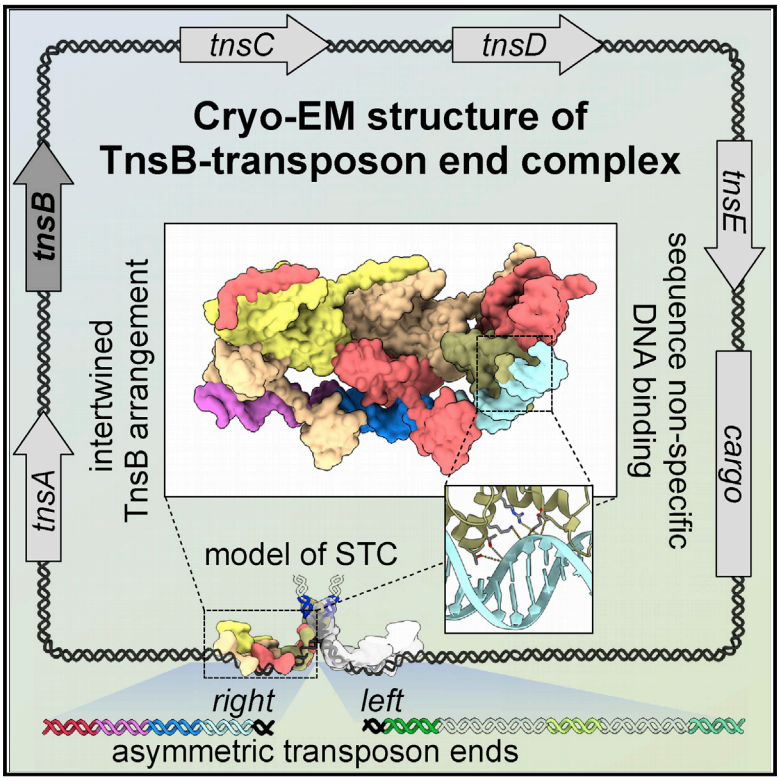
Molecular Cell

Article

Structural basis of transposon end recognition explains central features of Tn7 transposition systems

Graphical abstract



Highlights

- Cryo-EM structure of Tn7 transposase TnsB in complex with transposon end DNA
- TnsB chains interact with its binding sites in DNA in a tiled and intertwined fashion
- TnsB exhibits a DNA sequence preference rather than strict specificity
- Overlap of TnsB-binding sites in the right Tn7 end is key for its functionality

Authors

Zuzanna Kaczmarska, Mariusz Czarnocki-Cieciura, Karolina M. Górecka-Minakowska, ..., Timothy Grant, Joseph E. Peters, Marcin Nowotny

Correspondence

joe.peters@cornell.edu (J.E.P.), mnowotny@iimcb.gov.pl (M.N.)

In brief

Kaczmarska et al. present a structure of Tn7 transposon transposase TnsB bound to transposon end DNA. Tn7 is a bacterial transposon. Its relatives containing CRISPR-Cas machinery are promising tools for gene editing. TnsB molecules interact with repeating binding sites in the transposon end in an intertwined fashion for specific end recognition.

Kaczmarska et al., 2022, Molecular Cell *82*, 2618–2632
July 21, 2022
https://doi.org/10.1016/j.molcel.2022.05.005

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

Structural basis of transposon end recognition explains central features of Tn7 transposition systems

Zuzanna Kaczmarska,[1,7] Mariusz Czarnocki-Cieciura,[1,7] Karolina M. Górecka-Minakowska,[1,7,8] Robert J. Wingo,[2] Justyna Jackiewicz,[1] Weronika Zajko,[1] Jarosław T. Poznański,[3] Michał Rawski,[4] Timothy Grant,[5,6] Joseph E. Peters,[2,*] and Marcin Nowotny[1,9,*]

[1]Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, 02-109 Warsaw, Poland
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland
[4]Malopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland
[5]John and Jeanne Rowe Center for Research in Virology, Morgridge Institute for Research, Madison, WI 53715, USA
[6]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
[7]These authors contributed equally
[8]Present address: Captor Therapeutics Inc., Wroclaw, Poland
[9]Lead contact
*Correspondence: joe.peters@cornell.edu (J.E.P.), mnowotny@iimcb.gov.pl (M.N.)
https://doi.org/10.1016/j.molcel.2022.05.005

SUMMARY

Tn7 is a bacterial transposon with relatives containing element-encoded CRISPR-Cas systems mediating RNA-guided transposon insertion. Here, we present the 2.7 Å cryoelectron microscopy structure of prototypic Tn7 transposase TnsB interacting with the transposon end DNA. When TnsB interacts across repeating binding sites, it adopts a beads-on-a-string architecture, where the DNA-binding and catalytic domains are arranged in a tiled and intertwined fashion. The DNA-binding domains form few base-specific contacts leading to a binding preference that requires multiple weakly conserved sites at the appropriate spacing to achieve DNA sequence specificity. TnsB binding imparts differences in the global structure of the protein-bound DNA ends dictated by the spacing or overlap of binding sites explaining functional differences in the left and right ends of the element. We propose a model of the strand-transfer complex in which the terminal TnsB molecule is rearranged so that its catalytic domain is in a position conducive to transposition.

INTRODUCTION

Transposons are regions of DNA that can move within genomes increasing genetic diversity. Bacterial Tn7 elements are among the best-studied and most widespread DNA transposons and are detected in 10%–20% of sequenced bacterial genomes (Peters et al., 2017). Tn7 mobility is tightly controlled and mediated by five element-encoded proteins (Craig, 2002; Peters, 2014). Translocation occurs via a cut-and-paste mechanism executed by a heteromeric transposase TnsA+TnsB (Turlan and Chandler, 2000). The AAA+ ATPase TnsC coordinates transposase recruitment to the target site through an interaction with one of the two target selectors, TnsD or TnsE. TnsD is a sequence-specific DNA-binding protein that directs transposition into the conserved chromosomal *attTn7* site at a high frequency. TnsE targets mobile plasmids by recognizing features of conjugal DNA replication. The CRISPR-associated transposon (CAST) elements that use CRISPR-Cas systems for RNA-guided DNA transposition are all related to Tn7 and encode TnsB-like transposases (Faure et al., 2019; Peters et al., 2017). Tn7/CASTs provide some of the most promising systems for next-generation gene-editing tools.

The TnsB transposase belongs to the retroviral integrase superfamily with an RNase H fold catalytic domain with a DDE active site motif (Haren et al., 1999; Nowotny, 2009; Hickman et al., 2010). TnsB recognizes the left and right ends of the element, catalyzes cleavage at the 3′ ends of the transposon, and performs the strand-transfer reaction that leads to the insertion of the element into a target DNA. The TnsA endonuclease associates with TnsB and cleaves the 5′-ends allowing the element to be completely released from a donor site (Hickman et al., 2000; May and Craig, 1996; Sarnovsky et al., 1996). Imperfect inverted repeats delineate the ends of the element consisting of an 8 base pair (bp) end sequence and an array of internal 22 bp TnsB-binding sites (hereafter referred to as "sites") essential for both the positioning and orientation of integration events (Arciszewska et al., 1991; Arciszewska and Craig, 1991; Holder and Craig, 2010). Four partially overlapping sites are located in

the right end, and three sites (separated by 42 and 31 bp) are present in the left end, making the ends structurally asymmetric (Lichtenstein and Brenner, 1982). In addition, TnsB coordinates Tn7 end pairing, a step that must occur before DNA breakage and joining. TnsB and TnsC cooperate to integrate the transposon DNA at its target site. Despite these central roles, no structural information is currently available for TnsB limiting a full mechanistic understanding of the transposition process.

A notable aspect of Tn7 transposition is target immunity. Although an initial transposition event occurs at high frequency, target immunity strongly inhibits a second insertion from occurring into the same region, thereby promoting the distribution of the element to new preferred sites (Skelding et al., 2003; Stellwagen and Craig, 1997a,b). Target immunity is not universal but is also observed for some other transposons (e.g., Tn3) and Mu bacteriophage (Hauer and Shapiro, 1984; Robinson et al., 1977; Reyes et al., 1987; Arciszewska et al., 1989). A key feature of target immunity is the prevention of a stable interaction between TnsC and target DNA by a high local concentration of TnsB mediated by TnsB-binding sites in target DNA. Tn7 immunity is sufficient to protect a vector or large chromosomal region from subsequent integration events (DeBoy and Craig, 1996).

Here, we provide the first structural information for TnsB, a 2.7 Å cryo-EM structure of TnsB interacting with the right end of the element. TnsB adopts a three-domain beads-on-a-string architecture, with domains of adjacent TnsB chains arranged in a tiled and intertwined fashion. TnsB binds DNA with sequence preference rather than strict sequence specificity, and the formation of an array of TnsB molecules converts this preference into specific end recognition.

RESULTS

TnsB forms a tiled arrangement on DNA with repeating sites

We used cryoelectron microscopy to analyze complexes of prototypic *E. coli* Tn7 TnsB with double-stranded DNA substrates corresponding to the right end of the transposon. For one of the substrates, which comprised three sites (substrate 3BS, Figure 1A), we observed the formation of long filaments that we used to solve the cryo-EM structure of the TnsB-DNA complex (Figures 1B, 1C, S1, and S2; Tables S1 and S2).

In our structure, all three binding sites are occupied by TnsB proteins (Figure 1D). A full-length TnsB protomer contains three discrete domains that are separated by flexible linkers (Figure 2). Starting from the N terminus, we distinguished DNA-binding domain 1 (DBD1), which encompasses an SH3-like subdomain and a trihelical helix-turn-helix (HTH) motif. The next domain is DNA-binding domain 2 (DBD2), which consists of a winged helix-like (WH-like) domain that is similar to the one that was previously found in APC2 protein (PDB: 6NXK) (Watson et al., 2019). DBD2 is followed by the catalytic domain (CD), containing an RNase H fold subdomain with a transposase active site composed of D273, D361, and E396 (Sarnovsky et al., 1996) and a β barrel subdomain. A long α helix with a kink at residue 599 is present after the CD (Figures 1C and 2). For the DNA, we observed density for all base pairs of the three sites and one additional base pair from the 8 bp end sequence.

A prominent feature of the structure is DNA bending (Figure 1C). The total cumulative bend of the double helix axis is 70°. In the DNA, sites 1, 2, and 3 are bound by the DNA-binding domains of chains A, B, and C, respectively. DBD1 and DBD2 bind DNA 3 bp from each other and are joined by a linker that is positioned in the minor groove of the DNA. The catalytic domains are located on top of DBD1s creating extensive interaction surfaces (Figures 1B–1D and 3A) and do not form any contacts with DNA. The CDs of chains B, C, and D interact with DBD1s of chains A, B, and C, respectively; so, the domains of TnsB protomers form a tiled and intertwined arrangement. In the subsequent analysis, domains in the vicinity of a particular site are grouped into "modules" (Figure 1D).

The catalytic domains participate in additional interactions with SH3-like subdomains of DBD1 from a preceding module. A prominent interface is also formed by the C-terminal long α helix (Figure 3A). It binds the catalytic domain of the preceding TnsB chain, maintaining contacts mostly with the last two α helices of the RNase H fold subdomain and a loop between these helices. This part of the CD was previously shown to be required for an interaction with TnsA (Choi et al., 2013). It implies that TnsA binding to any of the TnsB molecules must be accompanied by the structural rearrangement of this region. TnsA may displace the C-terminal helix of the neighboring TnsB molecule from the position that is observed in our structure.

The DBD2 and CD linkage was not fully resolved in our reconstruction. It could have been explained by either of the following two possibilities: a tiled arrangement of domains between modules (Figures 1D and S2D) or an alternative arrangement with TnsB domains from one protein chain occupying one module (Figure S2E), which would require much longer DBD2-CD linker. We confirmed the tiled arrangement by showing that a variant of TnsB with a shorter DBD2-CD linker (Δ245-251) was able to bind 3BS DNA and form filaments similar to the wild-type protein (Figures S2F and S2G; see STAR Methods for details).

To further validate our structural model, we wanted to test the transposon end binding *in vivo* by TnsB variants with point substitutions designed based on the structure. So, we developed a new quantitative DNA-binding assay that takes advantage of the fact that TnsB binding to transposon end represses expression of the transposition genes. As the readout of this regulation, we used the reduction of expression and luminescence of NanoLuc, a small and highly luminescent luciferase variant. We expressed the effect observed for the wild-type TnsB as a 100% reduction of luminescence. The assay is schematically shown in Figures S3A–S3C and described in detail in STAR Methods.

To verify the interdomain contacts that were observed in the structure, we generated variants in which residues L43 and L525 at the DBD1 (SH3)-CD interface and P133 at the DBD1 (HTH)-CD interface (Figure 3A) were individually substituted with tryptophan. We reasoned that the presence of a bulky indole side chain would disrupt the interaction interfaces. We confirmed that these variants folded properly by measuring the infrared spectra of purified proteins (not shown). We next tested the DNA binding of the TnsB variants using the NanoLuc assay. Expression of L43W, P133W, and L525W variants caused smaller luminescence reduction than the expression of the

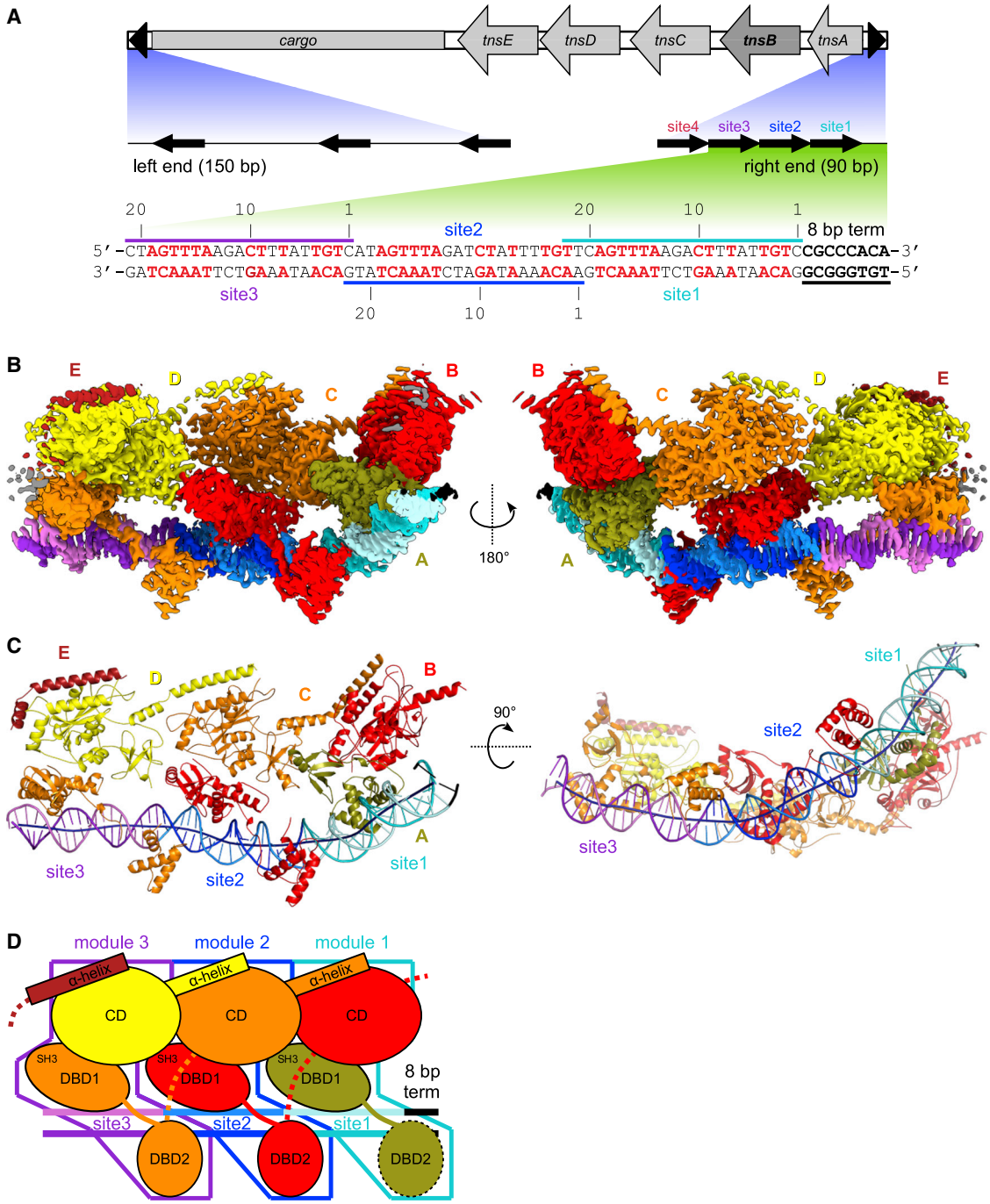


Figure 1. Beads-on-a-string organization of TnsB bound to the fragment of the right transposon end

(A) Schematic of Tn7 transposon and its ends. The sequence of 3BS DNA used to obtain the structure is shown at the bottom of the panel. The nucleotide numbering of each site (1, 10, and 20) is given, and the nucleotides conserved across all four sites of the right end are shown in red.

(B) 2.7 Å cryo-EM reconstruction of TnsB in complex with transposon end DNA. The map is colored according to TnsB chains (labeled with capital letters) and the DNA sites: olive for chain A, red for chain B, orange for chain C, yellow for chain D, burgundy for chain E, light cyan/cyan (top/bottom strands) for site 1, blue/dark blue for site 2, pink/purple for site 3, and black for transposon terminus.

(C) Atomic model of TnsB bound to DNA. All the components are colored as in (B). The axis of the DNA helix is shown as a dark blue tube.

(D) Schematic representation of the structure of TnsB bound to DNA. DBD2, which has a poorly defined density, is shown as a dashed oval. All the components are colored as in (B).

See also Figures S1 and S2 and Tables S1 and S2.

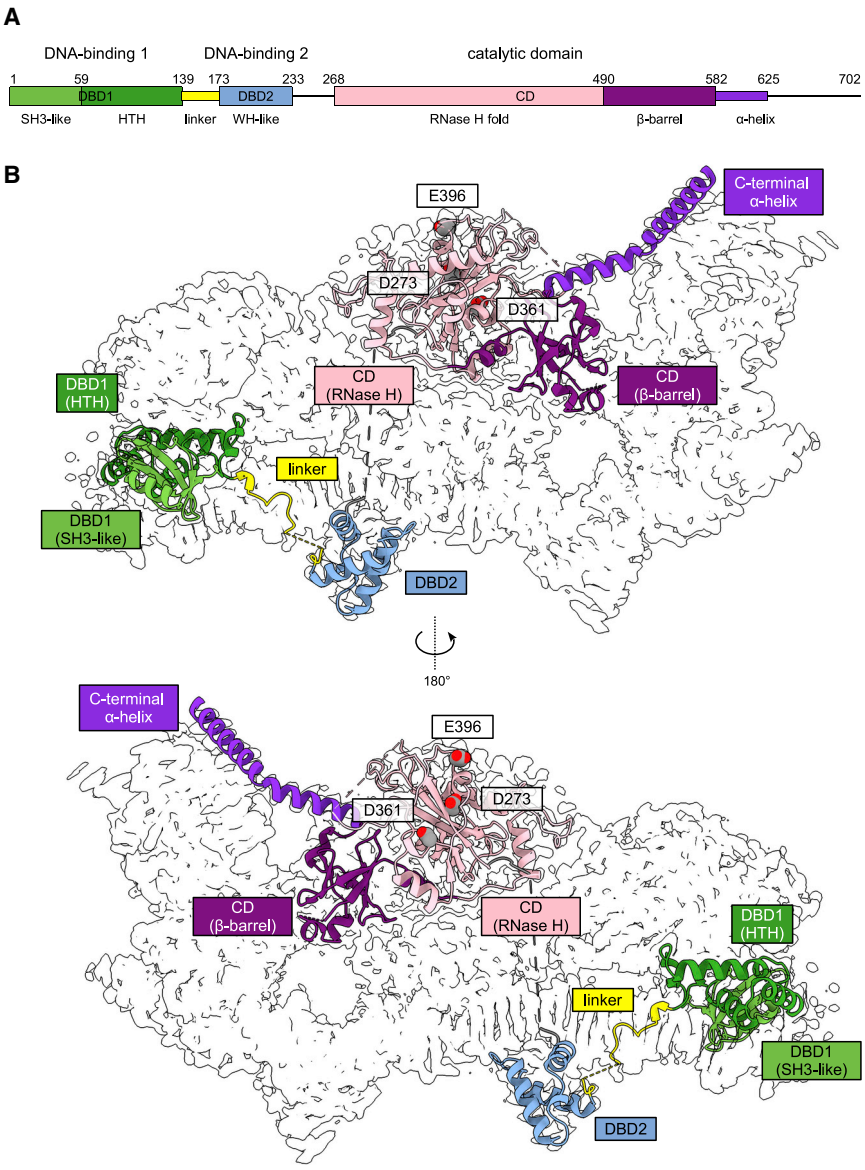


Figure 2. Structure of a single TnsB chain

(A) Domain architecture of TnsB. HTH, helix-turn-helix; WH-like, winged helix-like. Residue numbers at the domain boundaries are given. The domains/subdomains are colored: light and dark green for SH3-like and HTH subdomain of DBD1, respectively, yellow for the linker, light blue for DBD2, light pink for RNase H, violet for β barrel, and purple for C-terminal α helix.

(B) Structure of one TnsB protomer (chain C) shown in cartoon representation and colored as in (A). The EM potential map is shown as a white contour. Active site residues of the CD are shown as spheres.

wild-type TnsB (59%, 53%, and 83% of the wild-type protein effect, respectively, Figure 3B). This indicated that the DNA binding by these variants was perturbed. We also measured transposition frequency *in vivo*. The L43W, P133W, and L525W variants showed an ~80%, ~50%, and ~75% decrease in transposition efficiency, respectively (Figure 3C). We also designed a deletion mutant that comprised residues 1–589 and lacked the C-terminal helix (ΔC-term variant) to study its contribution to DNA complex stabilization. This mutant showed 64% of the luminescence reduction observed for the wild-type protein. Thus, this variant retained some DNA binding but is very unlikely to function *in vivo* because it lacks the key sequences from the extreme C terminus that are required for the interaction with TnsC (Skelding, 2002). Collectively, the NanoLuc and transposition assays results confirmed the importance of the interfaces that were identified in our structure.

Closer inspection of the structure of the TnsB-transposon end complex revealed that the positions of the CDs relative to DBDs and DNA are not identical among modules (Figures 4A–4C). When individual modules were superimposed using DBD1, the positions of CDs of chains B and D were very similar (Figure 4D), but the position of the CD of subunit C (module 2) differed (Figures 4E and 4F). We suggest that this alteration of TnsB arrangement most likely results from a larger overlap between TnsB-binding sites 1 and 2 (2 nt overlap) compared with sites 2 and 3 (1 nt overlap; Figure 1A). Since the core sequences of sites 1 and 2 are closer together, the distance between DBD1 of chain A (bound to site 1) and module 2 is smaller than the distance between DBD1 of chain B (bound to site 2) and module 3 (Figure 1D). This configuration drives multiple rearrangements in the structure. First, the CD of chain C is displaced from its position in the complex, pushed by DBD1 of chain A whose location is shifted due to the altered register of DNA binding. Second, for the DBD1 of chain A to maintain interactions with the CD of chain C, the bent DNA bound by this DBD1 has to change its trajectory (compare Figures 4B and 4C). Third, changes in the conformation of the C-terminal helices of chains C and D are required to accommodate this structural alteration. In multiple prototypical Tn7 elements, the 2 nt overlap between sites 1 and 2 is conserved (Figure S4), thus they may share a similar arrangement of TnsB molecules bound to sites 1 and 2.

TnsB forms contact mainly with the DNA backbone

We next analyzed TnsB interactions with the transposon end DNA. The contacts we identified are similar for all modules and are shown schematically in Figure 5A for module 2. The following three protein regions are involved in these interactions: DBD1, DBD2, and the linker between them with most contacts to the DNA backbone and few base-specific contacts.

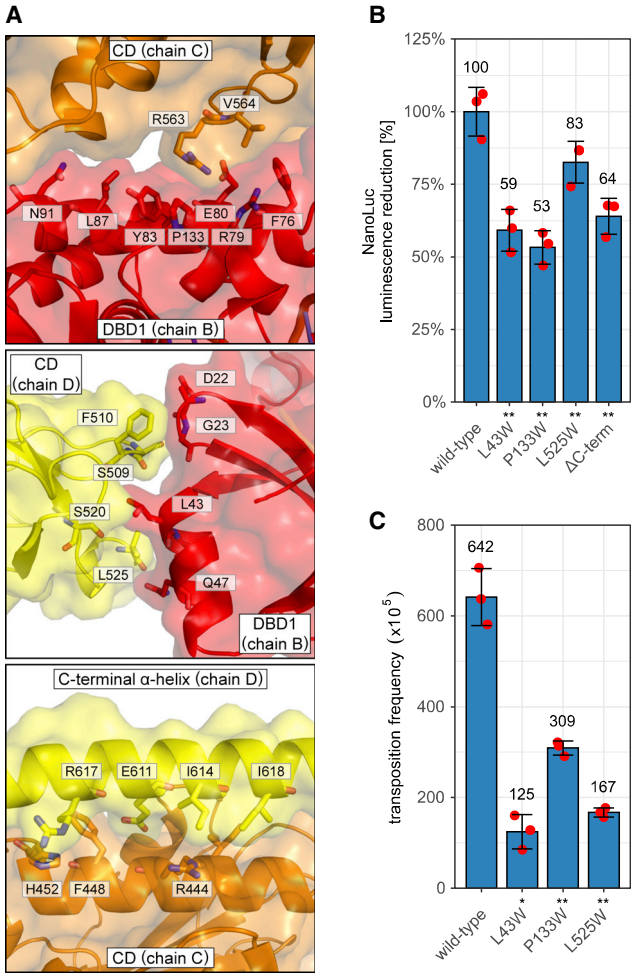


Figure 3. TnsB interdomain contacts

(A) Close up of interdomain contacts between DBD1 (HTH subdomain) and CD (top panel); DBD1 (SH3-like subdomain) and CD (middle panel); and C-terminal α helix and CD (bottom panel). The TnsB protein is in cartoon and transparent surface representation and colored as in Figure 1B. The residues involved in the contacts are shown as sticks and labeled.

(B) NanoLuc DNA-binding assay for wild-type TnsB, variants with substitutions at the interdomain interfaces or a C-terminal deletion (ΔC-term). The luminescence reduction upon expression of TnsB variants is plotted with the effect for the wild-type construct normalized to 100%.

(C) *In vivo* lambda hop transposition assay. Tn7 with indicated TnsB variants was introduced in *E. coli*, and the frequency of transposition events was measured.

(B and C) Values from three independent experiments are shown as mean ± SD. Asterisks indicate statistical significance between wild type and each mutant (* $p < 0.05$ and ** $p < 0.01$) by the paired sample t test.

See also Figures S2 and S3.

DBD1 approaches the DNA from the minor groove side, and nearly all DBD1-DNA contacts are with the backbone phosphates of the DNA (Figures 5A and 5B). The only exception is K116 interacting with O6 of a guanosine base, which is conserved in all sites (position 18 of the top strand of each binding site). K34 and G35 from the SH3-like subdomain as well as K78, R101, S102, T115, T118, Y120, K121, R125, and D139 from the HTH subdomain are involved in DNA phosphate binding (Figure 5B). The linker between DBD domains interacts with the minor grove of the DNA. Here, several interactions with bases are observed. S143 and R150 form hydrogen bonds with the bases of conserved thymidines (positions 9 and 7, top strand). R160 interacts with the base of an adenosine conserved in all four sites (position 4, bottom strand). Linker residues Y140, G147, R150, K157, R162, and E163 bind the DNA backbone, mostly via phosphate interactions (Figure 5B). DBD2 interacts with a DNA region that comprises the boundary between two sites and partially overlaps with the region bound by the linker. T221, Q224, and Y227 interact with the DNA phosphates, and R223 interacts with the bases in positions 2 and 3 of the binding site (Figure 5B). DBD2 residues T196, T197, R201, and R226 interact with the phosphates of nucleotides that are complementary to the DNA that is bound by DBD1 of the preceding module.

To validate protein-DNA interactions observed in our structure, we designed a series of TnsB variants with substitutions predicted to affect the binding to DNA. Mutated residues were selected based on their location in the vicinity of the DNA and conservation in TnsB proteins (Figures 5A and 5B; Data S1). The following five variants contained substitutions in the DBD1 domain: K99A/R101A, T115A/T118A, K116A, Y120A/K121A, and R124A/R125A. The three following variants carried substitutions in the linker between DBDs: S143A/R150A, K157A, and R160A. Three variants contained substitutions in the DBD2: R223A, Q224A/R226A, and Y227A/R231A. Three of the mutated residues (K99, R124, and R231) do not directly interact with the DNA in our structure but are located close to the DNA and with minor changes in the side-chain conformation could form such contacts. DNA binding of all these variants was assessed using the NanoLuc assay (Figure 5C). The expression of most of the tested variants led to a smaller reduction of the luminescence signal than the expression of the wild-type TnsB, indicating defects in DNA binding. The three least affected variants were K116A (81% of wild type), K157A (86% of wild type), and R160A (no effect on repression). More deleterious defects were observed in the *in vivo* transposition assay, where most of the mutants were completely deficient in transposition (Figure 5D). Only K116A, K157A, and R160A retained ~50%, ~10%, and ~25% of the transposition frequency of the wild-type element, respectively, which is in agreement with their weakest defects in the NanoLuc DNA-binding assay. A greater negative impact of amino acid substitutions on transposition than DNA binding implies that even small perturbations with transposon end binding lead to complete inhibition of transposition *in vivo*. Overall, the NanoLuc and *in vivo* transposition assays confirmed the importance of the residues that interact with DNA based on our structure.

To assess the functionality of the DNA-binding domains in isolation, we also constructed a variant that contained only two N-terminal DBDs (variant 1–235). In the NanoLuc assay with this variant, we observed 46% of the luminescence signal reduction of the wild-type TnsB (Figure 5C). This result suggests that the isolated DBDs can interact with the DNA, but the additional domains provide further stabilization of the nucleoprotein complex.

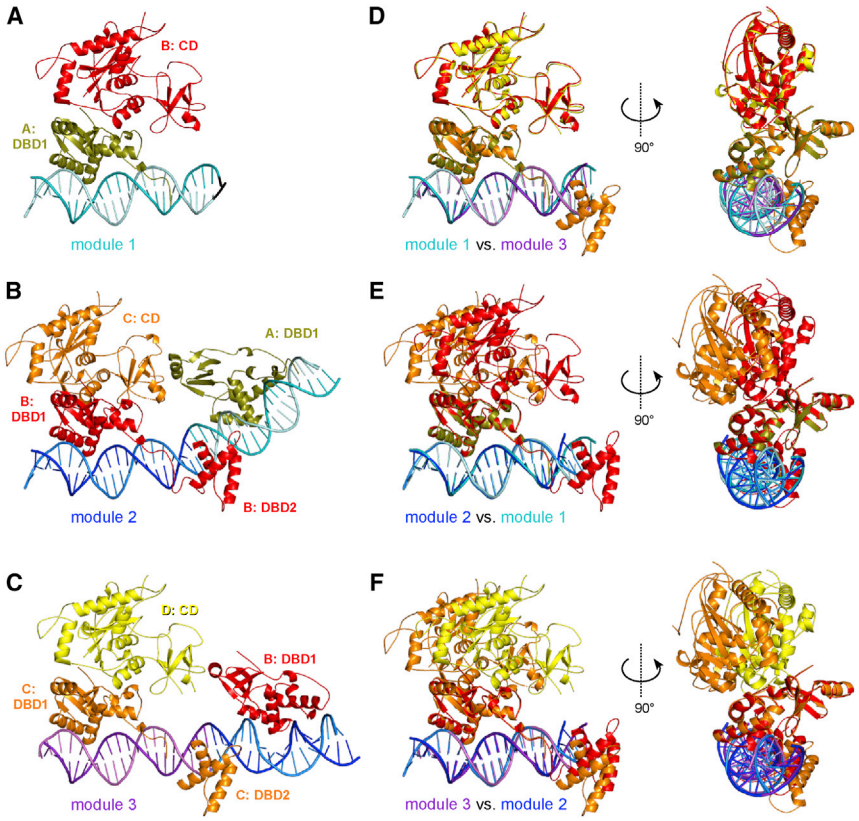


Figure 4. Comparison of TnsB modules

(A–C) Modules interacting with a particular site (1, 2, or 3, respectively). (B and C) The DBD1 that binds the adjacent site is also shown. The protein is in cartoon representation and colored as in Figure 1B. (D–F) Superimposition of modules 1 and 3 (D), modules 2 and 1 (E), and modules 2 and 3 (F); based on DBD1.

TnsB exhibits DNA sequence preference and not strict specificity

Most TnsB contacts were with the DNA phosphodiester backbone with relatively few contacts between the protein and bases of the DNA that could serve for sequence-specific recognition. This suggests that TnsB may have the ability to also bind random-sequence DNA, hinting at a possible functional role for nonspecific binding. DNA sequence specificity of TnsB was tested with NanoLuc reporter constructs where individual sites were scrambled: site 1 (variant s1); site 2 (s2); site 1 and site 2 (s12); site 2 and site 3 (s23); and sites 1–3 (s123) (Figure 6A, left). In the absence of exogenous TnsB, NanoLuc was efficiently expressed from all these reporter plasmid variants (Figure S3D). Upon TnsB expression, variants s1 and s2 showed 73% and 69% of the luminescence reduction observed with the wild-type construct with an intact transposon end. Variants with two sites scrambled, s12 and s23, showed 48% and 60% of the luminescence reduction of the wild-type construct, and a 50% reduction was observed for the s123 variant with three sites scrambled (Figure 6A, right). Therefore, gradual scrambling of the sites progressively reduced TnsB binding to site 4 for promoter control.

The NanoLuc assay provides a quantitative readout of the binding to site 4 (the last binding event in the transposon end recognition) and thus measures the cumulative effect of all sequence modifications introduced in the reporter plasmid. To monitor the individual binding events, we used electromobility shift assay (EMSA), which visualizes complexes with one to three TnsB molecules bound (Arciszewska and Craig, 1991). In these experiments, we used a series of fluorescently labeled variants of the 3BS DNA oligo in which the sequences of selected sites and/or terminal 8 bp were scrambled (schematically shown in Figure 6B, left). Consistent with previously published data (Arciszewska and Craig, 1991), unaltered 3BS DNA formed three distinct complexes containing one to three TnsB molecules (Figure S5). Exchanges of the cognate site sequence with a scrambled sequence led to a loss of higher order protein-DNA complexes. However, even for the substrate with the completely scrambled sequence (3BS s123) complexes with 1, 2, 3, or even 4 protein molecules were formed (Figure S5). A similar effect of sequence modification was observed for shorter substrates. For DNA with two binding sites (2BS), two TnsB complexes were observed in EMSA both for the DNA with native sequence and with both sites and 8 bp sequence scrambled (variant 2BS s12, Figure S5). For DNA corresponding to site 1 and 8 bp sequence (1BS), a single complex was formed with a similar affinity for both the native and scrambled sequence (Figure S5).

We next used densitometry of the EMSA results to quantify the amount of each complex and of the protein-DNA aggregate observed at the highest protein concentrations. The obtained values were fitted to a numerical model of TnsB binding that assumed a sequential binding of up to four TnsB proteins to the DNA substrate and included an empirical model of aggregate formation (Figure S5, see STAR Methods for details). Fitting of the model to our data provided the K_D values for the complexes with one, two, and three TnsB molecules (K_D1–3) (Figures 6B, right and S5). Given the complexity of this system, in particular, stemming from the aggregate formation, which precludes experiments with higher TnsB concentrations, this analysis, especially the determination of K_D2 and K_D3 values, needs to be treated as semi-quantitative. In the case of all 3BS substrates, the K_D1 was similar and in the ~150 nM range (Figures 6B, left and S5), K_D2 was ~200 nM for 3BS and had a similar value for 3BS s2. However, K_D2 was ~2-fold higher for 3BS s23 and 3BS s123 substrates. Finally, the highest values were obtained for K_D3 (365 nM for the 3BS DNA and similar values for 3BS s23 and 3BS s123). This constant was ~2-fold higher for 3BS s2. Therefore, K_D2 and K_D3 were more affected by any of the tested

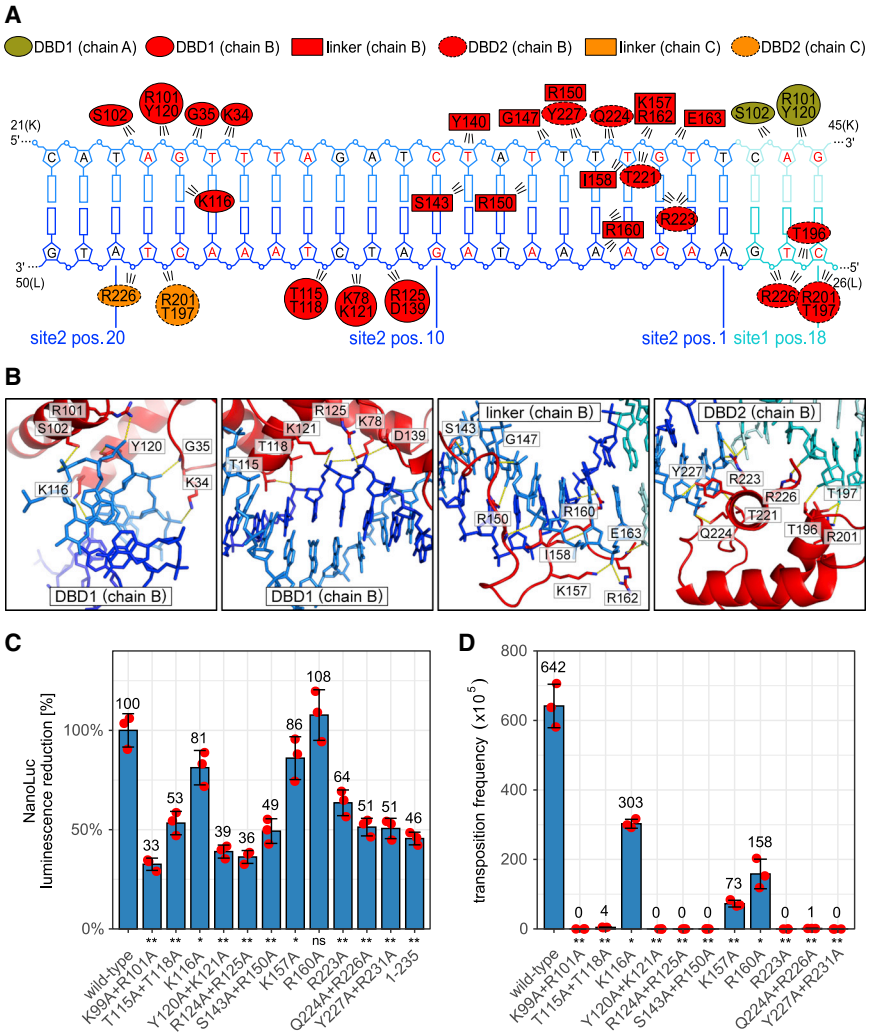


Figure 5. DNA binding by TnsB

(A) Schematic of TnsB-DNA contacts observed in module 2. Conserved nucleotides are shown in red.

(B) Close-up views of the protein-DNA contacts for DBD1, the linker between DBDs, and DBD2 (chain B).

(C) NanoLuc assay for TnsB variants with substitutions of DNA-binding residues. The luminescence reduction upon expression of TnsB variants is plotted with the effect for the wild-type construct normalized to 100%.

(D) *In vivo* lambda hop transposition assay. The frequency of Tn7 transposition is plotted for indicated TnsB variants.

In (C) and (D), values from three independent experiments are shown as mean ± SD. Asterisks indicate the statistical significance between wild type and each mutant (*$p < 0.05$ and **$p < 0.01$, ns, not significant) by the paired sample t test. See also Figure S3

modifications of the site sequence than K_D1. This suggests that changes in the DNA sequence influence the formation of the complex bearing two or more TnsB molecules the most. Similar results were obtained for DNAs with 2BS. For the DNA with 1BS, K_D values for both native and scrambled sequences were similar and higher than for substrates with two or three sites. The modest increase of K_D values upon sequence scrambling is in agreement with the results of the NanoLuc assay. Collectively, these experiments show that TnsB has a preference rather than strict specificity for a particular DNA sequence.

To explore transposon end specificity, we studied the interaction of TnsB with a fragment of the left end which natively has two binding sites separated by 42 bp (Figure S5, substrate 2BSL). We observed the formation of complexes with one, two, and even three TnsB molecules, which implies that the spacer can also be occupied by TnsB, in agreement with this protein's ability to bind DNA with non-cognate sequence. Quantification of EMSA results showed that 2BSL substrate was bound more strongly by TnsB than 3BS (Figure 6B, K_D1 = 85 nM, K_D2 = 109 nM, K_D3 = 293 nM) which may be a result of either higher affinity to the conserved sequences in the left transposon end or the fact that the complex with three TnsB molecules assembles more easily on the spaced binding sites (Figure 6B, right). We note that despite the fact that substrate 2BSL was longer than 3BS, a complex with four TnsB molecules could not be detected. One explanation for this finding could stem from the binding site being 22 nt and the spacer in 2BSL (and native to the left end) being 42 bp, which may be too short to accommodate two proteins. Collectively, these results are consistent with the idea that at higher levels of TnsB, even regions without TnsB-binding sites may be protected via the nonspecific binding activity of TnsB. Thus, the interaction of TnsB with non-conserved sites around the transposon end may contribute to the phenomenon of target immunity (Stellwagen and Craig, 1997a; see discussion). This hypothesis is also in line with the inhibition of transposition by rhamnose-induced TnsB overexpression (Figure S3F). Under such conditions, the large number of TnsB molecules present in the cell could nonspecifically bind the DNA and prevent transposition via an immunity-like process even without transposon ends (see discussion).

We developed a new mate-out transposition assay to verify the effect of modified TnsB-binding sites on *in vivo* transposition where only the right end was modified (STAR Methods and Figure S3G). The wild-type sequence was compared with the variants used for the NanoLuc assay (s1, s2, s12, s23, and s123) and ones with deletion of site 1 (Δ1), sites 1 and 2 (Δ12), and sites 1–3 (Δ123). The transposition frequency for the variants s2 and s23 was diminished by 1–2 orders of magnitude (Figure 6C, right). For variants s1, s12, and s123, the transposition was essentially abolished with a reduction of its frequency by 4–5 orders of magnitude. This demonstrated the essential role of an intact site 1. Deletion of site 1 or sites 1 and 2 (variants Δ1 and

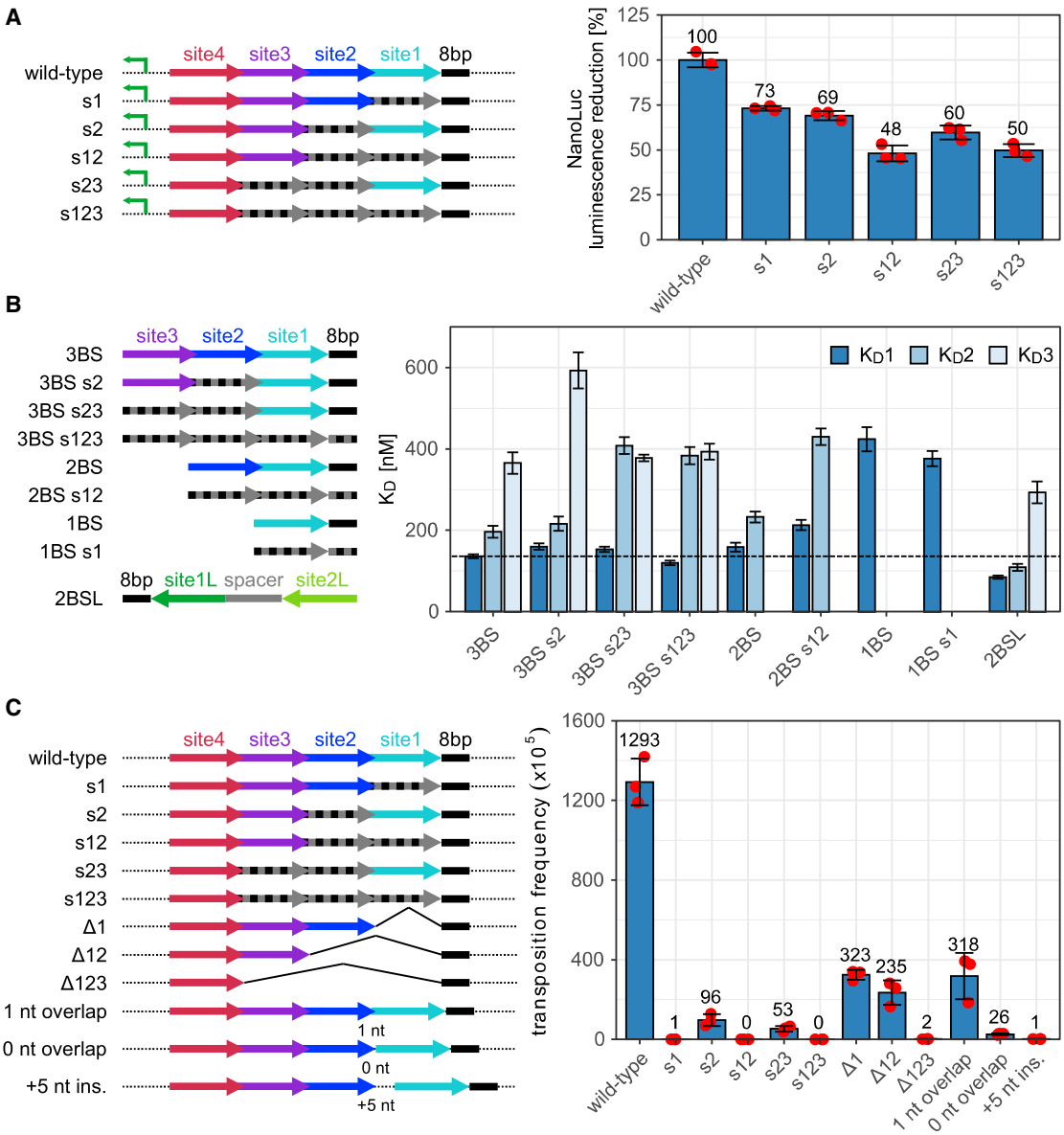


Figure 6. The impact of the sequence of the transposon end on TnsB binding and transposition

(A) NanoLuc assay. Expression constructs used in the experiments are shown schematically in the left panel. Regions in black/gray stripes are scrambled. The right panel shows the luminescence reduction upon expression of TnsB variants with the effect for the wild-type construct normalized to 100%.

(B) EMSA DNA binding assays. Cy5-labeled substrates used in the experiments are shown schematically in the left panel. The right panel shows the estimated K_Ds for each binding event calculated using densitometric quantification of EMSA assays and fitting to a model of binding.

(C) *In vivo* transposition monitored with the mate-out assay. The left panel schematically shows the constructs used in the assay. In the right panel, the transposition frequency is plotted for each variant.

In (A) and (C), values from three independent experiments are shown as mean ± SD. All differences between wild type and each mutant are statistically significant ($p < 0.01$, calculated using a paired sample t test). Values in (B) are calculated using a semi-quantitative numerical model and are plotted as mean ± SE of fits. See also Figures S3 and S5 and Table S4.

Δ12) led to a moderate reduction in transposition activity (~4- and ~6-fold, respectively), indicating that intact sites 2 or 3 can largely take over the role of site 1. Variant Δ123 was again devoid of the ability to transpose, which showed that one functional site is not sufficient. From these experiments, it is apparent that scrambling of the site sequence is more detrimental than deletion of individual sites, as long as two functional sites are present.

As described above, the TnsB-transposon end complex exhibits differences in TnsB conformation and shows differences in the trajectory of the DNA between modules, which are predicted to result from the difference in the overlaps between the

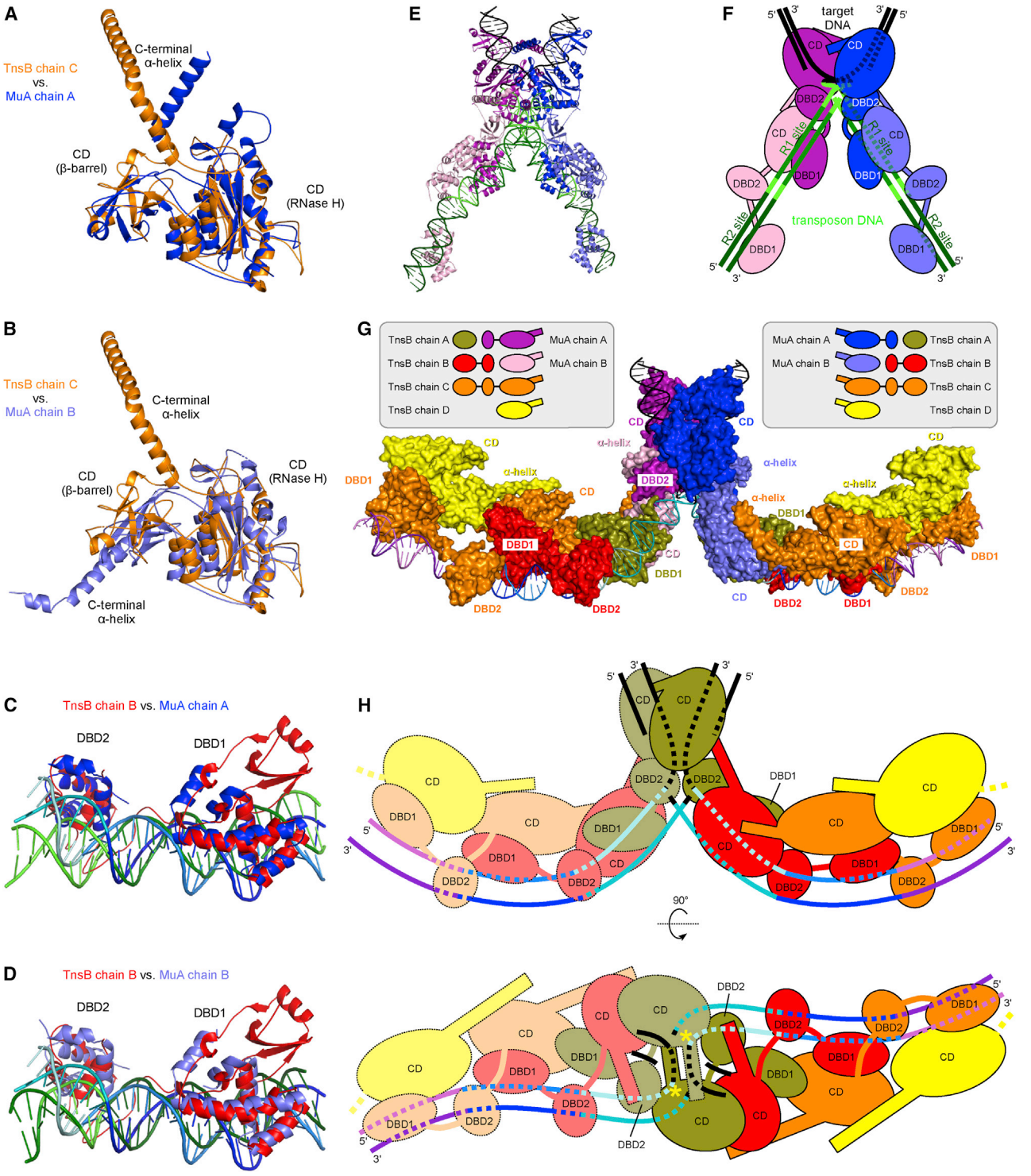


Figure 7. Comparison of TnsB and MuA and model of TnsB strand-transfer complex

(A and B) Superimposition of the CD of TnsB (orange, chain C) on CDs of MuA. For MuA, the catalytically engaged RNase H fold domain (R1-bound chain) is in a darker blue (A), and the non-catalytic domain (R2-bound chain) is in a lighter blue (B).

(C and D) Superimposition of DBDs. The DBD from TnsB subunit B is shown in red with DNA in cyan and blue. MuA chains are colored as in (A) and (B). The transposon DNA from the MuA structure is in green.

(legend continued on next page)

TnsB-binding sites. To test the importance of the site overlap, we generated variants of the mini-transposon in which the overlap between sites 1 and 2 was changed from the native 2 nt overlap to 1 nt or no overlap. We also included a variant with a 5-bp random-sequence insertion between sites 1 and 2. These modifications resulted in a progressive reduction in transposition frequency as follows: ~4-fold for 1 nt overlap, ~50-fold for 0 nt overlap (no overlap), and ~1,000-fold for 5 nt insertion (Figure 6C, right). This demonstrated that the 2 nt overlap characteristic for the right end, and likely the structural rearrangements it induces, are essential for transposition. We propose that these rearrangements serve as a transposon end specificity check. It would explain why transposition is very specific for transposon ends despite the lack of strict DNA sequence specificity of TnsB. These results also identify a feature of the right end which is critical for transposition—overlapping terminal TnsB-binding sites.

Few DNA-binding residues are conserved between canonical and CAST TnsB proteins

We compared *E. coli* TnsB with CAST TnsB proteins using multiple sequence alignments. For these analyses, we used TnsB sequences from diverse classes of CAST elements (Data S2). In the N terminus, two of the analyzed proteins (I-B1 CAST from *Anabaena variabilis* and I-B2 CAST from *Peltigera membranacea* cyanobiont) contain sequences equivalent to the SH3-like domain, but the other proteins lack this element, and the DBD1 is composed of only the helical part. Another feature that varies among the sequences is the length of the linker between DBD1 and DBD2. The linker is 22–23 residue shorter in CAST TnsB than in the *E. coli* variant with the exception of the protein from I-B1 CAST from *Anabaena variabilis*, which is the closest relative of *E. coli* TnsB among analyzed proteins. The shorter linker is a feature that CAST TnsB proteins share with MuA (see below).

The analysis of the alignment also indicated that relatively few residues involved in DNA binding are conserved between *E. coli* and CAST TnsBs. These observations suggest that, despite the fact that overall DNA-binding mode is likely shared among TnsB proteins, the residues involved in protein-DNA interactions may be different. This is in line with the fact that even with closely related elements (i.e., within the I-F3 CRISPR-Cas elements) ends can diverge sufficiently to allow them to be orthogonal (Klompe et al., 2022). This orthogonality is required for evolving the vast diversity of Tn7-like elements (Benler et al., 2021).

Prototypical TnsB is structurally similar to MuA

The prototypic Tn7 TnsB protein described here is distantly related to the Mu phage transposase MuA (Montaño et al., 2012). Both proteins are DDE transposases with activity regulated by interacting ATPase partners and bind to transposon end sequences that contain multiple binding sites. Despite very low sequence identity between TnsB and MuA (approximately 10%), both structures show strong resemblances (Figures 7A–7D).

The catalytic (RNase H fold) and β barrel domains of TnsB and MuA are very similar (Figures 7A and 7B). When the MuA and TnsB structures are superimposed using the RNase H fold, the β barrel domains of TnsB and MuA occupy different positions, related by an ~10 Å shift and ~30° rotation. The position of the β barrel in the MuA structure could not be accommodated in the TnsB-transposon end complex because it would create clashes with the C-terminal helix of TnsB. Differences in the structure and placement of the β barrel subdomain lead to distinct trajectories of the C-terminal helix in MuA and TnsB (Figures 7A and 7B). This helix plays a dual role in MuA, adopting different conformations in the two copies of MuA in the strand-transfer complex (STC). It may play a similar dual role in TnsB (see below).

The two DBDs of TnsB and MuA are very similar (Figures 7C and 7D). The main difference is the presence of the SH3-like domain in TnsB. The corresponding N-terminal region was not visualized in the available crystal structure of the MuA transpososome, but its nuclear magnetic resonance structure has been reported (PDB: 1TNS). The N-terminal domain adopts a winged HTH fold that is involved in DNA binding. As described above, TnsB binding introduces DNA bending, which was not observed in Mu ends that were bound by MuA. Because the two proteins differ by the presence of the SH3-like domain, perhaps it is partly responsible for DNA bending by forming contacts with the β barrel subdomain from the module that interacts with the neighboring site. The spacing between DBD1 and DBD2 on the DNA is the same between prototypical TnsB and MuA. Additionally, both proteins share the mode of DNA binding, despite the very low conservation of the residues that are involved in the interaction with nucleic acid (Data S1). The linker between DBD1 and DBD2 is much shorter in MuA (11 residues instead of 33 residues in TnsB).

Model of TnsB strand-transfer complex

The reported MuA STC structure (PDB: 4FCY, Figure 7E) is two-fold symmetric, and the complete assembly is generated by crystal symmetry. This assembly corresponds to the transposition product with two transposon 3′ ends that are covalently connected to different strands of the target DNA. The shape of the MuA transpososome has been likened to a pair of scissors in

(E) Structure of MuA STC (PDB: 4FCY). One copy of the MuA-DNA complex is shown with the protein in pink cartoon representation (darker shade for the chain with catalytically engaged RNase H fold domain). The second copy generated by crystal symmetry is shown with protein in blue. Transposon DNA is shown as a green ladder, and the target DNA is shown as a black ladder.
(F) Schematic representation of the MuA STC, with domains colored as in (E).
(G) Superimposition of TnsB end-recognition complex (this work) on MuA STC (PDB: 4FCY). Copies of the TnsB-DNA complex were superimposed on each half of the structure of the MuA STC. For this superimposition, the first module of TnsB bound at the transposon end and the corresponding module from MuA was used. For clarity, only selected domains of particular TnsB and MuA chains (indicated in the legend) are shown in surface representation. The DNA is shown as a ladder with the sites colored as in Figure 1 and the target DNA in black.
(H) Schematic representation of the model of TnsB STC based on the superimposition shown in (G) (two views). TnsB chains are colored as in Figure 1B and DNA is colored as in (G). Asterisks indicate the positions of the active sites of the catalytically competent CDs.
See also Data S1.

which the ends of phage DNA form the handles (Montaño et al., 2012) (Figures 7E and 7F). In the structure, these ends comprise the following two MuA binding sites: R1 and R2, numbered from the end of the phage DNA. The target DNA is sharply bent through ~140° and forms the blades of the scissors. In the STC, MuA forms a tetramer and the two protomers of each of the dimers interact with R1 and R2 in a similar manner, using both of their DBDs. However, the position of the CD of each MuA molecule differs. For the MuA chain whose DBDs bind the R2 site, the CD sits on top of the DBDs interacting with site R1 (Figure 7F). The CD of the R1-bound MuA chain is engaged at the phage DNA-target DNA junction. This engagement occurs in *trans*. In the molecule whose DBDs interact with the left end of phage DNA, the CD binds the junction of the right end with the target DNA and vice versa. The C-terminal helices of R1- and R2-bound MuA chains play different roles. In the R1-bound catalytic subunits, the helix stabilizes the bent target DNA and participates in the dimerization of the two catalytic domains. In the non-catalytic subunits, it interacts with DBD2 connected to the catalytic subunit, likely stabilizing its interaction with DNA.

We used the structure of the MuA STC (transposition intermediate) to model the STC of TnsB. The structure of the TnsB-transposon end complex was superimposed on the structure of the MuA transpososome using the TnsB module 1 (i.e., the RNase H fold subdomain of chain B and DBD1 of chain A) and the corresponding module from MuA, consisting of the RNase H domain of the non-catalytic (R2-bound) chain and DBD1 of the catalytic (R1-bound) chain. After this superimposition, the termini of DNA chains from Tn7 and Mu transposon ends were aligned to form proper connections, allowing the MuA module containing the CD of R2-bound MuA and DBD2 from R1-bound MuA, to be neatly accommodated in the portion of TnsB-transposon end complex that contained DBD1 domain from chain A followed by the modules interacting with site 2 and site 3 (Figures 7G and 7H). This procedure was also performed for the other half of the transpososome. The resulting assembly very likely represents the architecture of the STC formed with two right ends of the transposon. Tn7 can use two right ends to perform transposition *in vivo*, in contrast to the configuration with two left ends (Arciszewska et al., 1989). However, the physiological complex (containing the native right and left ends) will likely be asymmetric with two halves of the transposon adopting different conformations. The main difference between the STC found with MuA and TnsB is the trajectory of the DNA (Figures 7E–7H). In MuA, the DNA substrate runs straight, roughly parallel to the target sequence. In TnsB, the protein bends the DNA, so that it turns to run roughly perpendicular to the target DNA. The modeling also implies that TnsB, like MuA, performs the transposition reaction in *trans* (i.e., the subunit that binds one transposon end catalyzes the breakage and joining of the other transposon end). Based on the model, the catalytic TnsB chain would adopt a different conformation with its CD contacting both the transposon end and the target DNA to perform one of the cuts at the transposon end and to catalyze the joining of both DNA fragments. However, the signal that provokes only the TnsB protomer that is bound to the most terminal site to be rearranged and catalytically active remains unknown. We suspect that the driving force for that change may be the presence of the 8 bp sequence flanking both ends of the transposon.

DISCUSSION

A prominent feature of our structure is the bending of the DNA decorated by TnsB molecules. We note that the observed DNA curvature could also be enforced by the nature of the tube-like filaments used for EM studies. However, DNA bending upon TnsB binding has been shown previously in DNA gel mobility assays (Arciszewska and Craig, 1991), and its physiological importance can be supported by several features of Tn7-like systems. It is tempting to speculate that the bending of the DNA through nonspecific binding of closely spaced TnsB molecules could facilitate end engagement during transposition. Active Tn7 transposition complexes, despite their large size, do not require any DNA bending proteins (Bainton, 1993) or supercoiling (Gary et al., 1996) for full transposition activity *in vitro*. This is in contrast to the Mu transposition reaction which requires DNA supercoiling (Pato et al., 1990) and two DNA bending proteins, HU and IHF, for *in vitro* transposition activity (Harshey, 2014). DNA bending is important in many recombination systems (Arinkin et al., 2019). Not having a host factor requirement for DNA bending could contribute to the broad spread of Tn7 across diverse bacteria.

The spacing of TnsB-binding sites is the defining feature differentiating the left and right ends of the element. It is required for *E. coli* Tn7 transposition and is conserved among diverged Tn7 elements (Figure S4) and Tn7/CAST systems. Therefore, the differences in the arrangement of individual TnsB protomers, which result from different overlaps between TnsB-binding sites are key candidates for explaining the polarity of transposition events. We propose that partially overlapping binding sites in the right end of the element alter the configuration of the protomers and cause a distinct bend. This feature may explain why at least one right end is required for transposition.

One unexpected finding of our work is that TnsB exhibits DNA sequence preference rather than strict sequence specificity. The structure revealed very few base-specific protein-DNA contacts, and we showed that substrates with scrambled sequences were also bound by TnsB. The right transposon end comprises overlapping sites, whereas the sites in the left end are separated by intervening DNA sequences. The lack of strict sequence specificity may result in coverage of the whole left end by TnsB. Indeed, our results showed that a fragment of the left end comprising two sites separated by a segment of a random sequence could be bound by three TnsB molecules with affinities comparable to the substrate with three cognate sites. These observations are also in agreement with previous hydroxyl radical footprinting data showing that TnsB binding to the left end protected not only the sites but also, albeit less strongly, the region between the first and the second site (Arciszewska and Craig, 1991).

Nonspecific DNA binding over long stretches of DNA could contribute to the particularly expansive target site immunity found with Tn7 (DeBoy and Craig, 1996). It should be noted that experiments with Mu and Tn7 indicate that target immunity is not solely conducted through DNA (Adzuma and Mizuuchi,

1989; Stellwagen and Craig, 1997a) and instead involves the local concentration of TnsB in solution around a DNA comprising TnsB-binding sites. However, if sites specifically bound by TnsB can also increase the residence time of TnsB in the adjacent DNA by nonspecific DNA binding there should be an increase in the local concentration of TnsB allowing a larger region to be protected with immunity. In a related way, the nonspecific binding will limit the redistribution of TnsB to other parts of the chromosome. Although target immunity will discourage transposition across a wide (>190 kb) region around a previous insertion, transposition into a site ~2 Mb away is not affected (DeBoy and Craig, 1996). Therefore, nonspecific DNA binding could strengthen regional immunity while helping limit immunity at distal sites. Nonspecific TnsB binding could also contribute to a form of "immunity in *trans*" found *in vitro* at high concentrations of TnsB (Skelding, 2003). We find that gross overexpression of TnsB *in vivo* can also strongly inhibit transposition, consistent with the idea that at high levels of TnsB, even without TnsB-binding sites, nonspecific binding can protect DNAs that are distant from the donor element. Validation of these models will require further work.

It has been shown for the right transposon end that with increasing protein concentration, the sites are occupied sequentially by TnsB. Sites 1 and 3 are occupied first, with a slight preference for site 3, followed by site 2, which is bound with lower affinity. Finally, when the first three sites are covered by TnsB, site 4 is occupied (Arciszewska and Craig, 1991). Our model of STC implies that its formation strictly requires the simultaneous occupancy of sites 1 and 2 by TnsB, which are the binding event occurring at higher TnsB concentrations. Our results show that the dissociation constants for these later binding events (K_D2 and especially K_D3) are more sensitive to the modification of the transposon end sequence. Thus, it is possible that multiple sequential binding events serve to amplify a relatively weak TnsB sequence specificity to achieve proper transposon end recognition. Our structure shows that the binding of TnsB to sites 1 and 2 requires a different organization of TnsB molecules and an altered DNA trajectory. It is tempting to speculate that this constitutes another specificity check mechanism. Overcoming the energetic barrier required for these structural rearrangements will only be possible for a proper cognate site sequence and even small perturbations will affect this step.

Transposition of Tn7 elements requires the concerted action of several proteins including TnsA and TnsC, which interact directly with TnsB. The formation of TnsA-TnsB-TnsC assembly involves a triangle of binary contact points. TnsA-TnsB interaction engages TnsB residues 440–480 (Choi et al., 2013), but the structural basis of this binding is not known. The unstructured C-terminal part of TnsB interacts with the TnsB-binding motif of TnsC (Choi et al., 2014). This C-terminal flexible extension of TnsB was present in the protein that we used for the reconstruction of the transposon end-recognition complex, but it is not visible in our structure. In principle, it could be available for TnsC binding. TnsC has a mobile C-terminal extension harboring TnsB- and TnsA-binding motifs. Last but not least, TnsA forms a very stable complex with TnsC (Ronning et al., 2004).

It has been shown that the complex catalyzing DNA breakage and joining contains multiple copies of TnsB and an excess of TnsC and TnsA molecules (Holder and Craig, 2010). This implies that it contains several TnsC heptameric rings and presumably most TnsC protomers recruit TnsA molecules. Nearly all transposon end sites are occupied by TnsB, but we predict that only TnsB chains in the vicinity of the transposon terminus are associated with TnsA to execute transposon end cut. In our structure, the TnsA-binding region of TnsB is covered by the C-terminal helix of the adjacent TnsB chain. We propose that only in the catalytic TnsB subunits, this region is exposed, which enables them to bind TnsA. This in turn can recruit TnsC; therefore, the complete machinery is assembled only at the transposon boundary where excision occurs.

Among the differences between Mu and Tn7, the former uses replicative transposition, which requires the cleavage of only one DNA strand, whereas Tn7's cut-and-paste reaction involves the cleavage of the non-transferred strand by TnsA. We present a model of the STC, which is very likely similar between MuA and Tn7 but does not account for any changes in the architecture that may be imposed by TnsA. This is apparent from the fact that the DNA around the TnsA cut site is not accessible in our model of the STC. Moreover, it has been shown that the pre-transposition complex is less stable than the post-transposition complex (Holder and Craig, 2010; Skelding, 2002) and that transposon end protection changes between the TnsB complex alone and the post-transposition complex (Holder and Craig, 2010). Specific interactions between TnsA and TnsB likely manifest themselves in our finding that multiple substitutions in TnsB have a modest effect on DNA binding as shown by NanoLuc but completely abolish the transposition. This may suggest that, besides their effects on the proper formation of the intertwined TnsB array, the substituted residues also play key roles at other stages of the reaction. Currently, the knowledge about the synaptic complex assembly is very limited. The TnsA-TnsC complex is likely a part of the target complex before transposase recruitment. TnsA also stimulates synapse formation and therefore should be important for the donor complex formation (Choi et al., 2013). A model of this process has been proposed (Shen et al., 2022). All these considerations point to the fact that, very likely, the TnsA-TnsB complex adopts different conformations from that of the STC.

Comparisons of our TnsB complex structure with available structures of other DDE transposases show a remarkable diversity of how these proteins assemble at transposon ends (Hickman and Dyda, 2015). Examples of transposons that contain multiple transposase binding sites in the transposon ends are piggyBack (pB) and Hermes. In pB, both ends comprise terminal inverted repeats, but the sequence arrangement on the two ends differ. A model of end recognition by pB has been proposed (Chen et al., 2020), in which each pB dimer recognizes both ends. The C-terminal cysteine-rich domains form a roughly two-fold symmetric dimer binding one of the ends to a palindromic sequence. Interestingly, catalytic domains are also involved in DNA recognition. This is a very different organization from the head-to-tail arrangement observed for TnsB. Hermes element comprises three sequence repeats in each end. Here, the transposase forms a tetramer of dimers that is predicted to

recognize three repeated end sequences (Hickman et al., 2014). It has been proposed that one catalytic dimer within such octamer binds two ends, whereas other subunits contribute additional interactions with these ends in a complex arrangement.

In summary, our structures provide a mechanistic understanding of how Tn7 TnsB can interpret subtle differences in the spacing of diverged binding sites to impart structural differences in the left and right ends of the element. Full elucidation of these mechanisms will be required to harness Tn7 and Tn7-like elements for CRISPR-guided genome engineering.

Limitations of the study

Our work was limited to TnsB, TnsA-TnsB-TnsC structures will be needed for a comprehensive understanding of Tn7, including TnsA-dependent regulatory steps. Although we can show that a 2 nt overlap between TnsB-binding sites is essential for transposition, our proposal that this provides a critical specificity check will benefit from additional mechanistic interrogation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial cells
- METHOD DETAILS
  - Plasmid preparation, protein expression, purification and spectroscopic characterization
  - Plasmid preparation for *in vivo* assays
  - DNA substrate preparation
  - Cryo-EM and negative-stain EM sample preparation and data collection
  - Cryo-EM data processing
  - Cryo-EM map interpretation and model building
  - Electromobility shift assay
  - Nanoluciferase (NanoLuc) assay
  - *In Vivo* Mate-out Transposition Assays
  - *In vivo* Lambda Hop Transposition Assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.05.005.

ACKNOWLEDGMENTS

We thank Nancy Craig, Nikolaus Grigorieff, and Edward Egelman for critically reading the manuscript. Work in the M.N. lab was financed by the Foundation for Polish Science, TEAM program (POIR.04.04.00-00-20E7/16-00 to M.N.). Work in the J.E.P. lab was funded by the NIH (R01GM129118). This research has been performed using Centre for Preclinical Research and Technology (CePT) infrastructure (European Union project) (POIG.02.02.00-14-024/08-00). Funding for open access charges was obtained from the Foundation for Polish Science, TEAM program (POIR.04.04.00-00-20E7/16-00). This publication was developed under the provision of the Polish Ministry of Education and Science project: "Support for research and development with the use of research infrastructure of the National Synchrotron Radiation Centre SOLARIS" under contract no. 1/SOL/2021/2. We acknowledge SOLARIS Centre for access to the cryo-EM beamline, where the measurements were performed.

AUTHOR CONTRIBUTIONS

Z.K. prepared constructs, expressed and purified TnsB protein variants, prepared samples for cryo-EM imaging, performed biochemical experiments and DNA-binding assays, and built and refined the atomic model. M.C.-C. prepared and screened the cryo-EM grids and processed the cryo-EM images to obtain the 3D reconstructions. K.M.G.-M. established the protein purification protocols and prepared the samples for a pilot dataset. R.J.W. established and performed the *in vivo* binding and transposition assays. J.J. and W.Z. purified the TnsB protein variants. J.J. performed the DNA binding assays. M.R. collected the EM data. J.T.P. prepared the numerical model for EMSA analysis. T.G. performed the initial EM analysis. J.E.P. and M.N. supervised the project and wrote the manuscript with contributions from all authors.

DECLARATION OF INTERESTS

The J.E.P. lab has corporate funding for research that is not directly related to the work in this publication. Cornell University has filed patent applications with J.E.P. as the inventor involving CRISPR-Cas systems associated with transposons that are not directly related to this work.

Received: September 23, 2021
Revised: March 2, 2022
Accepted: May 3, 2022
Published: June 1, 2022

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

Adzuma, K., and Mizuuchi, K. (1989). Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. Cell *57*, 41–47. https://doi.org/10.1016/0092-8674(89)90170-0.

Arciszewska, L.K., and Craig, N.L. (1991). Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. Nucleic Acids Res. *19*, 5021–5029. https://doi.org/10.1093/nar/19.18.5021.

Arciszewska, L.K., Drake, D., and Craig, N.L. (1989). Transposon Tn7. cis-Acting sequences in transposition and transposition immunity. J. Mol. Biol. *207*, 35–52. https://doi.org/10.1016/0022-2836(89)90439-7.

Arciszewska, L.K., McKown, R.L., and Craig, N.L. (1991). Purification of TnsB, a transposition protein that binds to the ends of Tn7. J. Biol. Chem. *266*, 21736–21744. https://doi.org/10.1016/S0021-9258(18)54698-6.

Arinkin, V., Smyshlyaev, G., and Barabas, O. (2019). Jump ahead with a twist: DNA acrobatics drive transposition forward. Curr. Opin. Struct. Biol. *59*, 168–177. https://doi.org/10.1016/j.sbi.2019.08.006.

Asarnow, D., Palovcak, E., and Cheng, Y. (2019). asarnow/pyem: UCSF pyem v0.5. Zenodo. https://doi.org/10.5281/ZENODO.3576630.

Bainton, R.J., Kubo, K.M., Feng, J.N., and Craig, N.L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. Cell *72*, 931–943. https://doi.org/10.1016/0092-8674(93)90581-A.

Benler, S., Faure, G., Altae-Tran, H., Shmakov, S., Zheng, F., and Koonin, E. (2021). Cargo genes of Tn *7* -Like transposons comprise an enormous diversity of defense systems, mobile genetic elements, and antibiotic resistance genes. mBio *12*, e0293821. https://doi.org/10.1128/mBio.02938-21.

Blanchet, C., Pasi, M., Zakrzewska, K., and Lavery, R. (2011). CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. Nucleic Acids Res. *39*, W68–W73. https://doi.org/10.1093/nar/gkr316.

Chen, Q., Luo, W., Veach, R.A., Hickman, A.B., Wilson, M.H., and Dyda, F. (2020). Structural basis of seamless excision and specific targeting by piggyBac transposase. Nat. Commun. *11*, 3446. https://doi.org/10.1038/s41467-020-17128-1.

Choi, K.Y., Li, Y., Sarnovsky, R., and Craig, N.L. (2013). Direct interaction between the TnsA and TnsB subunits controls the heteromeric Tn7 transposase. Proc. Natl. Acad. Sci. USA *110*, E2038–E2045. https://doi.org/10.1073/pnas.1305716110.

Choi, K.Y., Spencer, J.M., and Craig, N.L. (2014). The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. Proc. Natl. Acad. Sci. USA *111*, E2858–E2865. https://doi.org/10.1073/pnas.1409869111.

Craig, N.L. (2002). Mobile DNA II (ASM Press).

DeBoy, R.T., and Craig, N.L. (1996). Tn7 transposition as a probe of cis interactions between widely separated (190 kilobases apart) DNA sites in the Escherichia coli chromosome. J. Bacteriol. *178*, 6184–6191. https://doi.org/10.1128/jb.178.21.6184-6191.1996.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501. https://doi.org/10.1107/S0907444910007493.

England, C.G., Ehlerding, E.B., and Cai, W. (2016). NanoLuc: A small luciferase is brightening up the field of bioluminescence. Bioconjug. Chem. *27*, 1175–1187. https://doi.org/10.1021/acs.bioconjchem.6b00112.

Faure, G., Shmakov, S.A., Yan, W.X., Cheng, D.R., Scott, D.A., Peters, J.E., Makarova, K.S., and Koonin, E.V. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. Nat. Rev. Microbiol. *17*, 513–525. https://doi.org/10.1038/s41579-019-0204-7.

Gary, P.A., Biery, M.C., Bainton, R.J., and Craig, N.L. (1996). Multiple DNA processing reactions underlie Tn7Transposition. J. Mol. Biol. *257*, 301–316. https://doi.org/10.1006/jmbi.1996.0164.

Goormaghtigh, E., Gasper, R., Bénard, A., Goldsztein, A., and Raussens, V. (2009). Protein secondary structure content in solution, films and tissues: redundancy and complementarity of the information content in circular dichroism, transmission and ATR FTIR spectra. Biochim. Biophys. Acta *1794*, 1332–1343. https://doi.org/10.1016/j.bbapap.2009.06.007.

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. *53*, 245–281. https://doi.org/10.1146/annurev.micro.53.1.245.

Harshey, R.M. (2014). Transposable phage Mu. Microbiol. Spectr. *2*. https://doi.org/10.1128/microbiolspec.MDNA3-0007-2014.

Hauer, B., and Shapiro, J.A. (1984). Control of Tn7 transposition. Mol. Gen. Genet. *194*, 149–158. https://doi.org/10.1007/BF00383510.

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. https://doi.org/10.3109/10409230903505596.

Hickman, A.B., and Dyda, F. (2015). Mechanisms of DNA transposition. Microbiol. Spectr. *3*. MDNA3–M0034. https://doi.org/10.1128/microbiolspec.MDNA3-0034-2014.

Hickman, A.B., Ewis, H.E., Li, X., Knapp, J.A., Laver, T., Doss, A.L., Tolun, G., Steven, A.C., Grishaev, A., Bax, A., et al. (2014). Structural basis of hAT transposon end recognition by Hermes, an octameric DNA transposase from Musca domestica. Cell *158*, 353–367. https://doi.org/10.1016/j.cell.2014.05.037.

Hickman, A.B., Li, Y., Mathew, S.V., May, E.W., Craig, N.L., and Dyda, F. (2000). Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. Mol. Cell *5*, 1025–1034. https://doi.org/10.1016/S1097-2765(00)80267-1.

Holder, J.W., and Craig, N.L. (2010). Architecture of the Tn7 posttransposition complex: an elaborate nucleoprotein structure. J. Mol. Biol. *401*, 167–181. https://doi.org/10.1016/j.jmb.2010.06.003.

Huber, S.T., Kuhm, T., and Sachse, C. (2018). Automated tracing of helical assemblies from electron cryo-micrographs. J. Struct. Biol. *202*, 1–12. https://doi.org/10.1016/j.jsb.2017.11.013.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858. https://doi.org/10.1038/nprot.2015.053.

Klompe, S.E., Jaber, N., Beh, L.Y., Mohabir, J.T., Bernheim, A., and Sternberg, S.H. (2022). Evolutionary and mechanistic diversity of Type I-F CRISPR-associated transposons. Mol. Cell *82*, 616–628.e5. https://doi.org/10.1016/j.molcel.2021.12.021.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature *571*, 219–225. https://doi.org/10.1038/s41586-019-1323-z.

Lichtenstein, C., and Brenner, S. (1982). Unique insertion site of Tn7 in the E. coli chromosome. Nature *297*, 601–603. https://doi.org/10.1038/297601a0.

May, E.W., and Craig, N.L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. Science *272*, 401–404. https://doi.org/10.1126/science.272.5260.401.

McKown, R.L., Orle, K.A., Chen, T., and Craig, N.L. (1988). Sequence requirements of Escherichia coli attTn7, a specific site of transposon Tn7 insertion. J. Bacteriol. *170*, 352–358. https://doi.org/10.1128/jb.170.1.352-358.1988.

Montaño, S.P., Pigli, Y.Z., and Rice, P.A. (2012). The μ transpososome structure sheds light on DDE recombinase evolution. Nature *491*, 413–417. https://doi.org/10.1038/nature11602.

Nowotny, M. (2009). Retroviral integrase superfamily: the structural perspective. EMBO Rep. *10*, 144–151. https://doi.org/10.1038/embor.2008.256.

Pato, M.L., Howe, M.M., and Higgins, N.P. (1990). A DNA gyrase-binding site at the center of the bacteriophage mu genome is required for efficient replicative transposition. Proc. Natl. Acad. Sci. USA *87*, 8716–8720. https://doi.org/10.1073/pnas.87.22.8716.

Pei, J., Kim, B.-H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. *36*, 2295–2300. https://doi.org/10.1093/nar/gkn072.

Petassi, M.T., Hsieh, S.-C., and Peters, J.E. (2020). Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons. Cell *183*, 1757–1771.e18. https://doi.org/10.1016/j.cell.2020.11.005.

Peters, J.E. (2014). Tn7. Microbiol. Spectr. *2*, 1–20. https://doi.org/10.1128/microbiolspec.MDNA3-0010-2014.

Peters, J.E., Makarova, K.S., Shmakov, S., and Koonin, E.V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc. Natl. Acad. Sci. USA *114*, E7358–E7366. https://doi.org/10.1073/pnas.1709035114.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX : structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

Reyes, O., Beyou, A., Mignotte-Vieux, C., and Richaud, F. (1987). Mini-Mu transduction: cis-inhibition of the insertion of Mud transposons. Plasmid *18*, 183–192. https://doi.org/10.1016/0147-619X(87)90061-8.

Robinson, M.K., Bennett, P.M., and Richmond, M.H. (1977). Inhibition of TnA translocation by TnA. J. Bacteriol. *129*, 407–414. https://doi.org/10.1128/jb.129.1.407-414.1977.

Ronning, D.R., Li, Y., Perez, Z.N., Ross, P.D., Hickman, A.B., Craig, N.L., and Dyda, F. (2004). The carboxy-terminal portion of TnsC activates the Tn7 transposase through a specific interaction with TnsA. EMBO J. *23*, 2972–2981. https://doi.org/10.1038/sj.emboj.7600311.

Saito, M., Ladha, A., Strecker, J., Faure, G., Neumann, E., Altae-Tran, H., Macrae, R.K., and Zhang, F. (2021). Dual modes of CRISPR-associated transposon homing. Cell *184*, 2441–2453.e18. https://doi.org/10.1016/j.cell.2021.03.006.

Sarnovsky, R.J., May, E.W., and Craig, N.L. (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. *15*, 6348–6361. https://doi.org/10.1002/j.1460-2075.1996.tb01024.x.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

Shen, Y., Gomez-Blanco, J., Petassi, M.T., Peters, J.E., Ortega, J., and Guarné, A. (2022). Structural basis for DNA targeting by the Tn7 transposon. Nat. Struct. Mol. Biol. *29*, 143–151. https://doi.org/10.1038/s41594-022-00724-8.

Skelding, Z., Queen-Baker, J., and Craig, N.L. (2003). Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. EMBO J. *22*, 5904–5917. https://doi.org/10.1093/emboj/cdg551.

Skelding, Z., Sarnovsky, R., and Craig, N.L. (2002). Formation of a nucleoprotein complex containing Tn7 and its target DNA regulates transposition initiation. EMBO J. *21*, 3494–3504. https://doi.org/10.1093/emboj/cdf347.

Stellwagen, A.E., and Craig, N.L. (1997a). Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. EMBO J. *16*, 6823–6834. https://doi.org/10.1093/emboj/16.22.6823.

Stellwagen, A.E., and Craig, N.L. (1997b). Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics *145*, 573–585.

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science *365*, 48–53. https://doi.org/10.1126/science.aax9181.

Turlan, C., and Chandler, M. (2000). Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. Trends Microbiol. *8*, 268–274. https://doi.org/10.1016/S0966-842X(00)01757-1.

Vo, P.L.H., Acree, C., Smith, M.L., and Sternberg, S.H. (2021). Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. Mobile DNA *12*, 13. https://doi.org/10.1186/s13100-021-00242-2.

Watson, E.R., Grace, C.R.R., Zhang, W., Miller, D.J., Davidson, I.F., Prabu, J.R., Yu, S., Bolhuis, D.L., Kulko, E.T., Vollrath, R., et al. (2019). Protein engineering of a ubiquitin-variant inhibitor of APC/C identifies a cryptic K48 ubiquitin chain binding site. Proc. Natl. Acad. Sci. USA *116*, 17280–17289. https://doi.org/10.1073/pnas.1902889116.

Wickham, H. (2016). ggplot2, use R! (Springer International Publishing). https://doi.org/10.1007/978-3-319-24277-4.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332. https://doi.org/10.1038/nmeth.4193.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166. https://doi.org/10.7554/eLife.42166.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli Top10	Invitrogen	N/A
E. coli BL21 Star (DE3)	Invitrogen	N/A
E. coli DH5α	J. Peters Lab	N/A
E. coli BW27783	Coli Genetic Stock Center (CGSC)	Cat#12119
E. coli CW51	J. Peters Lab	JP606
Chemicals, peptides, and recombinant proteins		
Auto induction media terrific broth base including trace elements	Formedium	Cat#AIMSB0260
Viscolase	A&A Biotechnology	Cat#1010-100
Lysosyme	Roth	Cat#8259.3
Leupeptin	Roth	Cat#CN33.2
Aprotinin	Roth	Cat# A162.3
PMSF	Roth	Cat# 6367.2
Pepstatin A	Roth	Cat# 2936.2
Uranyl acetate solution 2%	Electron Microscopy Sciences	Cat#22400-2
Glutaraldehyde solution	Sigma	Cat#G7651-10ML
EagI restriction enzyme	NEB	Cat#R0505S
AleI restriction enzyme	NEB	Cat#R0685S
Nano luciferase	Promega	Cat#N150A
Nano-Glo Luciferase Assay System reagent	Promega	Cat#N1610
Critical commercial assays		
KOD Hot Start DNA polymerase	Merck	Cat#71086-4
NEBuilder® HiFi DNA Assembly Master Mix	NEB	Cat#E2621L
Q5® High-Fidelity DNA polymerase	NEB	Cat#M0491L
Deposited data		
Atomic model for *E. coli* TnsB in complex with right transposon end	Protein Data Bank	7PIK
Cryo-EM reconstruction for *E. coli* TnsB in complex with right transposon end	EM Data Bank	EMD-13439
Helical reconstruction for *E. coli* TnsB in complex with right transposon end	EM Data Bank	EMD-13440
Oligonucleotides		
DNA oligos for structure determination and for EMSA binding assay	This study	Table S4
Recombinant DNA		
TnsB expression plasmids	This study	Table S3
Plasmids for *in vivo* transposition assays	This study	Table S5
Plasmids for NanoLuc assays	This study	Table S5
Software and algorithms		
cryoSPARC v3.3.1	(Punjani et al., 2017)	https://cryosparc.com/
RELION-3.1	(Zivanov et al., 2018)	https://github.com/3dem/relion
UCSF ChimeraX v1.2	(Pettersen et al., 2021)	https://www.cgl.ucsf.edu/chimerax/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PyMOL 2.3.1	Schrödinger, LCC	https://pymol.org
Curves+	(Blanchet et al., 2011)	http://curvesplus.bsc.es/
ImageJ 1.51q	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
Phyre2	(Kelley et al., 2015)	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
Coot 0.9.4.1 EL	(Emsley et al., 2010)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix 1.19.2-4158	(Adams et al., 2010)	http://www.phenix-online.org/
EPU_group_AFIS	Dustin Morado	https://github.com/DustinMorado/EPU_group_AFIS
R version 4.1.2	R Core Team	https://www.R-project.org/
ggplot2 package v3.3.5	(Wickham, 2016)	https://ggplot2.tidyverse.org/
PROMALS3D	(Pei et al., 2008)	http://prodata.swmed.edu/promals3d/promals3d.php
Other		
HisTrap HP column	Cytiva	Cat#17-5248-02
HiLoad 16/600 Superdex 200 pg	Cytiva	Cat#28-9893-35
HiLoad 16/600 Superose 6 pg	Cytiva	Cat#29-3239-52
Lacey carbon film on 300 Cooper mesh	Ted Pella	Cat#01895-F
carbon-coated 300-mesh cooper grid	Ted Pella	Cat#01843-F

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Marcin Nowotny (mnowotny@iimcb.gov.pl).

Material availability

All reagents generated in this study are available from the lead contact.

Data and code availability

- Atomic model and the corresponding cryo-EM reconstruction are available in the Protein Data Bank (PDB) and EM Data Bank under the accession codes: 7PIK and EMD-13439, respectively. The helical reconstruction is available in the EM Data Bank under the accession code: EMD-13440.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial cells

E. coli Top10 (Invitrogen) were grown in LB at 37 °C.

E. coli BL21 Star (DE3) (Invitrogen) were grown in LB at 37 °C. To induce protein expression, they were transformed by the selected plasmids listed in the Table S3, propagated at 37 °C in auto induction media at 37 °C, and induced at 18 °C.

E. coli DH5α were grown in LB at 37 °C.

E. coli BW27783 (Yale Coli Genetic Stock Center) were grown in LB at 37 °C.

METHOD DETAILS

Plasmid preparation, protein expression, purification and spectroscopic characterization

Wild-type *E. coli* TnsB (residues 1-702) was cloned into the pET28-6×His-SUMO expression vector. TnsB variants listed in the Table S3 were prepared via SLIC (Sequence and Ligation Independent Cloning) method. All proteins were expressed in *E. coli* BL21 Star (DE3) cells for 72 h in auto induction media terrific broth base including trace elements (Formedium) supplemented with 2% (v/v) glycerol at

18 °C. Cells were resuspended in buffer A (40 mM TRIS [pH 7.5], 500 mM NaCl, 10% [v/v] glycerol, 30 mM imidazole, and 1 mM DTT) with the addition of lysozyme (final concentration 1 mg/ml), viscolase (10 μl/50 ml lysate), and protease inhibitors (5 μg/ml leupeptin, 2 μg/ml aprotinin, 1 mM PMSF, 1 μM pepstatin A). The suspension was sonicated, and the soluble fraction was loaded on a 5 ml NiNTA column (GE Healthcare) using the AKTA Pure system (GE Healthcare). The column was washed with buffer A, and the protein was recovered with 30% of buffer B (40 mM TRIS [pH 7.5], 500 mM NaCl, 10% [v/v] glycerol, 500 mM imidazole, and 1 mM DTT). To remove the 6xHis-SUMO tag, the fusion protein was digested with SUMO protease and dialyzed overnight at 4 °C against buffer A without imidazole. The cleaved protein was recovered from the flow-through fraction after loading on a 5 ml HisTrap column. The protein was precipitated with $(NH_4)_2SO_4$ (0.4 g/ml), resuspended in buffer C (20 mM HEPES [pH 7.5], 10% [v/v] glycerol, and 0.5 mM TCEP), and loaded on a size exclusion column (HiLoad 16/600 Superdex 200 or HiLoad 16/600 Superose 6, GE Healthcare) equilibrated with buffer C. Pooled fractions were either stored at 4 °C or flash frozen in liquid nitrogen and stored at -20 °C.

Secondary structure content of selected variants was measured using Fourier transform infrared spectroscopy (TENSOR Series, Bruker). This method provides the same information as other spectroscopic methods, in particular circular dichroism (Goormaghtigh et al., 2009)

Plasmid preparation for *in vivo* assays

Cloning was performed in *E. coli* DH5α, and plasmids were transferred to RW0163 (BW27783 from Yale Coli Genetic Stock Center, #12119) for *in vivo* transposition assays. Overnight cultures were grown in LB media supplemented with 0.2% glycerol and 0.2% glucose. Strains with plasmids were maintained with antibiotic selection in LB medium supplemented with 0.2% glycerol and 0.2% glucose. Assembly of *tnsABCD* expression vector (pRW97) was accomplished using Gibson assembly (NEBuilder® HiFi, New England Biolabs, Ipswich, MA) of *tnsABCD* PCR product amplified from pCW4 and pBAD322 backbone. Mutations within *tnsB* were introduced using overlap PCR site-directed mutagenesis to construct pRW100-pRW112. To construct pRW116-pRW128, mutant versions of *tnsB* were PCR amplified from pRW100-pRW112, digested with EagI and AleI, and ligated into pRW97 digested with EagI and AleI.

Cloning of the Nano luciferase (Promega) reporter construct was performed using Gibson assembly with PCR products of Nano luciferase, the right end of a miniTn7(Kan) element, and pBBR backbone. Variants in the right end of the miniTn7(Kan) transposable element were generated using overlap PCR site-directed mutagenesis to construct pRW207-pRW211.

Construction of a miniTn7(Kan) donor (pRW221) for the mate-out assay was performed using Gibson assembly of a miniTn7(Kan) PCR product and pTSC29 backbone. Variants in the right end of the miniTn7(Kan) transposable element were generated using overlap PCR site-directed mutagenesis to construct pRW222-pRW232.

DNA substrate preparation

All double-stranded DNA substrates (Metabion, Table S4), either for structural studies or electrophoretic mobility shift assays (EMSAs), were prepared in the same way by annealing two complementary single-stranded DNA oligos in water. DNA was heated to 95 °C in a heat block, then slowly cooled to room temperature.

Cryo-EM and negative-stain EM sample preparation and data collection

The *E. coli* TnsB wild-type protein variant (2.4 mg/ml) was mixed with 3BS DNA at a 3:1 protein:DNA molar ratio, diluted 10× with 20 mM HEPES [pH 7.5] and 150 mM NaCl, and incubated overnight at 4 °C. Reconstituted filaments were fixed with 0.05% glutaraldehyde for 60 min by a 1:1 dilution with 20 mM HEPES [pH 7.5], 150 mM NaCl, and 0.1% glutaraldehyde, concentrated, and applied to a glow-discharged Lacey carbon film on 300 Cooper mesh (Ted Pella 01895-F). The sample was vitrified in liquid ethane using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) at 4 °C with 95% humidity, 4 s blot time, and -5 blot force. Data collection was performed on a Titan Krios G3i electron microscope (Thermo Fisher Scientific) operating at 300 kV and equipped with a BioQuantum energy filter (with 20 eV energy slit) and K3 camera (Gatan) at the SOLARIS National Synchrotron Radiation Centre in Krakow, Poland. Movies were collected with AFIS, with a nominal magnification of 105,000× (corresponding to a physical pixel size of 0.86 Å), 50 μm C2 aperture, and retracted objective aperture. The defocus range was set to -0.5 to -2.0 μm. The total dose (fractionated into 40 frames) was 40 e/Å^2. The dose rate was 15.5 e per pixel per second, measured without sample (in vacuum).

For negative-stain EM, TnsB-DNA complexes containing wild-type protein or deletion variant (Δ245-251) were prepared as described above, diluted 10× and applied to a glow-discharged carbon-coated 300-mesh cooper grid (TED PELLA 01843-F) and stained with 1% uranyl acetate. TEM images were acquired with Tecnai T12 BioTWIN 120 kV Electron Microscope (FEI) with acceleration voltage of 120 kV.

Cryo-EM data processing

Electron microscopy analysis of TnsB-DNA complexes revealed for one of the tested substrates the formation of long, well-defined filamentous structures (Figure S1A). To investigate the architecture of these filaments, we collected a pilot dataset and processed it with cryoSPARC software (Punjani et al., 2017). Briefly, 2D templates were created from 2D classifications of particles that were picked with template-free *Filament Tracer (BETA)* (Huber et al., 2018). The *ab initio* 3D model was created directly from the particle stack and refined with *Helical Refinement*. This preliminary helical reconstruction revealed the overall morphology of the TnsB-DNA filament. It is a tube-like, two-start, left-handed helix composed of repeating segments, each comprising of two copies of TnsB-DNA

complex. These two copies are related by C2 symmetry (Figure S1B). In the filament, each copy is arranged head-to-tail with TnsB-DNA complexes from adjacent subunits. The cross-section of the filament is roughly elliptical, with a minor and major axes of 150 and 180 Å, respectively. The calculated helical parameters (helical twist: +12.5°, helical rise: 112.9 Å) result in a very long helical pitch of the filament (163 nm, Figure S1C). This explains the varying thickness of the TnsB-DNA filaments that we observed in the raw micrographs (Figure S1A, blue/red circles) and in the 2D classes (Figure S1C, right). The low stability and high flexibility of the filament, along with the relatively low resolution (7-8 Å) of the obtained 2D classes, suggested that the interactions stabilizing the filament structure are weak. Indeed, averaging over helical symmetry reduced the resolution of the 3D reconstructions obtained from the pilot dataset. Therefore, the final reconstruction used for the model building was calculated without imposing any symmetry. In addition, the same set of particles was used to calculate the helical reconstruction that showed lower resolution, but allowed us to better visualize the interactions stabilizing the filament.

For the final reconstruction, a total of 6,692 movies were collected and processed with RELION-3.1 (Zivanov et al., 2018) and cryoSPARC (Figures S1D–S1F). Raw movies were motion-corrected and binned 2x using RELION's implementation of MotionCor2 software (Zheng et al., 2017). Micrographs were imported into cryoSPARC, and a contrast transfer function (CTF) was fitted with *Patch CTF* estimation. A total of 1.14 million particles were picked with *Filament Tracer* (BETA) based on 2D templates from pilot dataset. After three rounds of reference-free 2D classification, 510,000 selected particles were used for initial 3D refinement with C2 symmetry in cryoSPARC. The resulting 3.8 Å reconstruction contained a single segment of the filament with two TnsB-DNA complexes.

To improve the quality of the reconstruction, refined particles were subjected to two rounds of Bayesian polishing and CTF refinements (Figure S1E). Iterative switching between cryoSPARC and RELION software was required, because the polishing step is implemented only in RELION and we found that cryoSPARC performed better at global alignment of TnsB-DNA particles. First, particles were re-imported into RELION's pipeline with scripts from UCSF pyem (Asarnow et al., 2019) and subjected to the round I of *Bayesian polishing* (with re-extraction with an unbinned pixel size of 0.86 Å/pixel and a box size of 384 pixels). Resulting shiny particles were imported into cryoSPARC and cleaned with the final round of 2D classification. A *NU-Refinement* of 450,748 selected particles improved the resolution to 3.4 Å. Since Bayesian polishing is more effective for higher resolution maps, next refinement round was performed with a single TnsB-DNA complex. To this end, aligned particles were subjected to C2 *Symmetry Expansion* (extraction of asymmetric molecules representing single TnsB-DNA complexes), which doubled the number of unique, asymmetrical projections to 901,496. C2-expanded particles were re-imported into RELION and subjected to CTF refinements. Subsequent 3D refinement with local searches resulted in a reconstruction corresponding to a single DNA duplex decorated with TnsB protomers at 3.0 Å resolution. To avoid duplication of the data in the final helical refinement and to streamline data processing, the second round of *Bayesian polishing* was preceded with a symmetry de-expansion procedure (i.e., the removal of every second row from the STAR file, resulting in a reduction of the particle number from 901,496 to 450,748). After the second round of polishing (with re-extraction with a box size of 512 pixels) the resolution of C2-symmetrical *NU-Refinement* improved to 3.1 Å. This is a slightly worse resolution than for the asymmetrical reconstruction, but it shows that the two rounds of Bayesian polishing and CTF refinements significantly improved the alignment of particles, which can be assessed by the increase in the resolution of C2-symmetrical reconstruction from 3.8 to 3.1 Å.

To further improve the quality of the asymmetrical reconstruction and to facilitate the alignment of a single module during the final local 3D refinement, all C2-related copies of the TnsB-DNA complex from each picked segment were removed from the images with a *Signal Subtraction* procedure (Figure S1F, yellow boxes). To this end, aligned particles were again subjected to a C2 *Symmetry Expansion* followed by a *Local Refinement* that resulted in a 2.85 Å asymmetrical reconstruction. Refined particles were re-imported into RELION, divided into 15 optics groups based on the metadata from an EPU session with the aid of the EPU_Group_AFIS.py script (https://github.com/DustinMorado/EPU_group_AFIS), and subjected to a second round of CTF refinements, followed by a 3D refinement with local searches. The resulting reconstruction was removed from the images with a *Signal Subtraction* procedure. Subtracted particles were imported into cryoSPARC, divided into 15 optics groups, and subjected to a local and global CTF refinements. Final *Local Refinement* with a C2-rotated 3D reference and dynamic masking resulted in a 2.7 Å asymmetrical reconstruction containing a single TnsB-DNA complex. This final map was sharpened locally with *Local Filtering* tool in cryoSPARC with a B factor of -70 $Å^2$ and a maximum resolution for sharpening of 2.2 Å (Figure S2A).

The regions involved in the interactions between the individual TnsB-DNA complexes within the filament were not well resolved in the asymmetrical reconstruction, presumably due to the misalignment between the adjacent segments. To better visualize the molecular interactions stabilizing the filamentous structure, a helical reconstruction was calculated from the refined particles after the second round of Bayesian polishing (Figure S1F, pink box). To this end, refined particles were imported into cryoSPARC and subjected to local and global CTF refinements. Subsequent *Helical Refinement* resulted in a 3.8 Å reconstruction that was sharpened globally with a B factor of -116.6 $Å^2$ (Figure S2B). It is worth noting that the resolution of this helical reconstruction is worse than the resolution of all previous refinements, but since it contains three adjacent filament segments, it reveals additional information on the interactions between segments that was not available from the other reconstructions.

Analysis of the helical reconstruction confirmed that the filament is stabilized mainly by the intertwined nature of the TnsB-DNA complexes. The C-terminal α-helix connecting adjacent CD domains within a single TnsB-DNA complex is also involved in the interaction between adjacent segments (Figure S2B, upper inset). In addition, low resolution density of the DBD2 domain could be detected on the 8 bp terminal region of the DNA duplex (Figure S2B, lower inset, olive color density), in line with the assignment of

the DBD1 domain interacting with site 1 (chain A, olive color in Figure 1) to the protein chain containing CD domain interacting with site 3 of the adjacent segment (counterpart of the chain D, yellow color in Figure 1; see also Figure S2D). Importantly, we did not observe any interactions stabilizing the filament other than the interactions between the adjacent TnsB protomers already described in the results section.

All reported resolutions were estimated from gold-standard masked FSC at the 0.143 threshold. Data collection and processing statistics are presented in Table S1.

Cryo-EM map interpretation and model building

The 2.7 Å asymmetrical reconstruction of TnsB-DNA complex after local sharpening was used for *de novo* model building in Coot (Emsley et al., 2010) with the aid of a homology model generated on Phyre2 website (Kelley et al., 2015) based on the MuA structure (PDB: 4FCY) (Montaño et al., 2012) and refined using real space refinement in Phenix (Adams et al., 2010). The quality of EM potential map was sufficient to unambiguously determine the register of DNA sequence and build most of the side chains (Figures S2A–S2C). The local resolution was the highest for the core element of the TnsB proteins interacting with DNA (DBD1, linker, and β barrel part of CD domains along with all three sites of DNA) and worse in external regions of the reconstruction (DBD2, RNase H part of CD domains, and terminal 8 bp part of DNA) (Figures S2A and S2B).

Key for the interpretation of the EM maps was the link between DBD2 and the CD in each of the protein chains. The DBD2 and CD linkage was not fully resolved in our reconstruction and could have been explained by either of two possibilities. One possibility was a tiled arrangement of domains between modules (Figures 1D and S2D), where the distance between the C-terminus of DBD2 and N-terminus of the CD is ~45 Å. A second, alternative possibility was that TnsB domains from one protein chain would occupy one module (Figure S2E). Such an arrangement would require much longer DBD2-CD linker (~72 Å and ~66 Å for module 2 and module 3, respectively). To validate the arrangement of the domains in our structural model we designed a variant of TnsB in which residues 245-251 from the DBD2-CD linker were removed. Such a shortened linker could still be accommodated in the tiled domain arrangement, but it would be incompatible with an arrangement in which all domains of one TnsB chain reside in one module. The functionality of this TnsB variant was assessed in DNA binding studies by EMSA which showed that Δ245-251 variant was able to bind 3BS DNA similar to the wild-type protein forming three protein-DNA complexes (Figure S2F). Moreover, negative-stained EM images showed that the deletion variant formed filaments in the presence of 3BS DNA which were indistinguishable from those composed of the wild-type protein (Figure S2G). These results demonstrated that the shorter linker did not disrupt the architecture of the complex with the transposon end which confirmed the connection of domains in our structural model. Such arrangement of the domains also better explains the formation and stability of the tubular filament used for structure determination (compare Figures S2D and S2E).

Figures were prepared in UCSF ChimeraX (Pettersen et al., 2021) and Pymol (PyMOL Molecular Graphics System, version 2.0, Schrödinger). DNA geometry was analyzed with Curves+ (Blanchet et al., 2011). Refinement and validation statistics are presented in Table S2. Multiple sequence alignments of TnsB proteins were created with PROMALS3D (Pei et al., 2008). Sequences of CAST TsnB proteins were selected based on the literature (Klompe et al., 2019; Petassi et al., 2020; Saito et al., 2021; Strecker et al., 2019; Vo et al., 2021).

Electromobility shift assay

The substrate binding of TnsB variants was determined using 5'-Cy5-labeled dsDNAs. The substrate concentration was constant and equal to 50 nM. The protein concentration varied from 16 nM to 500 nM. Reactions were incubated at room temperature for 10 min in buffer that contained 20 mM HEPES [pH 7.5], 100 mM NaCl, and 2% (v/v) glycerol. The DNA-protein complexes were separated by 6% native polyacrylamide gel in 0.5 × TBE buffer at 4 °C and visualized by fluorescence readout.

For densitometry analysis, all bands on scanned gels were quantified with ImageJ software (Schneider et al., 2012) and assigned to one of the following categories: D0 (unbound DNA substrate), D1 (first shifted band), D2 (second shifted band), D3 (third shifted band), D4 (fourth shifted band), and Da (aggregate) indicated by arrows next to the scanned gels in Figure S5. Signal from each lane was normalized to the nominal DNA concentration (50 nM) and plotted against the TnsB concentration (dots on graphs, Figure S5).

For semi-quantitative analysis of TnsB-DNA binding we assigned the normalized signal to the concentrations of TnsB-DNA complexes containing one (D1), two (D2), three (D3) and four (D4) TnsB molecules and to the TnsB-DNA aggregates (Da). Quantified signal was fitted to a numerical model that assumed sequential binding events described with the following equations:

$$D1 = P0 \times D0 / K_D1$$

$$D2 = P0 \times D1 / K_D2$$

$$D3 = P0 \times D2 / K_D3$$

$$D4 = P0 \times D3 / K_D4$$

where K_D1-K_D4 are dissociation constants describing the formation of D1-D4 complexes and P0 is the concentration of unbound TnsB protein. Importantly, these K_D values do not refer to the binding to a particular site, because EMSA experiments carry information only on the concentration of complexes of a particular stoichiometry, but cannot discriminate between binding to different sites. To compensate for the effect of aggregation, we included the empirical model of TnsB-DNA aggregation:

$$Da = P0 \times r$$

where r is a constant describing the average DNA (Da) to protein (P0) ratio in an aggregate. The law of conservation of mass for the total DNA (Dtot) and protein (Ptot) leads to the following boundary conditions:

$$Dtot = D0 + D1 + D2 + D3 + D4 + Da$$

$$Ptot = P0 + 1 \times D1 + 2 \times D2 + 3 \times D3 + 4 \times D4 + r \times Da$$

Analytic polynomial equation for P0 was derived from the above system of equations and used for the calculation of predicted D1-D4 and Da values for a given set of model parameters (K_D1-K_D4 and r). Finally, normalized values from gel densitometry analysis were iteratively fitted to the above system of equations with the *nls*() function from the R language (https://www.R-project.org/). Fitting results were visualized with *ggplot2* package (Wickham, 2016, lines on graphs, Figure S5). Calculated model parameters along with the standard errors of the fit are plotted in Figure 6B.

Nanoluciferase (NanoLuc) assay

In Tn7 and Tn7-like elements, the TnsB binding site which is the most distant from the transposon end (site 4) overlaps with the predicted promoter in the right end of the element (Figures S3A–S3C) (Arciszewska et al., 1991). Therefore, TnsB binding to site 4 blocks this promoter allowing autoregulatory control of *tnsB* and the other transposition genes in the operon. According to the previously published data (Arciszewska et al., 1991), site 4 is the last one occupied by TnsB in the right end as the protein concentration increases, so the promoter blockage requires a formation of a complete end-recognition complex. We prepared a reporter plasmid in which an open reading frame coding for NanoLuc, a small and highly luminescent luciferase variant (England et al., 2016), was placed in-frame after the start codon of *tnsA* gene, the first gene in the operon. In this construct we also introduced a modified Shine-Dalgarno sequence to drive strong NanoLuc expression to ensure a highly sensitive assay (Figure S3D). We reasoned that for this construct TnsB binding to site 4 would block the promoter and reduce NanoLuc expression and that this effect would be proportional to the strength of TnsB-DNA interaction. Indeed, in the absence of exogenous TnsB, NanoLuc was efficiently expressed in *E. coli* as measured by its luminescence, but increasing the TnsB expression gradually reduced the luminescence signal (Figure S3E). In the studies of the TnsB variants, to facilitate the interpretation, we calculated the results as the relative reduction of NanoLuc luminescence signal upon expression of wild-type and mutant proteins. The effect observed for the wild-type protein was expressed as 100% reduction.

For experiments monitoring *in vivo* binding of TnsB *wild-type* to Tn7 right-end variants, RJW0163 was transformed with pRW99 or pRW171, and pRW182 or derivative (pRW207 – pRW211, Table S5) and plated onto LB agar supplemented with 8 μg/ml tetracycline, 10 μg/ml gentamicin, and 0.2% glucose. For experiments monitoring *in vivo* binding of TnsB mutants to Tn7 right-end *wild-type*, RJW0163 was transformed with pRW182 and pRW171, pRW99 or derivative (pRW100-pRW112, pRW215-pRW217) and plated onto LB agar supplemented with 8 μg/ml tetracycline, 10 μg/ml gentamicin, and 0.2% glucose. Overnight cultures were subcultured 1:100 in LB broth supplemented with 8 μg/ml tetracycline, 10 μg/ml gentamicin, and 0.2% rhamnose to induce expression of TnsB, and incubated at 37 °C with agitation for 2 hours. For luminescence measurements, 50 μl of cells were mixed with 50 μl of Nano-Glo Luciferase Assay System reagent according to the manufacturer's instructions (Promega, Madison, WI). Luminescence signals were measured over time using a BioTek Synergy H1 microplate reader and then normalized using OD_{600} measurements of experimental cultures. Experiments presented in Figures 3D, 5C, and 6A were performed simultaneously with a single control experiment for each replicate.

In Vivo Mate-out Transposition Assays

In vivo transposition frequencies of a miniTn7(Kan) transposable element (*wild-type* and right-end variants) with TnsB *wild-type* were evaluated using the mate-out assay (Figure 6C, right). RJW0163 with an F plasmid derivative was transformed with pRW97 and pRW221 or derivative (pRW222-232, Table S5) and plated onto LB agar supplemented with 100 μg/mL carbenicillin, 30 μg/mL chloramphenicol, and 0.2% glucose. After 16 hours incubation at 30 °C, transformants were collected from plates with LB broth and transferred to sterile microfuge tubes. Cells were subcultured 1:100 into LB broth supplemented with 100 μg/mL carbenicillin and 0.0002% arabinose to induce expression of TnsABCD and incubated at 37 °C with agitation. Following 2 hours incubation, a portion of induced cultures were washed once and resuspended with LB broth supplemented with 0.2% glucose. Washed cells were recovered for 2 hours at 37 °C with agitation and then mixed with a mid-log culture of CW51 recipient at a ratio of 1:5 donor:recipient. After incubation at 37 °C with gentle agitation for 90 minutes, cultures were vortexed, placed on ice, and serially diluted in LB supplemented with 0.2% glucose and plated on LB agar supplemented with 20 μg/mL nalidixic acid, 100 μg/mL rifampicin, 100 μg/mL spectinomycin, with or without 50 μg/mL kanamycin to select for transposition or sample the entire transconjugant population respectively. Plates were incubated at 37 °C for 36 hours before colonies were counted.

In vivo Lambda Hop Transposition Assays

In vivo transposition frequencies of TnsB *wild-type* and mutants were evaluated with the lambda hop assay where a conditional lambda KK1 phage vector is used to introduce a mini element with the left (166 bp) and right (199 bp) ends of the element flanking a kanamycin resistant determinant (miniTn7Kan^R) (McKown et al., 1988). Overnight cultures grown in LB media supplemented with 0.2% glycerol, 0.2% glucose, and appropriate antibiotic (100 μg/ml carbenicillin, 8 μg/ml tetracycline) were diluted 1:50 into LB supplemented with 0.2% glycerol, inducer (0.000002% arabinose) for expression of tnsABCD, inducer (0.00002%-0.2% rhamnose) for expression of additional tnsB where indicated, and appropriate antibiotic (100 μg/ml carbenicillin, 8 μg/ml tetracycline), and grown to mid-exponential phase at 37 °C. Cells were harvested, resuspended in 0.01 M $MgSO_4$ to a normalized OD_{600}, and infected with lambda KK1, a replication- and integration-defective lambda phage containing miniTn7Kan^R, at 37 °C static incubation. After 15 minutes, 0.01 M sodium citrate is added to inhibit new infections and cells were recovered for 60 minutes at 37 °C in LB media supplemented with 0.2% glycerol and 0.2% glucose. Cells were harvested and spread directly on LB agar medium supplemented with 0.2% glycerol, 0.2% glucose, 20 mM sodium citrate, and 100 μg/ml kanamycin. Plates were incubated at 37 °C for 18 hours. Transposition frequency was calculated by counting the number of kanamycin-resistant colonies per infectious phage particle. Transposition assays presented in Figures 3E and 5D were performed simultaneously with a single control experiment for each replicate.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical and numerical analysis were performed using R environment (version 4.1.2). EMSA experiments were performed in triplicates and quantified using semi-quantitative numerical model described above. Calculated K_D values presented in Figure 6B are plotted with the error bars corresponding to standard error of the fit. Nanoluc binding and *in vivo* transposition assays were performed independently three times and the results were reported with Mean ± SD. Asterisks below graphs indicate statistical significance of difference between wild-type and each mutant calculated using paired sample *t*-test.