IntDOT Interactions with Core- and Arm-Type Sites of the Conjugative Transposon CTnDOT[▽]

Jeanne M. DiChiara,^{1†} Aras N. Mattis,² and Jeffrey F. Gardner^{1*}

Departments of Microbiology¹ and Biochemistry,² University of Illinois, Urbana, Illinois 61801

Received 28 November 2006/Accepted 19 January 2007

CTnDOT is a *Bacteroides* conjugative transposon (CTn) that has facilitated the spread of antibiotic resistances among bacteria in the human gut in recent years. Although the integrase encoded by CTnDOT (IntDOT) carries the C-terminal set of conserved amino acids that is characteristic of the tyrosine family of recombinases, the reaction it catalyzes involves a novel step that creates a short region of heterology at the joined ends of the element during recombination. Also, in contrast to tyrosine recombinases, IntDOT catalyzes a reaction that is not site specific. To determine what types of contacts IntDOT makes with the DNA during excision and integration, we first developed an agarose gel-based assay for CTnDOT recombination, which facilitated the purification of the native IntDOT protein. The partially purified IntDOT was then used for DNase I footprinting analysis of the integration site *attDOT* and the excision sites *attL* and *attR*. Our results indicate that CTnDOT has five or six arm sites that are likely to be involved in forming higher-order nucleoprotein complexes necessary for synapsis. In addition, there are four core sites that flank the sites of strand exchange during recombination. Thus, despite the fact that the reaction catalyzed by IntDOT appears to be different from that typically catalyzed by tyrosine recombinases, the protein-DNA interactions required for higher-order structures and recombination appear to be similar.

The conjugative transposon CTnDOT is a large self-transmissible element that is normally integrated into the chromosome. In order to transfer to a new recipient, it first excises from the donor chromosome to form a nonreplicating circular intermediate. This is accomplished by a site-specific recombination reaction between specific sites, *attL* and *attR*, that flank the element. A single strand is then transferred to the new recipient by conjugation, where complementary second-strand synthesis and circularization of the element occur. Finally, the transferred copy integrates into the recipient chromosome by recombination between a specific site on CTnDOT (*attDOT*) and one of several similar sites (*attB*) in the recipient chromosome (2, 16).

CTnDOT has an unusual mechanism of excision and integration. During the excision process, the CTnDOT integrase, IntDOT, makes staggered cuts in the chromosomal DNA 7 bp from the ends of the element (Fig. 1). Since these 7-bp segments (coupling sequences) are not complements of each other, joining of the ends to produce the circular transfer intermediate creates a small region of heterology that is presumably resolved during the transfer process. A similar process occurs during integration (2, 11).

Although the CTnDOT integrase catalyzes unusual types of excision/integration reactions, it has the C-terminal amino acid signature (Arg-Lys-His-Arg-His-Tyr) that is found in lambdoid phage integrases and other members of the tyrosine recombi-

nase family (20). Another difference between the tyrosine recombinases and IntDOT is that integration is not entirely site specific. There is a 10-bp sequence on one side of the coupling sequence of *attDOT* that has partial sequence similarity to one side of the chromosomal *attB* attachment sites. The significance of the sequence is not known, but it is conserved in all six sequenced *attB* sites (2).

It is likely that the *attDOT* and *attB* sites involved in the integration reaction, and the *attL* and *attR* sites involved in the excision reaction, are bound by IntDOT and accessory factors during recombination in a manner similar to that of other elements that are members of the tyrosine recombinase family. These proteins form nucleoprotein complexes, called intasomes, that are necessary for the regulation of the directionality of the recombination reaction. The attachment sites typically contain two classes of DNA recognized by their cognate recombinase. One class, called core sites, is arranged as imperfect inverted repeats that flank the region of crossover. These sites are bound by the C-terminal catalytic domains of the proteins. A second class of sites, called arm sites, is located further away from the crossover site. These sites are bound by the N-terminal DNA-binding domain of the protein and, in conjunction with accessory factors, are involved in forming integrative or excisive intasomes (1, 13–15, 20).

Since the sites to which IntDOT might bind the DNA were not known, we decided to use footprint analysis to identify the core and arm binding sites bound by IntDOT. We report here the sequences of the binding sites IntDOT uses on the *attDOT*, *attL*, and *attR* sites and deduce consensus core and arm binding sequences for IntDOT.

MATERIALS AND METHODS

Strains and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. The concentrations of antibiotics used in this study

* Corresponding author. Mailing address: Department of Microbiology, University of Illinois at Urbana-Champaign, 601 S. Goodwin Ave., Urbana, IL 61801. Phone: (217) 333-7287. Fax: (217) 244-6697. E-mail: jeffgard@life.uiuc.edu.

† Present address: Molecular Genetics Program, Wadsworth Center, New York State Department of Health, Center for Medical Sciences, 150 New Scotland Avenue, Albany, NY 12208.

[▽] Published ahead of print on 2 February 2007.

FIG. 1. A schematic representation of the excision and integration reactions of CTnDOT. IntDOT makes 7-bp staggered cuts (11) adjacent to the coupling sequences (denoted by the bold arrows). The element then circularizes to form a covalently closed intermediate, with a region of heterology created by the noncomplementary coupling sequences. Replication through CTnDOT during transfer to a new recipient forms one of the two coupling sequences. (Here, the *attR* coupling sequence is shown for the rest of the reaction.) A new target site is selected in the recipient chromosome, where the element again integrates by using staggered cuts. The sequences that flank the coupling sequences as imperfect inverted repeats are designated D, D', B, and B'. (Modified from reference 4.)

were as follows: ampicillin (Amp), 100 µg/ml; chloramphenicol (Cam), 20 µg/ml; and kanamycin (Kan), 50 µg/ml. *Escherichia coli* strains were grown in LB medium at 37°C unless otherwise noted.

Purification of native IntDOT and *E. coli* IHF. The native *intDOT* gene (GenBank accession no. AJ311171) was cloned into the NdeI and HindIII sites of the vector pET-30(a)+ (Novagen), using the primers *intDOT*-NdeI-F and *intDOT*-HindIII-R (Table 2). The resulting plasmid, pT-*int*, was sequenced at the UIUC Core Sequencing Facility to confirm that there were no PCR mutations. This plasmid was then transformed into BL21(DE3)/pLysS. Cells were grown overnight at 30°C in LB-Cam-Kan then diluted 1:50 into 500 ml fresh LB-Cam-Kan. The cells were grown at 30°C until they reached an optical density at 600 nm of approximately 0.6 and then induced with IPTG (isopropyl-β-D-thiogalactopyranoside) to a final concentration of 1 mM. The cells were grown at room temperature for another 4 hours and then pelleted by centrifugation at 5,000 × g for 15 min at 4°C. Cell pellets were frozen at -80°C overnight. The cells were thawed on ice and then immediately resuspended in 20 ml low-salt sodium phosphate buffer (50 mM sodium phosphate, pH 7.0, 1 mM EDTA, 50

mM sodium chloride, 5% glycerol, 1 mM dithiothreitol [DTT]) plus a Roche Complete EDTA-free protease inhibitor tablet. Following resuspension, the cells were lysed by sonication. The cell lysate was then centrifuged at 4°C at 10,000 × g for 30 min. The cleared lysate was kept on ice until it was loaded onto a fast protein liquid chromatography column. The entire 20-ml lysate was loaded onto a 20-ml Amersham HiTrap heparin column. Following a 30-ml wash with low-salt phosphate buffer, 1-ml fractions were collected while a high-salt gradient (0 M NaCl to 2 M NaCl) was run over 150 ml. The fractions were subjected to electrophoresis by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis to determine where IntDOT eluted from the column. The peak fractions were then tested for activity using a gel-based *in vitro* integration assay (see below). The most active fractions were then pooled and dialyzed into storage buffer (low-salt phosphate buffer with 40% glycerol) overnight. The resulting partially purified IntDOT was stored at -80°C and used for footprinting analysis. Each aliquot of IntDOT was freeze-thawed no more than once prior to use.

IHF was purified as described previously (10) from BL21(DE3)/pLysS, except that a Mono S column was not used. The protein concentrations for both IHF

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype/phenotype	Description (source or reference)
<i>E. coli</i> strains		
DH5 α MCR	F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>dlacZ</i> Δ <i>M15</i> Δ (<i>lacZYA-argF</i>) <i>U169 endA1 recA1 supE44 l-thi-1 gyrAa96 relA1</i>	(Gibco BRL)
BL21(DE3)/pLysS	F ⁻ <i>ompT hsdS_B</i> (r _B ⁻ m _B ⁻) <i>gal dcm</i> (DE3)/pLysS Cam ^r	High-stringency protein expression strain (Novagen)
Plasmids		
pET-30a(+)	Kan ^r T7lac	An <i>E. coli</i> cloning vector for protein overexpression (Novagen)
pGEM-T	Amp ^r	An <i>E. coli</i> cloning vector for PCR products (Promega)
pGEM-T+ <i>attDOT</i>	Amp ^r	pGEM-T carrying a 618-bp <i>attDOT</i> site (Q. Cheng)
pPR204	Amp ^r	A plasmid carrying the native wild-type IHF genes (P. A. Rice)
pT ₇ <i>int</i>	Kan ^r T7lac	pET-30a(+) carrying the native IntDOT gene (this study)
pYS56	Amp ^r	A 537-bp <i>attL</i> product of DLJ/U487F and DLJ/INW-241R cloned into pGEM-T (19)
pYS57	Cam ^r <i>pir</i>	An <i>SpeI</i> - <i>SstII</i> fragment of pYS55 containing <i>attR</i> cloned into pEP185.2 (19)

and IntDOT were determined using the standard protocol for the DC protein assay (Bio-Rad). The purified IHF preparation used for footprinting was 1.5 mg/ml, while the partially purified IntDOT preparation was 2.0 mg/ml.

Development of a gel-based in vitro integration assay. We developed a gel-based in vitro integration assay that was a modification of the previously described in vitro integration assay (3). Rather than a two-plasmid assay that required electroporation, this modified gel assay used supercoiled plasmid containing *attDOT* and a linear radiolabeled *attB* (Fig. 2A). Previous experiments determined that two linear substrates are unable to perform in vitro integration or excision (19; J. F. Gardner, unpublished results). The plasmid containing *attDOT*, pGEM-T+*attDOT* (Table 1), was purified using a QIAGEN maxiprep kit. Radiolabeled linear *attB* substrates were prepared as follows: one *attB* oligonucleotide, #3-B-Top (Table 2), was labeled using T4 polynucleotide kinase (Fermentas) with [γ -³²P]dATP (PerkinElmer); the labeled #3-B-Top oligonucleotide was then annealed to its complementary unlabeled oligonucleotide (GC-#3-B-Top) (Table 2).

Integration assays (20- μ l volumes) were performed with 1 $\times$ reaction buffer (20 mM Tris-HCl [pH 7.4], 5 mM DTT, 0.05 mg/ml bovine serum albumin [BSA], 1% glycerol, 50 mM KCl), purified *E. coli* IHF (diluted 1:50 in 50 mM Tris-HCl, pH 8.0, 10% glycerol, 2 mg/ml BSA, and 50 mM KCl), and various concentrations of IntDOT. *E. coli* IHF is required for the integration reaction to proceed in vitro (3; J. F. Gardner and J. M. DiChiara, unpublished results). IntDOT was diluted using 50 mM Tris-HCl, pH 8, 1 mM DTT, 1 mM EDTA, 50 mM NaCl, 10% glycerol, and 2 mg/ml BSA. The reaction mixtures were then incubated at 37°C for 2 h, then stopped by the addition of stop solution (30% glycerol, 10% SDS, 0.5% xylene cyanol, 0.5% bromophenol blue), and loaded onto a 1% agarose gel. The gel was run in 1 $\times$ Tris-borate-EDTA buffer. The gel was dried, and the bands were visualized using a phosphorimager.

Footprinting substrates. Singly ³²P-end-labeled DNA substrates for footprinting *attL*, *attR*, and *attDOT* were prepared as described previously (4), using the primers and templates listed in Table 2. PCR was performed using the following protocol: 95°C for 2 min (denaturing step) and then 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds for a total of 15 cycles, followed by a final 72°C extension for 9 min. Four hundred nanograms of template was used per 100- μ l PCR volume. All attachment site DNA was quantified by spectrophotometry prior to footprinting.

DNase I footprinting. Binding reactions (100- μ l volumes) were carried out in the presence of 1 $\times$ binding buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 50 mM NaCl, 10% glycerol, 0.25 mg/ml BSA), 2.5 mM CaCl₂, 5 mM MgCl₂, and native IntDOT. IntDOT was diluted in IntDOT dilution buffer (see above). Footprinting substrates were added to a final concentration of 0.01 nM to 1 nM for each reaction. The mixtures of the binding reactions were incubated at room temperature for 30 min. DNase I (Worthington) was added to a final concentration of 0.1 mM per reaction and incubated at room temperature for 1 min. The DNase I digestion was stopped by adding 90 μ l of stop solution (200 mM NaCl, 30 mM EDTA, 1% SDS, 100 μ g/ml *Saccharomyces cerevisiae* tRNA [Sigma]), followed by vortexing. The reaction products were then cleaned up using standard protocols.

Equal amounts of counts were loaded for each sample on an 8% denaturing polyacrylamide gel. Gels were prerun for a minimum of 1 h at 55 to 60 W before the samples were loaded. The gels were then run at 55 to 60 W for 2 to 4 h, depending on the sample. Sequencing reactions were run adjacent to the footprinting sample. These were performed using a USB PCR sequencing kit following protocol no. 70170, excluding the first step. The PCR substrates for sequencing were produced using the same templates and primers as those used to make the footprinting substrates, but without any label. The PCR products

TABLE 2. PCR and sequencing primers used in this study

Primer	Sequence (5'-3')	Template
#3-B-Top ^a	CGT TGC TCG GAA ATT TGC AGT AAA TTT GCG CAA TTA AAA TAC TAA ACA GTA ATT ATA TCA TGG CAG C	NA ^b
GC-#3-B-Top ^a	GCT GCC ATG ATA TAA TTA CTG TTT AGT ATT TTA ATT GCG CAA ATT TAC TGC AAA TTT CCG AGC AAC G	NA
attR-FP-F2	GCC ATC CAT ACC CGT TTG T	pYS57
attR-FP-R2	TAT ATA ATT ACG CTT CGC CAT TTC TAC	pYS57
attL-FP-F2	GTT TTC AAA GAC TAT GGA GGC G	pYS56
attL-FP-R2	GAG GGG ATA AAA AAA CGG GAG T	pYS56
intDOT-NdeI-F	GAA ACC AAA AAA TTA AGG CAT ATG AA	Chromosomal DNA
intDOT-HindIII-R	GTC TCG CTT CAA AGC TTC TCT TTT TTA	Chromosomal DNA
DOT-FPSEQ-F	CCA TGG CCG CGG GAT GCC AT	pGEM-T+attDOT
DOT-FPSEQ-R	CTG CAG GCG GCC GCA CTA GT	pGEM-T+attDOT

^a These oligonucleotides were annealed for use in the gel-based in vitro integration assay.

^b NA, not applicable.

FIG. 2. (A) A schematic representation of the gel-based in vitro integration reaction for CTnDOT. A supercoiled plasmid carrying the *attDOT* junction was incubated with a linear radiolabeled *attB* in the presence of reaction buffer (see Materials and Methods), purified *E. coli* IHF, and native IntDOT at 37°C. The reaction mixture was then subjected to electrophoresis on a 1% agarose gel to separate the ~2-kb linear recombinants from the 67-bp *attB* substrate. (B) An example of the gel-based in vitro integration reaction. The reaction was performed as described above and in Materials and Methods, using either purified His₆-tagged IntDOT or a crude extract (CE) containing overexpressed native IntDOT. Recombinants were run on a 1% agarose gel as linear ~2-kb radiolabeled fragments.

were then cleaned up using a QIAquick spin kit (QIAGEN) and eluted with 30 μ l or 50 μ l EB (10 mM Tris-HCL, pH 8.5).

RESULTS

Partial purification of the native IntDOT protein. Previous in vitro studies of CTnDOT employed a His₆-tagged version of the IntDOT protein (3, 4, 11, 19). The tagged IntDOT performs recombination but is inefficient in vitro. For example, in vitro integration frequencies average approximately 1.5% recombination (3) while excision frequencies average approximately 0.1% recombination (4). Footprinting attempts with the tagged protein proved to be difficult and gave inconsistent results (A. Gupta and J. M. DiChiara, unpublished results). Therefore, we overexpressed and partially purified the native form of IntDOT to be used in our footprinting studies.

The native *intDOT* gene was cloned and expressed as described in Materials and Methods. The crude extract was prepared as described in Materials and Methods and tested for integration activity using a newly developed gel-based in vitro integration assay (Fig. 2A and B). Under conditions where both proteins showed their maximal activities, the native IntDOT protein displayed greater recombination activity than the His₆-tagged protein, as shown in Fig. 2B.

The crude extract containing the native IntDOT was then subjected to heparin agarose column purification, as described in Materials and Methods. The resulting peak fractions were analyzed by SDS-polyacrylamide gel electrophoresis and showed the presence of an approximately 48-kDa protein, which could be IntDOT (Fig. 3). The fractions also contain a protein that migrates at approximately 40 kDa, which could be

FIG. 3. Partial purification of the native IntDOT protein. The native form of IntDOT was overexpressed as described in Materials and Methods. The predicted molecular mass of native IntDOT is approximately 48 kDa. This SDS gel shows fractions collected from heparin column purification (see Materials and Methods). Lanes: 1, MultiMark ladder (Invitrogen); 2, BenchMark ladder (Invitrogen); 3 to 13, heparin fractions 37 to 47, respectively. The lanes marked with asterisks indicate the fractions with peak IntDOT activity.

a degradation product. These same fractions were then tested for integration activity using the gel-based *in vitro* integration assay. The results of these assays revealed that several fractions contained IntDOT activity, with the peak activity found between fractions 41 and 43 (data not shown). This partially purified preparation of native IntDOT was subsequently used for all the DNase I footprinting studies presented in this report.

DNase I footprinting of *attDOT*. DNase I footprinting was carried out using partially purified native IntDOT. The *attDOT* site located on CTnDOT was subjected to footprint analysis using PCR-generated DNA labeled on the top strand (Fig. 4) or the bottom strand (Fig. 5). Top- and bottom-strand end-labeled restriction fragments of *attDOT* were also analyzed (data not shown). The results of these analyses revealed protection adjacent to the coupling sequence on the top strand of *attDOT* (Fig. 4). This protection spans from bases -17 to $+18$ (bases are numbered by convention, with the middle base of the coupling sequence denoted as “0,” bases to the right are given negative numbers, and bases to the left are given positive numbers). Protection around the coupling sequence was expected, since IntDOT must bind there to perform the recombination chemistry. These sites, D and D’, therefore represent core-type binding sites. The *attDOT* top-strand footprints also revealed two arm-type sites located on the right end of the DNA (Fig. 4). The site closest to the coupling sequence is denoted as “R2” and shows protection from approximately bases -36 to -60 . This protection is most visible in lane 4 of Fig. 4. An additional arm-type site (L1) is located on the left end of the DNA; however, the exact location of this site cannot be determined because the resolution of the gel is not sensitive enough to resolve the sequence at the top of the gel (Fig. 4).

The bottom strand of *attDOT* also displayed protection around the coupling sequence, in the D and D’ sites, spanning bases -17 to $+19$ (Fig. 5A). The arm-type sites determined by this footprint include the R2, L1, and L2 sites (Fig. 5A and B). The bases protected by the L1 and L2 sites each span about 20 bases: from $+64$ to $+84$ in L1 and from $+106$ to $+126$ in L2.

FIG. 4. DNase I footprint of the top strand of *attDOT*, using PCR-generated DNA. Lanes: 1, DNA only; 2 to 3, DNase I; 4 to 7, IntDOT dilutions of 1:2, 1:4, 1:8, and 1:16, respectively. The lines ending with circles denote IntDOT protection. The coupling sequence is denoted as CS. See Materials and Methods for DNA and protein concentrations.

The footprint at R2 is subtle, but the disappearance of the DNase I-cleaved bands located at base -36 up to approximately base -60 indicates the endpoints of this site (Fig. 5A).

In a previous work, deletions were used to determine the minimal *attR* and *attL* sites required for excision *in vitro* (4). The minimal *attR* site was determined to be 179 bp, which includes 164 bp of CTnDOT DNA and 15 bp of bacterial DNA, including the 7-bp coupling sequence. The minimal *attL* site was determined to be 153 bp, which includes 138 bp of CTnDOT DNA and 15 bp of bacterial DNA, including the 7-bp coupling sequence. DNase I footprinting was performed on the full-length *attR* and *attL* sites in an attempt to confirm our deletion and mutagenesis studies and to independently confirm all of the IntDOT binding sites. The top strand of *attR* (data not shown) showed protection around the coupling sequence, although this protection was not as strong as that seen with *attDOT*. However, there is discernible protection surrounding the coupling sequence spanning from bases -17 to $+18$, particularly in the 1:2 and 1:4 IntDOT dilutions. This region includes the D site that was identified on *attDOT* (Fig. 1). Since the excision junctions contain sequences from both CTnDOT and the chromosome, the B’ core-type site present in the bacterial DNA of *attR* was also identified (Fig. 1). In

FIG. 5. DNase I footprint of the bottom strand of *attDOT*, using PCR-generated DNA. Lanes: 1, DNA only; 2 to 3, DNase I; 4 to 7, IntDOT dilutions of 1:2, 1:4, 1:8, and 1:16, respectively; 8, DNase I. IntDOT protection is denoted by the lines ending with circles. The coupling sequence is denoted as CS. (A) Bases ~ -60 to $+114$. (B) Bases $\sim +114$ to $+124$. These bottom-strand-labeled *attDOT* fragments are from the same footprinting gel. See Materials and Methods for DNA and protein concentrations.

addition, this footprint showed protection of the same R2 arm-type site that was found on the right end of *attDOT*, which spans bases -63 to -37 (Fig. 4, lane 4).

The bottom-strand footprint of *attR* also showed protection around the coupling sequence. This strand actually gave slightly clearer protection than what was found for the top strand (Fig. 6, lanes 4 to 7). Although DNase I cleavage in this area is still weak, the disappearance of bands at bases -8 through $+9$ can be seen. The two regions protected around the coupling sequence correspond to the D and B' sites. This bottom-strand footprint also revealed the R1 arm-type site (data not shown) located between bases -149 and -129 . Interestingly, the first arm-type site, R1, is located very close to endpoints of deletions made previously. In fact, the deletion of bases from -106 to -165 caused an approximately 10-fold drop in the in vitro excision frequency (4). Therefore, our footprinting analysis of *attR* not only has corroborated the binding sites of IntDOT with those found on *attDOT* but also has reinforced the results from previous deletion experiments.

FIG. 6. Detail of the *attR* bottom-strand DNase I footprint surrounding the coupling sequence. Lanes: 1, DNA only; 2 to 3, DNase I; 4 to 7, IntDOT dilutions of 1:2, 1:4, 1:8, and 1:16, respectively; 8, DNase I. IntDOT protection is denoted by the lines ending with circles. The coupling sequence is denoted as CS. See Materials and Methods for DNA and protein concentrations.

FIG. 7. *attL* top-strand DNase I footprint surrounding the coupling sequence. Lanes: 1, DNA only; 2 to 3, DNase I; 4 to 7, IntDOT dilutions of 1:2, 1:4, 1:8, and 1:16, respectively; 8, DNase I. IntDOT protection is denoted by the line ending with circles. The band exhibiting enhanced cleavage in the presence of IntDOT is marked by a circle only. The coupling sequence is denoted as CS. See Materials and Methods for DNA and protein concentrations.

The results of footprinting *attL* showed protection of the D' and B sites on both the top and the bottom strands (Fig. 7 and data not shown). The top strand showed protection corresponding to the D' and B cores, spanning bases +19 to -23 (Fig. 7). While this protection was seen only with the most concentrated IntDOT (Fig. 7, lane 4), there is one strongly protected band at base +9. This base was also protected in the D' site on *attDOT* (Fig. 4 and 5); however, it is much more dramatic on *attL* (Fig. 7). The bottom strand of *attL* also displayed protection surrounding the coupling sequence, corresponding to the B' and D cores (data not shown). Additionally, arm-type sites were identified on both the top and the bottom strands of *attL* (data not shown) that correspond to the L1 and L2 arm-type sites.

As demonstrated previously (3) and in this work, *E. coli* IHF can function with IntDOT to perform recombination in vitro. While the native host factor can be provided with a *Bacteroides thetaiotaomicron* crude extract (3), IHF is not the natural host factor for the reaction because *B. thetaiotaomicron* does not encode a protein related to IHF. Interestingly, Segall et al. showed that complexes of λ *att* DNA and integrase could be formed with proteins that bind the DNA nonspecifically (17). *E. coli* HU or eukaryotic HMG1 and HMG2 proteins can replace IHF with λ *att* sites to generate specific complexes with the DNA. These complexes have electrophoretic mobility and biochemical activity similar to those of complexes formed by IHF and λ integrase (λ Int). Segall et al. proposed that cooperation between HU (or HMG1 or HMG2) and λ Int promotes the formation of higher-order nucleoprotein complexes that depend upon the ability of the nonspecific binding proteins to bend the DNA (17). IHF interacts with specific sites

FIG. 8. DNase I footprint of the top strand of *attDOT*, using PCR-generated DNA. Lanes: 1, DNA only; 2, DNase I (no protein); 3 to 5, IHF dilutions of 1:5, 1:10, and 1:20, respectively; 6, DNase I. IHF protection is denoted by the line ending in diamonds, while bands enhanced in the presence of IHF are denoted by diamonds only. The coupling sequence is denoted as CS. See Materials and Methods for DNA and protein concentrations.

but also has the ability to bind nonspecifically to DNA (for a review, see reference 5). It is unlikely that there are specific IHF recognition sites at the appropriate positions in *attDOT* to allow IHF to bind specifically to form complexes with IntDOT. We believe that it is much more likely that IHF binds nonspecifically to *attDOT* DNA and cooperates with IntDOT to form nucleoprotein complexes that are active in recombination. Thus, nonspecific binding of IHF likely induces bends in the DNA that allow IntDOT and IHF to form nucleoprotein complexes that are active in recombination.

Figure 8 shows a footprint analysis with *E. coli* IHF and the top strand of *attDOT*. We observe protection in the region of the L1 site and several sites through *attDOT* where DNase I cleavage is enhanced. It is possible that there is a specific IHF recognition site in the L1 region, but there is no good match to the consensus IHF site in that region of *attDOT*. We interpret this result to indicate that IHF may bind specifically to one region of *attDOT* but that its interactions with other regions of the DNA are nonspecific.

FIG. 9. IntDOT protection of the *attDOT* site, as determined by DNase I footprinting analysis. The black brackets indicate the minimal DNA sequence required on the R and L ends. The gray bracket at base -105 indicates the endpoint of the *attR* deletion that causes an ~ 10 -fold drop in the in vitro excision frequency, and the gray bracket at base $+103$ indicates the endpoint of the *attL* deletion that causes an ~ 10 -fold drop in the in vitro excision frequency (4). Regions protected by IntDOT are shaded in gray, while the coupling sequence is boxed.

DISCUSSION

Detailed analysis of specific recombination systems that use heterobivalent integrases showed that the integration and excision reactions require highly coordinated nucleoprotein structures, called intasomes (1). The best studied of these systems is the phage λ site-specific recombination system. Footprinting studies of the *attP* site of λ employing DNase I, neocarzinostatin, and methylation protection revealed two distinct classes of λ Int binding sites: the high-affinity arm-type binding sites and the low-affinity core-type binding sites (13–15). The arm-type sites, which flank the crossover region, bind to the N-terminal domain of λ Int. In the integration reaction, λ Int and IHF assemble on the *attP* site to form the integrative intasome, which then captures the “naked” *attB* to allow integration to proceed (12). Similarly, λ Int, IHF, Xis, and an additional *E. coli*-encoded protein, factor for inversion stimulation (FIS), form intasomes on *attL* and *attR* that allow excision to proceed. The core-type sites interact with the core binding and catalytic domains of λ Int and perform the strand exchanges to form recombinant products during both reactions (1).

We have performed DNase I footprinting analyses of the *attDOT*, *attR*, and *attL* sites of CTnDOT to identify the IntDOT binding sites. These studies show that IntDOT binds the core-type binding sites D and D', comprised of CTnDOT sequences, and B and B', comprised of bacterial chromosome sequences (Fig. 9 and 10). Consistent with other well-studied

members of the λ Int family, the core-type sites identified on CTnDOT are located adjacent to the crossover region as imperfect inverted repeats (Fig. 9 to 11). One of these core-type sites, the D site, was identified previously (4) as a site required for efficient excision. Mutagenesis of bases -4 to -9 not only abolished in vitro excision but also virtually eliminated IntDOT binding to the *attR* junction, as determined by electrophoretic mobility shift assay (4). The core-type sites identified by footprinting contain a 13-base-pair consensus sequence, shown in Fig. 10A. This consensus contains a centrally located conserved GTA sequence found in all four of the core-type sites. There are also three conserved T residues found in all four sites. Other positions show less conservation.

A total of four arm-type regions were identified by the footprint analyses. Because the footprints for the regions called R2 and L2 are larger than those for the other two, it is possible that these regions each contain two arm sites. We were able to derive two putative arm-type sites for the R2 region. We have called these sites R2A and R2B (Fig. 10B). The L2 site does not contain a second site that appears to be related to the ones identified. A consensus sequence derived from a comparison of these five arm-type sites is shown in Fig. 10B. The length of the consensus sequence is similar to that of the P'1 arm site of λ *attP*, as defined by mutations that disrupt λ Int binding (8). This consensus sequence contains 3 out of 12 bases that are found in all five sites and may be important for IntDOT interactions with DNA. These sites each span approximately 20 bp,

FIG. 10. (A) The four core-type sites determined by DNase I footprinting. The 6-bp site within the D core that is required for excision in vitro is in bold and underlined (4). A consensus sequence derived from the alignment of all four of the core sites is shown at the bottom. The triangles represent the locations of IntDOT cleavage. (B) An alignment of the arm-type sites determined by DNase I footprinting reveals a 12-bp consensus sequence present in all five arm-type sites. D = A, T, or G; H = A, T, or C; K = T or G; M = A or C; R = A or G; W = A or T; Y = T or C.

which is comparable in size to footprints of the arm-type sites present on λ (14). All five arm-type sites lie within the previously determined minimal regions on the right and left ends of *attDOT* (Fig. 9). In particular, both the R1 and L2 sites are located very close to the end of the minimal DNA sequence required for excision in vitro. In fact, the deletion of either site reduces excision in vitro 10-fold, indicating that these arm-type sites are necessary for efficient excision to proceed. Presumably, binding of IntDOT to these sites coordinates the formation of the nucleoprotein excisive intasome complex (4).

The results presented for the CTnDOT attachment sites

contribute to the growing diversity and complexity of attachment sites used by tyrosine recombinases. The arrangement, spacing, orientation, and number of arm-type sites vary widely within the *att* sites of elements in the tyrosine recombinase family (1), as demonstrated by the comparison of attachment sites shown in Fig. 11. As detailed in these studies, CTnDOT contains five or six arm-type binding sites in two orientations, with spacing that varies among the systems shown in Fig. 11. One current hypothesis for the differences in the orientations and spatial arrangements of arm sites among attachment sites is that each element forms unique intasome structures. Thus, the overall structures would be quite different from one element to another. Alternatively, the higher-order structures of all the elements may have similar overall structures (1).

We previously performed mutational analysis to determine the sequence requirements necessary for excision in vitro. We found that 6 bp within the D site of *attR* (Fig. 1) was necessary for excision in vitro because mutagenesis of these 6 bp abolished excision to below-detectable levels. However, except for mutations that affected phasing of the DNA, many mutations made in the *attL* core (from bases +4 to +26) did not exhibit a great effect on excision. These phasing mutations within the *attDOT* DNA of *attL* (Fig. 9) indicated that a 4-bp insertion after base +15, which should cause a 180° turn in the downstream DNA, gave an approximately 30-fold drop in the in vitro excision frequency (4). We hypothesized that this dramatic turn in the downstream *attL* site adversely affected protein-protein interactions within the excisive intasome by placing one of the arm-type sites on the “wrong” face of the DNA to interact with the other proteins in the complex. Several other phasing mutations in the *attL* site also caused detrimental effects on excision: insertions after bases +9, +22, and +28 each caused about a 15-fold decrease in the excision frequency (4). Our current footprint results reinforce this interpretation, since we now know that both *attL* arm-type sites are located downstream of base +60 (Fig. 9).

Curiously, none of the *attR* phasing mutations previously studied, located between bases -4 and -27, caused a great effect on in vitro excision. Eight different phasing mutations were constructed and tested, and the most detrimental effect was only a sixfold drop in the excision frequency (4). The R1 site is approximately 130 base pairs from the site of recombination. This distance is close to the persistence length of DNA

FIG. 11. A schematic diagram of the integrase binding sites for the well-studied tyrosine recombinase family members λ , P2, P22, and Tn916 (1, 9, 18), as well as the newly defined CTnDOT sites. Arrows indicate the binding sites, with the names of the sites listed below them. All five elements contain core-type binding sites that flank the cut and crossover region (marked as “0”) as inverted repeats; however, the numbers, locations, and orientations of the arm-type binding sites differ significantly between the elements. The sizes of the sites are approximate and are not drawn to scale.

so that phasing mutants may not affect intramolecular bridging. However, it is possible that the R1 and R2 sites are not affected by phasing because they might participate in intermolecular bridging during the excision reaction. For instance, in the λ system, λ Int is capable of making intra- or intermolecular bridges between the same or different attachment sites, respectively (6, 7). If this is also true for IntDOT, a monomer of IntDOT could create an intermolecular bridge between *attL* and *attR* within the excisive intasome. If IntDOT binds to an arm-type site within *attR* with its N terminus and also binds to a core-type site, such as D', within *attL* with its C terminus, it would explain why phasing mutations in the *attR* site did not have a strong effect on the excision frequency.

We know very little about the function of the accessory proteins Orf2c, Orf2d, and Exc, which are required for excision (4, 19). It is likely that these proteins bind to *attL* or *attR* to form intasomes required for excision. Binding of these proteins could also aid in controlling the directionality of the site-specific recombination reactions. Further experiments will be required to provide information on the structures of *attL* and *attR* excisive intasomes and on the regulation of the CTnDOT site-specific recombination system.

ACKNOWLEDGMENTS

We thank Phoebe Rice for the gift of the IHF overexpression strain and the purification protocol. We thank Yanping Wang for constructing the plasmid pT₇-int. We also thank Sumiko Yoneji and Vince Kwok for extensive technical assistance and Wilma Ross for helpful discussions.

This work was supported by grant NIH-GM-28717.

REFERENCES

1. Azaro, M. A., and A. Landy. 2002. λ Integrase and the λ Int family, p. 118–148. In N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, DC.
2. Cheng, Q., B. J. Paszkiet, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers. 2000. Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. *J. Bacteriol.* **182**:4035–4043.
3. Cheng, Q., N. Wesslund, N. B. Shoemaker, A. A. Salyers, and J. F. Gardner. 2002. Development of an in vitro integration assay for the *Bacteroides* conjugative transposon CTnDOT. *J. Bacteriol.* **184**:4829–4837.
4. DiChiara, J. M., A. A. Salyers, and J. F. Gardner. 2005. In vitro analysis of sequence requirements for the excision reaction of the *Bacteroides* conjugative transposon, CTnDOT. *Mol. Microbiol.* **56**:1035–1048.
5. Johnson, R. C., L. Johnson, J. W. Schmidt, and J. F. Gardner. 2004. Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome, p. 65–134. In N. P. Higgins (ed.), *The bacterial chromosome*. ASM Press, Washington, DC.
6. Kim, S., L. Moitoso de Vargas, S. E. Nunes-Duby, and A. Landy. 1990. Mapping of a higher order protein-DNA complex: two kinds of long-range interactions in lambda attL. *Cell* **63**:773–781.
7. Kim, S., and A. Landy. 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci attL and attR. *Science* **256**:198–203.
8. Lee, E. C., R. I. Gumpert, and J. F. Gardner. 1990. Genetic analysis of bacteriophage λ integrase interactions with arm-type attachment site sequences. *J. Bacteriol.* **172**:1529–1538.
9. Lu, F., and G. Churchward. 1994. Conjugative transposition: Tn916 integrase contains two independent DNA binding domains that recognize different DNA sequences. *EMBO J.* **13**:1541–1548.
10. Lynch, T. W., E. K. Read, A. N. Mattis, J. F. Gardner, and P. A. Rice. 2003. Integration host factor: putting a twist on protein-DNA recognition. *J. Mol. Biol.* **330**:493–502.
11. Malanowska, K., A. A. Salyers, and J. F. Gardner. 2006. Characterization of a conjugative transposon integrase, IntDOT. *Mol. Microbiol.* **60**:1228–1240.
12. Richet, E., P. Abcarian, and H. A. Nash. 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* **52**:9–17.
13. Ross, W., A. Landy, Y. Kikuchi, and H. A. Nash. 1979. Interaction of Int protein with specific sites on λ att DNA. *Cell* **18**:297–307.
14. Ross, W., and A. Landy. 1982. Bacteriophage lambda Int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. *Proc. Natl. Acad. Sci. USA* **79**:7724–7728.
15. Ross, W., and A. Landy. 1983. Patterns of lambda Int recognition in the regions of strand exchange. *Cell* **33**:261–272.
16. Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L.-Y. Li. 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev.* **59**:579–590.
17. Segall, A. M., S. D. Goodman, and H. A. Nash. 1994. Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J.* **13**:4536–4548.
18. Smith-Mungo, L., I. T. Chan, and A. Landy. 1994. Structure of the P22 att site. Conservation and divergence in the lambda motif or recombinogenic complexes. *J. Biol. Chem.* **269**:20798–20805.
19. Sutanto, Y., J. M. DiChiara, N. B. Shoemaker, J. F. Gardner, and A. A. Salyers. 2004. Factors required in vitro for excision of the *Bacteroides* conjugative transposon, CTnDOT. *Plasmid* **52**:119–130.
20. Van Duyne, G. D. 2002. A structural view of tyrosine recombinase site-specific recombination, p. 93–117. In N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, DC.