Structure-Function Analysis of *Escherichia coli* DNA Helicase I Reveals Non-overlapping Transesterase and Helicase Domains*

Devon R. Byrd‡§, Juliana K. Sampson‡, Heather M. Ragonese‡, and Steven W. Matson‡¶‖**

From the ‡Department of Biology, the ¶Curriculum in Genetics and Molecular Biology, and the ‖Program in Molecular and Cellular Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599-3280

TraI (DNA helicase I) is an *Escherichia coli* F plasmid-encoded protein required for bacterial conjugative DNA transfer. The protein is a sequence-specific DNA transesterase that provides the site- and strand-specific nick required to initiate DNA strand transfer and a 5′ to 3′ DNA helicase that unwinds the F plasmid to provide the single-stranded DNA that is transferred from donor to recipient. Sequence comparisons with other transesterases and helicases suggest that these activities reside in the N- and C-terminal regions of TraI, respectively. Computer-assisted secondary structure probability analysis identified a potential interdomain region spanning residues 304–309. Proteins encoded by segments of *traI*, whose N or C terminus either flanked or coincided with this region, were purified and assessed for catalytic activity. Amino acids 1–306 contain the transesterase activity, whereas amino acids 309–1504 contain the helicase activity. The C-terminal 252 amino acids of the 1756-amino acid TraI protein are not required for either helicase or transesterase activity. Protein and nucleic acid sequence similarity searches indicate that the occurrence of both transesterase- and helicase-associated motifs in a conjugative DNA transfer initiator protein is rare. Only two examples (other than R100 plasmid TraI) were found: R388 plasmid TrwC and R46 plasmid (pKM101) TraH, belonging to the IncW and IncN groups of broad host range conjugative plasmids, respectively. The most significant structural difference between these proteins and TraI is that TraI contains an additional region of ~650 residues between the transesterase domain and the helicase-associated motifs. This region is required for helicase activity.

Bacterial conjugation is the primary mechanism by which many plasmids and conjugative transposons are spread throughout a bacterial population. The process begins with the formation of a stable mating pair involving a donor cell that contains a conjugative plasmid (or transposon) and a recipient cell that lacks the plasmid. This establishes the close cell-cell contact required for physical transfer of single-stranded DNA (ssDNA)[1] from the donor to the recipient. A site- and strand-specific nick is then introduced in *oriT* (origin of transfer), and the DNA is unwound to provide ssDNA for transfer to the recipient. Upon entering the recipient cell, the transferred ssDNA is converted into double-stranded DNA by host enzymes and either circularized to form a plasmid or recombined into the recipient chromosome. This stabilizes the transferred DNA in the recipient and ensures the transfer of genetic traits (for review, see Ref. 1).

The enzymology of DNA strand transfer has been of interest since conjugation was first discovered over 50 years ago (2). In the last decade, it has become clear that transmissible plasmids encode a conjugative DNA transfer (CDT) initiator protein that plays a key role in initiating DNA strand transfer. These proteins nick their cognate supercoiled DNA substrate via a site- and strand-specific transesterification, resulting in a covalent protein-DNA intermediate (for reviews, see Refs. 3 and 4). A small minority of the characterized CDT initiator proteins also catalyze a helicase reaction (5–13). Because only a single strand of the duplex plasmid DNA is transferred to a recipient cell, unwinding of the plasmid is required to produce the transferred DNA strand. This can be accomplished by a DNA helicase or, perhaps, through strand displacement synthesis by a DNA polymerase. Recently, the helicase activity of the *Escherichia coli* F plasmid CDT initiator protein (TraI protein) has been shown to be essential for DNA transfer (14).

Based on the above description, known conjugative transesterases may be grouped into two classes: (i) those lacking an intrinsic helicase activity (*e.g.* RP4 plasmid TraI) (15) and (ii) those in which transesterase and helicase activities have been shown to reside within a single protein, as is the case only for the R388 plasmid TrwC protein and the F plasmid TraI protein (the F and R100 plasmid *traI* genes and deduced amino acid sequences share >95% identity and will be treated as alleles of the same gene in this report) (8, 10, 12). The overwhelming majority of described CDT initiator proteins fall into the first class based on sequence homologies (16)[2] and known biochemical properties. However, regardless of class, the CDT transesterases exhibit structural motifs and functional similarities that reflect conservation of the catalytic mechanism (4, 16, 17). The majority of these do not have helicase motifs. Conversely, the vast majority of available helicase sequences do not exhibit transesterase motifs. Where present in proteins exhibiting both functions, it is tempting to predict that the two activities will reside in independently folding and potentially separable domains.

Support for this idea has been provided by the work of Llosa *et al.* (10) using the plasmid R388-encoded TrwC protein. Like F plasmid-encoded TraI (F-TraI), TrwC exhibits both 5′ to 3′

[1] The abbreviations used are: ssDNA, single-stranded DNA; CDT, conjugative DNA transfer; F-TraI, F plasmid-encoded TraI; ATPγS, adenosine 5′-*O*-(thiotriphosphate).

[2] D. R. Byrd, unpublished data.

Name	Sequence (5′ to 3′)	Description
St30	CTTGTTTTTGCGTGGGGTGTGGTGCTTTTG	F *oriT nic* site 30-mer
077	GCGGGTACCCATGGGCCCGCCGGAAAATGGTGTGATTGAACG	5′-PCR primer for 348C intein fusion construct
078	GCGGGTACCCATGGCTTCACAGGACGGGCCGGATGTG	5′-PCR primer for 309C intein fusion construct
079	GCGGGTACCCATGGGTGTACCGGTGGAGGCCTTTTCC	5′-PCR primer for 226C intein fusion construct
080	GTCTCCACCCAGGGTTTTCTCTTTCTGC	3′-PCR primer for *traI* helicase-intein fusions
081	GGTCTAGAGTGGGATTGATGCCGGGATATG	5′-PCR primer for unique *Xba*I site upstream of *traI* Shine-Dalgarno sequence
085	CATGCTCTTCCGCACCCCATGGGCGTCAGCGTGCGGAGATC	3′-PCR primer for N306 construct
086	GGTGGTCCCGGGGTAGAGCCTGCCAAAGG	3′-PCR primer for N235 construct
087	GGTGGTCCCGGGGTAGAGCCTGCCAAAGG	3′-PCR primer for N200 construct
088	TTTTTCATATGATGAGTATTGCGC	5′-PCR primer for TraIΔ252 construct
089	TTTGGATCCTCACTACGCCTGACGGAGAACAGC	3′-PCR primer for TraIΔ252 construct

DNA helicase and R388 *oriT*-specific transesterase activities in a single polypeptide chain (7, 18). Using recombinant DNA techniques, these investigators showed that the transesterase and helicase activities of TrwC could be separated into two overlapping segments of the protein. Coexpression of active overlapping segments of TrwC in a *trwC* mutant background resulted in poor functional complementation compared with an allele expressing native TrwC (10).

F-TraI is the 1756-amino acid product of the *traI* gene (19) and is essential for conjugation (20). Also known as *E. coli* DNA helicase I, TraI was initially purified based on its DNA-stimulated ATPase and DNA helicase activities (21, 22). TraI-catalyzed unwinding of duplex DNA was subsequently shown to occur with a 5′ to 3′ polarity (6, 23). The idea that TraI is the helicase that unwinds the F plasmid during conjugative transfer was suggested when the gene encoding helicase I was mapped to the *traI* gene on the F plasmid (19). This has now been shown to be the case (14). The transesterase activity of TraI was not discovered and characterized until sometime later (5, 8, 9, 24).

TraI and TrwC exhibit a significant degree of amino acid sequence similarity that includes both the transesterase and helicase motifs (4, 18). It is currently thought that the transesterase acts in initiation complex formation, whereas the helicase activity is involved in the subsequent unwinding stage of DNA transfer. Computer-assisted sequence analysis of TraI suggested the possibility of an interdomain segment spanning residues 304–309. Biochemical characterization of purified proteins encoded by segments of *traI* terminating at this point revealed the transesterase and helicase activities of F-TraI to reside in non-overlapping, physically separable domains.

EXPERIMENTAL PROCEDURES
Materials

Bacterial Strains—Bacteria were grown in LB medium (25) supplemented with 1.5% agar for plates. The medium was supplemented with antibiotics, as appropriate, at the following concentrations: ampicillin, 100 μg/ml; tetracycline, 12 μg/ml; and chloramphenicol, 20 μg/ml (in methanol). The donor strain for CDT complementation studies was a derivative of DBH10B (Invitrogen). A NaCl-inducible allele of phage T7 gene 1 (T7 RNA polymerase) was moved into DBH10B by P1 transduction. Several transductants were assessed for the presence of appropriate genetic markers and the ability to express T7 RNA polymerase upon the addition of 0.3 M NaCl to the growth medium. A representative isolate was designated DB51. The pOX38TΔ*traI* plasmid (14) was electroporated into DB51, and a representative isolate was designated DB52. DB52 was transformed with the appropriate complementation plasmids (see below) in single- and double-plasmid transformations. DB52-N contains only the plasmid expressing the N306 allele of *traI*; DB52-C contains only the plasmid expressing the 309C allele of *traI*; and DB52-NC contains both complementation plasmids.

DNA, Nucleotides, and Enzymes—DNA oligonucleotides are listed in Table I. Nucleic acids were quantified by spectroscopy at 260 nm. Unlabeled nucleoside 5′-triphosphates were from U. S. Biochemical Corp. [³²P]ATP and [³²P]dCTP were from Amersham Biosciences. Restriction enzymes, DNA polymerase I (large fragment), and Vent DNA polymerase were from New England Biolabs Inc. and were used as specified by the supplier. Phage T4 DNA ligase was from Roche Molecular Biochemicals.

Plasmid Constructions—Standard cloning techniques were employed essentially as described (25). With the exception of the fragment encoding N365, segments of *traI* were generated by PCR using Vent polymerase and pMP8 (26) as the template. Constructs were sequenced to ensure the absence of PCR-derived mutations.

The IMPACT protein purification system (New England Biolabs Inc.) was used to express and purify the segments of TraI used in this study. The vector pCYB3 was modified by replacing the *lac* promoter and sequences upstream of the *Nco*I site with the T7 RNA polymerase promoter and ribosome-binding site from pET11d (Novagen) by fragment exchange. PCR-generated segments of *traI* were cloned into the *Nco*I and *Sma*I sites of the modified vector. This results in the addition of a glycine (GGG) codon to the extreme 3′-end of the *traI* allele. The proteins were purified from HMS174(DE3) (Novagen) according to the manufacturer's instructions. Protein concentrations were determined using the Bradford protein assay (Bio-Rad) with bovine serum albumin as the standard.

N365 was generated by digestion of pET11d-traI with *Cla*I and *Hin*dIII, fill-in of the 5′-overhangs with *E. coli* DNA polymerase I (large fragment), gel purification, and self-ligation. These manipulations result in a segment of TraI with wild-type sequence up to residue 361, a conservative Asp-to-Glu substitution at position 362, and the non-native sequence LPM at position 363–365, followed by a TGA stop codon. The protein was purified essentially as described for TraI (6).

N200 and N235 were constructed using PCR to amplify the appropriate TraI segment from the N306 construct. The PCR primers used for these amplifications are listed in Table I. The amplified DNA fragments were digested to completion with *Nde*I and *Sma*I and cloned into pTYB2 (New England Biolabs Inc.). Both proteins were expressed in *E. coli* BL21(DE3) (Novagen) and purified according to the manufacturer's instructions (IMPACT system, New England Biolabs Inc.).

TraIΔ252 was constructed using PCR to amplify the *traI* gene on pET11d-traI (14). The downstream PCR primer contained two stop codons immediately following codon 1504 in the *traI* sequence to produce this C-terminal truncation of the *traI* gene. The protein was expressed in *E. coli* BL21(DE3) cells and purified using the standard TraI purification.

The plasmids used in CDT complementation studies were constructed as follows. The pBR322-derived replication origin present in pET11d-N306 was replaced with the pACYC184 origin by fragment exchange to generate an N306 expression plasmid with a p15A replication origin. The allele of *traI* encoding 309C was subcloned from the intein fusion expression plasmid into a pUC-derived plasmid encoding resistance to chloramphenicol. The resulting plasmid retained the phage T7 gene 10 transcription/translation signals to drive expression of the gene. Expression of the appropriate functional domain of TraI was verified for each of the constructs in the appropriate donor strain.

Methods

DNA Helicase Assays—A partial duplex unwinding substrate was made essentially as described (27). Briefly, a 91-nucleotide oligonucleotide was annealed to its complementary sequence on purified M13mp6 ssDNA at a molar ratio of 1:1. The 3′-end of the annealed oligonucleotide was extended using *E. coli* DNA polymerase I (large fragment) and [α-³²P]dCTP. The final length of the oligonucleotide was 93 nucleotides. The preparation was phenol/chloroform-extracted and passed over a

Bio-Gel A-5m column. The void volume fractions were pooled, ethanol-precipitated, and suspended in 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA to a final concentration of ~5 fmol/μl (DNA phosphate).

DNA unwinding reaction mixtures (typically 20 μl) contained 25 mM Tris-HCl (pH 7.5), 20 mM NaCl, 3 mM MgCl$_2$, 5 mM β-mercaptoethanol, 10 fmol of DNA substrate, and 2 mM ATP. Reaction mixtures were assembled at room temperature, and the reaction was initiated by the addition of enzyme. Incubations were carried out at 37 °C for 10 min. The reaction mixtures were quenched by the addition of EDTA to 25 mM and SDS to 1%. Reaction products were resolved on an 8% native polyacrylamide gel (20:1 cross-linking) and quantified using a Molecular Dynamics PhosphorImager.

ATPase Assays—Reaction mixtures (30 μl) were essentially identical to those used for helicase assays with the following exceptions. [γ-^{32}P]ATP (15 pmol) was added to each reaction mixture (final ATP concentration of 2 mM), which contained 0.75 μg of M13mp6 ssDNA instead of the partial duplex helicase substrate. Reactions were assembled at room temperature and initiated by the addition of enzyme, and incubation was carried out at 37 °C. Aliquots (5 μl) were removed at 1-min (TraI) or 2-min (all truncation mutants) intervals and quenched by the addition of 5 μl of 50 mM EDTA, 7 mM ADP, and 7 mM ATP. 5 μl was spotted onto a polyethyleneimine-cellulose TLC plate (J. T. Baker, Inc., Phillipsburg, NJ) and allowed to dry, and the plates were developed with a mobile phase consisting of 1.0 M formic acid and 0.8 M LiCl. The plates were allowed to dry, and the degree of ATP hydrolysis was quantified using phosphor-storage technology.

Duplex DNA Relaxation Assays—Assays were done in a manner slightly modified from that previously described (8). In addition to the protein (present at the indicated concentrations), a typical reaction mixture (16 μl) contained 7 nM supercoiled pBSoriT (or pBS) DNA, 40 mM Tris-HCl (pH, 7.5), 6 mM MgCl$_2$, and 15% glycerol. Reactions were assembled at room temperature and incubated at 37 °C for 20 min. Reactions were stopped by the addition of proteinase K (Roche Molecular Biochemicals) and SDS to final concentrations of 1 mg/ml and 0.25%, respectively, and allowed to incubate at 37 °C for an additional 20 min. The products were resolved on 0.8% agarose gels and visualized by ethidium bromide staining (0.5 μg/ml).

DNA Binding Assays—DNA binding assays using 226C, 309C, and 348C utilized the double-filter (nitrocellulose + DE81) technique previously described (38). Protein concentration was varied as indicated, and a 93-bp partial duplex DNA substrate was used as the ligand. Experiments were repeated three or four times, and the data were averaged and fit to a rectangular hyperbola to obtain an apparent K_d for DNA binding. The reaction conditions used were the same as described above for the DNA helicase reactions, except that ATPγS was substituted for ATP at a final concentration of 1 mM.

Gel retardation assays were used to measure the binding of N306 and N235 to a ssDNA oligonucleotide containing the relaxase recognition sequence (see Table I for sequence). The reaction conditions were those used for relaxation assays without the SDS/proteinase K incubation. After a 20-min incubation at room temperature, the binding reaction mixtures were loaded onto a 5% polyacrylamide and 0.125% bisacrylamide gel, and electrophoresis was performed at 200 V for 2 h at 4 °C. The running buffer was 2× Tris/glycine (50 mM Tris, 380 mM glycine, and 2 mM EDTA (pH 8.3)). Gels were visualized using a Molecular Dynamics PhosphorImager.

Genetic Assays—The liquid mating assay protocol was carried out as previously described (14). Briefly, DB52-N, DB52-C, and DB52-NC were used as donor strains; DH5α was used as the recipient strain. Donor and recipient strains were diluted 1:50 into LB medium from saturated overnight cultures grown under antibiotic selection and allowed to grow to mid- or late-log phase in the absence of selection at 37 °C. Donors and recipients were then mixed at a volume ratio of one donor to nine recipients and incubated at 37 °C. After 5 min, the mating mixtures were diluted 1:10 into LB medium and incubated at 37 °C for an additional 30 min. The mating mixtures were then vigorously vortexed to disrupt mating pairs, and 10-fold serial dilutions were prepared in phosphate-buffered saline. Appropriate dilutions were plated onto LB plates containing streptomycin and tetracycline to counterselect donors and unmated recipients while selecting for transconjugants. Aliquots of the unmated donor and recipient cultures were subjected to 10-fold serial dilution and plated onto LB plates containing the appropriate antibiotics to determine viable donor cell count and viable recipient cell count. Mating frequency was calculated as the number of transconjugants/100 viable donor cells.

Data Base Searching—A degenerate amino acid sequence (YY$X_{1,2}$(D/E)$X_{1,2}$(D/E)$X_{1,2}$YY) was used to search the Swiss Protein and TREMBL Sequence Databases for proteins encoded by bacterial conjugative plasmids with the N-terminal two-tyrosine doublet motif shared by TraI and TrwC.[3] Hits were manually scanned for the presence and position of the known transesterase and helicase motifs (16, 28, 29).

Computer-assisted Sequence Analysis—F-TraI, plasmid R388 TrwC, and plasmid R46 TraH were aligned in a binary fashion using the SIM algorithm (30). The gap opening and extension penalties were 10 and 2, respectively, and the comparison matrix was PAM 200. The amino acid sequence of TraI spanning residues 281–380 was submitted to the PSA server at the BioMolecular Engineering Research Center for secondary structure analysis.[4]

RESULTS

To begin an investigation of the functional domains of the F plasmid-encoded TraI protein, the sequence of F-TraI was compared with that of two related proteins, TrwC and TraH. This analysis was aided by a previous study (10) that identified the functional transesterase and helicase domains of TrwC in two separate and overlapping protein segments. The region of overlap (amino acids 192–348) brackets the point at which the N-terminal sequence similarity between TrwC and TraI falls off sharply (Fig. 1). We speculated that an interdomain segment in TraI would fall within this region, if present at all. The 200-amino acid sequence of TraI that spans this region was subjected to computer-assisted secondary structure analysis as described under "Experimental Procedures." The output of this analysis (Fig. 1) was inspected for regions that had a reasonable probability of being able to form a flexible linker. Constrained structures such as helices and sheets were excluded from consideration in favor of segments more likely to adopt an unconstrained loop/turn conformation. The TraI sequence in this region with the highest predicted probability of a loop/turn structure was LTPGPA (residues 304–309). These coordinates are nearly coincident with the end of significant sequence identity (as opposed to similarity) between TraI and TrwC. The sequence LTPGPA is not present in TrwC. Two additional regions, on either side of residues 304–309, with slightly lower probabilities of similar conformation were identified at positions 227–230 and 346–352.

Using these results as a guide, fragments of the *traI* gene were cloned, overexpressed, purified, and assessed for biochemical activity. Fig. 2 shows several of the purified proteins used in this study (full-length TraI, two N-terminal segments (N306 and N365), and three C-terminal segments (226C, 309C, and 348C)) resolved on an SDS-polyacrylamide gel. Purified proteins not shown in this figure (N235 and TraIΔ252) were of comparable purity. The names of the protein segments indicate the first native amino acid in the case of 226C, 309C, and 348C and the fact that the protein extends to the native C terminus. N235, N306, and N365 begin at the native N terminus and extend to the indicated native residue. Deviations from native amino acid sequence, where they exist, are described under "Experimental Procedures." TraIΔ252 begins at the native N terminus and extends to residue 1504 of the native protein located ~60 residues C-terminal to helicase-associated motif VI. Thus, this protein lacks the C-terminal 252 residues of TraI, but retains both the putative transesterase and helicase domains. The constructions, shown schematically, and the data derived from *in vitro* biochemical analysis of various segments of TraI are summarized in Fig. 3.

DNA Helicase Domain—The 226C, 309C, and 348C proteins were the logical candidates for an active helicase because the helicase-associated motifs are located within this portion of the protein. Each protein was purified to apparent homogeneity from an expression strain (see Fig. 2), and the purified protein was assayed for its ability to catalyze a helicase reaction using

---

[3] Available at www.expasy.ch/sprot/scnpsit2.html.

[4] Available at bmerc-www.bu.edu/psa/.

FIG. 1. **Schematic similarity comparison of TraI and TrwC and assignment of predicted interdomain segments.** A schematic comparison of full-length F plasmid TraI and R388 plasmid TrwC is depicted with approximate amino acid coordinates indicated. The active N348 and 192C segments of TrwC identified by Llosa *et al.* (10) are shown above the full-length schematic of that protein. The *bracket* indicates the approximate extent of overlapping amino acid sequence. The *black boxes* reflect the ~40% identity between the TraI and TrwC N-terminal regions and include the transesterase region of each protein. The *striped boxes* indicate the region of C-terminal similarity containing the helicase motifs (28). TraI and TrwC were aligned using the SIM algorithm (30). For this analysis, the gap opening and extension penalties were 10 and 3, respectively, and the comparison matrix was BLOSUM 100. The secondary structure probability contour plot generated for residues 210–400 of TraI (see "Experimental Procedures") is also shown. TraI sequences that correspond to high loop/turn probabilities are indicated at the bottom. The *numbers* correspond to the first and last residues of each string. *aa*, amino acids.
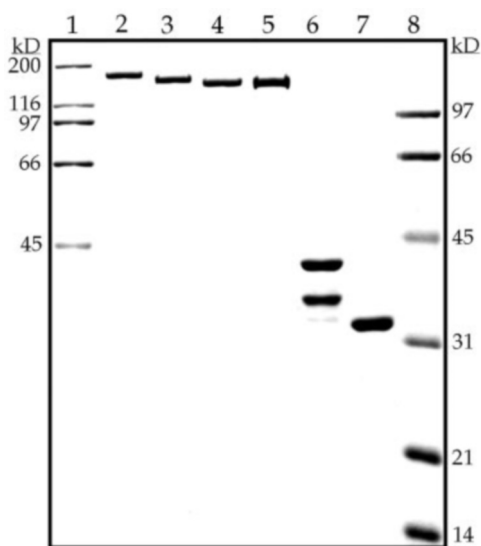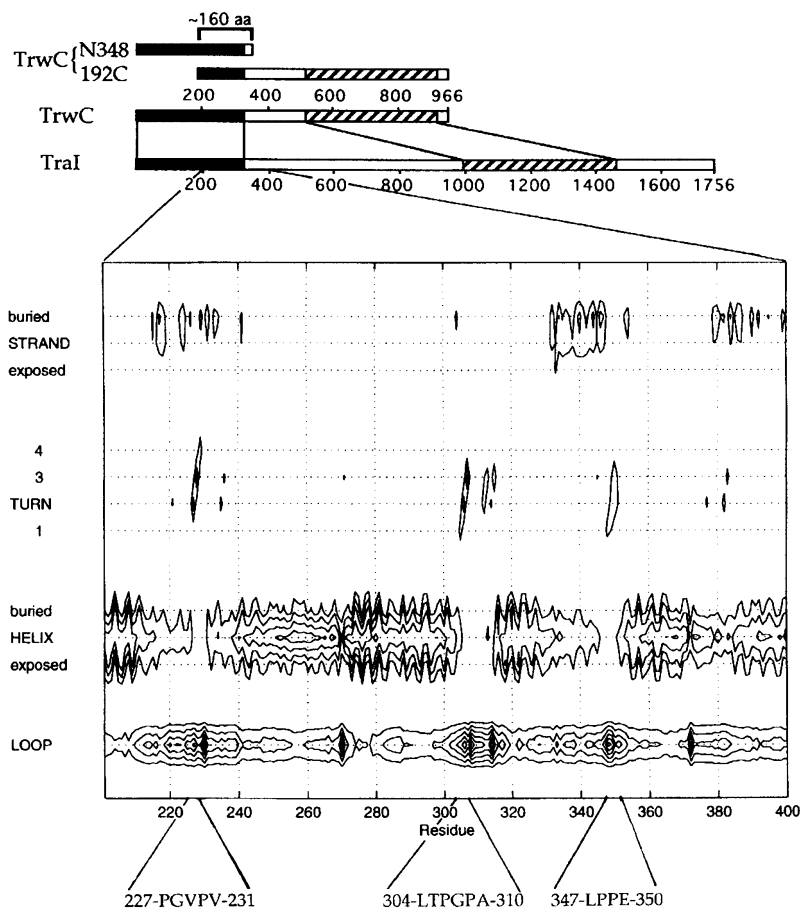

FIG. 2. **SDS-PAGE of purified TraI protein fragments.** Approximately 2 μg of each protein was resolved on a 9.6% SDS-polyacrylamide gel and stained with Coomassie Brilliant Blue. *Lane 1*, high range molecular mass standards (Bio-Rad) with sizes (in kilodaltons) indicated on the left; *lane 2*, native TraI; *lane 3*, 226C; *lane 4*, 309C; *lane 5*, 348C; *lane 6*, N365; *lane 7*, N306; *lane 8*, low range molecular mass standards (Bio-Rad) with sizes indicated on the right.

a 93-bp partial duplex substrate. Both 226C and 309C catalyzed unwinding of the 93-bp partial duplex substrate (Fig. 4*A*). The specific activities of 226C and 309C were 70 and 65%, respectively, that of the native protein based on the slope of the linear portion of the titration curve. On the other hand, the 348C protein was incapable of catalyzing detectable unwinding

despite the fact that it is only 39 residues shorter than 309C. Additional studies have shown that 226C and 309C also catalyze unwinding of long (851-bp) partial duplex substrates (Fig. 4*B*), albeit somewhat less efficiently than native TraI at low protein concentrations. The specific activities of 226C and 309C were 50 and 45%, respectively, that of the native protein determined as described above. However, the extent of the reaction at high protein concentrations was essentially identical (within experimental error) to that of the native protein. As expected, 348C was not able to catalyze unwinding the 851-bp partial duplex substrate.

TraIΔ252, lacking the C-terminal 252 residues, was also purified and analyzed in DNA helicase assays (data not shown). This protein was active as a DNA helicase, exhibiting a specific activity that was >50% that of the native protein. Thus, the C-terminal 252 residues, which lie just to the C-terminal side of the helicase-associated motifs, are not essential for the helicase activity of TraI.

Unwinding of a duplex nucleic acid substrate by a helicase is dependent upon nucleoside 5′-triphosphate hydrolysis. Therefore, we examined the DNA-stimulated ATP hydrolysis reaction catalyzed by 226C, 309C, 348C, and TraIΔ252 and compared this with the reaction catalyzed by the native protein. The k_{cat} for the native protein was 108 s^{-1} in the presence of ssDNA. Remarkably, the k_{cat} values for 226C, 309C, and TraIΔ252 were reduced to 9, 10, and 14 s^{-1}, respectively, although all three proteins exhibited helicase activity similar to that of the native protein (see Fig. 4). The 348C fragment of TraI exhibited a more profound defect in ATPase activity (k_{cat} < 1 s^{-1}). The significant defect in ATP hydrolysis exhibited by 348C is apparently sufficient to explain the lack of helicase activity and demonstrates that the N terminus of the functional helicase/ATPase domain of TraI lies within the 39
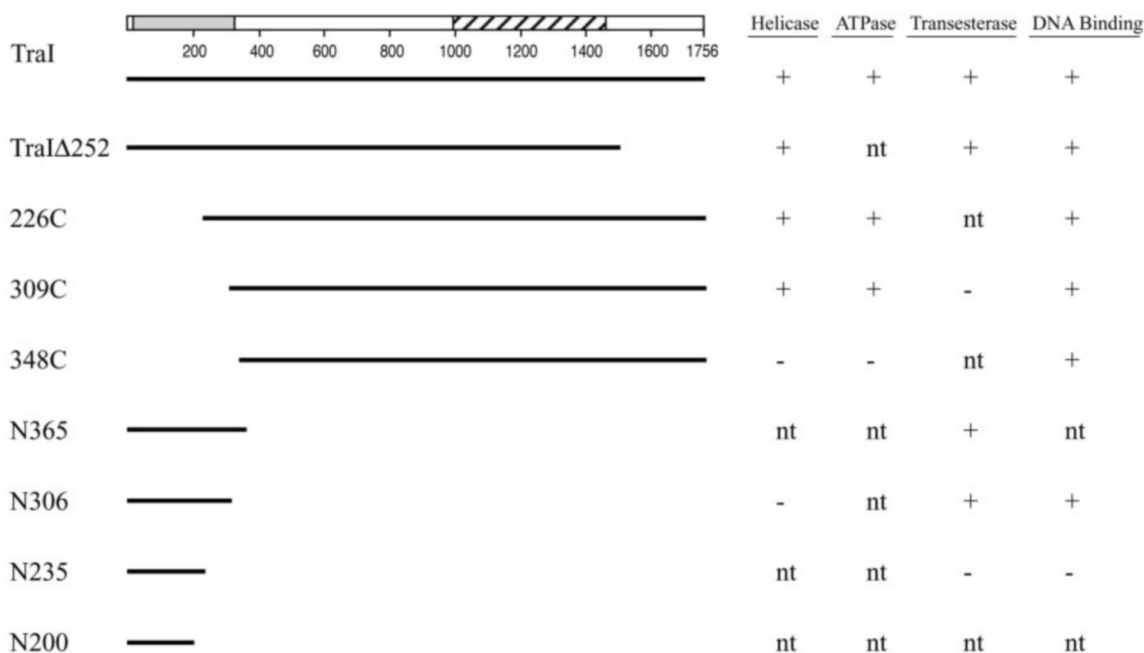
FIG. 3. **TraI segments and data summary.** The *solid horizontal lines* represent the segments of TraI that were purified and analyzed using biochemical assays. The names of the protein segments indicate the first native amino acid in the case of 226C, 309C, and 348C or the last native amino acid in the case of N200, N235, N306, and N365. TraIΔ252 is missing the last 252 amino acids of the native protein. The results are summarized on the right: *Helicase*, ability to unwind a 93-bp partial duplex substrate; *ATPase*, ability to catalyze ssDNA-stimulated hydrolysis of ATP; *Transesterase*, ability to catalyze conversion of supercoiled DNA to open circular DNA using pBSoriT as the DNA substrate; *DNA Binding*, ability to bind a 93-bp partial duplex DNA molecule (TraI, TraIΔ252, 226C, 309C, and 348C) or a ssDNA oligonucleotide containing the *nic* site (N306 and N235); +, positive result; −, negative result; *nt*, not tested.

amino acids that distinguish 309C from 348C.

We also measured the binding of 226C, 309C, and 348C to a partial duplex DNA ligand as described under "Experimental Procedures" (Fig. 5). Both 226C and 309C bound the DNA ligand with an apparent K_d of ~10 nM. The K_d measured for the 348C fragment was ~90 nM, indicating a defect in DNA binding. Although this binding defect is not likely to be sufficient to explain the lack of helicase activity, it does indicate that the region between residues 309 and 348 is important for DNA binding.

DNA Transesterase Domain—Three N-terminal segments of TraI were expressed and purified as outlined under "Experimental Procedures" (see Fig. 2). The C-terminal end of N365 was chosen based upon available restriction sites and not with regard to potential secondary structure. However, in conjunction with the largest helicase segment, 226C, these two proteins bracket the predicted interdomain region at residues 304–309. The overlap shared by these two segments of TraI is 139 amino acids, which is roughly comparable to that of the smallest functional segments of TrwC used by Llosa *et al.* (10).

Purification of N365 yielded both the expected protein and a significant proteolytic breakdown product that copurified with N365 (see Fig. 2, *lane 6*). Western blot analysis using polyclonal antibodies directed against TraI indicated that this proteolytic fragment was a fragment of TraI (data not shown). Incubation of this protein preparation with a 3′-end-labeled oligonucleotide whose sequence encompassed the F plasmid *oriT nic* site resulted in transfer of the labeled DNA to both protein species (data not shown). Thus, both proteins contain the active tyrosine present in the transesterase domain of TraI. Moreover, because the active-site tyrosine in TraI is within 23 residues (~2.4 kDa) of the N terminus of the full-length protein,[5] the proteolytic cleavage event, estimated at ~5 kDa or 40–50 amino acids, must remove the C-terminal end of N365. If the

cleavage event removed the N-terminal end of the protein, it would remove the transesterase active site, and this is not the case. The smaller active product appears to be comparatively stable to proteolysis as judged by the absence of faster migrating species. Thus, folding of the N-terminal region of TraI into a stable transesterase domain is independent of the rest of the protein.

Based upon the fact that the N terminus of the functional helicase (residue 309) coincided with the suspected interdomain region, N306 was constructed and purified (see Figs. 1 and 2). Both N306 and N365 were competent to nick a supercoiled *oriT*-containing DNA substrate (pBSoriT) (Fig. 6). In this semiquantitative analysis, the smaller transesterase domains exhibited slightly lower specific transesterase activity compared with native TraI. The presence of *oriT* in the substrate was required to observe transesterase activity; thus, conversion of the DNA from the supercoiled to the open circular form is not due to nonspecific cleavage (Fig. 6). N365 and N306 did not display detectably different activities from each other when present at equivalent concentrations (Fig. 6, compare *lanes 3* and *4* with *lanes 5* and *6*). These results demonstrate the *oriT*-specific DNA transesterase activity of TraI resides in the N-terminal 306 residues of the 1756-amino acid native protein.

To define the C-terminal end of the transesterase domain, two additional N-terminal fragments of TraI were expressed. N200 encompassed the first 200 amino acids of TraI and is homologous to the first 200 amino acids of TrwC, which has been shown to catalyze sequence-specific transesterification using an oligonucleotide substrate (10). N200 was completely insoluble and could not be analyzed further. The construction of N235 was based on limited proteolysis of N306, which suggested the presence of a stable domain encompassing the first ~235 amino acids of TraI.[6] N235 was expressed, purified, and
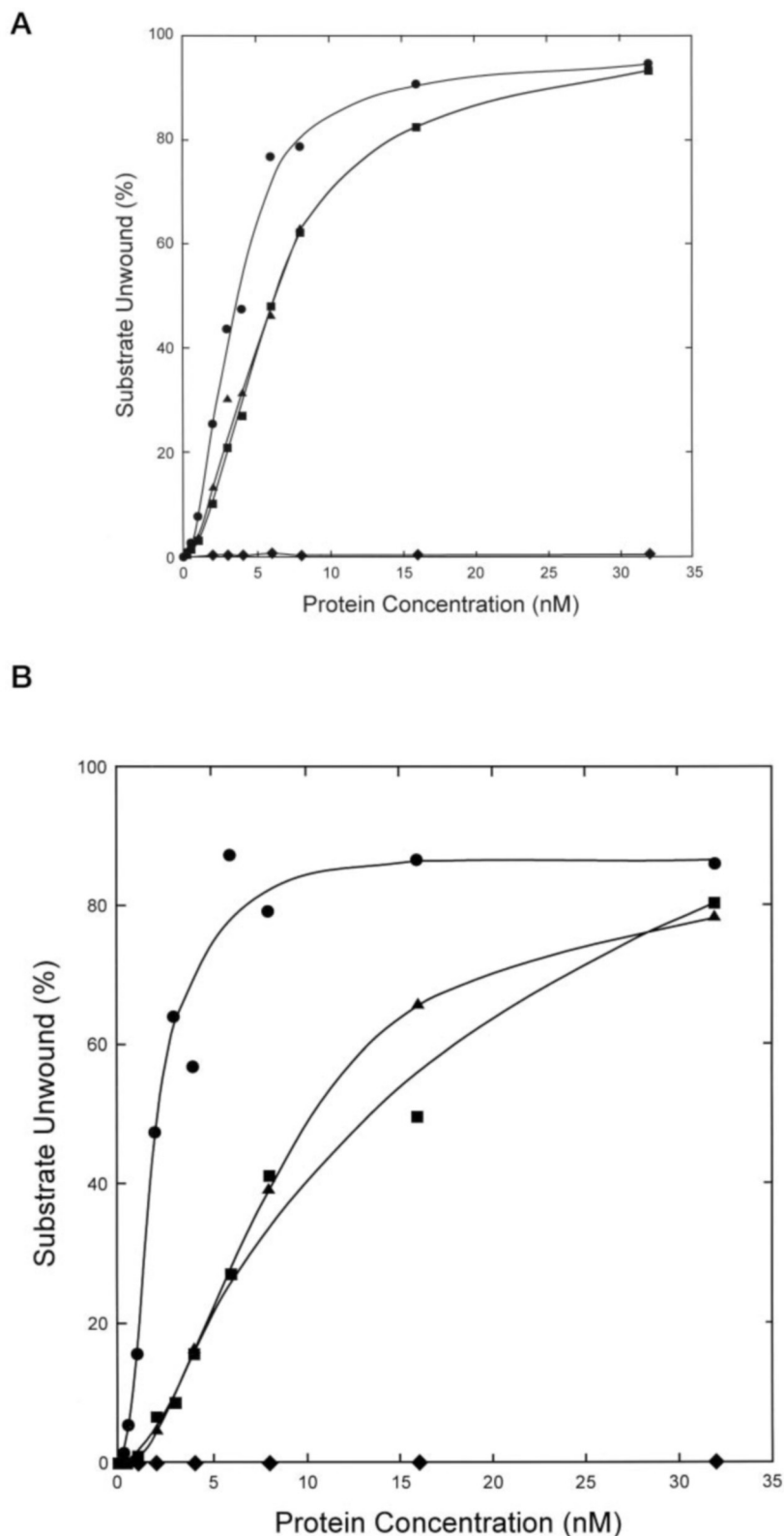
A

B

FIG. 4. **Helicase activity assays of TraI, 226C, 309C, and 348C.** *A*, helicase activity assays using either native TraI or the purified segments of TraI and the 93-bp partial duplex substrate were carried out as described under "Experimental Procedures" using the indicated amounts of each protein. ●, native TraI protein; ■, 226C; ▲, 309C; ◆, 348C. *B*, helicase activity assays using either native TraI or the purified segments of TraI and the 851-bp partial duplex substrate were carried out as described under "Experimental Procedures" using the indicated amounts of each protein. ●, native TraI protein; ■, 226C; ▲, 309C; ◆, 348C. The data represent the means of three to four determinations. S.D. values were omitted for clarity. In general, the S.D. was <10% of the mean.

analyzed for transesterase activity. The purified protein failed to catalyze sequence-specific transesterification and failed to bind a ssDNA oligonucleotide that contained the *nic* sequence (data not shown). On the other hand, N306 bound this oligonucleotide with high affinity as demonstrated using gel retardation assays. To ensure that N235 was properly folded, the secondary structure of the purified protein was examined by circular dichroism spectroscopy. The purified protein exhibited primarily α-helical structure, and its circular dichroism spectrum was comparable to that of N306.[6] Therefore, N235 appeared to be properly folded, but unable to bind a ssDNA oligonucleotide that contained *nic*. Thus, amino acids required for binding of the transesterase domain to its substrate are present within the 70-amino acid segment between residues 235 and 306. It is also clear from this analysis that the N-terminal 306 amino acids represent a minimal functional
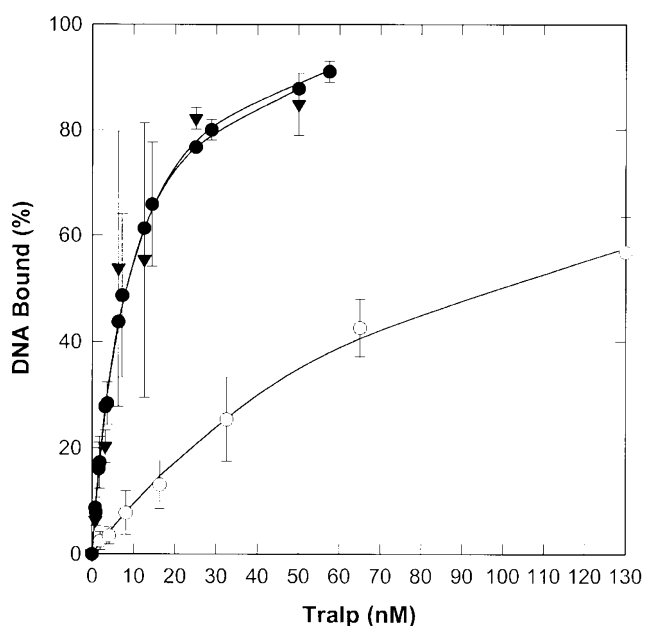
FIG. 5. **DNA binding activities of 226C, 309C, and 348C.** DNA binding assays were performed using the double-filter (nitrocellulose + DE81) technique as described under "Experimental Procedures." The DNA ligand (the 93-bp partial duplex DNA substrate) was incubated with the indicated amounts of each protein prior to filtration. ●, 226C; ▼, 309C; ○, 348C. The data represent the means ± S.D. of three to four determinations. The *line* in each plot represents the fit to a rectangular hyperbola.
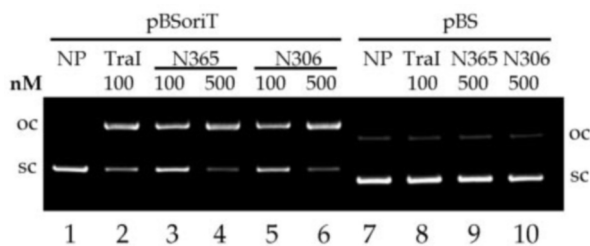

FIG. 6. **The N306 and N365 segments of TraI exhibit** *oriT*-**specific transesterase activity.** The ability of N-terminal segments of TraI to specifically nick supercoiled DNA containing *oriT* was assessed as described under "Experimental Procedures." The DNA substrate (pBSoriT, *lanes 1–6*; pBS, *lanes 7–10*) was present at 7 nM and was incubated with the indicated concentrations of each protein for 15 min at 37 °C prior to the addition of protein denaturants. Reaction products were resolved on a 0.8% agarose gel that was stained with EtBr to visualize the results. The pBS plasmid was identical to pBSoriT, except that it lacks the *oriT* sequence from the F plasmid (8). The supercoiled (*sc*) and open circular (*oc*) forms of the DNA substrate(s) are indicated on the right. *NP*, no protein.

transesterase domain, with the C-terminal end of the active transesterase located between residues 235 and 306.

Genetic Characterization of traI Alleles—The functional *traI* segments generated in this study were tested for their capacity to complement a strain containing a mini-F plasmid lacking the *traI* gene (DB52, Tra⁻) for CDT as described under "Experimental Procedures." Only the full-length *traI* gene was able to restore the Tra⁺ phenotype. The segmental *traI* alleles encoding functional domains of the protein, whether overlapping or abutting, failed to restore transfer in all cases tested (data not shown). This included expression of each functional domain singly and in combination with the other functional domain. This result was expected because previous studies have shown that both the transesterase and helicase activities of TraI are essential for F plasmid-mediated CDT (14). In addition, this result is consistent with the results presented for the analogous
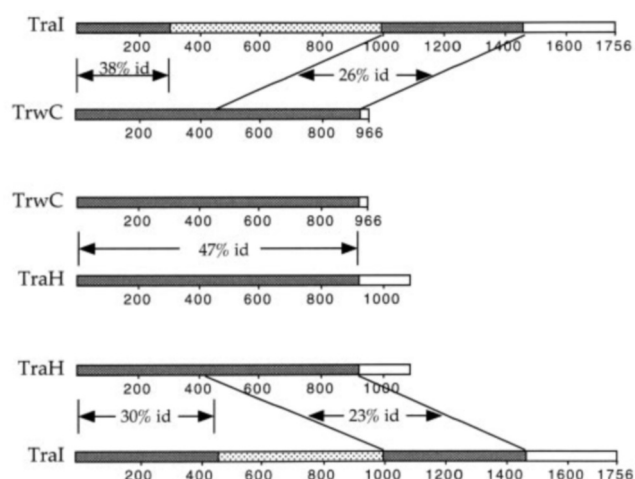

FIG. 7. **Binary alignments of TraI, TrwC, and TraH.** Amino acid sequences were aligned using the SIM algorithm (30) as described under "Experimental Procedures." The *shaded boxes* indicate well conserved regions among all three proteins. The extent and degree of identity (*id*) are indicated by the *double-headed arrows* and the accompanying numeric values. The *stippled boxes* indicate the region within TraI that is absent from the other two proteins. The *white box* at the C terminus of each protein represents a segment that is apparently unique to each, as no significant similarities were detected among three.

TrwC protein from plasmid R388 (10), where coexpression of the two functional domains on overlapping protein segments produced poor complementation.

Data Base Searches for Similar Proteins—The PROSITE and TREMBL Databases were searched with degenerate amino acid sequence patterns, and all hits were subjected to binary alignment with F-TraI as described under "Experimental Procedures." Only three proteins were uncovered (Fig. 7), underscoring the apparent scarcity of known proteins exhibiting the degree of similarity imposed by this approach. It should be noted that an apparent conjugative transesterase-helicase protein (*Agrobacterium tumefaciens* pTiC58-TraA) has been described (11). pTiC58-TraA was not identified in the data base search because it lacks the two N-terminal tyrosine doublets and the comparatively high amino acid identity exhibited by the other three proteins.

DISCUSSION

Previous studies have shown that the F plasmid-encoded TraI protein catalyzes two distinct biochemical reactions: a 5′ to 3′ DNA helicase reaction and a site- and strand-specific transesterase reaction (5, 6, 8, 9, 13). Both of these activities are essential to complete the strand transfer reaction associated with bacterial conjugation (14). The results presented here clearly demonstrate that the transesterase and helicase activities of TraI reside in separable domains of the full-length protein. The N-terminal domain (residues 1–306) harbors the transesterase activity associated with the TraI protein. The remainder of the protein (residues 309–1756) is an active 5′ to 3′ DNA helicase. Thus, the domains of TraI do not overlap. The fact that R388 plasmid TrwC, a protein from the conjugative plasmid R388 that is similar in organization and function to TraI, could be partially separated into these component activities (10) demonstrated a lack of obligatory interdependence of the two activities and raised the possibility that the activities reside in truly distinct domains. This has now been demonstrated for the F plasmid TraI protein, where the two domains can be fully separated without significant loss of biochemical activity.

The transesterase activity associated with purified N306 was

qualitatively similar to that of the full-length protein. Although the specific activity of this protein appeared to be somewhat reduced relative to that of the full-length protein, the purified N306 fragment of TraI catalyzed a robust transesterase reaction that was both site- and strand-specific and dependent on negatively supercoiled DNA. A smaller protein fragment (N235) lacked transesterase activity, suggesting that the C-terminal end of the functional transesterase lies within the region between residues 235 and 306. Indeed, the N235 protein failed to bind a ssDNA oligonucleotide containing *nic*, indicating that the region between amino acids 235 and 306 is important for binding of the transesterase to its DNA substrate. We also note that the prominent proteolytic fragment obtained when N365 was isolated is consistent with the existence of a stable domain. This protein is slightly larger than N306 and represents the N-terminal end of the protein (see "Results"). This suggests that the N-terminal portion of TraI folds into a stable domain that may be slightly larger than the domain defined by N306. The results reported here for the F plasmid relaxase are in contrast with the results reported for the R388 plasmid TrwC relaxase (10). In that case, a smaller protein fragment (residues 1–225) resulted in a protein that catalyzed transesterase activity using an oligonucleotide substrate, but failed to catalyze the same reaction with a supercoiled plasmid. We conclude that the F-TraI transesterase domain occupies the first ~310 residues, that it adopts a stable structure with nearly native transesterase activity, and that it is separable from the helicase domain associated with TraI.

The helicase activity associated with 309C was nearly identical to that of the native protein. The 309C fragment catalyzed a processive unwinding reaction with almost the same specific activity as full-length TraI. Remarkably, the 39-amino acid difference between 309C and 348C dictated whether the protein was a functional helicase or completely defective catalytically (*i.e.* 309C was a fully functional helicase, whereas 348C lacked detectable helicase activity). A similar result was obtained by Llosa *et al.* (10), in that the segment of TrwC beginning at residue 346 and extending to the native C terminus (346C) lacked ATPase activity, whereas the segment of TrwC beginning at residue 192 was a fully functional helicase. Thus, we have defined the N-terminal end of the minimal functional helicase as beginning between residues 309 and 348. The helicase-associated motifs are contained within a segment of the protein extending from approximately residues 990 to 1450. Thus, there are ~300 amino acids at the C-terminal end of TraI that lie outside the helicase-associated domains. We have removed the C-terminal 252 amino acids to construct TraIΔ252, which is also active as a DNA helicase. Thus, the protein fragment extending from residue 309 at the N-terminal end to residue 1504 at the C-terminal end is a functional DNA helicase. The role played by the C-terminal ~250 amino acids of TraI is not clear at present. However, preliminary results suggest that this region of the protein is essential for CDT.[7]

The size and complexity of the functional TraI helicase were unexpected. The helicase-associated motifs in TraI begin at residue ~990, and a putative restart protein (TraI*) whose coding sequence is located within the *traI* gene beginning at about residue 950 (31) seemed like a good candidate for an active helicase. There is a putative ribosome-binding site located just upstream of a methionine codon, and it has been speculated that TraI* is synthesized as a restart protein (31) much like the small form of the bacteriophage T7 gene 4 protein (32). Our results suggest that if TraI* is synthesized in the cell, it will not harbor helicase activity because it will be miss-ing the region from residues 309 to 950, which we have shown to be essential for helicase activity. In fact, we have directly tested the idea of helicase activity associated with the putative TraI* protein by expressing and purifying a TraI*/maltose-binding protein fusion. This protein was devoid of both ATPase and helicase activities.[5] Thus, the active helicase requires a large region between the N-terminal end of the helicase domain and the helicase-associated motifs.

There is a wealth of data supporting the notion that the evolutionarily conserved amino acids constituting the helicase-associated motifs in superfamily I DNA helicases serve to couple nucleoside 5′-triphosphate hydrolysis with translocation and unwinding of duplex nucleic acids (33–37). However, there is little information on the role of sequences outside the motifs, as the comparatively low level of sequence conservation makes selection of mutagenic targets difficult. Given the distance between the helicase-associated motifs and the N-terminal end of the functional helicase (~650 amino acids), perhaps this region of the protein is responsible for some activity associated with TraI that has yet to be recognized and defined. Alternatively, this region could play a strictly structural role. It is possible that the absence of the N-terminal 39 amino acids that distinguish 348C from 309C may have negative consequences for the global folding of the C-terminal 80% of TraI. However, this seems unlikely, as the 348C fragment of the protein was soluble and could be purified using the same protocol used to purify the native protein.

In comparing TraI with TrwC and TraH (the two proteins most closely related to TraI in sequence and organization), the most striking difference is the distance from the end of the N-terminal similarity to the beginning of the C-terminal similarity, the latter coinciding with the region containing the helicase-associated motifs (see Fig. 7). In TrwC and TraH, this distance is ~200 amino acids, whereas the analogous region of TraI comprises roughly 650 residues. Like TraI, TrwC has been characterized *in vitro* as a 5′ to 3′ DNA helicase (7), but this activity clearly does not require the extensive sequences present in TraI. Thus, there is over three times more "information potential" in TraI than presumably would be required just to enable helicase activity, raising the possibility of the central region having an activity or role distinct from the transesterase or helicase functions. The nature of this function, if any, has not yet been identified.

REFERENCES

1. Firth, N., Ippen-Ihler, K., and Skurray, R. A. (1996) in *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology* (Neidhardt, F. C., ed) pp. 2377–2401, American Society for Microbiology, Washington, D. C.
2. Lederberg, J., and Tatum, E. L. (1946) *Nature* **158,** 558
3. Lanka, E., and Wilkins, B. M. (1995) *Annu. Rev. Biochem.* **64,** 141–169
4. Byrd, D., and Matson, S. W. (1997) *Mol. Microbiol.* **25,** 1011–1022
5. Traxler, B., and Minkley, E. J. (1988) *J. Mol. Biol.* **204,** 205–209
6. Lahue, E., and Matson, S. W. (1988) *J. Biol. Chem.* **263,** 3208–3215
7. Grandoso, G., Llosa, M., Zabala, J., and de la Cruz, F. (1994) *Eur. J. Biochem.* **226,** 403–412
8. Matson, S. W., and Morton, B. (1991) *J. Biol. Chem.* **266,** 16232–16237
9. Reygers, U., Wessel, R., Muller, H., and Hoffman-Berling, H. (1991) *EMBO J.* **10,** 2689–2694
10. Llosa, M., Grandoso, G., Hernando, M., and de la Cruz, F. (1996) *J. Mol. Biol.* **264,** 56–67
11. Farrand, S., Hwang, I., and Cook, D. (1996) *J. Bacteriol.* **178,** 4233–4247
12. Fukuda, H., and Ohtsubo, E. (1997) *Genes Cells* **2,** 735–751
13. Fukuda, H., and Ohtsubo, E. (1995) *J. Biol. Chem.* **270,** 21319–21325
14. Matson, S. W., Sampson, J. K., and Byrd, D. R. N. (2001) *J. Biol. Chem.* **276,** 2372–2379
15. Pansegrau, W., Schroder, W., and Lanka, E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2925–2929
16. Ilyina, T., and Koonin, E. (1992) *Nucleic Acids Res.* **13,** 3279–3285
17. Pansegrau, W., and Lanka, E. (1991) *Nucleic Acids Res.* **19,** 3455–3463
18. Llosa, M., Bolland, S., and de la Cruz, F. (1994) *J. Mol. Biol.* **235,** 448–464
19. Abdel-Monem, M., Taucher-Scholz, G., and Klinkert, M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4659–4663

[7] S. W. Matson and H. Ragonese, manuscript in preparation.

20. Willetts, N., and Maule, J. (1979) *Mol. Gen. Genet.* **169,** 325–336
21. Abdel-Monem, A., and Hoffmann-Berling, H. (1976) *Eur. J. Biochem.* **65,** 431–440
22. Abdel-Monem, A., Durwald, H., and Hoffmann-Berling, H. (1976) *Eur. J. Biochem.* **65,** 441–449
23. Kuhn, B., Abdel-Monem, M., Krell, H., and Hoffmann-Berling, H. (1979) *J. Biol. Chem.* **254,** 11343–11350
24. Matson, S. W., Nelson, W., and Morton, B. (1993) *J. Bacteriol.* **175,** 2599–2606
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, p. A.1, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Panicker, M. (1985) *Role of TraD Gene Product in F Sex Factor-mediated Conjunction in E. Coli.* Ph.D. thesis, Carnegie-Mellon University
27. Matson, S. W. (1986) *J. Biol. Chem.* **261,** 10169–10175
28. Gorbalenya, A., Koonin, E., Donchenko, A., and Blinov, V. (1989) *Nucleic Acids Res.* **17,** 4713–4729
29. Koonin, E., and Ilyina, T. (1993) *Biosystems* **30,** 241–268
30. Xiaoquin, H., and Webb, M. (1991) *Adv. Appl. Math.* **12,** 337–357
31. Traxler, B., and Minkley, E. J. (1987) *J. Bacteriol.* **169,** 3251–3259
32. Dunn, J. J., and Studier, F. W. (1983) *J. Mol. Biol.* **166,** 477–535
33. Brosh, R. J., and Matson, S. W. (1995) *J. Bacteriol.* **177,** 5612–5621
34. Brosh, R. J., and Matson, S. W. (1996) *J. Biol. Chem.* **271,** 25360–25368
35. Brosh, R. J., and Matson, S. W. (1997) *J. Biol. Chem.* **272,** 572–579
36. Hall, M., and Matson, S. W. (1997) *J. Biol. Chem.* **272,** 18614–18620
37. Hall, M. C., and Matson, S. W. (1999) *Mol. Microbiol.* **34,** 867–877
38. Wong, I., and Lohman, T. M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 5428–5432