The *Bacteroides thetaiotaomicron* Protein Bacteroides Host Factor A Participates in Integration of the Integrative Conjugative Element CTnDOT into the Chromosome

Kenneth Ringwald, Jeffrey Gardner
Department of Microbiology, School of Molecular and Cellular Biology, University of Illinois at Urbana—Champaign, Urbana, Illinois, USA

**ABSTRACT**

CTnDOT is a conjugal 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 conjugal 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.

**IMPORTANCE**

*Bacteroides* species are normal members of the human colonic microbiota. These species can harbor and spread self-transmissible genetic elements (integrative conjugative elements [ICEs]) that contain antibiotic resistance genes. This work describes the role of a protein, BHFa, and its importance in the integration reaction required for the element CTnDOT to persist in *Bacteroides* host cells.

*Bacteroides* species are Gram-negative, obligate anaerobes that are part of the normal microbiota in the human colon (1). When the gut is punctured, *Bacteroides* can act as an opportunistic pathogen that may form abscesses in other regions of the body. Treatment of the abscesses is complicated by widespread resistance to tetracycline and erythromycin carried by integrative conjugal elements (ICEs) (also called conjugative transposons [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 (2). ICEs also carry genes to regulate and carry out their own transfer. Furthermore, some ICEs can mobilize coresident genetic elements that could not otherwise transfer (3). 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 (2).

CTnDOT is a well-characterized ICE found in *Bacteroides* species. It carries the *ermF* and *tetQ* genes that encode resistance 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 (4). IntDOT contains five of the six conserved amino acid residues that form the catalytic sites of tyrosine recombinases (5, 6). 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 *Bacteroides thetaiotaomicron* chromosome. During integration, IntDOT recombines the attDOT site in CTnDOT with an attB site in the bacterial chromosome to form the attL and 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 *Escherichia coli* chromosome. This requirement led to the original identification of IHF (7). 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, although there are no appropriately positioned IHF binding sites within attDOT (8, 9). Presumably, IHF binds CTnDOT DNA nonspecifically and bends the DNA into a favorable conformation for assembly of the inte-
some narrative to correct recombination (8). 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 Δ (BHFa). BHFa shares several conserved motifs with both E. coli HU and IHF, although the primary sequence is not similar to the sequences of those proteins. This is the first host factor identified for any of the ICEs in Bacteroides spp. BHFa binds specifically to four sites within the attDOT site. However, we found that other DNA binding proteins can substitute for BHFa 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: 100 µg/ml ampicillin, 50 µg/ml kanamycin, and 20 µg/ml gentamicin.

Growth of Bacteroides strains. All B. thetaiotaomicron cultures were grown anaerobically to an optical density at 650 nm (OD 650) of 0.8, as described previously (10). The cells in the cultures were then pelleted by centrifugation and frozen at −80°C. In all, 4 separate pellets (2 liters of culture) were combined for protein purification.

Purification of the Bacteroides host factor. 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 clarified by centrifugation. The resulting crude extract was fractionated by centrifugation and frozen at −80°C. In all, 4 separate pellets (2 liters of culture) were combined for protein purification.

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 to 20%)-SDS gel. The protein band that appeared in all the active fractions was roughly 10 kDa. The fractions remained active in the in vitro integration assay after heating at 70°C for 15 min and centrifugation at 4°C for 30 min. The 10-kDa band was excised from the gel, digested with trypsin, and analyzed by liquid chromatography-mass spectroscopy at the University of Illinois Protein Sciences Facility (University of Illinois Biotechnology Center). The best result (49% sequence coverage) was for the BT_1499 gene (NCBI GI:29345410) from the B. thetaiotaomicron VPI-5482 chromosomal sequence, annotated as encoding a putative HU-like protein. The secondary structure was predicted by using DSSP (11) and PsiPred (12, 13). We have named this gene bhfa, for Bacteroides host factor Δ.

In vitro recombination assay. The in vitro integration assay used was described previously (9, 14, 15). 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, HMBG1, and NHP6A; provided by Reid Johnson, University of California—Los Angeles). All proteins were diluted in IEF dilution buffer (50 mM Tris-Cl [pH 8], 10% glycerol, 2 mg/ml bovine serum albumin [BSA], and 200 mM KCl). The mixture was incubated for 2 h at 37°C, and the reactions were stopped by the addition of 5 µl of stop solution (30% glycerol, 10% SDS, 0.25% xylene cyanol, and bromophenol blue) to the mixture. The samples were then subjected to electrophoresis on a 1% agarose gel for 2 h at 140 V and then dried and analyzed as described previously (14).

Construction of 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 named Bhu2-F and Bhu2-R in Table S1 in the supplemental material. Both the PCR product and the PET-30 vector were digested sequentially with NdeI and 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 Δhfa strain in LB medium supplemented with kanamycin. The cultures were grown to an OD600 of 0.6 at 37°C, and the production of BHFa was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The cultures were then grown at 37°C for 4 h, and cells were pelleted by centrifugation. The resulting pellet was suspended in a buffer containing 50 mM NaHPO4 (pH 7.2), 1 mM EDTA, 50 mM NaCl, 5% glycerol, and 1 mM DTT, 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 ~95% pure and stored at −80°C.

Determination of molecular mass. 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 the manufacturer’s instructions. Blue dextran 2000 and the low-molecular-weight protein standards were resuspended in a solution containing 50 mM NaHPO4 (pH 7.2), 1 mM EDTA, 150 mM NaCl, 5% glycerol, and 1 mM DTT. Purified BHFa eluted with ~68 µl of elution buffer, which is consistent with a molecular mass of ~22 kDa and the predicted dimer molecular mass of 22.2 kDa.

Electrophoretic mobility shift assays. Fragments containing attDOT, attL, attR, and the BHFa binding sites (H1, H2, and H3-H4) or DNA containing a PCR fragment from pUC19 for the electrophoretic mobility shift assays (EMSA) was produced by PCR. The DNA was radiolabeled with [γ-32P]ATP by using poly nucleotide kinase (Fermentas). Primers for these fragments are listed in Table S1 in the supplemental material.

Reaction mixtures included BHFa, the DNA fragment (2.6 nM for attDOT- and pUC19-containing sites, 0.6 nM for attL- and attR-containing sites, and 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 (16). Gels were then dried, exposed to a phosphorimager screen, and scanned in the same way as that described above for the in vitro integration assay.

Footprinting. Footprinting reactions were performed as described previously (17). 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. A total of 1 pmol of DNA was used in each footprinting reaction mixture.

DNase I (Worthington Biochemical Corporation) was suspended in 5 ml of a solution containing 10 mM Tris-HCl (pH 7.5), 5 mM MgCl2, and...
BHFa Participates in Integration of CTnDOT

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 Tris- HCl (pH 7.4), and 25 μg/ml BSA. The total reaction volume was 20 μl. Incubations were carried out at room temperature for 30 min. 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 min, and the mixture was quenched by the addition of 1 volume of 0.5 M EDTA. The reaction mixtures were purified by using the Qiagen PCR purification kit. Each digest containing BHFa was tested in duplicate and compared against a control reaction mixture lacking BHFa to identify protected bases. After digestion and purification, samples were submitted to the University of Illinois core sequencing facility and analyzed by 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 as those used to generate the attDOT DNA for footprinting. Reaction mixtures were likewise submitted to the University of Illinois core sequencing facility and analyzed by 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 (9, 15). 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 nonspecifically to attDOT and bends the DNA to promote intasome formation so that IntDOT can carry out integration. Bacteroides species do not have a closely related IHF homolog, although there are several putative DNA binding proteins that are annotated as being related to HU based on sequence similarity within the B. thetaiotaomicron genome (18).

Column chromatography was used to fractionate B. thetaiotaomicron DNA binding proteins. The DNA binding proteins were purified from crude extracts of B. thetaiotaomicron strain 4001, which lacks any known Ices. The pellet wells from 2 liters of anaerobically grown 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 in vitro integration assay as described in Materials and Methods. The integration assay uses a supercoiled 3.5-kb plasmid containing the attDOT site and a linear 32P-labeled 67-bp attB site. Integration produces a 3,567-bp linear product containing the radioactively labeled DNA, which can be visualized on a 1% agarose gel (Fig. 1A). In this assay, a small amount of recombination was observed in reaction mixtures 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 ~1 M NaCl) that stimulated activity in the in vitro integration assay (fractions 41 and 43) (Fig. 1A). The active fractions were subjected to SDS gel electrophoresis as described in Materials and Methods. One band, of ~10 kDa, appeared to correlate with the integration activity (fractions 41 to 43) (Fig. 1B, open arrow). In addition, the active fractions were subjected to heat treatment, and the supernatant fractions were assayed for 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 GenBank accession number NC_004663), which is annotated as the DNA binding protein HU. Due to its role in CTnDOT integration, the BT_1499 gene has been designated bhfa, for Bacteroides host factor A.

The bhfa gene was amplified from the B. thetaiotaomicron genome and cloned into pET-30 as described in Materials and Methods. BHFa was then overexpressed in an E. coli strain lacking the ihfa gene and was purified to 95% purity by using heparin-agarose column chromatography followed by Superdex-75 size exclusion chromatography (Fig. 1C, closed arrow). The 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 E. coli protein. Based on elution from a Superdex-75 size exclusion chromatography column, BHFa has a molecular mass of ~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 (Fig. 2A).

Incubation of attDOT DNA with BHFa forms one complex at a concentration of 600 nM BHFa. Further dilution of BHFa led to the formation of 2 distinct complexes with similar mobilities as the BHFa concentration decreased to 300 nM (Fig. 2A). As BHFa was diluted to 5 nM, four shifts appeared, 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 depends on the relative location of the binding site within the DNA fragment (19–21). 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 different mobilities.

BHFa was further tested with an unrelated DNA sequence from plasmid pUC19. BHFa bound the nonspecific DNA sequences at a concentration of 500 nM but failed to form complexes at concentrations of <50 to 100 nM (Fig. 2B). A second complex was faintly visible at a concentration of 100 nM BHFa. The complexes observed at high concentrations of BHFa (600 nM in Fig. 2A and 500 nM in Fig. 2B) were similar between both the attDOT and pUC19 DNAs. However, as BHFa was diluted, the initial complex and the faint secondary complex with pUC19 DNA disappeared. Therefore, we conclude that the complex formed with pUC19 at 500 nM BHFa is nonspecific and that BHFa binding to attDOT DNA is specific (Fig. 2A and B).

When BHFa was incubated with attR DNA, only one shift was observed across the range of concentrations from 500 nM to 5 nM,
which suggests that there is only one binding site present (Fig. 2C). Over the same range, BHFα binding to attL showed a pattern of two distinct shifts similar to that of the full attDOT fragment, suggesting that the attL site contains multiple BHFα binding sites (Fig. 2D).

**Complexes with BHFα and IntDOT.** Binding of both IntDOT and BHFα 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, BHFα alone, or both BHFα and
IntDOT (Fig. 2E). BHFa was used at a constant concentration of 5 nM, where it formed two complexes (Fig. 2E, lane 2). At a concentration of 175 nM, IntDOT alone shifted a small amount of attDOT DNA (Fig. 2E, filled arrow), while some DNA remained in the wells (Fig. 2E, lane 9). When the concentration of IntDOT varied (110 to 430 nM) and the concentration of BHFa remained constant (5 nM), two supershifts were observed (Fig. 2E, lanes 3 to 8). The higher band was stronger and remained as IntDOT was diluted (Fig. 2E, open arrow). 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 attDOT DNA. Further dilution of IntDOT caused the lower supershift to disappear, while the two BHFa-associated shifts reappeared (Fig. 2E, lanes 4 to 8). The supershifts formed with both proteins were distinct from the complexes formed with either the IntDOT or BHFa protein on its own.
(Fig. 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 multiple IntDOT monomers may be bound to the same bent attDOT DNA.

Identification of the binding sites of BHFa. Since the binding of BHFa appears to be specific, we wanted to identify the binding sites. The attDOT site was labeled with the fluorescent dye 6-FAM and PCR amplified as described in Materials and Methods. The sequence of the attDOT site used for these protection studies is shown in Fig. 3A, along with the core and arm-type sites that are bound by IntDOT. The resulting chromatograms showed a pattern of protection over three regions of the DNA on the bottom strand, designated H1, H2, and H3-H4 (Fig. 3B). The corresponding regions were also protected on the top strand, although the protection of H3-H4 was 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 bases −59 to −44 on the bottom strand (Fig. 3A). It also overlaps 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 bases +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 cleavage between bases +63 and +66 was observed. Enhanced cleavage between bases +90 and +92 was also observed, which may define the border between the two binding sites. The H3 and H4 sites showed some protection on the top strand, but the footprint was much less distinct than that on the bottom strand (data not shown). The H3 site overlaps the entire L1 arm-type site, and the H4 site overlaps 7 of the 10 bases of the L2 arm-type site.

At 545 nM BHFa, the entire attDOT region showed slightly decreased digestion compared to that with the no-protein control, which may be due to nonspecific binding of BHFa. As BHFa was serially diluted to 180 nM, the H1 and H3-H4 sites lost some protection (Fig. 3B). The H2 site was the last site to lose protection (at concentrations of <110 nM), and it appears that BHFa had the highest affinity for this site. In addition, we observed enhanced DNA cleavage outside the protected regions on attDOT (Fig. 3B, 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 occurred when the H1 site was occupied by BHFa.

EMSA 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 a comparison of relative binding affinities of BHFa to the sites.

A PCR product containing the H1 site extending from positions −101 to −6 was incubated with BHFa. Only one shift was observed, indicating that there was one binding site present (Fig. 4A and B). The PCR-generated DNA fragment containing the H2 site extends from positions −26 to +75 (Fig. 4C). Incubation with BHFa also showed a single shift, indicating that there was one BHFa site on the fragment (Fig. 4D). In agreement with the footprinting results, BHFa showed a higher affinity for the H2 site than for either the H1 or H3-H4 site (Fig. 4B, D, and F).

The PCR product containing the H3 and H4 sites includes positions 57 to 145 (Fig. 4E). When EMSAs were performed with DNA containing this region, two distinct shifts were observed (Fig. 4F). The highest shift, visible from 2,160 to 270 nM 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 (Fig. 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 (22). Based on complexes of IHF and HU proteins with DNA, we would expect that relatively few bases of each footprint are contacted directly by residues of BHFa, while the rest of the DNA is bent around the protein (23–27).

DNA bending proteins promote the in vitro integration reaction. Previous studies showed that E. coli IHF could substitute for the B. thetaiotaomicron host factor in in vitro integration assays. 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 an in vitro integration assay. We found that entirely unrelated DNA bending proteins can substitute for BHFa in the in vitro integration assay (Fig. 5). E. coli IHF is a heterodimeric nucleoid-associated protein that is closely related to IHF but binds nonspecifically (20). E. coli IHF can substitute for IHF in the in vitro integration assay (Fig. 5, lanes 3 to 5). E. coli Fis is also a nucleoid-associated protein but binds DNA more specifically than does IHF and plays a role in the excision of bacteriophage λ in E. coli. In Salmonella enterica serovar Typhimurium, Fis binds to DNA and interacts with the Hin recombinase in the Hin inversion system (28). Fis cannot substitute for BHFa or IHF in the integration assay (Fig. 6, lanes 6 to 8). HMGB1 is a eukaryotic protein (from the high-mobility-group family of proteins) that binds DNA nonspecifically and bends DNA. NHP6A is a related protein but was originally found in Saccharomyces cerevisiae (29). Although both these proteins are eukaryotic proteins and have structures that are very different from those of bacterial IHF and HU proteins, they can enable integration in the in vitro integration reaction (Fig. 5, lanes 9 to 11 and 12 to 14). These proteins have also been shown to substitute for IHF in other recombination systems (30).

DISCUSSION

In this study, we identified the protein BHFa, a Bacteroides host factor required for the integration and likely the excision of CTnDOT. The purification of BHFa was carried out by using a crude extract and column chromatography. Identification of BHFa was based on the activity of the purified fractions in an in vitro integration assay. However, it is possible that other proteins of B. thetaiotaomicron also enable the 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 the E. coli Fis, IHF, and HU proteins (31).

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 (31–35). The exact role of BHFa in chromosome mainte-

FIG 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 zero. The overlap region (GCCTAGT) includes bases –3 to +3. Bases that are upstream of the central bases have negative numbers, and downstream bases have positive numbers. 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 boldface type were protected from DNase I digestion in the presence of BHFa. H sites (H1, H2, and H3-H4) are marked with the labeled boxes. (B) Chromatograms of DNase I protection of the bottom strand of attDOT. Concentrations of BHFa are indicated on the left. The top panel shows a reaction mixture that contained no BHFa. Regions of protection are marked with black rectangles. Gray rectangles indicate intermediate protection, and empty rectangles indicate a loss of protection. Sites of enhanced DNA cleavages are marked with asterisks.
A

B

C

D

E

F

Table: BHFα (nM)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table: BHFβ (nM)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure: Diagrams A-F showing sequences and corresponding gel images.
nance 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 (Fig. 6A) (36–38). BHFa is 34% identical to the IHF alpha subunit and 32% identical to the IHF beta subunit. However, the IHF alpha and beta subunits are 33% identical to one another, so BHFa is as identical to each subunit as the subunits are to one another. In addition, the HU subunits show similar levels of identity to BHFa, at ~30% and 32%, respectively (percent identities are shown in Table S2 in the supplemental material).

The secondary structure is much more conserved for these proteins, and the predicted secondary structure of BHFa is similar to those of the *E. coli* proteins IHF and HU. The actual secondary structures of HU and IHF are shown below the protein sequences and are conserved despite primary sequence variations between both IHF and HU and their subunits (Fig. 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. This region includes a proline residue within the flexible arm region that contacts DNA and introduces a kink into the backbone (position 65) (Fig. 6A, open arrow) (23, 25, 26). 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 (39). Interestingly, an *E. coli* mutant IHF protein that lacks the conserved proline still binds DNA but shows decreased DNA binding specificity (40). 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.

Although BHFa has several differences from the well-studied IHF and HU proteins, these variations appear to be consistent among other predicted DNA binding proteins found in related *Bacteroides* species (Fig. 6B) (36, 37). Specifically, the different amino acids within the conserved arm region found in other IHF/HU superfamily proteins are consistent among *Bacteroides* spp. (residues 60 to 63) (Fig. 6B). Although not as strictly conserved as the proline at position 65 (position 65 in Fig. 6A), the preceding residues 60 to 62 (GRN) are also very well conserved among the superfamily. However, the *Bacteroides* DNA binding proteins all include the residues ARN instead (Fig. 6B, black arrows). Furthermore, the primary sequences of the remainder of the other *Bacteroides* species DNA binding proteins are very highly conserved, with only 6 positions showing variation. It is likely that these related proteins enable the recombination of CTnDOT and related elements into the host chromosomes of species related to *B. thetaiotaomicron*. This is supported by previously reported work showing that CTnDOT is capable of transfer and integration in a variety of *Bacteroides* species and close relatives such as *Prevotella* (2).

The occupancy of the different binding sites of BHFa found in

---

**FIG 4** Binding of BHFa to H site DNA fragments. (A) The 96-bp DNA sequence used for 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 boldface type and gray. (B) EMSA of DNA containing the H1 site. The DNA concentration was 2.3 nM in each reaction mixture. (C) The 102-bp DNA sequence used for 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 solid lines. Protected bases are in boldface type and gray. (D) EMSA of the DNA containing the H2 site. The DNA concentration was 2.3 nM in each reaction mixture. (D) The 103-bp DNA sequence containing the H3 and H4 sites used for EMSA. The L1 and L2 arm-type sites are marked with solid boxes. Protected bases are in boldface type and gray. (F) EMSA of the DNA containing the H3 and H4 sites on the same fragment of DNA. The DNA concentration was 2.3 nM in each reaction mixture.
the attDOT region depends on the concentration of BHFa. Since three of the four BHFa binding sites are adjacent to or overlap arm-type sites, this may lead to a differential availability of the arm-type sites for IntDOT binding. The H2 site is the only BHFa binding site that does not overlap any other known protein binding sites within the attDOT sequence. The lower-affinity sites (H1, H3, and H4) overlap 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 the R2 and R2’ arm-type sites, while the H3 site overlaps the L1 arm-type site, and the H4 site partially overlaps the L2 arm-type site. This overlap between the lower-affinity binding sites of BHFa and the arm-type sites of IntDOT suggests that nonspecific 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, 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 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 the 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 this family. It is a homodimer that binds specifically to 4 sites found in the attDOT region and varies at several conserved residues. However, specific binding to these sites does not appear to be required for CTnDOT integration in an in vitro integration assay. Unrelated proteins enable the integration reaction in vitro and may enable the 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 is a suitable 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 Bacteroides spp.

ACKNOWLEDGMENTS

We thank Abigail Salyers, Carolyn Keeton, Margaret Wood, Jennifer Laprise, Crystal Hopp, and Adam Kolakowski for suggestions and helpful comments. We thank Sumiko Yoneji for purified IntDOT. We also 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

7. Nash H, Robertson C. 1981. Purification and properties of the Esche-


