Published in final edited form as:

Proteins. 2012 August ; 80(9): 2250–2261. doi:10.1002/prot.24114.

Solution structure and small angle scattering analysis of TraI (381-569)

Nathan T. Wright^{a,c}, Madushi Raththagala^a, Casey W. Hemmis^a, Sheldon Edwards^a, Joseph E. Curtis^b, Susan Krueger^b, and Joel F. Schildbach^a

^aDepartment of Biology, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218

^bNational Institute of Standards and Technology, Gaithersburg, MD 20899

Abstract

TraI, the F plasmid-encoded nickase, is a 1756 amino acid protein essential for conjugative transfer of plasmid DNA from one bacterium to another. While crystal structures of N- and C-terminal domains of F TraI have been determined, central domains of the protein are structurally unexplored. The central region (between residues 306 and 1520) is known to both bind single-stranded DNA (ssDNA) and unwind DNA through a highly processive helicase activity. Here, we show that the ssDNA binding site is located between residues 381 and 858, and we also present the high-resolution solution structure of the N-terminus of this region (residues 381 to 569). This fragment folds into a four-strand parallel β sheet surrounded by α helices, and resembles the structure of the N-terminus of helicases such as RecD and RecQ despite little sequence similarity. The structure supports the model that F TraI resulted from duplication of a RecD-like domain and subsequent specialization of domains into the more N-terminal ssDNA binding domain and the more C-terminal domain containing helicase motifs. In addition, we provide evidence that the nickase and ssDNA binding domains of TraI are held close together by an 80-residue linker sequence that connects the two domains. These results suggest a possible physical explanation for the apparent negative cooperativity between the nickase and ssDNA binding domain.

Keywords

bacterial conjugation; NMR solution structure; SAXS; SANS; ssDNA binding protein

Introduction

Bacterial conjugation allows direct transfer of genetic information between bacteria ¹. During conjugation, plasmid DNA in single-stranded form is unidirectionally transported through a pore complex from a donor to a recipient cell ^{2,3}. This process of DNA transfer, first observed almost 60 years ago, is a major mechanism by which diverse species of bacteria acquire both antibacterial resistance and metabolic cassettes, and is a driving force of bacterial evolution ^{4–7}. Conjugation is common in the bacterial world, occurring between different bacterial species and even reported to occur between bacterial and human cells ^{7–13}. The prospects of manipulating horizontal gene transfer to curtail the spread of antibacterial resistance, and more broadly to understand bacterial evolution, are predicated in part on a detailed understanding of the molecular mechanism of this process.

Corresponding author: Joel F. Schildbach, Department of Biology, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA joel@jhu.edu.

^cCurrent Address: Department of Chemistry and Biochemistry, James Madison University, 800 S. Main Street, Harrisonburg, VA 22807

Bacterial F-plasmid transfer requires a specialized set of proteins, named Tra or transfer proteins, that are encoded on the conjugative plasmid¹⁴. These proteins form a conjugative pilus (TraA) and type IV-like secretory pore (TraB, TraK, TraV, TraC, TraE, TraL, TraG) to contact recipient cells and transport protein and ssDNA^{14,15}. They also bind, cleave, unwind, and move plasmid DNA from one cell to another (TraI, TraY, TraM)^{16,17}. To transport DNA, TraY and the chromosomally-encoded protein integration host factor (IHF) must first bind to and distort the conjugative plasmid DNA at the origin of transfer (*oriT*) with the effect of creating a short ssDNA region^{18,19}. TraI then binds sequence-specifically to this ssDNA, cleaves the DNA in a metal-dependant manner, and creates a stable nucleoprotein intermediate through the covalent phosphodiester bond between Tyr16 on TraI and the 3' phosphate of T141' on the DNA²⁰⁻²³. TraI also possesses helicase activity, and in one model of activity this helicase is required to unwind the plasmid and separate the two DNA strands, allowing DNA to shuttle from the donor cell through the conjugative pore and into the recipient cell^{21,24}. Once the transfer process is completed, the DNA in the recipient can be re-circularized by TraI and second strand synthesis can be performed in donor and recipient, resulting in two cells capable of additional rounds of conjugative transfer.

TraI can be functionally divided into at least three segments (figure 1)²⁵⁻²⁷. The first 306 amino acids comprise the nickase domain, whose function is to specifically bind to and cleave the origin of transfer (*oriT*)²⁸. The adjacent ~600 amino acids have been recently shown to possess high-affinity nonspecific ssDNA binding activity²⁹. Interestingly, TraI can bind either *oriT* ssDNA at the nickase site or nonspecific ssDNA at this second site, but not to both simultaneously²⁹. This apparent negative cooperativity between the two binding sites suggests that the two domains are close together in physical space. Despite exerting influence on one another, however, each domain is functional and stable outside the context of the rest of the protein^{26,28,30}. It is unclear what function the large middle ssDNA binding region performs in the context of DNA transport during conjugation; mutations that relieve this apparent negative cooperativity of binding, presumably through disruption of the ssDNA-binding domain structure, often display higher than wild-type mating efficiencies²⁶. This ssDNA binding site has been dubbed a helicase-like ssDNA-binding domain because it displays weak homology to the *E. coli* helicase RecD, even though this TraI domain contains no Walker A or B boxes, no ATPase activity, and exhibits no helicase activity. The C-terminal half of TraI (858-1520) contains another more conserved helicase motif that shows sequence homology to RecD. Both the ssDNA binding domain and the helicase domain contain translocation sequences (TSA and TSB) that facilitate protein movement from host to recipient cell³¹. It is currently unclear whether these sequences are functional if separated from the domain fold, or if the RecD-like fold itself is a TS for this system.

TraI must be able to bind and unwind DNA, and likely unfolds in order to traverse the 20 Å diameter pore in the type IV secretion system connecting the two bacteria³¹⁻³³. Once in the recipient, TraI must refold and ligate the ssDNA back into a circular plasmid^{31,33}. While the protein itself is by necessity functionally complex and highly dynamic, some individual domains of TraI have proven to be well-behaved for crystallographic or NMR structural determination^{28,30,34}. We initially set out to determine the structure of TraI (381-858), the helicase-related ssDNA binding domain, but were unable to crystallize the fragment and found the fragment was unsuitable for NMR analysis. Here, we describe the high-resolution solution structure of the well-folded N-terminus (residues 381 to 569) of the ssDNA binding domain of F-plasmid TraI. We show that this fragment, despite limited sequence similarity, folds into a globular structure that closely resembles the N-termini of other helicases such as RecD and RecQ. Using NMR and small angle scattering experiments we also demonstrate that in a larger TraI fragment the nickase and ssDNA binding domains are close together. The proximity of the domains occurs despite evidence that the 80-a.a. linker between them

is highly dynamic. Thus, we here provide a structural context for the observed negative cooperativity of ssDNA binding between these two regions.

Results

In total, the solution structure of TraI (381-569) contains 2204 distance restraints, including 519 intraresidue, 596 sequential, 468 medium, 481 long range, 140 hydrogen bonds, 241 dihedral, and 96 residual dipolar coupling restraints. Together, this gives an average of greater than 13 constraints per residue. NOE correlations were assigned on the basis of the previously reported ^1H , ^{13}C , and ^{15}N side chain and backbone chemical shifts for this fragment of TraI³⁰. Almost all of the N-terminal 10 residues, including the leading Met, are invisible in all NMR experiments. When taken as a whole, the NOE correlations and the chemical shift indices show that TraI (381-569) contains four β strands (S432-Q439, G460-A469, S503-Q509, and V530-G538) and six helices (L390-R404, S420-R430, Q445-E458, R470-S483, E511-R527, S545-S553; figure S1 and S2).

TraI (381-569) assembles into a single structural domain, defined by a central four-strand parallel β sheet that is flanked by helices. NOE data unambiguously demonstrate this. For example, a plane from the 4D ^{13}C , ^{13}C -edited NOESY experiment (figure 2a) shows one of the methyl groups of residue L426, located near the end of helix 2. NOE correlations are visible to other residues in helix 2 (V423), at the beginning of helix 1 (S391), at the turn between helix 2 and sheet 1 (R430, P431), in sheet 1 (A434, V436), helix 3 (M455), sheet 3 (I506), and in sheet 4 (L533). Taken together these data indicate that L426 comprises part of the hydrophobic core of the molecule, packing against strands 1, 3, and 4. Its presence within a hydrophobic region is further verified by the high upfield shift (0.132 ppm) of the δL426 hydrogen moiety. The orientation of the helices relative to one another became clear early in the structural calculations of TraI (381-569) because of the structural constraints imposed by the central parallel β sheet. In fact, when all long-range inter-helical and helix-sheet restraints are removed (but all inter-strand NOEs are left in place), all helices except helix 1 remain in roughly the same orientation. The presence of interhelical NOEs, such as those visible in figure 2a from helix 2 (L426) to helix 1 (S391) and helix 3 (M455), further defines the exact location of the helices relative to each other.

In order to refine and independently verify the three-dimensional structure of TraI (381-569), residual dipolar coupling (RDC) data were collected (figure 2b)³⁵. Here, 96 $^1\text{D}_{\text{N-H}}$ values for TraI (381-569) were recorded from a ^{15}N - ^1H -2D correlation spectrum in the absence of ^1H decoupling during the t_1 evolution period, and the upfield and downfield components of the ^{15}N - ^1H doublets were separated into two spectra in an interleaved manner as described previously³⁶. These measurements fit well into the TraI fragment structure, and when a random subset was removed for Q-value calculation the average value was 0.27 ± 0.05 . To ensure that we did not bias the data by taking out only well-fitting RDC values, a second random subset of RDC values was removed, another ensemble of structures was calculated, and the Q-value of this second group was 0.30 ± 0.03 .

An ensemble of 200 structures was calculated using RDC, distance, and angle constraints. More than 120 of these had no NOE violations greater than 0.40 Å and no dihedral violations of greater than 5°. Of these, 75 had no Ramachandran plot violations as determined by the program Procheck³⁷. Q-values were calculated on these last 75 structures. Figure 3a shows the superposition of the 20 that had the lowest Q-values and the lowest backbone rmsd. Overall, the ensemble has a reasonably good rmsd when calculated from the average structure (< 0.86 Å for the backbone and < 1.43 Å for the heavy atoms; Table 1). Only residues 390 to 553 are shown in figure 3a, since these residues represent the

first and last residue that have long-range NOE connectivities. When only regions of the molecule that are in regular secondary structure (blue sheets and red helices) are considered, the rmsd of backbone heavy atoms calculated from the average structure drops to 0.66 Å. In particular, the random coil region from K477 to P501 is relatively poorly defined and shows considerable deviation between the structures. Peaks in the NOESY spectrum from this region tended to be weak and low in number, likely due to some amount of internal motion.

Figure 3b shows a cartoon version of the NMR model that is the closest to the average. This structure has good values in all categories present in Table 1, with a majority of the residues in the most favorable region of the Ramachandran plot (79.8 %) and no residues in the disallowed region (figure S3). The structural fold is most reminiscent of a large family of helicases (RecQ, RecD, Dbp5), as assessed by the program DALI (avg. Z-score=5.9; avg. C α rmsd=3.3 Å), even though *E. coli* RecD at only 10% identity demonstrates the most sequence conservation.

Based on limited proteolysis studies (figure S4)³⁰, TraI (381-569) comprises the N-terminal region of a larger domain that can bind ssDNA with a low nM K_D and low sequence specificity. Based on results generated using the structure prediction program Phyre³⁸, this whole domain (381-858) is expected to resemble the *E. coli* RecD helicase. However unlike RecD, this region of TraI lacks essential helicase motifs and possesses no helicase activity; presumably the region plays a role in either DNA unwinding or regulation of ssDNA transfer from host to recipient cell. TraI (381-569) itself does not possess any measurable DNA binding affinity (Table 2). The intact domain (381-858), though, possesses high affinity for the 22-base TAMRA-labeled G144'C *oriT* oligonucleotide used previously to assess binding by the TraI helicase-associated ssDNA binding site³⁹. Interestingly, a stable fragment from residue 541 to 858 also possesses no ssDNA binding (data not shown), suggesting that both the N and C-terminal regions of this domain are needed for proper function. A second region that contributes detectably to ssDNA binding is located between residues 309 and 331 as judged by an approximate 7-fold affinity decrease in affinity when the region is not present (Table 2). The contribution by this region could result from a single polar or electrostatic interaction. To test if any single side chain in this region was important for this interaction, we created a series of substitutions from positive or polar residues to Ala (Q317A, K318A, K331A). None of these substitutions reduce binding to the extent that deleting the region does, consistent with the segment providing multiple weak interactions with the DNA.

Although the TraI (309-380) region contributes to ssDNA binding, results from limited proteolysis studies³⁰ suggest the region may be dynamic. To examine the structural state of this region, we collected an HSQC spectrum of TraI (309-569). Peaks corresponding to residues 381 to 569 overlaid with this spectrum quite well, suggesting that the core fold is unchanged with the addition of this N-terminal segment (figure 4a). However, the additional peaks corresponding to residues 309 to 380 generally display a profile consistent with an unfolded or highly dynamic domain. This is particularly visible in the middle of the spectrum, where the peaks severely overlap. We tested a large range of conditions, and although a small subset of amide peaks display unique shifts, we concluded that the region was poorly ordered and sequence-specific chemical shift assignment of this spectrum was not possible.

As a further structural examination of this region, we ran a ¹H-¹⁵N heteronuclear NOE (hetNOE) experiment on the 309-569 domain. Again, all peaks from the previously-assigned 381-569 region overlapped with this longer segment. Peaks from the β sheet region (residues 432-438, 463-467, 502-508, and 530-538) demonstrated typically well-ordered hetNOE ratios of 0.79 (+/- 0.13). The large standard deviation comes from a pronounced

dip of those residues that contact 509; this peak is broadened by exchange and the flexibility this indicates is apparently partially propagated to its neighbors as well. In contrast to the β sheet, residues 481-498, which are in a random coil and have a higher RMSD in our structure, show hetNOE values of 0.49, indicating a more dynamic region.

In the 309-569 data there were roughly 36 extra peaks relative to the 389-569 data, presumably belonging to the 309-380 fragment. We expected 68 peaks, and the missing peaks are either obscured or are in some intermediate exchange. Unassigned peaks have an average hetNOE value of 0.35. 6 peaks show values about 0.80, indicating there may be some small secondary structural element in an overall dynamic region. These values demonstrate that the region is not completely disordered, but has intermediate dynamics. This is consistent with the idea of this region having intrinsic disorder as proposed elsewhere in this manuscript.

One formal possibility for the messy 309-569 HSQC spectrum is that the fragment is aggregating or oligomerizing. During molecular weight exclusion chromatography, TraI (309-569) eluted with an apparent molecular weight of 31.7 kDa, compared to its calculated molecular weight of 28.4 kDa (figure S5). We therefore conclude that there is either no or minimal aggregation of this sample.

Because the 309-380 region contains some residues that participate in ssDNA binding, and because it is located between two domains of TraI that compete for binding to DNA²⁹, the conformation of this region affects both the orientation of the domains and the function of TraI. To better understand the conformation, and ultimately the function, of TraI, we turned to a combination of small angle scattering and molecular modeling. Although small angle X-ray and neutron scattering (SAXS and SANS, respectively) yield models that are, at best, low resolution, the results should both provide us some insight into TraI function and generate models that can be tested and refined experimentally. We initially constructed a model of TraI (1-330). This fragment was generated using limited proteolysis²⁰ and its structure previously determined. In the crystal structures, however, the C-terminus of this fragment is disordered. We used TraI (1-330) as an initial test case for our modeling procedure. We constructed the model using residues 1-306 from the crystal structure of TraI (1-330)⁴⁰ and a C-terminal random coil from residue 307 to 330. Each phi and psi angle of the C-terminus was randomized so that the segment could extensively sample allowed conformational space. In this way, 5795 models with different C-terminal tails were generated. We then conducted a series of small angle neutron scattering (SANS) experiments on the TraI (1-330) fragment (figure S6). (SANS data utilizes an absolute scale and is superior for detecting small amounts of aggregation, however we did not see aggregation with any constructs tested using SANS. Because SAXS measurements give a higher signal to noise ratio, most subsequent experiments utilized SAXS.) The experimentally-determined SANS data were compared to the theoretical SANS profiles for each TraI (1-330) model. Figure 4b shows the crystal structure of TraI (1-306) in red, with the protein shell of residues 307 to 330 of all models colored in black mesh. The protein shell of residues 307 to 330 of those structures that had the lowest χ^2 values (< 5.0 from a maximum of 65.5) is colored in solid green. From this analysis, it is clear that the green protein shell represents structures that are not folded next to the rest of the nickase domain, nor entirely extended, nor are they folded into a small sub-domain. These results are consistent with the 309 to 330 region being mostly unfolded in this context. SAXS analysis yielded similar results.

Extending the studies to larger TraI fragments, SAXS data were collected on TraI (1-569), both in the presence and absence of a small piece of ssDNA (*oriT*) (figure S8). While TraI (1-569) comprises only half of the nonspecific ssDNA binding domain, it was used here

because a) there are high resolution structures of both 1-306 and 381-569, b) it permits a stepwise increase in the complexity of the modeling, and c) TraI (381-858) is predicted to be a single compact structure with two independently folded subdomains (much like the structure of RecD; figure S9), and therefore an accurate model of the 1-569 construct will be an excellent starting point for modeling the 1-858 fragment. The radius of gyration (R_g) of TraI (1-569) is similar both with and without *oriT* bound (35-38 Å), with no aggregation detected in the Guinier plot (figure S10-a,b). Using methods similar to those described in figure 4, ensembles of $\approx 10,000$ TraI (1-569) and TraI (1-569)-*oriT* models were generated using the high-resolution x-ray structure of the nickase (residues 1 to 306) and the solution structure of TraI (390-569) as structured bookends to a flexible middle region. Theoretical SAXS decay curves for each model were then back-calculated, and these curves were compared to the experimental SAXS profile of either TraI (1-569) or TraI (1-569)-*oriT*. χ^2 values of the best-fitting TraI (1-569) models were higher than anticipated (≈ 17), with a broad lower χ^2 boundary from 35-38 Å. Also, the $P(r)$ distribution shows a complex protein conformation (figure S10c)⁴¹. We interpret this as an indication of a dynamic structure; one model cannot completely describe the system. In contrast, the best models of TraI (1-569)-*oriT* agree better with the SAXS data ($\chi^2 = 3.0$). We therefore used TraI (1-569)-*oriT* for the remainder of the SAXS analysis. As shown in figure 5a, the best models are those with an $R_g \approx 34$ Å. R_g values in the ensemble of models ranged from 28 Å to over 140 Å. We reasoned that if the linker region were completely flexible, the best-fitting models would have intermediate R_g values, and would have a broad minimum in the R_g vs χ^2 graph. Therefore, such a small experimental R_g value indicates that TraI (1-569)-*oriT* is compact in solution. Figure 5b is a visualization of the model ensemble. Here, TraI (1-305) is depicted in ribbon form in red. The protein shell of those models with a low χ^2 value (< 15) is shown in blue, while the protein shell of all models is depicted in black mesh. Figure 5c shows the model of TraI (1-569)-*oriT* that agrees best with the experimental SAXS data. From this depiction, it is clear that the two domains are near one another. Such close proximity of these two domains is also evident in the TraI (1-569) analysis. While we are confident that the domains are close, SAXS and SANS analysis does not allow generation of high resolution models. Determining, for example, the relative orientation of each domain to its neighbor will require considerable additional experimentation.

The small R_g value was unexpected, given that the NMR data (fig 4a) suggests the linker region is flexible. We propose three potential explanations for this inconsistency: a) the two flanking domains interact with each other, b) the linker region folds correctly in the presence of the nickase domain, or c) the linker region is dynamic and/or partially folded but contains enough structure to keep the flanking domains near each other. To search for evidence of a direct interaction between regions 1-330 and 381-569, we attempted formaldehyde crosslinking of the domains, both in the intact 1-569 chain and in a mixture of the individual domains, without positive results (data not shown). Both of the remaining hypotheses are plausible, and future experiments will focus on distinguishing between them.

Discussion

A recent paper by Cheng *et al.* proposes a model of the overall TraI structure where each functionally independent TraI domain is connected to the next domain via medium-length flexible linkers⁴². While our current study only details the high-resolution structure of TraI (381-569), our NMR and SAXS data together allow further refinement of this model. The nickase domain (1 to 306) structure is known, and based on the data presented here, TraI (381-858) very likely folds into one large helicase-like ssDNA-binding motif. SAXS data suggests that these two domains are held close to one another. While we cannot speculate on the relative orientation of the two domains, this proximity may help explain the coupling of ssDNA binding of one domain to the inability to bind ssDNA to its neighboring domain; a

mechanism of binding-site occlusion would be the most obvious explanation, although others are possible.

While the small angle scattering analysis clearly shows that the nickase and ssDNA binding domains exist close together in space, NMR, crosslinking, and protease digestion studies are all consistent with the region linking the domains being disordered in solution^{22,30}. One possible explanation, consistent with the hetNOE data presented, is that the linker region has a defined albeit flexible structure, which in turn can aid in ssDNA binding and can keep the two flanking domains near each other.

TraI (381-569) shares only limited sequence identity (19 residues; 10%) with its closest structural homologue, RecD from *E. coli* (residue 125-355), but the two domains are very similar in secondary structure and tertiary topology (C_{α} rmsd ≈ 3.4 Å). Residues that are conserved between the structures are colored in red in figure 6a. Interestingly, while some of these regions comprise the hydrophobic core of the protein (V392, S437, G438, L452, V507, D508, V532, and A546), most are more peripheral. Similar residues between the two proteins are colored in yellow in figure 6a, and are almost all either part of the core parallel β sheet or residues that abut this sheet. Thus despite the two proteins being only 28% similar, they are most alike in their core 4-strand β sheet; this fold can clearly tolerate a large number of substitutions outside of its core structure and still adopt the same general conformation. Compared to TraI, RecD has several insertions of four or more amino acids (after residues 405, 457, 481, 492, and 553; TraI numbering). All of these insertions occur either in a loop or at the very end of a helix. Because of the sites of insertion, there are relatively limited structural consequences of these insertions: TraI (381-569) helix 3 is roughly one turn shorter than its corresponding helix in RecD⁴³. The largest sequence deviation occurs after residue 481 in TraI, and consists of an additional 17 amino acids in the RecD sequence. Interestingly, this region makes up part of the ssDNA binding site in RecD. TraI has few interresidue NOEs here, presumably due to some amount of internal motion. Similarly, this region of RecD has significantly higher B-factors when compared to the rest of the protein⁴³. Therefore these two similar protein structures may also exhibit similar dynamic regions, despite having low sequence homology.

A 2006 paper by Haft *et al.* identified locations within TraI at which 31 amino acid insertions could be made without significantly affecting steady-state protein levels (and, arguably, protein folding and stability) but yielded a range of effects on conjugation efficiency²⁶. While many insertions in the helicase region strongly inhibited conjugation, most of those in the helicase-associated ssDNA binding region actually increased conjugation efficiency. Three of the nine insertions that resulted in higher transfer activity (Q439, V505, I506) and one insertion yielding close-to-wild-type transfer (E517) are within or are abutting the well-ordered region of the current NMR structure (figure 6b). For V505 and I506 especially, in the middle of a central β strand, such large insertions (which begin and end with non-hydrophobic residues) should significantly affect the structure or the structural integrity of this domain⁴⁴. Haft *et al.* conclude that the insertions that yield higher efficiency transfer work by releasing a negative regulatory influence on conjugation. Results from this and other work suggest that the ssDNA binding region, and in particular the binding of ssDNA or the stability of the domain, is the cause of this negative conjugation regulation.

Conclusions

Here we present the high-resolution solution structure of the N-terminal fragment of the F TraI helicase-associated ssDNA binding domain. We show that this structure forms a well-folded globular motif consisting of a core of 4 parallel β strands that are surrounded by

helices. This fold closely resembles that of the N-terminus of RecD, with which TraI shares a low degree of sequence similarity. We also show how this region of TraI is situated in the context of the entire N-terminal third of the TraI molecule. Although TraI is a modular protein, the nickase and ssDNA binding domains are in close proximity to each other. This orientation of domains may have functional consequences, in that it gives the apparent negative cooperativity of ssDNA binding a structural basis.

Materials and Methods

Sample preparations

Recombinant ^{15}N -, ^{13}C -, ^{15}N -labeled, and unlabeled TraI (381-569) were purified after overexpression in *Escherichia coli* (HMS174(DE3)) at 37 °C as previously described³⁰. Briefly, after sonication (in 50 mM Tris-base pH 7.5, 50 mM NaCl, and 1 mM EDTA) and centrifugation, the cell lysate was passed over a SP-heparin column (GE Lifesciences, Piscataway, NJ)¹, and the protein was eluted in 0.2 M to 1 M NaCl using a salt gradient produced by an ÄKTAprius plus (GE Healthcare). Fractions containing the TraI fragment were then run over an SP-blue column, and highly purified protein eluted between 1.5 M and 2 M NaCl. Unlabeled TraI (1-330), TraI (1-569), TraI (309-858) and all other TraI fragments were purified using a similar strategy. Proteins were concentrated with Centricon concentrators (10 kDa cut-off; Millipore, Billerica, MA) and dialyzed into either NMR buffer (25 mM Tris-HCl, pH 7.2, 25 mM NaCl, 1mM EDTA, 0.34 mM NaN_3 , 10% $^2\text{H}_2\text{O}$) or a similar buffer without $^2\text{H}_2\text{O}$ or with 100% $^2\text{H}_2\text{O}$. Final TraI (381-569) NMR samples were 0.5 – 1 mM, and SAS samples were 0.5 – 2.0 mg/mL. Fresh acrylamide solutions used in the dipolar coupling experiments were made at 30% acrylamide (1:20 bis) concentration. Compressed gels contained 5% acrylamide solution, 0.015% TEMED, and 0.015% ammonium persulfate (APS), and were prepared as previously described⁴⁵.

To determine the oligomeric state of TraI (309-569) by molecular weight exclusion chromatography, the protein in NMR buffer (0.5 mL of 0.6 mM) was applied to a 10/300 Superose 6 column with a flow rate of 0.2 mL/min. 0.5 mL fractions were collected over 30 mL with a 1 min lag time prior to the start of sample collection.

NMR spectroscopy

NMR spectra were collected at 37 °C with a 600 MHz Bruker AVANCE spectrometer (600.13 MHz for protons) and a Bruker AVANCE 800 NMR spectrometer (800.27 MHz for protons) each equipped with four frequency channels and 5 mm triple-resonance z-axis gradient cryogenic probeheads. The backbone and side chain ^1H , ^{13}C , and ^{15}N chemical shift assignments have been previously reported (BMRB ID 16971)³⁰. A 3D ^{15}N -edited NOESY-HSQC⁴⁶ and 4D ^{13}C , ^{13}C -edited NOESY HSQC spectra with a mixing times of 120 and 130 ms, respectively, were collected on ^{15}N - and ^{13}C , ^{15}N -TraI (381-569) (40). Residual dipolar coupling values were collected via a 2D IPAP ^1H - ^{15}N HSQC with the protein in both an isotropic and aligned environment, as previously described^{45,47}. In the final processed data, the peak position was determined using the pick peaking software of SPARKY⁴⁸.

¹Certain commercial materials, instruments, and equipment are identified in this manuscript in order to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Structure calculation

Interproton distance constraints were derived from 3D and 4D NOESY experiments (^{15}N -edited 3D NOESY and ^{13}C , ^{13}C -edited 4D NOESY) as described previously⁴⁵. Dihedral constraints $\phi \pm 20^\circ$ and $\psi \pm 15^\circ$ for α -helix and $\phi \pm 40^\circ$ and $\psi \pm 40^\circ$ for β -sheet were included based on TALOS+ and the chemical shift index of $^1\text{H}\alpha$ and $^{13}\text{C}\alpha$ atoms⁴⁹. An ensemble of structures without dihedral restraints had an rmsd of 1.13 Å when compared to structures with dihedral constraints. While slightly larger than the overall rmsd, most of this is caused by deviations in the structure between K477 and P501, which have a low number of NOE/residue and are thus more influenced by dihedral and RDC restraints. Hydrogen bond constraints were added into the structure only after the secondary structure was completely determined. Structures calculated without hydrogen bonds had an rmsd of 0.84 Å when compared to those calculated with hydrogen bonds, indicating that inclusion of these bonds did not drastically influence the overall structure. Hydrogen bond constraints of $r_{\text{HN-O}} = 1.5 \text{ Å}$ to 2.8 Å and $r_{\text{N-O}} = 2.4 \text{ Å}$ to 3.5 Å were included in the final stage of structure calculations, and were based off regions that were clearly in well-defined secondary structural motifs. Pseudopotentials for secondary $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts and a conformational database potential were included in the final simulated annealing structural calculations using the computer program XPLOR^{50,51}. Structures run with and without these pseudopotentials show an rmsd of 0.92 Å, indicating that the inclusion of this data does not significantly change the structure. The internuclear dipolar coupling (in Hz) were determined from the difference in J splitting between isotropic and radially compressed polyacrylamide-aligned phases using the 2D IPAP ^1H - ^{15}N HSQC to record N- ^1H splittings. The axial ($\mathbf{D}_a$; 12.5 Hz) and rhombic ($\mathbf{R}$; 0.40) portions of the alignment tensor were obtained from a histogram of the dipolar couplings, and were incorporated into the final structure calculation as previously described using an energy constant of 0.60^{36,45}. The inclusion of these data slightly increased the percentage of residues in the most-favorable region of the Ramachandran plot (figure S3). Q-factors were calculated by randomly removing $\approx 10\%$ of the N- ^1H RDC data, and then comparing these values to those back-calculated from the structure. The final 20 structures were selected (from 200) based on lowest Q-values and lowest rmsd from the average, and were of high quality based on the statistical criteria listed in Table 1. The coordinates of the TraI (381-569) structure have been deposited in the Protein Data Base (PDB accession number [2l8b](#)).

ssDNA affinity

The oligonucleotide used in this study (G144'C: 5'-TTT GCG TGG GCT GTG GTG CTT T-3') was 3'-carboxytetramethyl-rhodamine (TAMRA)-labeled and HPLC-purified (Integrated DNA Technologies, Coralville, IA). The affinities of protein fragments for this oligonucleotide were measured in 50 mM Tris-base pH 7.5, 50 mM NaCl, and 1 mM EDTA at 25 °C, by following the fluorescent intensity changes using an ATF-105 automatic titrating fluorometer (Aviv Biomedical, Inc., Lakewood, NJ) and a single-site binding model was fit to the resulting data, as previously described²⁹.

SANS

Small-angle neutron scattering experiments were performed on TraI (1-330) on the 30 m SANS instruments⁵² at the National Institute of Standards and Technology (NIST) Center for Neutron Research (NCNR). Neutrons of wavelength $\lambda = 6 \text{ Å}$ with a wavelength spread of $\Delta\lambda/\lambda = 0.11$ were used. Measurements were made using a sample-to-detector distance of 1.5 m in order to cover a wavevector transfer range, $0.02 \text{ Å}^{-1} < q < 0.3 \text{ Å}^{-1}$, where $q = 4\pi(\sin(\theta)/\lambda)$ and 2θ is the scattering angle. Samples (TraI (1-330) at 1, 2 or 5 mg/ml) were measured in H_2O and $^2\text{H}_2\text{O}$ buffers. After brief centrifugation, samples were loaded into banjo-shaped quartz cuvettes (Hellma USA, Plainville, NY) of 1- or 2-mm path length (for

H₂O or 100% ²H₂O buffer, respectively). Neutron exposure time was 1 h to 1.5 h. Scattered neutrons were detected with a 64 cm × 64 cm two-dimensional position-sensitive detector with 128 × 128 pixels. Data reduction was accomplished using Igor Pro software (WaveMetrics, Lake Oswego, OR) with SANS macros developed at the NCNR⁵³. The total two-dimensional scattering was corrected for the scattering from the quartz cell, stray neutrons in the room and non-uniform detector response. Then, the scattering was placed on an absolute scale by normalization to the incident beam flux and radially averaged to obtain the scattering intensity, $I(q)$ versus q . This scattering intensity was further corrected for background scattering from the buffers and incoherent scattering from hydrogen in the sample to obtain the final scattering intensity for the proteins in solution. The SANS data are shown in supplementary figure S7. The radius of gyration (R_g) and the scattering at $q=0$, $I(0)$, were calculated from the neutron scattering at low q using Guinier plots⁵⁴. Since the data were obtained on an absolute scale, the Mw of the sample was obtained from $I(0)$ using the relation

$$\frac{I(0)}{c} = \frac{(\Delta\rho)^2}{N_A d^2} M_w \quad (1)$$

where c is the concentration, $\Delta\rho$ is the scattering length difference between the protein and solvent, d is the mass density of the protein and N_A is Avogadro's number. From the Mw calculations, it was clear that only the samples in 100% ²H₂O buffer were monomeric. Thus, structure models were compared only to these data. SAXS and SANS parameters derived from the Guinier analysis are shown in supplementary Table S1.

Structure Modeling of TraI (1-330)

A starting structure for modeling was constructed from the apo TraI (1-330) X-ray crystal structure (PDB ID 1P4D). Residues 236 to 267, which were missing in the 1P4D structure, were added by using the coordinate information from the 2A0I structure for TraI (1-330) bound to 10-bp *oriT* ssDNA. The missing TraI (1-330) residues were inserted into the 1P4D structure, along with 24 missing C-terminal residues, and the structure was energy minimized. Several simulated annealing runs were then performed on region 236 to 267 to determine the range of conformations allowable. Model SANS profiles were calculated for representative samples of the resultant structures using the program Xtal2sas⁵⁵ and references therein). There was no significant difference in these profiles. Because the conformation of region 236 to 267 did not alter the model SANS profiles, further modeling was performed on the modified 1P4D structure with residues 236 to 267 in the conformation from the 2A0I structure. An ensemble of ~10,000 TraI (1-330) model structures was calculated using the program SASSIE⁵⁶ (<http://www.smallangles.net/sassie/SASSIE/>). The 24 added C-terminal residues were permitted to freely and randomly sample allowed conformational space, and the model SANS profile corresponding to each of these model structures was calculated using Xtal2sas. Chi squared analysis and protein density map calculations were run via SASSIE. Protein shells were visualized using VMD⁵⁷.

SAXS

Small angle X-ray scattering was performed on the F2 beamline at the Cornell High Energy Synchrotron Source (CHESS), using an X-ray source with a beam edge of 9.881 keV (1.2563 Å) and an area of 250 μm². After centrifugation, ~20 μL of protein sample with or without 1.05x ssDNA (5'-TTT GCG TGG GGT-3') was loaded into a horizontal capillary tube. Based on the K_D (nM or lower) of the nickase for *oriT* ssDNA⁵⁸, 99.5% of TraI (1-569) is *oriT*-bound under these conditions. Additionally, and importantly, based on the μM K_D of TraI (1-569) for the G144'C oligonucleotide, which binds poorly if at all to the nickase

(Table 2), no more than 2% of TraI (1-569) is bound to *oriT* at residues 309-330. In the beam line, the sample was oscillated during data collection to avoid sample radiation damage. Data was collected for 3 3-minute cycles. To ensure against aggregation, data from multiple concentrations were collected. Dark field and buffer samples were collected, and these were then subtracted from the raw protein data to obtain an $I(q)$ versus q plot using the program RAW⁵⁹. The SAXS data are shown in supplementary figures S6 and S7. Guinier plots for use in estimating the R_g of the (1-569) TraI and *oriT*-(1-569) TraI fragment were then calculated. 8755 model structures of *oriT*-TraI (1-569) were generated using SASSIE, where the bound DNA, plus residues 1 to 306 and 390 to 553 were kept constant and residues 307 to 389 and 554 to 569 were allowed to sample all allowed conformational space. CRY SOL was then used to back-calculate how well these structures fit with the actual SAXS data for *oriT*-TraI (1-569)⁶⁰. Chi squared analysis and protein density map generation were calculated using SASSIE, and the protein shell was visualized using VMD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health GM61017 (to J.F.S.), the Dimitri V. d'Arbeloff fellowship (to N.T.W.), and by American Recovery and Reinvestment Act (ARRA) funds through grant number R01 GM61017 (to J.F.S.) from the National Institute of General Medical Sciences, National Institutes of Health. This work utilized facilities supported in part by the National Science Foundation under Agreement No. DMR-0944772. We also thank the JHU Biomolecular NMR Center, CHESS, and the University of Maryland NMR center for providing facilities, resources and consultation.

References

1. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000; 405(6784):299–304. [PubMed: 10830951]
2. Lanka E, Wilkins BM. DNA processing reactions in bacterial conjugation. *Annu Rev Biochem*. 1995; 64:141–169. [PubMed: 7574478]
3. Llosa M, Gomis-Ruth FX, Coll M, de la Cruz Fd F. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol*. 2002; 45(1):1–8. [PubMed: 12100543]
4. Tatum EL, Lederberg J. Gene recombination in the bacterium *Escherichia coli*. *J Bacteriol*. 1947; 53(6):673–684.
5. Barlow M. What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol Biol*. 2009; 532:397–411. [PubMed: 19271198]
6. Tenover FC. Mechanisms of antimicrobial resistance in bacteria. *Am J Infect Control*. 2006; 34(5 Suppl 1):S3–10. discussion S64–73. [PubMed: 16813980]
7. Davies J. Inactivation of antibiotics and the dissemination of resistance genes. *Science*. 1994; 264(5157):375–382. [PubMed: 8153624]
8. Yoon YG, Koob MD. Transformation of isolated mammalian mitochondria by bacterial conjugation. *Nucleic Acids Res*. 2005; 33(16):e139. [PubMed: 16157861]
9. Yoon YG, Koob MD, Yoo YH. Re-engineering the mitochondrial genomes in mammalian cells. *Anat Cell Biol*. 2010; 43(2):97–109. [PubMed: 21189990]
10. Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, Citovsky V. Genetic transformation of HeLa cells by *Agrobacterium*. *Proc Natl Acad Sci U S A*. 2001; 98(4):1871–1876. [PubMed: 11172043]
11. Waters VL. Conjugation between bacterial and mammalian cells. *Nat Genet*. 2001; 29(4):375–376. [PubMed: 11726922]
12. Rayamajhi N, Cha SB, Kang ML, Lee SI, Lee HS, Yoo HS. Inter- and intraspecies plasmid-mediated transfer of florfenicol resistance in Enterobacteriaceae isolates from swine. *Appl Environ Microbiol*. 2009; 75(17):5700–5703. [PubMed: 19592530]

13. Wolska KI. Horizontal DNA transfer between bacteria in the environment. *Acta Microbiol Pol.* 2003; 52(3):233–243. [PubMed: 14743976]
14. Frost LS, Ippen-Ihler K, Skurray RA. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev.* 1994; 58(2):162–210. [PubMed: 7915817]
15. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett.* 2003; 224(1):1–15. [PubMed: 12855161]
16. Williams SL, Schildbach JF. TraY and integration host factor oriT binding sites and F conjugal transfer: sequence variations, but not altered spacing, are tolerated. *J Bacteriol.* 2007; 189(10):3813–3823. [PubMed: 17351033]
17. Matson SW, Morton BS. Escherichia coli DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid oriT. *J Biol Chem.* 1991; 266(24):16232–16237. [PubMed: 1651938]
18. Luo Y, Gao Q, Deonier RC. Mutational and physical analysis of F plasmid traY protein binding to oriT. *Mol Microbiol.* 1994; 11(3):459–469. [PubMed: 8152370]
19. Nelson WC, Howard MT, Sherman JA, Matson SW. The traY gene product and integration host factor stimulate Escherichia coli DNA helicase I-catalyzed nicking at the F plasmid oriT. *J Biol Chem.* 1995; 270(47):28374–28380. [PubMed: 7499339]
20. Street LM, Harley MJ, Stern JC, Larkin C, Williams SL, Miller DL, Dohm JA, Rodgers ME, Schildbach JF. Subdomain organization and catalytic residues of the F factor TraI relaxase domain. *Biochim Biophys Acta.* 2003; 1646(1–2):86–99. [PubMed: 12637015]
21. Larkin C, Datta S, Harley MJ, Anderson BJ, Ebie A, Hargreaves V, Schildbach JF. Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. *Structure.* 2005; 13(10):1533–1544. [PubMed: 16216584]
22. Larkin C, Datta S, Nezami A, Dohm JA, Schildbach JF. Crystallization and preliminary X-ray characterization of the relaxase domain of F factor TraI. *Acta Crystallogr D Biol Crystallogr.* 2003; 59(Pt 8):1514–1516. [PubMed: 12876370]
23. Larkin C, Haft RJ, Harley MJ, Traxler B, Schildbach JF. Roles of active site residues and the HUH motif of F plasmid TraI relaxase. *J Biol Chem.* 2007; 282(46):33707–33713. [PubMed: 17890221]
24. Lahue EE, Matson SW. Escherichia coli DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. *J Biol Chem.* 1988; 263(7):3208–3215. [PubMed: 2830275]
25. Matson SW, Ragonese H. The F-plasmid TraI protein contains three functional domains required for conjugative DNA strand transfer. *J Bacteriol.* 2005; 187(2):697–706. [PubMed: 15629940]
26. Haft RJ, Palacios G, Nguyen T, Mally M, Gachelet EG, Zechner EL, Traxler B. General mutagenesis of F plasmid TraI reveals its role in conjugative regulation. *J Bacteriol.* 2006; 188(17):6346–6353. [PubMed: 16923902]
27. Byrd DR, Sampson JK, Ragonese HM, Matson SW. Structure-function analysis of Escherichia coli DNA helicase I reveals non-overlapping transesterase and helicase domains. *J Biol Chem.* 2002; 277(45):42645–42653. [PubMed: 12207019]
28. Datta S, Larkin C, Schildbach JF. Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure.* 2003; 11(11):1369–1379. [PubMed: 14604527]
29. Dostal L, Schildbach JF. Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated. *J Bacteriol.* 2010; 192(14):3620–3628. [PubMed: 20435720]
30. Wright NT, Majumdar A, Schildbach JF. Chemical shift assignments for F-plasmid TraI (381–569). *Biomol NMR Assign.* 2010; 5(1):67–70. [PubMed: 20936510]
31. Lang S, Gruber K, Mihajlovic S, Arnold R, Gruber CJ, Steinlechner S, Jehl MA, Rattei T, Frohlich KU, Zechner EL. Molecular recognition determinants for type IV secretion of diverse families of conjugative relaxases. *Mol Microbiol.* 2010; 78(6):1539–1555. [PubMed: 21143323]
32. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, Waksman G. Structure of the outer membrane complex of a type IV secretion system. *Nature.* 2009; 462(7276):1011–1015. [PubMed: 19946264]
33. Dostal L, Shao S, Schildbach JF. Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. *Nucleic Acids Res.* 2011; 39(7):2658–2670. [PubMed: 21109533]

34. Guogas LM, Kennedy SA, Lee JH, Redinbo MR. A novel fold in the TraI relaxase-helicase c-terminal domain is essential for conjugative DNA transfer. *J Mol Biol.* 2009; 386(2):554–568. [PubMed: 19136009]
35. Bax A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* 2003; 12(1):1–16. [PubMed: 12493823]
36. Wright NT, Margolis JW, Margolis FL, Weber DJ. Refinement of the solution structure of rat olfactory marker protein (OMP). *J Biomol NMR.* 2005; 33(1):63–68. [PubMed: 16222559]
37. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol.* 1993; 231(4):1049–1067. [PubMed: 8515464]
38. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009; 4(3):363–371. [PubMed: 19247286]
39. Dostal L, Schildbach JF. Single-stranded DNA binding by F TraI relaxase and helicase domains are coordinately regulated. *J Bacteriol.* 2010
40. Datta S, Larkin C, Schildbach JF. Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure (Camb).* 2003; 11(11):1369–1379. [PubMed: 14604527]
41. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr.* 1992; 25:495–503.
42. Cheng Y, McNamara DE, Miley MJ, Nash RP, Redinbo MR. Functional characterization of the multi-domain F plasmid TraI relaxase-helicase. *J Biol Chem.* 2011
43. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature.* 2004; 432(7014):187–193. [PubMed: 15538360]
44. Manoil C, Traxler B. Insertion of in-frame sequence tags into proteins using transposons. *Methods.* 2000; 20(1):55–61. [PubMed: 10610804]
45. Wright NT, Varney KM, Ellis KC, Markowitz J, Gitti RK, Zimmer DB, Weber DJ. The three-dimensional solution structure of Ca(2+)-bound S100A1 as determined by NMR spectroscopy. *J Mol Biol.* 2005; 353(2):410–426. [PubMed: 16169012]
46. Muhandiram DR, Guang YX, Kay LE. An enhanced-sensitivity pure absorption gradient 4D ¹⁵N, ¹³C-edited NOESY experiment. *J Biomol NMR.* 1993; 3:463–470.
47. Wright NT, Prosser BL, Varney KM, Zimmer DB, Schneider MF, Weber DJ. S100A1 and calmodulin compete for the same binding site on ryanodine receptor. *J Biol Chem.* 2008; 283(39):26676–26683. [PubMed: 18650434]
48. Kelley, TGA. SPARKY. Vol. 3. San Francisco: University of San Francisco; 2005.
49. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* 2009; 44(4):213–223. [PubMed: 19548092]
50. Schwieters CD, Clore GM. The VMD-XPLOR visualization package for NMR structure refinement. *J Magn Reson.* 2001; 149(2):239–244. [PubMed: 11318623]
51. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson.* 2003; 160(1):65–73. [PubMed: 12565051]
52. Glinka CJ, Barker JG, Hammouda B, Krueger S, Moyer JJ, Orts WJ. The 30 m Small-Angle Neutron Scattering Instruments at the National Institute of Standards and Technology. *J Appl Crystallogr.* 1998; 31:430–445.
53. Kline SR. Reduction and analysis of SANS and USANS data using IGOR Pro. *J Appl Crystallogr.* 2006; 39(6):895–900.
54. Guinier, A.; Fournet, G. *Small Angle Scattering of X-rays.* New York: John Wiley and Sons; 1955.
55. Krueger S, Gorshkova I, Brown J, Hoskins J, McKenney KH, Schwarz FP. Determination of the conformations of cAMP receptor protein and its T127L, S128A mutant with and without cAMP from small angle neutron scattering measurements. *J Biol Chem.* 1998; 273(32):20001–20006. [PubMed: 9685337]
56. Curtis JE, Raghunandan S, Nanda H, Krueger S. SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comp Phys Comm.* 2012; 183:382–389.

57. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996; 14(1): 33–38. 27–38. [PubMed: 8744570]
58. Stern JC, Schildbach JF. DNA recognition by F Factor TraI36: Highly sequence-specific binding of single-stranded DNA. *Biochemistry.* 2001; 40(38):11586–11595. [PubMed: 11560509]
59. Nielsen SS, Toft KN, Snakenborg D, Jeppesen MG, Jacobsen JK, Vestergaard B, Kutter JP, Arleth L. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J Appl Cryst.* 2009; 42:959–964.
60. Svergun D, Barerato C, Koch MHJ. CRY SOL- a Program to Evaluate X-ray Solution Scattering of Bioogical Macromolecules and Atomic Coordinates. *J Appl Cryst.* 1995; 28:768–773.
61. Wuthrich, K. *NMR of Proteins and Nucleic Acids.* New York: Wiley; 1986.
62. Wright NT, Majumdar A, Schildbach JF. Chemical shift assignments for F-plasmid TraI. *Biomol NMR Assign.* 2010:381–569.

Figure 1.

Schematic of F-plasmid TraI protein. Regions of known structure are colored in red (PDB ID 1P4D), blue (current manuscript; PDB ID 2L8B), and orange (PDB ID 3FLD). TSA and TSB denote Translocation Sequences A and B; regions containing sequences necessary for protein transfer to recipient cell. Walker A/B identifies the ATP binding site in the helicase domain. Hatched areas indicate no currently known structure or function. The boundaries of the fragments used in this study are depicted underneath, along with the relative abilities of the fragments to bind an ssDNA *oriT* oligonucleotide or the (nonspecific) G144'C variant. The *oriT* oligonucleotide is bound by both nickase and helicase-associated ssDNA binding sites, while only the latter exhibits nonspecific ssDNA binding.

Figure 2.

NMR data for the TraI fragment (381-569). (A) A plane from the 4D ^{13}C , ^{13}C -edited NOESY-HSQC showing NOE correlations of the $\delta 1$ methyl hydrogen of L426 to medium and long-range resonances. In particular, long-range NOE correlations are visible to A434 in strand 1, M455 in helix 3, I506 in strand 3, and L533, in strand 4. (B) A representative region of the 2D ^1H - ^{15}N HSQC IPAP NMR spectra used to measure residual dipolar coupling data (RDCs). This region illustrates splitting of I534. Dipolar couplings were determined by comparing the splitting in the absence (right) or in the presence (left) of the aligning acrylamide medium.

Figure 3. 3D structure of TraI (390-569). (A) Stereoview of the 20 best structures of TraI (390-569). α -helices are colored in red and β -strands are colored in blue. Random coils and turns are colored in grey. (B) Ribbon diagram of the best TraI (390-569) structure. The region from H2 to H6 alternates between helices and sheets, forming a three-dimensional structure that closely resembles the fold of the N-terminal region of RecD (PDB ID 1W36).

Figure 4.

Structure of the first TraI linking region (306 to 381). (A) HSQC overlay between F-TraI (381-569) in black and F-TraI (309-569) in red. The red HSQC suggests that part of this construct is well-folded (mostly, those whose peaks overlay with (381-569)), while part of this construct is poorly folded or unfolded. (B) Shown is a density plot of the structures that best agree with SANS data collected on TraI (1-330). Residues 1 to 306 of the modified 1P4D crystal structure are represented in red. Using the program SASSIE, 5795 structures were generated where residues 307 to 330 were allowed to move at random. The protein shell of residue 306 to 330 from this simulation is represented in black mesh. The SANS profile of these structures were then back-calculated and compared to actual SANS data of TraI (1-330). The protein shell of residues 306 to 330 from those models that most closely

agreed to the actual SANS data ($\chi^2 < 5$) is represented in green. This shows that residue 307 to 330 does not exist in a well-formed structure in this construct.

Figure 5.

Analysis of SAXS data from TraI (1-569). 8755 models of TraI (1-569)-*oriT* ssDNA were generated using SASSIE, where residues 306 to 389 were allowed to randomly move. These models were then back-calculated to SAXS data collected on (1-569) using CRY SOL. (A) A plot of the χ^2 versus radius of gyration (R_g) shows that models with a R_g value close to 34 Å fit best with the collected SAXS data. Data colored in blue represent those models with a χ^2 value of less than 15. (B) TraI (1-305) is shown in red, with the protein shell of residues 306 to 569 of all models represented in black mesh. Those models with a low χ^2 value (240 models) are depicted in blue. (C) The structural model of 1-569 that most closely agrees with the SAXS data. In this and all other well-agreeing model structures, the nickase (red) and (390-569) (blue) are close to each other in physical space. In green is the 12 bp ssDNA from the crystal structure of nickase-ssDNA, and is the fragment that was bound to the nickase domain in these experiments.

Figure 6. Functional and evolutionary implications of the TraI (381-569) structure. (A) Residues of TraI (381-569) that are conserved with RecD (red) or similar to RecD (yellow). (B) TraI (381-569), highlighting residues where a 31 amino acid insertion causes a > 2000-fold gain of function in bacterial conjugation (orange; Q439, I506, V507, V555) or where bacterial conjugation is roughly equivalent to wild-type levels (green; E517) ²⁶.

Table 1

NMR-derived restraints and statistics of 20 NMR structures²

rmsd from distance constraints (Å)³	<20>	best
total (2204)	0.037 ± .002	0.037
intraresidue (519)	0.001 ± .001	0.000
sequential ($ i-j = 1$) (596)	0.027 ± .002	0.027
medium range ($1 < i-j $) (468)	0.037 ± .005	0.036
long range ($ i-j > 5$) (481)	0.048 ± .003	0.052
hydrogen bonds (140)	0.072 ± .005	0.070
rmsd from exptl dihedral restraints (°)		
ϕ, ψ (241)	0.209 ± 0.105	0.030
rmsd from dipolar coupling restraints (Hz)		
D_{NH} (96)	1.52 ± .08	1.51
rmsd from experimental¹³C chemical shifts		
¹³ Ca(ppm)	1.389 ± 0.039	1.44
¹³ Cβ(ppm)	1.218 ± 0.045	1.3
rmsd from idealized geometry		
bonds (Å)	0.005 ± .001	0.005
angles (°)	0.768 ± 0.017	0.754
impropers (°)	0.713 ± 0.039	0.672
Lennard-Jones potential energy (kcal/mol)⁴	-627 ± 16	-622
Q-factor⁵	27.4 ± 5.4	23.0
R, Pearson corr coeff on all calc and exp RDC	0.98	
% residues in the most favorable region of the Ramachandram plot⁶	78.7 ± 1.6	79.8
rmsd to the mean structure (Å)⁷		
all backbone atoms (390-553)	0.86 ± 0.12	0.67
all heavy atoms (390-553)	1.43 ± 0.11	1.44

²The 20 ensemble structures, <20>, are the results of simulated annealing calculations. The best structure is the closest to the average structure. The values shown for the <20> are the mean ± standard deviation.

³None of the 20 structures has a distance violation > 0.4 Å or a dihedral angle violation of > 5°. The force constants used in the SA calculations are as follows: 1000 kcal mol⁻¹ Å² for bond length, 500 kcal mol⁻¹ rad⁻² for angles and improper torsions, 4 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals (vdw) repulsion term (hard-sphere effective vdw set to 0.8 times their values in CHARMM (Chemistry at HARvard Molecular Mechanics) parameters), 50 kcal mole⁻¹ Å⁻² for experimental distance constraints, 100 kcal mol⁻¹ Å⁻² for non-crystallographic symmetry, 0.5 kcal mol⁻¹ ppm⁻² for the ¹³C chemical shift constraints, and 1.0 for the conformational database potential. The force constants (in kcal Hz⁻²) used for dipolar coupling restraints was 0.50 for ¹⁵N-¹HN.

⁴Lennard-Jones van der Waals energies were calculated using CHARMM parameters and were not used in any stage of the structure determination

⁵Q-values were determined by randomly removing 10% of RDC values. To ensure accuracy, a second ensemble of structures with a second randomly removed subset of RDC values was generated.

⁶PROCHECK was utilized to generate the Ramachandran plot.

⁷Backbone calculations include C^α, N, and C' atoms. Only residues 390 to 553 are included since no secondary structure or long-range NOE correlations were observed for residues 381 to 389 or 554 to 569.

Table 2

Affinities of TraI fragments for the 3'-TAMRA-labeled 22-base G144'C oligonucleotide used previously to measure TraI nonspecific ssDNA binding. All experiments were performed at least 3 times, and the mean and standard deviation of the measurements reported.

TraI fragment	KD (nM)
1-569	1300 ± 800
381-569	> 5500
309-569	1600 ± 1300
309-858	6.9 ± 1.3
330-858	47.0 ± 5.8
381-858	38.6 ± 5.0
309-858 (K331A)	18.7 ± 4.2
309-858 (K318A)	4.7 ± 0.7
309-858 (Q317A)	10.9 ± 1.7