved folding pattern in the secondary structures. They contain three beta strands $\beta_1$, $\beta_2$, $\beta_3$, and two helices, H1 and H2 (Figure 1.8).

In the lambda Int, the N domain has higher binding affinity for the arm-type sites than the CB and CAT have for the core-type sites (38). A crystal structure (6) and an NMR study (23) of the lambda arm domain revealed that the three beta sheets recognize the sequences in the major groove of the arm-type sites. Two sequence-specific interactions are well defined by the NMR study (23). R19 in the strand $\beta_1$ and E34 in the strand $\beta_3$ form hydrogen bonds with specific nucleotides in the major groove. The second sequence specific contact is made by N21 in the loop between the strand $\beta_1$ and $\beta_2$. Side chain of N21 enables to donate and accept hydrogen bonds with specific sequences in the arm-type sites. These interactions could explain the higher binding affinity of the N domain to the arm-type sites than the CB and CAT domain to the core-type sites.
One recent study on the lambda Int N domain (23) revealed that the first 11 residues of protein are required for recombination and facilitate recognition of the minor groove in the arm-type sites. The structure of this region of lambda Int is flexible depending on the presence or absence of DNA. In the absence of DNA, this region is disordered and unstructured but becomes ordered and immobilized in the minor groove in the presence of the DNA substrate. The beginning of protein binds to the minor groove of DNA which assists DNA recognition by three β strands to the major groove of the arm-type site.

Little is known about the function of the H1 helix but E47 of lambda Int located in H1 interacts with the excision protein, Xis. This indicates that H1 is likely to have a role in directionality. The N domain of lambda Int has an interdomain region, helix H2. The H2 is a linker between the N and CB domain and is called “the coupler”. A mutational study of the lambda coupler showed that H2 is required for recombination and has a role in protein-protein interactions involved in HJ resolution (64). Lee et al. (2005) showed that mutant proteins of lambda Int with the reverse charged residue, D71R in the coupler and R30D in the loop between the β 2 and β 3 strands in the N domain were defective in HJ resolution. However, mutant protein containing the double substitutions of R30D:D71R was proficient in HJ resolution because the R30D:D71R mutant protein restored an ion pair by the substituted residues. This result indicates that R30 interacts with D71 by a charge interaction (45). This is an intermolecular interaction where the coupler of one monomer contacts the loop between β2 and β3 in the N domain of its neighboring monomer.
The CB domain

The CB domains of tyrosine recombinases contain four to five helices and no β strands. An alignment study of the CB domains of 25 tyrosine recombinases suggested that the CB domains contain a similar folding pattern of five helices (A, B, C, D, and E) (74). However, crystal structures of lambda Int (6) and XerC/D (71) show four helices in the CB domain and no helix E. For those proteins containing four helixes, the lack of helix E gives more space to the CAT domain when the CAT domain rotates dynamically during recombination (6). The CB domain functions as a supporter for the CAT domain to perform catalysis. The CB domain provides specificity for the recombination reaction and binding affinity to the core-type sites (3).

The CB domains of tyrosine recombinases recognize the core-type site specifically and non-specifically (38). The lambda Int crystal structure (3) shows that helices B and D contact non-specifically with the phosphate backbone of DNA. The Cre crystal structure (28, 31) shows that the CB domain contains five helices (A, B, C, D, and E) and recognizes DNA. Helices B and D directly contact with the major groove of the core-type site. Helix E is positioned differently in active and non-active subunits. Another study has shown that the Cre CB domain has a function in protein-protein interactions (27). The A36V mutant protein of Cre containing a substitution in the loop between helices A and B, was shown to slow the rate of recombination although it is proficient in binding and cleaving DNA (27). It is likely that the A36V protein is defective in synapsis due to the lack of protein-protein interactions. Like Cre, the Flp CB domain contains five helices A, B, C, D, and E. The crystal structure of Flp shows that helix A has four segments A1, A2, A3, and A4. Helices A1 and B bind to DNA non-
specifically by phosphate contacts whereas Y60 in helix B contacts DNA with a base specific interaction (27).

For the factor assisted tyrosine recombinases, the CB domain of the integrases of phage lambda and HK022 have been well studied. Phage lambda and HK022 are two closely related phages and their integrases have ~ 70 % sequence identity and the same biochemical mechanisms of recombination (22). However, these two proteins specifically recognize their own core-type sites. Neither protein can perform recombination on sites containing the others’ core-type sites. Nagaraja et al. isolated lambda Int mutants that can recombine HK022 att sites (50). They found that lambda Int mutant protein with a substitution of N99D in helix B (21, 22). A subsequent study showed that the relaxed specificity of the N99D protein of lambda Int is caused by altering binding specificity of the protein to the core-type site (14, 21)). This study showed that the CB domain of lambda Int can be a determinant of recombination specificity.

The CAT domain

The CAT domains of tyrosine recombinases perform the catalytic reactions such as cleavage, strand exchange, and ligation at the core-type sites. Most members of the tyrosine recombinase family contain the six invariant residues in the CAT domain (52). They are called the “signature sequences”: arginine R_I, lysine K, histidine H_I, arginine R_{II}, histidine H_{II}/typtophan W, and tyrosine Y. These residues of tyrosine recombinases are clustered at the end of the CAT domain which is the strongest conserved domain among members of tyrosine recombinase family. The location of these residues is often well-conserved. As shown in Figure 1.9, the first arginines, R_I are located in the loop between
helices G and H except for in IntDOT. In the case of IntDOT, a previous alignment study was unable to identify the R₁ of IntDOT at the conserved position. The lysines are in the loop between strands β2 and β3. The first histidines, H₁ are in the loop between helices J and K. The second arginines, R₁I are in helix K. The second histidines, H₁/tryptophan W in Cre and Flp are in the loop between helices L and M. The catalytic tyrosines Y are in helix M in most tyrosine recombinases. The Y of lambda Int is located in the β8 strand followed by helix L. Work presented in this thesis indicates that the arginine residue that functions as the R₁ of IntDOT is found in an unusual location in the CAT domain, R285 (36), that is closer to the catalytic lysine K while other proteins have a larger distance between the R₁ and K.

The roles for some of the catalytic residues have been proposed. As shown in Figure 1.10, the cleavage reaction is mediated by nucleophilic action of the catalytic tyrosine and with the help of the other five residues in the active site. The two arginines, R₁ and R₁I in this complex are thought to stabilize this intermediate by contact with two non-bridging oxygens (76). A proton released from hydroxyl group is presumably accepted by a basic residue. In Cre, H₁ (H289) is thought to be the prime candidate that can carry out this function (76). In the case of vaccinia virus topoisomerase, K has been proposed to act as a general acid protonating the leaving 5-hydroxy group when a phophodiester bond is reformed and protein is released (39). The role of H₁I/ W is uncertain (29).

Besides the conserved locations of these signature residues, the CAT domain shows other structural conservations. The glycine next to H₁I in the loop between helices L and M is a strongly conserved residue and it is also likely to be involved in catalysis.
These two residues are called ‘the GH_{II} dyad’. The number of amino acids between the GH_{II} dyad and the catalytic Y is also conserved (52).

A crystal structure of lambda Int (3) has shown that the tail containing β9 strand (residues 350-356) can interact with the β7 strand (residues 239-243) either in the same protein (intramolecular) or in the neighboring protein (intermolecular). When the tail of lambda Int contacts the β7 strand of same monomer, it reduces cleavage activity. On the other hand, when the tail contacts the β7 strand of the neighboring protein in a recombination complex, the inhibition is relaxed and the recombination complex performs coordinated HJ resolution. A truncated protein of lambda Int, the W350 protein that is missing the tail (43) was tested in cleavage of half att site substrates or in cleavage of full att site substrates to determine whether monomers can cleave DNA. The W350 protein was also tested in HJ resolution (35). The result was that monomeric W350 protein increases cleavage activity but multimeric W350 protein reduces cleavage and resolution activity. The monomeric W350 protein is more active in cleavage than the wild type protein but defective in making protein-protein interactions (32). These results indicate that the tail has two functions: the regulation of catalysis and coordination of HJ resolution by protein-protein interactions (3).

The Cre CAT domain contains nine helices and five beta strands (31). In the Cre crystal structure, the majority of CAT domain interactions with DNA are made by direct contacts, of the helix J that interacts with the major groove of DNA (31). In Figure 1.11, the last helix N in the CAT domain is flexible and embedded in the hydrophobic pocket of an adjacent monomer (31). This contact forms a cyclic tetramer required for efficient
recombination. Cre helix N is structurally and functionally equivalent to the lambda Int tail.

The Flp CAT domain is composed of fourteen helices and seven strands (11). In Flp crystal structure, the position of Y343 is well ordered in a pair of active monomers whereas the hydroxyl group of catalytic tyrosine is positioned 10 Å away from the target phosphates in the inactive monomers (11). The most distinct feature of recombination by Flp is that the catalytic tyrosine is provided “in trans” by adjacent monomer of Flp (10). One monomer provides the RHRKW residues while the other provides the catalytic tyrosine. Thus, a recombinase monomer bound to the core site at the site of cleavage is not the monomer that actually cleaves phosphate of the backbone (44).

1.6 Diversity in tyrosine recombinases

Autonomous tyrosine recombinases

Cre recombinase

The Cre recombinase is encoded by the E. coli bacteriophage P1. P1 exists as a plasmid in the cell. The function of Cre is to resolve dimers of the P1 genome to monomers to facilitate segregation during cell division. Cre contains 343 amino acids and binds to a specific DNA site called loxP site when it performs site specific recombination (76). The loxP site contains 32-bp that consists of 6 bp of overlap region flanked by two 13-bp inverted repeats. The loxP-Cre system is an autonomous recombination system and requires no host factors or additional DNA sequences. Cre-loxP system is often used as a genetic tool to engineer DNA rearrangements.
Warren et al. recently constructed a chimeric Cre protein that contains the lambda N domain, and the Cre CB and CAT domains. This protein makes the reaction unidirectional and it becomes regulated by accessory proteins such as IHF, lambda Xis, and FIS (79). This study shows that Cre recombinase can gain a new function when the protein is modified to have a lambda N domain. It is surprising that an autonomous recombinase can accommodate a N domain and act as a factor assisted recombinase. It might indicate that this type of adaptation of autonomous proteins contributes to the evolutionary diversity of tyrosine recombinases.

**Flp recombinase**

Flp is an eukaryotic member of the tyrosine recombinase family encoded by the yeast *Saccharomyces cerevisiae* 2-µm plasmid. Flp contains 423 amino acids and is responsible for control of the copy number of the plasmid during cell division. Flp recombines the *FRT* sites in the plasmid that contains 8 bp of overlap flanked by two 13 bp inverted repeats (65). The Flp monomer contains a tail (residues 362-423) that is about 100 amino acids longer than those of any other tyrosine recombinases. The Flp tail includes helices P and Q and the last two β strands, β 6 and β 7. These helices and strands of the tail provide flexible inter-domain connections.

**XerC/D**

All bacteria have XerC and XerD. XerC and XerD are encoded by the bacterial chromosome and resolve chromosomes after DNA replication. XerC and XerD both contain 298 amino acids and have 36% identity at the amino acid sequence level (7). XerC/D recombination differs from other tyrosine recombinase systems in that the reaction is performed by two related proteins rather than by a single protein. XerC and
XerD perform a variety of SSRs in the cell depending on the sites that they react with: the \textit{dif} site on chromosome, the \textit{cer} site on the plasmid ColE1 (17), and the \textit{psi} site on the plasmid pSC101 (16). In the chromosome, XerC/D recombines two intramolecular \textit{dif} sites that contain 28 bp core-type sites. The \textit{dif} site contains two half sites with different binding specificities: one for XerC and the other one for XerD. The role of XerC/D and the \textit{dif} system is to maintain faithful segregation of the \textit{E. coli} chromosomal DNA after cell division. In the case of plasmids ColE1 and pSC101, the recombination reaction requires accessory proteins and 200 bp of additional DNA sequences adjacent to the \textit{cer} or \textit{psi} attachment sites to resolves dimers of the plasmids to monomers (16, 47). Since XerC/D are encoded by the chromosome, they are thought to be an ancestor of other phage-encoded recombinases.

\textbf{Factor assisted recombinases}

\textbf{Lambda Integrase}

Lambda Int is encoded by phage lambda and is one of the best studied tyrosine recombinases. It contains 356 amino acids. Integration of lambda requires integrase and a host factor, IHF. Recombination occurs between the \textit{attP} site on the phage lambda genome and the \textit{attB} site in the host chromosome. Figure 1.12A shows integration and excision of phage lambda. The \textit{attB} site contains only two core site flanking the overlap, BOB’. The \textit{attP} arm contains the core (COC’) and the arms, the P arm on the left and the P’ arm on the right of the core (41). The P arm contains the P1 and P2 site and the P’ arm contains the P’1, P’2, and P’3 site. The specific IHF binding sites are the H1 and H2 site in the P arm and the H’ site in the P’ arm (84). All necessary protein-DNA complexes are formed on the \textit{attP} and the naked \textit{attB} is captured by the \textit{attP} intasome.
Two recombinant sites, \textit{attL} and \textit{attR}, are generated by integration at the junctions of the lambda genome and host chromosome.

The excision reaction requires lambda Int, IHF, and the phage-encoded Xis. Excision is stimulated by host-encoded Fis. Xis and Fis are DNA bending proteins and have an architectural role. Three molecules of Xis, two molecules of IHF, one molecule of Fis, and Int assemble an intasome on \textit{attR}. Two molecules of Xis bind to X1 and X2 and the third Xis binds to bent DNA site between X1 and X2 (Figure 1.12B). They bend and wrap DNA forming nucleoprotein filaments (1, 54). The function of these filaments is to stabilize the recombination complex. Xis is a recombination directionality factor of lambda system (26) in that DNA bending by Xis stimulates excision and inhibits integration.

\textbf{IntDOT}

IntDOT is encoded by CTnDOT and contains 411 amino acids. The strand exchange mechanism by IntDOT is different than other members of the tyrosine recombinases family. Integrative recombination by IntDOT occurs by two sequential rounds of strand exchanges between \textit{attDOT} and \textit{attB} (46). The first strand exchanges occur by cleavage between the T and GC dinucleotide of the \textit{attDOT} and \textit{attB} sites (Figure 1.13) and this reaction is homology dependent. The second strand exchanges occur where sequences are not complementary 7 bases away from the first strand exchange sites. Thus, the second set of strand exchanges is not affected by the lack of complementary DNA, and this reaction is a homology independent reaction. Interestingly, when the GC dinucleotide located in the top strand of \textit{attB} is moved to the bottom strand of \textit{attB}, IntDOT cleaves and exchanges the bottom strand containing the
GC dinucleotide first (42). A distinct feature of the IntDOT recombination mechanism is that sequences in the core-type site can change the order of strand exchange depending on the location of the GC dinucleotide. It also indicates both orientations of attB (BOB’ and B’OB) can form intasome complexes capable of undergoing recombination (42).

**Integrate of Tn916**

Tn916 was the first CTn discovered. It was found in the gram positive *Enterococcus faecalis*. It carries a gene, tetM, that encodes a tetracycline resistance protein (25). Excision and integration of Tn916 occurs by recombination in a manner similar to the lambda system. However, unlike lambda, Tn916 integrates randomly and the Tn916-encoded integrase is the only required enzyme in the integration reaction. No host factor is required for integration. This may explain that Tn916 has a broad host range that includes both gram positive and gram negative bacteria. Excision of the element requires integrase and the Tn916 encoded Xis (33) and is enhanced by the host factor HU (18). Like CTnDOT, the overlap region where strand exchange occurs lacks sequence identity and recombination of Tn916 tolerates mismatches.

**NBU1 Integrate**

NBU is a mobilizable transposon found in *Bacteroides* spp. The two recombining sites share sequence identity within overlap region similar to what is seen in the lambda system. Excision of NBU1 is dependent on a co-resident CTn in the chromosome. Excision is complicated involving five element-encoded proteins whose functions are under investigation. Integration of NBU1 is independent of a CTn. The target sites and NBU1 encoded integrase, IntN1, and uncharacterized host factors are required for
integration. IntN1 has been characterized as a tyrosine recombinase (59) and mediated NBU1 integration occurs site-specifically into the 3’ end of the Leu tRNA gene (67).

Thesis Outline

My project is to perform a structure-function analysis of the IntDOT protein, the integrase of conjugative transposon, CTnDOT. IntDOT is an enzyme that mediates site-selective recombination during the movement of CTnDOT. It has been classified as a member of the tyrosine recombinase family. Previous studies have shown that the mechanism of recombination mediated by IntDOT is different from those of other tyrosine recombinases such as lambda Int and Cre.

In Chapter 2, I present a mutational analysis of IntDOT. I used a genetic approach to isolate IntDOT mutants that are defective for integration. I developed an in vivo integration assay that uses the lacI gene as the reporter in E. coli. This assay was used to screen for IntDOT mutant proteins from a large population of random mutations. The int DOT gene was mutagenized with hydroxylamine and the mutated int DOT genes were expressed in the E. coli indicator strain. I screened approximately 3,000 colonies and isolated twenty five mutants that produced defective IntDOT proteins. Twenty three of the mutants changed residues that were not substituted in previous studies. Four residues, R13, S38, V95, and G101 in the N domain, two, T184 and P209 in the CB domain and seventeen, T256, W280, R285, R295, V319, C325, L326, R330, S347, A352, T354, T365, S367, S368, G371, A382, and L389 in the CAT domain were newly identified as
important residues for integration. Mutants with substitutions of two previously known catalytic residues, R348 and H372 were also isolated in this study. I tested the 25 mutant proteins for activity in three distinct steps in the recombination reaction. These included DNA binding, cleavage, and ligation. I found that mutants with substitutions in the N domain were active in catalysis but were defective for binding to the arm-type sites to forming an intasome. The A352 residue in the catalytic domain is likely to stabilize the position of the catalytic tyrosine during cleavage. The biochemical data and a homology based model proposed that R285 might act as the first catalytic arginine of IntDOT that was previously inidentified. This work has been published (36).

In Chapter 3, I present an analysis of HJ resolution by wild type and mutant IntDOT proteins. For wild type IntDOT, I constructed two types of HJs containing the core-type sites. One contained a heterologous overlap and the other contains a homologous overlap. I demonstrated that IntDOT resolved synthetic HJs in a manner similar to lambda Int. IntDOT resolves a HJ with identical overlap sequences in both directions but a HJ with heterologous overlap was resolved in only one direction with bias to substrates. This result indicates that in order for IntDOT to recombine sites containing heterology, intasome formation with the arm-type sites is required. Then, I tested mutant IntDOT proteins on HJ resolution. I found that the V95 in the coupler and the R295 and S368 in the CAT domain are important for protein-protein interactions. The residues in the coupler and N domain of lambda Int have been shown to function in protein-protein interactions by a charge interaction. In order to determine whether IntDOT has a similar interaction, I mutated charged residues in the arm domain and coupler to identify residues that may make a protein-protein interaction required for HJ
resolution. I found that substitution of negative charged residues at four different locations, D19, D31, D44, and D49 in the arm domain had no effect on resolution. However, D19 and D44 seem to be important for the arm binding and intasome formation. In the coupler, K94 and K96 are also important for intasome formation and resolution. This indicates that the coupler is important for HJ resolution but negatively charged residues in the N domain are not involved in resolution. Therefore, protein-protein interactions during HJ resolution by IntDOT may be made by hydrophobic clusters.

IntDOT is a tyrosine recombinase that uses a different mechanism than other tyrosine recombinases. During recombination, IntDOT appears to use a very different mechanism for HJ resolution that is independent on homology. The coupler of IntDOT appears to be important for a protein-protein interaction and the interaction seems to be made by hydrophobic residues.
Figures and Tables

Figure 1.1 Movement of Bacteroides conjugative transposons. The CTn integrated in the chromosome is excised and forms a covalently closed circular intermediate. The intermediate is nicked at the oriT and a single strand of CTnDOT is transferred by conjugation. The complementary strand is replicated by rolling circle replication in both donor and recipient and then newly formed intermediates are integrated into chromosome. Figure is taken from (81).
CTnDOT

Figure 1.2 Organization of the CTnDOT. CTnDOT is approximately 65 kb in size. It carries tetracycline resistance gene, \textit{tetQ} and an erythromycin resistance gene, \textit{ermF}. Figure is taken from (80).
Figure 1.3 Excision and integration of CTns. The attL and attR junctions between CTnDOT and the bacterial genome are cleaved 7bp apart. The sequences in the staggered cuts in the both ends are not complementary. The joined ends contain a bubble of five mismatched pairs. This mismatch is repaired by replication during transfer of one strand of CTnDOT from the donor to recipient. In integration, mismatches are created temporarily in the both junctions due to non complementary sequences in the overlaps of attB and attDOT. These mismatched bubbles are resolved by a mismatch repair system or replication of chromosome. The arrows indicate cleavage sites by IntDOT. CTnDOT genome is shown in red and chromosome in grey. Figure taken from (58).
The model of regulation of CTnDOT excision. Tetracycline stimulates CTnDOT excision by 100 to 1000-fold. The tetQ gene expression is controlled by translational attenuation (TA) of the tetQ promoter. TetQ activates a two component regulatory system, rteA and rteB. RteB activates the rteC protein which acts as a transcriptional regulator of the excision operon containing the orf2c, orf2d, and exc genes. IntDOT is expressed under its own promoter. Figure taken from (49).
Figure 1.5 The arm-type sites of CTnDOT. The attDOT site contains the R1’, R1, R2, R2’ arm-type sites on the right and the L1 and L2 arm-type sites on the left side of the core-type sites, D and D’ shown as grey ovals. The R1’ and L1 sites shown in dark boxes are required for integration of CTnDOT. The R1, R2, and R2’ sites shown as grey boxes are important for intramolecular interactions in an intasome complex. The L2 site shown as a white box is not required for either integration or excision. In excision, B and B’ are represented by dark ovals. The R1’, R1, and L1 sites shown as grey boxes stimulate excision and R2 and R2’ sites are not important to excision of CTnDOT. Figure taken from (82).
Figure 1.6 Sequences of the overlap sites of \textit{attDOT} and \textit{attB}. The GC dinucleotide is indicated in bold. The two overlaps are flanked by two imperfect inverted repeat core-type sites, the D and D’ sites of \textit{attDOT} and the B and B’ sites of \textit{attB}. Five out of seven base pairs are not complementary. The site of first strand exchange shown as arrows are homology dependent and the second strand exchange sites shown as open arrows are homology independent.
Figure 1.7 Diagram of the mechanism of recombination. The overlap regions of two recombining sites are shown in red and blue. Four recombinase monomers are in yellow and an activated pair of monomers is shown in dark yellow inside a circle. Two sets of strand exchanges occur at the one side of overlap at a time. The first strand cleavages form 3’ phosphoryl linkages with the enzymes leaving adjacent free 5’hydroxyl groups. The exposed free 5’hydroxyl groups mediate DNA ligation with the phosphotyrosyl bonds in the partner strands. The first pair of strand exchanges forms a Holliday Junction (HJ) intermediate. In the reaction catalyzed by lambda Int and most tyrosine recombinases, the HJ undergoes a series of conformational changes also called isomerization that require sequence identity in the overlap. The HJ is resolved by the second strand exchanges at the intact strands to produce recombinants containing the hybrid overlaps.
Figure 1.8 The arm domain of the N terminus of factor assisted tyrosine recombinases. A conserved secondary structure contains three strands $\beta_1$, $\beta_2$, $\beta_3$, and helices $H_1$ and $H_2$. $H_2$ is a linker and is also called the coupler located between the N domain and the CB domain.
Figure 1.9 The catalytic domains of the tyrosine recombinase family. The secondary structures of IntDOT (36), Cre (31), lambda Int (40), XerD (71), and Flp (11) are arranged by helices G, H, I, J, K, L, M, N shown in boxes and strands shown in arrows. “Six signature sequences residues”, RIKHIY in the CAT domain are indicated in blue at their corresponding locations. IntDOT contains R1285, K287, H345, Rll348, Hll372, and Y381. Cre has R1173, K201, H1289, Rll292, W315, and Y324. Lambda Int has R212, K235, H1308, Rll311, Hll333, and Y342. XerC/D has R1148, K172, H1244, Rll247, Hll270, and Y279. Flp has R1191, K223, H1305, Rll308, W330, and Y343.
Figure 1.10 **Active site of tyrosine recombinases.** Six catalytically important residues are shown to interact with a scissile phosphodiester bond. Two arginine residues, R₁ and R₂, and a tryptophan/ H₂ residue interact with non-bridging oxygens. Tyrosine acts as a nucleophile containing the hydroxyl group. The H₁ of Cre has been shown to act as general bases and lysine of vaccinia topoisomerase has been proposed to be act a general acid in cleavage reaction. Figure taken from (76).
Figure 1.11  Crystal structure of Cre with HJ. Two pairs of recombinase monomer shown in pink and green bind to a HJ intermediate shown in gold. The helix N of one monomer is donated to the adjacent monomer. These intermolecular contacts help to stabilize the Cre tetramer. Figure taken from (28).
Figure 1.12 Integration and excision of phage lambda. A. Integration and excision are reciprocal reactions depending on the life cycle of phage lambda. Supercoiled phage lambda DNA is integrated by site specific recombination between **attP** in lambda and **attB** in the chromosome. B. Three molecules of excision protein, Xis, bind to the X1, X1.5, and X2 sites in **attR** during excision. Figure taken from (2).
Figure 1.13 The model of recombination mediated by IntDOT. Wild type \textit{attB} contains the GC dinucleotide in the top strand of \textit{attB} and inverted overlap \textit{attB} contains the GC dinucleotide in the bottom strand. IntDOT cleaves top strand of wild type \textit{attB} first whereas IntDOT cleaves bottom strand of inverted overlap \textit{attB} site first adjacent to the GC dinucleotide. Figure taken from (42).
References


78. **Wang, Y., N. B. Shoemaker, and A. A. Salyers.** 2004. Regulation of a Bacteroides operon that controls excision and transfer of the conjugative transposon CTnDOT. J Bacteriol 186:2548-57.


Chapter 2. Structure – function analysis of IntDOT

2.1 Introduction

Conjugative transposons (CTns) are mobile DNA segments that use conjugation and site-specific recombination to transfer a copy of their DNA from a donor to a recipient strain. CTnDOT was originally discovered in a strain of *Bacteroides thetaiotaomicron* that was capable of transferring resistances to tetracycline and erythromycin (4, 35). Upon exposure to tetracycline, CTnDOT excises from the donor chromosome, copies its DNA by rolling circle replication, and transfers its DNA into the recipient cell where it circularizes and is integrated into the recipient chromosome by site-specific recombination. In the last 30 years the frequency of tetracycline-resistant *Bacteroides* isolates has risen dramatically to around 80% of isolates (35). Much of the spread of tetracycline resistance is due to the conjugative transposon CTnDOT and its close relatives (37).

Previous work showed that the integration and excision reactions require the CTnDOT-encoded integrase (IntDOT) and an uncharacterized *Bacteroides* host factor (8, 9, 30, 39). Analysis of the IntDOT amino acid sequence indicated it was a member of the tyrosine recombinase family. It contains five of the six signature residues required for catalysis of the tyrosine recombination reactions (8, 30, 33). I previously constructed and characterized mutants that contain alanine substitutions of residues in the catalytic (CAT) domain that are conserved amongst tyrosine recombinases. The results supported inclusion of IntDOT within the tyrosine family of recombinases. However, the catalytic core seemed to have a somewhat different organization than other tyrosine recombinases.
In addition, I used a homology modeling method to identify residues in the core-binding (CB) domain that are predicted to be near the DNA (29). The results of alanine substitutions of several residues indicated that some residues in the CB domain are likely involved in catalysis.

The information-directed mutagenesis approaches I used previously with IntDOT are useful for producing amino acid substitutions at positions predicted to be important for protein function based on methods such as sequence analysis or homology modeling. Because IntDOT has an arm-binding (N) domain in the N-terminus of the protein about which relatively little is known and because the CAT domain appeared to have an unusual structure not found in other family members, the in vitro approach is limited in its ability to produce useful substitution mutations affecting the functions of these domains. In order to complement our earlier work, I chose to use a structure-function approach similar to one used previously with l Int (21). The strategy I used was to isolate substitution mutants of IntDOT generated by random mutagenesis using an in vivo screen for recombination activity. This approach produced amino acid substitutions in all three of the domains of IntDOT. Analysis of the mutants has uncovered novel amino acid substitutions that cause defects in different steps in the recombination pathway such as DNA binding, DNA cleavage or DNA ligation.

2.2 Materials and Methods

Growth media, strains, chemicals, and enzymes

The strains, plasmids, and oligonucleotides used in this study are listed in Table 1. 

_E. coli_ strains were grown in Luria-Bertani (LB) medium (Difco). _E. coli_ DH5α was
used for cloning and plasmid maintenances. *E. coli* W3110 Δ(*lacI*) (described below) was used for construction of the indicator strain. *E. coli* MC1061 containing pSK2 and pSK2 derivatives were used for production of wild-type and mutant IntDOT proteins. Antibiotics, arabinose, and hydroxylamine were purchased from Sigma. Antibiotic concentrations used were 100 ug/ml augmentin (Aug), 100 ug/ml ampicillin (Amp), 10 ug/ml chloramphenicol (Cam), and 50 ug/ml kanamycin (Kan). X-GAL was obtained from RPI and used at a concentration of 80 ug/ml. DNA modifying enzymes were supplied by New England Biolabs or Invitrogen and reactions were performed as recommended by the manufacturers. Gamma [\(^{32}\text{P}\)]-ATP was purchased from Perkin Elmer, Inc. and T4 DNA kinase from Fermentas. All oligonucleotides were synthesized by Integrated DNA Technologies.

**Construction of pSK1 and pSK2**

A *pir*-dependent plasmid, pJMD100 (12), was constructed that contains the CTnDOT *attB* and *attDOT* sites, the phage λ *attP* site containing the P’3 TEN mutation (32), a chloramphenicol resistance gene (*camR*), the *oriR6K* origin of replication, and the *lacZ*\(^+\) gene. In order to construct pSK1, two DNA fragments produced by PCR amplification were ligated using XbaI restriction sites introduced at the ends of oligonucleotide primers (Table 2.1). The first DNA fragment contained the *lacI*\(^+\) gene amplified from pAH55 (20). The second DNA fragment was amplified from pJMD100 and contained: the CTnDOT *attB* and *attDOT* sites, the phage λ *attP* P’3TEN site, the *camR* gene, and the *oriR6K* origin amplified from pJMD100. The DNA sequences of the λ *attP* P’3TEN site in pSK1 was changed to the wild type *attP* P’3 DNA sequence by site-directed mutagenesis (Stratagene QuikChange kit) so that the pSK1 plasmid could
integrate into the $\lambda$ attB site in *E. coli* W3110 $\Delta$(lacI) through recombination catalyzed by $\lambda$ Int (see below). In order to construct pSK2, the wild-type intDOT gene from plasmid pT7 int (11) was subcloned into pBAD18 between the XbaI and HindIII sites.

**Construction of *E. coli* integration indicator strain**

The lacI$^+$ gene in *E. coli* W3110 was substituted with a gene encoding kanamycin resistance as described by Datsenko et al. (10) using oligonucleotides in Table 2.1 to produce strain JG19000. The deletion of the lacI gene was confirmed by the ability of the mutant strain to form dark blue colonies on LB agar plates supplemented with X-GAL in the absence of IPTG. JG19001 contains the plasmid pBMS2, which contains a pBR322 origin of replication, a gene encoding ampicillin resistance, and the $\lambda$ int gene under control of the arabinose-inducible $P_{BAD}$ promoter as a source of $\lambda$ Int. Plasmid pSK1, which contains the lacI$^+$ gene flanked by the CTnDOT attB and attDOT sites (see above), was integrated into the JG19001 chromosome by site-specific recombination between the $\lambda$ attP site on pSK1 and the attB site in the chromosome of JG19001.

Chloramphenicol resistant colonies containing the integrated pSK1 were white on LB X-GAL plates in the absence of IPTG due to expression on the Lac repressor carried by pSK1. The integrated plasmid was transduced by phage P1 vir into JG19000. A white P1 transductant (JG19002) that was resistant to chloramphenicol and kanamycin and sensitive to ampicillin was selected to be the indicator strain used to screen for IntDOT mutants.
Hydroxylamine-induced random mutagenesis of *intDOT* and Isolation of IntDOT mutants

Plasmid pSK2 DNA (10 ug) was incubated with 7M hydroxylamine in sodium phosphate-EDTA buffer (0.1 M potassium phosphate, 1 mM EDTA [pH 6]) (21) at 37° C for 7 hr. The reaction was stopped by drop-dialysis for 1 hr against TE buffer (10 mM Tris [pH 8]) on 0.025 um Millipore filters. Competent *E. coli* JG19002 cells were electroporated with the hydroxylamine-treated pSK2 DNA. After 2 hr incubation at 37° C in the presence of 0.02 % arabinose to induce IntDOT expression, cells were plated on LB plates containing X-GAL, 0.02 % (w/v) arabinose, and Aug and grown at room temperature for 48 hr. Loss of the integrated *lacI*+ gene due to IntDOT catalyzed recombination was determined by colony color and PCR using oligonucleotides in Table 2.1. Expression of *lacZ* was monitored by screening for blue or white colonies on media containing X-GAL. Colonies expressing active IntDOT protein were blue and gave a PCR product of approximately 500 base pairs due to loss of the *lacI*+ gene. Colonies expressing inactive IntDOT protein were white and produced 1.8 kb PCR products because of the presence of the *lacI*+ gene. White colonies with blue papillae yielded both PCR products. The *intDOT* genes from white colonies containing putative mutants were subcloned into a fresh parental vector and electroporated into fresh JG19002 cells to confirm the recombination-defective phenotype. The *intDOT* gene from each putative mutant subclone was sequenced to confirm the presence of the mutation and to identify the amino acid substitutions.
**Site-directed mutagenesis**

Three pSK2 derivatives containing the R285A, R285D, and R285K substitutions were constructed using the Stratagene QuikChange site-directed mutagenesis kit. Oligonucleotides containing specific mutations are listed in Table 2.1. Mutagenized plasmids were sequenced to confirm the desired mutation. The plasmids were transformed into JG19002 for in vivo integration assays and into MC1061 for overproduction of mutant proteins for biochemical assays.

**Overproduction of proteins**

To express wild type and defective IntDOT proteins, MC1061 cells containing pSK2 and pSK2 derivatives were induced with arabinose. A 50-ml culture was grown in LB supplemented with ampicillin to an OD 0.4 at 37° C and arabinose was added to a final concentration of 0.4%. The cells were induced for 4 hr at room temperature. Cells were harvested and suspended in 0.5 ml of low salt IntDOT lysis buffer (50 mM NaHPO₄, 1 mM EDTA, 50 mM NaCl, 10 % glycerol, 1 mM DTT, [pH 8]). After sonication and centrifugation at 4000 g for 20 min at 4° C, the supernatants containing IntDOT protein were used for in vitro assays described below. Quick Start Bradford protein assays (Bio-Rad) were used to determine the amount of protein in cell extracts. The IntDOT proteins expressed from the wild type of mutant genes were not detected on the SDS polyacrylamide gels.

**DNA binding assay**

A 324 bp fragment containing the functional attDOT site was amplified from the plasmid pJDE2.3 (8) using oligonucleotides in Table 1. The DNA was labeled as described (29) and gel-purified. Thirteen ul of each binding reaction containing 25 nM
labeled attDOT DNA, 15 nM purified E. coli IHF, and 15 ug of total cell extract containing wild type or mutant IntDOT proteins and IHF were performed at room temperature for 20 min in GSBA75 buffer (50 mM Tris-HCl [pH 8], 1 mM EDTA, 50 mM NaCl, 10 % glycerol, 75 ng/ml herring sperm DNA). The mixtures were loaded on native 5% polyacrylamide gels and subjected to electrophoresis for 2 hr at 200 V. The gels were dried and exposed to phosphorimager screens and results were analyzed using Fujifilm Image Gauge software (Macintosh v.3.4). The (++) designation indicates complexes formed as efficiently or nearly as efficiently as with the wild-type protein. The (+) designation indicates decreased efficiency. The (-) designation indicates severely decreased efficiency.

**Phosphorothiolate cleavage assay**

Phosphorothiolate cleavage substrates (30) containing a bridging sulfur at the site of cleavage were prepared by annealing the radio-labeled JG01T-ops and unlabeled JG01B at a 1:10 ratio in an annealing buffer (0.1 M KCl, 10 mM Tris-HCl, [pH 8]) by heating to 90° C and cooling to 22° C at a rate of 5° min⁻¹ using a thermocycler (MJ Research, Inc). Double stranded cleavage substrate (43 pmol) was incubated with 15 ug of cell extract containing wild type or mutant IntDOT proteins in a 12 ul of cleavage reaction mixture (20 mM Tris-HCl [pH 8], 5 mM DTT, 0.05 mg/ml BSA, 1 % glycerol) at 37 °C for 2 hr. Reactions were stopped by addition of 4 ul of 4X sample buffer (200 mM Tris-HCl [pH6.8], 400 mM dithiothreitol, 8 % SDS, 0.4 % bromophenol blue, 40 % glycerol). Samples were boiled for 5 min and protein-DNA complexes were resolved on 4-20 % Tris-glycine gels (30). The gels were exposed to phosphorimager screens and
analyzed as described below in DNA ligation assay. The assays were performed at least in triplicate.

**DNA ligation assay**

The *attDOT* ligation substrate (30) was double stranded DNA that mimics the intermediate produced by DNA cleavage performed by IntDOT. The oligonucleotides used (see Table 1) were annealed to form a double stranded substrate with a 3’- para-nitrophenol (pNP) and 5’–OH at the site of ligation. The end labeled JG02 DNA contains a pNP group at the 3’ end and was annealed with JG02B and JG02T at a ratio of 1:5:10 as described in cleavage assay. Assays were performed in 14 ul of reactions containing 6.5 pmol $^{32}$P-labeled *attDOT* ligation substrate, 50 mM Tris-HCl [pH 8], 1 mM EDTA 65 mM KCl, 10% glycerol, 200 ug/ml BSA, and 15 ug of cell extract as a source of IntDOT proteins. The reactions were stopped after 1 hr incubation at 37$^\circ$ C by adding 14 ul of 90 % deionized formamide in 1X TE (10 mM Tris-HCl, 1 mM EDTA, [pH 8]). Samples were heated for 5 min and the ligated DNA products were detected after electrophoresis through 15 % polyacrylamide TBE buffer-urea gels at 600 V for 2 hr. Gels were exposed to phosphorimager screens and analyzed as described below.

Results for each sample were calculated by dividing the total amount of counts present in the cleavage or ligation product bands by the total amount of counts in the substrate and product bands. Results were reported as the average of at least three independent experiments. (**++) designation indicates 20 or above percent of cleavage or ligation product, (+) designation indicates 0.1 to 20 percent, and (-) designation indicates no detectable product.
Homology Modeling of Int DOT

A three-dimensional structure for the CAT domain of IntDOT was produced essentially as described (29, 40). Briefly the mGenTHREADER fold-recognition algorithm (24, 31) was used to identify sequence-structure alignments between the amino-acid sequence of the IntDOT CAT domain and experimentally determined structures in the Protein Data Bank (PDB). Significant structural conservation was detected with numerous tyrosine recombinases. The most closely related proteins were Integron IntI (25.3 % identical; PDB:2a3v) (28), HP1 Int (20.6 % identical; PDB: 1aih) (23), λ Int (18.4 % identical; PDB :1ae9) (27), XerD (17.1 % identical; PDB:1aop) (38), and Cre (12.1 % identical; PDB: 1xo0) (16). Each pair-wise alignment produced by mGenTHREADER was generally consistent and similarly aligned the active-site residues of each crystallized protein with IntDOT. Each sequence-structure alignment was evaluated for use as a modeling template by threading the IntDOT amino-acid sequence onto each crystal structure using Deep View Swiss-PdbViewer, and analyzing its calculated threading energy. The 4crx Cre structure (18) was selected for model building (also see Results). Because one short region of the alignment (IntDOT residues 284 –292) yielded high threading energy, I produced a hybrid template structure for model building as described previously (40) in which the high-energy portion of the 4crx template structure was replaced with the low-energy portion of the 1p7d structure. This yielded a much lower threading energy for IntDOT in this region, while also preserving the interactions of the template structure with DNA. A three-dimensional all-atom model for the IntDOT CAT domain bound to a core-type DNA site was then constructed from the hybrid template structure using Swiss Model (36), which incorporates molecular
dynamics simulations to minimize the energy of polypeptide backbone and side-chain conformations, and a loop-building process to predict the best position and orientation of amino-acid residues in flexible protein regions. All model calculations were performed using default parameters.

2.3 Results and Discussion

Isolation of IntDOT Mutants

In previous studies, I used the conservation of catalytic residues or homology modeling of tyrosine recombinases as guides to construct IntDOT mutants in vitro. Although these methods can be used to test specific hypotheses, these approaches do not uncover mutants with interesting phenotypes that cannot be predicted. In order to circumvent this limitation, I chose to complement the in vitro studies by using a structure-function approach based on random mutagenesis that I used previously with λ Int (21). In order to perform a structure – function analysis of the IntDOT protein I first needed to isolate mutants generated by random mutagenesis that were deficient in recombination in an in vivo integration assay. Subsequent analysis of the mutants could identify ones that are defective in different steps in the recombination pathway such as DNA binding, DNA cleavage or DNA ligation.

Because of technical difficulties involved with genetic manipulations in Bacteroides, I developed a genetic screen for CTnDOT integrative recombination that functions in E. coli. Since the in vitro reaction functions with IntDOT, IHF, and attDOT and attB substrates (9), I anticipated that the reaction would also function in E. coli cells in vivo. Using the Lac repressor (lacI+) gene in E. coli as a reporter, I developed an
indicator strain (JG19002; See Materials and Methods) to assay for integrative recombination in vivo (Figure 2.1). The indicator strain was engineered to have a single copy of the lacI+ gene flanked by direct repeats of the attDOT and attB sites. Introduction of a plasmid containing a functional intDOT gene into the indicator strain results in recombination between attDOT and attB. This reaction excises the lacI+ gene as a circular product which cannot replicate. The lacI+ gene is lost from the population as the cells divide to form a colony. β-galactosidase is expressed constitutively in colonies with cells that have lost the lacI+ allele and can be detected by plating cells on LB plates supplemented with X-GAL. Since the E. coli in vivo integration assays uses indicator plates to screen for mutants, it is practical to screen for mutants out of a large population of colonies.

I used the in vivo genetic screen to isolate a large collection of recombination-defective IntDOT mutants produced by random mutagenesis with hydroxylamine. After electroporation of the JG19002 indicator strain with mutagenized plasmid DNA, a wide range of colony phenotypes was observed: blue colonies, white colonies, and white colonies containing varying amounts of blue papillae. Colonies containing wild-type IntDOT were blue or white with blue papillae. Colonies containing severely defective IntDOT proteins were white or mostly white because the lacI+ gene remained in the genome of all cells in the colony and Lac repressor repressed synthesis of β-galactosidase. As expected, a plasmid containing an intDOT gene with a substitution of the catalytic tyrosine for phenylalanine (Y381F) produced white colonies in this assay. I picked the white regions of colonies that contained few or no papillae and streaked them on plates supplemented with X-GAL and observed the color patterns of several
independent colonies. Colonies that yielded white or a mixture of white and papillated colonies when streaked were saved as potential mutants. Colonies that produced both blue and sectored colonies were assumed to contain plasmids with wild-type intDOT genes and were not analyzed further. To ensure that these phenotypes result from IntDOT recombination, the presence or absence of the \textit{lacI} \textsuperscript{+} gene was confirmed by colony PCR. All putative mutants were backcrossed into JG19002 to confirm that the phenotypes were associated with the plasmid.

Using this approach I isolated twenty-five mutants with single substitution mutations in the \textit{intDOT} gene from approximately 3,000 transformants screened. This resulted in a 0.83\% mutation frequency. As predicted by the known in vitro specificity of hydroxylamine, all of mutants contained C:G to T:A transition mutations. Twenty four of the mutants contained a single missense mutation and one contained an amber mutation. Two mutants, R348H and H372Y, affected residues that are conserved in the catalytic domain of tyrosine family recombinases (13, 30, 33). The other twenty-three substitution mutations were distributed throughout the three domains of the protein. The mutants and their amino acid substitutions are listed in Table 2.2.

**IntDOT secondary structure**

Tyrosine recombinases often contain three domains: an arm-binding (N) domain, a core-binding (CB) domain, and a catalytic (CAT) domain. Because the integration and excision reactions of IntDOT are regulated I expected that IntDOT would also contain an N domain. I predicted the secondary structure of IntDOT using the PROF sec (34). This prediction agreed with the previous secondary structure prediction I reported for the IntDOT CB domain using the PSIPRED algorithm (25, 29). Figure 2.2 summarizes the
predicted secondary-structure elements of IntDOT, the organization of these elements into the three domains typically found in several tyrosine recombinases, and the location in the protein of each amino acid substitution shown in Table 2.2. The putative IntDOT N domain spans approximately 90 residues and is predicted to contain a three-stranded sheet (β1, β2, β3) followed by an α-helix (H1). This arrangement is also found in both λ Int and Tn916 Int although the primary sequences of this region of the three proteins are not related. The protein also contains a second putative α-helix (H2) extending from residues 92 to 107, which is analogous to the “coupler” region of λ Int (5).

I previously predicted that the IntDOT CB domain comprises residues 108 through 220 and contains four major alpha-helices, (A, B, C, D) (29), arranged in the orthogonally crossed conformation that is typical of tyrosine recombinases (40). It is not clear whether the IntDOT CB domain also contains a fifth α-helix (E, near residues 209-213) similar to that found in Cre (19) and Flp (7); therefore in this report I have not labeled any of the helices “E” in order to avoid conflicts in nomenclature.

The predicted CAT domain of IntDOT contains residues 222 to 411. It contains nine helices (F, G, H, I, J, K, L, M, and N) and three β-strands (β3, β4, and β5). The λ Int (27), Cre (19), and Flp (7) proteins contain between 7 and 14 helices and between 5 and 7 strands in their CAT domains, depending on the protein. Through multiple-sequence alignment, I previously identified five of the six signature active-site residues of tyrosine recombinases within the CAT domain of IntDOT (8, 30). These include K287 in the loop between helices H and I, H345 in the loop between helices J and K, R348 in the helix K, H372 in the loop between helices L and M and the catalytic tyrosine Y381 in helix M. However, IntDOT is missing the first conserved arginine (Arg I) which is present in
other tyrosine recombinases (30). The residue at the equivalent position (259) in IntDOT is serine. Substitution of the serine with alanine has no effect on recombination when mutated (30). Our experimental and modeling results suggest that the Arg I function of IntDOT is provided by an arginine from another location in the CAT domain (see below).

**Homology Modeling of the CAT Domain**

A homology model for the IntDOT CAT domain was constructed based on the structural conservation I detected between IntDOT and several other tyrosine recombinases. Cre was selected for model building because (1) it provided a low overall threading energy, (2) known active-site residues were aligned correctly (except for Arg I; see below), (3) insertions and gaps in the alignment did not disrupt secondary-structure elements that are strongly conserved among tyrosine recombinases (33), (4) co-crystal structures with Cre allow interactions between the modeled protein and a bound DNA site to be investigated, and (5) I previously obtained good results using Cre to model the IntDOT CB domain (29). Although Cre shares a low sequence identity with IntDOT CAT domain (12.1%), I found previously that modeling of tyrosine recombinases is useful at this level of identity (29, 40). However, the model has certain limitations; for example, the true path of the polypeptide backbone may differ by as much as 3-4 angstroms (on average) from the model, and exact conformations of amino acid side chains in variable protein regions (e.g. on DNA-binding surfaces) can not be predicted reliably (40). Nevertheless, the strong conservation of CAT domain structure and active-site configurations among tyrosine recombinases (33, 45) suggests that the model can provides a useful approximate representation of the overall IntDOT CAT domain and the position of its active-site residues within this structure.
Functional Analyses of IntDOT Mutants

I analyzed the mutant proteins for their ability to bind DNA and perform different steps in the recombination reaction using assays described previously (30). In order to undergo recombination, IntDOT, the host factor, and attDOT DNA form a nucleoprotein structure called the integrative intasome. The host factor is an unknown Bacteroides protein but I showed that E. coli IHF substitutes for the host factor in an in vitro recombination assay (8, 9). IHF likely binds non-specifically and bends attDOT DNA to act in concert with IntDOT to form the integrative intasome. I also showed previously that IntDOT and IHF form a complex with attDOT DNA that can be detected by a gel-shift assay (Figure 2.3). IntDOT alone does not shift attDOT efficiently. In the presence of both IntDOT and IHF, a super-shift containing both IntDOT and IHF is seen (Figure 2.3). Because the complex requires IHF which bends DNA, it is possible that the complex contains a monomer of IntDOT that interacts with a single DNA molecule containing attDOT through binding of its N-domain to an arm-type site and binding of its CB-domain to a core-type site (See below). The complex could contain one or more other monomers of IntDOT. The gel-shifts performed with the mutant proteins and IHF are shown in Figure 2.3.

To analyze the abilities of the proteins to carry out catalytic steps in the recombination reaction, I used DNA cleavage and ligation assays. The cleavage assay utilizes a suicide substrate containing a bridging Phosphorothiolate at one of the cleavage sites where strand exchange is initiated (Figure 2.4A). Cleavage of the substrate by IntDOT produces a 5’-phosphotyrosine linkage with the protein leaving behind a 5’SH group that is a poor nucleophile. Thus, it cannot attack a 3’ phosphotyrosyl bond to
release the enzyme from the DNA. This reaction results in the irreversible covalent attachment of the protein to DNA through its catalytic Tyr 381 (30). The cleavage assays performed by the wild type and mutant IntDOT proteins are shown in Figure 2.4B.

The ligation assay utilizes an activated substrate containing a 3’-para-nitrophenol adjacent to a free hydroxyl at the site of catalysis in the DNA (Figure 2.5A). Ligation of the substrate releases the para-nitrophenol group and forms a covalent phosphodiester bond in the DNA. This ligation product can be resolved from unreacted DNA in a denaturing gel. This ligation reaction does not require the catalytic Tyr381 (30). The ligation reactions performed by the wild type and mutant IntDOT proteins are shown in Figure 2.5B.

**Mutants with Substitutions in the N-domain**

Among mutants with substitutions in the N-domain, the R13C protein shows a strong phenotype and S38N a leaky phenotype in the in vivo integration assay. Although the R13C and S38N proteins show a diminished ability to form the supershifted complex, the mutants retain substantial cleavage and ligation activities indicating that they are still catalytically active. The observation that these mutants are defective in forming the supershifted complex with IHF and attDOT DNA is consistent with a defect in DNA binding or protein–protein interactions. NMR studies of the λ Int N domain show that it contains an unstructured N-terminal tail and three β-strands that interact with an arm-type site (14, 44) by making sequence-specific contacts with the major groove of the DNA. Our secondary structure prediction for the N domain of IntDOT suggests that it contains a similar three β-strand motif but lacks an N-terminal tail. Thus it is possible that the β 1
strand containing R13 and the β 3 strand containing S38 may have a role in binding to arm-type sites by IntDOT.

The V95M mutant is defective and G101R mutant is leaky in the in vivo integration assay. The V95M substitution is located in a predicted α-helix (H1) that joins the N domain to the CB domain, and the G101R is also near that helix. In λ Int this region contains an α-helix and is known as the “coupler” region (5). The V95M and G101R proteins also showed a diminished ability to shift attDOT DNA in the presence of IHF. However, the V95M and G101R proteins did form complexes with mobilities similar to the complex with wild type protein and IHF and were catalytically active in the cleavage and ligation assays. Warren et al (42) showed that residues in the coupler region of λ Int are involved in protein-protein interactions that are essential for cooperative binding to arm sites. Taken together, the position of V95M and G101R in a putative “coupler” region of IntDOT and their inability to form nucleoprotein complexes suggests that these mutants are defective for protein-protein interactions or interactions with DNA.

**Mutants with Substitutions in the CB domain**

The T184I and P209L proteins contain substitutions that introduce hydrophobic residues into the CB domain of IntDOT. The T184I and P209L mutant proteins are partially defective in all of the in vitro assays. However, they are distinguished by their effect on in vivo integration; T184 is a strong mutant while P209 is leaky mutant. T184 lies in helix D of the CB domain and is equivalent to residue S139 of λ Int (40). This position is the highly conserved amongst tyrosine recombinases where the preferred residues are serine or threonine (40). In our homology based model of λ Int CB domain,
λ. Int S139 is very close to the DNA backbone directly across from the active site, and this position was predicted to play an important mechanistic role during the recombination process (40). The strong effect of the T184I substitution on recombination catalyzed by IntDOT supports this prediction and provides additional experimental evidence for the importance of this conserved residue in recombination.

The P209L protein is also defective in all of the in vivo and in vitro assays. Interestingly the P209 residue is located in the “linker” region connecting the CB and CAT domains. In the XerD crystal structures (38), this linker region is disordered and may differ in structure depending on interactions with DNA or other proteins in the intasome. Removal of a proline in this region is expected to shift the path of the peptide backbone significantly, and may thus have a significant effect on the linker region structure.

**Mutants with Substitutions in the CAT Domain**

The W280 amber mutant is predicted to form a truncated protein that contains both the N and the CB domains but lacks a large portion of the CAT domain. The mutant is completely inactive in the in vivo integration reaction as well as the cleavage and ligation reactions as expected. The protein did not form the supershift with IHF and attDOT DNA (Fig. 2.3), which is consistent with the predicted requirement for binding of the CAT domain to a core-type DNA site in the supershifted complex. Since the complex requires IHF, which bends DNA, I previously suggested that bending of the DNA by IHF allows a single IntDOT monomer to simultaneously form an intramolecular bridge with an arm-type site through the N domain and a core-type site through interactions with the
CB and CAT domains (29). However, I cannot rule out the possibility that the truncated protein is rapidly degraded by cellular proteases.

All of the remaining substitutions, which comprise the majority of the mutants isolated in this study, lie in the CAT domain. The T256I, V319I, C325Y, L326F, R330H and S347N proteins have defects in all of the in vivo and vitro assays. These proteins are defective in forming specific complexes with IHF and some (T256I, W280, C325Y, L326F, and R330H) actually appear to prevent the formation of the specific complex that contains only IHF and \textit{attDOT} DNA (Fig. 2.3). This result is reproducible and has been obtained for all five mutants with several independent extracts. I do not know the mechanistic basis for these results but the proteins could interact non-specifically with \textit{attDOT} DNA to inhibit binding of IHF or form complexes that are not stable. Most of the mutants isolated appear to have specific effects on discrete protein functions necessary for recombination. However, the V319I and S347N mutant proteins were defective to some extent in all assayed functions, suggesting that the proteins might not be expressed well, misfolded, or degraded.

The substitutions in the A352T and T354I proteins lie within putative \(\alpha\)-helix K. The A352T protein is defective and the T354I protein is leaky in the in vivo integration assay, but both of these mutant proteins formed super-shift complexes with IHF and ligate DNA. Interestingly, the A352T protein is defective in cleavage activity without apparent alteration of its ligation activity whereas T354I protein cleaves DNA as well as wild type protein. In the model, the A352 residue is located in the hydrophobic core of the protein. Therefore, it appears that the non-conservative polar A352T substitution alters the CAT domain structure by interfering with hydrophobic packing in the domain.
Because A352 is located near the active site, the A352T substitution probably disrupts recombination by altering the position of active-site residues, such as H345 and R348 that reside nearby in the same α-helix. The T354I substitution is also located in the same α-helix and may have a similar effect; however, the effect of T354I may be diminished because this position is located farther from the active site. It is unlikely that substitutions in the A352T and T354I proteins destabilize the CAT domain since the mutant proteins bind DNA as well as the wild-type protein.

The T365M, S367N, and S368F substitutions change residues that are located in putative α-helix L. The T365M is defective and the S367N and S368F proteins are leaky mutants in in vivo integration assay. The T365N and S368F proteins are active in the gel-shift, cleavage, and ligation assays. The observation that these mutants retain significant activity in most assays indicates that the substituted residues are not severely disruptive. The S367N protein forms a super-shifted complex but shows a significant reduction of cleavage and ligation activity. In the model these three residues are located proximal to the conserved catalytic H372 residue, and thus the substituted residues may affect recombination and catalysis by altering the position of H372. Additionally, the three residues are located on or near the exterior surface of the protein and the substituted residues may therefore interfere with inter-protein interactions within the intasome.

The G371E and H372Y proteins are defective in most assays and contain substitutions that affect adjacent residues in IntDOT. This glycine–histidine dyad, located in a loop between helix L and M in IntDOT, is conserved in λ Int and other tyrosine recombinases (33), and in λ Int the glycine–histidine dyad is known to interact with DNA (27). H372 aligns with the conserved His II residues in other tyrosine
recombinases and thus is likely to be a catalytic residue located in the active site as shown in the model (Figure 2.7). I showed previously that the H372A mutant protein is also inactive in recombination and lacks detectible cleavage and ligation activities (30).

Substitution of the G371 residue is expected to reduce the flexibility of the protein backbone at this location. Due to the proximity of G371 to the active site residue H372, it seems probably that the glycine to glutamic acid substitution disrupts recombination by altering the position of H372. However, it is also possible that introduction of the glutamic acid residue, and not loss of the glycine, is responsible for the mutant phenotype. Our model suggests that the mutant glutamic acid residue is proximal to the DNA and may thus interfere with DNA binding through electrostatic repulsion with the negatively charged DNA backbone. Alternatively, the negatively charged glutamic acid might interact with the nearby positively charged residues of the active site to disrupt reaction chemistry.

The substitution in the A382V protein affects the residue adjacent to the catalytic tyrosine Y381 in helix M. A382V is defective for recombination in vivo but is proficient in most of the in vitro assays. In the model, A382 is located on the back side of the helix containing Y381, and the valine substitution could alter the positioning of this critical residue without disrupting DNA binding by other parts of the protein.

The L389F substitution affects a residue in the C-tail of the protein. The L389F protein showed reduced recombination in vivo integration and normal activity in the DNA binding and cleavage and ligation assays. In our model, L389F is located on the extensible “tail” that interacts with another recombinase monomer in the IntDOT tetramer, so the L389F substitution appears unlikely to have a direct effect on catalytic
activity. In λ Int, a mutant protein lacking the last 8 C-terminal residues is defective for recombination but has increased topoisomerase activity suggesting that the C-terminal tail is involved in regulation of catalysis (26). Subsequent work (22) on the C-terminal tail of λ Int shows that this region is essential for inter-molecular protein-protein interactions required for coordinated Holliday Junction (HJ) resolution.

**R285 may function in catalysis**

The active sites of tyrosine recombinases contain six signature residues (RKHRHY) that are involved in catalysis. The two conserved arginines (Arg I and Arg II) stabilize an intermediate in cleavage reaction and are essential to recombinase activity (3, 41). Our previous attempt to identify the Arg I residue through multiple-sequence alignment was unsuccessful because the position that corresponds to Arg I in other tyrosine recombinases is a serine (S259) in IntDOT (30). The S259 residue does not appear to be involved in catalysis because an alanine substitution (S259A) at this position did not affect recombination in vivo, while an arginine substitution (S259R) interfered with recombination (30). A second nearby arginine residue was also identified (R247), but an alanine substitution at this position showed that R247 also was not required for recombination (30). The lack of a conserved Arg I residue in IntDOT was particularly interesting because Arg I substitutions in λ Int, Cre and Flp are all defective in catalysis (1, 2, 6, 21, 30, 33, 43). Our conclusion was that an arginine residue in another region of the CAT domain might substitute for the missing one (30). In this study, I identified two non-conserved arginine residues in our mutant collection, R295 and R285, which are candidates for the missing catalytic arginine of IntDOT.
The R295H protein was defective for integration in vivo but was proficient in the DNA binding, cleavage and ligation activities in vitro. Since the catalytic activities of R295H seem to be as functional as the wild-type protein, it is unlikely that R295 is involved in catalysis. This prediction is supported by the model, which shows that R295 is on the opposite side of the CAT domain from the active site and DNA binding interface.

On the other hand, the R285H mutant protein was defective for recombination, cleavage, and ligation. However it formed the supershift complex like the wild-type protein. These results suggest that R285 is involved in catalysis and could be functioning in a manner similar to the Arg I residues in other tyrosine recombinases. To investigate the effects of other residues at position 285, I constructed and characterized proteins with substitutions to alanine (R285A), aspartic acid (R285D) and lysine (R285K; see Figure 2.6). All three mutant proteins were defective in recombination in vivo and were unable to cleave DNA. However, the mutant proteins were able to form the IHF-dependent complexes in the gel-shift experiment similar to the wild-type protein, indicating they are capable of binding DNA and forming higher-order nucleoprotein structures.

Mutant proteins with substitutions of the Arg I residues of Cre (173) and Flp (191) have been constructed and analyzed. Most of phenotypes were similar to those of the IntDOT R285K protein. The Cre R173K protein was defective in recombination and was able to bind to DNA (1). The Flp R191K protein bound DNA but was defective in ligation and in completing recombination (15). However, the Flp R191K protein was able to cleave a full site efficiently (6) while the IntDOT R285K protein cleaved a full site inefficiently. Since the substrates used in the Flp studies did not contain a nick while
the substrate used in our assay with IntDOT R285K contained a nick in the core, the results cannot be directly compared. Thus, with the possible exception of cleavage activity, the IntDOT R285K protein has a phenotype similar to the Cre R173K and the Flp R191K proteins.

The Arg I residues of Cre (R173), Flp (R191), and λ Int (R212) are located in the loop between α-helices G and H (Fig. 2.2). In contrast, R285 of IntDOT is predicted to be at the end of β4 in the CAT domain. Despite the unusual position of R285 in the polypeptide backbone, our homology model predicts that R285 can enter the active site and interact with bound core-type DNA in a manner similar to the Arg I residue of other tyrosine recombinases (Figure 2.7). The position and predicted role of R285 provided independently by the model agrees with our biochemical evidence demonstrating the importance of R285 for catalysis by IntDOT. If our model is correct and Arg285 substitutes for Arg I of other tyrosine recombinases, this might be an example of a fundamental difference in the architecture of the catalytic site of IntDOT relative to other tyrosine recombinases.

It is not unexpected to find major structural differences between IntDOT and other recombinases in the region containing IntDOT-R285. Previous alignment of tyrosine recombinase sequences showed weak amino acid conservation in this region (33), despite the proximity of a conserved active site residue, λ Int-K235 (IntDOT-K287). Inspection of co-crystal structures containing recombinase bound to DNA (e.g. Cre (17) and IntI (28)) shows that this region comprises a pair of anti-parallel beta-strands (β 4 and β 5 in Figure 2.2) that directly interact with DNA near the active site. This protein structure is apparently flexible because it can assume different conformations in different
protein-DNA complexes. For example, IntI-K160 (which corresponds to λ.Int-K235 and IntDOT-K287) is either found interacting with DNA or exposed to solvent, depending upon the recombinase monomer that is examined (28). In another example, this region of Cre was disordered in a co-crystal with a Holliday junction (17). Collectively, this information indicates that the IntDOT region containing residues 279 to 297 is likely to diverge from other recombinase sequences and may therefore contribute to the novel active site configuration proposed in this report.

Another example of variation in active-site architecture is displayed by Flp protein. In most tyrosine recombinases, the six conserved active site residues are located on the same monomer and the protein cleaves DNA “in cis” during the recombination reaction. In Flp the tyrosine nucleophile of the active site is located on a different monomer from the other five conserved residues. Thus the active site is formed by residues from two different monomers and the DNA is cleaved “in trans” during the recombination reaction (7).

2.4 Conclusion

This Chapter 2 presents a genetic approach to identify functional residues that are essential to various steps of recombination catalyzed by IntDOT. Using random mutagenesis, I have identified residues that are important for IntDOT function throughout the entire protein. Surprisingly, the Arg I active-site residue that is absolutely conserved in other tyrosine recombinases could be provided by a non-conserved arginine residue, R285, that is located in a different part of the protein. This intrinsic difference in protein structure between IntDOT and other tyrosine family members could contribute to the
unique mechanism of IntDOT mediated recombination. It will be interesting to solve the structure of the IntDOT catalytic site to better understand the role of the unusual R285 residue in catalysis.
**Figures and Tables**

**Figure 2.1** Screen for IntDOT mutant proteins with the *E. coli* indicator strain. (A) Structure of the plasmid pSK1. Plasmid pSK1, a pir-dependent plasmid, contains the lacI gene between direct repeats of *attDOT* and *attB<sup>DOT</sup>, camR, and λ *attP* site. This plasmid was integrated into JG19001 by site-specific recombination between λ *attP* site in pSK1 and λ *attB* on chromosome resulting in white chloramphenicol-resistant colonies. (B) Integrated version of pSK1. JG19002 contains the *lacI* + gene flanked by direct repeats of *attB<sup>DOT</sup>* and *attDOT*. When IntDOT-mediated recombination occurs between the *attB<sup>DOT</sup>* and *attDOT* sites, the *lacI* + gene is excised on a circular element that cannot replicate. The resulting colony is blue on a plate supplemented with X-GAL because the *lacI* + gene is lost from the cell. Cells expressing defective IntDOT proteins remain white (if the mutant protein is inactive) or white with blue papillae (if the mutant protein retains some activity). To ensure that phenotypes result from IntDOT recombination, the presence or absence of the *lacI* + gene in the chromosome was analyzed by colony PCR as described in the Materials and Methods.

71
Figure 2.2 Distribution of Mutational Changes in IntDOT. The secondary structure of IntDOT was predicted by PROF sec. The IntDOT secondary structure elements have been named to correlate with those of the Cre recombinase. The IntDOT protein is predicted to have three DNA binding domains called the N (purple), CB (dark blue) and CAT (light blue). The β1, β2, β3 strands and H1 and H2 α-helices lie in the N-domain. The CB domain contains four α-helices (helices A, B, C, and D). The CAT domain contains nine α-helices (helices F, G, H, I, J, K, L, M, and N) and three β-strands (β4, β5, and β6). The positions at which mutations were obtained are shown. Y381F was constructed by site-directed mutagenesis. Leaky mutants are represented in blue and strongly defective mutants in red.
Figure 2.3  *attDOT* binding by the mutant IntDOT proteins. The 324 bp *attDOT* DNA fragment contains the minimal *attDOT* site required for recombination. Protein-DNA complexes containing bound IntDOT and IHF migrate more slowly that complexes containing only IHF. The (++) designation indicates complexes with mutant proteins were formed as efficiently or nearly as efficiently as with the wild-type protein. The (+) designation indicates a decreased efficiency of complex formation. The (-) designation indicates severely decreased efficiency of complex formation. Dots above each protein name indicate whether the substituted position is conserved among tyrosine recombinases. Leaky mutants are designated in blue type and strong mutants in red type.
**Figure 2.4** DNA cleavage by the mutant IntDOT proteins. (A) A synthetic oligonucleotide containing a bridging phosphorothiolate at the site of top strand cleavage was labeled at 5'-end with $^{32}$P(*) and annealed with a complementary bottom strand to generate a “suicide” substrate (see Materials & Methods). When the substrate reacts with IntDOT, the cleavage product containing IntDOT covalently bonded to the DNA accumulates because the reaction is irreversible. (B) DNA substrate and protein-DNA products were resolved on 4-20% Tris-Glycine gels. The (++), (+), and (-) designations are described in the Materials and Methods. Leaky mutants are designated in blue type and strong mutants in red type. Dots above each protein name indicate whether the substituted position is conserved among tyrosine recombinases.
Figure 2.5 DNA ligation by the mutant IntDOT proteins. (A) An oligonucleotide containing a 3’- para-nitrophenyl group (pNP) was radio-labeled at the 5’-end (*). The labeled oligo containing pNP was annealed with the indicated top and bottom strands to form the DNA ligation substrate. When IntDOT ligates the DNA ends to form a phosphodiester bond the pNP group is released. The 44 base-pair labeled DNA molecule is detected on a 15% denaturing polyacrylamide gel. (B) The (++), (+), and (-) designations are described in the Materials and Methods. Dots above each protein name indicate whether the substituted position is conserved among tyrosine recombinases. Leaky mutants are designated in blue type and strong mutants in red type.
Figure 2.6 DNA binding and cleavage of the IntDOT variants altered at Arg-285. (A) DNA binding by Arg-285 mutants of IntDOT protein containing substitutions of alanine (A), aspartic acid (D), lysine (K), and histidine (H) at residue 285. Cell extracts containing mutant proteins and DNA substrates used in gel-shift assays were used as indicated in Fig. 2.3 and the Materials and Methods. Super-shifted products in the presence of IHF were analyzed on a 5% native polyacrylamide gel. (B) DNA cleavage by Arg-285 mutants of IntDOT. This substrate contained a nick at the cleavage site of top strand as described in the Materials and Methods. Cleavage products were analyzed on a 4-20 % Tris-glycine gradient gel.
Figure 2.7  Predicted amino acid residues and conformation of the IntDOT active site. A cartoon representation is shown of the IntDOT CAT domain bound to a core-type DNA site. The DNA backbone is shown as a grey tube. The protein backbone is shown in blue and yellow, with α-helices represented by cylinders, β-strands by arrows, and unstructured coil regions by tubes. The region of IntDOT (residues 279 to 297) that interacts with DNA and contains active-site residues R285 and K287 is shown in yellow. The side chains of each predicted active site residue in IntDOT are shown in orange, except for R285, which is shown in red. The structure and position of a typical conserved “Arg I” residue, in this case R212 from λ Int (which aligns with IntDOT-S259), has been superimposed on the IntDOT structure and is shown in purple. It can be seen that the IntDOT-R285 and the superimposed λ Int-R212 residue enter the active site from different positions on the peptide backbone; however, they both can interact with a similar part of the DNA backbone. A single, specific orientation of the IntDOT-R285 side chain is shown in the model; however, multiple conformations of the R285 side chain are possible that can interact with DNA.
Table 2.1 Strains and oligonucleotides used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relative genotype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>W3110</td>
<td>Wild type</td>
<td>Lab strain</td>
</tr>
<tr>
<td>JG19000</td>
<td>W3110(ΔlacI)Kan</td>
<td>This work</td>
</tr>
<tr>
<td>JG19001</td>
<td>W3110(ΔlacI)Kan/pBMS2</td>
<td>This work</td>
</tr>
<tr>
<td>JG19002</td>
<td>W3110(ΔlacI)Kan:pSK1</td>
<td>This work</td>
</tr>
<tr>
<td>MC1061</td>
<td>hsdR2 hsdM+ hsdS+ araD139 Δ(ara-leu)7697 Δ(lac)X74 galE15 galK16 rpsL (StrR) mcrA mcrB1</td>
<td>Lab strain</td>
</tr>
<tr>
<td>DH5α</td>
<td>F' endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+, λ–)</td>
<td>GibcoBRL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJMD100</td>
<td>Contains attDOT, attB&lt;sup&gt;DOT&lt;/sup&gt;, λ attP&lt;sup&gt;3&lt;/sup&gt;, oriR6K, camR</td>
</tr>
<tr>
<td>pAH55</td>
<td>Contains lacI</td>
</tr>
<tr>
<td>pBMS2</td>
<td>Contains λ int downstream of the ara pBAD promoter and ampR</td>
</tr>
<tr>
<td>pT&lt;sub&gt;7&lt;/sub&gt; int</td>
<td>pT&lt;sub&gt;7&lt;/sub&gt; int is pET-30(a) plasmid that contains intDOT gene between NdeI and HindIII sites</td>
</tr>
<tr>
<td>pSK1</td>
<td>attDOT, attB&lt;sup&gt;DOT&lt;/sup&gt;, λ attP, oriR6K, cam, lacI,</td>
</tr>
<tr>
<td>pSK2</td>
<td>The pSK2 is pBAD18 containing the intDOT gene from the plasmid</td>
</tr>
<tr>
<td>pSK2 derivatives</td>
<td>The pSK2 derivatives are pBAD18 containing the wild-type or mutant intDOT genes between Xbal and HindIII sites cloned downstream of the ara pBAD promoter.</td>
</tr>
</tbody>
</table>

**Oligonucleotide Sequences (5'→3')**

Primers used for amplification of the 324 bp attDOT region

<table>
<thead>
<tr>
<th>Oligonucleotides used in cleavage assay&lt;sup&gt;1,2&lt;/sup&gt;</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRJ/MM160F</td>
<td>TCGGGCATGTCAGTCAGAC</td>
</tr>
<tr>
<td>DRJ/AG161R</td>
<td>CTCGAATTAAATAGCTCTTTTG</td>
</tr>
</tbody>
</table>

Oligonucleotides used in ligation assay<sup>1,2</sup>

| JG01T-ops                                           | *CAGCTTTGGTATGGTAACCTG-TG -ops-GCttagGAAAGTTAAGAGAGT | (30) |
| attDOT -T4                                         | *CAGCTTTGGTATGGAACCTG | (30) |
| JG01T                                              | TTAGCTGAAAGTTACGAGAGAGT | This study |
| JG01B                                              | ACTCTTTTCGTTAATCTCGTcaataGCAAGGTTTACTCAAAAAAGTG | (30) |

Oligonucleotides used to change λ P<sup>3</sup> ten site to wild type λ attP site

<table>
<thead>
<tr>
<th>Primers used to change λ P&lt;sup&gt;3&lt;/sup&gt; to wild type λ attP site</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P&lt;sup&gt;3&lt;/sup&gt; to wt F</td>
<td>AGG TCA CTA TCA GTC AAA ATA AAA TCA TTA TTT GAT TTC AAT TTT GTC CCA CAG GCC GGC CTG TGG GAC AAA ATT GAA ATC AAA TAA TGA TTT TAT TTT GAC TGA TAG TGA CCT (underlines indicate wild type λ attP sites)</td>
</tr>
<tr>
<td>P&lt;sup&gt;3&lt;/sup&gt; to wt R</td>
<td>This work</td>
</tr>
</tbody>
</table>

Primers used for amplification of 1.5 kb camR region of pJMD100

<table>
<thead>
<tr>
<th>attDOT-Xbal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJMD99/100</td>
<td>5'GGTCCTAGAACCGCAATCCGACAGCAAAAGTG</td>
</tr>
<tr>
<td>Rev-Xbal</td>
<td>5'GCCTTAGAACCGCCTTGACGCAAAAGTG</td>
</tr>
</tbody>
</table>

(underlines indicate Xbal sites)
Table 2.1 (continued)

<table>
<thead>
<tr>
<th>Primers used for amplification of the lacI gene of pAH55</th>
</tr>
</thead>
<tbody>
<tr>
<td>F:lacI-XbaI</td>
</tr>
<tr>
<td>R:lacI-XbaI</td>
</tr>
<tr>
<td>GCT CTAGAGCCTCCGGAGCTGACATGTGTC</td>
</tr>
<tr>
<td>GCTCTAGAGCTCATT ACTGGACCGGATCTG</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>(underlines indicate XbaI sites)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers used for amplification of kanR flanked by lacI gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEL_lacI:kan(F)</td>
</tr>
<tr>
<td>CCG GTG TCT CTT ATC AGA CCG TTT CCC GCG GTG</td>
</tr>
<tr>
<td>TAG GCT GGA GCT GCT TC</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>DEL_lacI:kan(R)</td>
</tr>
<tr>
<td>TCA CTG CCC GCT TTT CAG TCG GGA AAC CAT CAT</td>
</tr>
<tr>
<td>ATG AAT ATC CTC CTT AG</td>
</tr>
<tr>
<td>This work</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers used for colony PCR to detect the lacI in the chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>F:pSK1_2571</td>
</tr>
<tr>
<td>CGGAAATTCCGGGTATTCGATGTGTT</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>R:pSK1_600</td>
</tr>
<tr>
<td>CGGAAATTCCGGGATCCTCCTAGTA</td>
</tr>
<tr>
<td>This work</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers used for site-directed mutagenesis of R285 variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_R285A</td>
</tr>
<tr>
<td>TGG ATA AAC ATC AAG CAC AAA ACG GGT GTG</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>R_R285A</td>
</tr>
<tr>
<td>CAC ACC CTG TTT CTG GGC GTT GAT GTT TAT CCA</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>F_R285K</td>
</tr>
<tr>
<td>TGG ATA AAC ATC AAG CAA CAG AAA ACG GGT GTG</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>R_R285K</td>
</tr>
<tr>
<td>CAC ACC CTG TTT CTG GTT GAT TAT CCA</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>F_R285D</td>
</tr>
<tr>
<td>TGG ATA AAC ATC AAC GAC CAG AAA ACG GGT GTG</td>
</tr>
<tr>
<td>This work</td>
</tr>
<tr>
<td>R_R285D</td>
</tr>
<tr>
<td>CAC ACC CTG TTT CTG GTT GAT TAT CCA</td>
</tr>
<tr>
<td>This work</td>
</tr>
</tbody>
</table>

1) Oligonucleotides containing a $^{32}$P label are represented by an asterisk (*).
2) Coupling sequences are shown in lower case.
<table>
<thead>
<tr>
<th>Domain</th>
<th>IntDOT mutants(^1)</th>
<th>Base changes(^2)</th>
<th>(E.\text{coli}) in vivo integration</th>
<th>binding</th>
<th>cleavage</th>
<th>ligation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>pBAD18 only</td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>R13C</td>
<td>CGT→TGT</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>S38N</td>
<td>AGC→AAC</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>V95M</td>
<td>GTG→ATG</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>G101R</td>
<td>GGA→AGA</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>CB</td>
<td>T184I</td>
<td>ACC→ATC</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>P209L</td>
<td>CCG→CTG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CAT</td>
<td>T256I</td>
<td>ACC→ATC</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>W280</td>
<td>TGG→TAG</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R285H</td>
<td>CGC→CAC</td>
<td>-</td>
<td>+/++</td>
<td>-/+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R285A (SDM)</td>
<td>CGC→GCC</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-/+</td>
</tr>
<tr>
<td></td>
<td>R285D (SDM)</td>
<td>CGC→GAC</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R285K (SDM)</td>
<td>CGC→AAA</td>
<td>-</td>
<td>++</td>
<td>-/+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R295H</td>
<td>CGC→CAC</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>V319I</td>
<td>GTC→ATC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>C325Y</td>
<td>TGC→TAC</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>L326F</td>
<td>CTT→TTT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R330H</td>
<td>CGT→CAT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S347N</td>
<td>AGC→AAC</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R348H</td>
<td>CGC→CAC</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A352T</td>
<td>GCC→ACC</td>
<td>-</td>
<td>++</td>
<td>-/+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>T354I</td>
<td>ACC→ATC</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>T365M</td>
<td>ACG→ATG</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>S367N</td>
<td>AGC→AAC</td>
<td>+</td>
<td>+</td>
<td>-/+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>S368F</td>
<td>TCC→TTT</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>G371E</td>
<td>GGA→GAA</td>
<td>-</td>
<td>-/+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>H372Y</td>
<td>CAC→TAC</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A382V</td>
<td>GCG→GTG</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>L389F</td>
<td>CTC→TTC</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>Y381F (SDM)</td>
<td>TAC→TTC</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>++</td>
</tr>
</tbody>
</table>

1) The first letter indicates the amino acid in wild type IntDOT, followed by the position and the amino acid substituted for this amino acid. If there is only the first letter, the mutation creates a nonsense codon.
2) Base pair changes are shown in bold letters.
References


Chapter 3. Analysis of HJ resolution by IntDOT

3.1 Introduction

IntDOT is an integrase encoded by CTnDOT, a Bacteroides integrative and conjugative element (ICE). IntDOT is a member of the tyrosine family of recombinases (11) which includes lambda integrase (Int), FLP, CRE, and the XerC and XerD proteins. Lambda Int and IntDOT are called factor assisted tyrosine recombinases because they show regulated directionality of their recombination reactions. They contain three DNA binding domains: the N-terminal arm-binding domain (N), the core-binding domain (CB), and the catalytic domain (CAT). These recombinases bind to two different types of DNA sites: the arm-type sites and core-type sites. For lambda Int, it has been shown that the N domain binds to the arm-type sites and functions in the directionality and regulation of catalysis. The CB and CAT domains of lambda Int bind to the core-type sites and perform catalysis (For a review see(8, 13)).

In recombination reactions mediated by tyrosine recombinases, four recombinase monomers bind to the core-type sites of the recombining partner sites. The core-type sites are imperfect inverted repeats that surround the overlap regions where two sets of strand exchanges occur to form recombinant products. The four recombinase monomers that perform recombination communicate via interactions mediated within the DNA and by protein-protein interactions so that only one pair of recombinase monomers is activated at a time. For the first round of strand exchanges, two recombinases out of four become activated and cleave and exchange a strand of DNA from each site to form an intermediate called a Holliday Junction (HJ). The HJ-protein complexes undergo a
conformational change or an isomerization step that allows the second round of strand exchanges to occur to form products (For a review see (1, 16)).

Most tyrosine recombinases require strict DNA sequence identity in the overlap region (17). For example, a single mismatch in the overlap region of a lambda site inhibits recombination with a partner site (3). The sequence identity is required for the DNA isomerization step that occurs between the sets of strand exchanges (16). However, unlike other tyrosine recombinases, IntDOT recombines two sites that contain five base pairs of mismatches in the overlap region (12). This observation indicates that IntDOT uses a somewhat different mechanism of recombination from that of most of other tyrosine recombinases.

This study focuses on characteristics of HJ resolution by wild type IntDOT and the roles of the N domain, the coupler region, and other residues of IntDOT in resolution of HJs. Wild type IntDOT responds differently to two types of synthetic HJs depending upon whether the overlap regions contain identical or mismatched sequences. Interestingly, wild type IntDOT cannot tolerate heterology in the overlaps of synthetic HJs in the absence of the arm-type sites. This observation indicates that intasome formation with a substrate containing IntDOT bound to the arm-type sites and a host factor is important for IntDOT to complete recombination with natural sites.

Analysis of HJ resolution by mutant IntDOT proteins indicates that the V95 residue in the coupler region is important for HJ resolution. Mutational analysis of the N domain and the coupler indicate that it may interact with another region of the protein by hydrophobic interactions. In addition, residues R295 and S368 in the CAT domain are

86
likely important for protein-protein interactions required for resolution of HJs but not for catalysis.

3.2 Materials and Methods

Growth media, strains, chemicals, and enzymes

The strains, plasmids, and oligonucleotides used in this study are listed in Table 3.1. *E. coli* strains were grown in Luria-Bertani (LB) medium (Difco). *E. coli* DH5alpha MCR was used for site directed mutagenesis and plasmid maintenance. JG19002 (7) was used for the in vivo integration assay. Antibiotics and arabinose were purchased from Sigma and X-GAL was obtained from RPI and used at a concentration of 80 ug/ml. Antibiotic concentrations used were 100 ug/ml augmentin (Aug) and arabinose was used at the concentrations indicated. Gamma $^{32}$P-ATP was purchased from Perkin Elmer, Inc. and T4 DNA kinase from Fermentas. All oligonucleotides shown in Table 3.1 were synthesized by Integrated DNA Technologies (IDT).

Site-directed mutagenesis

Eight pSK2 plasmids containing a single substitution; D19K, D31K, D44K, D49K, E93K, K94D, K94E, and K96E were made using the Stratagene QuikChange Site-directed Mutagenesis kit. Primers containing specific mutations are listed in Table 3.1. Mutagenized plasmids were sequenced to confirm that the desired substitutions were obtained. The plasmids were transformed into JG19002 for in vivo integration assays and into MC1061 for overproduction of mutant proteins and for biochemical assays.
Overproduction of proteins

To express wild type and defective IntDOT proteins, MC1061 cells containing pSK2 and pSK2 derivatives were induced by addition of arabinose to a final concentration of 0.4%. A 50-ml culture was grown in LB supplemented with ampicillin to an OD 0.4 at 37 °C and arabinose. The cells were induced for 4 hr at room temperature. Cells were harvested and suspended in 0.5 ml of low salt IntDOT lysis buffer (50 mM NaHPO$_4$ pH[7], 1 mM EDTA, 50 mM NaCl, 10 % glycerol, 1 mM DTT). After sonication and centrifugation, the supernatants were used as sources of IntDOT proteins for in vitro assays described below. Quick Start Bradford protein assays (Bio-Rad) were used to determine the amount of protein in cell extracts. Fifteen ug of cell extract was used for the DNA binding, ligation, and resolution assays.

Construction of HJ intermediates

HJs were constructed by annealing four strands; 50 bases of attB, 59 bases of attR, 85 bases of attL, and 94 bases of attDOT. All four strands were radio labeled with $^{32}$P and mixed at an equal molar ratio (60 pmole/ ul per each strand) in an annealing buffer (0.1 M KCl, 10 mM Tris-HCl, pH[8]). The annealing reaction was heated to 95 °C and cooled to 22 °C at a rate of 5 ° min$^{-1}$ using a PCR machine. The mixture was electrophoresed on a 5 % polyacrylamide gel at 200V for 2hr. The band containing the four annealed strands was purified and eluted in low salt buffer (10 mM Tris-Cl pH 8, 100 mM NaCl, 1 mM EDTA) for 1-2 days. The solution containing HJ DNA was passed through a DE 52 column to remove gel slices and to concentrate DNA. The HJs were ethanol precipitated, dissolved in water, and used as substrates in HJ resolution assays.
HJ resolution assay

The resolution assays were carried out in a 20 μl of reaction mixture of 15 μg/μl of crude extract containing IntDOT protein, 20 ng of HJ DNA, 50 mM Tris-HCl [pH8], 50 mM NaCl, 1 mM EDTA, 5 mM DDT at 37°C for 2 hr. The reaction was phenol extracted and subjected to electrophoresis on an 8 % polyacrylamide gel at 200 V for 2 hr. The resolution products in the gels were exposed to phosphoimager screens and results were analyzed using Fujifilm Image Gauge software (Machitoch v.3.4).

E. coli in vivo integration and in vitro assays

In vivo integration, DNA binding, and ligation assays were performed as described previously (6).

3.3 Results and Discussion

Resolution of Holliday Junctions by wild type IntDOT

Site specific recombination reactions catalyzed by tyrosine recombinases occur by two sequential sets of strand exchanges. The sites of exchange border a region of DNA, called the overlap region, that is usually identical in both recombination sites. The complex that performs the recombination reaction contain four recombinase monomers bound to the four core-type sites of the partner DNA substrates. Two of the recombinase monomers are catalytically active while the other two recombinase monomers are inactive. The first strand exchanges are carried out by the two active recombinase monomers to form a four-way branch structure HJ intermediate. (Since the two sites of strand exchanges mediated by tyrosine recombinases are usually 7 base pairs apart at the ends of the overlap region, conformational changes in DNA structure must occur between
two sets of strand exchanges.) In order to execute the second set of strand exchanges to form products, the HJ intermediate undergoes a conformational change that results in the activation of the second pair of recombinase monomers that perform the strand exchanges to form products. Presumably, recombinase monomers in the complex also interact with each other through protein–protein interactions to help regulate and coordinate the strand exchanges. Thus the isomerization of the HJ in addition to protein–protein contacts are responsible for aligning the sites of cleavage and exchange in the active sites of two monomers. If the strands that were exchanged initially to form the HJ are exchanged a second time, the HJ is resolved to reform the substrates and the recombination reaction moves backward. If the strands that remained intact during HJ formation are exchanged, the HJ is resolved into recombinants and the recombination reaction moves forward.

The strand exchange mechanism and resolution of HJs by lambda Int have been studied in detail (2, 14). Like most other tyrosine recombinases, lambda recombination requires strict sequence identity in the overlap region for efficient recombination (3). However, IntDOT does not require extensive DNA sequence identity in the overlap region. The 7 bp overlap regions of attDOT and attB sites recombined by IntDOT contain two base pairs of identity (the conserved GC dinucleotide) adjacent to the first cleavage sites in the top strand and 5 bp that can be completely mismatched extending from the GC dinucleotide to the position of the second cleavage sites in the bottom strands (12) (Figure 3.1). As long as the two base pairs at the left side of the overlap are the same in the both sites, the first strand exchange occurs (9). Thus, during a recombination reaction IntDOT needs 2 bp of identity at the end of each overlap to
resolve a HJ but does not need sequence identity for HJ resolution (9). Thus that the first strand exchanges are dependent on sequence identity in the recombining sites but the second strand exchanges can bypass a barrier of heterology (9).

In a previous study, I characterized a collection of IntDOT mutants I isolated for defects in DNA binding, DNA cleavage and DNA ligation assays to identify specific defects of the mutant proteins in the recombination pathway (7). Since IntDOT is a tyrosine recombinase, it would be expected that the enzyme would be able to process synthetic HJs like other family members. As a result, I needed to determine whether IntDOT has the ability to resolve a synthetic HJ intermediate into substrates or products. In order to study resolution of HJs without IntDOT interaction with the arm-type sites that could complicate interpretation of the data, I constructed two types of HJs that lacked the arm-type sites in this study. One was a HJ that contains identical overlap sequences (Figure 3.2A). The branch point of this HJ can migrate unimpeded throughout the overlap. IntDOT has been shown to recombine attB and attDOT sites containing the identical sequences as efficiently as sites that contain mismatched sequences (Laprise J. unpublished). A second HJ contains mismatches in the overlap like a natural HJ made by IntDOT during a reaction between natural attDOT and attB sites (Figure 3.2C). The branch point of this HJ is frozen at the left side of the overlap and cannot branch migrate to the right of the overlap region due to mismatches.

Both synthetic HJs in this study mimic HJs formed after the first strand exchanges of integrative recombination. The HJs were assembled by annealing the four appropriate radio-labeled strands containing the core-type sites. Synthetic HJs contain two top strands of attL and attR, and two bottom strands of attB and attDOT (Figures 3.2A and
C). The final constructs contain 60 bp of the longest north arm (N), 34 bp of the east arm (E), and 25 bp of the west (W) and south (S) arms (Figures 3.2A and C). The differences in the lengths of the arms allow us to distinguish one direction of resolution from the other and to detect all four att sites after resolution.

Figures 3.2B and D show that wild type IntDOT resolves both types of synthetic HJs. In Figure 3.2B, a HJ with identical overlap sequences is resolved in both directions indicating that resolution by wild type IntDOT produces all four att sites, the substrates and recombinant products. The branch point of this HJ can move freely between the sites of the first and second strand exchanges which can form two types of HJ conformers. One conformer is favorable to the top strand exchange and the other one is favorable to the bottom strand exchange. In this case, IntDOT exchanges one pair of strands at either side of the overlap depending on which of the HJ conformers is present. This result shows that IntDOT resolves a HJ with identical overlap sequences in both directions.

In contrast, a HJ with mismatched overlap sequences contains an immobile junction like a natural HJ formed by wild type IntDOT (Figure 3.2C). The junction can migrate within the two base pairs of ‘the GC dinucleotide’ adjacent to the site of the first strand exchange but cannot move through the region of heterology to the site where the bottom strand exchange would occur. In this case, the HJ can form the conformer which is favorable to top strand exchange. The results (Figure 3.2D) show that IntDOT resolves this HJ in only one direction which is strongly biased to regenerate substrates, attDOT and attB. This indicates that a barrier of heterology inside the overlap of the immobile synthetic HJ that lacks the arm-type sites prevents resolution of the immobile synthetic HJ into recombinant products.
Presumably, during an integrative recombination reaction the intasome composed of IntDOT and a host factor is assembled on \textit{attDOT} DNA. When the intasome captures an \textit{attB} site, synapsis and the first strand exchanges occur. The complex is able to bypass the barrier of heterology in the overlap and perform the second set of strand exchanges. In the experiments presented here, intasome formation does not occur because the synthetic HJs lack the arm-type sites. When IntDOT binds to the synthetic HJs with a mismatched overlap, it cannot bypass the barrier of heterology and produces only \textit{attDOT} and \textit{attB} substrates. On the other hand, when IntDOT binds to synthetic HJs containing identical overlap sequences, it produces both \textit{attDOT} and \textit{attB} substrates, and \textit{attL} and \textit{attR} recombinants because unimpeded branch migrations of this HJ can change HJ structure without the arm-type binding or intasome formation. This result indicates that, unlike most tyrosine recombinases, IntDOT appears to have gained a function that allows it to recombine mismatched overlap sequences by forming intasomes with the arm-type sites.

\textbf{Resolution of HJs by mutant IntDOT proteins}

IntDOT mutants deficient in integration listed in Table 3.2 were characterized previously in DNA binding, cleavage and ligation assays (7). I showed previously that IntDOT and IHF from \textit{E. coli} form a complex with \textit{attDOT} DNA (7). Presumably one or more IHF dimers bind nonspecifically to \textit{attDOT} DNA and at least one monomer of IntDOT binds to form the complex. I do not know the exact number of IntDOT monomers bound but it is likely that at least one binds simultaneously to a core and arm-type sites because formation of the complex requires IHF. The cleavage and ligation
assays measure the ability of the wild-type or mutant IntDOT to perform catalytic functions.

I showed above that the wild type IntDOT protein behaves differently in the resolution of HJs depending on the presence of identical or mismatched sequences in the overlap. It was of interest to see whether these IntDOT mutant proteins resolve the two types of HJs in the same way as wild type IntDOT or if some behave differently. For example, a mutant might resolve a HJ with identical overlap sequences but not a HJ with an immobile junction. Another type of mutant might be able to resolve the immobile HJ into products which would be a new function that wild type IntDOT cannot perform. In this case, the substituted residue would be predicted to be important in overcoming the barrier produced by the immobile junction. However, none of mutants that were able to resolve HJs were found to behave differently with the two types of HJs tested. In other words, mutants were either unable to resolve HJs or resolved them like wild type IntDOT.

For HJ resolution assays with IntDOT mutant proteins, I constructed the HJ shown in Figure 3.3C. The attDOT and attL strand are 94 bases and the attB and attR strand are 59 bases. This HJ contains mismatched overlap sequences that are the same as the one in Figure 3.2C but in this case only the attL strand was labeled. Either the attL or the attB sites could be visualized as resolution products that indicate the direction of resolution (Figure 3.3B).

Several mutants with substitutions in the CB domain (T184I and P209L) and in the CAT domain (T256I, C325Y, L326F, R330H, S347N, and G371E) were defective in DNA binding, cleavage and ligation (7). Thus, it is not surprising that these mutant
proteins are also defective in HJ resolution because HJ resolution requires cleavage and ligation of DNA. Mutants with substitutions of catalytic residues such as the R348H, H372Y, and Y381F proteins were also defective in HJ resolution as expected.

In the initial analysis of IntDOT, the residue equivalent to the first catalytic arginine of IntDOT was missing (11). Recent in vitro experiments and homology modeling indicated that R285 might provide the function of the missing arginine (7). If this residue were required for catalysis, I would predict that a mutant with a substitution of R285 would be unable to resolve HJs. As expected, the R285H protein was seriously defective in HJ resolution. In the same study, I also proposed that the A352 residue was involved in positioning of the catalytic tyrosine and the A352T protein was defective in HJ resolution too (Figure 3.3B).

Several mutants that are defective in integration were able to resolve HJs. These include three mutants with substitutions in the N domain (R13C, S38N, and G101R). These mutants affect residues implicated in binding to arm-type sites and are not likely to be directly involved with HJ resolution. Six mutants with substitutions in the CAT domain (V319I, T354I, T365M, S367N, A382V, and L389F) also resolve HJs efficiently. Several of the mutants contain substitutions that are not predicted to be near the active site. The causes of the defects in the recombination pathway for the V319I, T354I, T365M, A382V, and L389F proteins are not obvious. The V319 residue was predicted to be located the β6 strand of IntDOT (7). In the previous homology model, the T354 residue in helix K and the T365 residue in helix L (7) are located far from the active site and show no effect on cleavage, ligation, or HJ resolution. The A382 and L389 residues are near the catalytic tyrosine Y381 in helix M (7). In the homology model, the A382
residue is predicted to lie on the backside of the helix containing Y381. The L389 residue is predicted to lie in the extensible “tail” that interacts with another recombinase monomer in the IntDOT tetramer so it may not have a direct effect on catalytic activity. Interestingly, the S367 residue has been predicted to be near the catalytic H372 and the S367N protein was defective in cleavage and ligation but was active in HJ resolution. This mutant was leaky in the in vivo integration assay although I do not know the mechanism behind its phenotypes in the cleavage, ligation, and resolution assays. The exact roles of these residues await structural information that may be forthcoming from crystallization studies.

The V95M, R295H, and S368F proteins are especially interesting. All three were proficient in cleavage and ligation reactions but were defective or have a reduced activity in HJ resolution. I think it is likely that the V95, R295, and S368 residues are involved in protein-protein interactions. For example, the R295 residue was predicted to be located in the end of the β5 strand of IntDOT which was aligned to the region equivalent to the β7 (residues 239-243) of lambda Int (Figure 1.9). In lambda Int, the β7 strand is in the center of the protein and interacts with the tail (residues 350-356 including β9) of a neighboring protein in an intermolecular interaction (5). Intermolecular contacts between the β7 and the β9 strand of lambda Int are required for coordination of HJ resolution (4). The A241V protein with a substitution in the β7 of lambda Int has been characterized (15). The phenotypes of the A241V protein of lambda Int and the R295H protein of IntDOT are very similar. Both were recombination deficient proteins that showed enhanced catalytic activity such as cleavage and ligation but were defective in HJ resolution. The disruption of an intermolecular contact between the β7 strand and tail of
lambda Int affects regulation of catalysis and coordination of HJ resolution. It is possible that the R295 residue of IntDOT could be involved in a protein-protein interaction as shown in lambda Int.

The S368 residue in helix L is likely to be located in the surface of IntDOT and could also be capable of making a protein-protein interaction (7). However, I note that the mutant IntDOT protein with substitution of S367 next to S368 also has been predicted to be located in the helix L and was active in HJ resolution. This result indicates that two adjacent serine residues have different roles in recombination steps.

Among mutants with substitutions in the N domain, the V95M mutant protein showed a drastically different phenotype than other mutants with N domain substitutions such as the R13C, S38N, and G101R proteins. The V95 residue is located in a region that corresponds to the lambda Int coupler (Figure 3.4). The lambda coupler has been shown to be required for HJ resolution and interact with the N domain of a different Int monomer. Thus, the V95M protein could also be defective in resolution that requires protein-protein interactions. To confirm that the putative coupler region of IntDOT is important for HJ resolution and to identify the residues that interact in the N domain and the coupler, a site directed mutational analysis was performed.

**Role of the N domain in HJ resolution**

The lambda Int coupler region has been shown to participate in HJ resolution by a protein-protein interaction. Substitution of either the D71 residue in the coupler or the R30 residue in the N domain results in loss of HJ resolution activity. However a double mutant with the D71R and R30D substitutions regains the ability to resolve HJs (10). From this observation and other work the authors proposed that he D71 residue in the
coupler region makes an intermolecular ion pair with the R30 residue of an adjacent monomer (10).

Interestingly, the location of the V95 residue in the putative IntDOT coupler region and inability of the V95M mutant protein to resolve a HJ indicated that the IntDOT coupler region could play a role in HJ resolution similar to the lambda Int coupler. The mechanism could be through a protein-protein interaction by a charge interaction as shown in lambda Int or by interactions by hydrophobic clusters. In an attempt to determine whether there is an ion pair in the IntDOT N domain and coupler region, I made substitutions of all the charged residues in the IntDOT coupler to the reverse charged residue by site directed mutagenesis. The substitution mutants I made in the coupler are E93K, K94D, K94E, and K96E. The E93K mutant protein is active in in vivo integration, DNA binding, ligation, and resolution indicating that the E93 residue is not important for DNA binding or catalysis. At the position of K94, two negatively charged residues D and E were tested to determine if substitutions of similar negatively charged amino acids have different or similar phenotypes. Using the vivo integration assay, I found that the K94D, K94E and K96E proteins are strongly defective.

I expected that the K94D and K94E proteins would also likely have similar DNA binding, ligation and HJ resolution phenotypes because of the conserved nature of the substitutions. However, both are active in ligation (Figure 3.3) but the K94D protein is active for DNA binding and resolution while the K94E protein is inactive in DNA binding and HJ resolution (Figure 3.5 and Table 3.3). Thus the results show that that two similar negatively charged residues at a same position showed different effects on DNA binding and HJ resolution. Since the K94D protein has a slightly smaller side chain than
the K94E protein, it is possible that the bulkier side chain of glutamine in the K94E protein is responsible for the different phenotypes of the proteins. The glutamic acid side chain could disrupt an interaction between the coupler region and the N domain that is not disrupted by the shorter aspartic acid side chain. The glutamic acid could destabilize an interaction with another region of the protein that is not disrupted by the aspartic acid side chain. However, it is difficult to visualize a simple mechanism that would explain how the glutamic acid substitution could also affect DNA binding. Perhaps the normal K94 residue plays a role in DNA binding in an early part of the reaction and plays a role in HJ resolution during a later step of the reaction. The K96E mutant protein with a substitution of a glutamine showed the same phenotype as the K94E protein in that it simultaneously affected HJ resolution and DNA binding. Taken together, the results are consistent with the hypothesis that the coupler is important for HJ resolution through a protein-protein interaction. However, the role of charged residues to the interaction remains uncertain because it is possible that defectiveness of the K94E and K96E proteins in HJ resolution are not necessarily due to reversed charge residues.

The secondary structure prediction of the region of IntDOT that includes residues 91 to 97 is an alpha helix. In Figure 3.6, I have projected the amino acid residues of this region of the protein onto a helical wheel. Examination of the wheel reveals that several of the residues, including V95 are on a side of the wheel that contains mainly hydrophobic residues (T91, A92, V95, A98, and F99). The other side of the wheel contains K94 and K96 as well as the charges E93. The organization of amino acid residues in this region is intriguing but more information will be required to determine if the residues play roles in DNA binding and / or HJ resolution.
Substitutions of charged residues in the N domain

To further investigate the possibility that mechanism of contact between the N domain and coupler involves a charge interaction, I made additional substitutions in the N domain that produced reverse charged residues. Since mutants with substitutions of the two lysine residues in the coupler, K94E and K96E, are defective in resolution, I hypothesized that substitution of a negatively charged residue with a positively charged residue in the N domain could restore the putative charge interaction. I mutated the aspartic acids at four different positions in the N domain to lysines, D19K, D31K, D44K, and D49K (Figure 3.4). These could be potential candidates for residues that interact with substituted lysine residues in the coupler. I found that three substitution mutants, the D19K, D31K, and D49K proteins are strongly defective and the D44K protein is leaky in vivo integration. The D19K and D44K proteins were also defective in DNA binding and intasome formation in the gel shift assay indicating that these residues are directly or indirectly involved in DNA binding. The D31K and D49K mutants were proficient in DNA binding, ligation and HJ resolution. However, the observation that these proteins are defective for recombination in vivo indicates that the D31 and D49 residues are required for a step in recombination pathway.

Regardless of DNA binding phenotypes, all four substitution mutant proteins in the N domain were active in HJ resolution (Figure 3.5). These observations indicate that these D residues are not involved in a charge interaction with residues in the coupler. These phenotypes are also similar to those of other substitution mutants in the N domain such as R13C and S38N mutants indicating that these IntDOT residues in the N domain are not involved in HJ resolution.
I find no evidence that there is a charge interaction between any residues in the coupler region and the N domain. It is possible that the HJ resolution activity of substitution mutants is not affected enough by disruptive protein-protein interactions to produce a phenotype. It is possible that V95 in IntDOT coupler and the adjacent hydrophobic residues in the helical wheel interact with the N domain.

3.4 Conclusion

This chapter demonstrates that wild type and mutant IntDOT proteins can resolve synthetic HJs. Comparison of the resolution of HJs containing identical or mismatched overlap sequences indicates that IntDOT requires intasome formation with the arm-type sites to overcome a barrier of heterology in the overlap of synthetic HJs that lack the arm-type sites. Thus, the unusual ability of IntDOT to recombine the mismatched overlap sequences is provided from the intasome during recombination. I also found that the V95, R295, and S368 residues are likely important in protein-protein interactions. Further mutational analysis of the N domain and coupler of IntDOT indicated that the IntDOT coupler is important for HJ resolution. The possible protein-protein interactions between the N domain and coupler could occur by hydrophobic clusters containing V95 residue.
Figures and Tables

**Figure 3.1 The attDOT and attB attachment site of IntDOT.** The two overlap regions of the attDOT and attB contain 7 bp of DNA sequences that are flanked by D and D’ of attDOT and B and B’ of attB site. The overlap sequences compose of 2 bp of sequences identity (the GC dinucleotide) shown in bold and five base pairs of mismatched sequences in the overlap region. The first strand exchanges occur at the adjacent site to the GC dinucleotide of the top strand and the second strand exchanges occur at the other side of the overlap at the bottom strand.
Figure 3.2 HJ resolution by wild type InDOT. (A) A HJ that contains the core-type sites and identical sequences in the overlap. The top strands of attL and attR are shown as solid lines and the bottom strands of attDOT and attB are shown as dotted lines. All four strands are labeled with γ-32P. Small arrows indicate the first strand exchange sites and the big open arrows indicate the second strand exchange sites. The identical sequences in the overlaps are in bold. (B) Resolution of the HJ in (A) with identical overlap sequences by wild type IntDOT. Wild type IntDOT resolves this HJ into products and substrates containing all four sites attDOT, attB, attL, and attR. (C) A HJ that contains the core-type sites and an immobile junction in the overlap where 2 bp of GC identity and 5 bp of heterology are present. This HJ mimics the natural HJ. (D) Resolution of a HJ in (C) with an immobile junction by wild type IntDOT. This HJ in Figure 3.2C is resolved into substrates, attDOT and attB only.
Figure 3.3 HJ resolution by mutant IntDOT proteins. (A) A HJ that was used to test resolution abilities of IntDOT mutant proteins. The HJ used in this assay contains an immobile junction and the attL strand was labeled and two directions of resolution are distinguishable by the size of resolved products. (B) Resolution by wild type and mutant IntDOT proteins. The HJ substrate shown in (A) was incubated with 15 ug of extracts containing defective IntDOT proteins at 37 °C for 2 hr. The reaction was stopped by phenol extraction and resolved products were detected on an 8 % polyacrylamide gel. The results are summarized in Table 3.2. M is 50 bp DNA ladder.
The N domain of λ Integrate

\[\beta_1 \quad \beta_2 \quad \beta_3 \quad \alpha H1 \quad \alpha H2\]

The coupler
...

Figure 3.4 The secondary structures and amino acid sequences of the N domains of lambda Int and IntDOT. The N domains of the two proteins share a common folding pattern. They contain three \(\beta\) strands (\(\beta_1, \beta_2, \beta_3\)), and two helices (\(\alpha H1, \alpha H2\)). \(\alpha H2\) is called the “coupler” in lambda Int. The big arrows indicate that the residues R30 and D71 of lambda Int form a protein-protein interaction in lambda Int. In IntDOT, the V95 residue is located in the region equivalent to the coupler of lambda Int. The thin arrows indicate aspartic acids at four different positions in the IntDOT N domain that were substituted with lysines in this study by site directed mutagenesis.
Figure 3.5 Resolution of HJ by IntDOT mutant proteins with substitutions in the N domain and coupler. The HJ resolution assays were performed same as in Figure 3.3B. M is 50 bp DNA ladder.
Figure 3.6 Helical wheel projection of the putative α-helix spanning amino acid residues 91 to 99 in IntDOT. The V95 residue is on a side of the wheel that contains mainly hydrophobic residues (T91, A92, V95, A98, and F99). The other side of the wheel contains K94 and K96 as well as the charges E93. Box indicates the residue that shows defect in integration in vivo when substituted.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Relative genotype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>JG19002</td>
<td>W3110(ΔlacI)Kan::pSK1</td>
<td>Lab strain</td>
</tr>
<tr>
<td>MC1061</td>
<td>hsdR2 hsdM+ hsdS+ araD139 Δ(ara-leu)7697 Δ(lac)X74 galE15</td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F- endA1 glvV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–</td>
<td>Lab strain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSK2</td>
<td>pSK2 is pBAD18 containing the intDOT gene from the plasmid pSY cloned between XbaI and HindIII sites.</td>
</tr>
<tr>
<td>pSK2 derivatives</td>
<td>The pSK2 derivatives are pBAD 18 containing the wild-type or mutant intDOT genes between XbaI and HindIII sites cloned downstream of the ara pBAD promoter.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Oligonucleotide Sequences (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primers used for site-directed mutagenesis to substitute a reverse charged residue</td>
</tr>
</tbody>
</table>

| F_D19K | CTCAAGCGTCAGGTTAGTGA AAAAAGAAGGACAGTTCCCCGTCAC TGGGACGC | This work |
| R_D19K | GCGTCCCAATGACGCAGGAACTGTCCTTTTTTTTTCTACACCTGAC GCTTGGAG | This work |
| F_D31K | CCCGTCATGGGACGCATCACGTTGAAGGCACGACACGACGACGTTCAGCAGCTGC | This work |
| R_D31K | GCAAGCTGAAACTCTGCACGGTTTGTGACAGTCAGTTTTGCAAGCCTGAC TGCAGG | This work |
| F_D44K | ACCGAGTGCTAGCAGCTCAAAACTGACTGTCAAAACGGAGGCTGTCGG ACACCAAA | This work |
| R_D44K | TTTTGGGTTCCTCACTGCACGGTTTGTGACAGTCAGTTTTGCAAGCCTGAC ACTGTCG | This work |
| F_D49K | CTCGACTGTCAAGGCGAAGCTGTTGAAGAAACCAAGGGGACGTGACGTCAGCAGTACGTTGCA GCACGGG | This work |
| R_D49K | GCCCGTGACAGTGCTGCCCTTTTGGTTTTTCCACGCTTCGGGTGCA CAGTCAG | This work |
| F_E93K | AACCTTGGTGATCAGGCGAAGGAAGGTAGAAGACGCC | This work |
| R_E93K | GGCGGCTCTTACCTCTTCCTCAGGCGCGTACGAGG | This work |
| F_K94D | CGTGACACACTGCTCACGGCGGAGGACGTGCAAGAAGACGCCTTCC TGGATTG | This work |
| R_K94D | CAATCAGAAAGAAGGCTTCTTACGTCCTCAGCGGCTGAGCAAG AAGGCGG | This work |
| F_K94E | TCCGTCAGGGGCGGAAAGGTAGTGAAGAAGACGGCTTCC TGGATTG | This work |
| R_K94E | AACGGGGTTTCTTACCTCCTCGCGGTCGACGAAGA | This work |
| F_K96E | ACCGCGGGAAGGGAGGAGGAGGACGGCTTCTCGGA | This work |
| R_K96E | TCGAGAAAGGCGGTTCTCAGACGCCTTCCGCGCTG | This work |
Table 3.1 (continued)

<table>
<thead>
<tr>
<th>Oligonucleotides used to construct HJs (5’→3’)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bs <em>attB</em></td>
<td><em>AAT TAC TGT TTA GTA TTT TAA TTG CGC AAA TTT ACT GCA AAT TTC CGA GC</em>&lt;br&gt; This work</td>
</tr>
<tr>
<td>bs <em>attDOT</em></td>
<td><em>CAA GAG CGA TTA ACC TAC GCT CAT TTC CAA TAA ATT ACA CTC TTT TCG TAA CTT CAT TAA GCA AAG TTA CTA CAA AAA AGT GAA ATG CGG AAA G</em>&lt;br&gt; This work</td>
</tr>
<tr>
<td>ts <em>attL</em></td>
<td><em>GCT CGG AAA TTT GCA GTA AAT TTG CTG TT AGT GAA GTT ACG AAA AGA GTG TAA TTT ATT GGA AAT GAG CGT AGG TTA ATC GCT CTT G</em>&lt;br&gt; This work</td>
</tr>
<tr>
<td>ts <em>attR</em></td>
<td><em>CTT TCC GCA TTT CAC TTT TTT GTA GTA ACT TTG CGC AAT TAA AAT ACT AAA CAG TAA TT</em>&lt;br&gt; This work</td>
</tr>
<tr>
<td>bs <em>attB</em> homology</td>
<td><em>AAT TAC TGT TTA GTA TTT TAA CTA AGC AAA TTT ACT GCA AAT TTC CGA GC</em>&lt;br&gt; This work</td>
</tr>
<tr>
<td>bs <em>attR</em> homology</td>
<td><em>CTT TCC GCA TTT CAC TTT TTT GTA GTA ACT TTG CTT AGT TAA AAT ACT AAA CAG TAA TT</em>&lt;br&gt; This work</td>
</tr>
</tbody>
</table>

Sequences in the overlap are underlined

Oligonucleotides contain a $^{32}$P label are represented by an asterisk (*).
Table 3.2 IntDOT mutants isolated in this study

<table>
<thead>
<tr>
<th>Domain</th>
<th>IntDOT mutants$^1$</th>
<th>E.coli in vivo integration$^2$</th>
<th>HJ resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>pBAD18 only</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>R13C</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>S38N</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>V95M</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>G101R</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>CB</td>
<td>T184I</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>P209L</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CAT</td>
<td>T256I</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>W280</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R285H</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>R295H</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>V319I</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>C325Y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>L326F</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R330H</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S347N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>R348H</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A352T</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>T354I</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>T365M</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>S367N</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>S368F</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>G371E</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>H372Y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A382V</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>L389F</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>Y381F</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1) The first letter indicates the amino acid in wild type IntDOT, followed by the position and the amino acid substituted for this amino acid. If there is only the first letter, the mutation creates a nonsense codon.

2) Base pair changes are shown in bold letters.
Table 3.3 Mutational analysis of the arm domain and coupler of IntDOT

<table>
<thead>
<tr>
<th>Domain</th>
<th>mutants</th>
<th>In vivo integration</th>
<th>binding</th>
<th>Ligation</th>
<th>HJ resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>pBAD18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>The coupler</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E93K</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>K94D</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>K94E</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>K96E</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>The arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D19K</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>D31K</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>D44K</td>
<td>+</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>D49K</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>
References


Summary

Integrase of CTnDOT called IntDOT is a member of tyrosine recombinase family. An alignment and mutational analysis of the catalytic domain identified five of six signature sequences of tyrosine recombinases. The first catalytic arginine residue was not found at the conserved position. To extend the mutational analysis of IntDOT and identify important residues in the arm and core binding domain addition to the catalytic domain, random mutagenesis was performed. Using the LacI repressor as a reporter, I constructed an *E. coli* in vivo indicator strain and isolated twenty five IntDOT mutants that are deficient in integration; R13C, S38N, V95M, and G101R in the N domain, T184I and P209L in the CB domain, T256I, W280, R285H, R295H, V319I, C325Y, L326F, R330H, S347N, R348H, A352V, T354I, T365M, S367N, S368F, G371E, H372Y, A382V, and L389F in the CAT domain. To find out which steps were affected during recombination, these mutant proteins were characterized in three different steps of recombination like DNA binding, cleavage and ligation. Substitutions in the N domain had no effect on catalysis and substitutions in the CB domain showed reduced activity in all of the assays. In the CAT domain, most mutants with a substitution were active in all of three steps. The most interesting mutant in the CAT domain is arginine at 285 of IntDOT. It apparently functions as the first catalytic arginine that was not found in previous study. Alanine at 352 of IntDOT also has been proposed to be important for positioning of the catalytic tyrosine.

Unlike other tyrosine recombinases, IntDOT recombines sites that contain mismatched DNA in the overlap. Analysis of resolution by wild type IntDOT of two
types of HJs containing mismatched and identical overlap sequences is performed to see how IntDOT responds to heterology and homology in the overlap. The results showed that wild type IntDOT resolves a heterologous HJ in one direction due to blockage of branch migration within the overlap sequences but resolves a HJ with identical sequences in both directions. This indicates that wild type IntDOT is sensitive to heterology and cannot recombine mismatched DNA in the absence of the arm sites. In order to overcome the barrier of heterology in the overlap of IntDOT, intasome formation involving in the arm type binding are required.

The analysis of IntDOT mutant proteins revealed that the V95 residue in the coupler, R295 and S368 residues in the CAT domain are important for HJ resolution that is required for a protein-protein interaction. Since the lambda Int coupler has been shown to interact with the N domain of adjacent protein by a charge interaction, I attempted to detect protein-protein interactions between the N and coupler of IntDOT by isolating charged residues that are important to HJ resolution. Eight substitutions with a reverse charged residues in the N domain and coupler were made, D19K, D31K, D44K, D49K, E93K, K94D, K94E, and K96E. D19 and D44 are important to the arm binding and intasome formation but all of the negatively charged residues in the N domain are not important for protein-protein interactions in HJ resolution. In the coupler, K94 and K96 two lysine residues seem to be involved in HJ resolution. However, interestingly the K94D protein and K96E protein have a similar reversed charged residue showed different phenotypes. The K94E is defective but the K94D protein is active in in HJ resolution. This expected result can be explained that the bulky side chain of the K94D protein may disrupt interactions made in IntDOT coupler region. In the N domain, four mutants that
substituted aspartic acids to lysine are active in HJ resolution indicating that aspartic acids in the N domain are not involved in HJ resolution. Therefore, the interacting residues with IntDOT coupler and the mechanism of interaction are not kown yet.
Chapter 4. Future directions

4.1 Crystallization of IntDOT

Most tyrosine recombinases contain six invariant signature residues, R\textsubscript{I}, K, H\textsubscript{I}, R\textsubscript{II}, H\textsubscript{II}, Y at conserved positions in the CAT domains. These six residues compose a catalytic active site. A previous alignment and mutational study of IntDOT did not find the first catalytic arginine, R\textsubscript{I} of IntDOT at the conserved position in the protein. This indicates that IntDOT has a different architecture of the active site. Interestingly, an arginine residue at position 285 of InDOT was proposed to function as the first catalytic arginine (2). R285 is located in an unusual position which is not aligned with the R\textsubscript{I} residues of other tyrosine recombinases. In order to determine whether R285 is part of the catalytic site, the IntDOT protein is being crystallized with DNA substrates.

The *intDOT* gene was cloned into the pET30a plasmid as a source of IntDOT crystallization. Sumiko Yoneji, a research assistant in our lab, overexpressed the IntDOT protein and purified by heparin column chromatography. Purified IntDOT is incubated with a suicide *attDOT* DNA substrate in an appropriate buffer. In this reaction, IntDOT-DNA complexes containing a phosphotyrosyl linkage can be crystallized and the X-ray diffraction data will show the IntDOT structures. Interactions of IntDOT with DNA and with other IntDOT molecules such as a IntDOT dimer can be studied. Crystal structure studies will provide important molecular features of IntDOT. Also, it will be interesting if the R285 residue is confirmed as the first catalytic arginine of IntDOT in crystal structures.
4.2 Analysis of the role of IntDOT in excision

Excision of CTnDOT is a complicated reaction requiring by four different proteins that are expressed from CTnDOT: Orf2c, Orf2d, Exc, IntDOT, and a host factor. IntDOT is an essential enzyme for excision as well as integration. To better understand the function and mechanism of IntDOT in CTnDOT excision, it is important to know which amino acids of IntDOT are required for excision. For in vivo excision assays, an *E. coli* excision indicator strain is being developed. Carolyn Keeton, a PhD student in our lab is working on this project. She integrated the chloramphenicol resistance gene (*camR*) as a reporter gene flanked by *attL* and *attR* into *E. coli* chromosome that contains the kanamycin resistance gene (*kanR*). Individually cloned *orf2c- orf2d, exc*, and *intDOT* are induced by IPTG in this indicator strain. Transformants are grown on replicated plates, one supplemented with chloramphenicol and the other one with kanamycin. In a positive control, cells performing excision by wild type IntDOT will survive only on plate with kanamycin due to loss of the *camR* gene. Excision can be confirmed by a colony PCR to amplify the *attDOT* site. To identify important residues of IntDOT on excision, random mutagenesis of the *intDOT* gene can be performed by hydroxylamine or error-prone PCR. Cells containing mutated IntDOT proteins cannot perform excision and will survive on both Cam and Kan plates. Twenty five integration defective IntDOT mutant proteins and newly isolated excision defective IntDOT mutant proteins will be characterized in DNA binding to *attL* and *attR*, cleavage and ligation, and HJ resolution. The excision indicator strain can also be used to isolate mutants of excision proteins Orf2c, Orf2d, and Exc and *Bacteroides* host factor (BHF).
4.3 Roles of the arm-type sites of CTnDOT in HJ resolution

Tyrosine recombinases form a HJ intermediates during recombination. A HJ is formed by the first pair of strand exchanges and resolved by the second pair of strand exchanges. The HJ resolution mechanism by IntDOT is different from that of other tyrosine recombinases in that IntDOT does not require identical sequences at the site of the second set of strand exchanges and can tolerate a barrier of heterology in the overlap region (3). HJ resolution by the second set of strand exchanges requires a protein-protein interaction. Identification of important residues for protein-protein interaction will provide a better understanding about resolution mechanism mediated by IntDOT.

In Chapter 3, I analyzed the effect of HJ conformational changes on HJ resolution by wild type IntDOT in the presence and absence of identical sequences in the overlap. The results showed that IntDOT cannot resolve a HJ with mismatched overlap sequences in the absence of the arm-type sites indicating that intasome formation with IHF and the arm-type sites is required for IntDOT to recombine sites containing non identical sequences. To investigate the roles of the arm-type sites in HJ resolution, a HJ containing the core- and arm-type sites will be constructed. Four att sites can be PCR amplified from the plasmids containing each att site. PCR reactions to amplify the attDOT site are performed using a set of an unlabeled forward primer and a labeled reverse primer. The attB, attL, and attR sites can be amplified using a set of an unlabeled forward and reverse primer. Four double stranded att sites are denatured and reannealed in an annealing mixture. The gel purified HJ containing four strands will be used as substrates. In the same way, HJs containing the arm-type site mutants with base substitution mutations can be constructed. Since the bottom strand of attDOT is labeled,
only a HJ containing the top strands of *attL* and *attR* and bottom strands of *attDOT* and *attB* is visualized on a gel. The wild type and mutant IntDOT proteins are tested with these HJs in the resolution assay. An accessory protein like a *Bacteroide* host factor would be required for this reaction to provide an appropriate architecture that is capable to resolve HJ.

The roles of the arm-type sites in HJ resolution can be analyzed by providing the unlabeled arm-type sites “*in trans*” in the reaction including a labeled HJ with the core-type site. Previously, footprinting analysis (1) and gels shift assays (4) revealed six arm-type sites in *attDOT*: the R1, R1’ R2, R2’, L1, and L2 sites. Since the L2 site is not important in integration and excision, five different unlabeled arm-type sites can be prepared separately. One wild type arm-type site can be substituted with one mutant arm-type site with base substitutions in HJ resolution assays.

**References**


Curriculum Vitae

SEYEUN KIM

I. EDUCATION

2010  Ph.D. Microbiology, University of Illinois at Urbana-Champaign
      Thesis advisor: Dr. Jeffrey F. Gardner
      Thesis title: Structure-Function Analysis of CTnDOT integrase, IntDOT

2007  M.S. Microbiology, University of Illinois at Urbana-Champaign

1998  M.S. Genetics, Ewha University, Seoul Korea

1996  B.S. Educational Science Biology, Ewha University, Seoul Korea

II. PUBLICATIONS


III. PRESENTATIONS
