JBC RESEARCH ARTICLE

Characterization of the specific DNA-binding properties of Tnp26, the transposase of insertion sequence IS*26*

Received for publication, July 19, 2021, and in revised form, August 31, 2021 Published, Papers in Press, September 3, 2021,
https://doi.org/10.1016/j.jbc.2021.101165

Carol H. Pong, Christopher J. Harmer, Janine K. Flores, Sandro F. Ataide[‡], and Ruth M. Hall[*,‡]
From the School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia

Edited by Patrick Sung

The bacterial insertion sequence (IS) IS*26* mobilizes and disseminates antibiotic resistance genes. It differs from bacterial IS that have been studied to date as it exclusively forms cointegrates *via* either a copy-in (replicative) or a recently discovered targeted conservative mode. To investigate how the Tnp26 transposase recognizes the 14-bp terminal inverted repeats (TIRs) that bound the IS, amino acids in two domains in the N-terminal (amino acids M1–P56) region were replaced. These changes substantially reduced cointegration in both modes. Tnp26 was purified as a maltose-binding fusion protein and shown to bind specifically to dsDNA fragments that included an IS*26* TIR. However, Tnp26 with an R49A or a W50A substitution in helix 3 of a predicted trihelical helix–turn–helix domain (amino acids I13–R53) or an F4A or F9A substitution replacing the conserved amino acids in a unique disordered N-terminal domain (amino acids M1–D12) did not bind. The N-terminal M1–P56 fragment also bound to the TIR but only at substantially higher concentrations, indicating that other parts of Tnp26 enhance the binding affinity. The binding site was confined to the internal part of the TIR, and a G to T nucleotide substitution in the TGT at positions 6 to 8 of the TIR that is conserved in most IS*26* family members abolished binding of both Tnp26 (M1–M234) and Tnp26 M1–P56 fragment. These findings indicate that the helix–turn–helix and disordered domains of Tnp26 play a role in Tnp26–TIR complex formation. Both domains are conserved in all members of the IS*26* family.

Bacterial insertion sequences (ISs) are mobile genetic elements that play a significant role in generating diversity although the mechanism of action has been described in detail for relatively few (1–3). ISs are genetically compact, and by definition only carry genes and sites required for their mobilization (4). Movement of the IS is catalyzed by the encoded transposase (Tnpase), which has three distinct functions: DNA binding, multimerization, and catalysis (5–7). The majority of IS encode Tnpases that are related to the retroviral integrases. The catalytic core has an RNaseH-like fold and is defined by three characteristic acidic residues: two aspartate (D) and one glutamate (E) (1, 7). ISs that encode an aspartate–aspartate–glutamate (DDE) Tnpase are bounded by terminal inverted repeats (TIRs) that are recognized and acted on by the Tnpase. A number of IS-encoded Tnpases are predicted to possess a helix–turn–helix (HTH) that acts as the DNA-binding domain (DBD) facilitating sequence-specific binding to the TIR (5, 8–14). The TIR can be subdivided into one or more DNA-recognition sites and a cleavage site at the terminus (3, 12). The interaction between an IS and its Tnpase must occur in a highly coordinated manner to form a transpososome and enable movement of the IS (7).

The IS*26* family, defined as the subgroup of the bacterial IS in the IS*6* family most closely related to IS*26* (15), includes the three ISs that play the most significant roles in the evolution of antibiotic resistance, namely IS*26* in Gram-negative bacteria (16–20) and IS*257*/IS*431* and IS*1216* in Gram-positive bacteria (17, 21–25). IS*26*, the best studied member of the IS*26* family, is 820 bp in length, and the *tnp26* gene (705 bp) encodes a 234-amino acid DDE Tnpase, Tnp26 (26). The DDE catalytic residues of the Tnp26 Tnpase have been demonstrated to be necessary for IS*26* movement (27). IS*26* is bounded by 14-bp perfect TIRs. The TIRs are highly conserved across the IS*26* family with GG as the outermost residues, a completely conserved G residue at position 7 flanked by highly conserved T residues, and a highly conserved G residue at position 10 (15). However, the IS*26* mechanism of movement has long been recognized to be different from the mechanism used by IS that are members of most other families.

IS*26* is unable to move from one position to a new one as a discrete unit, that is, move alone, and instead exclusively forms cointegrates in which the two product ISs each remain attached at one end to the original location (28–30). Early studies (26, 29, 30) showed that IS*26* forms cointegrates *via* copy-in cointegration, previously called "replicative transposition." More recent studies have demonstrated that IS*26* also performs an IS*26*-targeted cointegration reaction when a second IS*26* copy is present in a separate DNA molecule or replicon (27). The targeted reaction is conservative and occurs at 100- to 1000-fold higher frequency than the copy-in reaction (27, 31). To date, this unique reaction has only been demonstrated for further members of the IS*26* family (32, 33). The significant difference between the two reaction modes is that both IS ends are involved in the copy-in reaction, whereas the targeted conservative reaction occurs at one end (either left or right) (31, 32). These dual capabilities have significant

‡ These authors have contributed equally to this work.
* For correspondence: Ruth M. Hall, ruth.hall@sydney.edu.au.

implications for the spread of antibiotic resistance genes and genome diversification. However, to date, little is known about the features that distinguish IS*26* and related IS from the IS that have been studied to date, most of which use a cut and paste or a copy-out-paste-in mechanism. Comparison of the sequences of Tnpases that are the closest relatives of IS*26* revealed a short segment at the N terminus that includes two highly conserved phenylalanine (or tyrosine) residues (15). This motif is not seen in the Tnpases encoded by any other known IS family, and this finding may provide a first clue as to what makes IS*26* so different.

Here, we have characterized the interactions between the TIR of IS*26* and Tnp26. First, computational modeling of Tnp26 was performed to visualize the putative functional domains. The effect of introducing alterations to a predicted trihelical HTH domain, the potential TIR recognition and binding domain (DBD) of Tnp26, and replacement of conserved residues in the short N-terminal region that is unique to the Tnp26 Tnpase family were examined both *in vivo* and *in vitro*. The effect of substitution of bases in the IS*26* TIR sequence on Tnp26 binding was also examined, and a completely conserved residue was found to be critical for TIR-specific binding.

Results

A structural model of Tnp26

In a Phyre2 model of Tnp26, 90% of the amino acids (211 of 234) were modeled with over 90% confidence, and a C-terminal DDE catalytic domain, described previously (34), and an N-terminal trihelical bundle connected by a flexible linker were observed (Fig. 1). The two domains were flanked by a short disordered region at the extreme N terminus that includes two phenylalanine residues separated by four amino acids, which are highly conserved in the IS*26* family members (Fig. 2*A*, (15)), and an extended C-terminal tail with no identifiable functional motifs. The DDE catalytic domain modeled closely with several DDE-containing retroviral integrases (35) and HIV-2 integrase (Protein Data Bank [PDB]: 3F9K) with confidence >99.74% (10–15% identity).

The N-terminal region of Tnp26 modeled closely to the trihelical HTH domain of HIV-2 integrase (99.79% confidence and 15% identity). The sequence of the Tnp26 N-terminal region was also separately modeled using Phyre2, and amino acids 5 to 56 modeled with >90% confidence with a trihelical HTH in the OrfA portion of Tnpases encoded by two IS*3* family members, namely IS*629* (PDB: 2RN7; 92.89% confidence and 21% identity) and an IS that is a close relative of IS*1206* ("Cgl2762"; PDB: 2JN6; 96% confidence and 20% identity) (36), as well as the HTH domain of ISCth4 (IS*256* family) Tnpase (PDB: 6XGX; 95.15% confidence and 22% identity (13)).

Cointegration activity of Tnp26 with amino-acid substitutions disrupting the HTH

Conservation of the N terminus of the Tnpase across the IS*26* family is shown in Figure 2*A*. Mutations predicted to introduce changes that disrupt the secondary structure of the

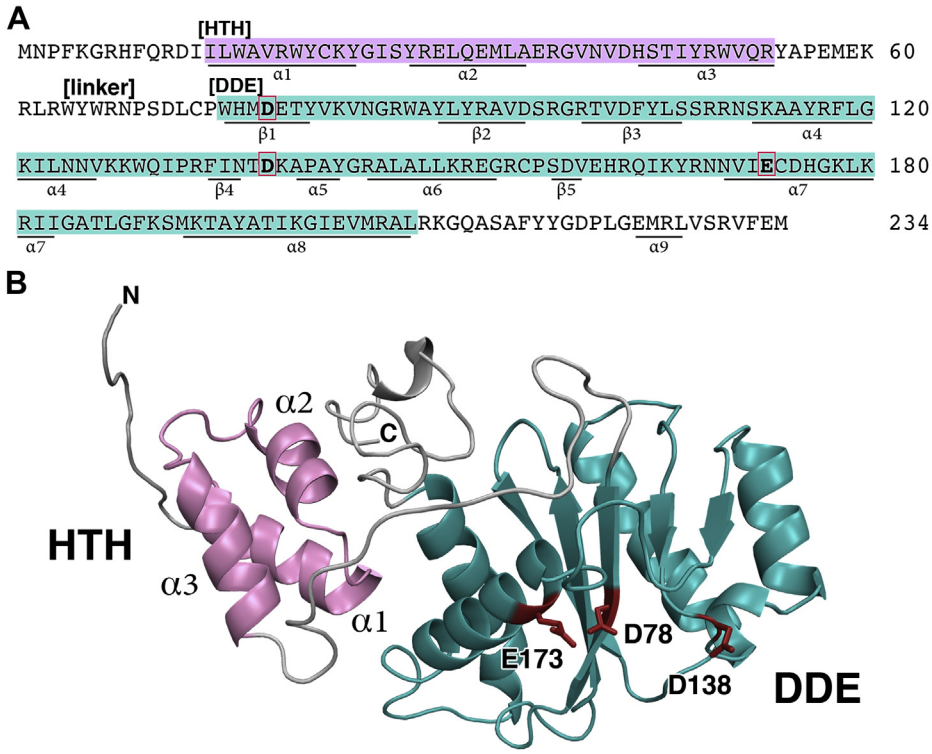

Figure 1. Tnp26 transposase. *A*, Tnp26 sequence (234 amino acids) with predicted secondary structure features of the HTH motif, highlighted in *pink*, and DDE catalytic core, highlighted in *blue*, and predicted α-helices and β-strands *underlined*. The catalytic residues (D78, D138, and E173) are *bold* and *boxed* in *red*. The linker between the two domains is indicated. *B*, Tnp26 modeled by Phyre2 is shown with the HTH (α1–3) and DDE catalytic core colored *pink* and *blue*, respectively. The DDE catalytic residues are colored *maroon* and labeled. DDE, aspartate–aspartate–glutamate; HTH, helix–turn–helix.

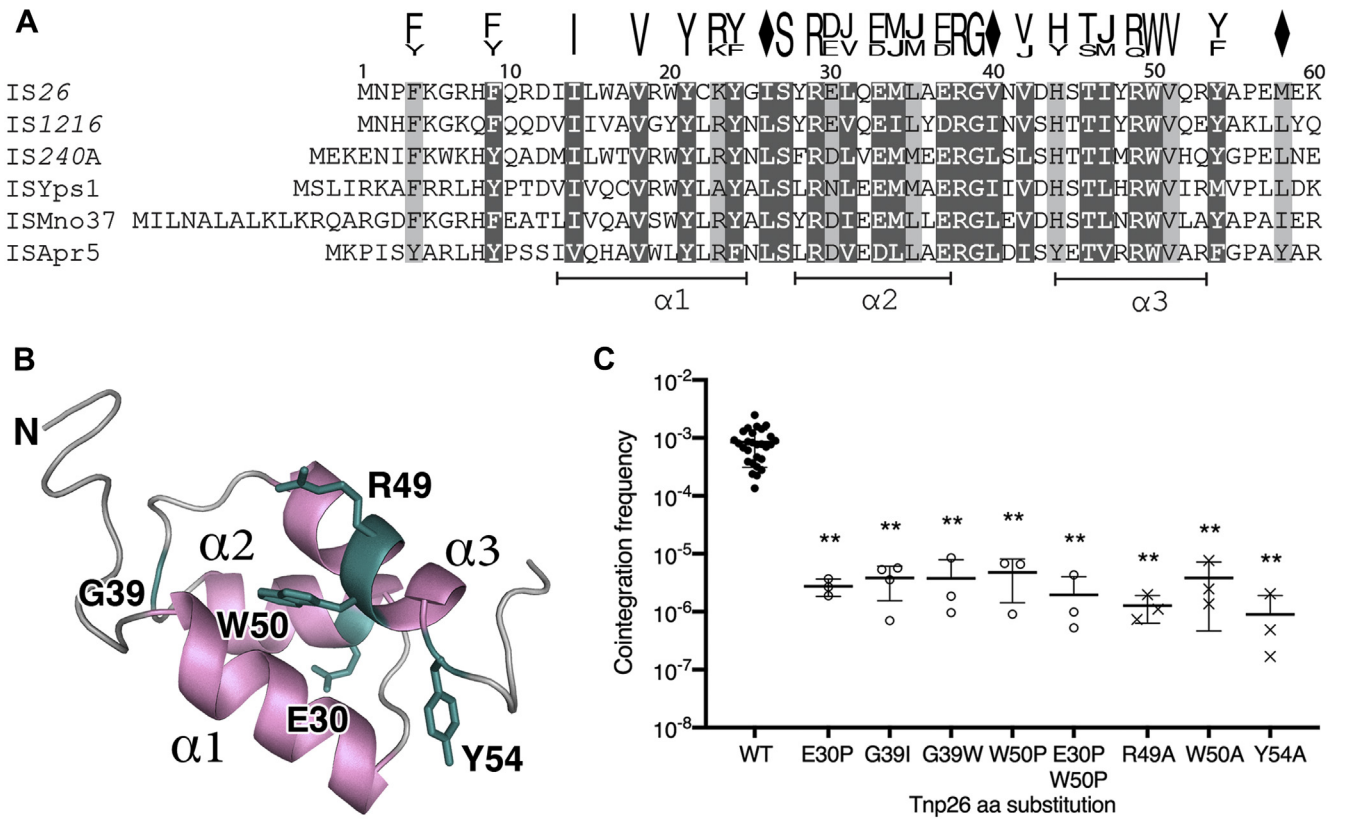

Figure 2. Properties of the N-terminal region of Tnp26. *A*, alignment showing conservation in amino acids M1–K60 of Tnp26. A single representative of each clade (I–VI) is shown with amino acid conserved in ≥55, highlighted in *light gray*, or ≥60, highlighted in *dark gray*, of the 65 curated sequences identified as members of the IS*26* family (15). The consensus sequence is shown above the alignment, with a *diamond symbol* (◆) representing amino acids I, L, M, and V; and J representing amino acids I and L. Where alternate residues are found, the font size indicates their approximate relative abundance. For full alignment, see Ref. (15). *B*, model of Tnp26 M1–K60 showing positions of residues targeted for substitution, E30, G39, R49, W50, and Y54, shown in *teal*. α1–3 and residues are labeled. *C*, frequencies of cointegration between pRMH977 derivatives and the corresponding R388::IS*26* derivative, encoding Tnp26 (WT; "•" *black filled circles*) or Tnp26 with amino acids substitutions E30P, G39I, G39I/W, W50P, E30PW50P ("○" *open circles*), R49A, W50A, or Y54A ("×" *crosses*). Cointegration frequencies are expressed as the mean ± SD (n ≥ 3). Significance (**$p < 0.01$) was determined using a one-way ANOVA and Šidák multiple comparison test. IS, insertion sequence.

conserved HTH fold were introduced into IS*26* in plasmid pRMH977 (Ap^R), which carries one IS*26* copy. Five pRMH977 derivatives encoding Tnp26 derivatives with the substitutions E30P, G39I, G39W, W50P, or E30PW50P (see Fig. 2*B* and Table 1 for location in Tnp26) were generated and tested in mating-out assays to measure cointegrate formation with the conjugative plasmid R388 (Tp^R) by the low-frequency untargeted copy-in mechanism. WT IS*26* in pRMH977 (Ap^R) formed cointegrates (Ap^RTp^R) with R388 (Tp^R) at a frequency of $1.8 \pm 1.6 \times 10^{-7}$ cointegrates per transconjugant (mean of n = 12). However, cointegrate formation between the pRMH977 derivatives and R388 was not detected with a limit of detection of $<1.3 \times 10^{-10}$ to $<4.9 \times 10^{-10}$ cointegrates per transconjugant (Table 1), corresponding to <0.1% to <0.3% of WT activity, and indicating that the variants may be unable to support cointegrate formation *via* the copy-in route.

Table 1
Copy-in cointegration between R388 and pRMH977 derivatives encoding Tnp26 with amino-acid substitutions in the HTH

Tnp26 amino-acid substitution	Location	No.[a]	Cointegration frequency (cointegrate/transconjugant)	
			Range	Mean ± SD/detection limit[b]
None	—	12	1.13×10^{-8}–5.09×10^{-7}	$1.8 \pm 1.6 \times 10^{-7}$
E30P	α2	3	$<2.10 \times 10^{-10}$–$<1.37 \times 10^{-9}$	$<1.3 \times 10^{-10}$
G39I	Turn	3	$<4.83 \times 10^{-10}$–$<8.47 \times 10^{-9}$	$<3.9 \times 10^{-10}$
G39W	Turn	3	$<4.42 \times 10^{-10}$–$<5.05 \times 10^{-10}$	$<1.6 \times 10^{-10}$
W50P	α3	3	$<3.92 \times 10^{-10}$–$<1.85 \times 10^{-9}$	$<2.3 \times 10^{-10}$
E30PW50P	α2&α3	3	$<7.52 \times 10^{-10}$–$<2.86 \times 10^{-9}$	$<4.9 \times 10^{-10}$
R49A	α3	3	$<6.49 \times 10^{-9}$–$<9.80 \times 10^{-9}$	$<2.6 \times 10^{-9}$
W50A	α3	3	$<2.11 \times 10^{-9}$–$<3.31 \times 10^{-9}$	$<8.2 \times 10^{-10}$
Y54A	α3	3	$<2.44 \times 10^{-9}$–$<9.26 \times 10^{-9}$	$<1.4 \times 10^{-9}$

[a] Number of independent determinations.
[b] Mean calculated when cointegrates were detected. When no cointegrates were detected, the total number of $Ap^SSm^RTp^R$ transconjugants was used to calculate the limit of detection.

Targeted conservative cointegration was also examined by introducing the IS*26* variants into R388::IS*26* and measuring cointegrate formation between each pRMH977 derivative and the corresponding R388::IS*26* derivative (Fig. 2*C*). WT IS*26* (pRMH977 with R388::IS*26*) formed cointegrates at a frequency of 8.4 ± 5.3 × 10^{-4} cointegrates per transconjugant (mean ± SD, n = 28). The cointegration frequency of the pRMH977 derivatives and the corresponding R388::IS*26* derivatives (Fig. 2*C*, *open circles*) was over two orders of magnitude lower ($p < 0.1$, n ≥ 3). The potential cointegrates detected were carefully analyzed and shown to be true cointegrates with the relevant mutation in both copies of the IS indicating that sufficient folding to support low-level activity had occurred.

Taken together, these *in vivo* findings indicate that the disruptions to the secondary structure of the HTH fold had substantial impacts on Tnp26 activity *in vivo*.

Cointegration activity of Tnp26 with amino-acid changes to α3

Residues that have potential DNA-binding properties and are conserved in the IS*26* family (Fig. 2*A*) and may have a direct role in specific binding of Tnp26 to the TIRs were also examined. Residues R49 and W50 in α3, and Y54 at the end of α3, were replaced with alanine. Copy-in cointegration for pRMH977 derivatives encoding Tnp26 R49A, Tnp26 W50A, or Tnp26 Y54A was assessed, and cointegrate formation between these pRMH977 derivatives and R388 was not detected (Table 1). The cointegration frequency of these pRMH977 derivatives and the corresponding R388::IS*26* derivative (Fig. 2*C*, *crosses*) was reduced by over two orders of magnitude ($p < 0.1$, n = 3), relative to the WT cointegration frequency. Together, these results demonstrate that the modifications to helix 3 of the trihelical HTH of Tnp26 reduced activity to <1% of WT activity. In these cases, the residual activity may be

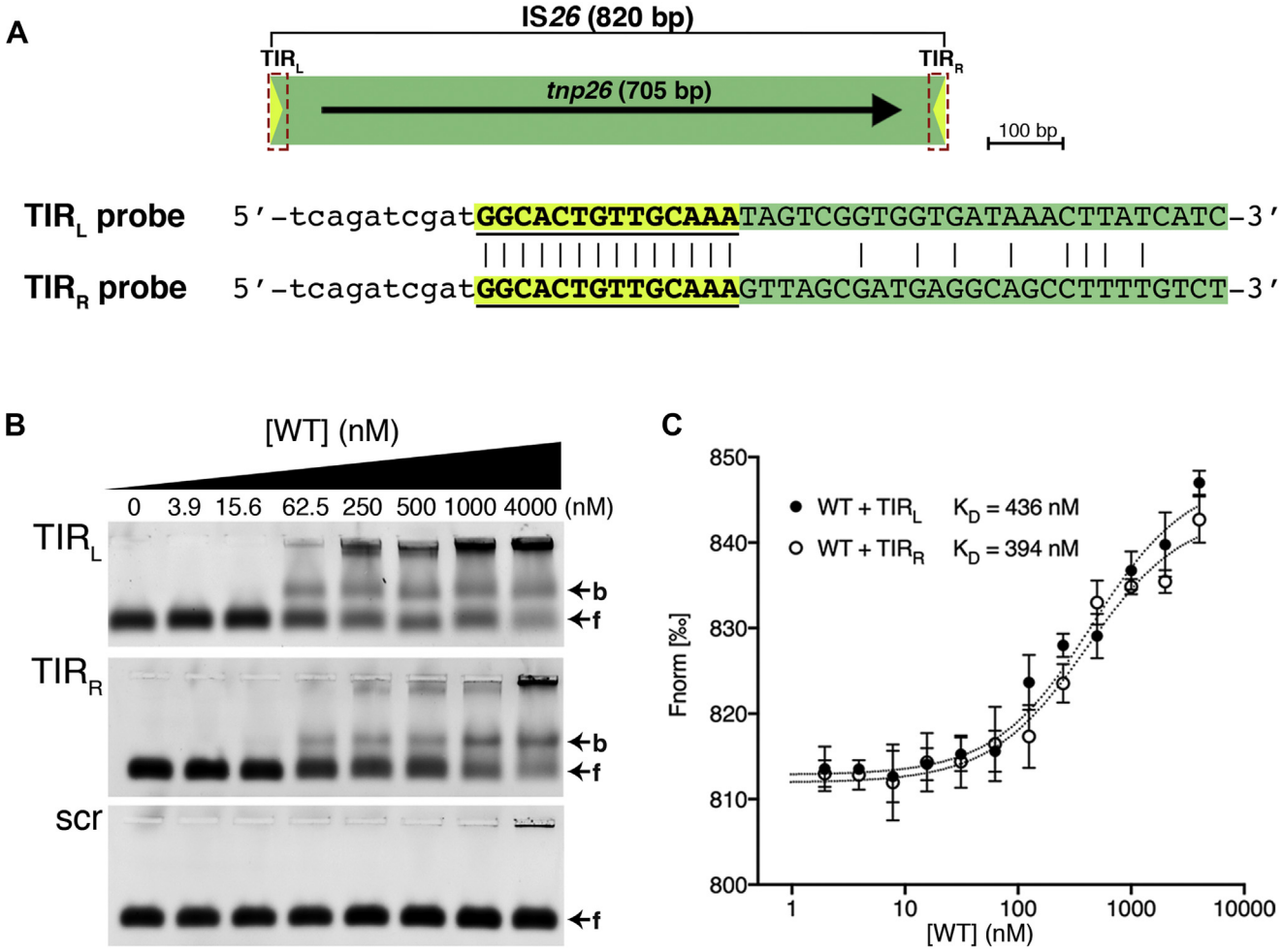

Figure 3. Binding of Tnp26 WT to IS*26* TIR sequence. *A*, schematic of IS*26* and sequence of one strand of the Cy5-labeled 50-bp TIR_L and TIR_R probes. The extent and orientation of *tnp26* is shown by the *black arrow*, and the TIRs are represented by *yellow arrows* and *boxed* in *red*. In the probe, bases not derived from the IS are in *lower case*. IS*26* sequence is in *upper case*, with the 14-bp TIRs highlighted in *yellow* and the internal sequence in *green*. *B*, EMSAs of Cy5-labeled DNA probes containing TIR_L, TIR_R, or a scrambled version of TIR_R (5′-Cy5-TTCGTGACTACATCGATCTAGGCTAGCTAGTCAGCTGATCGACTAGCGAT-3′), titrated with increasing concentrations of Tnp26 WT. Representative gels from n = 3 are shown. The *arrows* labeled "f" and "b" indicate free DNA and protein–DNA complexes, respectively. *C*, MST binding curve of TIR_L or TIR_R titrated with Tnp26 WT (n = 3) at protein concentrations up to 4000 nM. IS, insertion sequence; MST, microscale thermophoresis; TIR, terminal inverted repeat; TIR_L, left terminal inverted repeat; TIR_R, right terminal inverted repeat.

explained by the fact that more than one interaction is expected to contribute to specific binding to the TIR.

Tnp26 binds to the IS26 TIRs in a sequence-specific manner

In order to directly examine binding of Tnp26 to the IS*26* TIR sequence, the monomeric form of N-terminally His_6-maltose-binding protein (MBP)–tagged full-length Tnp26 (His_6-MBP-Tnp26 WT, hereafter "Tnp26 WT") was purified as described in the Experimental procedures section (Fig. S1). Purified Tnp26 WT was used in EMSA to test for sequence-specific binding to Cy5-labeled 50-bp dsDNA probes composed of 40 bp from either the left or the right IS*26* end with an additional 10 bp adjacent to the outer end (designated "TIR_L" or "TIR_R" DNA probe, respectively; Fig. 3*A*). A scrambled DNA sequence with the same base composition as the TIR_R probe was used to control for nonspecific binding. A fixed concentration of the DNA probe (25 nM) was titrated with Tnp26 WT (0–4000 nM) in the presence of DNA-competitor poly-d(I-C) to eliminate nonspecific binding.

Binding of Tnp26 WT to TIR_L and TIR_R was observed at protein concentrations of 15.6 nM or 62.5 nM and above, as demonstrated by the appearance of a single shifted band migrating above the free DNA probe and a sequence-specific aggregate that did not enter the gel (Fig. 3*B* and Fig. S2). Binding to scrambled DNA was not detected at any Tnp26 WT protein concentration tested (Fig. 3*B*) although some nonspecific interactions of binding components was typically observed at 4000 nM protein. Therefore, protein concentrations above 4000 nM were not examined. Shorter Cy5-labeled DNA probes 25, 30, 35, and 40 bp in length (including the additional 10 bp at the 5′ end) were also examined (Fig. S2). As the binding interaction decreased in stability as the probe length was shortened and no binding to the 25-bp probe was detected, 50-bp dsDNA probes were used hereafter.

Using microscale thermophoresis (MST), 50 bp TIR_L and TIR_R DNA probes were titrated with Tnp26 WT (Fig. 3*C*), and the dissociation constant (K_D) was determined to be 436 (n = 3) and 394 nM (n = 3), respectively. No binding to the scrambled DNA probe was detected. Together, these results indicate that Tnp26 WT recognizes and binds to the IS*26* 14-bp TIR sequence, which is the only IS-derived segment shared by the TIR_L and TIR_R probes (Fig. 3*A*).

Modifications to α3 impact Tnp26 binding

Three His_6-MBP-Tnp26 variants with amino-acid changes in or near α3, R49A, W50A, and Y54A (hereafter, Tnp26 R49A, Tnp26 W50A, and Tnp26 Y54A, respectively) were purified and tested for binding to the IS*26* TIR probes. In EMSAs conducted as described above, binding of Tnp26 Y54A to TIR_L and TIR_R was detected when protein concentrations reached 62.5 nM, although the band intensity of the Tnp26 Y54A–TIR complex was lower than with Tnp26 WT (Fig. 4*A*). Tnp26 Y54A binding to TIR_L and TIR_R was quantified by MST, and the K_D was determined to be 802 and 821 nM, respectively (Fig. 4*A*), approximately twofold higher than WT. Hence, Tnp26 Y54A is able to bind to the IS*26* TIRs in a sequence-specific manner but at a slightly lower affinity compared with WT. This contrasts with the *in vivo* data where Tnp26 Y54A cointegration activity was reduced to <1% WT (Fig. 2*C*), suggesting that although Y54 may not have a substantial role in TIR recognition, it is important for Tnp26 activity.

No binding of Tnp26 R49A and Tnp26 W50A to TIR_L, TIR_R, or the scrambled probe was detected with protein concentrations up to 4000 nM (Fig. 4, *B* and *C*), indicating that replacement of R49 or W50 with an A residue interfered with the DNA-binding ability of Tnp26. These results indicate that the HTH is the Tnp26 DBD.

Tnp26 M1-P56 is sufficient for TIR sequence recognition

Two truncation derivatives of Tnp26 fused to His_6-MBP were generated to assess the DNA-binding properties of the N-terminal region. The first derivative spans Tnp26 M1–P56, hereafter $Tnp26_{1–56}$, and contains the entire predicted HTH motif (I13–R53). Tnp26 M1–D71, hereafter $Tnp26_{1–71}$, included the "linker" between the HTH and catalytic domains.

$Tnp26_{1–56}$ was used in EMSAs to test for binding to the IS*26* TIR_L, TIR_R, or scrambled DNA probes as described above. In contrast to the full-length Tnp26, no $Tnp26_{1–56}$ binding was detected up to protein concentrations of 4000 nM (not shown). However, as no aggregation of binding components was detected in this concentration range, the titration was extended up to 40 μM of $Tnp26_{1–56}$. A shift in the TIR_L and TIR_R probes was observed when the protein concentration reached 5 to 10 μM (Fig. 4*D*), notably higher than the concentration required for aggregation or binding to be detected for full-length Tnp26 (∼62.5 nM). The fluorescence of the shifted band was also less pronounced in comparison to the full-length Tnp26, and the majority of the DNA probe remained unbound in the protein concentration range tested. As no binding to the scrambled DNA probe was detected, the binding was specific for the TIR sequence. The K_D for these interactions was unable to be quantified with confidence by MST because the binding interaction did not reach saturation at the highest protein concentration.

The $Tnp26_{1–71}$ derivative was incubated with TIR_R probe, and aggregation of reaction components was detected when $Tnp26_{1–71}$ reached 250 to 500 nM and continued to accumulate as the protein concentration increased (not shown). However, no distinct shifted band was observed, and we were unable to obtain evidence that the $Tnp26_{1–71}$–DNA aggregate was sequence specific. The behavior of the $Tnp26_{1–71}$ derivative suggests that the presence of the linker may be interfering with DNA binding or correct protein folding.

These results indicate that Tnp26 M1–P56 containing the predicted HTH domain of Tnp26 is sufficient for sequence recognition and binding to the IS*26* TIR sequence. However, the remainder of the protein, containing the DDE catalytic domain, likely plays a role in increasing the affinity of binding between Tnp26 and the IS*26* TIRs.

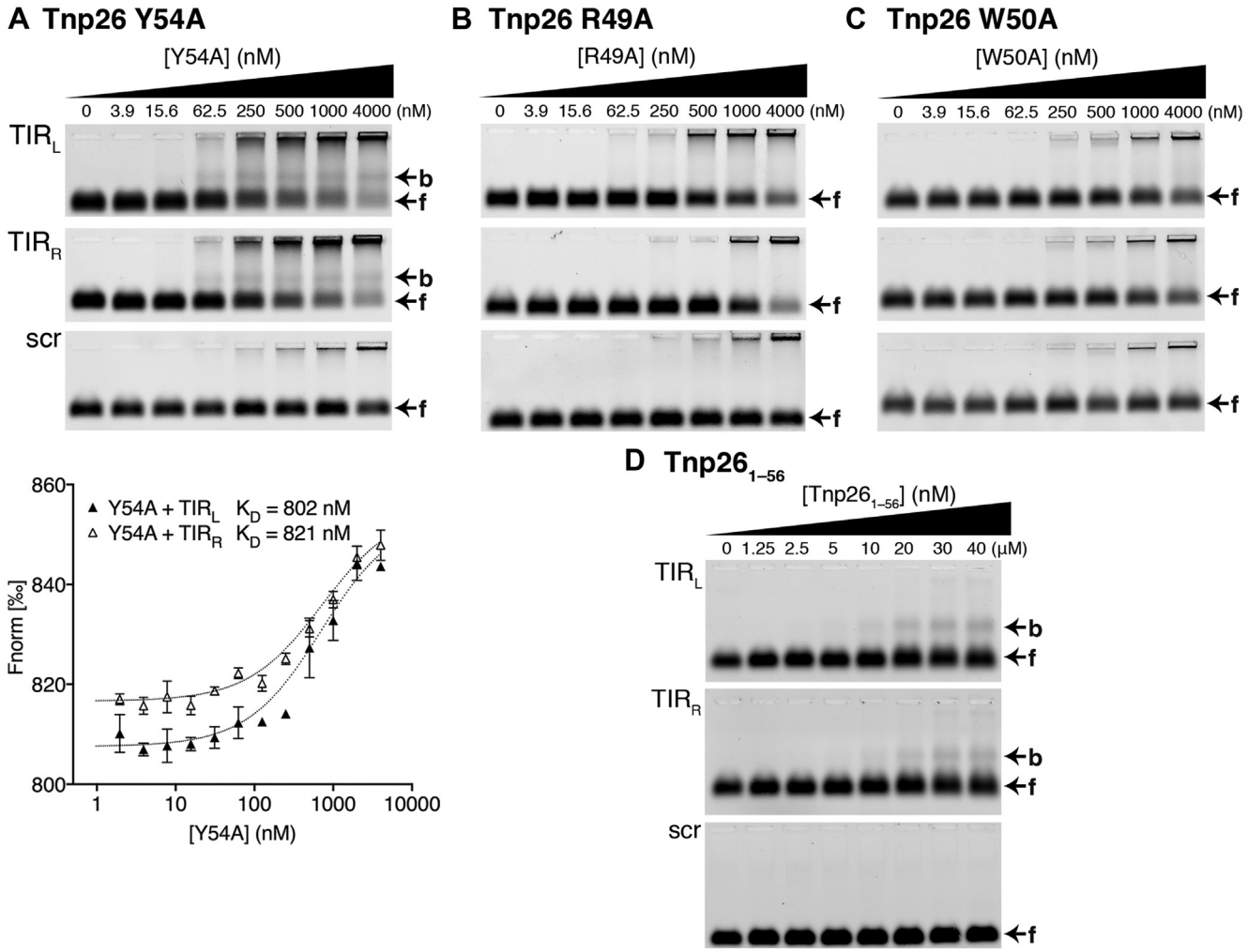

Figure 4. Binding of Tnp26 modified in α3 and Tnp26$_{1–56}$ to IS*26* TIR probes. *A*, EMSAs of Cy5-labeled dsDNA probes containing TIR$_L$, TIR$_R$, or scrambled titrated with Tnp26 Y54A (n = 2; *top*) and the MST-binding curves of TIR$_L$ or TIR$_R$ titrated with up to 4000 nM Tnp26 Y54A (n = 3; *bottom*). EMSAs of Tnp26 R49A (*B*; n = 2), Tnp26 W50A (*C*; n = 2), and Tnp26$_{1–56}$ (*D*; n = 3) titrated against TIR$_L$, TIR$_R$, and scrambled. Representative gels are shown. The *arrows* labeled "f" and "b" indicate free DNA and protein–DNA complexes, respectively. IS, insertion sequence; TIR, terminal inverted repeat; TIR$_L$, left terminal inverted repeat; TIR$_R$, right terminal inverted repeat.

Replacement of conserved bases in TIR$_R$ affects cointegrate formation

An alignment of the TIR sequences of a curated set of 65 IS*26* family members showed high conservation of 14 bases (15). Bases 1 to 2 (GG) and 7G are completely or almost completely conserved, and 6 to 8 (TGT) and 10G are highly conserved (15). At the left IS*26* end, 10G forms part of the –35 box of the *tnp26* promoter sequence, but it is also conserved at the right end. The effect of replacing highly or completely conserved G bases in the right-hand IS*26* TIR with T was examined. The transversion of the terminal two G bases was previously found to abolish IS*26* copy-in cointegration (31). Additional derivatives of pRMH977 with transversions at bases 7 (G > T), 10 (G > T), or 6 to 8 (TGT > GTG) in the IS*26* TIR$_R$ were generated, and their effect on copy-in cointegration with R388, which requires both IS ends, was investigated. No cointegrates of any of these pRMH977 derivatives with R388 were detected (detection limit, $<9.2 \times 10^{-11}$–$<1.4 \times 10^{-10}$ cointegrates/transconjugant; Table S1), a reduction of at least three orders of magnitude indicating that these highly conserved bases are important. Further pRMH977 derivatives with transversions introduced at the remaining bases of the IS*26* TIR$_R$ three bases at a time (bases 3–5, 9–11, and 12–14) were generated, and no cointegrates were detected when these derivatives were assayed for copy-in cointegration with R388 (detection limit, $<1.5 \times 10^{-10}$–$<2.2 \times 10^{-10}$ cointegrates/transconjugant; Table S1).

Tnp26 binding to TIR$_R$ with modified bases

To assess the involvement of the conserved bases in the IS*26* TIRs in Tnp26 binding, three derivatives of the IS*26* TIR$_R$ 50-bp DNA probe were generated, containing transversions of the terminal GG of the TIR (TIR$_R$ 1G > T 2G > T), base 7G (TIR$_R$ 7G > T), and base 10G (TIR$_R$ 10G > T) (Fig. 5*A*). Protein–DNA binding assays were conducted with Tnp26 WT as described previously. When Tnp26 WT (up to 4000 nM) was incubated with TIR$_R$ 1G > T 2G > T probe, a shift of the DNA probe was detected (Fig. 5*B*) when the protein concentration

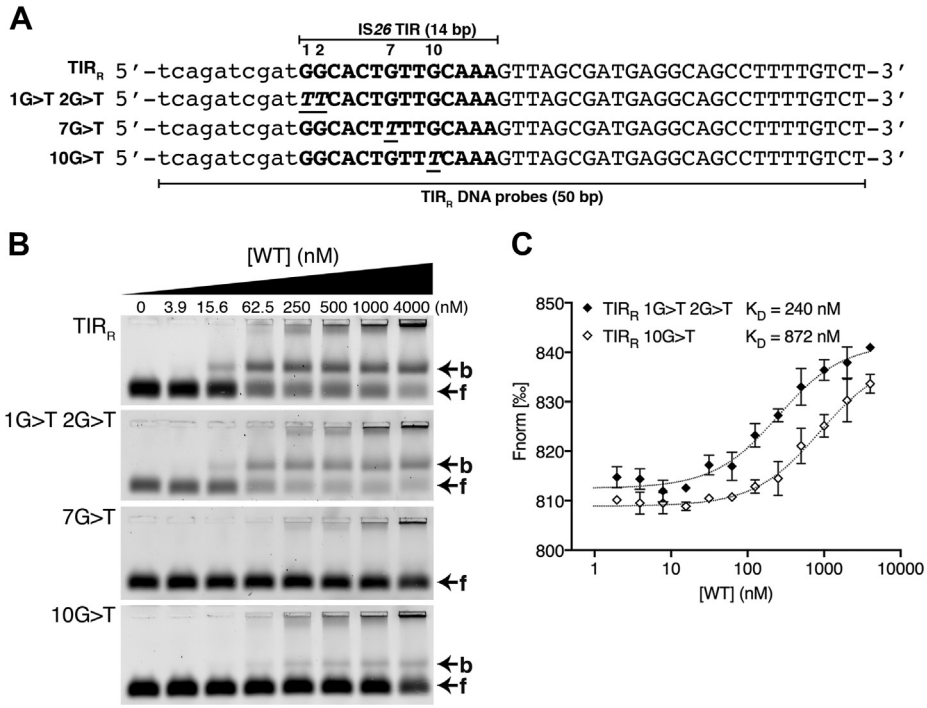

Figure 5. Binding of Tnp26 WT to IS*26* TIR$_R$ with altered bases. *A*, the top strand of Cy5-labeled DNA probes containing TIR$_R$ or TIR$_R$ sequence with base changes, indicated by *italicized* and *underlined text*. Protein binding to DNA probes titrated with Tnp26 WT examined *via B*, EMSA (n = 2; representative gels shown) and *C*, MST (n = 3). Error bars represent SD. IS, insertion sequence; MST, microscale thermophoresis; TIR, terminal inverted repeat; TIR$_L$, left terminal inverted repeat; TIR$_R$, right terminal inverted repeat.

reached 15.6 nM and the K_D determined by MST was 240 nM (Fig. 5*C*), indicating that the binding of Tnp26 WT to TIR$_R$ 1G > T 2G > T was tighter than with TIR$_R$. Hence, the terminal bases do not play a significant role in Tnp26 sequence recognition, consistent with the expectation that the Tnp26 DNA-recognition site would not include the cleavage site at the IS*26* boundaries.

Binding of Tnp26 WT to TIR$_R$ 10G > T probe was detected when the protein concentration reached 62.5 nM (Fig. 5*B*), although the intensity of the Tnp26 WT-TIR$_R$ 10G > T complex was lower than with Tnp26 WT-TIR$_R$ (Fig. 3*C*). The K_D of Tnp26 WT binding to TIR$_R$ 10G > T quantified by MST was 872 nM, approximately twofold higher than Tnp26 WT-TIR$_R$. Hence, 10G may contribute to but is not essential for TIR recognition.

However, binding of Tnp26 WT to the TIR$_R$ 7G > T probe was not detected in EMSAs (Fig. 5*B*) or by MST (not shown). Furthermore, no binding between Tnp26$_{1-56}$ and TIR$_R$ 7G > T probe was detected (not shown) under the conditions where Tnp26$_{1-56}$–TIR binding (Fig. 4*D*) was observed. Hence, 7G of the IS*26* TIRs forms an essential part of the Tnp26 binding site.

Role of the two conserved phenylalanine residues at the Tnp26 N terminus

Aside from the HTH DBD and DDE catalytic domain, no distinct motif or folded structure was identified in Tnp26 by *in silico* modeling or secondary structure prediction tools. However, the high conservation of two aromatic residues, usually phenylalanine or alternatively tyrosine, with four amino-acid spacing at the extreme N terminus of the IS*26* family Tnpases (Fig. 2*A*, (15)) is a feature unique to this group. Therefore, the roles of these residues at F4 and F9 in Tnp26 in IS*26*-mediated cointegrate formation were investigated.

Copy-in cointegration for pRMH977 derivatives encoding Tnp26 F4A, Tnp26 F9A, and Tnp26 F4AF9A was measured, and no cointegrates between the pRMH977 derivatives and R388 were detected (detection limit, $<1.1 \times 10^{-10}$–$<1.8 \times 10^{-10}$ cointegrates/transconjugant; Table S2) representing a reduction of at least three orders of magnitude. Targeted conservative cointegration between pRMH977 derivatives and R388::IS*26* derivatives with the corresponding mutant (*i.e.*, both plasmids encode either Tnp26 F4A, Tnp26 F9A, or Tnp26 F4AF9A) was also reduced (Fig. 6*A*). When plasmids encoded Tnp26 F4A or Tnp26 F9A, the cointegration activity was significantly reduced to <1% of WT activity ($p < 0.1$, n = 3; Table S3). However, no cointegrates were detected when the plasmids encoded Tnp26 F4AF9A corresponding to a reduction of over three orders of magnitude, indicating these residues are required for Tnp26 activity. These results indicate F4 and F9 play an important role in cointegrate formation.

We considered the possibility that the phenylalanine residues from different Tnp26 molecules may interact by π–π interactions (37) and be involved in dimer formation. Therefore, targeted conservative cointegration between plasmids encoding different variants (*i.e.*, Tnp26 F4A encoded by one plasmid and Tnp26 F9A encoded by the other) was assessed, and the activity was further reduced to <0.1% of WT ($p < 0.1$, n = 3; Table S3 and Fig. 6*A*). The larger reduction in IS*26* cointegration when Tnp26 F4A and Tnp26 F9A are in the same reaction suggests that the

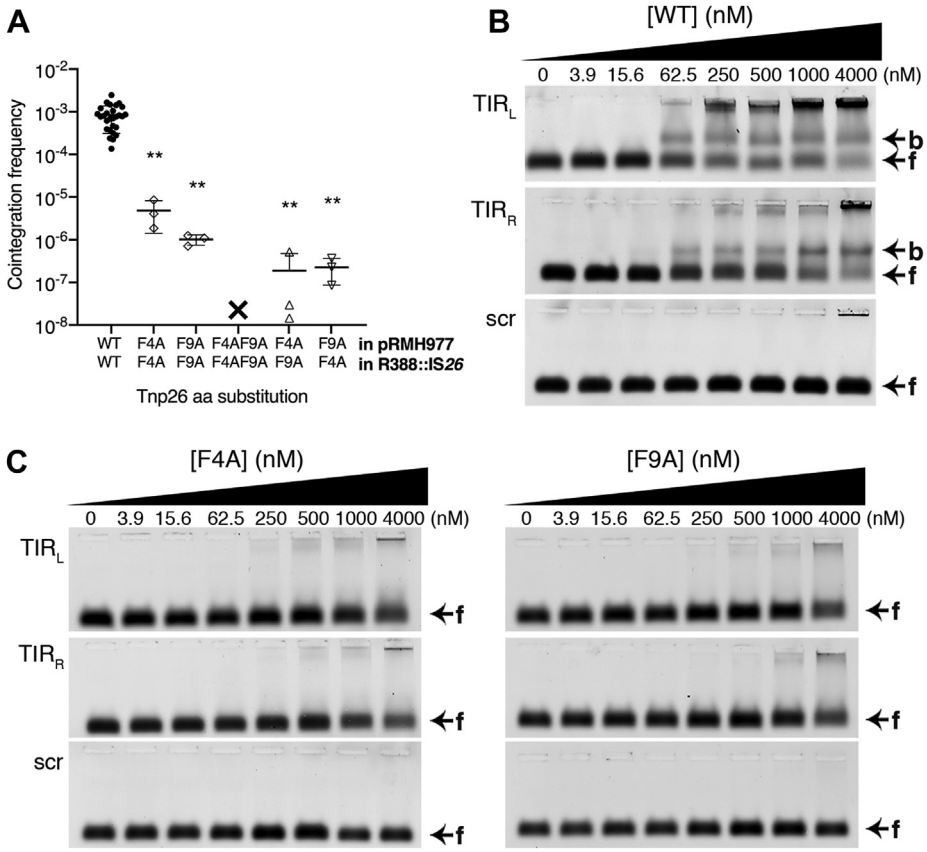

Figure 6. Activity of Tnp26 with F4A and/or F9A changes. *A*, frequencies of cointegration between pRMH977 derivatives and R388::IS*26* derivatives, encoding Tnp26 with F4A or F9A substitutions ("◇" *open diamond symbol*) compared with the frequency for pRMH977 with R388::IS*26* (WT; "●" *black filled circles*; included here for comparison and is as for Fig. 2*C*). The cross "×" indicates that no cointegrates were detected. Cointegration frequencies are expressed as the mean ± SD. Significance (**$p < 0.01$, n ≥ 3) was determined using a one-way ANOVA and Šidák multiple comparison test. Binding to TIR_L, TIR_R, or scrambled DNA probes. *B*, Tnp26 WT and *C*, Tnp26 F4A (*left*) or Tnp26 F9A (*right*) examined *via* EMSA (n = 2; representative gels shown). Panel *B* is from Figure 3*B*. and is reproduced here to facilitate comparison with *C*. IS, insertion sequence; TIR_L, left terminal inverted repeat; TIR_R, right terminal inverted repeat.

phenylalanine residues may form two interacting pairs (F4 with F4 and F9 with F9) across different Tnp26 molecules and the absence of one of the paired F residues removes this interaction. Though there is evidence that Tnp26 preferentially acts in *cis* (31), the residual cointegration activity detected here could be due to the Tnp26 variants acting in *trans*, such that two of the same Tnp26 variants are able to form one F-pair and enables some cointegrate formation

Alterations to the Tnp26 N terminus affect binding to the TIR

To investigate the impact of the altered F residues on end recognition, two additional His_6–MBP–Tnp26 fusions, Tnp26 F4A and Tnp26 F9A, were generated. No binding of Tnp26 F4A or Tnp26 F9A to TIR_L, TIR_R, or the scrambled probe was detected in EMSAs at protein concentrations between 0 and 4000 nM (Fig. 6*C*), indicating that replacement of either phenylalanine residue interfered with end recognition and binding by Tnp26. Hence, the two conserved phenylalanine residues are important for both Tnp26 activity and binding to the TIR, though their exact role remains to be established.

Discussion

The analyses presented here demonstrated that the N-terminal domain of Tnp26 (M1–P56), which contains an HTH domain and a short N-terminal extension that includes 2F residues, is important for recognition of the IS*26* TIR sequence by Tnp26. The WT Tnp26 bound to probes containing TIR_L or TIR_R but not to a scrambled probe or to the TIR_R probe with the conserved G residue at position 7 in the TIR replaced by T. Replacement of this G residue with a T also abolished copy-in cointegrate formation. Hence, the observed Tnp26 WT DNA binding was TIR specific. Introduction of various amino-acid substitutions into the Tnp26 HTH reduced *in vivo* cointegration activity to <1% of WT activity, and Tnp26 proteins with an A residue replacing R49 or W50 in α3 of the HTH did not bind to the TIR_R probe, consistent with a significant role for the HTH domain in TIR recognition and binding. However, TIR binding is clearly more complex as the Tnp26 M1–P56 bound the same probes only weakly suggesting that other parts of Tnp26 such as the catalytic core contribute to binding affinity.

In addition, replacement of the F4 or the F9 residue with A also substantially reduced cointegrate formation *in vivo* and TIR binding *in vitro*. The presence and spacing of two planar aromatic residues (F or Y) in the short N-terminal region that precedes the HTH is highly conserved in the IS*26* family (Fig. 2*A* and (15)), but we were unable to find further examples of this motif in disordered regions of other IS families. Hence, this motif may be a determinant of the unique mechanistic

features of IS*26* that lead to the inability to move as a discrete unit and instead only form cointegrates and to the ability to target a second copy of the IS in a conservative reaction (27, 38). The finding that the targeted conservative reaction was reduced by over four orders of magnitude to below the limit of detection when both F4 and F9 were substituted, the largest reduction seen here, speaks to the importance of these residues or this reaction. The *in vivo* data indicate that F4 interacts with F4, and F9 with F9, an observation that could potentially indicate a role in multimerization. Multimerization is a necessary step in transpositional movement that brings the reacting entities, the IS ends, and the target in the copy-in mode and two like ends in the targeted conservative mode (31, 32), together to form the synaptic complex or transpososome. However, further work will be needed to determine if these interactions are important for multimer formation, or if these residues are also directly involved in TIR recognition and binding.

Some interesting parallels can be drawn from the recently published analysis of the structure of the ISCth4 Tnpase bound to the TIR of the IS (13). This Tnpase includes two DBDs that each recognize part of a longer TIR and the HTH of Tnp26 modeled on the HTH domain of the ISCth4 Tnpase. The TIR of ISCth4 (TIR_R: 5′-**GGCAG**<u>**TGT**</u>AAATA-3′, TIR_L: 5′-**G**AGA**G**<u>**TGT**</u>AAAAT-3′) includes a subregion with some similarities to the TIR of IS*26* (TIR_L and TIR_R: 5′-**GGCAC**<u>**TGT**</u>TGCAAA-3′), and Tnpase binding to this region, which is the region bound by the HTH, induces a bend centered on the TGT. Most of the amino acids in the HTH of the ISCth4 Tnpase that interact with the TIR interact with the backbone phosphates. Several of these amino acids correlate with conserved residues in the Tnp26 HTH found in helices 1, 2, and 3, including R49 (K148 in ISCth4 Tnpase). The R127 near the beginning of α2 of the ISCth4 Tnpase HTH that interacts directly with the GT bases of the TGT at positions 6 to 8 in the TIR is present in Tnp26 (R29 in Fig. 2*A*) and conserved in the Tnpase of IS*26* family members (Fig. 2*A*; (15)). In addition, several amino acids near the first D or the E of the DDE motif in the catalytic domain of ISCth4 Tnpase interact with phosphates in the backbone of this portion of the TIR, and if this also occurs in Tnp26 binding, it would explain the weak binding of the amino acid M1–P56 fragment of Tnp26. The Tnp26 trihelical HTH bundle also modeled to HIV-2 integrase. However, we note that the histidine and cysteine residues involved in Zn^{2+} binding that stabilizes the structure (39) are not present in Tnp26 or the Tnp of other IS*26* family members.

All the changes introduced into the TIR_R of IS*26* had a substantial effect on cointegrate formation in the copy-in mode (the targeted conservative mode was not tested as one pair of active ends is sufficient in that mode (31)). However, the TGT motif may be a key element in the TIR of IS in other IS families as we have noticed that it is at least partly conserved (TGT, NGT, or TGN) in multiple members of other families, as noted previously (40) for the IS*256* family, which includes ISCth4. The IS*903* group in the IS*5* family (*e.g.*, 5′-GGCTT<u>**TGT**</u>TGAATA-3′ in IS*903*) and the IS*1* group in the IS*1* family include this motif. The IS*26* family members are among the shortest IS and have relatively short TIR with only 14 bp highly conserved (15). Their Tnpases do not include multiple potential DBDs suggesting that the HTH examined here is likely to provide at least some of the specificity determinants for end recognition. Further work could investigate the effect *in vivo* and on DNA binding of replacements in amino acids of helices 1 and 2, corresponding to those in helices 3 and 4 of ISCth4 Tnpase that interact directly with the TIR. Some of these residues in the HTH of the IS*256* Tnpase have been shown to abolish DNA binding (11). The contribution of the catalytic domain to binding affinity is also of interest. One of these regions that interact with the ISCth4 TIR is a loop beyond the helix that includes the E of the DDE in ISCth4 Tnpase. This raises the possibility that the enhanced cointegration activity observed when the G residue at 184, which is in this region in Tnp26, is replaced by an N, as found in the Tnp26 variant encoded by IS*26*-v1 (34), is due to enhanced TIR binding. However, further work is needed to determine if the binding of Tnp26 with this substitution to the TIR is enhanced.

IS*26* is one of the most important forces shaping the resistance regions of Gram-negative bacteria, but until recently, little had been done to examine what properties make it so much more effective than the many other ISs found in clinically relevant species. A better understanding of its properties should help to explain this role.

Experimental procedures

Bacterial strains and growth media

Escherichia coli strains DH5α (*supE44 ΔlacU169* Φ80*lacZ*ΔM15 *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*, nalidixic resistant Nx^R), UB5201 (F^- *pro met recA56 gyrA*, Nx^R), and UB1637 (F^- *lys his trp lac recA56*, streptomycin [Sm]-resistant, Sm^R) were used for the propagation of plasmids and in cointegration assays, as previously described (27). *E. coli* BL21-CodonPlus(DE3)-RIPL (F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tc^R (tetracycline-resistant) *gal* λ(DE3) *endA* Hte [*argU proL*, chloramphenicol (Cm)-resistant Cm^R] [*argU ileY leuW* Sm^R/spectinomycin resistant Sp^R]) and Rosetta 2 (F^- *ompT* *hsdS*$_B$(r_B^- m_B^-) *gal dcm*$^+$ [pRARE2, Cm^R]) were used as protein expression cell lines.

Bacterial strains were routinely cultured in LB broth (0.5% [w/v] yeast extract, 1% [w/v] NaCl, 1% [w/v] tryptone) or solid media (0.5% [w/v] yeast extract, 1% [w/v] NaCl, 1% [w/v] tryptone, and 1.5% [w/v] agarose) at 37 °C. Mueller–Hinton broth 2 (Becton Dickinson) or Mueller–Hinton agar (Oxoid) was used when selecting for trimethoprim resistance (Tp^R). Growth media were supplemented with antibiotics (Sigma–Aldrich) at the following concentrations: ampicillin (Ap): 100 μg/ml, Cm: 34 μg/ml, kanamycin: 25 μg/ml, nalidixic acid: 25 μg/ml, Sm: 25 μg/ml, Tc: 10 μg/ml, and trimethoprim (Tp): 20 μg/ml.

Plasmids

Plasmids used and generated in this study are listed in Table 2. Plasmid DNA was extracted by alkaline lysis (41) for

Table 2
Plasmids used and generated in this study

Plasmid	Description	Resistance phenotype[a]	Reference
pRMH977	Single IS*26* cloned into pUC19[b]	Ap	(27)
pRMH1016	pRMH977 with 7G>T transversion in TIR_R	Ap	This study
pRMH1017	pRMH977 with 10G>T transversion in TIR_R	Ap	This study
pRMH1003	pRMH977 with transversions of bp 1–2 of the IS*26* TIR_R	Ap	(31)
pRMH1018	pRMH977 with transversions of bp 3–5 of the IS*26* TIR_R	Ap	This study
pRMH1019	pRMH977 with transversions of bp 6–8 of the IS*26* TIR_R	Ap	This study
pRMH1020	pRMH977 with transversions of bp 9–11 of the IS*26* TIR_R	Ap	This study
pRMH1021	pRMH977 with transversions of bp 12–14 of the IS*26* TIR_R	Ap	This study
pRMH977-F4A	pRMH977 encoding Tnp26 F4A	Ap	This study
pRMH977-F9A	pRMH977 encoding Tnp26 F9A	Ap	This study
pRMH977-F4AF9A	pRMH977 encoding Tnp26 F4AF9A	Ap	This study
pRMH977-E30P	pRMH977 encoding Tnp26 E30P	Ap	This study
pRMH977-E30PW50P	pRMH977 encoding Tnp26 E30PW50P	Ap	This study
pRMH977-G39I	pRMH977 encoding Tnp26 G39I	Ap	This study
pRMH977-G39W	pRMH977 encoding Tnp26 G39W	Ap	This study
pRMH977-R49A	pRMH977 encoding Tnp26 R49A	Ap	This study
pRMH977-W50A	pRMH977 encoding Tnp26 W50A	Ap	This study
pRMH977-W50P	pRMH977 encoding Tnp26 W50P	Ap	This study
pRMH977-Y54A	pRMH977 encoding Tnp26 Y54A	Ap	This study
R388	Conjugative IncW plasmid	SuTp	(43)
R388::IS*26*	R388 with IS*26*[c]	SuTp	(27)
R388::IS*26*-E30P	R388::IS*26* encoding Tnp26 E30P	SuTp	This study
R388::IS*26*-E30PW50P	R388::IS*26* encoding Tnp26 E30PW50P	SuTp	This study
R388::IS*26*-G39I	R388::IS*26* encoding Tnp26 G39I	SuTp	This study
R388::IS*26*-G39W	R388::IS*26* encoding Tnp26 G39W	SuTp	This study
R388::IS*26*-R49A	R388::IS*26* encoding Tnp26 R49A	SuTp	This study
R388::IS*26*-W50A	R388::IS*26* encoding Tnp26 W50A	SuTp	This study
R388::IS*26*-W50P	R388::IS*26* encoding Tnp26 W50P	SuTp	This study
R388::IS*26*-Y54A	R388::IS*26* encoding Tnp26 Y54A	SuTp	This study
1C	*E. coli* expression vector carrying an ORF encoding N-terminal His_6-MBP-Asn_{10}-TEV[d]	Km	Scott Gradia[e]
1C::$Tnp26_{1-56}$	1C encoding Tnp26 M1–P56	Km	This study
1C::$Tnp26_{1-71}$	1C encoding Tnp26 M1–P71	Km	This study
1C::$Tnp26_{1-234}$	1C encoding Tnp26 M1–M234	Km	This study
1C::$Tnp26_{1-234}$ F4A	1C encoding Tnp26 M1–M234 F4A	Km	This study
1C::$Tnp26_{1-234}$ F9A	1C encoding Tnp26 M1–M234 F9A	Km	This study
1C::$Tnp26_{1-234}$ R49A	1C encoding Tnp26 M1–M234 R49A	Km	This study
1C::$Tnp26_{1-234}$ W50A	1C encoding Tnp26 M1–M234 W50A	Km	This study
1C::$Tnp26_{1-234}$ Y54A	1C encoding Tnp26 M1–M234 Y54A	Km	This study

[a] Ap, ampicillin; Km, kanamycin; Su: sulfamethoxazole; and Tp, trimethoprim.
[b] IS*26* together with bases 119,362 to 119,454 and 122,137 to 122,225 from GenBank accession number KF976462. Cloned the insert with *tnp26* facing toward P_{lac} in pUC19.
[c] IS*26* 8-bp duplication of bases 26,745 to 26,752 in R388 (GenBank accession no.: BR000038).
[d] MBP, maltose-binding protein; TEV, tobacco etch virus cleavage site.
[e] pET His6 MBP Asn10 TEV LIC cloning vector (1C) was a gift from Scott Gradia (Addgene plasmid #29654).

analysis by PCR and gel electrophoresis. Standard PCR was conducted as previously described (16). High-fidelity PCR was conducted using Q5 DNA polymerase (New England Biolabs), according to the manufacturer's instructions. PCR amplicons purified using a QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions and sequenced by the Sydney node of the Australian Genome Research Facility Ltd on an Applied Biosystems 3720*xl* DNA Analyzer using the Big Dye Terminator system. Primers generated in this study were synthesized by Integrated DNA Technologies, Inc and are listed in Table S4.

Construction of pRMH977 and R388::IS26 derivatives

Derivatives of the nonconjugative pRMH977 (IS*26* cloned into pUC19; Ap^R) plasmid (27) were generated using site-directed mutagenesis as previously described (27) using primers listed in Table S4. The mutations were introduced into the conjugative R388::IS*26* (Su^RTp^R) plasmid as follows. Cointegrates generated from the fusion of pRMH77 derivatives and R388::IS*26* were verified by PCR screening across each of the two IS*26* copies using primer pairs RH1451 (5′-ACCCAGATACGGCTGATGTC-3′) with RH1472 (5′-ATCGGAAATGGTTGTGAAGC-3′) and RH1452 (5′-TTCGTTCCTGGTCGATTTTC-3′) with RH1471 (5′-CCGCTCCAAAAACTATCCAC-3′) (see Fig. 5 in (27)) and sequencing the amplicons. Cointegrates that carried IS*26* mutations in the left IS*26* copy were digested with *Swa*I, to excise the pRMH977-containing fragment, and religated to reform R388::IS*26* derivatives containing the desired IS*26* mutation. The successful generation of pRMH977 derivatives and R388::IS*26* derivatives was verified by PCR amplifying across IS*26* using primer pairs RH1451/RH1452 and RH1471/RH1472 and sequencing the product, respectively.

Cointegration assays and statistical analyses

A mating-out assay was used to detect cointegrates formed between nonconjugative pRMH977 (Ap^R) and conjugative R388 (Tp^R) or R388::IS*26* (Tp^R), and their derivatives, in a recombination-deficient background as previously described (27). The cointegration frequency was calculated as the ratio of Sm^R cointegrates (Ap^RTp^R) per Sm^R transconjugant (Tp^R). One-way ANOVA and Šidák multiple comparison tests were performed using GraphPad Prism, version 8.3.0 (GraphPad

Software). p values of <0.01 were considered statistically significant. Exact p and t statistic values are provided in Table S5.

Protein homology modeling

Structural homology modeling of full-length and domain truncations of Tnp26 was performed using Phyre2 (42). Protein structures and models were visualized using the PyMOL Molecular Graphics System version 1.7.2.1 (Schrödinger, LLC).

Construction of 1C derivatives

Expression vector 1C (pET His_6 MBP Asn10 tobacco etch virus ligase-independent cloning vector; Addgene plasmid #29654; kanamycin^R), a gift from Scott Gradia, was used to generate recombinants containing the entire or a segment of *tnp26*. The full-length or shorter segments of the *tnp26* sequence was PCR amplified from pRMH977, and the amplicons were cloned into 1C at the *Ssp*I site using the Gibson Assembly Kit (New England Biolabs). For 1C derivatives encoding full-length Tnp26 with amino-acid substitutions, pRMH977 derivatives (listed in Table 2) were used as the DNA template for cloning into the 1C vector. For the 1C derivatives encoding Tnp26 M1–P56 and Tnp26 M1–D71, the P56 end point was selected as proline is less likely to be found in secondary structures, and the D71 end point was selected to avoid the inclusion of cysteine C73. Successful construction of 1C derivatives was verified by PCR amplification across the ORF using primer pairs T7 forward (5′-TAATACGACTCACTATAGGG-3′) and T7 reverse (5′-GCTAGTTATTGCTCAGCGG-3′). The 1C derivatives generated are listed in Table 2.

Protein expression and purification

E. coli BL21-CodonPlus(DE3)-RIPL or Rosetta 2 cell lines carrying 1C recombinants were grown at 37 °C in LB media supplemented with appropriate antibiotics to an absorbance of ∼0.6 at 600 nm and induced with 0.5 mM or 1.0 mM IPTG at 30 °C for 3.5 to 4 h. Cells were harvested, and pellets were stored at −20 °C. Cells were lysed by French press in the presence of 0.5 mM EDTA, 10% (v/v) glycerol, 50 mM Hepes, 5 mM imidazole, 1 M NaCl, 1 mM PMSF, 1 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP–HCl; pH 7.5) and 2 M urea (pH 7.5), with the addition of one cOmplete Protease Inhibitor Cocktail tablet (Roche Diagnostics) per 40-ml volume. Whole cell lysate was clarified by centrifugation at 40,000*g* for 45 min, and the supernatant was applied to 5 ml nickel–nitrilotriacetic acid resin (Invitrogen) to capture His_6-tagged proteins. His_6-tagged proteins were eluted using 400 mM imidazole, and eluted fractions were incubated with protamine sulfate (0.24% [w/v]) to precipitate nucleic acids, and the precipitate was removed by centrifugation at 10,000*g* for 10 min. The soluble portion was shown to be free of DNA using agarose or acrylamide gels stained with SYBR Gold and was applied to amylose resin (New England Biolabs) to immobilize MBP-tagged proteins and eluted with 50 mM maltose. Eluted fractions were further purified by size-exclusion chromatography using the HiLoad 16/600 Superdex 75 in the presence of 0.5 mM EDTA, 10% (v/v) glycerol, 50 mM Hepes, 500 mM NaCl, 1 mM TCEP–HCl (pH 7.5), and the peak fractions containing the monomeric form of the protein and a lower molecular-weight protein containing MBP, which was not separated in this step were pooled (indicated by a bar in Fig. S1). Higher molecular-weight forms eluted with or close to the void volume (at 38–40 ml) precluding estimation of their molecular weights. Tagged proteins were incubated with tobacco etch virus protease at overnight, but this resulted in loss of purified protein to precipitation, and the tag was therefore not removed. All steps were conducted at 4 °C or on ice.

Final protein concentration was estimated with a NanoDropND-1000 Spectrophotometer (Thermo Fisher Scientific) using molar extinction coefficients, and molecular weight values were calculated using ExPASy ProtParam tool (https://web.expasy.org/protparam), namely 72.6 kDa for Tnp26 WT, and 51.6 kDa and 53.7 kDa, respectively, for $Tnp26_{1-56}$ and $Tnp26_{1-71}$. Calculations accounted for lower molecular-weight forms in the purified samples (Fig. S1*C*).

Fluorescently labeled DNA probes

Unlabeled and Cy5-labeled oligonucleotides were synthesized and purchased from Integrated DNA Technologies, Inc and are listed in Table S6. To generate fluorescently labeled dsDNA probes for protein–DNA binding assays, Cy5-labeled oligonucleotides were annealed to their complementary strands at a 1:1 ratio by heating equimolar amounts of each oligonucleotide at 95 °C for 10 min and allowing the mixture to cool over a period of >4 h. Cy5-labeled dsDNA probes were stored at −20 °C.

EMSA and analyses

To detect protein–DNA binding, purified Tnp26 protein was titrated against 25 nM dsDNA probe in the presence of 50 mM Hepes, 150 mM KCl, 1.5 mM $MgCl_2$, 1 mM TCEP, and 20% (v/v) glycerol (pH 7.5). DNA competitor poly-d(I-C) (0.005% [w/v]; Roche) was also included to eliminate nonspecific binding, which was observed in its absence. Binding reactions were incubated for 20 min on ice before glycerol was added to a final concentration of 25% (v/v) and run on 0.8% (w/v) agarose gel (UltraPure Agarose; Invitrogen) in 19.2 mM glycine and 25 mM Tris–HCl (pH 8.0) buffer at 50 V at 4 °C for 50 to 70 min. At least two replicates were performed for each assay. Gels were visualized using the Typhoon FLA 9000 imager and program, with excitation wavelength set at 635 nm.

MST and analyses

Purified Tnp26 proteins were titrated against 25 nM fluorescently labeled dsDNA probes in the presence of 0.01% (w/v) bovine serum albumin, 50 mM Hepes, 150 mM KCl, 1.5 mM $MgCl_2$, 20% (v/v) glycerol, and 0.005% (w/v) poly-d(I-C) DNA competitor (pH 7.5). Binding reactions were incubated for 20 min before loading into Monolith NT. 115 Series Standard Treated Capillaries (NanoTemper Technologies GmbH). MST was conducted using Monolith NT. 115 Control, version 2.0.2.29 (NanoTemper Technologies GmbH), using a light-

emitting diode power of 90% and IR-laser power of 60%. MST measurements were analyzed using MO.Affinity Analysis, version 2.3 (NanoTemper Technologies GmbH) and are tabulated in Table S7, and MST traces are shown in Fig. S3. Three independent replicates were used to determine the K_D and accompanying quality values at an MST-on time of 2.5 s.

Data availability

All data presented and discussed in the article are contained within the main body of the article and in the accompanying supporting figures and tables.

Supporting information—This article contains supporting information (31).

Acknowledgments—We thank Willem Baris for his contributions to the project during his internship with the Ataide laboratory.

Author contributions—R. M. H. conceptualization; J. K. F. and S. F. A. methodology; R. M. H. validation; C. H. P. formal analysis; C. H. P., C. J. H., and J. K. F. investigation; S. F. A. and R. M. H. resources; C. H. P., C. J. H., J. K. F., and R. M. H. data curation; C. H. P. and R. M. H. writing–original draft; C. H. P., S. F. A., and R. M. H. writing–review and editing; C. H. P. and C. J. H. visualization; C. J. H., S. F. A., and R. M. H. supervision; R. M. H. project administration; R. M. H. and S. F. A. funding acquisition.

Funding and additional information—C. J. H. and J. K. F. was supported by the National Health and Medical Research Council project grants GNT1032465 and GNT1141540 to R. M. H. and the University of Sydney Bridging Support grant G199779 to S. F. A. C. H. P. was supported by a Research Training Program stipend and the National Health and Medical Research Council project grant GNT1141540.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: Ap, ampicillin; Ap^R, ampicillin resistant; Cm, chloramphenicol; DBD, DNA-binding domain; DDE, aspartate–aspartate–glutamate; HTH, helix–turn–helix; IS, insertion sequence; MBP, maltose-binding protein; MST, microscale thermophoresis; PDB, Protein Data Bank; Sm, streptomycin; Tc, tetracycline; TCEP, Tris(2-carboxyethyl)phosphine; TIR, terminal inverted repeat; TIR_L, left terminal inverted repeat; TIR_R, right terminal inverted repeat; Tnpase, transposase.

References

1. Siguier, P., Gourbeyre, E., Varani, A., Ton-Hoang, B., and Chandler, M. (2015) Everyman's guide to bacterial insertion sequences. *Microbiol. Spectr.* **3**, 555–590
2. Curcio, M. J., and Derbyshire, K. M. (2003) The outs and ins of transposition: From Mu to Kangaroo. *Nat. Rev. Mol. Cell Biol.* **4**, 865–877
3. Mahillon, J., and Chandler, M. (1998) Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**, 725–774
4. Campbell, A., Berg, D. E., Botstein, D., Lederberg, E. M., Novick, R. P., Starlinger, P., and Szybalski, W. (1979) Nomenclature of transposable elements in prokaryotes. *Gene* **5**, 197–206
5. Rice, P. A., and Baker, T. A. (2001) Comparative architecture of transposase and integrase complexes. *Nat. Struct. Mol. Biol.* **8**, 302–307
6. Hickman, A. B., Chandler, M., and Dyda, F. (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. Biochem. Mol. Biol.* **45**, 50–69
7. Dyda, F., Chandler, M., and Hickman, A. B. (2012) The emerging diversity of transpososome architectures. *Q. Rev. Biophys.* **45**, 493–521
8. Ohta, S., Yoshimura, E., and Ohtsubo, E. (2004) Involvement of two domains with helix–turn–helix and zinc finger motifs in the binding of IS*1* transposase to terminal inverted repeats. *Mol. Microbiol.* **53**, 193–202
9. Rousseau, P., Gueguen, E., Duval-Valentin, G., and Chandler, M. (2004) The helix-turn-helix motif of bacterial insertion sequence IS*911* transposase is required for DNA binding. *Nucleic Acids Res.* **32**, 1335–1344
10. Nagy, Z., Szabó, M., Chandler, M., and Olasz, F. (2004) Analysis of the N-terminal DNA binding domain of the IS*30* transposase. *Mol. Microbiol.* **54**, 478–488
11. Hennig, S., and Ziebuhr, W. (2010) Characterization of the transposase encoded by IS*256*, the prototype of a major family of bacterial insertion sequence elements. *J. Bacteriol.* **192**, 4153–4163
12. Tavakoli, N. P., DeVost, J., and Derbyshire, K. M. (1997) Defining functional regions of the IS*903* transposase. *J. Mol. Biol.* **274**, 491–504
13. Kosek, D., Hickman, A. B., Ghirlando, R., He, S., and Dyda, F. (2021) Structures of ISCth4 transpososomes reveal the role of asymmetry in copy-out/paste-in DNA transposition. *EMBO J.* **40**, e105666
14. Lewis, L. A., Astatke, M., Umekubo, P. T., Alvi, S., Saby, R., Afrose, J., Oliveira, P. H., Monteiro, G. A., and Prazeres, D. M. (2012) Protein-DNA interactions define the mechanistic aspects of circle formation and insertion reactions in IS*2* transposition. *Mob. DNA* **3**, 1–25
15. Harmer, C. J., and Hall, R. M. (2019) An analysis of the IS6/IS*26* family of insertion sequences: Is it a single family? *Microb. Genom.* **5**, e000291
16. Cain, A. K., Liu, X., Djordjevic, S. P., and Hall, R. M. (2010) Transposons related to Tn*1696* in IncHI2 plasmids in multiply antibiotic resistant *Salmonella enterica* serovar Typhimurium from Australian animals. *Microb. Drug Resist.* **16**, 197–202
17. Partridge, S. R., Kwong, S. M., Firth, N., and Jensen, S. O. (2018) Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* **31**, e00088-17
18. Dawes, F. E., Kuzevski, A., Bettelheim, K. A., Hornitzky, M. A., Djordjevic, S. P., and Walker, M. J. (2010) Distribution of class 1 integrons with IS*26*-mediated deletions in their 3′-conserved segments in *Escherichia coli* of human and animal origin. *PLoS One* **5**, e12754
19. Blackwell, G. A., Nigro, S. J., and Hall, R. M. (2015) Evolution of AbGRI2-0, the progenitor of the AbGRI2 resistance island in global clone 2 of *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* **60**, 1421–1429
20. de Curraize, C., Siebor, E., Varin, V., Neuwirth, C., and Hall, R. M. (2020) Two new SGI1-LK variants found in *Proteus mirabilis* and evolution of the SGI1-HKL group of *Salmonella* genomic islands. *mSphere* **5**, e00875-19
21. Stewart, P. R., Dubin, D. T., Chikramane, S. G., Inglis, B., Matthews, P. R., and Poston, S. M. (1994) IS*257* and small plasmid insertions in the *mec* region of the chromosome of *Staphylococcus aureus*. *Plasmid* **31**, 12–20
22. Yui Eto, K., Firth, N., Davis, A. M., Kwong, S. M., Krysiak, M., Lee, Y. T., O'Brien, F. G., Grubb, W. B., Coombs, G. W., Bond, C. S., and Ramsay, J. P. (2019) Evolution of a 72-kilobase cointegrant, conjugative multiresistance plasmid in community-associated methicillin-resistant Staphylococcus aureus isolates from the early 1990s. *Antimicrob. Agents Chemother.* **63**, e01560-19
23. Barile, S., Devirgiliis, C., and Perozzi, G. (2012) Molecular characterization of a novel mosaic *tet*(S/M) gene encoding tetracycline resistance in foodborne strains of *Streptococcus bovis*. *Microbiology* **158**, 2353–2362
24. D'Andrea, M. M., Antonelli, A., Brenciani, A., Di Pilato, V., Morroni, G., Pollini, S., Fioriti, S., Giovanetti, E., and Rossolini, G. M. (2019) Characterization of Tn*6349*, a novel mosaic transposon carrying *poxtA*, *cfr* and other resistance determinants, inserted in the chromosome of an ST5-MRSA-II strain of clinical origin. *J. Antimicrob. Chemother.* **74**, 2870–2875
25. Di Sante, L., Morroni, G., Brenciani, A., Vignaroli, C., Antonelli, A., D'Andrea, M. M., Di Cesare, A., Giovanetti, E., Varaldo, P. E., Rossolini, G. M., and Biavasco, F. (2017) pHTβ-promoted mobilization of non-conjugative resistance plasmids from *Enterococcus faecium* to *Enterococcus faecalis*. *J. Antimicrob. Chemother.* **72**, 2447–2453

26. Mollet, B., Iida, S., and Arber, W. (1985) Gene organization and target specificity of the prokaryotic mobile genetic element IS*26*. *Mol. Gen. Genet.* **201**, 198–203
27. Harmer, C. J., Moran, R. A., and Hall, R. M. (2014) Movement of IS*26*-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS*26* and preferentially inserts adjacent to another IS*26*. *mBio* **5**, e01801-14
28. Harmer, C. J., and Hall, R. M. (2021) IS*26* cannot move alone. *J. Antimicrob. Chemother.* **76**, 1428–1432
29. Trieu-Cuot, P., and Courvalin, P. (1985) Transposition behavior of IS*15* and its progenitor IS*15*-Δ: Are cointegrates exclusive end products? *Plasmid* **14**, 80–89
30. Brown, A. M., Coupland, G. M., and Willetts, N. S. (1984) Characterization of IS*46*, an insertion sequence found on two IncN plasmids. *J. Bacteriol.* **159**, 472–481
31. Harmer, C. J., and Hall, R. M. (2017) Targeted conservative formation of cointegrates between two DNA molecules containing IS*26* occurs via strand exchange at either IS end. *Mol. Microbiol.* **106**, 409–418
32. Harmer, C. J., and Hall, R. M. (2021) Targeted conservative cointegrate formation mediated by IS*26* family members requires sequence identity at the reacting end. *mSphere* **6**, e01321-20
33. Harmer, C. J., and Hall, R. M. (2020) IS*26* family members IS*257* and IS*1216* also form cointegrates by copy-in and targeted conservative routes. *mSphere* **5**, e00811-19
34. Pong, C. H., Harmer, C. J., Ataide, S. F., and Hall, R. M. (2019) An IS*26* variant with enhanced activity. *FEMS Microbiol. Lett.* **366**, fnz031
35. Li, X., Krishnan, L., Cherepanov, P., and Engelman, A. (2011) Structural biology of retroviral DNA integration. *Virology* **411**, 194–205
36. Singarapu, K. K., Xiao, R., Sukumaran, D. K., Acton, T., Montelione, G. T., and Szyperski, T. (2008) NMR structure of protein Cgl2762 from *Corynebacterium glutamicum* implicated in DNA transposition reveals a helix-turn-helix motif attached to a flexibly disordered leucine zipper. *Proteins* **70**, 1650–1654
37. Makwana, K. M., and Mahalakshmi, R. (2015) Implications of aromatic-aromatic interactions: From protein structures to peptide models. *Protein Sci.* **24**, 1920–1933
38. Harmer, C. J., Pong, C. H., and Hall, R. M. (2020) Structures bounded by directly-oriented members of the IS*26* family are pseudo-compound transposons. *Plasmid* **111**, 102530
39. Eijkelenboom, A. P., van den Ent, F. M., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T. D., Plasterk, R. H., Kaptein, R., and Boelens, R. (1997) The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: A three-helix bundle stabilized by zinc. *Curr. Biol.* **7**, 739–746
40. Dodd, H. M., Horn, N., and Gasson, M. J. (1994) Characterization of IS*905*, a new multicopy insertion sequence identified in *Lactococci*. *J. Bacteriol.* **176**, 3393–3396
41. Sambrook, J., and Russell, D. (1989) *Molecular Cloning: A Laboratory Manual* (Vol 3). Cold Spring Harbor Laboratory Press, New York, NY
42. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858
43. Revilla, C., Garcillán-Barcia, M. P., Fernández-López, R., Thomson, N. R., Sanders, M., Cheung, M., Thomas, C. M., and de la Cruz, F. (2008) Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. *Antimicrob. Agents Chemother.* **52**, 1472–1480