IntDOT Interactions with Core Sites During Integrative Recombination

Jennifer Laprise#, Sumiko Yoneji, Jeffrey F. Gardner

Department of Microbiology
University of Illinois at Urbana-Champaign, Urbana, IL 61801

#Corresponding author. Mailing address: Department of Microbiology, University of Illinois at Urbana-Champaign, 601 S. Goodwin Avenue, Urbana, IL 61801. Phone: (217) 333-7289. Fax: (217) 244-6697. E-mail: jlapri2@illinois.edu

Abstract

Integrative and conjugative elements (ICEs), formerly called conjugative transposons, have been implicated in the proliferation of antibiotic resistance genes. CTnDOT is an extensively studied ICE found in Bacteroides spp. In addition to carrying resistance genes to both erythromycin and tetracycline, CTnDOT encodes a tyrosine recombinase called IntDOT that catalyzes integration into and excision out of the bacterial host chromosome. CTnDOT integrates into one of several known attB sites in the bacterial chromosome that consists of a pair of inverted repeat core sites called B and B’ in attB. The attDOT site contains the core sites and D and D’. These sites flank the overlap regions where strand exchanges occur. A notable feature of all known attB sites is the conservation of the B core site sequence, which is also found in the D core site of attDOT. In this study, we used a mutational analysis to establish the importance of this conserved sequence for integration, and characterize the interaction of IntDOT with individual base pairs. We identified important T-A basepairs at position -5 in the B and D core sites and position +5 in the poorly conserved B’ core site that are important for integrative recombination. Base analog studies suggest that IntDOT may make specific contacts with the A residues in the major groove at positions -5 and +5. IntDOT interaction with the A at position -5 in the B core site is required for the first strand exchange.


Introduction

*Bacteroides* spp. are Gram negative obligate anaerobes that comprise over 40% of the human gut microbiota (1). Although *Bacteroides* normally function as commensals, they can cause serious infections including abscesses and septicemia if released from their normal gut environment. These infections are becoming increasingly difficult to treat due to the high incidence of antibiotic resistance (2-4). By 2001 over 80% of *Bacteroides* strains were resistant to tetracycline, up from 30% in the 1970’s (4). There is evidence that *Bacteroides* is involved in horizontal gene transfer with Gram positive pathogens such as *Clostridium perfringens*, *Streptococcus pneumoniae* and *Enterococcus faecalis*, which suggests that *Bacteroides* is acting as a reservoir for antibiotic resistance genes (4). Resistance genes can be carried on mobile genetic elements such as integrative and conjugative elements (ICEs), which are often referred to as conjugative transposons (CTns). CTnDOT is an ICE found in *Bacteroides* spp. that carries resistance genes for both tetracycline (*tetQ*) and erythromycin (*ermF*). The presence of tetracycline actually induces CTnDOT excision and transfer by 1,000-10,000 fold (5).

The excision and integration reactions of CTnDOT are catalyzed by the element-encoded integrase called IntDOT. IntDOT is a member of the tyrosine recombinase family, which includes lambda Int, XerC/D, Cre and Flp. There are two types of systems within this family: autonomous and factor-assisted. Cre and Flp are examples of autonomous recombinases, which do not require additional proteins to catalyze recombination (5-7). These proteins have core binding (CB) and catalytic (CAT) domains and only recognize core sites that directly flank the overlap region where cleavage, strand exchange and ligation occur. IntDOT and lambda Int are examples of
factor-assisted recombinases which show directionality, and require additional proteins to form a higher order nucleoprotein complex called an intasome. Factor-assisted recombinases have three domains: the CB, CAT, and an N-terminal arm-binding domain (N) that binds to DNA sequences called arm-type sites which flank both sides of the core region. Formation of the intasome requires interactions between the recombinase and both core and arm-type DNA sites. Additional proteins such as host factors help facilitate these interactions (8-12).

The core sites recognized by IntDOT are called B and B' in the bacterial attachment site \textit{attB}, and D and D' in the \textit{attDOT} site (Figure 1). There are multiple \textit{attB} sites in the \textit{Bacteroides} chromosome that have been identified (13). All of these sites share a conserved sequence in the B core site - GTANNTTT. This sequence is also present in the D core site of \textit{attDOT}, which suggests that this sequence is important for IntDOT recognition of the core sites. In this study we systematically examined mutant core sequences in the \textit{attB} site to determine which base pairs are important for IntDOT interactions with the DNA. We also used base analogs to examine potential interactions between IntDOT and the major or minor groove of the DNA. The information gained allows us to make predictions that will complement future crystal structures of IntDOT complexed with DNA.

\textbf{Materials and Methods}

\textit{Preparation of Radiolabeled DNA Substrates}

Single stranded DNA oligonucleotides were obtained through Integrated DNA Technologies (IDT). One strand of each complementary pair was 5'-end labeled with $[\gamma$-
\[^{32}\text{P}\text{-ATP (Perkin-Elmer) using T4 polynucleotide kinase (Fermentas). Excess [\gamma-^{32}\text{P}] \text{ ATP was removed with Illustra Microspin G-25 columns (GE Healthcare). The oligonucleotides were annealed by combining the labeled strand and the complementary unlabeled strand at a 1:5 molar ratio in an annealing buffer consisting of 0.1 M KCl, 10 mM Tris-HCl (pH 8), 5 mM EDTA. The annealing mixture was heated to 90°C for two minutes then slowly cooled to 25°C.}

In vitro Integration Assay

The DNA substrates containing \textit{attDOT} and \textit{attB} were combined at a final concentration of 2 nM each in a 40 μl reaction volume that also contained 0.17 μM \textit{E. coli} IHF, 1 unit of IntDOT, 30 mM Tris-HCl (pH 7.4), 5 mM DTT, 0.1 mg/ml tRNA, 0.07 mg/ml BSA, 2.6% glycerol and 50 mM KCl. A unit of IntDOT is defined as the minimum amount of IntDOT needed to produce maximum recombination between \textit{attDOT} and \textit{attB} (12, 14, 15). The reaction is known to be slow, so samples were incubated overnight at 37°C and quenched the following morning by adding 10 μl of a stop solution containing 30% glycerol, 10% SDS, 0.25% xylene cyanol, and 0.25% bromophenol blue. A 10 μl aliquot of each sample was electrophoresed on a 1% agarose gel at 100V for 2 hours, then dried on a vacuum slab drier and exposed onto a phosphorimager screen. Recombination was quantified using a Fujifilm FLA-3000 phosphorimager and Fujifilm Image Gauge software (Macintosh v.3.4). To account for day-to-day variation (approximately 6% standard deviation) in IntDOT activity, a standard reaction containing wild type \textit{attDOT} and \textit{attB1} sites was included in each experimental group.
Restriction Digest of the Recombinants

A standard in vitro integration assay was performed but was terminated by heating at 60°C for 20 minutes instead of by the addition of stop solution. MgCl₂ was added to a final concentration of 10 mM. A volume of 18 μl of the reaction was transferred to a new microcentrifuge tube and 20 units of SspI (Fermentas) were added. The sample was incubated at 37°C for two hours then stopped by heating at 60°C for 20 minutes. The digested and non-digested samples were electrophoresed alongside each other on a 1% agarose gel at 100V for 2 hours. Gels were analyzed as described above.

Cleavage Assay

A cleavage assay using suicide substrates has been previously developed (16, 17). We used a variation of this assay in which IntDOT binds to and cleaves a substrate containing a nick in the middle of the overlap region. These substrates were prepared as described previously (14). The ³²P labeled top strand contains the wild type or mutant D core site and two bases of the overlap region. The reaction buffer is the same used in the integration assay. After a 2 hour incubation, the samples were boiled for five minutes in Tris-glycerol SDS sample buffer (63 mM Tris-HCl, 10% glycerol, 2% SDS, 0.0025% bromophenol blue) and electrophoresed on a 4-20% Tris-glycine gradient gel. Gels were analyzed as described above.

Results

Core Site Variants
The core sites are a pair of inverted repeats that directly flank the overlap region in both the \textit{attDOT} and \textit{attB} sites. In \textit{attDOT}, the core sites are D and D', and in \textit{attB} the core sites are B and B'. During integrative recombination, IntDOT is predicted to bind to these sites with the CB and CAT domains. By analogy to bacteriophage lambda and other tyrosine recombinase systems, it is likely that each of the four sites are bound by single IntDOT monomers assembled in a nucleoprotein complex on \textit{attDOT} called an intasome. The intasome undergoes synapsis with a naked \textit{attB} site where strand exchange occurs. It is notable that all of the \textit{attB} sites that have been identified to date share a nearly identical B core site sequence (GTANNTTT). This sequence is also conserved in the D core site of \textit{attDOT} (GTANNTTT) (Figure 2) (13, 14). This conserved sequence includes the first two base pairs of the overlap region - the GC dinucleotide that provides the only region of required sequence identity within the overlap regions (Figure 1). Based on this highly conserved sequence, some or all of the conserved base pairs should be important for making direct or indirect contacts with the protein, which would be required for the initiation and completion of recombination.

In order to identify specific base pairs in the core sequences that are important for IntDOT binding, we first made mutations in the conserved B sequence of \textit{attB1} and measured their effects on integration. If we disrupt important contact positions, we should observe a resulting decrease in recombination levels. To start, we designed complementary \textit{attB} oligonucleotides containing contiguous three base pair substitutions in the B core site and measured the effect of these changes in the \textit{in vitro} integration assay. Recombination with wild type \textit{attDOT} and \textit{attB} substrates averages between 35-40% due to day-to-day variation in IntDOT activity. When base pairs at
positions -6 to -4 of attB are changed to the complementary sequence (TTT to AAA), the recombination levels dropped to about 2%, suggesting important interactions had been disrupted. Mutation of the non conserved base pairs at positions -9 to -7 resulted in a reduction of recombination by 10% but with a relatively large margin of error that suggests these bases are likely not as important for binding. When base pairs at positions -11 to -9 are changed, recombination plummets to < 1%, which indicates that one or more of these positions is important for IntDOT interactions (Figure 3).

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

Analysis of the sequence of attB1 showed an imperfect direct repeat of the B core site sequence just upstream of the B site (Figure 4). This arrangement is similar to the FRT site recognized by Flp recombinase. The FRT site contains a direct repeat upstream of the left core site that is not required for recombination (3). The repeat in
attB1 could potentially be acting as an alternative core site and might increase or decrease IntDOT binding at the B core site during recombination. In order to determine what role, if any, this potential site may play in integration we disrupted three bases at a time and measured the effect in the in vitro integration assay (Figure 4). None of the mutations had any significant effect on recombination levels, indicating that the site is not required for integration.

Effects of Core Site Single Base Pair Mutations on Recombination

The results of the earlier mutagenesis experiments suggest that at least some of the conserved base pairs in the B and D core sites are important for IntDOT recognition and binding. We wanted to identify the specific bases within this sequence that are likely to be important for IntDOT recognition of the sequence. To identify important bases, we made single base pair changes along the 8 bases of the 10 bp conserved sequence that are not part of the overlap region. Each base was changed to its complement and the mutated substrate tested for recombination in the in vitro integration assay. The results showed that mutations at positions -5, -9 and -10 had the greatest effect on recombination (Figure 5). Positions -7 and -8 are not conserved amongst the core sites and mutations at these positions had no effect. We chose to analyze more substitutions positions -5, -9 and -10 by changing the remaining two base pairs at those positions so that all four base pairs would be tested at each location. If IntDOT is making direct contact with a specific base pair, then changing the bases at that position should disrupt interactions with IntDOT and result in substantially decreased recombination levels. Alternatively, if IntDOT interacts with the phosphate backbone, we would not expect to see dramatic changes in integration regardless of which base pair is present at that
position, unless we are disrupting structural features of the DNA that IntDOT recognizes (18). A third option is that IntDOT is making water mediated interactions with the DNA. We expect these interactions to be more tolerant of base pair substitutions, as water molecules are free to reposition or re-orient themselves for favorable hydrogen bonding interactions (18). Therefore, it is difficult to say with certainty whether or not water mediated interactions exist at a given position as opposed to direct or backbone interactions.

We substituted A, C, and G in place of the T that is normally present at position -10. As shown in Figure 5A, recombination levels with a T-A base pair at position -10 average 35%. When a G-C base pair is substituted at this position, recombination decreases by 10%. An A-T or C-G base pair at this position is not well tolerated and reduces recombination levels to 5-10%. Variants with substitutions at position -9 had minimal effects. The least effective base pair, C-G, only decreased recombination by 5% (Figure 6B). This leads us to conclude attB sites with changes at position -9 do not disrupt favorable interactions with IntDOT, possibly suggesting interactions with the phosphate backbone instead of direct contact with the bases. In contrast, position -5 appears to be the most important base of the conserved sequence. All substitutions at this position result in a significant loss of recombination (Figure 6C) with the greatest effect seen when a G-C base pair is present (<1%) followed by a C-G base pair (2%). The dramatic results seen with substitutions at position -5 are consistent with direct interactions between IntDOT and the T-A base pair at that position.

We confirmed the importance of the T-A base pair at position -5 in attDOT by making mutations at positions -4, -5, and -6. Single base mutations to the
complementary sequence at positions -4 and -6 did not affect recombination levels but
the single base mutation (-5A) at position -5 reduced recombination threefold (Figure 7).
Double (-4A, -5A and -5A, -6A) and triple (-4A, -5A, -6A) base mutations that include
position -5 completely abolished recombination. This supports our earlier hypothesis
that base pairs at position -5 in both attDOT and attB are interacting directly with
IntDOT.

Effect of Core Site Mutations on Cleavage

The integration reaction is a complex, multistep process consisting of DNA
binding and intasome formation, synopsis, cleavage, strand exchange, and ligation. In
principle, it is possible to interrupt any step within the reaction. To determine whether
substitutions at position -5 disrupt the IntDOT binding step, we performed electro-
mobility shift assays (EMSA). Unfortunately, the attB site does not shift, and the attDOT
core shifts poorly at best (12). Because of this, we could not accurately compare binding
between the wild type sequence and sites with substitutions at position -5.

Although we were unable to demonstrate that the DNA binding step was
disrupted, by using a simple cleavage assay that only requires two steps – binding and
cleavage, we could determine whether substitutions at position -5 inhibit one of the early
steps of recombination. In our cleavage assay, IntDOT binds and irreversibly cleaves a
suicide substrate to become covalently linked to the DNA. This product can be detected
by denaturing gel electrophoresis. We tested D site variants containing either G, A, or T
(WT) at position -5. As a control, we also tested a substitution at position -7 that was not
expected to have an effect.
Assays done with the wild type substrate averaged 55% cleavage. When position -5 is changed from T to A, cleavage is reduced to 22%. When a G is present at position -5, there is no detectable cleavage. The control reaction with a substitution at position -7, where C is changed to A, was identical to wild type with 55% cleavage. Because this reaction requires only binding and cleavage, it is likely that the base pair at position -5 is either important for binding or for positioning of the catalytic tyrosine for catalysis.

Effect of Core Site Mutations on Orientation of Recombinant Products

The results described above indicate that IntDOT interactions at position -5 of the B core site and +5 of the B’ core site are important for integrative recombination.

Sequence alignment of the six known attB sites reveals that position +5 of the B’ core sites contains a conserved T-A base pair one base away from the bottom strand cleavage site. This T-A base pair is also present at position +5 on the attDOT site. This location is spatially equivalent to the T-A base pair at position -5, yet triple mutations of the B’ core site containing a change at position +5 do not have the same effect as triple mutations containing a change at position -5. This could suggest that the primary recognition and binding sequence of the attB site is contained in the B core site.

What role then does the B’ core site, specifically the conserved position +5 play in integration? We propose that a T-A base pair in the core site adjacent to the GC dinucleotide within the overlap region may be important for a productive first strand exchange and allow recombination to proceed. Synapsis between the intasome and the attB can occur in one of two orientations, but the GC dinucleotides in the overlap regions of attDOT and attB must be properly aligned for a productive initial strand
exchange. This determines the alignment of the \textit{attDOT} and \textit{attB} sites that form the recombinant products. We have shown previously that the location of the conserved GC dinucleotide within the overlap region dictates the site of the first productive strand exchange by providing the homology required for the first strand exchange (14). In the same paper, we showed that an \textit{attB} site containing a symmetric overlap, where the GC dinucleotide is present adjacent to both the top and bottom strand cleavage sites, produced recombinants resulting from synapsis in both orientations.

Using this symmetric overlap \textit{attB} site as a template, we changed the conserved T-A base pairs of the B or B’ core sites to G-C at positions -5, +5, and a double mutant containing G-C at both positions. We used SspI restriction digest of the recombinant product to determine the orientation of the integrated element. In a normal reaction between wild type \textit{attDOT} and top strand labeled \textit{attB} substrates, the restriction digest produces a 1.1 kb fragment. In a reaction between a wild type \textit{attDOT} and an \textit{attB} site containing a symmetric overlap region, there are two restriction fragments: 1.1 kb and 2.5 kb. The 2.5 kb fragment corresponds to an integration event that occurred with the \textit{attB} site in the inverted orientation.

Figure 8 shows reactions containing a wild type \textit{attDOT} site and an \textit{attB} site with a symmetric overlap. Recombination with wild type \textit{attB} forms a single recombinant product because only one alignment during synapsis allows the exchange of the GC dinucleotide. An \textit{attB} site containing a symmetric overlap region has the GC dinucleotide present at both ends of the overlap region. Recombination with a symmetric overlap substrate forms recombinants from the two possible alignments of \textit{attB} relative to \textit{attDOT} as shown previously (14). However, recombination with an \textit{attB}...
site containing a symmetric overlap and a mutation at position -5 of the B core site produces only a 2.5 kb restriction fragment. The reaction containing a mutation at position +5 of the B' core site, instead of position -5 of the B core site, produces only a 1.1 kb restriction fragment. When both positions -5 and +5 are mutated, no detectible recombination occurs. From these results, we conclude that productive recombination requires a T-A base pair at position -5 of the B site or +5 of the B' core site. Furthermore, initial strand exchange will occur at whichever of these positions is adjacent to the homology within the overlap, explaining how an attB site with a symmetric overlap region can form recombinants in both orientations.

Base Analog Substitutions in attB

2-aminopurine

Base analogs are modified versions of the canonical bases. They can be used to introduce or remove functional groups into the major or minor groove of the DNA, which can potentially identify important contact positions of DNA binding proteins. We used base analogs to help determine whether IntDOT is interacting with the major groove or the minor groove to identify core binding sites. The base analog 2-aminopurine (2AP) is structurally similar to A and forms base pairs with T (Figure 9). The difference between A and 2AP is the location of the amino group in the DNA. In A, the amino group is located in the major groove while in 2AP the amino group is located in the minor groove. We inserted 2AP where A is the normal base at position -5 in the bottom strand, at position -10 on the bottom strand, and at position -7 on the top strand to act as a control. Position -7 is not conserved and single base mutations at that site have no
effect on recombination. In all cases, the 2AP was paired to a T in the complementary strand.

Having a 2AP at position -10 did not significantly reduce recombination levels (25% compared to 25-30% with the A at position -10) (Figure 9). In contrast however, the 2AP at position -7 actually increased recombination levels by twofold to 55%. Sequence comparisons show that most attB sites contain either an A or a T at position -7. Perhaps the presence of the base analog produced more favorable interactions between IntDOT at that location than a canonical A-T base pair. The 2AP at position -5 also had a significant effect on recombination levels, reducing them threefold to less than 9% (Figure 9). These results by themselves do not show whether the loss in recombination is due to the loss of the amino group from the major groove or the addition of the amino group to the minor groove.

2,6-diaminopurine

To further investigate whether the decrease in recombination seen with 2AP at position -5 was possibly due to the loss of a functional group from the major groove or the gain of a functional group in the minor groove, we used another base analog: 2,6-diaminopurine (DAP). This base analog contains amino groups in both the major and the minor grooves and pairs with T. The T/DAP base pairing also resembles a G/C base pair with three sets of hydrogen bonds instead of two (Figure 10). Together with the 2AP results, the DAP should help distinguish between the major and minor groove interactions because it structurally resembles A but has an additional amino group in the minor groove. We designed attB sites with DAP on the bottom strand of position -5, the top strand at position -7, and at the bottom strand of position -10.
Recombination levels with an attB containing a DAP at positions -7 and -10 ranged from 20-25% which is similar to attB sites with A at positions -7 and -10. Having a DAP at position -5 increased recombination to wild type levels (Figure 10), with recombination at 15-20%, similar to wild type reactions. In contrast, having a 2AP at this position decreased recombination to less than 9% (Figure 9). This suggests that the presence of the amino group in the major groove is important for binding of IntDOT and suggests that there is direct contact between the protein and the DNA at this position.

Discussion

Interactions with both core- and arm-type sites are necessary for formation of the catalytically active CTnDOT intasome complex (9, 19-21). We are particularly interested in the intasome formed by IntDOT because the protein is able to tolerate extensive heterology between the overlap regions of the attDOT and attB DNA substrates. While it is likely that the DOT intasome is structurally very similar to other intasomes, like the lambda intasome for example, the mechanisms for strand exchange and Holliday junction processing differ. Lambda integrase (and most tyrosine recombinases) recombines attB and attP sites efficiently only when the overlap regions contain complete homology. It has been proposed that sequential homology-dependent steps, rather than classic branch migration, occur between the strand exchanges at the top and bottom strands of the substrates (22, 23). The attDOT site of CTnDOT integrates site-selectively into multiple attB sites in the Bacteroides chromosome despite extensive heterology between the overlap regions. The only homology requirements are that the two nucleotides adjacent to the site of the first strand exchange in the attDOT and attB
site needs to be the same in both substrates. The remaining 5 bp in \textit{attDOT} and \textit{attB} can differ. These differences in homology requirements between most tyrosine recombinases and IntDOT indicate that processing of HJs is different for each pathway. Another difference between most tyrosine recombinase systems and the IntDOT system is the nature of the core sites. How IntDOT interacts with core sites may provide insight as to the differences between the two types of reactions. Each of the \textit{attB} sites recognized by IntDOT contains a conserved sequence in the B core site (GTANNTTT), which is also conserved in the D core site of the \textit{attDOT} site. There is poor conservation of the B' core site sequences. We suspect that the conserved sites may be important for recognition of and binding to target \textit{attB} sites during formation of the intasome and capture of the \textit{attB} site.

To confirm the importance of the conserved core sequence, we made a series of triple mutations in the B core site from positions -11 to -4 changing each trio of bases to the complementary sequence. We did not expect mutations containing positions -7 and -8 to have much effect, as the sequence at those two positions is not conserved amongst the core sites. The remaining six base pairs at positions -11 to -9, and -6 to -4 are strongly conserved, and triple mutations that disrupt positions -9, -10, and -11 or positions -4, -5, and -6 of the B core site abolish recombination. These results support our hypothesis that the conserved sequence is important for IntDOT recognition and binding. By substituting the complementary base pairs at those positions, we have introduced enough unfavorable interactions between the DNA and IntDOT that recombination does not occur at normal levels.
To further narrow down which bases are important for interactions with IntDOT, we made single base mutations of the six bases identified as important from the triple mutations (-11, -10, -9, -6, -5, and -4) with each position changed to its complementary sequence. Base pairs at positions -4, -6, and -11 do not appear to be important, as single changes at these positions showed no phenotype. These three positions were not further studied.

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

The T-A base pair normally present at position -5 is the only base pair that is well tolerated, as substitutions at this position reduce recombination from undetectable to about 50% of wild type levels depending on the base pair (Figure 6C). This appears to be the single most important position of the entire conserved sequence. Base analog experiments substituting a 2-aminopurine in place of the A normally present reduces recombination to about 10%. When the amino group is restored in the major groove with a substitution of diaminopurine, recombination returns to wild type levels. This suggests
an important role for the amino group of A in the major groove, which may be involved in direct interactions with IntDOT. Unlike most other tyrosine recombinase systems, the sequence of one of the core sites recognized by IntDOT is conserved in both sites. The B core sequence on \textit{attB} is also shared by the D core site on \textit{attDOT}. The D' and B' core sequences are very poorly conserved. It is interesting to note that the T-A base pair present at position -5 of the B core sites is also conserved at the equivalent position, +5, among all of the B' core sites. Despite the poor overall sequence conservation of the B' core site, could position +5 be functionally equivalent to position -5 on the B core site?

The experiments with the \textit{attB} site containing a symmetric overlap region support the idea that the T-A base pair at position -5 is important for integration. Mutations changing this position to a G-C base pair abolish recombination in the normal orientation of the \textit{attB} site, but still allow recombination in the inverted orientation, suggesting that the T-A base pair located at position +5 on the B' core site can compensate as long as homology exists within the overlap region adjacent to the B' cleavage site.

Similarly, a substitution at position +5 in the B' core site prevented integration in the inverted orientation even when there was homology in the adjacent overlap region. This suggests that interactions between IntDOT and positions +5 and -5 occur during strand exchange. It is important to note, however, that a substitution at either position does not block recombination if it is at the site distal to the first strand exchange. Presumably, the intasome containing four IntDOT monomers is still able to form and
perform recombination with one of the monomers bound to the mutated site. Thus the second strand exchange does not require the interaction at position +5 or -5.

Because recombination is a complex, multistep process, there are several steps that could be affected by a substitution at position +5 or -5. As discussed above, one possibility is that the IntDOT monomers interact differently with the core sites adjacent to the site of initial strand exchange as opposed to the interaction of the monomer that performs the second strand exchange. Alternatively, interactions at these positions may be required for proper positioning of the catalytic tyrosine. Until we have a crystal structure, we cannot say for certain which step is affected by loss of interactions with positions +5 or -5.

Studies done with Fip, Cre, and lambda Int show that these proteins do not make many direct contacts with the DNA. Crystallization studies done with Fip showed that most of the contacts made between the protein and the DNA are through the phosphate backbone as opposed to the bases (24). The direct contact that Fip does make with the FRT DNA site appears to be with the T-A base pair at position -13 and the C-G base pair at position -11 (24). Both of these base pairs are conserved across both core sites. When position -11 was changed, recombination levels were severely affected (25). Guo et al. showed that most interactions between Cre and loxP were water mediated (26).

Studies with Cre showed that there is only one direct contact between the protein and the major groove of the DNA – a C-G base pair at position 10 which is conserved in both core sites (27). The same study also showed that Cre makes some contacts with bases via the minor groove at positions -16 and -17 (27). Crystallographic studies done with lambda Int showed numerous contacts with the phosphate backbone but few direct
interactions with the DNA bases (28). There are some van der Waals interactions between the C5 methyl group of 4 Ts and two residues that directly contact base pairs that are conserved among the four core sites (28). These limited interactions may help explain why lambda Int has such a low affinity for its core sites. The stronger interactions seen with IntDOT and its core sites more closely resembles the interactions between Flp and it's target FRT, and Cre and it's target loxP.

Acknowledgements

The authors would like to thank Jillian Waters and Margaret Wood for their comments on the manuscript.

This work was supported by the National Institutes of Health (GM-28717).


Figure 1

A. 

B. 

Figure: A) Schematic representation of the circular CTnDOT element and the bacterial attachment site attB located on the Bacteroides chromosome. The core sites are shown in the grey box. B) Sequences of the attD (top) and attB (bottom) core sites. IntDOT catalyzes sequential strand exchanges in the overlap regions (boxed) during integration. This reaction requires IntDOT and a host factor (HIF).
Figure 2

D: GTAACCTTT
B1: GTAAATTT
B2: GTATTTTT
B3: GTACTTTTT
B4: GTAGATTT
B5: GTTGTTTTT
B6: GTATATTTT

Consensus:

GTANWTTTT

Figure 2: Alignment of the D site in attDOT and the B sites of all the known attB sites (5). The consensus sequence is shown underlined at the bottom. W = A or T, N = any base.
Figure 3. A) The core sites and overlap region of \textit{attB}1. Core sites B and B' are shown as inverted repeats. The boxed area is the overlap region. The C in the center of the overlap region is designated as position 0. B) Results of integration assays with \textit{attB} substrates containing triple base pair substitutions in the B and B' core sites. The bases were changed to their complement.
Figure 4

A. 

[Diagram showing DNA sequences and substitutions]

B. 

[Bar graph showing average % recombination]

Figure 4. A) The direct repeat sequence of the 10 bp conserved sequence just upstream of the B-core site is shown as a thick arrow covering positions -22 to -13. The indicated triple substitutions shown below the sequence were constructed and the effects on integration were measured. B) Results from integration assays done using attB substrates containing the upstream substitutions.
Figure 5. Integration assay results of attB sites with single base substitutions along the B core site. Each base was changed to its complement.
Figure 6

A) Integration assay results of attB sites with single base substitutions of the B core site at position -10. B) Results with an attB containing substitutions at position -9. C) Results with an attB containing substitutions at position -5.
Figure 7: Integration assay results from attDOT substrates with the D core site containing single substitutions (-4A, -5A, and -6A), double substitutions (-4A, -5A and -5A, -6A), and a triple substitutions (-4A, -5A, -6A).
Figure 8

A) Possible orientations of an attB substrate relative to the attD site and resulting recombinant products. After recombination, the orientation of attD relative to attB can be determined by cutting with SspI. If the orientation of attB is as shown on the left, a 1.1 kb fragment is produced. If the orientation of attB is as shown on the right, a 2.5 kb labelled fragment is produced. B) Integration assay with substrates containing attB sites containing a symmetric overlap region with substitutions at positions -5 or +5 and both -5 & +5.
Figure 9. Integration assay results of 2-aminopurine substitutions in attB1 at positions -5, -7, and -10 of the B core site. The 2-aminopurine is on the bottom strand at positions -5 and -10 and on the top strand at position -7 so that there is always a 2AP-T base pairing.
Figure 10. Integration assay results of attB1 substrates containing diaminopurine (DAP) substitutions at positions -5, -7, and -10 of the B core site. Substitutions at positions -5 and -10 are on the bottom strand while -7 and -9 are on the top strand. In all cases, the DAP is paired to a thymine.