THE ROLE OF BACTEROIDES HOST FACTOR A IN CTnDOT INTEGRATION AND
HOLLIDAY JUNCTION RESOLUTION

BY
KENNETH WAYNE RINGWALD

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, 2016

Urbana, Illinois

Doctoral Committee:
Professor Jeffrey Gardner, Chair
Professor James Imlay
Professor Andrei Kuzminov
Professor Peter Orlean
Abstract

*Bacteroides* are a common genus of Gram negative anaerobic bacteria within the human gut. Under normal conditions, *Bacteroides* spp. benefit the host by breaking down complex polysaccharides. However, when the gut is punctured, these species may escape to cause abscesses or infections within the blood. In the past, tetracycline was used as a treatment for these infections. However, *Bacteroides* opportunistic infections are now much more difficult and costly to treat due to the widespread presence of mobile genetic elements carrying antibiotic resistance genes.

One of these mobile genetic elements, CTnDOT, moves by conjugation and site-selective recombination. It is an example of a Conjugative Transposon (CTn). These elements were later classified as Integrated Conjugative Elements (ICEs). They frequently carry resistance to antibiotics and always contain the genes to mediate their own transfer by conjugation. CTnDOT is 65 kb and carries genes for mobilization, transfer, integration, and excision in addition to resistance to tetracycline and erythromycin. Once integrated into the chromosome, ICEs such as CTnDOT are stably maintained.

The integration reaction into the *Bacteroides* chromosome and the excision reaction necessary for transfer are catalyzed by an integrase, IntDOT. IntDOT is a member of the tyrosine recombinase family. It was previously known that a host factor was required for integration and that *Escherichia coli* Integration Host Factor (IHF) could substitute for the host factor in an *in vitro* integration assay. I purified DNA binding proteins from a *Bacteroides thetaiotaomicron* strain lacking known ICEs. The purified
fractions were then tested in the in vitro integration assay. The active fraction contained a protein that we named BHFa for Bacteroides Host Factor A. Subsequent electrophoretic mobility shift assays and fluorescent footprinting assays revealed four BHFa binding sites within the attDOT DNA sequence of CTnDOT. Surprisingly, further experiments showed that other DNA bending proteins could effectively substitute for BHFa in the in vitro integration assay even when distantly related or entirely unrelated.

The integration and excision reactions of CTnDOT and other tyrosine recombinases proceed through ordered strand exchanges. The first set of strand exchanges generates a Holliday junction (HJ) intermediate that isomerizes through an overlap region and which is resolved by a second set of strand exchanges. Most tyrosine recombinases require identical DNA sequences in the overlap region where the strand exchanges occur. However, IntDOT can resolve HJs containing mismatched bases in the overlap region in vivo. This ability implies a difference in the overall protein-DNA complex, called an intasome. In integration, the intasome involves at least two different proteins and two double stranded DNA molecules, attDOT and attB. For excision, the attL and attR DNA sites are required for the assembly of two separate excisive intasomes before being brought together to reform the circular element. As many as five proteins may participate in excision and four are required.

In order to study the mechanisms of the integration and excision reaction, we constructed synthetic HJs. HJ intermediates were constructed by annealing four oligonucleotides from the different products and substrates. This creates the intermediate formed after the first set of strand exchanges. The synthetic HJs can be constructed with either identical or mismatched overlap regions. It had been previously shown that
synthetic HJs containing only the IntDOT core sites could be processed to both products and substrates if the overlaps are identical \textit{in vitro}.

However, if the overlaps contain mismatches, the HJs are resolved back to substrates. This inability of IntDOT to process mismatched HJs to products may be due to the lack of arm-type sites (which IntDOT binds) and because other protein participants were absent. I hypothesized that BHFa (or another DNA-binding protein) and the arm-type sites are necessary to form the higher-ordered complexes called intasomes and to enable IntDOT to catalyze recombination through the mismatched bases.

To test this, I have constructed larger synthetic HJs (composed of four annealed oligonucleotides) containing the arm-type sites. As with core-only HJs, they were resolved into either products or substrates when the overlap contains identical bases. I have also constructed core plus arm-type HJs with a mismatched overlap. When these HJs are incubated with both IntDOT and BHFa they are resolved to both substrates and products. Accordingly, it appears that both arm-types sites and BHFa are required to enable IntDOT to resolve a HJ with a mismatched overlap region.
Table of Contents

CHAPTER 1: Introduction .............................................................................................................. 1
  Bacteroides and Conjugative Elements ....................................................................................... 1
CTnDOT ...................................................................................................................................... 3
  Integration .................................................................................................................................. 3
  Excision and transfer .............................................................................................................. 5
  Regulation of the excision, transfer, and mobilization operons .............................................. 7
Host Factor Proteins ..................................................................................................................... 8
Tyrosine Recombinases ............................................................................................................... 9
  Lambda Int ............................................................................................................................. 11
  Arm-type sites in the lambda system and in CTnDOT .............................................................. 12
  Homology .............................................................................................................................. 13
Synthetic Holliday Junctions ......................................................................................................... 15
Thesis Outline ............................................................................................................................ 16
References .................................................................................................................................. 18
Figures ....................................................................................................................................... 26

CHAPTER 2: The Bacteroides thetaiotaomicron encoded protein Bacteroides
Host Factor A participates in integration of the Integrative Conjugative Element
CTnDOT into the chromosome ..................................................................................................... 34
Abstract ...................................................................................................................................... 34
Introduction ................................................................................................................................ 34
Materials and Methods ............................................................................................................... 37
Results ........................................................................................................................................ 42
  Identification of Bacteroides host factor .................................................................................. 42
  BHFa binds specifically to the attDOT DNA sequence .......................................................... 45
  Complexes with BHFa and IntDOT ....................................................................................... 46
  Identification of the binding sites of BHFa .......................................................................... 47
  EMSA Analysis of BHFa binding to the H1, H2, and H3-H4 sites ......................................... 49
  DNA bending proteins promote the in vitro integration reaction ....................................... 50
Discussion .................................................................................................................................. 51
Acknowledgments ...................................................................................................................... 55
References .................................................................................................................................. 56
Figures ....................................................................................................................................... 59
Tables ......................................................................................................................................... 69
CHAPTER 1: Introduction

_Bacteroides and Conjugative Elements_

_Bacteroides_ species are Gram negative, obligate anaerobes that compose 20-30% of the normal microbiota in the human colon (3, 15). In general, _Bacteroides_ spp. benefit their hosts by breaking down complex polysaccharides and by limiting the ability of pathogens to colonize the gut (97, 100). However, when the tissue of the gut is punctured, _Bacteroides_ may act as opportunistic pathogens capable of causing abscesses in other regions of the body or infections in the blood (bacteremia). _B. fragilis_ is the most frequently isolated anaerobic cause of bacteremia (9, 97).

The resulting infections are difficult to treat due to the widespread resistance to antibiotics found in _Bacteroides_ spp. Many _Bacteroides_ species are resistant to cefoxitin and clindamycin (28, 68). The resistance genes for tetracycline (tetQ) and erythromycin (ermF) are carried by conjugative transposons found in _Bacteroides_, and the frequency of resistance has increased dramatically in the last thirty years (82). Certain IS elements, another type of mobile genetic element, can also activate the cfiA and nimD genes that confer resistance to more recently preferred treatments, such as metronidazole and meropenem (29).

Conjugative transposons, also known and subsequently referred to as Integrated Conjugative Elements (ICEs), are horizontally transferred genetic elements (11). ICEs are clinically relevant due to their ability to transfer antibiotic resistance or pathogenicity islands between bacterial species via conjugation (17). ICEs encode genes to integrate into the host chromosome, excise from the chromosome, and ultimately transfer a copy
into a new recipient cell. Mechanistically, this life cycle means that ICEs have traits in common with both conjugative plasmids and temperate bacteriophages.

In order to transfer, ICEs excise from the host chromosome to form a circular intermediate in a manner similar to lysogenic phages, such as lambda. Both excision and transfer are regulated by the element and require specialized proteins. The excised circular intermediate then transfers a single strand via conjugation (as with conjugative plasmids) using a Type IV secretion system as the mating pore. After the transfer, the single strand presumably replicates and the double stranded circular intermediate integrates into a new site in the recipient chromosome (Figure 1.1) (36, 77). After integration, the ICE is replicated along with the host DNA (12). Integration requires an element-encoded integrase (either a tyrosine or serine recombinase) and a DNA-bending protein. Tyrosine recombinases are more common and, in another similarity to phages, are members of the same family as the bacteriophage lambda integrase. The necessary DNA-bending protein is often a nucleoid-associated host factor that has been repurposed by the ICE.

ICEs can vary widely in size, from about 20 kb to over 500 kb and frequently include genes for additional functions besides transfer, integration, and excision (29). The genes (sometimes called cargo genes) confer phenotypes to the host cell, such as antibiotic resistance, heavy metal resistance, toxin-antitoxin systems, or metabolic pathways (36). Historically, ICEs were discovered due to experimentally obvious phenotypes provided to the host cells, particularly antibiotic resistance (24, 81). Sequencing of chromosomes has made it possible to identify ICE-like sequences even when the cargo genes do not provide obvious phenotypes or are genes of unknown
function. Further, ICEs such as CTnDOT regulate the expression of numerous chromosomal genes of their host (60). In this way, ICEs can further influence the behavior of the host cell. ICEs can mobilize other co-resident genetic elements in the cell (54, 76). These genetic elements, such as NBU1 in Bacteroides, may encode their own integrase proteins but lack functional excision or transfer operons (79, 83, 99).

Due to their utility to the bacterial host and ability to transfer between strains or even species within the same phylum, ICEs are widespread in both Gram-positive and Gram-negative bacterial populations (82). An extreme case, Tn916, is capable of transferring between Gram-positive and Gram-negative species despite the different physiologies involved (6). In general, ICEs can transfer among diverse organisms, though the limitations on their host range are not yet well understood. After integration into the host chromosome, they are retained even in the absence of selective pressure and are thus inherited vertically, as well as horizontally.

**CTnDOT**

**Integration**

After CTnDOT has transferred to a recipient cell, the element must integrate into the host chromosome to persist. Integration requires the tyrosine recombinase IntDOT and a host factor, BHFa, to serve as the DNA binding protein. The required DNA substrates, called att sites, are the joined ends of the circular element, attDOT (324 bp), and a suitable attB site within the bacterial chromosome (Figure 1.2). It was initially thought that there were 10 or fewer attB locations within the *B. thetaiotaomicron*
chromosome (18). However, more recent work suggests there are up to 20 additional
attB sites (Hopp, unpublished results). Both attDOT and attB contain core sites which
are bound by IntDOT. IntDOT also interacts with arm-type sites found within attDOT,
described later. The core sites are imperfect inverted repeats that flank the overlap
region. The overlap region is where the steps of cleavage, strand exchange, and ligation
occur to generate the HJ intermediate. The core sites on attDOT are arranged as D –
overlap - D’ and the attB is arranged as B – overlap - B’ (Figure 1.2). All known B and D
core sites contain the sequence (5’ - GTANNTTT -3′) (18, 51). The B’ sites are less
strictly conserved.

The attDOT DNA must be double-stranded and supercoiled for integration to
proceed (19). First, the host factor and IntDOT bind to the attDOT DNA to form the
higher order nucleoprotein complex called the intasome. The intasome then captures the
attB to make the initial set of strand exchanges and form the HJ intermediate. IntDOT
makes two sets of strand exchanges 7 bp apart in both attDOT and in the attB sites.
These strand exchanges flank the overlap region, which is boxed in Figure 1.2. Due to
the GC dinucleotide, there are 2 bp of identity on the left side (18, 50, 56) (Figure 1.2).
Homology is required at this location, presumably for the ligation to proceed.

The rest of the attB and attDOT sequences may not be complementary, which
generates heterology in the HJ intermediate. After isomerization, IntDOT is able to carry
out the second strand exchange even in the presence of heterology, but it is unusual
among tyrosine recombinases in this respect. The second strand exchanges yield the
recombinant attL and attR products (Figure 1.3). If the sequences were not
complementary, heteroduplexes are generated at \textit{attL} and \textit{attR}, which must be resolved later by either mismatch repair or chromosomal replication (Figure 1.3) (56, 72).

**Excision and transfer**

CTnDOT excision is more complex than other ICEs, involving five proteins: IntDOT, Xis2c, Xis2d, Exc, and a host factor (91, 92). Other ICEs such as Tn916 are comparatively simple, requiring just one element-encoded protein and host-encoded proteins in order to excise (21). The DNA substrates of CTnDOT excision are the ends of the integrated element, \textit{attL} (420 bp) and \textit{attR} (220 bp), which are separated by 65 kb in the chromosome (23).

Excisive intasomes are assembled at both \textit{attL} and \textit{attR}. The intasomes are then brought together to undergo synapsis. Similar to integration, IntDOT carries out two sets of strand exchanges 7 bp apart at both \textit{attL} and \textit{attR} during excision. Some of these bases may not be complementary once brought together in \textit{attDOT}, creating 5 bp of heterology (Figure 1.3). Excision restores the circular intermediate of CTnDOT and the \textit{attB} site in the \textit{Bacteroides} chromosome (Figure 1.3).

Xis2c and Xis2d are small, basic proteins that act as recombination directionality factors (53). Both Xis2c and Xis2d participate in the intasomes and could interact with IntDOT during excision. The \textit{attR} excisive intasome contains Xis2c, Xis2d, and IntDOT (42). In the \textit{in vitro} excision assays, a host factor is required. While \textit{E. coli} IHF and other DNA-bending proteins can substitute in the \textit{in vitro} integration reaction, only BHFa enables \textit{in vitro} excision.
Exc is a type IA topoisomerase that can relax supercoiled DNA\(^\text{84}\). Exc was originally thought to be essential for excision \textit{in vivo} but later studies showed that it plays a stimulatory role (41, 84). Despite being identified as a topoisomerase, an Exc mutant lacking the catalytic tyrosine and therefore incapable of relaxing DNA still stimulates excision (84). Therefore, the role of Exc in excision may be to coordinate other excision proteins or to stabilize intermediate complexes. Exc has not been shown to bind \textit{attL}, \textit{attR}, \textit{mob}, or \textit{tra} DNA by itself, but it does interact with Xis2d when stimulating the \textit{mob} operon (32).

The process of transfer by conjugation requires both mobilization (\textit{mob}) and transfer (\textit{tra}) operons in order to construct the mating pore and move the DNA into the recipient cell. The mating pore, encoded by an operon consisting of \textit{traA-traQ} in CTnDOT, is part of a type IV secretion system that transfers a single DNA strand of the ICE into a recipient (35). The protein products of the \textit{mob} operon (consisting of \textit{mobA}, \textit{mobB}, and \textit{mobC}) are responsible for shuttling the circular intermediate of CTnDOT to the mating pore and nicking the DNA at the \textit{oriT} site (55, 95). Presumably, CTnDOT acts similarly to well-studied conjugative plasmids, such as F and Ti, and is then transferred through the mating pore as a single strand. The transferred strand is replicated in the recipient cell and the resulting double stranded circular intermediate integrates into the recipient cell’s chromosome.
Regulation of the excision, transfer, and mobilization operons

To regulate initiation of excision and transfer, CTnDOT controls expression of the excision proteins Exc, Xis2c, and Xis2d. These proteins are part of the excision operon containing \( \text{xis2c}, \text{xis2d}, \text{exc} \). The excision operon is indirectly induced by tetracycline and by a two-component regulatory system consisting of the proteins RteA and RteB. The regulatory operon contains \( \text{tetQ-rteA-rteB} \) and is activated by a translational attenuation mechanism (Figure 1.4); the presence of tetracycline slows the ribosome and causes the formation of alternate hairpins within the mRNA (90, 91). RteB activates transcription of \( \text{rteC} \), then RteC activates the transcription of the excision operon. The overall effect is that tetracycline stimulates transfer of the element 1000-fold.

In addition to their direct roles in excision, Xis2c, Xis2d, and Exc positively regulate the \( \text{tra} \) and \( \text{mob} \) operons. Xis2c activates transcription of the \( \text{tra} \) operon (7, 42). Xis2d activates transcription of both the \( \text{tra} \) and \( \text{mob} \) operons, while Exc is required for transcriptional activation of the \( \text{mob} \) operon (Figure 1.4) (32, 35, 43, 63, 95). CTnDOT also uses negative regulation in the form of the small RNA, RteR. The DNA encoding RteR overlaps with a portion of \( \text{exc} \) but has its own promoter and is expressed constitutively rather than being regulated by the presence of tetracycline (94). RteR negatively regulates full transcription of the \( \text{tra} \) operon, possibly by increasing the frequency of transcriptional termination (94). However, RteR does not affect the excision or the mobilization genes.
Host Factor Proteins

Many transposition and site-specific recombination systems require host factors. Often the host factors for recombination systems are nucleoid-associated proteins within the cell. These proteins share several qualities and help maintain the bacterial chromosome. In general, nucleoid-associated proteins are small and basic in order to facilitate interactions with DNA (10). They may bind to specific DNA sequences or characteristic structures and introduce sharp bends to DNA. Two well-studied examples of nucleoid-associated proteins are IHF and HU. In *E. coli* and many other bacteria that contain both proteins, IHF and HU are homologs. Along with other small, basic proteins such as H-NS and FIS, both IHF and HU are present at high concentrations in the cell, though the exact concentration often varies by growth phase. For transposable elements, the differing concentrations and compositions of nucleoid proteins may provide a way to sense the growth phase of the host or the level of supercoiling of the host chromosome (48). In general, these proteins play important roles in chromosome partitioning, transcription, regulation, DNA replication, transposition, and protection or repair of DNA (8, 20, 34, 75).

Several nucleoid-associated proteins contribute to other DNA recombination reactions. IHF was first identified because it is necessary for lambda integration (62). IHF is a dimer composed of one α (11 kDa) and one β (9 kDa) subunit with slightly different sequences (61). IHF helps position DNA properly for the formation of intasomes during lambda recombination (62). Under specific conditions, IHF may be replaced by HU in this reaction. Alternatively, if the target DNA sequence encodes specific bends or an Int mutant is used, no DNA bending protein is necessary (25). IHF
binds the specific consensus sequence WATCAANNNNTTR (26). However, not all of these bases are required for IHF binding, and at high concentrations it binds nonspecifically (101).

HU is most often found as a 20 kDa heterodimer of α and β subunits (9.5 kDa each) in E. coli, though αα and ββ homodimers are also present and can bind DNA. Salmonella and Serratia marcescens have heterodimeric HU as well, but many bacterial species have a homodimeric form instead (67). HU is generally considered a nonspecific DNA binding protein (70). However, HU has a higher affinity for DNA containing single strand breaks or recombination intermediates (39, 40). HU plays a role in the S. typhimurium hin inversion system and in Tn916 excision. For Tn916, HU enhances excision from the host chromosome in E. coli (21). In the hin system, three separate sites are brought together to form the invertasome complex: hixL, hixR, and the enhancer region. In the wild type reaction, HU is required to bring the enhancer into place. In the wild type arrangement of these sites, the enhancer is found 99 bp from hixL. If the distance between the enhancer and hixL is greatly increased, HU is no longer required (30). Likewise, FIS (Factor for Inversion Stimulation) is named for its role in DNA inversion in the Salmonella typhimurium and phage Mu systems (37, 38). It plays a role in bacteriophage lambda excision as well (86).

**Tyrosine Recombinases**

IntDOT is a tyrosine recombinase in the same family of enzymes as lambda Int, Flp, XerC, XerD, and Cre (72). IntDOT contains five of the six conserved amino acid
residues that form the catalytic sites of tyrosine recombinases. The conserved amino acid residues are: Arginine, Lysine, Histidine or Lysine, Arginine, Histidine or Tryptophan, and Tyrosine (45, 57, 66). The missing conserved residue of IntDOT is the first arginine, but a non-conserved arginine in another part of the protein is predicted to fulfill that role, giving IntDOT a distinct active site compared to other tyrosine recombinases (45). These enzymes perform strand exchanges by a site specific topoisomerase activity, generating a four-stranded DNA intermediate called a Holliday junction. Cre and Flp are considered simple tyrosine recombinases; they do not require accessory proteins or additional DNA sites (such as arm-type sites) to carry out recombination reactions.

The recombination reaction required for integration or excision of an ICE requires four tyrosine recombinase monomers (Figure 1.5). Two of the monomers cleave the DNA substrates, creating a covalent 3’ phosphotyrosyl intermediate. The resulting free 5’ hydroxyl can carry out a nucleophilic attack on the partner phosphotyrosyl DNA-protein intermediate to form a new phosphodiester bond and generate the Holliday junction (88). The HJ complex can then isomerize by a strand-swapping mechanism until it reaches the necessary conformation for the second strand exchange, carried out by the other pair of monomers. The second strand exchange and subsequent ligation results in the formation of the recombinant products. In some systems (CTnDOT, lambda, Cre) the order of strand exchanges is defined and sequential; the same strand is always exchanged first (31, 33, 46, 47, 57). In other systems, such as Flp, the order of strand exchange is sequentially nonbiased (2).
Bacteriophage lambda has one of the best-studied site-specific recombinase systems, both biochemically and genetically. Based on a complex genetic switch, bacteriophage lambda can integrate into the host *E. coli* chromosome site-specifically to form a lysogen. When conditions change, lambda can then excise from the chromosome again as part of the lytic cycle. The lambda integration and excision reactions involve accessory proteins (both host and phage-encoded) and additional DNA sites called arm-type sites. The lambda integrase, lambda Int, can interact with both the core and arm-type DNA sequences because it is a heterobivalent protein containing three domains. The amino terminal domain (NTD) consists of residues 1-74 and interacts with the arm-type sites described below. The remaining domains, the core-binding and catalytic domains, can act independently of the NTD. The core-binding domain consists of residues 75-160, while the catalytic domain is composed of the remaining residues, 170-356. The core-binding and catalytic domains are connected by a linker region, residues 160-176 (49).

The integration reaction of bacteriophage lambda requires both lambda Int and IHF to bind *attP* of the phage. The core sites flank the overlap region (Figure 1.6). The assembled *attP*-lambda Int-IHF intasome then captures the *attB* site within the *E. coli* chromosome to undergo synapsis (59, 62, 73). Excision requires lambda Int, Xis, and the host proteins IHF and Fis (Figure 1.6) (1, 13, 86). Xis inhibits the integration reaction. Fis was not originally predicted to participate, since its binding site partially overlaps with one of the Xis binding sites, but later studies demonstrated its involvement in excision (69).
Arm-type sites in the lambda system and in CTnDOT

In the bacteriophage lambda system, the arm-type sites regulate the directionality of recombination; different sites are used for integration as opposed to excision. There are a total of five arm-type sites: P1 and P2 (which flank one of the IHF binding sites) and P′1, P′2, and P′3 (Figure 1.6) (74). Only the first 64 residues of the lambda Int NTD are strictly required to bind the arm-type sites. However, cooperative binding to adjacent sites (P′1, P′2, P′3) is not possible without residues 1-70 (78). During recombination, lambda Int determines the order of strand exchanges based on interactions with the arm-type sites (46).

For integration, the P1, P′2, and P′3 sites are required and presumably occupied on attP. Both the H1 and H′ sites are bound by IHF (58, 62). The lambda Int-attP intasome assembles on the phage DNA and captures the naked attB within the E. coli chromosome (73). During excision, the attL and attR sites recombine in order to regenerate attP and attB. attL contains the H′, P′1, P′2, and P′3 sites. Of these, H′, P′1, and P′2 are occupied during excision. attR contains the P1, H1, Xis, Fis, and H2 binding sites. Only P2, the Xis and Fis sites, and H2 are used during excision (Figure 1.6) (14, 64, 85, 87).

The arm-type sites found within the attDOT region (or attL and attR after integration) affect the directionality of CTnDOT recombination. As in the bacteriophage lambda system, IntDOT affinity for the DNA core sites was enhanced when IntDOT is simultaneously binding arm-type DNA (98). There are a total of 6 IntDOT arm-type sites: R1′, R1, R2, R2′, L1, and L2 (Figure 1.7). All of the sites except R1′ were identified based on footprinting studies with IntDOT (22). R1′ was identified by
comparing the sequences of the other arm-type sites and searching the \textit{attDOT} sequence (98). The sites were altered by site-directed mutations and tested for their importance in integration, excision, and binding by IntDOT. For integration, the R1’ and L1 sites were required. Additionally, the cooperative combinations of R1 with R2, R1 with R2’, or R1/R2/R2’ were required for integration. In the excision reaction, no single site was absolutely required for excision. Mutations in R1’, R1, or L1 caused the greatest decrease. Substrates with combined mutations in R1 and R2 or R1 and R2’ showed the strongest decrease in excision. The L2 site was not found to affect either integration or excision (98). Similar defects in excision frequency were observed subsequently with a different \textit{in vitro} excision assay using mutated R1’, R1, and L1 arm-type sites (42).

**Homology**

Most well-studied tyrosine recombinases, such as Int, Flp, and Cre have overlap regions with identical bases in both DNA participants and are unable to resolve HJs containing mismatches. These observations led to the assumption that identity between overlaps of the recombining sites (homology) was strictly required for resolution of the HJ intermediate by site-specific recombinases. As part of this model, pairs of recombinase monomers made the first set of strand exchanges. The resulting junction could move, base by base, through branch migration until it reached the site of the second strand exchange. At that site, two additional recombinase monomers completed the second set of strand exchanges, and the intermediate resolved to products (96). Later experiments based on both genetics and crystal structures challenged the branch migration model.
Instead, a strand swapping isomerization model was proposed to explain the movement of the Holliday junction intermediate in the lambda system (65). In this model, two sequential swaps of three base pairs precede the second strand exchange (89). In order to generate the HJ intermediate, recombinases must bind DNA, cleave at the first strand exchange site, form the phosphotyrosine intermediate and ultimately ligate the DNA strands back together. Subsequent experiments tested which of these steps required homology. Initial cleavage did not require homology in lambda, Cre, or Flp (10, 27, 52, 65).

For example, Lambda Int can catalyze the first set of strand exchanges in the presence of heterology. However, if the heterology is to the right of the first strand exchanges lambda Int cannot complete the strand swapping isomerization step and proceed to the second set of strand exchanges (5, 96). Subsequent experiments showed that lambda Int is unable to ligate mismatched bases. Presumably, when the bases cannot anneal, they cannot rotate to form the proper contacts for Int to catalyze ligation (65).

In the Flp system, the presence of heterology does not inhibit recombination, but prevents resolution to products. Supercoiled substrates containing heterology still underwent two rounds of Flp-mediated recombination, but only the non-recombinant conformation resulted (4). This indicates that, like lambda Int, Flp can cleave mismatched DNA, but cannot ligate if the adjacent bases are heterologous (52).

IntDOT differs profoundly from lambda Int and other tyrosine recombinases because IntDOT can completely resolve HJs containing mismatched bases. IntDOT requires only two base pairs of homology at the site of the first strand exchange (23, 50, 57). The second strand exchange can occur despite up to five base pairs of heterology to
the right of the first strand exchange. In the *in vitro* integration reaction, the reaction proceeds equally well with a matched or mismatched overlap region. Int\textsuperscript{TN916} (from the Tn916/Tn1545 family of ICEs) is another tyrosine recombinase that can resolve mismatched overlap regions (16, 71, 80). In *B. thetaiotaomicron*, the unusual NBU1 integrase IntN1 is actually stimulated by heterology in the overlap region (79, 81).

In the CTnDOT *in vitro* integration assay, whether the overlap region is identical or mismatched, does not affect the efficiency or speed of the reaction. However, in the *in vitro* excision reaction, homology of the overlap region dramatically affects excision frequencies. Without Exc, only about 2\% of the *attL* and *attR* sites recombined when the overlap regions contained heterology. Adding Exc improved the excision to about 14\%, suggesting that Exc can assist in resolution of mismatched bases. However, the yield of the excision reaction increased dramatically when the overlap region was entirely homologous, at about 90\% excision even without adding Exc. With Exc, the efficiency of the excision reaction approached 100\% (41).

**Synthetic Holliday Junctions**

Even in the *in vitro* integration and excision assays, there are numerous participants: two *att* sites (possibly on different molecules) and up to five proteins are required to form intasomes and complete the recombination reactions. In order to better understand the roles of these individual participants, synthetic Holliday junctions were constructed by annealing together labeled oligonucleotides representing the *attB*, *attDOT*, *attL* and *attR* strands of the HJ intermediate (44). For *attB* and *attDOT*, the bottom
strands were used; for \textit{attL} and \textit{attR}, the top strands. The resulting HJs contained either identical or mismatched overlap regions (Figure 1.8). When the overlap region was identical, IntDOT alone could resolve the HJ into either products (\textit{attL} and \textit{attR}) or back to substrates (\textit{attDOT} and \textit{attB}). However, if the overlap region contained mismatches, the HJs were only resolved back to substrates (\textit{attDOT} and \textit{attB}). The inability of IntDOT to process mismatched HJs to products may be due to the lack of arm-type sites in the synthetic HJs.

\textbf{Thesis Outline}

This thesis covers two primary projects. The first project was to identify and characterize the host factor required for integration of CTnDOT using an \textit{in vitro} assay. The second project used the identified host factor, BHFa, along with synthetic HJs containing arm-type sites to gain further insights into how IntDOT can resolve HJs containing mismatched bases in the overlap region.

In chapter 2, I describe the initial purification of DNA binding proteins from a \textit{Bacteroides} strain lacking CTnDOT. These fractions were tested for activity in the \textit{in vitro} integration assay. After identifying the active fraction, I further characterized that protein, which we have named BHFa. I used electrophoretic mobility shift assays (EMSAs) to test BHFa binding to \textit{attDOT} DNA. In addition, I used fluorescent footprinting to identify the four BHFa specific binding sites within \textit{attDOT}. Finally, I tested unrelated DNA binding proteins in the \textit{in vitro} integration assay. I was surprised to
find that, despite the fact that BHFa binds specifically to *attDOT* DNA, the unrelated proteins were able to bind non-specifically and still facilitated integration.

In chapter 3, I describe experiments that test the importance of the arm-type sites and proteins involved in integration and excision on synthetic HJ resolution. Based on our current understanding of the arm-type sites in CTnDOT integration and excision, I was able to design both wild-type and altered arm-type sites in the synthetic HJs. In each case, I constructed HJs that contained either identical or mismatched overlap regions in order to determine the importance of the DNA sites and protein participants in the resolution reaction.
References


60. **Moon K, Sonnenburg J, Salyers A.** 2007. Unexpected effect of a Bacteroides conjugative transposon, CTnDOT, on chromosomal gene expression in its bacterial host. Mol Microbiol **64:**1562-1571.


69. **Papagiannis CV, Sam MD, Abbani MA, Yoo D, Cascio D, Clubb RT, Johnson RC.** 2007. Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. J Mol Biol **367:**328-343.


93. Waters JL, Salyers AA. 2013. Regulation of CTnDOT conjugative transfer is a complex and highly coordinated series of events. MBio 4:e00569-00513.


Figure 1.1: Overview of Integrated Conjugative Elements.

In the top panel, the cell on the left contains an ICE integrated into its chromosome.
In the second panel, the ICE has excised and formed a circular intermediate.
In the third panel, one strand of the ICE is transferred via conjugation through the mating bridge.
In the fourth panel, both single stranded circular forms of the ICE replicate before integrating into the respective host cell chromosomes in the bottom panel.

Figure taken from (93).
Figure 1.2: Recombination between \textit{attDOT} and \textit{attB} core sites.

\textbf{A.} The joined ends of CTnDOT comprise the \textit{attDOT} sequence.

The core sites (D and D') flank the overlap (O) region.

The arrows indicate that the sequences are imperfect inverted repeats.

Within the \textit{Bacteroides} chromosome, B and B' sites (also imperfect inverted repeats) flank the \textit{attB} overlap region.

Numbers are based on the central base of the overlap region, which is 0.

\textbf{B.} Specific sequences for the \textit{attDOT} and \textit{attB} core sites and overlap region.

The leftmost GC within the overlap region is required for the first strand exchange.

B and D sites contain the conserved sequence GTANNTTT.

Figure taken from (51).
Figure 1.3: Overview of CTnDOT excision, intermediate, and integration.

At the top, CTnDOT is integrated into the *Bacteroides* chromosome. After assembly of the excisive intasomes at *attL* and *attR*, CTnDOT excises to form a circular intermediate. This intermediate may include a region of heterology depending on the original *attL* and *attR* sequence. The heterology can be resolved by replication of the element.

After transfer to a recipient cell by conjugation, CTnDOT must integrate into the host chromosome. Integration requires IntDOT and BHFa and may result in heteroduplexes at both the *attL* and *attR* sites flanking the integrated CTnDOT. These may also be resolved either by replication or by mismatch repair.

Figure adapted from (72).
Figure 1.4: Regulation of the Excision, Mobilization, and Transfer operons.

Tetracycline affects the regulatory operon, shown in pale blue. RteA-RteB act as a two component regulatory system.

RteC (activated by RteB) activates transcription of the excision operon, shown in purple. The excision proteins Xis2c, Xis2d, and Exc are expressed.

Xis2c and Xis2d stimulate the transfer operon (orange).

Exc and 2d stimulate the mobilization operon (green).

RteR (shown in black) negatively regulates transcription of the transfer operon. IntDOT (dark blue) is expressed constitutively.

Numbers indicate order of activation.

Figure taken from (93).
Figure 1.5: Resolution of a Holliday junction intermediate.

In the top left, the first two tyrosine recombinases (with catalytic tyrosines marked) cleave the DNA, creating the 3’ phosphotyrosyl intermediates.

The free 5’ OHs (marked in the top center) can carry out the nucleophilic attack on the partner strands, generating the Holliday junction intermediate (top right).

The HJ can then isomerize by strand swapping (bottom right) until the second pair of tyrosine recombinases carry out the second set of cleavages (bottom center).

After resolution, two recombinant molecules are generated (bottom left).

Figure taken from (72).
Figure 1.6: Overview of the bacteriophage lambda system.

The top illustration shows the attP DNA of lambda. C and C’ (blue) are the core sites, flanking the overlap region.

H sites represent IHF binding sites.

X sites (gold) indicate Xis binding.

The F site (pink) represents Fis binding.

P and P’ sites (green boxes) are lambda Int arm-type sites.

B’ and B (blue) are the attB core type sites, flanking the overlap region.

Arrows facing the same direction indicate direct repeats.

Arrows facing each other (as in CoC’) indicate inverted repeats.

Arrows in the same direction (as in P1, P2, P3) indicate direct repeats.

Filled boxes indicate that the site is used in that reaction (for example, P1 on attP in integration).

Empty boxes indicate that the site is not used in that reaction.

Figure taken from (87).
Figure 1.7: Arm-type sites involved in integration and excision of CTnDOT.

The top two lines represent \textit{attDOT} (solid line) and \textit{attB} (dashed line) before integration. White circles represent the D and D’ core sites. Pale grey circles represent B and B’ core sites.

Black boxes indicate arm-type sites that are required for integration. Grey boxes indicate arm-type sites that have a cooperative role in integration. White boxes are not required.

The bottom two lines represent the \textit{attL} and \textit{attR} of the integrated CTnDOT.

Grey boxes stimulate excision.

White boxes are not required for excision.

Based on findings from (98) and (41).
Figure 1.8: Resolution of Holliday junction intermediates in vitro.

A. Synthetic HJ containing the core sites and a homologous overlap region. No arm-type sites are included. Asterisks indicate 5′ 32P labels; all four oligonucleotides are labeled. Small arrows indicate the site of the first set of strand exchanges. Open arrows indicate the site of the second strand exchange.

B. Resolution of the homologous overlap HJ. Both products (attL, attR) and substrates (attB, attDOT) result.

C. Synthetic HJ containing the core sites and a heterologous overlap region (south arm). No arm-type sites are included. Asterisks and arrows are as in A.

D. Resolution of the heterologous overlap HJ. Only substrates (attB, attDOT) result.

Figure taken from (44).
CHAPTER 2: The *Bacteroides thetaiotaomicron* encoded protein *Bacteroides* Host Factor A participates in integration of the Integrative Conjugative Element CTnDOT into the chromosome

**Abstract**

CTnDOT is a conjugative transposon found in *Bacteroides* species. It encodes multiple antibiotic resistances and is stimulated to transfer by exposure to tetracycline. CTnDOT integration into the host chromosome requires IntDOT and a previously unknown host factor. We have identified a protein, designated BHFa (*Bacteroides* host factor A), that participates in integrative recombination. BHFa is the first host factor identified for a site-specific recombination reaction in the CTnDOT family of integrative and conjugative elements. Based on the amino acid sequence of BHFa, the ability to bind specifically to 4 sites in the attDOT DNA, and its activity in the integration reaction, BHFa is a member of the IHF/HU family of nucleoid-associated proteins. Other DNA bending proteins that bind DNA nonspecifically can substitute for BHFa in the integration reaction.

**Introduction**

*Bacteroides* species are Gram negative, obligate anaerobes that are part of the normal microbiota in the human colon (3). When the gut is punctured, *Bacteroides* can act as opportunistic pathogens that may form abscesses in other regions of the body. Treating the abscesses is complicated by widespread resistance to tetracycline and erythromycin carried by Integrative Conjugative Elements (ICEs, also called conjugative
transposons or CTns) found in *Bacteroides*. These elements are capable of transferring between diverse organisms by conjugation and integrate into the host bacterium’s chromosome. Due to this integration, ICEs are maintained stably even in the absence of selective pressure. As a result, the frequency of antibiotic resistance in *Bacteroides* has increased dramatically over the last 30 years (29). ICEs also encode genes to regulate and carry out their own transfer. Further, some ICEs can mobilize co-resident genetic elements that could not otherwise transfer (28). Because of their benefit to the bacterial host and their ability to transfer among organisms, ICEs are widespread in both Gram positive and negative bacterial populations (29).

CTnDOT is a well characterized ICE found in *Bacteroides* species. It carries the *ermF* and *tetQ* genes that encode resistances to erythromycin and tetracycline respectively. Exposure to tetracycline induces the excision and transfer of CTnDOT. CTnDOT integration and excision require an integrase, IntDOT, and a host encoded protein factor. IntDOT is a tyrosine recombinase and is in the same family of enzymes as λ Int, Flp, XerC, XerD, and Cre (25). IntDOT contains five of the six conserved amino acid residues that form the catalytic sites of tyrosine recombinases (14, 19). These enzymes perform strand exchanges by a site-specific topoisomerase activity. Unlike bacteriophage lambda, IntDOT mediates site-selective integration at one of several sites within the *B. thetaiotaomicron* chromosome. During integration, IntDOT recombines the *attDOT* site in CTnDOT with an *attB* site in the bacterial chromosome to form the *attL* and the *attR* sites of the integrated element. A host factor is also required for integration. During excision from the bacterial chromosome, higher-order nucleoprotein complexes, called intasomes, are formed on the *attL* and *attR* sites. In addition to IntDOT and the
host factor, the CTnDOT-encoded accessory proteins Xis2c, Xis2d, and Exc participate in the excision reaction.

Many transposition and site specific recombination systems require host factors. For example bacteriophage lambda requires Integration Host Factor (IHF) for both integration into and excision from the *E. coli* chromosome. This requirement led to the original identification of IHF (21). In the lambda system, IHF binds to specific sites and bends DNA. We showed previously that *Escherichia coli* IHF can substitute for the *B. thetaiotaomicron* host factor in the CTnDOT integration reaction though there are no appropriately positioned IHF binding sites within *attDOT* (6, 7). Presumably IHF binds CTnDOT DNA nonspecifically and bends the DNA into a favorable conformation for assembly of the intasomes necessary for recombination (7). Based on the ability of IHF to substitute in the CTnDOT integration assays, it was expected that the *Bacteroides* host factor would also introduce bends into DNA after binding. In this paper we have identified and purified a host factor called *Bacteroides Host Factor A* (*bhfA*). BHFa shares several conserved motifs with both *E. coli* HU and IHF, though the primary sequence is not similar to those proteins. This is the first host factor identified for any of the ICEs in the *Bacteroides* spp. BHFa binds specifically to four sites within the *attDOT* site. However, we found that other DNA binding proteins can substitute in the *in vitro* integration assay.
Materials and Methods

Media and antibiotics

*Escherichia coli* strains were grown in Luria-Bertani (LB) medium (Difco). Antibiotics were purchased from Sigma and used at the following concentrations: ampicillin, 100 μg/ml, kanamycin, 50 μg/ml, and gentamycin, 20 μg/ml.

Growth of Bacteroides strains

All *B. thetaiotaomicron* cultures were grown anaerobically to OD$_{650}$ of 0.8 as described previously (13). The cells in the cultures were then pelleted by centrifugation and frozen at -80°. In all, 4 separate pellets (2 L of culture) were combined for protein purification.

Purification of the Bacteroides host factor

The *B. thetaiotaomicron* BT4001 pellets were resuspended in *Bacteroides* suspension buffer (50 mM Tris-Cl [pH 7.4], 10% sucrose). The cells were then lysed by sonication and the extracts were clarified by centrifugation. The resulting crude extract was fractionated by heparin agarose column chromatography using a buffer consisting of 50 mM NaHPO$_4$, 1 mM EDTA, 5% glycerol, and 1 mM DTT [pH 7.5]. We used a NaCl gradient from 50 mM NaCl to 2 M NaCl for elution on a GE AKTA-Purifier Box 900 system. Protein fractions were stored at 4° and tested in the *in vitro* integration assay.
Identification of the *Bacteroides* host factor

The fractions that were active in the *in vitro* integration assay were subjected to electrophoresis on a denaturing Tricine (10-20%) SDS gel. The protein band that appeared in all the active fractions was roughly 10 kDa in size. The fractions remained active in the *in vitro* integration assay after heating at 70° for 15 minutes and centrifugation at 4° for 30 minutes. The 10 kDa band was excised from the gel, digested with trypsin, and analyzed using liquid chromatography mass spectroscopy at the University of Illinois Protein Sciences Facility (U of I Biotechnology Center). The best result (49% sequence coverage) was the gene BT_1499 (NCBI GI: 29345410) from the *B. thetaiotaomicron* VPI-5482 chromosomal sequence, annotated as a putative HU-like protein. The secondary structure was predicted using DSSP (12) and PsiPred (11, 18). We have named this gene *bhfA*, for *Bacteroides Host Factor A*.

In *vitro* recombination assay

The *in vitro* integration assay has been described previously (4, 6, 16). A supercoiled substrate containing *attDOT* and radiolabeled, annealed complementary oligonucleotides containing the *attB* site were incubated with IntDOT protein, along with IHF, BHFa, or other DNA binding proteins (*E. coli* HU and Fis, HMGB1, and NHP6A; provided by Reid Johnson, University of California Los Angeles). All proteins were diluted in IHF dilution buffer (50 mM Tris HCl [pH 8], 10% glycerol, 2 mg/ml BSA, and 200 mM KCl). The mixture was incubated for 20 hours at 37° and the reactions were stopped by adding 5 μl of stop solution (30% glycerol, 10% SDS, 0.25% Xylene cyanol and bromophenol blue). The samples were then subjected to electrophoresis on a 1%
agarose gel for 2 hours at 140 volts, then dried, and analyzed as described previously (16).

**Construction of the BHFa expression plasmid, pKWR33**

The *bhfA* gene was amplified by PCR using primers that introduced NdeI and HindIII restriction sites at the ends of the fragment. The PCR product was subcloned into the pET-30 expression vector between the NdeI and HindIII restriction sites. Primer sequences are listed in Supplemental Materials, Table 1 as Bhu2-F and Bhu2-R. Both the PCR product and the pET-30 vector were digested sequentially with NdeI then HindIII and the products were gel purified and ligated into the vector. The insert was detected by colony PCR and the plasmid was then sequenced to confirm that the BHFa gene contained the correct sequence. Sequencing reactions were performed by the University of Illinois core sequencing facility. The plasmid was named pKWR33 and contains the wild type *bhfA* gene adjacent to an *E. coli* ribosome binding site.

**Protein overexpression and purification**

pKWR33 was transformed into the BL21 DE3 star ΔihfA strain, in LB supplemented with kanamycin. The cultures were grown to an OD$_{600}$ of 0.6 at 37$^\circ$ and production of BHFa was induced by addition of IPTG to a final concentration of 0.5 mM. The cultures were then grown at 30$^\circ$ for 4 hours and cells were pelleted by centrifugation. The resulting pellet was suspended in a 50 mM NaHPO$_4$ [pH 7.2], 1 mM EDTA, 50 mM NaCl, 5% glycerol, 1 mM DTT buffer and the cells were lysed by sonication. The crude extract was clarified by centrifugation before purification.
A GE heparin agarose column was used for the first purification step as described above. Because other proteins co-purified with BHFa, a second column chromatography step was necessary. A GE HiLoad 16/60 Superdex-75 size exclusion column was used to remove the larger proteins. The active fractions from the heparin agarose column purification were dialyzed into low salt (50 mM NaCl) buffer. For the 16/60 Superdex 75 elution, the elution buffer used contained 50 mM NaHPO₄ [pH 7.2], 1 mM EDTA, 150 mM NaCl, 5% glycerol, and 1 mM DTT. The resulting fractions were tested for activity in the in vitro integration assay and dialyzed into storage buffer (50 mM NaHPO₄ [pH 7.2], 1 mM EDTA, 50 mM NaCl, 40% glycerol, and 1 mM DTT. After dialysis, BHFa was about 95% pure and stored at -80°C.

**Determination of molecular weight**

A GE HiLoad 16/60 Superdex-75 size exclusion column was also used to determine whether BHFa is a monomer or a dimer in solution. The GE filtration calibration kit (low molecular weight) was used as the standard according to manufacturer’s instructions. Blue Dextran 2000 and the Low Molecular Weight protein standards were resuspended in 50 mM NaHPO₄ [pH 7.2], 1 mM EDTA, 150 mM NaCl, 5% glycerol, and 1 mM DTT. Purified BHFa eluted at about 68 ml of elution buffer, which is consistent with a size of about 22 kDa and the predicted dimer size of 22.2 kDa.

**Electrophoretic mobility shift assays**

Fragments containing attDOT, attL, attR, the BHFa binding sites (H1, H2, H3-H4), or DNA containing a PCR fragment from pUC19 for the electrophoretic mobility
Shift assays were produced by PCR. The DNA was radiolabeled with \([\gamma^{-32}P]ATP\) using polynucleotide kinase (Fermentas). Primers for these fragments are listed in Table 2.1.

Reactions included BHFa, the DNA fragment (2.6 nM for \(attDOT\) and pUC19; 0.6 nM for \(attL\) and \(attR\); 2.3 nM for H1, H2, and H3-H4 containing sites), and gel shift binding buffer (GSBA 75), consisting of 50 mM Tris-HCl [pH 8], 1 mM EDTA, 50 mM NaCl, 10% glycerol, and 0.075 μg/μL herring sperm DNA. Binding reactions and electrophoresis were performed as described previously (38). Gels were then dried, exposed to a phosphoimager screen, and scanned in the same way as in the \textit{in vitro} integration assay.

**Footprinting**

Footprinting reactions were performed as described previously (39). Footprinting primers (\(attDOT\) top and bottom strands) were purchased from Integrated DNA Technologies with 6-carboxyfluorescein (6-FAM™) at the 5′ end. DNA fragments were produced by PCR so that each fragment contained only one 6-FAM labeled strand and gel-purified. One pmole of DNA was used in each footprinting reaction.

Deoxyribonuclease I (Worthington Biochemical corporation) was suspended in 5 ml of 10 mM Tris-HCl [pH 7.5], 5 mM MgCl\(_2\) and 50% glycerol to a concentration of 1 mg/ml and diluted for subsequent digestions in DNase I buffer (2.5 mM MgCl\(_2\), 0.5 mM CaCl\(_2\), 10 mM Tris HCl [pH 7.6], and 0.1 mg/ml BSA).

In addition to the \(attDOT\) DNA and BHFa, the digestion mix consisted of 3 mM CaCl\(_2\), 7 mM MgCl\(_2\), 2 μL IHF dilution buffer, 1 pmol of 6-FAM labeled DNA, 9.5%
glycerol, 50 mM TrisHCl [pH 7.4], and 25 μg/ml BSA. The total reaction volume was 20 μL. Incubations were carried out at room temperature for 30 minutes. DNase I was then added (at a final concentration of 0.1 μg/μl for top strand digestions and 0.2 μg/μl for bottom strand digestions) for 2 minutes and the mixture was quenched by addition of 1 volume of 0.5 M EDTA. The reactions were purified using the Qiagen PCR purification kit. Each digest containing BHFa was tested in duplicate and compared against a control reaction lacking BHFa to identify protected bases. After digestion and purification, samples were submitted to the University of Illinois core sequencing facility and analyzed using the Applied Biosciences Genemapper program (version 3.7).

To align regions of protection by BHFa to specific bases in the DNA sequence, the USB Thermo Sequenase kit was used according to the manufacturer’s instructions using the same primers used to generate the attDOT DNA for footprinting. Reactions were likewise submitted to the University of Illinois core sequencing facility and analyzed using the Genemapper program.

**Results**

**Identification of Bacteroides host factor**

Previous studies showed that *E. coli* IHF can substitute for the *B. thetaiotaomicron* host factor in *in vitro* integration reactions (4, 6). IHF binding to the attDOT site appears to be nonspecific due to the high concentrations of IHF required for the *in vitro* recombination reaction to proceed. We proposed that IHF binds non-
specifically to \textit{attDOT} and bends the DNA to promote intasome formation so IntDOT can carry out integration. \textit{Bacteroides} species do not encode a closely-related IHF homolog although there are several putative DNA-binding proteins that are annotated as related to HU based on sequence similarity within the \textit{B. thetaiotaomicron} genome (40).

Column chromatography was used to fractionate \textit{B. thetaiotaomicron} DNA-binding proteins. The DNA-binding proteins were purified from crude extracts from \textit{B. thetaiotaomicron} strain 4001, which lacks any known ICEs. The cell pellets from two liters of anaerobically grown \textit{B. thetaiotaomicron} cultures were utilized to make an extract that was subjected to heparin agarose column chromatography as described in Materials and Methods.

The resulting fractions were tested in the \textit{in vitro} integration assay as described in the Materials and Methods. The integration assay uses a supercoiled 3.5 kb plasmid containing the \textit{attDOT} site and a linear $^{32}$P labeled 67 bp \textit{attB} site. Integration produces a 3567 bp linear product containing the radioactively labeled DNA, which can be visualized on a 1\% agarose gel (Figure 2.1A). In this assay, a small amount of recombination was observed in reactions containing only IntDOT and IHF dilution buffer. This background activity is due to contaminating IHF present in the IntDOT preparation.

By this method, we identified several active fractions (which eluted at approximately 1M NaCl) that stimulated activity in the \textit{in vitro} integration assay (Fractions 41 and 43, Figure 2.1A). The active fractions were subjected to SDS gel electrophoresis as described in Materials and Methods. One band, about 10 kDa in size,
appeared to correlate with the integration activity (Fractions 41-43, marked with open arrow, Figure 2.1B). In addition, the active fractions were subjected to heat treatment and the supernatant fractions were assayed for \textit{in vitro} integration activity. The same 10 kDa band remained stable in the active fraction while the other bands from the same fractions disappeared due to the heat treatment (data not shown). The 10 kDa protein was subjected to liquid chromatography mass spectroscopy. Based on the primary sequence of the protein, it was identified as BT_1499 (NCBI accession: NC_004663), which is annotated as a DNA-binding protein HU. Due to its role in CTnDOT integration, the gene BT_1499 has been designated \textit{bhfA}, for \textit{Bacteroides} \textit{Host} \textit{Factor A}.

The \textit{bhfA} gene was amplified from the \textit{B. thetaiotaomicron} genome and cloned into pET-30 as described in Materials and Methods. BHFa was then overexpressed in an \textit{E. coli} strain lacking the \textit{ihfA} gene and was purified to 95% purity by using heparin agarose column chromatography followed by SuperDex-75 size exclusion chromatography (marked with closed arrow, Figure 2.1C). The \textit{in vitro} integration assay was used to verify that the purified protein remained active after each step. After purification, the protein was resubmitted for liquid chromatography mass spectroscopy analysis to confirm that it was the cloned BHFa and not an \textit{E. coli} protein. Based on elution from a SuperDex-75 size-exclusion chromatography column, BHFa has a molecular weight of about 22 kDa (data not shown). This size corresponds to a homodimer, similar to HU proteins in most species.
BHFa binds specifically to the attDOT DNA sequence

Presumably, BHFa participates in the integration reaction by binding and bending DNA and, in conjunction with IntDOT, facilitating the formation of the integrative intasome which catalyzes the recombination reaction. In order to demonstrate that BHFa binds DNA, electrophoretic mobility shift assays (EMSAs) were used. The target DNA was a 324 bp fragment containing the attDOT site (Figure 2.2A).

Incubation of attDOT DNA with BHFa forms one complex at a concentration of 600 nM BHFa. Further dilution of BHFa led to formation of 2 distinct complexes of similar mobility as the BHFa concentration decreased to 300 nM (Figure 2.2A). As the BHFa is diluted to 5 nM, four shifts appear, suggesting that multiple binding sites are present. The single shift at 600 nM likely reflects all binding sites being occupied by BHFa. The other shifts may be complexes where BHFa dimers have bound to only some of the binding sites in the DNA. Each BHFa dimer bound to attDOT presumably bends the DNA around itself. The mobility of the resulting complex would depend on the relative location of the binding site within the DNA fragment (1, 20, 37). Therefore, occupancy of one site, multiple sites, or all available sites may introduce a different degree of DNA bending and cause the formation of complexes with differing mobilities.

BHFa was further tested with an unrelated DNA sequence from the plasmid pUC19. BHFa bound the nonspecific DNA sequences at a concentration of 500 nM but failed to form complexes at concentrations lower than 50-100 nM (Figure 2.2B). A second complex is faintly visible at a concentration of 100 nM of BHFa. The complex observed at high concentrations of BHFa (600 nM in Figure 2.2A, 500 nM in Figure 2.2B)
2.2B) is similar between both the* attDOT* and pUC19 DNA. However, as BHFa was diluted, the initial complex and faint secondary complex with pUC19 DNA disappeared. Therefore, we conclude that the complex formed with pUC19 at 500 nM of BHFa is nonspecific and that BHFa binding to *attDOT* DNA is specific (Figure 2.2A and B).

When BHFa was incubated with *attR* DNA, only one shift is observed across the range of concentrations from 500 nM to 5 nM which suggests that there is only one binding site present (Figure 2.2C). Over the same range, BHFa binding to *attL* shows a pattern of two distinct shifts similar to the full *attDOT* fragment, suggesting that the *attL* site contains multiple BHFa binding sites (Figure 2.2D).

**Complexes with BHFa and IntDOT**

Binding of both IntDOT and BHFa to *attDOT* DNA is necessary in order to form the intasomes required for recombination. In order to detect and characterize these complexes, we performed EMSAs using *attDOT* DNA incubated with IntDOT alone, BHFa alone, or both BHFa and IntDOT (Figure 2.2E). BHFa was used at a constant concentration of 5 nM, where it formed two complexes (Figure 2.2E, Lane 2). At a concentration of 175 nM, IntDOT alone shifted a small amount of *attDOT* DNA (marked with filled arrow) while some DNA remained in the wells (Figure 2.2E, lane 9). When the concentration of IntDOT was varied (110 to 430 nM) and the concentration of BHFa remained constant (5 nM), two super shifts were observed (Figure 2.2E, lanes 3-8). The higher band was stronger and remained as IntDOT is diluted (marked with open arrow, Figure 2.2E). The lower band was fainter and was rapidly lost with IntDOT dilution. The presence of two supershifts may reflect IntDOT interacting with alternate arm-type
sites, made possible by BHFa bending the \textit{attDOT} DNA. Further dilution of IntDOT caused the lower supershift to disappear while the two BHFa associated shifts reappear (Figure 2.2E, lanes 4-8). The super shifts formed with both proteins are distinct from the complexes formed with either IntDOT or BHFa protein on their own (Figure 2.2E, lanes 2, 6, and 9). This result also suggests that either IntDOT is interacting with different arm-type sites due to BHFa binding or that multiple IntDOT monomers may be bound to the same bent \textit{attDOT} DNA.

\textbf{Identification of the binding sites of BHFa}

Since the binding of BHFa appears to be specific, we wanted to identify the binding sites. The \textit{attDOT} site was labeled with fluorescent 6-FAM and PCR amplified as described in Materials and Methods. The sequence of the \textit{attDOT} site used for these protection studies is shown in Figure 2.3A, along with the core and arm-type sites which are bound by IntDOT. The resulting chromatograms show a pattern of protection over three regions of the DNA on the bottom strand, designated H1, H2, and H3-H4 (Figure 2.3B). The corresponding regions are also protected on the top strand, though the protection of H3-H4 is more difficult to observe since DNase I does not cut as effectively in that area (data not shown).

The H1 site includes bases -67 to -46 on the top strand and -59 to -44 on the bottom strand (Figure 2.3A). It also overlaps with the entire IntDOT R2 arm-type site and about half of the R2′ arm-type site. The H2 site includes bases +23 to +42 on the top strand and +30 to +45 on the bottom strand. The H2 site is located downstream of the
core region and D’core-type site but upstream of the L1 arm-type site and does not appear to overlap any other known protein binding sites within *attDOT*.

The H3 and H4 sites are the largest region of protection, encompassing bases +53 to +118 on the bottom strand. Due to its size, we believe that there are two BHFa binding sites immediately adjacent to one another. Enhanced cleavages were observed between bases +63 and +66. Enhanced cleavage was also observed between bases +90 and +92, which may define the border between the two binding sites. The H3 and H4 sites show some protection on the top strand, but the footprint is much less distinct than on the bottom strand (data not shown). The H3 site overlaps with the entire L1 arm-type site and the H4 site overlaps with seven of the ten bases of the L2 arm-type site.

At 545 nM of BHFa, the entire *attDOT* region shows slightly decreased digestion compared to the no protein control, which may be due to nonspecific binding of BHFa. As BHFa is serially diluted to 180 nM, the H1 and H3-H4 sites lose some protection (Figure 2.3B). The H2 site is the last to lose protection (at concentrations lower than 110 nM) and it appears that BHFa has the highest affinity for this site. In addition, we observed enhanced DNA cleavage outside of the protected regions on *attDOT* (Figure 2.3B; specific sites marked with asterisks). This enhanced cleavage may be due to BHFa binding causing the *attDOT* DNA to bend in a way that makes the peripheral regions more accessible to DNase I because the enhanced cleavages occur when the H1 site is occupied by BHFa.
EMSA Analysis of BHFa binding to the H1, H2, and H3-H4 sites

After identifying the three regions of protection on attDOT by DNase I footprinting, we used EMSA as an independent method to demonstrate BHFa binding. EMSAs also allowed the comparison of relative binding affinities of BHFa to the sites.

A PCR product containing the H1 site extending from position -101 to -6 was incubated with BHFa. Only one shift was observed, indicating that there is one binding site present (Figure 2.4A and 2.4B). The PCR-generated DNA fragment containing the H2 site extends from position -26 to +75 (Figure 2.4C). Incubation with BHFa also shows a single shift, indicating that there is one BHFa site on the fragment (Figure 2.4D). In agreement with the footprinting results, BHFa shows a higher affinity for the H2 site than either H1 or H3-H4 sites (Figure 2.4B, 2.4D, and 2.4F).

The PCR product containing the H3 and H4 sites includes position 57 to 145 (Figure 2.4E). When EMSAs are performed with DNA containing this region, two distinct shifts were observed (Figure 2.4F). The highest shift, visible from 2160 to 270 nM of BHFa, likely corresponds to both sites being occupied at once. The lower shift likely corresponds to only one of the two sites being occupied. This pattern is consistent with the footprint results suggesting that there are two binding sites separated by a small region of enhanced cleavage (Figure 2.3B).

Comparison of the protected binding sites did not reveal an obvious consensus binding site. It may be that, like IHF, the binding site of BHFa is relatively degenerate and more independent sites will need to be examined to determine the consensus binding sequence (8). Based on complexes of IHF and HU proteins with DNA, we would expect
that relatively few bases of each footprint (probably less than 10) are contacted directly by residues of BHFa, while the rest of the DNA is bent around the protein (26, 32, 34-36).

**DNA bending proteins promote the in vitro integration reaction**

Previous studies had showed that *E. coli* IHF could substitute for the *B. thetaiotaomicron* host factor in the *in vitro* integration assay. However, there are no appropriately positioned IHF binding sites within the *attDOT* sequence, suggesting that nonspecific binding of IHF is sufficient for integration. Accordingly, we tested other DNA bending proteins in the *in vitro* integration assay. We found that entirely unrelated DNA bending proteins can substitute for BHFa in the *in vitro* integration assay (Figure 2.5). *E. coli* HU is a heterodimeric nucleoid associated protein that is closely related to IHF but binds nonspecifically (37). *E. coli* HU can substitute in the *in vitro* integration assay (Figure 2.5, lanes 3-5). *E. coli* FIS is also a nucleoid-associated protein but binds DNA more specifically than HU and plays a role in excision of bacteriophage λ in *E. coli*. In *Salmonella typhimurium*, FIS binds to DNA and interacts with the Hin recombinase in the Hin inversion system (22). FIS cannot substitute for BHFa or IHF in the integration assay (Figure 2.6, lanes 6-8). HMGB1 is a eukaryotic protein (from the High Mobility Group family of proteins) that binds DNA non-specifically and bends DNA. NHP6A is a related protein but was originally found in *S. cerevisiae* (15). Though these are both eukaryotic proteins and have very different structures from bacterial IHF and HU proteins, they can enable integration in the *in vitro* integration reaction (Figure 2.5, lanes 9-11 and 12-14). These proteins have also been shown to substitute for HU in other recombination systems (24).
Discussion

In this study, we identified the protein BHFa, a *Bacteroides* host factor required for integration and likely excision of CTnDOT. The purification of BHFa was carried out using a crude extract and column chromatography. Identification of BHFa was based on activity of the purified fractions in the *in vitro* integration assay. However, it is possible that other proteins encoded by *B. thetaiotaomicron* might also enable integration or excision of CTnDOT. These proteins may have been excluded from our initial analysis if they are less stable than BHFa or if their expression is associated with a specific growth phase, as with *E. coli* FIS, IHF, and HU proteins (10).

Along with other small, basic proteins (such as FIS) both IHF and HU are considered nucleoid associated proteins and may be found at very high concentrations in the cell. While these proteins are often involved in the recombination of temperate bacteriophages and mobile genetic elements, they also play important roles in the host cell. Nucleoid associated proteins are necessary for chromosome partitioning, transcription, regulation, DNA replication, and protection or repair of DNA (2, 5, 9, 10, 27). The exact role of BHFa in chromosome maintenance and other processes in the *B. thetaiotaomicron* host is not yet known. Presumably, it is involved in at least some of the same processes as the other nucleoid associated proteins discussed above.

At the primary sequence level, BHFa diverges extensively from these other nucleoid associated proteins (Figure 2.6A) (23, 30, 31). BHFa is 34% identical to the IHF alpha subunit and 32% identical to the IHF beta. Yet the IHF alpha and beta subunits are 33% identical to one another, so BHFa is as identical to each as the subunits
are to one another. In addition, the HU subunits show a similar identity to BHFa at about 30% and 32%, respectively subunit (Percent identities are shown in Table 2.2).

The secondary structure is much more conserved for these proteins and the predicted secondary structure of BHFa is similar to the *E. coli* proteins IHF and HU. The actual secondary structure for HU and IHF is shown below the protein sequences and is conserved despite primary sequence variations between both IHF and HU and their subunits (Figure 2.6A). BHFa is similar to many HU proteins since it is a homodimer, but BHFa binds specific DNA sequences while HU proteins bind nonspecifically. While both BHFa and IHF bind to specific DNA sequences, BHFa has several unusual amino acid residues within a very highly conserved region among HU and IHF family proteins. The region includes a proline residue within the flexible arm region that contacts DNA and introduces a kink into the backbone (position 65 in the Figure 2.6A alignment, marked with an open arrow) (26, 34, 35). The crystal structures of several HU proteins and IHF have demonstrated that a proline at this position is found in virtually all proteins within the IHF/HU superfamily (33). Interestingly, an *E. coli* mutant IHF that lacks the conserved proline still binds DNA but shows decreased DNA binding specificity (17). In BHFa, there is an isoleucine at this position instead. Presumably, this difference would contribute to the specific binding of BHFa and could possibly influence how the protein bends DNA.

Though BHFa has several differences from well-studied IHF and HU proteins, these variations appear consistent among other predicted DNA-binding proteins found in related *Bacteroides* species (Figure 2.6B) (30, 31). Specifically, the different amino
acids within the conserved arm region found in other IHF-HU superfamily proteins are consistent among the *Bacteroides* spp (Figure 2.6B, residues 60-63). Though not as strictly conserved as the position 63 proline (position 65 in Figure 2.6A), the preceding residues 60-62 (GRN) are also very well conserved among the superfamily. However, the *Bacteroides* DNA-binding proteins all include the residues ARN instead (marked with black arrows, figure 2.6B). Further, the remainder of the other *Bacteroides* spp. DNA binding proteins’ primary sequence is very highly conserved, with only 6 positions showing variation. It is likely that these related proteins would enable the recombination of CTnDOT and related elements into the host chromosomes of species related to *B. thetaiotaomicron*. This is supported by previous work showing that CTnDOT is capable of transfer and integration into a variety of *Bacteroides* species and close relatives such as *Prevotella* (29).

Occupancy of the different binding sites of BHFa found in the *attDOT* region depends on the concentration of BHFa. Since three of the four BHFa binding sites are adjacent or overlap with arm-type sites, this may lead to differential availability of the arm-type sites for IntDOT binding. The H2 site is the only BHFa binding site that does not overlap with any other known protein-binding sites within the *attDOT* sequence. The lower affinity sites (H1, H3, and H4) overlap with portions of the arm-type sites, some of which are required for CTnDOT integration into and excision from the host chromosome. The H1 site overlaps with the R2 and R2’ arm-type sites, while the H3 site overlaps with the L1 arm-type site and H4 site partially overlaps with the L2 arm-type site. This overlap between the lower affinity binding sites of BHFa and the arm-type sites of IntDOT suggests that non-specific binding of *attDOT* could be important for integration.
This argument is consistent with the observation that several DNA binding and bending proteins that are unrelated to BHFa can substitute for BHFa in the integration reaction. For example, once CTnDOT enters a new host, and is converted to the double stranded circular form, then a DNA binding and bending protein could help form the integrative intasome. If IntDOT is expressed in the recipient, the CTnDOT element could recombine with an \textit{attB} site in the recipient chromosome where it would be stably maintained.

In summary, BHFa is the first example of a host factor that promotes integration of CTnDOT into a host cell chromosome. This protein may play a part in the transfer of other genetic elements, such as mobilizable transposons like NBU1 and NBU2. While BHFa is a member of the IHF/HU superfamily, it also differs from the well-studied members of that family. It is a homodimer that binds specifically to 4 sites found in the \textit{attDOT} region and varies at several conserved residues. However, specific binding to these sites does not appear to be required for CTnDOT integration in the \textit{in vitro} integration assay. Unrelated proteins enable the integration reaction \textit{in vitro} and may enable integration of CTnDOT (or related elements) into a recipient cell’s chromosome as long as IntDOT has been expressed. Besides the presence of a DNA bending protein, the only other requirement for integration and long term maintenance would be a suitable \textit{attB} site within the recipient chromosome. Overall, the identification of this host factor illustrates a new member of the IHF/HU superfamily of proteins and provides new insights into a reaction that spreads antibiotic resistance carried by conjugative elements among the medically relevant \textit{Bacteroides spp.}
Acknowledgments

We thank Abigail Salyers, Margaret Wood, Carolyn Keeton, Sumiko Yoneji, and Jennifer Laprise for suggestions and helpful comments. We would also like to thank Reid Johnson for generously providing samples of the proteins HU, HMGB1, Fis, and NHP6A.

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


Figure 2.1: Identification and purification of BHFα.

A. Activity of purified *B. thetaiotaomicron* protein fractions in the *in vitro* integration assay. Percent integration for individual reactions is shown above each lane and was calculated as described in Materials and Methods.

Reactions were supplemented with IHF dilution buffer (B), *B. thetaiotaomicron* crude extract (CE), and various fractions from the heparin agarose column (F). Fractions are numbered by order of elution.

IHF is included as a positive control.
Figure 2.1 (continued)

B. Silver stained SDS-PAGE gel showing \textit{B. thetaiotaomicron} protein fractions. Molecular weight (MW) standards are shown on the left. Other lanes contain crude extract (CE1 and CE2), flowthrough (FT) from the heparin agarose column, and fractions (F) from the heparin agarose column. The active fractions (F41 and F43 in 1A) contain the heat-stable protein that enables integration in the \textit{in vitro} integration assay, indicated with an open arrow. The 10 kDa protein from F43 was analyzed by liquid chromatography mass spectroscopy.

C. BHFa was purified to homogeneity as shown on the SDS-PAGE gel. The BHFa monomer is 10.1 kDa in size (arrow).
Figure 2.2: DNA binding by BHFa.

A. Specific binding of BHFa on the 325 bp attDOT fragment. Each reaction contains 2.6 nM of DNA substrate.

B. Non-specific binding of BHFa on a 330 bp fragment of pUC19. Each reaction contains 2.6 nM of DNA substrate.

C. Binding of BHFa on attR. Each reaction contains 0.6 nM of DNA.

D. Binding of BHFa on attL. Each reaction contains 0.6 nM of DNA.

E. Binding of BHFa and IntDOT to attDOT DNA. A 325 bp attDOT fragment was incubated with 5 nM BHFa (lane 5) or 175 nM of IntDOT (lane 9). In lanes 3-8 the concentration of BHFa is 5 nM and the concentrations of IntDOT are varied.
Figure 2.3: DNA base pairs of the attDOT region protected by BHFa.

A. The sequence of attDOT and regions bound by BHFa. Bases are numbered relative to the overlap region of attDOT; the central base is 0. The overlap region (GCTTAGT) includes bases -3 to +3. Bases that are upstream of the central bases are numbered as negative and downstream numbers are positive.

The core-type sites (D and D') contacted by the core-binding and catalytic domains of IntDOT are immediately adjacent to the overlap and are marked with dashed boxes. The arm-type sites (R1', R1, R2, R2', L1, and L2) which are contacted by the N-terminal domain of IntDOT are marked with solid boxes. Bases marked in bold were protected from DNase I digestion in the presence of BHFa. H sites (H1, H2, and H3-H4) are marked with the labeled boxes.
Figure 2.3 (continued)

B. Chromatograms of DNase I protection of the bottom strand of \textit{attDOT}.

Concentrations of BHFa are indicated to the left of each panel. The top panel shows a reaction that contained no BHFa.

Regions of protection are marked with black rectangles.

Grey rectangles indicate intermediate protection and empty rectangles indicate a loss of protection.

Sites of enhanced DNA cleavages are marked with asterisks.
Figure 2.4: Binding of BHFa to H site DNA fragments.

A. The 96 bp DNA sequence used for the EMSA. The R2 and R2' arm-type sites are marked with solid boxes. Bases of DNA containing the H1 site protected by BHFa are in bold and grey.
Figure 2.4 (continued)

B. EMSA of DNA containing the H1 site. DNA concentration was 2.3 nM in each reaction.

C. The 102 bp DNA sequence used for the EMSA of DNA containing the H2 site.
The overlap region is marked with solid lines.
The D and D’ core-type sites are marked with dashed lines.
The L1 arm-type site is marked with a solid lines.
Protected bases are in bold and grey.

D. EMSA of the DNA containing the H2 site. DNA concentration was 2.3 nM in each reaction.

E. The 103 bp DNA sequence containing the H3 and H4 sites used for the EMSA.
The L1 and L2 arm-type sites are marked with solid boxes.
Protected bases are in bold and grey.

F. EMSA of the DNA containing the H3 and H4 sites on the same fragment of DNA.
DNA concentration was 2.3 nM in each reaction.
Figure 2.5: Non-\textit{Bacteroides} proteins can substitute for BHFα in the \textit{in vitro} integration assay.

This figure shows the \textit{in vitro} integration assay using proteins from non-\textit{Bacteroides} organisms. Lane 1 contains IHF dilution buffer. Lane 2 contains \textit{E. coli} IHF.

Lanes 3-5 contain \textit{E. coli} HU protein.

Lanes 6-8 contain \textit{E. coli} Fis protein, which did not enable integration at the concentrations tested.

Lanes 9-11 contain bovine HMG-1 protein.

Lanes 12-14 contain \textit{S. cerevisiae} NHP6A.

All of the proteins except Fis enable some integration in the assay.
Figure 2.6: Alignments of BHFa with *E. coli* IHF, HU subunits and other *Bacteroides* proteins.

A. Alignment of BHFa with *E. coli* IHF and HU proteins (alpha and beta subunits). The secondary structure of IHF and HU proteins is shown below the sequence; indicating the three alpha helices, the beta strands, and the flexible arm region that contacts DNA. The open arrow indicates a conserved proline residue that interacts directly with the DNA.

B. Alignment of *Bacteroides* putative DNA binding proteins. Though BHFa differs substantially at the primary sequence level from *E. coli* IHF and HU proteins, other *Bacteroides* DNA-binding proteins are highly conserved among various species. Black arrows represent variations from the consensus for the arm region of IHF/HU proteins that are consistent amongst *Bacteroides* spp. DNA binding proteins.

Identical proteins from different species have been combined into single entries; accession numbers follow:

1. MULTISPECIES: DNA-binding protein [*Bacteroides*]
   Accession: WP_008762277.1
   GI: 496037770
   >gi|496037770|ref|WP_008762277.1| MULTISPECIES: DNA-binding protein [*Bacteroides*]
   MTKADIVNEITKTKGIDKTTVTVTTVEAFMEAEVGDSNENYVRGFSFVKKRAQKTARNISK NTTII
   IPEHNNIPAFKPAKFTISVKK

2. MULTISPECIES: DNA-binding protein [*Bacteroides*]
   Accession: WP_005677943.1
   GI: 491931000
   >gi|491931000|ref|WP_005677943.1| MULTISPECIES: DNA-binding protein [*Bacteroides*]
   MTKADIVNEITKTKGIDKTTVTVTTVEAFMDAVKDSNENYVRGFSFVKKRAQKTARNISK NTTII
   IPEHNNIPAFKPAKFTISVKK
3. MULTISPECIES: DNA-binding protein [Bacteroides]
Accession: WP_004300646.1
GI: 490428497
>gi|490428497|ref|WP_004300646.1| MULTISPECIES: DNA-binding protein [Bacteroides]
MTKADIVNEITKKTGIKDQTVLTTEAFMDAVKDSLNDENVYLRGFGSFVVKKRAQKTARNISK
NTTII
IPEHNIPAFPKAKFTTISVKK

4. DNA-binding protein [Bacteroides faecichinllae]
Accession: WP_025074398.1
GI: 640647623
>gi|640647623|ref|WP_025074398.1| DNA-binding protein [Bacteroides faecichinllae]
MTKADIVNEITKKTGIKDQTVLTTEAFMDAVKDSLNDENVYLRGFGSFVVKKRAQKTARNISK
NTTII
IPEHNIPAFPKAKFTTISVKK

5. DNA-binding protein [Bacteroides reticulotermitis JCM 10512]
Accession: GAE83694.1
GI: 576503644
>gi|576503644|dbj|GAE83694.1| DNA-binding protein [Bacteroides reticulotermitis JCM 10512]
MTKADIVNEITKKTGIKDQTVLTTEAFMDAVKDSLANNENVYLRGFGSFIVKKRAQKTARNISK
NTTII
IPEHNIPAFPKAKFTTISVKK
### Table 2.1: Primer sequences used in cloning, EMSAs, and footprinting.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bhu2-F</strong></td>
<td>GGA GAT ATA CAT ATG ACT AAA GCA GAT ATT GTA AAC GAG ATT ACA AAG AAA ACC GG</td>
<td></td>
</tr>
<tr>
<td><strong>Bhu2-R</strong></td>
<td>CCC GCT AAG CTT TTA TTT CTT TAC TGA AAT TGT AAA TGT CTT AGC CGG TTT GAA GG</td>
<td></td>
</tr>
<tr>
<td><strong>GC+3-B-TOP</strong></td>
<td>GCT GCC ATG ATA TAA TTA CTG TTT AGT ATT TTA ATT GCG CAA ATT TAC TGC AAA TTT CCG AGC AAC G</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>#3-B-TOP</strong></td>
<td>CGT TGC TCG GAA ATT TGC AGT AAA TTT GCG CAA TTA AAA TAC TAA ACA GTA ATT ATA TCA TGG CAG C</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>DRJ/MM160F</strong></td>
<td>TCG GGC ATG GTT ACG AAC</td>
<td>Margaret Wood, unpublished</td>
</tr>
<tr>
<td><strong>DLJ/AG161R</strong></td>
<td>CTC GAA TTA AAT AGC TAC TTT TGC</td>
<td>Margaret Wood, unpublished</td>
</tr>
<tr>
<td><strong>FP-T DRJ/MM160-FL</strong></td>
<td>/56-FAM//iMe-isodC/TC GGG CAT GGT TAC GAA CAA GTA ACG TTG TGG</td>
<td></td>
</tr>
<tr>
<td><strong>FP-B DLJ/AG161-R</strong></td>
<td>/56-FAM//iMe-isodC/CT CGA ATT AAA TAG CTA CTT TGG C</td>
<td></td>
</tr>
<tr>
<td><strong>FP-B DRJ/MM160-FL</strong></td>
<td>TCG GGC ATG GTT ACG AAC AAG TAA CGT TGT GG</td>
<td></td>
</tr>
<tr>
<td><strong>DLJ/106F</strong></td>
<td>GTA AAT TTG CGC AAT GAA GTT ACG</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>attL GS-L</strong></td>
<td>GCA AGC GTT GCT CGG AAA TTT GCA</td>
<td>Margaret Wood, unpublished</td>
</tr>
<tr>
<td><strong>DLJ/INW-1 /R127R-GS-L</strong></td>
<td>GCT ACT TTT GCA TAC GAA GAG TTC TTT GAA GGT TAC G</td>
<td>Margaret Wood, unpublished</td>
</tr>
<tr>
<td><strong>DRJ/2255F</strong></td>
<td>GCA TTA TCG GGC ATG GTT A</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>DRJ/S433R</strong></td>
<td>TAG TAT TTT AGA ATT GCA AAG TTA CTA CAA</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>H1-R</strong></td>
<td>AGT TAC TAC AAA AAA GTG AAA TGC</td>
<td></td>
</tr>
<tr>
<td><strong>H2-F</strong></td>
<td>TTT CAC TTT TTT GTA GTA ACT TTG C</td>
<td></td>
</tr>
<tr>
<td><strong>H2-R</strong></td>
<td>TCG GTA ACC TAT TAC TAT CAA GAG CG</td>
<td></td>
</tr>
<tr>
<td><strong>H3-F</strong></td>
<td>TTT CAC TTT TTT GTA GTA ACT TTG C</td>
<td></td>
</tr>
<tr>
<td><strong>H3-R</strong></td>
<td>TTT GCA TAC GAA GAG TTC TTT GAA GG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BHFa</td>
<td>IHF-α</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>BHFa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHF-α</td>
<td>34.4%</td>
<td></td>
</tr>
<tr>
<td>IHF-β</td>
<td>31.5%</td>
<td>32.6%</td>
</tr>
<tr>
<td>HU-α</td>
<td>30.0%</td>
<td>36.7%</td>
</tr>
<tr>
<td>HU-β</td>
<td>32.1%</td>
<td>32.6%</td>
</tr>
</tbody>
</table>

Table 2.2: Percent identity among *E. coli* subunits of IHF and HU and BHFa.

Percent identities were compared using lalign.

(http://www.ch.embnet.org/software/LALIGN_form.html)
CHAPTER 3: Importance of host factor protein BHFa and arm-type sites in resolution of Holliday Junction intermediates containing heterology by the CTnDOT integrase IntDOT

Abstract

CTnDOT is an integrated conjugative element found in Bacteroides species. CTnDOT encodes and transfers antibiotic resistance genes. The element integrates into and excises from the host chromosome via a Holliday junction intermediate as part of a site-specific recombination mechanism. The CTnDOT integrase, IntDOT, is a tyrosine recombinase with core-binding, catalytic and amino terminal domains. Unlike well-studied tyrosine recombinases such as lambda Int, IntDOT is able to resolve Holliday junctions containing heterology (mismatched bases) between the sites of strand exchange.

However, previous work showed that IntDOT was unable to resolve synthetic Holliday junctions containing mismatched bases to products in the absence of the arm-type sites and a DNA bending protein. We constructed synthetic HJs with the arm-type sites and tested them with the Bacteroides host factor. We found that the L1 site is required for directionality of the reaction, particularly when the HJ contains mismatches. The Bacteroides host factor is required for efficient resolution to products when the overlap region contains mismatches and stimulates resolution to products when the overlap region is identical.
**Introduction**

CTnDOT is an integrated conjugative element (ICE, formerly known as conjugative transposons) found within *Bacteroides* species (8, 10). CTnDOT integrates into the *Bacteroides* host chromosome and carries genes for resistance to tetracycline and erythromycin. The element is induced to transfer via conjugation by tetracycline (11). Since *Bacteroides* spp. are major components (20-30%) of the gut microbiota and can act as opportunistic pathogens, these species may cause abscesses or anaerobic bacteremia if the gut is punctured (1, 7, 36). CTnDOT and similar elements have contributed to the increase in antibiotic resistance among *Bacteroides* spp. in the last forty years (32).

CTnDOT integrates into and excises from the host chromosome by a site-selective recombination mechanism. The recombination reaction of CTnDOT is catalyzed by the integrase, IntDOT, a member of the tyrosine recombinase family. Integration into the chromosome requires the joined ends of the element (*attDOT*), IntDOT, a host-encoded DNA-bending protein, and a suitable *attB* site within the chromosome (10, 20, 21). Excision from the chromosome is tightly regulated and involves the additional excisionase proteins Xis2c and Xis2d. A topoisomerase, Exc, stimulates the excision reaction (11, 12, 15, 23, 33).

One example of a DNA-bending protein that can assist with CTnDOT integration is *Bacteroides* Host Factor A (BHFa), a host-encoded, nucleoid-associated protein that enables integration in the CTnDOT *in vitro* integration reaction. BHFa has four binding sites within the *attDOT* region (28). More on the discovery and characterization of BHFa can be found in Chapter 2 of this work.
In both integration and excision, the necessary proteins and DNA sites form a higher-order nucleoprotein complex called the intasome. While the integrative and excisive intasomes are composed of different participants and DNA sites, both proceed through a series of ordered strand exchanges by IntDOT monomers that generate a Holliday junction (HJ) intermediate. Two IntDOT monomers make the initial cleavages at the core-type sites, D and D’, located seven base pairs apart (Figure 3.1) (12, 20). The region between the core-type sites is called the overlap region.

Two IntDOT monomers each create a covalent 3’ phosphotyrosyl intermediate, leaving a 5’ hydroxyl to attack the partner strand. Once formed, the HJ intermediate isomerizes by a strand-swapping mechanism through the overlap region until a second set of strand exchanges by an additional two IntDOT monomers resolve the junction (34). Unlike other tyrosine recombinases, IntDOT can resolve HJ intermediates that contain mismatched bases in the overlap region \textit{in vivo} (Figure 3.2). Lambda Int, in contrast, is unable to resolve HJ intermediates containing a single mismatch (2, 35).

Like lambda Int, IntDOT is a heterobivalent protein with three domains. The core binding (CB) domain of IntDOT interacts with the core-type sites: D and D’ on \textit{attDOT} and B and B’ on \textit{attB}. The catalytic domain is responsible for the cleavage and ligation steps necessary for strand exchange between the DNA sites. The amino-terminal (N) domain of IntDOT interacts with the arm-type sites, which are positioned 40 base pairs (or more) away from the core-type sites. The N domain of lambda Int determines the order of strand exchanges and the directionality (integration versus excision) of the recombination reaction based on interactions with the arm-type sites (19, 29, 37).
In CTnDOT recombination, the role of the arm-type sites is more complex. Two of the sites (the L1 and R1′ sites) are required for both integration and excision. Three additional sites act cooperatively in integration (R1, R2, and R2′) but appear dispensable in excision (Figure 3.3) (16, 38). While lambda Int has a higher affinity for the arm-type sites than the core-type sites, IntDOT has a higher affinity for the core-type sites (27, 38). In the absence of core-type DNA, IntDOT shifts the L1 site weakly and does not shift the other arm-type sites without assistance from DNA bending proteins. The order of strand exchanges in IntDOT is determined by a conserved GC dinucleotide within the D core-type site rather than by interactions with the arm-type sites (20). These two base pairs of homology are essential, but the remaining five base pairs of the overlap region may be mismatched, particularly in recombination between natural sites (Figure 3.2).

Synthetic HJ intermediates containing the IntDOT core-type sites can be processed to both products and substrates if the overlaps are identical *in vitro* (Figure 3.1) (17). However, if the overlaps contain mismatches, the HJs are resolved back to substrates (Figure 3.2). This inability of IntDOT to process mismatched HJ intermediates to products may be due to the lack of arm-type sites which are bound by the N domain of IntDOT. Even if the sites were present, it is unlikely that the same IntDOT monomer could contact both a core and arm-type site without the assistance of a DNA bending protein.

I hypothesized that BHFa (or another DNA-binding protein) and the arm-type sites are necessary to form intasomes and enable IntDOT to catalyze recombination through the mismatched bases. To test this, I constructed larger synthetic HJs containing the arm-type sites. As with core-only HJs, they were resolved into either products or
substrates when the overlap region contained identical bases. I also constructed core plus arm-type HJs with a mismatched overlap. When these HJ intermediates were incubated with both IntDOT and BHFa they were resolved to both substrates and products. I conclude that intasome formation on the HJ intermediate containing mismatches is required for resolution to products.

Unlike bacteriophage lambda, the arm-type sites of attDOT seem to be more important for resolving HJ intermediates containing mismatched overlap regions than for controlling the directionality of the reaction. Accordingly, I also tested mutations in the arm-type sites to determine which of the arm-type sites are necessary when the overlap region is mismatched. The L1 arm-type site was required for efficient resolution to products, but mutations in the R1’ and R2-R2’ arm-type sites did not affect resolution levels or directionality. In fact, DNA bending by BHFa appears more important than the presence or absence of the attR arm sites tested for resolution of HJ intermediates.

**Materials and Methods**

*Enzymes, Reagents, and PCR Primers*

All oligonucleotides were ordered from IDT. Oligonucleotides were suspended in TE [pH 8.0] and stored at -80°.

\[\gamma^{32}\text{P}]\text{ATP} \text{ was ordered from Perkin-Elmer. T4 Polynucleotide Kinase was obtained from Fermentas.}
KOD PCR Master Mix was purchased from Novagen. Gel extractions were performed using Qiagen kits according to manufacturer’s instructions. PCR primer sequences can be found in Table 3.5.

Generating Holliday junction intermediates

Holliday junction intermediates were prepared similarly to the method of Kim and Gardner (17). Only the \textit{attB} oligonucleotides (SK1 or SK7) were labeled with [\(\gamma\)-\(^{32}\)P]ATP using polynucleotide kinase (Fermentas). Sequences for all oligonucleotides can be found in Table 3.6.

After labeling, equal amounts (24 picomoles) of the other oligonucleotides (\textit{attL}, \textit{attR}, and \textit{attDOT}) were mixed in the annealing buffer (10 mM Tris-HCl [pH 8], 100 mM KCl, 5 mM EDTA). The mixture was heated to 95\(^{\circ}\) in a PCR machine and cooled by 1\(^{\circ}\) per 30 seconds until it reached 50\(^{\circ}\). It was then cooled by 1\(^{\circ}\) per minute until reaching room temperature.

Holliday junction intermediates were gel purified by loading on a pre-run 8% polyacrylamide gel. The HJ intermediates were electrophoresed for about 2 hours at 150 V. After that time they were exposed to X-Ray film and excised from the gel. The HJ intermediates were ethanol precipitated, resuspended in 0.5 M KCl and stored at -20\(^{\circ}\). HJ Intermediates were quantified using a ThermoFisher Qubit 2.0 and ssDNA quantification kit according to the manufacturer’s instructions.
Resolving and visualizing Holliday junction intermediates and resolution results

The HJ resolution assays were performed in 50 uL reaction volumes containing 50 mM Tris-HCl [pH 8], 50 mM KCl, 1 mM EDTA, 0.1 mg/mL BSA, 5 mM DTT, and 6 ng/uL of Herring Sperm DNA. Herring sperm DNA was present to reduce non-specific DNA binding activity of IntDOT and BHFa to substrate DNA. Herring Sperm DNA was also found to increase resolution rates (data not shown).

Amounts of each HJ intermediate were standardized based on the Qubit 2.0 quantifications. Each reaction contained 17 femtomoles of the purified HJ intermediate. When included, IntDOT is present at 110 nM. BHFa is present at 90 nM.

Reactions were allowed to proceed for 2 hours at 37\(^\circ\). They were stopped by the addition of SDS to a final concentration of 0.4% and loaded onto a pre-run 8% polyacrylamide gel. The gels were run for 5-6 hours at 100 V. Gels were exposed to phosphoimager screens overnight, scanned, and results were analyzed and quantified using Fujifilm Image Gauge software (Macintosh v. 3.4).

Percent product was calculated by dividing the number of counts in the product band (attR) by the total counts of the product band, the substrate band (attB), and the unresolved HJ band combined.

Percent substrate was calculated by dividing the number of counts in the substrate band by the total counts of the product band, the substrate band, and the unresolved HJ band combined.
**Proteins**

IntDOT was purified as described in (12) except that an *ihfA* overexpression strain was used. BHFa was purified as described in (28) and diluted in IHF dilution buffer (50 mM Tris HCl [pH 8], 10% glycerol, 2 mg/ml BSA, and 200 mM KCl). Both IntDOT and BHFa proteins were stored at -80°. Proteins were quantified using a Qubit 2.0 and Qubit Protein Assay kit according to manufacturer’s instructions.

**Results**

Arm-type Site DNA provided in *trans* stimulates resolution of mismatched overlap HJ intermediates to products

It was previously shown that HJ resolution intermediates with only the core-type sites cannot be resolved efficiently to both substrates and products if the overlap region is mismatched (17). The core-only HJ intermediates lack any of the six arm-type sites which would be present *in vivo*. Unlike bacteriophage lambda Int, IntDOT has a higher affinity for core-type sites than arm-type sites (38). Since IntDOT can bind and resolve HJ intermediates with only the core-type sites, it is expected that the CB domain of up to four IntDOT monomers will bind to the HJ intermediate. However, the IntDOT amino terminal domains will be unoccupied and able to bind DNA in *trans*.

To test the importance of the arm-type sites in HJ resolution, we incubated the core-only HJ intermediates with the following DNA fragments in *trans*: one containing the two *attL* arm sites (L1 and L2), another containing the four *attR* arm sites (R1’, R1, R2’, R2).
R2, R2'), and a 156 bp PCR product from the pUC19 plasmid that contains none of the arm-type site sequences (Figure 3.4 A-C).

When the overlap region of the HJ intermediate is identical, IntDOT can resolve the intermediate into either products (24%) or back to substrates (54%) (Figure 3.5). The addition of either arm-type or nonspecific DNA in trans has little effect on resolution (Figure 3.5, lanes 3-5). While overall resolution is increased slightly (5-10%) in the presence of additional DNA, the directionality is entirely towards the substrate (Figure 3.6, columns C, D, and E). The slight bias towards substrates occurs whether attL arm DNA, attR arm DNA, or nonspecific DNA are added. Higher concentrations of these DNA substrates than the ones shown (200 nM) do not increase overall resolution or the bias towards products (data not shown).

If the arm-type sites are provided in trans, HJ intermediates with a mismatched overlap can be resolved to either products or to substrates. Without arm-type DNA, we saw extremely low levels of resolution to product (around 5%) when the overlap was mismatched (Figure 3.7, lane 2). The attL arm DNA stimulated IntDOT resolution of mismatched overlap HJs to products, up to 11% (Figure 3.7, lane 3).

The four attR arm-type sites also stimulated resolution of products to 13% (Figure 3.7, lane 4). Adding both the attL and attR arm-type sites simultaneously did not stimulate resolution compared to adding them singly (data not shown). This may be because of the strong consensus between the L1 and R1' arm-type sites; eight of the ten bases are identical (Figure 3.3). In the absence of DNA-bending proteins, the N domain of IntDOT is probably interacting with only the highest affinity sites, the L1 or R1' sites. It was previously shown that DNA bending is important for IntDOT interactions with the
R1, R2, and R2′ arm-type site DNA in trans, so it is likely that these sites are contributing less than the R1′ site (38). As a result, IntDOT may not be able to make full contact with multiple attR arm-type sites, and so the result is similar to the fragment containing the L1 arm-type site.

The addition of non-specific DNA also stimulates low levels of resolution to product, about 7% (Figure 3.7, lane 5). The herring sperm DNA in the resolution buffer, which is of varying lengths, yields 5% resolution to products (Figure 3.7, lane 2). DNA binding (including nonspecific binding) by the N domain of IntDOT may stimulate the catalytic domain and explain this result. However, the levels of resolution with nonspecific DNA were lower than with either attL or attR arm-type sites and primarily towards substrates (Figure 3.8, columns C and D compared with E).

Even when the arm-type sites are provided in trans to mismatched overlap core-only HJ intermediates, the resulting levels of product are lower than identical overlap HJ intermediates, as shown in Table 3.1. It is likely that contact with arm-type DNA by the N domain of IntDOT stimulates either protein-protein interactions with other IntDOT monomers or that it affects the catalytic activity of same monomer. However, it is unlikely that adding the arm-type DNA in trans fully mimics the coordination found when the different arm-type sites are present in cis as part of the same molecule as the core-type sites. This lack of coordination may explain the lower resolution levels, especially when the overlap region is mismatched. Overall, it seems that DNA provided in trans does stimulate resolution of mismatched overlap HJ intermediates. The arm-type site sequences bias resolution towards products but the effect is modest.
BHFa is required for resolution to products even when arm-type DNA is present in cis

Larger Holliday junction intermediates were constructed, containing the R1’, R1, R2, and R2’ arm-type sites and the L1 site. The L2 site was omitted for technical reasons and because it was shown previously that it is not required for integration or excision (39). One version of the HJ intermediates contained an identical overlap region (Figure 3.9) and the other contained a mismatched overlap region (Figure 3.10). These HJ intermediates were tested for resolution, with and without the DNA-bending protein BHFa. BHFa was previously identified for its role in integration of CTnDOT. There are four BHFa binding sites within the full *attDOT* sequence, but only the H2 site is entirely included within the core-only HJ intermediate, since it is located between the D’ core site and the L1 arm-type site (Figure 3.11). Five base pairs of the BHFa H3 site are included in the core-only HJ intermediate (28). As a DNA-bending protein, the addition of BHFa alone to HJ intermediates does not result in any resolution (data not shown).

Core-only HJ intermediates contain only the BHFa H2 site. However, the addition of BHFa to core-only HJ intermediates reduces overall resolution (Figure 3.12, lanes 3 and 6). When the overlap region is identical, total resolution of the core-only HJ intermediate was reduced from 67% to 47% when BHFa was added (Figure 3.13, columns B and C; Table 3.2). When the overlap region is mismatched, resolution was reduced from 65% to 33% (Figure 3.13, columns E and F; Table 3.2). BHFa binding to core-only HJs may compete with IntDOT binding, though previous IntDOT footprints do
not overlap with the H2 site (12). It is also possible that BHFa binding bends the smaller HJs into non-productive conformations.

In the core plus arm-type site HJs intermediates, additional BHFa binding sites are present. In addition to the H2 site, the H3 site and a small portion of the H4 site are present in the north arm of the HJ intermediate. The H3 site overlaps with the L1 site. The H1 site is present in the east arm, overlapping the R2-R2′ arm-type sites (Figure 3.14).

Similarly to the core-only HJ intermediates, the identical overlap core plus arm-type site HJ intermediate was resolved to either substrate (16%) or products (56%) with IntDOT alone (Figure 3.12, lane 8). Adding BHFa increased resolution towards products to 31% (Figure 3.12, lane 9), presumably by bringing the arm-type sites into a favorable conformation for the resolution complex (Figure 3.13, columns H and I).

When the HJ intermediate containing the arm-type sites and a mismatched overlap was incubated with IntDOT alone, almost no resolution to product (3%) was seen; the mismatched overlap HJ intermediate was primarily (61%) resolved to substrate (Figure 3.12, lane 11). The addition of BHFa increased resolution to products to 15% (Figure 3.12, lane 12). As with the HJ intermediate with an identical overlap region, total resolution remained similar when BHFa is added (Figure 3.13, columns K and L). These results are summarized in Table 3.2. Based on these results, the presence of the arm-type sites in cis is not sufficient to confer directionality to resolution of HJ intermediates. In the absence of BHFa, both directionality and total resolution are similar between core-only and core plus arm-type HJ intermediates (Figure 3.13; columns B and H, E and K).
Both arm-type sites in *cis* and BHFa are necessary to form the higher-order complexes that increase resolution to products.

**L1 arm-type mutants are defective in resolution to products when the overlap region is identical**

In order to establish the importance of the individual arm-type sites in resolution of HJ intermediates, different versions of each HJ intermediate were constructed. The L1, R1’, and R2-R2’ sites were selected based on previous studies indicating their importance in integration and excision of CTnDOT (16, 38). The L1 and R1’ sites are required for integration and excision. Either the R2 site or R2’ site are required in combination with the R1 site for integration (Figure 3.3). Mutating both the R2 and R2’ sites did not eliminate integration or excision previously, but they were selected to test the cooperativity of the R arm-type sites in HJ resolution.

As in the previous work, each of the selected sites was changed to a HindIII sequence (Figure 3.15). The resulting arm-type site mutant HJ intermediates were constructed with either an identical or mismatched overlap region and tested for resolution. While some of the altered bases are within BHFa binding sites (H1 for the R2-R2’ mutant and H3 for the L1 mutant), the BHFa binding site appears to be degenerate and involve contacts with numerous bases, much like IHF. In addition, BHFa has the highest affinity for the H2 site, which was unaffected by any arm-type site mutations (13). Based on mutations to these sites tested in the *in vitro* integration assay, we do not believe that these changes (four to six base pairs) affected BHFa binding (data
not shown). Only interactions between the IntDOT N domain and the arm-type site DNA should be affected.

In HJ intermediates with an identical overlap region, when the L1 arm-type site is mutated but BHFa is absent, resolution to products dropped from 16% to 12% for the wild-type sequence (Figure 3.16, lanes 2 and 5). Overall resolution actually increased by about 12% but was biased towards substrates (Figure 3.17, column E). Adding BHFa when the L1 site was mutated did not compensate for the mutated arm-type site (Figure 3.16, lanes 5 and 6). While resolution to products improved with BHFa present, up to 18%, it was still less than the 31% product when the wild-type L1 site was present (Figure 3.16, lanes 3 and 6). The increased resolution was pushed entirely towards substrate, suggesting the L1 site is important in the directionality of the reaction (Figure 3.17, columns E and F). Without the wild-type L1 site, IntDOT is less capable of resolving even an identical overlap HJ intermediate towards products. The addition of BHFa may restore some of the disrupted higher-order structure, but the loss of the interactions between the L1 site and the N domain of IntDOT reduces the amount of resolution to products.

Mutations in either the R1’ site or the R2-R2’ sites together did not substantially affect resolution levels or directionality (Figure 3.16, lanes 7-12). Increased resolution in R1’ mutants was very slight and offset by the inclusion of BHFa (Figure 3.17, columns H and I). The ratio of resolution to products and substrates remained the same. The R2-R2’ mutants followed the same pattern (Figure 3.16, lanes 11 and 12). The results for the arm-type site mutants of identical overlap HJ intermediates are summarized in Table 3.3.
The L1 Arm-Type site is required for resolution of HJ intermediates with mismatched overlap regions to products

The arm-type site mutants were also tested with HJ intermediates containing mismatched overlap regions. In the mismatched overlap HJ intermediates, L1 mutants without BHFa returned to background levels of products (Figure 3.18, lane 5). The addition of BHFa resulted in low (4%) levels of product (Figure 3.18, lane 6), but actually decreased overall resolution to 50% (Figure 3.19, columns E and F; Table 3.4). Based on this result, we conclude that BHFa still binds to HJ intermediates with a mutated L1 site but is unable to restore the necessary contacts between the N domain of IntDOT and the \textit{attL} arm when the L1 site is mutated. These contacts seem to be especially important when the overlap region is mismatched (compare Figure 3.17, columns E and F to Figure 3.19, columns E and F). Overall, it appears that the contact between the L1 site and the N domain of IntDOT is a key element for the directionality of HJ resolution.

As in the identical overlap HJ intermediates, both the R1’ and R2-R2’ site mutants were not substantially different from the wild-type arm-type sites (Figure 3.18, lanes 7-12). Mismatched overlap HJ intermediates with R1’ mutants had the same ratios of product and substrate as the wild-type HJ intermediates (Figure 3.19, columns H and I).

R2-R2’ mutants had slightly higher resolution to products and higher overall resolution (Figure 3.19, columns K and L). This result suggests that the R2-R2’ sites are redundant with other sites or that they are not necessary for resolution of HJ intermediates. The \textit{attR} arm-type sites have previously been shown to be cooperative, so
it may be that the R1 or R1′ sites must simultaneously be mutated in addition to the R2-
R2′ sites to see an effect (38). The results of these HJ intermediates with arm-type site
mutants and a mismatched overlap region are summarized in Table 3.4.

The results of mismatched overlap region HJ intermediates with arm-type site
mutants parallel the identical overlap region HJ intermediates, particularly with the attR
arm sites. None of the attR sites tested are required for resolution or directionality.
However, the L1 site is required; without this site resolution to products drops to
background levels. The addition of BHFa does not offset the effect of the mutated L1
site, making this the most important site tested for resolution of HJ intermediates to
products.

Discussion

In this study, we examined the requirements for forming intasomes with synthetic
HJ intermediates. We tested HJ intermediates with either identical or mismatched
overlap regions and tested the role of the arm-type sites in trans or in cis. When the arm-
type sites were present in trans, neither attL nor attR arm DNA stimulated resolution to
products when the overlap region was identical. Either attL or attR arm DNA stimulated
low levels of resolution to products when the overlap region was mismatched, but so did
a nonspecific DNA sequence.

While the structure of the IntDOT intasome (either integrative or excisive) is not
known, it is likely that efficient resolution is based on both protein-protein interactions
between the IntDOT monomers and regulation from the N domain of each monomer. For
example, a crystal structure of lambda Int tetramers associated with a Holliday junction indicated extensive interactions between the monomers (4). In the experiments with core-only HJ intermediates and arm-type site DNA in trans, the arm-type site DNA was presumably bound by the N domain of IntDOT, but any higher-order coordination was absent. For example, DNA bending proteins increase the affinity of IntDOT for a DNA fragment containing R1′, R1, R2, and R2′ sites (38). Without these interactions, it is not surprising that levels of resolution to product are relatively low.

The fact that either the attL or the attR arm-type site DNA provided in trans can stimulate resolution to products may reflect the needs of the excisive intasome. It is possible that one site is used during integration but blocked by an accessory factor (such as Xis2c or Xis2d) during excision. The other site may be blocked during integration (most likely by BHFa or another DNA bending protein) but available during assembly of the excisive intasome.

The HJ intermediates (as annealed oligonucleotides) lack the DNA supercoiling that would be present in vivo. Integration requires the attDOT substrate to be supercoiled in vitro (10, 22). The requirements for supercoiled substrates in excision are not known. The lack of supercoiling may be one of the reasons that IntDOT does not resolve 100% of HJ intermediates the way that other recombinases (such as lambda Int) do in in vitro HJ resolution studies (6).

Surprisingly, we found that, of the arm-type sites tested, only the L1 arm-type site was required for resolution to products when the overlap region is mismatched. When the overlap region is identical, resolution was decreased but not eliminated when the L1 site was mutated. It was already known that IntDOT could resolve identical
overlap HJ intermediates in the absence of any arm-type sites (18). These findings suggest that contact between the N domain of IntDOT and the L1 site help drive resolution to products of a mismatched overlap HJ intermediate. It is possible that contact with the L1 site is involved in IntDOT protein-protein interactions by correctly positioning one monomer to interact with another in a complex. Alternatively, N domain binding to the L1 site may influence the catalytic domain of IntDOT. For instance, the arm-type domains interact with the N domain of lambda Int in order to regulate the activity of the core binding and catalytic domains (4, 5, 26, 30). Either one of these possibilities could explain the drop in resolution of the identical overlap HJ intermediates when the L1 site was mutated. The contact with the L1 site is essential rather than stimulatory when the overlap region is mismatched, so that when L1 was mutated, the mismatched core HJ intermediates could only be resolved back to substrates.

We found that BHFa is required for IntDOT to interact with arm-type sites in cis and stimulates resolution to products when the overlap region is identical. As with the L1 site, these interactions are more important when the overlap region is mismatched. In the core-type only HJs, the addition of BHFa did not stimulate resolution to products and inhibited overall resolution. This suggests that there is no inherent property of BHFa or contacts with IntDOT that increase resolution towards products. Thus, the role of BHFa is to bend DNA into suitable conformations for IntDOT monomers to contact both the core and arm-type sites. Overall, the addition of BHFa more dramatically affected resolution of HJ intermediates to products than the presence of any arm-type site tested except the L1 site. This suggests that the higher-order dynamics of the intasome are extremely important in HJ resolution by IntDOT. BHFa appears to induce the DNA
bends necessary for the catalytic, core-binding, and N domains of IntDOT to interact productively in resolving HJ intermediates.

Based on these results, the attR arm-type sites appear to be redundant in HJ resolution, suggesting that single and even double sites may be mutated without affecting directionality or overall resolution. This redundancy may be an evolutionary adaptation by the CTnDOT element in order to maintain a wider host range. In Bacteroides spp., BHFa appears to bend the DNA into a favorable conformation for resolution to products, but other potential hosts will have other nucleoid-associated proteins, with different binding affinities and bend angles. It is possible that the cooperative attR sites stabilize these less than ideal DNA complexes and are less critical when BHFa is used. Alternatively, the suitability of the host nucleoid-associated proteins may represent a limit on the CTnDOT host range. Another possibility is that (as suggested in previous work) that individual attR arm-type sites may be dispensable due to the accessory factors (such as Exc, Xis2c, and Xis2c) used in CTnDOT recombination (38). A final possibility is that, similar to experiments in the lambda system, it is necessary to mutate multiple attR arm-type sites before observing a drastic effect in resolution of HJ intermediates (3, 24).

Overall, IntDOT resolution of HJ intermediates containing mismatched bases represents an unusual ability among tyrosine recombinases. Tn916 is another integrase from an ICE that may be able to resolve mismatched bases in a HJ intermediate (9, 14, 25, 31). While Tn916 has sequences that are functionally similar to arm-type sites, they have not been shown to play a part in resolution of mismatched bases. We believe that
this is the first instance of the arm-type sites providing the ability for a tyrosine recombinase to resolve through mismatched bases of a HJ intermediate.
References


Figure 3.1: Core-type only Holliday junction intermediate with an identical overlap region.

The top portion shows the two double-stranded substrates for CTnDOT integration that will lead to an identical overlap region.

The lower portion shows the Holliday Junction intermediate containing only core sites and with an identical overlap region.

In both, small black arrows indicate the sites of the first strand exchange. The large, open arrows are the sites for the second strand exchange.

The circle at the 5’ end of attB represents the $^{32}$P label. Only attB (substrate) and attR (product) will be visible due to this label.

Solid lines (attL, attR) indicate the top strand. Dashed lines (attDOT, attB) indicate the bottom strand.

Bent arrows to the right indicate possible results of resolution by IntDOT. Shaded boxes indicate the presence of the $^{32}$P label.

Figure adapted and modified from (17).
Figure 3.2: Core-type only Holliday junction intermediate with a mismatched overlap region.

The top portion shows two double-stranded substrates for CTnDOT integration that will lead to a mismatched overlap region. Mismatched bases are capitalized and in red.

The lower portion shows the Holliday Junction intermediate containing only core sites and with an identical overlap region.

In both, small black arrows indicate the sites of the first strand exchange. The large, open arrows are the sites for the second strand exchange.

The circle at the 5′ end of attB represents the ^32P label. Only attB (substrate) and attR (product) will be visible due to this label.

Solid lines (attL, attR) indicate the top strand. Dashed lines (attDOT, attB) indicate the bottom strand.

Bent arrows to the right indicate possible results of resolution by IntDOT. Shaded boxes indicate the presence of the ^32P label.

Figure adapted and modified from (17).
Figure 3.3: Arm-type sites within attDOT.

The top portion shows the spatial arrangement of the arm-type sites found in attDOT. The attR arm contains the R1′, R1, R2 and R2′ sites. The attL arm contains L1 and L2.

The core-type sites (D and D′) are indicated by circles and flank the overlap region.

Positions are relative to the center base of the overlap region, which is zero.

Boxes in black (R1′, L1 sites) are required for integration and stimulate excision.

Boxes in grey (R1, R2, R2′ sites) are required cooperatively for integration and excision.

The white box (L2 site) has showed no effect in either integration or excision.

The lower portion shows the sequences of each arm-type site and the consensus sequence. The grey boxes indicate cooperative interactions required for integration. If either light grey box (R2, R2′) is mutated along with the dark grey box (R1) integration will not occur.

Figure adapted from (38) and findings from (38) and (16).
Figure 3.4: DNA sequences added to core-only HJ intermediates in trans.

A. The *attL* arm DNA fragment. L1 and L2 arm-type sites are indicated with solid boxes. This fragment is 155 base pairs long and contains 38% GC bases. The D’ site was included because smaller regions did not amplify well and is marked with a dashed box.

B. The *attR* arm DNA fragment. R1’, R1, R2, and R2’ arm-type sites are indicated with solid boxes. The fragment is 149 base pairs long and contains 42% GC bases.

C. A portion of the pUC19 plasmid amplified to serve as a non-specific control. This fragment contains no arm-type site bases. It is 156 basepairs in length and contains 48% GC bases.
Figure 3.5: Resolution of identical overlap HJ intermediates with arm-type site DNA in *trans*.

L-ATS indicates the *attL* arm DNA fragment containing the L1 and L2 sites.  
R-ATS indicates the *attR* arm DNA fragment containing the R1’, R1, R2, and R2’ sites.  
NS is the non-specific control.

Resolution to both product (*attR*) and substrate (*attB*) is seen in lanes 2-5.

All DNA substrates in *trans* are present at 200 nM.  IntDOT is present at 110 nM.  HJ intermediates are present at 0.34 nM.
Figure 3.6: Quantification of resolution of identical overlap HJ intermediates with arm-type site DNA in \textit{trans}.

Product indicates \textit{attR}. Substrate indicates \textit{attB}.

L-ATS indicates the \textit{attL} arm DNA fragment containing the L1 and L2 sites.

R-ATS indicates the \textit{attR} arm DNA fragment containing the R1', R1, R2, and R2' sites.

NS is the non-specific control.

“Percent signal” is calculated using the intensity of the bands visualized using the phosphoimager. The total signal is determined by adding together the substrate band (\textit{attB}), the product band (\textit{attR}), and any unresolved HJ.

<table>
<thead>
<tr>
<th>IntDOT DNA</th>
<th>Percent signal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- A</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>+ B</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>+ C</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>+ D</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>+ E</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.7: Resolution of mismatched overlap HJ intermediates with arm-type site DNA in *trans*.

L-ATS indicates the *attL* arm DNA fragment containing the L1 and L2 sites.

R-ATS indicates the *attR* arm DNA fragment containing the R1′, R1, R2, and R2′ sites.

NS is the non-specific control.

Resolution to product is extremely low in the absence of DNA in *trans* (lane 2).

All DNA substrates in *trans* are present at 200 nM. IntDOT is present at 110 nM. HJ intermediates are present at 0.34 nM.
Figure 3.8: Quantification of resolution of mismatched overlap HJ intermediates with arm-type site DNA in *trans*.

L-ATS indicates the *attL* arm DNA fragment containing the L1 and L2 sites.

R-ATS indicates the *attR* arm DNA fragment containing the R1’, R1, R2, and R2’ sites.

NS is the non-specific control.

All DNA substrates in *trans* are present at 200 nM. IntDOT is present at 110 nM. HJ intermediates are present at 0.34 nM.

“Percent signal” is calculated using the intensity of the bands visualized using the phoshoimager. The total signal is determined by adding together the substrate band (*attB*), the product band (*attR*), and any unresolved HJ.
Table 3.1: Summary of resolution of core-only HJ intermediates with arm-type DNA sites in *trans*.

<table>
<thead>
<tr>
<th>Overlap</th>
<th>IntDOT</th>
<th>DNA in <em>trans</em></th>
<th>Product (%)</th>
<th>Substrate (%)</th>
<th>Resolution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical</td>
<td>-</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>L-ATS</td>
<td>24</td>
<td>54</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>R-ATS</td>
<td>23</td>
<td>60</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>NS</td>
<td>24</td>
<td>58</td>
<td>82</td>
</tr>
<tr>
<td>Mismatched</td>
<td>-</td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>L-ATS</td>
<td>5</td>
<td>82</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>R-ATS</td>
<td>11</td>
<td>74</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>NS</td>
<td>7</td>
<td>80</td>
<td>86</td>
</tr>
</tbody>
</table>

L-ATS indicates the *attL* arm DNA fragment containing the L1 and L2 sites.
R-ATS indicates the *attR* arm DNA fragment containing the R1’, R1, R2, and R2’ sites.
NS is the non-specific control.
Figure 3.9: Core plus arm-type site HJ intermediate with an identical overlap region.

Arm-type sites are marked with grey boxes.

The circle at the 5’ end of attB indicates the $^{32}$P label. Only attB (substrate) or attR (product) will be visible after resolution.

Small black arrows indicate the sites of the first strand exchange. Large open arrows indicate the sites of the second strand exchange.

Solid lines indicate top strands and dashed lines indicate bottom strands.

For this HJ intermediate, the visible substrate attB is 50 base pairs long. The other substrate attDOT is 249 base pairs long.

The product attL is 114 base pairs long. The visible product attR is 185 base pairs long.
Figure 3.10: Core plus arm-type site HJ intermediate with a mismatched overlap region.

Arm-type sites are marked with grey boxes.

Mismatched bases in the overlap are capitalized and in red.

The circle at the 5′ end of attB indicates the $^{32}$P label. Only attB (substrate) or attR (product) will be visible after resolution.

Small black arrows indicate the sites of the first strand exchange. Large open arrows indicate the sites of the second strand exchange.

Solid lines indicate top strands and dashed lines indicate bottom strands.

For this HJ intermediate, the visible substrate attB is 50 base pairs long. The other substrate attDOT is 249 base pairs long.

The product attL is 114 base pairs long. The visible product attR is 185 base pairs in length.
Figure 3.11: Locations of BHFa binding to core-only HJ intermediates.

Blue boxes indicate BHFa binding sites within the HJ intermediate. Numbers indicate bases protected by BHFa in footprinting trials.

“0” indicates the center base of the overlap region and is the second base north of the branch point.

Positive numbers indicate the attL (north) arm.

Solid lines indicate top strands. Dashed lines indicate bottom strands.
### Figure 3.12: Comparison of core-only and core plus arm-type site HJ intermediates.

Resolution results of core-only HJ intermediates (lanes 1-6) or core plus arm-type site HJ intermediates (lanes 7-12).

IntDOT is present at 110 nM. BHFa is present at 90 nM. HJ intermediates are present at 0.34 nM.

<table>
<thead>
<tr>
<th>Overlap</th>
<th>Core only</th>
<th>Core plus arm-type sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identical</td>
<td>Mismatch</td>
</tr>
<tr>
<td>IntDOT</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>BHFa</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

\[ attR (59 \text{ bp}) \]
\[ attB (50 \text{ bp}) \]
Figure 3.13: Quantification of resolution of core-only and core plus arm-type site HJ intermediates.

Product indicates \textit{attR}. Substrate indicates \textit{attB}.

Quantifications are based on three separate trials, averaged together. Error bars are based on the standard error from the three trials.

“Percent signal” is calculated using the intensity of the bands visualized using the phosphoimager. The total signal is determined by adding together the substrate band (\textit{attB}), the product band (\textit{attR}), and any unresolved HJ.
Figure 3.14: Locations of BHFa binding to core plus arm-type site HJ intermediates.

Blue boxes indicate BHFa binding sites within the HJ intermediate. Numbers indicate bases protected by BHFa in footprinting trials.

“0” indicates the center base of the overlap region and is the second base north of the branch point. Negative numbers indicate the attR (east) arm while positive numbers indicate the attL (north) arm.

The arm-type sites include the following bases:

L1 site: +68 to +77
R1′ site: -141 to -150
R1 site: -131 to -140
R2 site: -50 to -59
R2′ site: -40 to -49

Solid lines indicate top strands and dashed lines indicate bottom strands.
<table>
<thead>
<tr>
<th>Overlap</th>
<th>ATS</th>
<th>BHFa</th>
<th>Product (%)</th>
<th>Substrate (%)</th>
<th>Resolution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identical</td>
<td>Absent</td>
<td>-</td>
<td>21</td>
<td>46</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>14</td>
<td>33</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Mismatched</td>
<td>Absent</td>
<td>-</td>
<td>5</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>3</td>
<td>30</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Core plus arm-type sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identical</td>
<td>R1', R1, R2, R2', L1</td>
<td>-</td>
<td>16</td>
<td>56</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>31</td>
<td>44</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Mismatched</td>
<td>R1', R1, R2, R2', L1</td>
<td>-</td>
<td>3</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>15</td>
<td>45</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Summary of resolution results for core-only and core plus arm-type site HJ intermediates.

ATS indicates arm-type sites.
Figure 3.15: Mutations in the arm-type sites.

Four arm-type sites were mutated to include HindIII sites to test IntDOT interactions with the site in the context of HJ resolution.

The original sequences are listed on the left.

The mutated sequences are listed on the right. The changed bases (4-6 bases) are underlined and bolded.

Due to their adjacent locations and cooperativity, the R2-R2′ sites were both changed.

<table>
<thead>
<tr>
<th></th>
<th>Original Sequence</th>
<th>Altered Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1′</td>
<td>5′- GGTTACGAAC</td>
<td>5′- G<strong>AAGCTT</strong>AAC</td>
</tr>
<tr>
<td>R2 and R2’</td>
<td>5′- TGTAACTCAA</td>
<td>5′- T<strong>AAGCTT</strong>CAA</td>
</tr>
<tr>
<td></td>
<td>5′- AGTGATAGAA</td>
<td>5′- A<strong>AAGCTT</strong>GAA</td>
</tr>
<tr>
<td>L1</td>
<td>5′- GGTTACGATT</td>
<td>5′- <strong>G</strong>AAG<strong>C</strong>TTATT</td>
</tr>
</tbody>
</table>
Figure 3.16: Resolution of identical core HJ intermediates with arm-type site mutations.

Lanes 1-3 contain the wild-type arm-type sites and are reproduced from Figure 3.11.
Lanes 4-6 contain a HJ intermediate with a mutated L1 site.
Lanes 7-9 contain a HJ intermediate with a mutated R1’ site.
Lanes 10-12 contain a HJ intermediate with mutated R2 and R2’ sites.

IntDOT is present at 110 nM. BHFa is present at 90 nM. HJ intermediates are present at 0.34 nM.
Figure 3.17: Quantification of resolution of identical core HJ intermediates with arm-type site mutations.

Product indicates \textit{attR}. Substrate indicates \textit{attB}.

Quantifications are based on three separate trials, averaged together. Error bars are based on the standard error from the three trials.

“Percent signal” is calculated using the intensity of the bands visualized using the phosphoimager. The total signal is determined by adding together the substrate band (\textit{attB}), the product band (\textit{attR}), and any unresolved HJ.
Table 3.3: Summary of resolution results from identical overlap HJ intermediates containing arm-type site mutations.

<table>
<thead>
<tr>
<th>Overlap</th>
<th>Mutation</th>
<th>BHFa</th>
<th>Product (%)</th>
<th>Substrate (%)</th>
<th>Resolution (%)</th>
<th>Percent of WT Product</th>
<th>Percent of WT Substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical</td>
<td>WT</td>
<td>-</td>
<td>16</td>
<td>56</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>31</td>
<td>44</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identical</td>
<td>L1</td>
<td>-</td>
<td>12</td>
<td>72</td>
<td>84</td>
<td>73</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>18</td>
<td>66</td>
<td>82</td>
<td>58</td>
<td>150</td>
</tr>
<tr>
<td>Identical</td>
<td>R1’</td>
<td>-</td>
<td>17</td>
<td>58</td>
<td>75</td>
<td>107</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>29</td>
<td>43</td>
<td>72</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>Identical</td>
<td>R2-R2’</td>
<td>-</td>
<td>17</td>
<td>61</td>
<td>78</td>
<td>104</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>30</td>
<td>43</td>
<td>73</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

Core plus arm-type site HJ intermediates contain five arm-type sites: the R1’, R1, R2, R2’, and the L1 site.
Figure 3.18: Mismatched core HJ intermediates with arm-type site mutations.

Lanes 1-3 contain the wild-type arm-type sites and are reproduced from Figure 3.11.

Lanes 4-6 contain a HJ intermediate with a mutated L1 site.
Lanes 7-9 contain a HJ intermediate with a mutated R1’ site.
Lanes 10-12 contain a HJ intermediate with mutated R2 and R2’ sites.

IntDOT is present at 110 nM. BHFa is present at 90 nM.
HJ intermediates are present at 0.34 nM.
Figure 3.19: Quantification of resolution of mismatched core HJ intermediates with arm-type site mutations.

Product indicates $\text{attR}$. Substrate indicates $\text{attB}$.

Quantifications are based on three separate trials, averaged together. Error bars are based on the standard error from the three trials.

“Percent signal” is calculated using the intensity of the bands visualized using the phosphoimager. The total signal is determined by adding together the substrate band ($\text{attB}$), the product band ($\text{attR}$), and any unresolved HJ.
Table 3.4: Summary of resolution results from mismatched overlap HJ intermediates containing arm-type site mutations.

Core plus arm-type site HJ intermediates contain five arm-type sites: the R1’, R1, R2, R2’, and the L1 site.
<table>
<thead>
<tr>
<th>Name</th>
<th>Length</th>
<th>PCR product</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>P19 156 F1</td>
<td>20 bp</td>
<td>Nonspecific DNA</td>
<td>GAT TAG CAG AGC GAG GTA TGT AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fragment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P19 156 R1</td>
<td>21 bp</td>
<td>Nonspecific DNA</td>
<td>CTC TTG ATC CGG CAA ACA AAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fragment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRJ/MM 176 F</td>
<td>21 bp</td>
<td>R-ATS</td>
<td>CCC GTT TGT CGC ATT ATC GGG</td>
<td>(38)</td>
</tr>
<tr>
<td>DLJ/MM 25R</td>
<td>21 bp</td>
<td>R-ATS</td>
<td>CTA TCA CTA TAT CTT TCC GCA</td>
<td>(38)</td>
</tr>
<tr>
<td>Thermo 2F</td>
<td>20 bp</td>
<td>L-ATS</td>
<td>GCT CAA GGA AGT TAC GAA AA</td>
<td></td>
</tr>
<tr>
<td>DLJ/INW-1/R127R</td>
<td>21 bp</td>
<td>L-ATS</td>
<td>CTC TTC GTA TGC AAA AGT AGC</td>
<td>(38)</td>
</tr>
</tbody>
</table>

Table 3.5: PCR Primers.
<table>
<thead>
<tr>
<th>Name</th>
<th>Length</th>
<th>Site</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK1</td>
<td>50 bp</td>
<td>bs <em>attB</em> (mismatched overlap)</td>
<td><em>AAT TAC TGT TTA GTA TTT TAA TTG CGC AAA TTT ACT GCA AAT TTC CGA GC</em></td>
<td>(17)</td>
</tr>
<tr>
<td>SK 3</td>
<td>94 bp</td>
<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 CAC TAA GCA AAG TTA CTA CAA AAA AGT GAA ATG CGT AGG TTA ATC GCT CTT G</em></td>
<td>(17)</td>
</tr>
<tr>
<td>SK 4</td>
<td>85 bp</td>
<td>ts <em>attL</em></td>
<td><em>GCT CGG AAA TTT GCA GTA AAT TTG CTT AGT GAA GTT ACG AAA AGA GTG TAA TTT ATT GGA AAT GAG CGT AGG TTA ATC GCT CTT G</em></td>
<td>(17)</td>
</tr>
<tr>
<td>SK 6</td>
<td>59 bp</td>
<td>ts <em>attR</em> (mismatched overlap)</td>
<td><em>CTT TCC GCA TTT CAC TTT TTT GTA GTA ACT TTG CGC AAT TAA AAT ACT AAA CAG TAA TT</em></td>
<td>(17)</td>
</tr>
<tr>
<td>SK 7</td>
<td>50 bp</td>
<td>bs <em>attB</em> (identical overlap)</td>
<td><em>AAT TAC TGT TTA GTA TTT TAA CTA AGC AAA TTT ACT GCA AAT TTC CGA GC</em></td>
<td>(17)</td>
</tr>
<tr>
<td>SK 8</td>
<td>59 bp</td>
<td>ts <em>attR</em> (identical overlap)</td>
<td><em>CTT TCC GCA TTT CAC TTT TTT GTA GTA ACT TTG CTT AGT TAA AAT ACT AAA CAG TAA TT</em></td>
<td>(17)</td>
</tr>
<tr>
<td>L114</td>
<td>114 bp</td>
<td>ts <em>attL</em></td>
<td><em>GCT CGG AAA TTT GCA GTA AAT TTG CTT AGT GAA GTT ACG AAA AGA GTG TAA TTT ATT GGA AAT GAG CGT AGG TTA ATC GCT CTT GAT AGT TAG TTA</em></td>
<td></td>
</tr>
<tr>
<td>DOT 249</td>
<td>249 bp</td>
<td>bs <em>attDOT</em></td>
<td><em>TAA CTA ACT AAA TCG TAA CCT ATT ACT ATC AAG AGC GAT TAA CCT ACG CTC ATT TCC AAT AAA TTA CAC TCT TTT CGT AAC TTT ACT AAG CAA AGT TAC TAC AAA AAA GTG TAA TGC GAA AAG ATA TAG TGA TAG AAT TGA GTT ACA AGG GTT TTA GTT GAA GTG GCT ATA ATT CGT GAA GCC AGT ACA ATC ACA GCA AAA GCC AAA TCC TGA CGC CAC AAG AGT ACT TGT TCG TAA CTA TGC CGC ATA</em></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.6: HJ intermediate oligonucleotides.**

Asterisks indicate 5' [γ-32P] labels.

Underlines indicate the overlap region.

Bolded bases indicate changes from wild-type sequence to HindIII sites in mutated arm-type sites.

All sequences are listed from 5’ to 3’.
Table 3.6 (continued)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R185</strong></td>
<td><strong>185 bp</strong></td>
<td><strong>ts attR</strong> (identical overlap)</td>
<td>* TAT CGG GCA TGG TTA CGA ACA AGT AAC GTT GTG GCG TCA GGA TTT GGC TTT TGC TGT GAT TGT ACT GGC TTC AGCAATTATAGCCACTTCAACTAAAAAC CTTTGTAACTCAATTCTATCATCTATAT CTTTTCGATTTTTACTTTTTTTGAGTA ACTTTGCTTTAGTTAAATACTAAACAG TAAATT</td>
<td></td>
</tr>
<tr>
<td><strong>R185 MM</strong></td>
<td><strong>185 bp</strong></td>
<td><strong>ts attR</strong> (mismatched overlap)</td>
<td>* TAT CGG GCA TGG TTA CGA ACA AGT AAC GTT GTG GCG TCA GGA TTT GGC TTT TGC TGT GAT TGT ACT GGC TTC AGC CAC TTC AAC TAA AAC CCT TGT AAC TCA ATT CTA TCA CTA TAT CTT TTC GCA TTT CAC TTT TTT GTA GTA ACT TTG CGC AAT TAA AAT ACT AAA CAG TAA TT</td>
<td></td>
</tr>
<tr>
<td><strong>L1</strong></td>
<td><strong>114 bp</strong></td>
<td><strong>ts attL</strong> (mutant L1 site)</td>
<td>* GCT CGG AAA TTT GCA GTA AAT TTG CTT AGT GAA GTT ACG AAA AGA GTG TAA TTT ATT GGA AAT GAG CGT AGG TTA ATC GCT CTT GAT AGT AAT AGA AGC TTA TTT AGT TAG TTA</td>
<td></td>
</tr>
<tr>
<td><strong>DOT L1</strong></td>
<td><strong>249 bp</strong></td>
<td><strong>bs attDOT</strong> (mutant L1 site)</td>
<td>* TAA CTA ACT AAA TAA GCT TCT ATT ACT ATC AAG AGC GAT TAA CCT ACG CTC ATT TCC AAT AAA TTA CAC TCT TTT CTT AAC TCT ACT AAG CAA TTC AAT CCT ACG ACA AAA GCA AAA GCA AAA TCC TGA CGC CAC AAC GTC ACT TGT TCG TAA CCA TGC CGG ATA TGA AGC TTA</td>
<td></td>
</tr>
<tr>
<td><strong>R1 id</strong></td>
<td><strong>185 bp</strong></td>
<td><strong>ts attR</strong> (mutant R1′ site; identical overlap)</td>
<td>* TAT CGG GCA TGA AGC TTA ACA AGT AAC GTT GTG GCG TCA GGA TTT GGC TTT TGC TGT GAT TGT ACT GGC TTC AGC CAC TTC AAC TAA AAC CCT TGT AAC TCA ATT CTA TCA CTA TAT CTT TTC GCA TTT CAC TTT TTT GTA GTA ACT TTG CTT AGT TAA AAT ACT AAA CAG TAA TT</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.6 (continued)

<table>
<thead>
<tr>
<th>R1</th>
<th>185 bp</th>
<th>ts <code>attR</code></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td></td>
<td>(mutant R1’ site; mismatched overlap)</td>
<td>* TAT CGG GCA TGA AGC TTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACA AGT AAC GTT GTG GCG TCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGA TTT GGC TTT TGC TGT GAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGT ACT GGC TTC ACG AAT TAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AGC CAT TAC AAG TAA AAG GAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAT CTT TTC GCA TTT CAT TTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TTT GTA GTA ACT TTG CGC AAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAA AAT ACT AAA CAG TAA TT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R1</th>
<th>249 bp</th>
<th>bs <code>attDOT</code></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DOT</td>
<td></td>
<td>(mutant R1’ site)</td>
<td>* TAA CTA ACT AAA TCG TAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCT ATT ACT ATC AAG AGC GAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAA CCT ACG CTC ATT TCC AAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAA TTA CAC TCT TTT CGT AAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TTC ACT AAG CAA AGT TAC TAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AAA AAA GTG AAA TGC GAA AAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATA TAG TGA TAG AAT TGA GTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACA AGG GTT TTA GTT GAA GTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCT ATA ATT CGT GAA GCC AGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACA ATC ACA GCA AAA GCC AAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TCC TGA CGC CAC AAC GTT ACT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGT TAA GCT TCA TGC CGG AAT</td>
</tr>
</tbody>
</table>

| R2-2’ | 185 bp | ts `attR` |  |
| id   |        | (mutant R2 and R2’ sites; identical overlap) | * TAT CGG GCA TGG TTA CGA |
|      |        |             | ACA AGT AAC GTG GTG GCG TCA |
|      |        |             | GGA TTT GGC TTT TGC TGT GAT |
|      |        |             | TGT ACT GGC TTC ACG AAT TAT |
|      |        |             | AGC CAT TAC AAG TAA AAG GAG |
|      |        |             | TTA GCT TCA ATT CAA GCT TTA |
|      |        |             | TAT CTT TTC GCA TTT CAT TTT |
|      |        |             | TTT GTA GTA ACT TTG CTT AGT |
|      |        |             | TAA AAT ACT AAA CAG TAA TT |

| R2-2’ | 185 bp | ts `attR` |  |
| MM   |        | (mutant R2 and R2’ sites; mismatched overlap) | * TAT CGG GCA TGG TTA CGA |
|      |        |             | ACA AGT AAC GTG GTG GCG TCA |
|      |        |             | GGA TTT GGC TTT TGC TGT GAT |
|      |        |             | TGT ACT GGC TTC ACG AAT TAT |
|      |        |             | AGC CAT TAC AAG TAA AAG GAG |
|      |        |             | TTA GCT TCA ATT CAA GCT TTA |
|      |        |             | TAT CTT TTC GCA TTT CAT TTT |
|      |        |             | TTT GTA GTA ACT TTG CGC AAT |
|      |        |             | TAA AAT ACT AAA CAG TAA TT |
| R2-2’ DOT | 249 bp | bs *att*DOT (mutant R2 and R2’ sites) | * | TAA CTA ACT AAA TCG TAA  
|     |       |  
|     |       | CCT ATT ACT ATC AAG AGC GAT  
|     |       | TAA CCT ACG CTC ATT TCC AAT  
|     |       | AAA TTA CAC TCT TTT CGT AAC  
|     |       | TTC ACT AAG CAA AGT TAC TAC  
|     |       | AAA AAA GTG AAA TGC GAA AAG  
|     |       | ATA TA AAG TTA GGG TAG TGG  
|     |       | CTA AAG TTA GTA TTA AAG TGG  
|     |       | CAA TGG TAG TCT AAA CAG AA  
|     |       | CAA TCA CAG CAA AAG CCA AAT  
|     |       | CCT GAC GCC ACA ACG TTA CTT  
<p>|     |       | GTT CGT AAC CAT GCC CGA TA |</p>
<table>
<thead>
<tr>
<th>HJ</th>
<th>Overlap</th>
<th>ATS (wild-type)</th>
<th>Mutation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK-1</td>
<td>Identical</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>SK-2</td>
<td>Mismatch</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>KR-3</td>
<td>Identical</td>
<td>L1, R1’, R1, R2, R2’</td>
<td>None</td>
</tr>
<tr>
<td>KR-4</td>
<td>Mismatch</td>
<td>L1, R1’, R1, R2, R2’</td>
<td>None</td>
</tr>
<tr>
<td>KR-5</td>
<td>Identical</td>
<td>R1’, R1, R2, R2’</td>
<td>L1</td>
</tr>
<tr>
<td>KR-6</td>
<td>Mismatch</td>
<td>R1’, R1, R2, R2’</td>
<td>L1</td>
</tr>
<tr>
<td>KR-7</td>
<td>Identical</td>
<td>L1, R1, R2, R2’</td>
<td>R1’</td>
</tr>
<tr>
<td>KR-8</td>
<td>Mismatch</td>
<td>L1, R1, R2, R2’</td>
<td>R1’</td>
</tr>
<tr>
<td>KR-9</td>
<td>Identical</td>
<td>L1, R1’, R1</td>
<td>R2, R2’</td>
</tr>
<tr>
<td>KR-10</td>
<td>Mismatch</td>
<td>L1, R1’, R1</td>
<td>R2, R2’</td>
</tr>
</tbody>
</table>

Table 3.7: Oligonucleotides used in HJ intermediates and arm-type sites present.
CHAPTER 4: Summary and Future Directions

Summary

As an ICE, CTnDOT must integrate into the host chromosome to persist after transfer via conjugation to a new host cell (5). In order to integrate, the element must recombine the joined ends (attDOT) with a partner site in the chromosome (attB). The tyrosine recombinase IntDOT and a DNA-bending protein are required for integrative recombination (5, 7).

Through my studies, I identified and characterized the Bacteroides Host Factor A (BHFa) protein, the first Bacteroides host protein identified for its role in a recombination pathway. I identified BHFa by purifying DNA-binding proteins from B. thetaiotaomicron strains lacking known ICEs and testing them in the in vitro integration assay. BHFa is a 22 kDa, basic protein with domains in common with the IHF-HU superfamily and a predicted secondary structure similar to those proteins. It is likely that BHFa is a Bacteroides nucleoid-associated protein and it is conserved among Bacteroides species.

Through the use of electrophoretic mobility shift assays, I determined that both IntDOT and BHFa bind to attDOT DNA. Using fluorescent footprinting, I determined the binding sites of BHFa within the attDOT sequence and determined that there are four sites: H1, H2, and the adjacent H3-H4 sites. Despite the fact that BHFa binds specifically within attDOT, IHF, HU, and other unrelated proteins can substitute for BHFa in the in vitro integration assay. This suggests that the specific binding sites of BHFa are less
important than the need for a DNA-bending protein in the *in vitro* integration assay. Accordingly, it is unlikely that the host factor protein limits the host range of CTnDOT.

In my second project, I examined the unusual ability of IntDOT to resolve HJ intermediates containing mismatched bases. It had been previously established that IntDOT could resolve core-only HJ intermediates to both products and substrates when the overlap region was identical. However, when the overlap region was mismatched, IntDOT was only able to resolve the HJ intermediate back to substrates (12). I found that adding nonspecific herring sperm DNA to the reaction allowed resolution of a mismatched overlap HJ intermediate to products, though at extremely low levels.

IntDOT is a heterobivalent protein that can interact with both arm and core-type sites (8, 19). The role of the core-type sites in strand exchange and integration was well studied (8, 16, 17). The importance of the arm-type sites in integration and excision suggested that these sites would be significant in HJ resolution (10, 28). Since the core-only HJs lacked either arm-type sites or DNA-bending proteins, the higher-order intasomes could not be formed.

After designing, assembling, and purifying larger HJ intermediates with the arm-type sites, I tested them for resolution. I found that BHFa biased resolution to products when the overlap region was identical. When the overlap region was mismatched, BHFa was required for efficient resolution of the HJ intermediates to products. Given that other proteins could substitute for BHFa in the *in vitro* integration assay, it was surprising that BHFa contributed so strongly to the directionality of HJ intermediates.
In order to test the importance of individual arm-type sites, I designed HJ intermediates with mutated arm-type site sequences. The R1′ site and R2-R2′ site did not affect resolution of HJ intermediates, whether the overlap region was identical or mismatched. The L1 site was important in both situations. When the overlap region was identical and the L1 site was mutated, resolution towards products decreased. When the overlap region was mismatched and the L1 site was mutated, resolution to products returned to background levels. Levels of resolution to product were only slightly increased by the addition of BHFa.

These results clarify the requirements for IntDOT to resolve mismatched bases in a HJ intermediate. IntDOT varies from the conserved tyrosine recombinase residues, suggesting that IntDOT may have an altered active site (13, 23). The affinity of IntDOT for core and arm-type sites is quite different from lambda Int as well (18, 28). With the findings from this work, we are closer to understanding the minimum requirements for an IntDOT intasome that can resolve mismatched bases.

**Future Directions**

**Peptide inhibitors of resolution**

Short, synthetic peptides have been developed that trap HJ intermediates (14). These peptides have been used in the capture and subsequent study of lambda HJ intermediates for both the integration and excision pathways (4). In particular, trapped pathways have been useful in identifying factors that can control directionality of the recombination reaction, such as spermidine (2, 3). We have received two of these
synthetic peptides from Anca Segall and plan to use them to trap integrative and excisive intasomes.

While the strand exchange order and participants of the CTnDOT integration reaction have been well-studied, the excision reaction is understood in less detail (20). If we are able to accumulate excision intermediates with the short peptides, we would be able to study the excisive complex and determine the differences from the integration reaction.

In addition, the HJ intermediates tested thus far lack any DNA supercoiling. Since DNA supercoiling is required for the attDOT substrate in the in vitro integration assay, it is likely that the lack of supercoiling affects overall resolution in other assays (7). Successful use of the peptide inhibitors could capture HJ intermediates between supercoiled substrates. These supercoiled HJ intermediates would better represent conditions in vivo and might yield higher resolution rates overall.

The role of accessory factors in resolution of HJ intermediates

As noted above, the CTnDOT excision reaction is not as well characterized as the integration assay. One reason for this is the number of participants. Xis 2c and Xis2d are required, while Exc appears to be stimulatory (5, 6, 26, 27). BHFa or another host factor are likely involved, at least in a stimulatory role. Two other Xis-like proteins, Orf2A and Orf2B may play a part as well (C. Hopp, unpublished data).

The role of these proteins has been established in regulating the mob and tra operons, but the purpose they serve in excisive recombination is not yet established (9,
11, 21). Testing these proteins in HJ resolution is complicated by the instability of Xis2c, which has proven difficult to purify. However, Xis2d and Exc should affect the directionality of HJ resolution. Further studies with HJ intermediates and the excision proteins may allow us to untangle the individual contributions of these numerous participants in excision.

The roles and redundancy of the attR arm-type sites

It was surprising to find that the R1′ site was not required for resolution of HJ intermediates to products, since the R1′ site is required in integration and in excision. The fact that it is not required in resolution of HJ intermediates poses several questions.

It is possible that the R1′ site is important for another stage of recombination, particularly since R1′ site binding by IntDOT seems to stimulate IntDOT interactions with the other attR arm-type sites (28). Synthetic HJ intermediates with core sites arranged around a branch point may bypass the need for these interactions.

In this case, the attR arm-type sites would be more important for generating HJ intermediates than for resolving them. Based on previous work, it will be necessary to mutate two to three of the four attR arm-type sites at once to see a measurable effect, however (1, 22, 28). It also possible that the immediately adjacent positioning of both the R1′-R1 pair and the R2-R2′ pair allows IntDOT to interact with one of the pair even when the other is unavailable.
Further characterization of IntDOT

As noted above, IntDOT varies in many respects from well-studied tyrosine recombinases such as lambda Int. While we have gained a better appreciation for the important residues and structure of IntDOT, there are still many questions about the N domain, protein-protein interactions, and the participation of individual amino acids in its structure and function (12, 13, 19).

We have previously attempted to generate crystal structures of IntDOT bound to DNA substrates. A crystal structure of IntDOT bound to a HJ intermediate would greatly enhance our understanding of the intasome complex.

The N domain and core-binding/catalytic domains of lambda Int have been separately purified. These experiments showed that the lambda Int N domain inhibits the topoisomerase, DNA binding, cleavage, and HJ resolution activities of the catalytic domain when they are part of the same protein. However, when present in trans, the N fragment enhanced all of these activities (15, 24, 25). This surprising interaction would be intriguing to test with IntDOT. While we have not yet generated separate IntDOT N domain or CB domain fragments, either one (if stable and functional) could provide further insights into interactions between each domain and its DNA binding sites. Separate domains would also make it possible to determine if and how the domains regulate one another.
References


