Nucleic Acids Research, 2025, **53**, gkaf472
https://doi.org/10.1093/nar/gkaf472
Molecular and Structural Biology

Asymmetric loading of TnsE regulates Tn7 targeting of DNA replication structures

Shreya S. Krishnan[1,2], Yao Shen[1,2,3], Treasa B. O'Hagan[1,2,4], Lindsay A. Matthews[1,2], Nuwani W. Weerasinghe[2,5], Rodolfo Ghirlando[6], Christopher J. Thibodeaux[2,5], Alba Guarné[1,2,*]

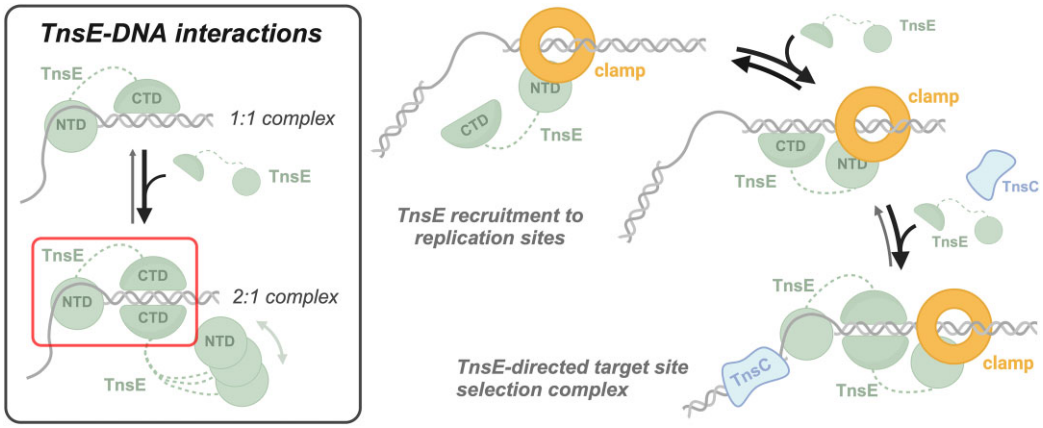
[1]Department of Biochemistry, McGill University, Montreal, QC H3G 0B1, Canada
[2]Centre de Recherche en Biologie Structurale, McGill University, Montreal, QC H3G 0B1, Canada
[3]Present address: Division of Structural Biology, Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom
[4]Present address: Kavli Institute for Nanoscience Discovery, Dorothy Crowfoot Hodgkin Building, University of Oxford, Oxford OX1 3QU, United Kingdom
[5]Department of Chemistry, McGill University, Montreal, QC H3A 0B8, Canada
[6]Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, United States

[*]To whom correspondence should be addressed. Email: alba.guarne@mcgill.ca

Abstract

Tn7 transposable elements are known for their sophisticated target-site selection mechanisms. For the prototypical Tn7 element, dedicated transposon-encoded proteins direct insertions to either a conserved site in the chromosome or replicating DNA structures in conjugal plasmids, ensuring the vertical and horizontal spread of the element. While the pathway targeting the *attTn7* site in the bacterial chromosome has been extensively studied, the pathway targeting DNA replication structures remains poorly understood. We have used an integrative structural biology approach to elucidate how the Tn7-encoded protein TnsE recognizes replication sites. Using native mass spectrometry, we found that TnsE forms 1:1 and 2:1 (TnsE:DNA) complexes on 3′-recessed DNA, with gain-of-function TnsE variants favoring the formation of 2:1 complexes. Structural characterization confirms that two TnsE molecules bind to DNA with the C-terminal domain of the protein recognizing duplex DNA, leaving the N-terminal domain to impose DNA substrate specificity and recruit the core transposition machinery. Collectively, our work is consistent with a model where TnsE-mediated target-site selection relies on the formation of an asymmetric TnsE:DNA complex to recruit the Tn7 transposase to DNA replication structures.

Graphical abstract



Introduction

DNA transposons are mobile elements found in virtually all organisms. While most transposons move with little specificity, some have developed sophisticated targeting pathways to minimize the impact on the host and ensure their spread. Bacterial Tn7 transposons are a prime example for mobile elements with elaborate target-site selection pathways [1]. The canonical Tn7 element encodes five proteins, and it is delimited by structurally asymmetric inverted repeats (Fig. 1A). TnsA and TnsB form a heteromeric transposase that

Received: August 21, 2024. **Revised:** April 25, 2025. **Editorial Decision:** April 30, 2025. **Accepted:** May 20, 2025

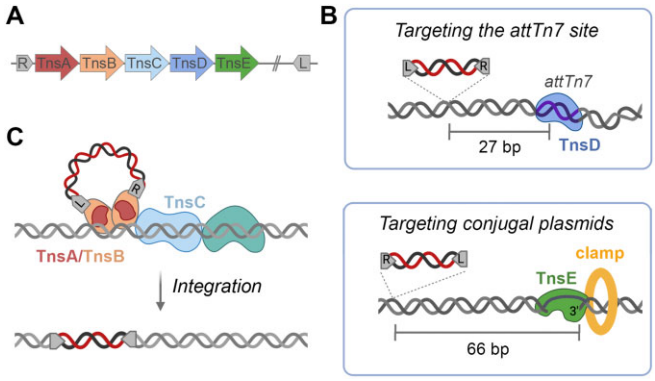


Figure 1. The Tn7 element. (**A**) Organization of the Tn7 element. (**B**) Recruitment of the transposase machinery (TnsABC) to target sites and integration. (**C**) Target-site selection pathways of the prototypical Tn7 element. TnsD targets a conserved site in the bacterial chromosome (*attTn7*) and TnsE targets replication structures in conjugal plasmids. These two target-site selection mechanisms ensure the vertical (top panel) and horizontal (bottom panel) spread of the element.

recognizes inverted repeats at the ends of the element, excise it from its original location, and integrate it into a pre-identified target site [2, 3]. The TnsAB transposase on its own is not active but its association with the TnsC adaptor defines the core transposition machinery (Fig. 1B). TnsC is a AAA+ ATPase that coordinates recruitment of the transposase to target sites. Moreover, TnsC imposes the strict spacing between the target and insertion sites [4]. TnsABC, however, remains inactive until recruited to a specific DNA site by a target-site selector protein (Fig. 1C). This targeting results in insertions at either a specific chromosomal site, *attTn7*, when recruited by TnsD or replicating conjugal plasmids when recruited by TnsE [5].

All elements in the Tn7 family encode at least one target selector protein from the TnsD/TniQ family [6, 7]. These proteins always include a conserved TniQ domain that is either fused to additional domains for DNA-binding specificity or naturally combined with CRISPR–Cas systems for RNA-guided transposition [8, 9]. In the prototypical Tn7 element, TnsD is a sequence-specific DNA-binding protein that recognizes a highly conserved sequence at the end of the *glmS* gene and imposes an asymmetrical DNA distortion to recruit and activate the TnsABC machinery to its cognate *attTn7* site [10, 11]. Since the Tn7 transposon ends are not identical, Tn7 insertions occur with a specific orientation [1]. TnsD/TniQ-driven Tn7 insertions occur with a left-to-right orientation (Fig. 1C). *In vitro* transposition experiments using a DNA distortion and a gain-of-function variant of TnsC demonstrated that the TnsD/TniQ target-site selector is not required to determine the orientation bias of the insertions [11, 12]. Instead, TnsD asymmetrically distorts DNA and functions as a nucleotide-exchange factor for TnsC, in turn facilitating DNA-dependent oligomerization of TnsC at target sites [13].

In contrast to TnsD/TniQ, TnsE is a sequence-independent but structure-specific DNA-binding protein that recognizes actively replicating DNA sites [14]. TnsE-driven insertions occur with an opposite orientation to TnsD-directed insertions indicating that the recruitment of the TnsABC machinery must be different when targeting DNA replication structures (Fig. 1C). TnsE preferentially targets conjugal plasmids ensuring the horizontal spread of the Tn7 element [15], but TnsE-directed Tn7 insertions can also be observed within the bacterial chromosome at low frequency [14, 16]. In this case, they preferentially target termination sites or sites where replication forks stall [17]. Tn7 transposition into replicating DNA structures is facilitated by the interaction with the β-sliding clamp [18], highlighting the importance of the DNA structure and the interplay with the replication machinery to form TnsE-directed Tn7 targeting complexes.

TnsD-mediated insertions occur at about a 1000-fold higher frequency than those mediated by TnsE [5]. However, TnsE gain-of-activity mutants allowed for transposition levels comparable to those of TnsD [14]. These TnsE variants showed the same target site preference as wild-type TnsE but enhanced DNA binding affinity. Structural characterization of the C-terminal domain of TnsE, which harbors virtually all the gain-of-activity mutations, unveiled a unique switch where conformational flexibility would target DNA engagement until an appropriate insertion site is identified [19]. However, the lack of structural information in the presence of DNA has prevented the understanding of how TnsE recognizes specific DNA structures and recruits the transposition machinery to target sites. Here, we present an integrative structural biology approach to characterize how TnsE recognizes 3′-recessed end DNA substrates. We find that TnsE forms an antiparallel, asymmetric complex upon binding to 3′-recessed end DNA substrates in which 1:1 and 2:1 (TnsE:DNA) complexes coexist. Gain-of-activity variants of TnsE stabilize the 2:1 form of the complex, allowing us to propose a model for how TnsE may coordinate the interactions with the β-clamp and the TnsC adaptor to target replicating DNA structures.

Materials and methods

Cloning of TnsE variants

Thiopseudomonas alkaliphila TnsE (*Ta*TnsE) (NCBI ID: WP_053111219.1) was purchased from Bio Basic Inc. (Ontario, Canada). The gene was subcloned into the pET21b expression vector encoding a non-removable, C-terminal 6x His-tag using the NdeI and XhoI restriction sites (pAG9187). The C-terminal domain of *Ta*TnsE (*Ta*TnsE-CTD, residues Asp377-Glu534) was generated by site-directed mutagenesis using the Q5® kit (New England Biolabs) (pAG9214). Plasmids with genes coding for *Ec*TnsE-A453V/D523N and wild-type *Escherichia coli* TnsE (*Ec*TnsE) were obtained from Dr Joseph Peters lab (Cornell University) and subcloned into pET21b using NdeI and XhoI sites to generate pAG8723 and pAG8721, respectively. *Ec*TnsE-D523N/E516K (pAG9154), *Ec*TnsE-DR2 [Δ(Arg332-Thr379), pAG9309], *Ec*TnsE-DR1G4-DR2 (where residues Phe270–Asp307 are replaced by four glycine residues and harbor the DR2 deletion, pAG9483), *Ec*TnsE-DR1G6-DR2 (residues Phe270–Asp307 replaced by six glycine residues and DR2 deletion, pAG9484), and *Ec*TnsE-CTD (encompassing residues Thr379–His538, pAG9240) were generated by site-directed mutagenesis using the Q5® kit. All constructs were verified by DNA sequencing (Genome Quebec).

Protein expression and purification

Plasmids were transformed in BL21 Star™ (DE3) pLysS cells (Life Technologies) supplemented with a plasmid encoding for rare transfer RNAs equivalent to that from the Rosetta2 (DE3) cell line (Novagen) and incubated at 37 °C with orbital agitation up to OD_{600} ∼ 0.7. Protein expression

was then induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After induction, cell cultures were grown for 5 h at 25°C (full-length TnsE variants and *Ec*TnsE-DR2), overnight at 16 °C (*Ec*TnsE-DR1G4-DR2 and *Ec*TnsE-DR1G6-DR2), or 3 h at 37 °C (C-terminal domain TnsE variants). Cell pellets were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris–HCl, pH 8, 0.5 M NaCl, 30 mM imidazole, 1.4 mM 2-mercaptoethanol, and 5% glycerol) supplemented with protease inhibitors (2 mM benzamidine, 2 mM phenylmethylsulfonyl fluoride (PMSF), 250 μg leupeptin, and 35 μg pepstatin A). Cells were lysed by sonication and spun at 39 000 × *g* for 40 min. The clarified lysate was loaded onto a Ni-chelating Hi-Trap column (GE HealthCare) pre-equilibrated with lysis buffer. Impurities were eliminated by washing with 45 mM imidazole, and protein was eluted with 180 mM imidazole. To eliminate genomic DNA, a Q-Sepharose Hi-Trap column (GE HealthCare) was attached to the Ni-chelating Hi-Trap column during the elution step. For full-length TnsE and linker deletion variants, the fractions containing protein were pooled and diluted in IEX buffer A [20 mM Tris, pH 8, 1 mM ethylenediaminetetraacetic acid (EDTA), 1.4 mM 2-mercaptoethanol, and 5% glycerol] to decrease the NaCl concentration to 125 mM and the protein was further purified over a Mono Q 10/100 GL anion exchange column (GE HealthCare) using a linear gradient to 0.5 M NaCl. For *Ta*TnsE^{CTD} and *Ec*TnsE^{CTD}, the fractions from Ni-affinity chromatography were pooled and diluted to 125 mM NaCl and purified by Capto HiRes S 5/50 GL cation exchange column (GE Life Sciences). Pure proteins were concentrated and stored in storage buffer (20 mM Tris–HCl, pH 8, 0.15 mM NaCl, 0.1 mM EDTA, 1.4 mM 2-mercaptoethanol, and 5% glycerol). All purification and concentration steps were done at 4 °C.

Table 1. X-ray data collection and processing statistics

	*Ta*TnsE^{CTD}–DNA
Data collection	
Beamline	08B1; Canadian Light Source (CLS)
Detector	Rayonix MX300HE
Space group	$P6_122$
Cell dimensions	
a, *b*, *c* (Å)	63.77, 63.77, 220.51
α, β, γ (°)	90, 90, 120
Wavelength (Å)	1.0332
Resolution (Å)	49.38–2.8 (2.9–2.8)
Total reflections	275 635 (18 891)
Unique reflections	7169 (702)
Completeness (%)	99.6 (98.9)
Redundancy	38.4 (26.9)
I/σ_I	9.03 (0.56)
R_{meas}	0.67 (6.65)
R_{pim}	0.11 (1.27)
$CC_{1/2}$	0.995 (0.42)
Refinement	
Resolution (Å)	49.38–2.8 (2.9–2.8)
No. of reflections	7143 (695)
R_{work}/R_{free}	0.2368/0.2779 (0.4235/0.4688)
No. of atoms	
Macromolecules	1523
Ligands	5
Solvent	19
Protein residues	157
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.47
Ramachandran plot	
Favored (%)	94.19
Allowed (%)	5.81
Disallowed (%)	0

Protein crystallization and structure determination

To assemble the TnsE:DNA complex, *Ta*TnsE (4 mg/ml) was incubated overnight at 4 °C with a 3′-recessed DNA substrate (24ss + 24ds) prepared by annealing the SK037 (48-mer) and SK013 (24-mer) oligonucleotides (Supplementary Table S1). Diffraction quality crystals grew at room temperature in 0.2 M NaI, 0.1 M Bis-Tris–HCl propane (pH 6.5), and 20% PEG 3350 using a 1:1 (protein:DNA) ratio. A complete dataset was collected on the CMCF-BM (08B1) beamline at the Canadian Light Source (Saskatoon, Canada). The data were processed and merged with XDS [20, 21]. The structure was determined by molecular replacement using *Ec*TnsE^{AVDN} C-terminal domain (PDB ID: 5D16) as the search model. The model was improved through iterative cycles of manual model building in COOT and refinement in Phenix (Table 1) [22, 23].

DNA binding assays

The DNA-binding specificities of *Ta*TnsE, *Ec*TnsE, and their C-terminal domains were determined by electrophoretic mobility shift assays (EMSAs). We used fluorescein-labeled single-strand [FAM-YS002 (15-mer)], duplex [FAM-YS002 (15-mer) + YS003 (15-mer)], and 3′-recessed end [YS005 (30-mer) + FAM-YS006 (15-mer)] DNA substrates (Supplementary Table S1). DNA duplexes were prepared by resuspending complementary oligonucleotides in Milli-Q ultrapure water mixed in a 1:1 ratio, boiled at 95 °C for 5 min and cooled slowly to room temperature. Reactions (10 μl) were prepared by mixing equal volumes of DNA (final concentration 10 nM) with increasing concentrations of protein (10–1280 nM) in EMSA buffer (25 mM HEPES, pH 7.6, 3 mM Tris–HCl, pH 7.6, 50 μg/ml bovine serum albumin, 2 mM ATP, 0.7 mM 2-mercaptoethanol, 0.02 mM EDTA, 4 mM KCl, and 6% glycerol) [14]. The reactions were incubated for 30 min at room temperature and resolved in 5% native TBE polyacrylamide gels run at 80 V in an ice bath with 0.5× TBE as the running buffer. Gels were pre-conditioned by running them for 30 min in the same running conditions (80 V in an ice bath). All gels were run in triplicate, imaged using Azure Biosystems Sapphire RGBNIR Gel Imaging System and quantified with Fiji [24].

To assess DNA binding cooperativity, a 3′-recessed end 30ss + 30ds DNA substrate [FAM-YS004 (60-mer) + YS005 (30-mer), 10 nM] was incubated with increasing concentrations of *Ec*TnsE (2.5, 7.5, 10, 15, 20, 25, 30, 35, 40, and 50 nM) and EMSAs were quantified to determine the amounts of free DNA, 1:1 and 2:1 TnsE:DNA complexes. The Hill coefficient was determined by plotting the fraction of DNA bound [$Y = ([1{:}1] \times 0.5) + ([2{:}1] \times 1)$] against protein concentration and fitting the Hill equation ($Y = B_{max} \times X^h/(K_d^h + X^h)$, where B_{max} is the maximum specific binding in the same units as Y and h is the Hill slope) to the results in PRISM.

Small-angle X-ray scattering experiments

The oligomeric states of *Ec*TnsE-A453V/D523N and *Ec*TnsE-A453V/D523N bound to DNA were analyzed by size-exclusion chromatography coupled to small-angle

X-ray scattering (SEC–SAXS). Scattering data were collected at the 18-ID beamline of the Advanced Photon Source (APS). Protein and protein–DNA samples were resolved over a Superdex 200 HR 10/30 column (Cytiva) pre-equilibrated with storage buffer. Data processing was performed using the ATSAS 3.0.1 program suite [25]. Specifically, scattering frames corresponding to buffer and sample were selected using CHROMIXS, and linear Guinier region was estimated using PRIMUS [25, 26]. GNOM was used to compute the radius of gyration, R_g; pairwise distribution function, $P(r)$; and the maximum particle dimension, D_{max} [27]. Ten independent *ab initio* models were generated using DAMMIF, averaged, and optimized using DAMMIN [28, 29].

Scattering data for the 3′-recessed DNA samples [15ss + 15ds (YS005 + YS006) and 30ss + 30ds (YS004 + YS005)] were collected using a Rigaku BioSAXS-1000 instrument. Scans were conducted over 3 h at concentrations of 50, 100, and 150 μM. Data quality was assessed by comparing scattering curves across different concentrations. Subsequent processing used the ATSAS 3.0.1 software suite as described above.

Preparation of gold-coated borosilicate nanospray emitters

The nanospray emitters were prepared in-house using borosilicate glass capillaries [1.0 mm (outside diameter) × 0.78 mm (inside diameter) × 100 mm (length); Harvard Apparatus GC100T-10] that were pulled using Sutter Instrument model P-2000 tip puller with the following instrumental parameters: heat = 350, filament = 4, velocity = 60, delay = 255, and pull = 0. The tip diameters are ∼1–2 μm estimated using light microscopy. The tips were then coated with 40 nm of gold using Quorum Q150T S plus.

Native mass spectrometry

*Ec*TnsE and gain-of-activity variants at 3 mg/ml were incubated with 3′-recessed end DNA substrates generated by annealing either 30-mer and 15-mer (YS005 and YS006; Supplementary Table S1) or 60-mer and 30-mer (TOH02 and YS005; Supplementary Table S1) at 2:1 ratio (protein:DNA) for 30 min at room temperature in storage buffer. The TnsE:DNA complexes were buffer exchanged into 200 mM ammonium acetate (pH 8.5; pH adjusted with ammonium hydroxide) by applying the 50 μl of the sample sequentially through two pre-equilibrated Micro Bio-spin 6 columns (Bio-Rad). Samples (10 μl) were applied to the nanospray emitter using Eppendorf 20 μl gel loader tips. The emitters were then fastened into a tip holder (Thermo Scientific, part ES286), centrifuged for 2–3 s to push the liquid to the end of the tips, and mounted onto a Synapt G2-Si ion mobility mass spectrometer (Waters) equipped with a nanospray ESI source. All data were acquired in positive ion and sensitivity modes with an m/z range of 100–8000 and a scan rate of 1 scan/s. The time-of-flight (TOF) detector was calibrated over the same m/z range using sodium iodide calibrant. Data were collected in TOF mode with the following settings: capillary voltage = 1.5 kV, cone voltage = 10 V, source offset = 10 V, source temperature = 35 °C, and trap gas (argon) flow = 10 ml/min. To quantify the DNA-binding stoichiometry differences between TnsE variants, we calculated the percentage of *Ec*TnsE (or the *Ec*TnsE^{AVDN} and *Ec*TnsE^{DNEK} variants) bound to DNA at 1:1 and 2:1 ratio by integrating the area of the individual peaks that could unequivocally be assigned to a single species. Namely, we used the 14+ to 16+ ions (1:1) and the 21+ to 26+ ions (2:1) for the 15ss + 15ds DNA substrate, and the 15+ to 19+ ions (1:1) and the 20+ to 26+ ions (2:1) for the 30ss + 30ds DNA substrate. The percentage of each species was then calculated by the sum of the integrated peak area of the individual species over the total DNA-bound population.

Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity experiments on *Ec*TnsE (∼5 μM), 15ss + 15ds DNA (YS005 + YS006, ∼5 μM), and their mixtures (∼5 μM DNA + 1, 2, 4 equivalents of *Ec*TnsE) were carried out at 50 000 rpm (201 600 × g at 7.20 cm) and 20 °C on a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge and An50-Ti rotor following standard protocols [30]. Stock solutions in either storage or mass spectrometry buffer were diluted in the same buffer to prepare samples for analysis. Samples were loaded in 12- or 3-mm two-channel centerpiece cells, and sedimentation data were collected using the absorbance (260 nm) and interference optical detection systems. Data were analyzed in SEDFIT in terms of a continuous $c(s)$ distribution of sedimenting species [31]. The solution densities and viscosities were measured experimentally on an Anton Paar DMA 5000 density meter and Anton Paar AMVn automated micro-viscometer at 20 °C. The protein's partial specific volume was calculated in SEDNTERP [32], and a partial specific volume of 0.55 cm^3/g was used for DNA. Additivity rules were used to determine the partial specific volumes of the protein–nucleic acid complexes. As the species have different partial specific volumes, experimental sedimentation coefficients are presented.

Cryo-electron microscopy

A solution containing 100 μM (6.2 mg/ml) *Ec*TnsE^{AVDN} was incubated with 50 μM 30ss + 30ds 3′-recessed DNA substrate (TOH02 + YS005, Supplementary Table S1) overnight at 4 °C in storage buffer. The TnsE:DNA complex was resolved over a S200 Increase 10/300 GL column (GE HealthCare) pre-equilibrated in cryo-electron microscopy (cryo-EM) buffer (20 mM Tris–HCl, pH 8.0, 0.15 M NaCl, 1.4 mM 2-mercaptoethanol, and 0.1 mM EDTA). Peak fractions were pooled and concentrated to ∼1.2 mg/ml (20 μM) right before being deposited onto C-flat CF-2/2–3Cu holey carbon grids. Grids were pre-treated with chloroform and glow discharged in air at 15 mA for 30 s. The TnsE:DNA complex (3.5 μl) was directly applied onto the grids, and vitrified in liquid ethane using Vitrobot Mark IV (Thermo Fisher Scientific) blotting one time for 5 s with a blot force of +1. The Vitrobot chamber was set to 25 °C and 100% relative humidity. Datasets were collected using SerialEM software [33], in a Titan Krios microscope operated at 300 kV and equipped with a Gatan BioQuantum LS K3 electron detector. Images were collected in counting mode according to the parameters described in Supplementary Table S2.

Cryo-electron microscopy image processing

All cryo-EM data processing steps were performed using CryoSPARC v4.1.1 [34]. Movies were corrected for beam-induced motion using patch motion correction with default settings. All frames in the movies were used to produce merged micrographs. The contrast transfer function (CTF) parameters

were estimated using the Patch CTF estimation using default settings. The minimum and maximum resolution considered to estimate the CTF parameters were 25 and 4 Å, respectively, and the minimum and maximum defocus values were set up at 1000 and 50 000 Å. Images with an estimated resolution of 6.5 Å or better were kept for further processing.

Particle picking and selection were done in three steps: (i) manual and blob picker followed by 2D class averaging; (ii) template picking followed by 2D class averaging; and (iii) Topaz [34, 35] training followed by 2D class averaging. On the first step, we used six micrographs to manually select ~800 particles that were input into blob tuner, which optimizes blob sizes, shapes, and thresholds by comparing picks to the selected examples. Blob tuner found a single set of parameters for an ellipse shape with minimum and maximum diameters of 64 and 143 Å, respectively. These parameters were then used to select particles in 500 micrographs. The selected particles were curated through 2D classification, requesting 100 classes and using default settings. The best 2D classes were selected for the second picking step using Template Picking in 5978 micrographs with default parameters and a particle diameter of 150 Å. The selected particles were curated through another round of 2D classification, requesting 200 classes and using default settings. In the third step, we used selected particles from the 2D classification and 280 micrographs for Topaz training. We set up two training jobs differing in the expected number of particles that was set up as 400 and 450 particles, respectively. The non-default setup parameters were the same in both training jobs and included a downsampling factor of 8 and 30 epochs. These trained models were used to pick particles on the subset of 280 micrographs, setting the radius of extracted regions to 14. The picked particles were extracted with an extraction box size of 400 pixels and subsequently Fourier-cropped to 100 pixels. Extracted particles were curated using 2D classification, requesting 50 classes and increasing the number of online-EM iterations to 30. The rest of the parameters were used in the default settings. The curated particles were used in another round of Topaz training, 2D classification, and particle curation. This process of reselecting particles for Topaz training was repeated two more times prior to using the final trained models for picking particles on the complete dataset. Particles picked from different training models were extracted, and duplicates with a minimum separation distance of 70 Å were removed using the "remove duplicate particles" job prior to 2D classification. A total of 2.5 million particles were then subject to 2D classification requesting 200 class averages, and the number of online-EM iterations and batch size per class were modified to 40 and 300, respectively. We did one round of reference-free 2D classification to remove incorrectly picked or damaged particles. The curated set of particles (2 292 812 particles) binned by 4 was used to generate 3D *ab initio* models.

For all *ab initio* reconstruction routines, we modified the initial minibatch size and final minibatch size to 300 and 1000, respectively. Five classes were generated in the first round of 3D *ab initio* reconstruction (Supplementary Fig. S1). Classes 1 and 2 (1 113 501 particles) had the most defined features with the least amount of noise. These particles were subjected to another round of 3D *ab initio* reconstruction generating five maps that were used as initial maps in a subsequent heterogeneous refinement with default settings to better classify all remaining particles. From this refinement step, classes were inspected visually to identify features resembling the crystal structure TnsE^{CTD}:DNA, with the overall dimensions predicted from our SAXS measurements of the complex. Classes 2.2, 2.3, and 2.4 were either too small (class 2.3) or did not have enough density (classes 2.2 and 2.4) to represent the 2:1 TnsE:DNA complex and were not considered further. Classes 2.0 and 2.1, in which we could clearly identify protein densities for at least two TnsE^{CTD} and DNA features, were subjected to two more rounds of 3D *ab initio* reconstruction and one round of heterogeneous refinement to remove damaged particles. The curated particles for each class were re-extracted with a box size of 384 pixels and used as full-size images for nonuniform refinement with the map for class 5.1 as the initial reference (Supplementary Figs S1 and S2).

Fitting the TaTnsE^{CTD} structure onto the cryo-EM maps

The "antiparallel" configuration of the TaTnsE^{CTD} bound to DNA (2:1 stoichiometry) was generated from the crystal structure. The "parallel" configuration was modeled from the "antiparallel" configuration by rotating one of the protomers by 180° around itself and translating it along the DNA duplex until the protein:DNA interactions seen in the crystal structure were recapitulated. This placed the N-termini of both protomers on the same side of the complex. The "antiparallel" and "parallel" configurations were then fitted into the cryo-EM map using the fit in map tool in ChimeraX [35]. We fitted each configuration in two orientations resulting from fitting protomer 1 either on the top-half of the density or rotating the complex by 180° around the DNA axis, thereby positioning protomer 1 on the bottom-half density (Supplementary Fig. S3). The best-scoring conformation for each configuration was subjected to a cycle of rigid-body refinement in Phenix treating the entire complex as a single body to preserve the protein:DNA interactions seen in the crystal structure (Supplementary Fig. S3).

AlphaFold predictions

The predicted model for *Ec*TnsE (AF-P05845-F1) was downloaded from the AlphaFold Protein Structure Database [36]. The models for *Ec*TnsE bound DNA were predicted with AlphaFold3 using full-length *Ec*TnsE and either 15ss + 15ds or 30ss + 30ds DNA substrates [37]. Comparisons of the AlphFold predictions were done using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC). All figures depicting 3D structures or density maps were also generated with PyMOL.

Results

The C-terminal domain of *Ta*TnsE^{CTD} binds double-stranded DNA

TnsE binds specifically to 3′-recessed DNA structures and promotes Tn7 transposition into sites of active DNA replication [14]. To characterize how TnsE specifically recognizes 3′-recessed DNA substrates, we screened crystallization conditions for several bacterial TnsE proteins bound to various 3′-recessed DNA substrates. We obtained crystals of *Ta*TnsE^{CTD} bound to a 3′-recessed DNA substrate formed by annealing a 48-nt to a 24-nt oligonucleotide (Supplementary Table S1), which we herein refer to as 24ss + 24ds substrate. The crystals diffracted X-rays to 2.8 Å resolution, but the processing of the diffraction data revealed that the unit cell was

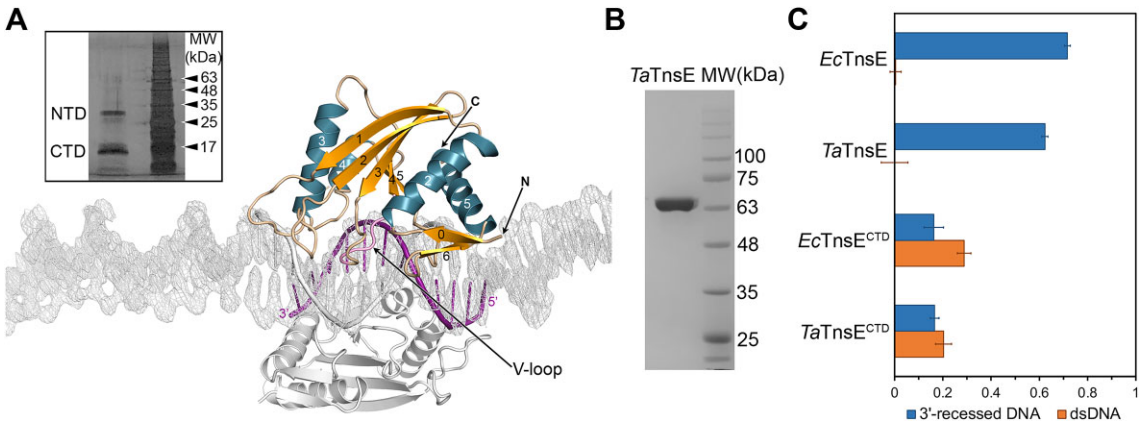


Figure 2. The C-terminal domain of TnsE binds double-stranded DNA (dsDNA). (**A**) Crystal structure of *Ta*TnsE^{CTD} bound to DNA. The asymmetric unit contains one *Ta*TnsE^{CTD} monomer and one DNA strand (colored), with the duplex forming through crystallographic symmetry (gray). The contents of a related asymmetric unit are shown in gray. The 2Fo–Fc (1σ) map for the pseudo-continuous DNA duplex is shown as a gray mesh. The inset shows a silver-stained sodium dodecyl sulfate (SDS)–polyacrylamide gel of the *Ta*TnsE:DNA crystallization mixture showing that the protein is broken down into its N- and C-terminal domains during the crystallization process. Molecular weight (MW) markers are indicated in kDa. (**B**) Coomassie-blue stained SDS–polyacrylamide gel of purified *Ta*TnsE. MW markers are indicated in kDa. (**C**) DNA binding and specificity of *Ec*TnsE, *Ta*TnsE, and their C-terminal domains to a 15ss + 15ds 3′-recessed DNA substrate or a blunt-ended 15bp DNA duplex (dsDNA). DNA (10 nM) was incubated with protein (40 nM). The *x*-axis corresponds to fraction of DNA bound. Errors bars represent standard error of the mean ($n = 3$).

too small to accommodate the full-length protein bound to DNA. Using the C-terminal domain of *Ec*TnsE as the search model for molecular replacement (PDB: 5D17), we found that the asymmetric unit contained one C-terminal domain of *Ta*TnsE (*Ta*TnsE^{CTD}, residues Tyr381–His537) and one DNA strand that defined a DNA duplex by crystallographic symmetry. This arrangement resulted in a complex with 2:1 (TnsE^{CTD}:DNA) stoichiometry (Fig. 2A).

Although the purified *Ta*TnsE sample used for crystallization was devoid of degradation products (Fig. 2B), there was no visible electron density to accommodate the N-terminal region of *Ta*TnsE suggesting that the protein was proteolytically cleaved during crystallization. We dissolved the remaining crystals in the crystallization solution and analyzed the mixture by SDS–polyacrylamide gel electrophoresis and found that the crystallization mixture contained two species (Fig. 2A, inset). The smallest product migrated at a molecular weight consistent with the *Ta*TnsE^{CTD} (~19 kDa), and the largest was a fragment of the N-terminal region (29–36 kDa). However, the full-length protein (62 kDa) was not present (Fig. 2B). The presence of species corresponding to the N-terminal in the crystallization solution suggested that *Ta*TnsE had been proteolytically cleaved during the crystallization process.

The asymmetric unit contains one molecule of *Ta*TnsE^{CTD} and a DNA strand (12 nt) that forms a DNA duplex by crystal symmetry, resulting in two molecules of *Ta*TnsE^{CTD} bound to a DNA duplex (Fig. 2A). This duplex propagates through crystallographic symmetry forming a pseudo-continuous DNA duplex. As for other sequence-independent protein:DNA complexes [38, 39], this arrangement resulted in the asymmetric unit representing an average of the protein binding along the entire duplex (Supplementary Fig. S4A and B), thereby precluding the assignment of the DNA register and washing out the signal for the 5′ single-strand DNA tail. While there is no obvious electron density for the 5′ single-strand tail, we found an isolated density near the duplex consistent with the size and geometry of a phosphate ion but with a partial coordination shell (Supplementary Fig. S4C). Since neither the sample nor the crystallization solution contained phosphate, it is plausible that this density corresponds to a phosphate moiety from the single-strand DNA tail of the substrate stabilized through electrostatic interactions with the protein. However, we could not model it unambiguously or exclude the possibility that the tail was degraded during the crystallization process.

The crystal structure of *Ta*TnsE^{CTD} bound to DNA implies that the C-terminal domain of the protein suffices to bind dsDNA. As expected from the high conservation between the two proteins (82% sequence identity), EMSAs with the full-length protein showed that both *Ec*TnsE and *Ta*TnsE preferentially bind 3′-recessed DNA substrates (Fig. 2C and Supplementary Fig. S5). The full-length proteins also bind blunt-ended dsDNA when present in excess (Supplementary Fig. S5). In contrast, the isolated C-terminal domains of *Ec*TnsE and *Ta*TnsE bind DNA but do not discriminate between 3′-recessed and blunt-ended DNA substrates (Fig. 2C and Supplementary Fig. S5). These results, together with the crystal structure, indicate that the C-terminal domain of TnsE binds dsDNA, leaving the N-terminal domain to impose the DNA substrate specificity.

DNA binding stabilizes the V-loop in a single conformation

The structure of *Ec*TnsE^{CTD} revealed that the V-shaped loop connecting β3 and β4 (Val451–Ser461) adopted two different conformations (Fig. 3A) [19]. In contrast, this loop adopted a single conformation in the structure of the C-terminal domain of the gain-of-function variant *Ec*TnsE-A453V/D523N (*Ec*TnsE$^{CTD–AVDN}$) (Fig. 3B), suggesting that it may act as a DNA-binding switch. Indeed, the V-loop in the structure of *Ta*TnsE^{CTD} bound to DNA adopts a similar "on" conformation to that seen in the structure of the *Ec*TnsE$^{CTD–AVDN}$ (Fig. 3B and Supplementary Fig. S6A–E) [19]. Modeling the alternative conformation seen in the structure of *Ec*TnsE^{CTD} resulted in major clashes with the phosphate backbone (Supplementary Fig. S6B), thereby confirming the idea that toggling between the two conformations regulates DNA binding [19].

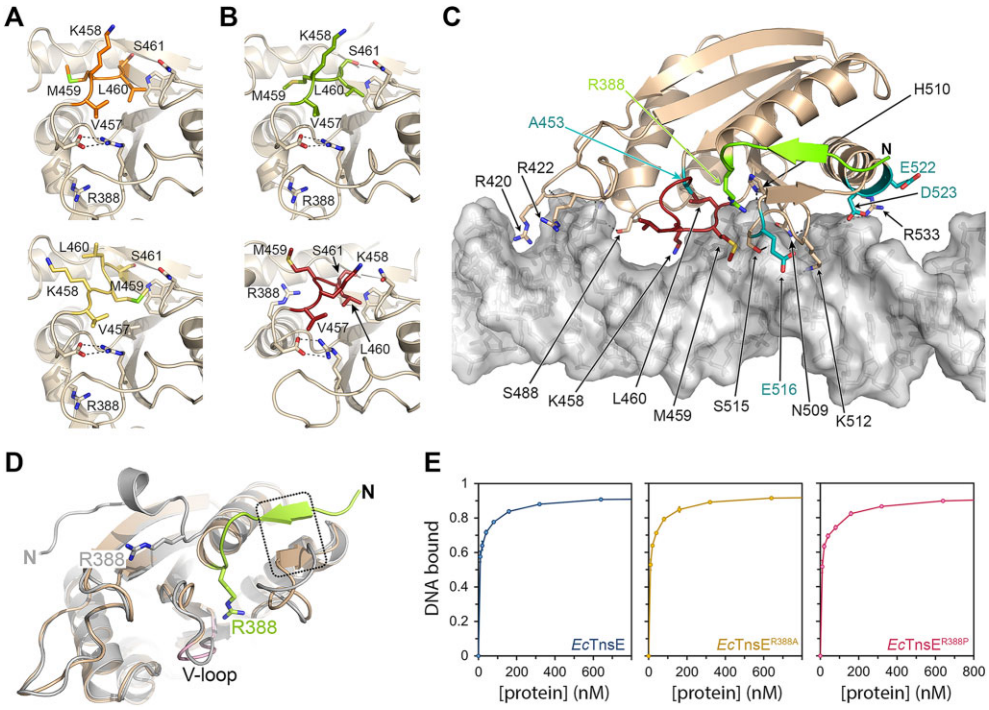


Figure 3. DNA-binding stabilizes a single conformation of the V-loop. (**A**) Alternate conformations of the V-loop (orange and yellow) on the structure of wild-type *Ec*TnsE^{CTD} (PDB ID: 5D17). (**B**) Conformations of the V-loop in the structures of *Ec*TnsE^{CTD}-A453VD523N (PDB ID: 5D16, top panel) and *Ta*TnsE^{CTD} bound to DNA (bottom panel). (**C**) Interactions between *Ta*TnsE^{CTD} and DNA. Residues involved in interactions are shown as sticks colored in dark red (V-loop), teal (residues that lead to gain of activity when mutated), or bright green (Arg388). The DNA is shown as a semi-transparent surface. Hydrogen bonds between protein and DNA atoms are shown as black dashed lines. (**D**) Superimposition of *Ec*TnsE^{CTD}-A453V/D523N (gray) onto *Ta*TnsE^{CTD} (sand with N-termini in green and V-loop in pink) depicting the reorientation of the N-terminal end of the domain (residues Y381–R388) around Arg388. (**E**) DNA binding of *Ec*TnsE (left), *Ec*TnsE-R388A (centre), and *Ec*TnsE-R388P (right) to a 15ss + 15ds 3′-recessed end DNA substrate (10 nM). Error bars represent standard error of the mean ($n = 3$).

Consistent with previously published data showing that the interaction between TnsE is sequence-independent [14, 19], the structure does not show any sequence specific contacts between *Ta*TnsE^{CTD} and DNA. The C-terminal arm of the V-loop straddles the phosphate backbone with Lys458 protruding into the major groove of the DNA (Fig. 3C and Supplementary Fig. S6A–C). The phosphate backbone is cradled by the side chain of Met459 and the carbonyl group of Leu460 on the major groove, and the side chains of Asn509, His510, and Lys512 at the end of the β5 strand on the minor groove (Fig. 3C and Supplementary Fig. S6F). This interaction is further stabilized by Ser488 interacting with the phosphate of the complementary DNA strand, with Arg420 and Arg422 projecting into the adjacent minor groove and providing favorable electrostatics (Fig. 3C and Supplementary Fig. S6G). As opposed to the structure of *Ec*TnsE^{CTD}, the side chain of Arg388 extends into the protein core toward the phosphate backbone (Fig. 3C and D). Interestingly, this conformation of Arg388 reorients the N-terminal extension of the domain allowing residues Asn382–Phe385 to form a two-strand β-sheet with residues Leu517–Val519 (Fig. 3D). We pondered whether the rotation around Arg388 might be important to position the N-terminal domain for DNA binding. To test this idea, we generated two variants of *Ec*TnsE, *Ec*TnsE-R388A and *Ec*TnsE-R388P, to alter the flexibility and contacts of this residue. These variants bound 3′-recessed DNA similarly to *Ec*TnsE (Fig. 3E), indicating that the different orientation of the N-terminus is not critical for DNA binding.

Secondary structure and domain-boundary predictions indicated that the N- and C-terminal domains of TnsE are connected by two variable and unstructured regions bridged by short β-strands (Fig. 4A). Arg388 is preceded by the second unstructured region spanning residues 332–379. Since a point mutation at Arg388 may not suffice in restricting the movement between the two domains, we also generated the *Ec*TnsE-DR2 variant that deleted residues 332–379. *Ec*TnsE-DR2 showed similar DNA binding specificity to *Ec*TnsE (Fig. 4B). Simultaneous deletion of the two unstructured regions *Ec*TnsE-DR1-DR2 (DR1: Δ270–307 and DR2: Δ332–379) resulted in insoluble protein, suggesting that the first disordered region might be part of the N-terminal domain rather than the flexible linker connecting the two domains. Therefore, we replaced residues 270–307 with either four or six glycines to allow for proper folding of the N-terminal domain. While the *Ec*TnsE-DR1G4-DR2 and *Ec*TnsE-DR1G6-DR2 variants were less stable than *Ec*TnsE, as indicated by a 5 °C decrease in T_m, they showed the same selective binding to the 15ss + 15ds 3′-recessed end DNA substrate as *Ec*TnsE (Fig. 4B). These results imply that restricting the relative orientation between the N- and C-terminal domains of TnsE is not critical for DNA binding. In fact, TnsE variants lacking both disordered regions bound DNA marginally better than wild type and formed well defined 1:1 and 2:1 (TnsE:DNA) complexes (Fig. 4B and C), indicating that the linkers are not required for DNA binding or substrate specificity *in vitro*. However, we cannot rule out the possibility that they add plasticity to target-site selection *in vivo*.

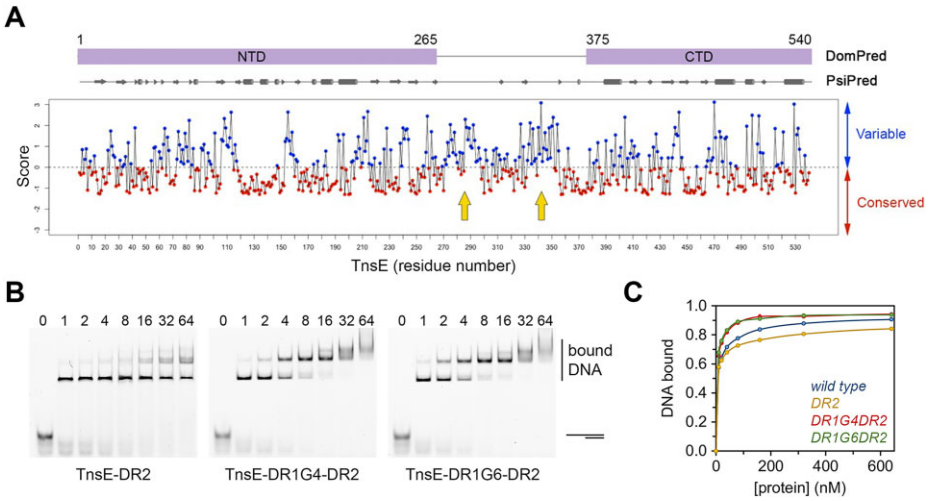


Figure 4. Flexibility between domains is not necessary for DNA binding. (**A**) Sequence conservation analysis and domain prediction for TnsE calculated with DomPred (top), PsiPred (center), and Consurf (bottom). The arrows indicate the variable regions of the predicted interdomain linker: DR1 (270–307) and DR2 (332–379). (**B**) EMSAs comparing the binding of increasing concentrations of *Ec*TnsE-DR2 [EcTnsE-Δ(332–379)], *Ec*TnsE-DR1G4-DR2 [*Ec*TnsE-Δ(270–307)Gly4-Δ(332–379)], and *Ec*TnsE-DR1G6-DR2 [*Ec*TnsE-Δ(270–307)Gly6-Δ(332–379)] to the 15ss + 15ds 3′-recessed end DNA substrate. For all gels, DNA (10 nM) was incubated with increasing concentrations of protein (10–640 nM). The protein to DNA fold excess is indicated above each lane. (**C**) DNA binding quantification. Error bars indicate standard error of the mean ($n = 3$).

Table 2. SAXS molecular weight analysis of the TnsE:DNA complex

Sample	MW^{SAXS} (kDa)	$MW^{theoretical}$	Stoichiometry
TnsE	59.1	62.3	Monomer
DNA15 (15ss + 15ds)	14.9	13.8	Monomer
DNA30 (30ss + 30ds)	27.6	27.7	Monomer
	142.7	138.3	2:1
		76.1	1:1
TnsE:DNA30	167.6	152.2	2:1
		89.9	1:1

TnsE forms 1:1 and 2:1 complexes with DNA

The crystal structure of *Ta*TnsE^{CTD} shows two molecules bound to a single DNA; we next sought to characterize how TnsE interacts with 3′-recessed DNA substrates in solution. We used the gain-of-activity variant *Ec*TnsE^{AVDN} to ensure that the V-loop was locked in the DNA-binding conformation (Fig. 3B). SEC–SAXS measurements of *Ec*TnsE^{AVDN} returned an estimated molecular weight of 59 kDa that closely matches the calculated molecular weight of the monomer and is consistent with the protein behaving as a monomer (Table 2 and Supplementary Fig. S7). *Ec*TnsE^{AVDN} mixtures with either 15ss + 15ds or 30ss + 30ds 3′-recessed DNA substrates prepared at a 1:1 stoichiometry eluted as two species from the SEC column consistent with the formation of a 2:1 (TnsE:DNA) complex and the presence of free DNA. In good agreement, SAXS curves returned molecular weights consistent with the formation of TnsE:DNA complexes with 2:1 stoichiometry (Table 2). Modeling of the scattering curves for the complex assembled using the 15ss + 15ds DNA substrate returned elongated shapes consistent with two TnsE molecules sandwiching one DNA molecule (Supplementary Fig. S8).

While this analysis confirmed that TnsE binds DNA with a 2:1 stoichiometry, scattering data are measured at high protein concentrations and this may have favored the formation of higher-order complexes. We observed multiple DNA-bound TnsE species in the EMSAs (Supplementary Fig. S5), suggesting that several DNA-bound TnsE species may coexist at lower protein concentrations. To test this idea, we carried out a series of sedimentation velocity experiments of the 15ss + 15ds DNA substrate in the presence of increasing amounts of *Ec*TnsE. As expected, both the DNA and protein on their own were monodisperse and sedimented as single species at 2.03 and 3.56 S (Fig. 5A and B), respectively, with estimated molar masses of 13 and 61 kDa, confirming that both are monomers in solution. The complex prepared with equimolar ratios of TnsE and DNA also mainly resulted in a single species and a sedimentation coefficient of 4.67 S corresponding to an experimental mass of 74 kDa, suggestive of a 1:1 protein:DNA complex (Fig. 5C, red trace). The complex prepared at a 4:1 (TnsE:DNA) ratio showed a mixture of free protein at 3.64 S and a faster sedimenting 110-kDa species at 6.28 S (Fig. 5C, gray trace), which we interpreted as a mixture of 2:1 complex and excess TnsE. The complex prepared at a 2:1 ratio showed mainly a species at 5.7 S, which could correspond to a reaction boundary composed of 2:1 complex in exchange with the 1:1 complex and free species (Fig. 5C, blue trace).

We then used native mass spectrometry (native MS) to examine these intermediate species. Under gentle nanoelectrospray ionization conditions, the native mass spectra of *Ec*TnsE:DNA complex assembled with the 15ss + 15ds DNA substrate at a 2:1 (protein:DNA) stoichiometry showed ions corresponding to free TnsE and TnsE bound to DNA as a monomer (Fig. 5D, left). In contrast, the native mass spectra of the complex assembled using the 30ss + 30ds DNA showed ions corresponding to free TnsE and TnsE bound to DNA both at 1:1 and 2:1 stoichiometries (Fig. 5D, right), confirming that these two TnsE:DNA species coexist in solution. Additionally, EMSAs recapitulated this behavior and returned a Hill coefficient of 1.4, indicating that TnsE binds DNA cooperatively.

Formation of higher-order complexes enhances Tn7 transposition

Gain-of-activity variants of TnsE bind DNA better than wild type [19]. To assess whether this was related to favoring

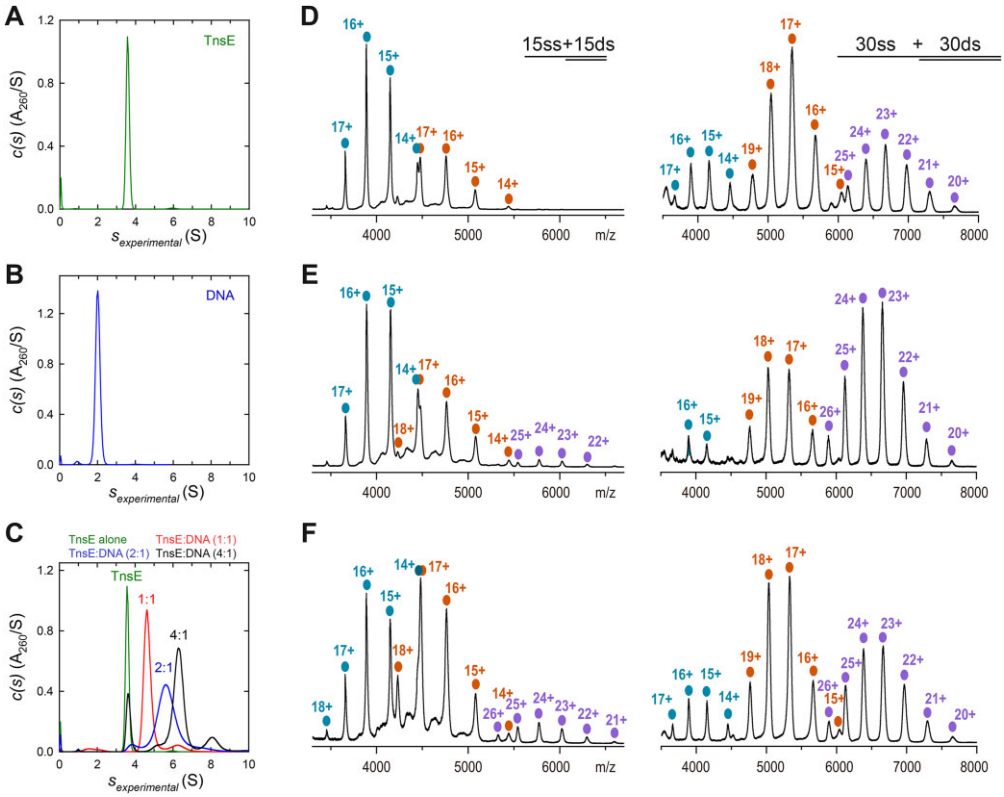


Figure 5. TnsE forms higher-order complexes on DNA. Sedimentation velocity *c*(*s*) distributions for EcTnsE (4.4 μM) (**A**), 15ss + 15ds DNA (3.6 μM) (**B**), and TnsE:DNA mixtures in 200 mM ammonium acetate buffer (**C**). Traces for the complexes assembled at 1:1, 2:1, and 4:1 (TnsE:DNA) ratios and 5 μM DNA are shown and labeled. The buffer system was chosen to match the native mass spectrometry experiments, but identical results were obtained in storage buffer. Native mass spectra of *Ec*TnsE (**D**), *Ec*TnsE-A453V/D523N (**E**), and *Ec*TnsE-D523N/E516K (**F**) in the presence of either a 15ss + 15ds (left) or a 30ss + 30ds (right) DNA substrates. The peaks corresponding to protein alone, one protein molecule bound to DNA, and two protein molecules bound to DNA are marked with color-coded circles. The charge states of each unique peak are indicated with the similar color scheme. The *y*-axis indicates the relative intensity of the peaks.

the formation of higher-order complexes on DNA, we characterized how two different gain-of-function variants interacted with the 3′-recessed DNA substrates by native MS. We selected gain-of-activity variants with similar DNA binding abilities but different transposition frequencies [14, 19]. The native MS spectra for *Ec*TnsE-A453V/D523N ($TnsE^{AVDN}$) with the 15ss + 15ds DNA substrate showed ions corresponding to free protein (monomer) and protein bound to DNA with both 1:1 and 2:1 ($TnsE^{AVDN}$:DNA) stoichiometries (Fig. 5E, left), although the fraction of 2:1 complex was small. In contrast, complexes with 2:1 ($TnsE^{AVDN}$:DNA) stoichiometry were the predominant species for the complex assembled with the 30ss + 30ds DNA substrate (Fig. 5E, right). Next, we carried out the same analysis using the *Ec*TnsE-D523N/E516K ($TnsE^{DNEK}$) variant. $TnsE^{DNEK}$ has lower transposition frequency than $TnsE^{AVDN}$ but binds DNA similarly [19]. We found that $TnsE^{DNEK}$ also formed 2:1 complexes on DNA more readily than wild-type TnsE (Fig. 5F). To quantify the differences between TnsE variants, we calculated the percentage of TnsE bound to DNA at 1:1 and 2:1 ratios by integrating the area of the individual peaks that could unequivocally be assigned to a single species. While only 4.5% ± 1.7% corresponded to 2:1 complex for wild-type TnsE bound to 15ss + 15ds DNA, this percentage increased to 11.4% ± 2.6% and 16.2% ± 4.5% for $TnsE^{AVDN}$ and $TnsE^{DNEK}$ variants, respectively (Fig. 6A). The longer DNA substrate favoured the association of two TnsE molecules to DNA as evidenced by the increased percentages of 2:1 complex when using a 30ss + 30ds DNA substrate. This finding was consistent with the EMSAs showing slower migrating species with this substrate (Fig. 6B and C). Interestingly, the $TnsE^{AVDN}$ variant causes a three-fold increase in 2:1 complex with respect to wild-type TnsE, while the increase is more modest for the $TnsE^{DNEK}$ variant (Fig. 6B and C). The differences observed by native MS recapitulate the electrophoretic mobility shift differences observed for these two gain-of-activity variants (Fig. 6). Since $TnsE^{AVDN}$ is the most efficient gain-of-activity variant identified thus far with nearly 1000-fold increase in transposition frequency [19], these results strongly suggest that the formation of a 2:1 complex may be critical to target replication structures.

TnsE:DNA higher-order complexes are asymmetric

We predicted that TnsE would form higher-order complexes on DNA by engaging the C-terminal domain on dsDNA as seen in the crystal structure (Fig. 2A). However, in the crystal structure, the two $TaTnsE^{CTD}$ molecules sandwiching the DNA are related by crystallographic symmetry, causing them to bind opposite faces of the DNA duplex and adopt an "antiparallel" configuration (Fig. 7A, left). As a result, the N-terminus of each $TnsE^{CTD}$ protomer points in opposite directions. Since we could not exclude that this arrangement was an artifact of the crystal packing, we constructed a model with the two protomers bound to DNA in a "parallel" configura-

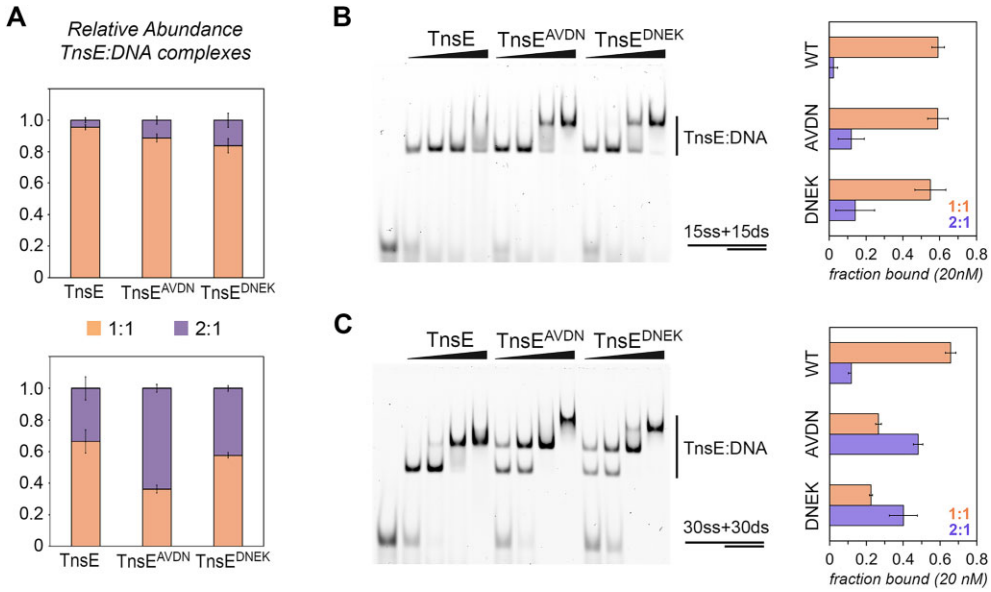


Figure 6. Gain-of-activity TnsE variants promote DNA-mediated oligomerization. (**A**) Quantification of the relative abundance of single (1:1) and double (2:1) TnsE protomers bound to DNA from the native mass spectra of *Ec*TnsE, *Ec*TnsE-A453V/D523N, and *Ec*TnsE-D523N/E516K with the 15ss + 15ds (top) and 30ss + 30ds (bottom). Error bars indicate standard error of the mean ($n = 3$). Electrophoretic mobility assay gels comparing the migration of the 15ss + 15ds (**B**) and 30ss + 30ds (**C**) DNA substrates (10 nM) in the presence of increasing concentrations of *Ec*TnsE (10, 20, 40, and 80 nM), or the gain-of-activity variants, *Ec*TnsE-A453V/D523N and *Ec*TnsE-D523N/E516K, as indicated. The right-hand side panels show the quantification of the EMSAs at a protein concentration of 20 mM. Error bars indicate standard error of the mean ($n = 3$).

tion. In this model, we staggered the TnsE^{CTD} protomers by 4 bp to maintain identical interactions with the duplex DNA to those seen in the crystal structure while keeping the protomers on opposite faces of the DNA duplex (Fig. 7A, right). We then use cryo-EM to probe the two binding modes.

We assembled the protein:DNA complex with *Ec*TnsE^{AVDN} and the 30ss + 30ds DNA substrate at a 2:1 ratio, resolved it over SEC, and applied it to cryo-EM grids. The particles on the cryo-EM micrographs were elongated and resulted in elongated two-dimensional (2D) class averages (Fig. 7B). Several 2D class averages showed high-resolution features resembling two TnsE^{CTD} molecules sandwiching a DNA duplex (Fig. 7B), but with minimal additional densities. For 2D class averages with additional densities, the structural details were less prominent indicating an inherent flexibility between the two domains of the protein. *Ab initio* reconstruction followed by heterogeneous and nonuniform refinements resulted in a final map with an overall resolution of 6.1 Å from 110k particles (Fig. 7C and Supplementary Figs S1 and S2). The map has an overall dumbbell shape with dimensions consistent with those estimated from SAXS (Fig. 7C, Table 2, and Supplementary Figs S7 and S8).

One of the globular densities of the cryo-EM map is better defined than the other and resembles two C-terminal domains sandwiching the DNA (Fig. 7C). We fitted the "antiparallel" and "parallel" configurations into this region of the cryo-EM map using the ChimeraX "fit in map" tool. We found that the "parallel" configuration resulted in twice as many atoms outside the volume as for the "antiparallel" configuration (Supplementary Fig. S3). Subsequent rigid-body refinement also returned better correlation coefficients for the "antiparallel" configuration (Supplementary Fig. S3), suggesting that full-length TnsE most likely binds to 3′-recessed end DNA substrates in the same "antiparallel" configuration observed in the crystal structure for TnsE^{CTD} (Fig. 7C).

AlphaFold predicts the N- and C-terminal domains of *Ec*TnsE (AF-P05845-F1) with high confidence (Fig. 7D). Conversely, the linker connecting the two domains (DR2, residues 332–379) and a disordered loop within the N-terminal domain (DR1, residues 270–307) are predicted with much lower confidence. Superimposition of the predicted TnsE^{CTD} model onto the crystal structures of *Ec*TnsE^{CTD} and *Ta*TnsE^{CTD} returns root mean square deviations of 0.73 and 0.71 Å, respectively, confirming that AlphaFold predicts this domain correctly. Therefore, we assumed that the prediction for the N-terminal domain (*Ec*TnsE^{NTD}) would also be correct (Fig. 7E). The overall dimensions of the TnsE^{NTD} domain closely resemble the additional density observed in the cryo-EM map (Fig. 7C–E). Although the limited features of the map in this region made it difficult to fit this domain, its size suggests that only one of the N-terminal domains in the 2:1 (TnsE:DNA) complex is bound to the single-stranded portion of the DNA. This, in turn, suggests that the TnsE^{NTD} domain is flexible when not bound to DNA, likely explaining why the N-terminal domain for the second protomer is not visible in our cryo-EM maps. However, this hypothesis still requires experimental validation.

Efforts to predict the structure of TnsE bound to DNA using AlphaFold3 failed to produce consistent models in which both domains were bound to DNA (Supplementary Fig. S9). However, these models provided a framework to propose how TnsE^{NTD} interacts with 3′-recessed end DNA substrates. Most of the models calculated using a 1:1 (TnsE:DNA) complex showed a stretch of four to five nucleotides anchored in a shallow groove delimited by positive charged residues (His18, Arg40, Lys41, and Lys98) and lined with aromatic residues (Phe20, Phe31, Trp33, Trp42, and Phe92) that stabilize the nucleobase moieties (Fig. 8A). This consistent binding feature suggests that electrostatic interactions may favor the single strand portion of the substrate to wrap around TnsE^{NTD} ,

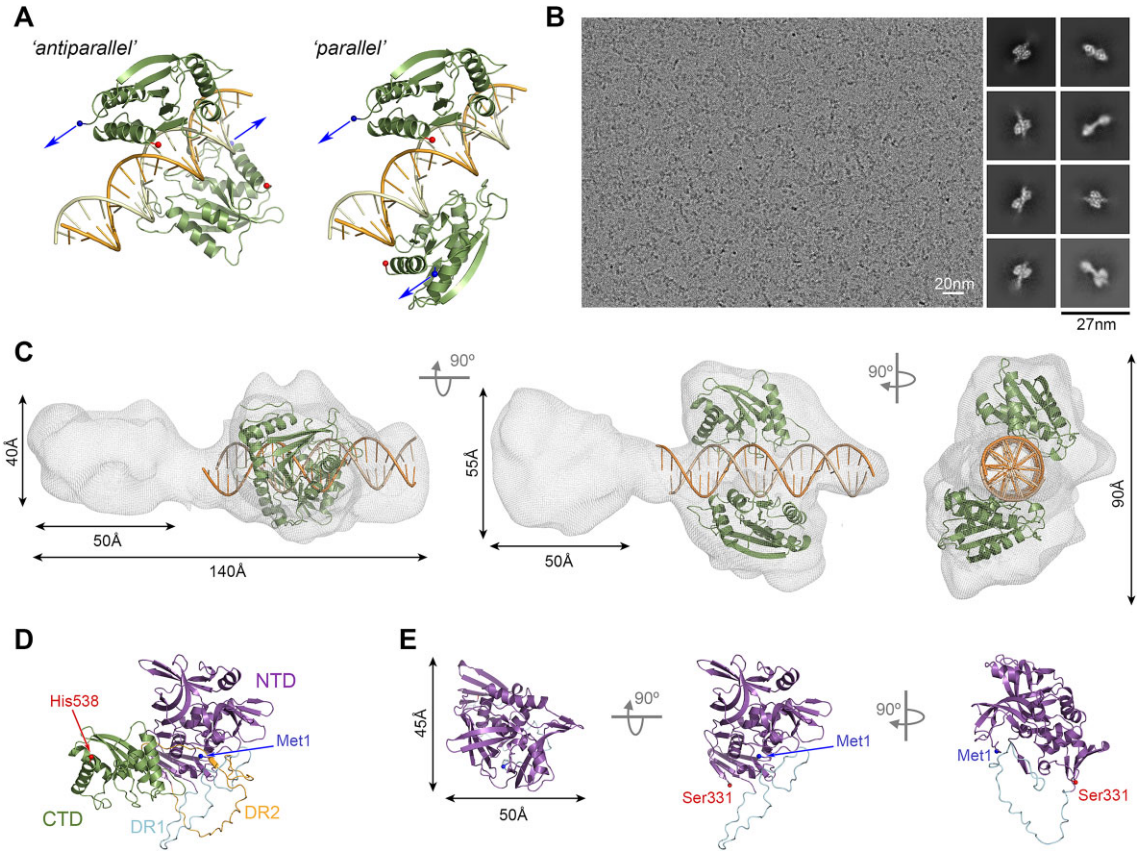


Figure 7. TnsE binds DNA in an antiparallel configuration. (**A**) Antiparallel configuration observed in the crystal structure of *Ta*TnsE^{CTD} bound to DNA (left). Model of the parallel configuration generated by rotating one of the *Ta*TnsE^{CTD} molecules by 180° and translating by 4 bp to maintain identical DNA contacts (right). The N- and C-terminal residues of the domain are marked with blue and red spheres, respectively. (**B**) Representative cryo-EM and 2D classes from the cryo-EM dataset of *Ec*TnsE^{AVDN} bound to 30ss + 30ds DNA substrate. Scale bars represent 20 and 27 nm, respectively. (**C**) Three orthogonal views of the final cryo-EM map shown as a gray mesh contoured at 3σ. The crystal structure of TnsE^{CTD} bound to DNA is fitted into the distal globular densities of the map (see also Supplementary Fig. S8). The dimensions of the map are shown for reference. (**D**) AlphaFold prediction of full-length TnsE. The N-terminal domain (NTD; residues 1–331) is shown in purple, the internal disordered loop (DR1; residues 271–315) in blue, the linker connecting the two domains (DR2; residues 332–380) in gold, and the C-terminal domain (CTD; residues 381–538) in green. (**E**) Orthogonal views of the AlphaFold prediction of the N-terminal domain of TnsE shown with the same color coding as in panel (D). The approximate dimensions are indicated for comparison to the cryo-EM map.

while hydrophobic and π–π stacking interactions may introduce a kink to recruit TnsC in an analogous manner to TnsD [13].

Discussion

The integrative approach presented here describes how TnsE recognizes 3′-recessed end DNA substrates. We find that TnsE combines a switch loop to regulate DNA binding with DNA-dependent formation of higher-order complexes to select its DNA binding substrates. Wild-type TnsE binds 3′-recessed end DNA substrates at 1:1 and 2:1 stoichiometries but gain-of-function variants of TnsE favor the formation of the 2:1 complex. This suggests that formation of higher-order complexes of TnsE on DNA is important to promote TnsE-directed insertions of Tn7. Intriguingly, the two TnsE protomers in the 2:1 complex interact with DNA in an antiparallel and asymmetric manner, presumably with only one of the TnsE^{NTD} domains bound to DNA (Fig. 7C and 8A). AlphaFold predicts a compact conformation for TnsE, which does not align with the extended conformation of TnsE observed by SAXS (Supplementary Fig. S7) or that of the TnsE:DNA complex seen in cryo-EM (Fig. 7D). Although AlphaFold predictions do not always hold up against experimental verification, it is tempting to speculate that additional conformational changes mediated by the disordered regions of the protein could also restrict the frequency of TnsE-directed transposition events [1].

It has been previously shown that the interaction between TnsE and the sliding β-clamp is essential for transposition *in vivo* [18]. This interaction is mediated by the N-terminal domain of the protein and determines the spacing and orientation of Tn7 insertions [18]. Therefore, it is plausible that the free TnsE^{NTD} of the 2:1 TnsE:DNA complex interacts with the β-clamp to enhance DNA binding. This would indicate a sequential assembly where the interaction with the β-clamp tethers the first TnsE protomer to actively replicating DNA sites, facilitating binding of its C-terminal domain to duplex DNA, and cooperatively promoting the binding of a second TnsE protomer. The N-terminal domain of the second TnsE protomer would then be able to bind and remodel the single-strand end of the substrate to recruit TnsC to the target site (Fig. 8B).

Previous work from us and others revealed how TnsC helps determine the spacing between the target and insertion sites for different Tn7 elements [4, 40, 41]. A rough spacing cal-

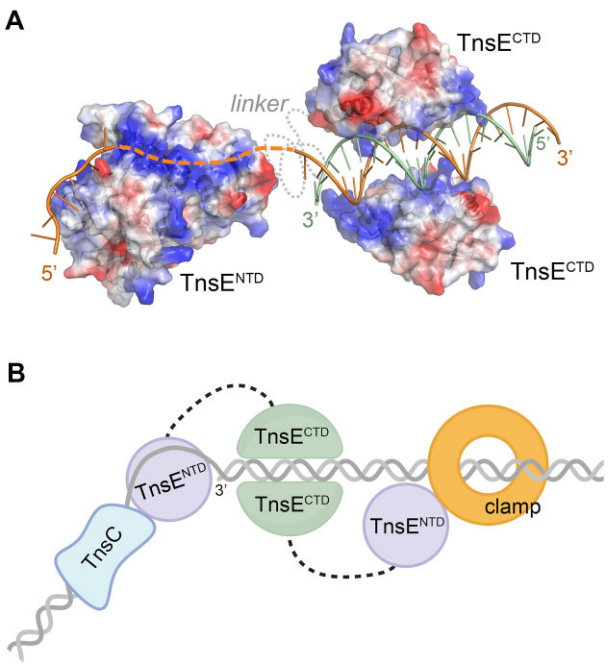


Figure 8. TnsE target-site selection complex. (**A**) Model of the TnsE dimer bound to 3′-recessed end DNA built from the crystal structure of *Ta*TnsE^{CTD} bound to DNA and the AlphaFold model of *Ec*TnsE bound to DNA. The protein moieties are shown as a semi-transparent electrostatic surface (with red and blue indicating negative and positive charges, respectively). The disordered protein linker is shown as a gray dotted line and the absent intervening DNA strand as a dashed orange line. (**B**) Cartoon representation of TnsE being recruited to target sites through the interaction on its N-terminal with the sliding β-clamp, in turn facilitating binding of the C-terminal domain of TnsE to DNA and promoting DNA-mediated dimerization. The N-terminal domain of the second protomer would then be able to bind and remodel the single-strand end of the substrate to recruit TnsC to the target site. The disordered linker between the two domains is represented as dashed lines.

culation based on the cryo-EM reconstructions of *Ec*TnsC bound to DNA [4] and *Ec*TnsE bound to DNA (Fig. 7C), as well as the *Ec*TnsE^{NTD}–DNA interactions predicted by AlphaFold (Supplementary Fig. S9), places the Tn7 insertion site ∼60 bp away from the free 3′ end. This distance is comparable to TnsABC + E transposition assays showing that the β-clamp preferentially directs Tn7 insertions 66 bp away from the free 3′ end of an ssDNA gap with a left-to-right orientation [18] (Fig. 1). Our model, wherein the two N-terminal domains of TnsE have asymmetric functions within the targeting complex, is further supported by experimental evidence showing that in the absence of the β-clamp *in vitro*, insertions occur randomly throughout the circular DNA in different orientations [18].

The modular and directional assembly is a common feature for other target-site selection pathways in the Tn7 family. For TnsD-directed insertions, the C-terminal domain of the protein specifically recognizes the *glmS* site, while its N-terminal domain bends DNA to recruit TnsC [13]. In other Tn7-like elements, targeting relies on the coordinated binding of a TniQ/TnsD protein and other target selectors to recognize a specific site and recruit TnsC [40–42]. TniQ/TnsD target-site selectors promote insertions in a right-to-left orientation. This orientation bias can be recapitulated by a distortion on the DNA when using a gain-of-function variant of TnsC [12]. This indicates that the TnsABC are the only transposon-encoded proteins required to drive insertions in this orientation. In contrast, TnsE-directed insertions occur in a left-to-right orientation highlighting the importance of target-site selection complex assembly at reversing the orientation of the insertions. The involvement of host factors is also a common feature across Tn7 elements. The ribosomal protein uL29 and the acyl carrier protein work together to enhance TnsD binding to its cognate binding site, in turn stimulating Tn7 transposition [43]. Furthermore, the ribosomal protein uS15 stabilizes the target-site selection complex in type V-K Tn7-like elements [41]. It is conceivable that the interaction with β-clamp restricts the flexibility of the TnsE^{NTD} domain that is not bound to ssDNA, explaining why one of the *Ec*TnsE^{NTD} in the 2:1 complex is invisible in our cryo-EM maps. Understanding the exact role of the β-clamp at recruiting TnsE, however, awaits further investigation.

Together with recent studies on other Tn7 target-site selection pathways, this work highlights two emerging features common to all Tn7 targeting mechanisms. First, the strict spacing between the target and insertion sites is defined by the minimal DNA length required to assemble a modular target-site selection complex. Second, target-site specificity is influenced by the ability to form a compact complex. Future research will help clarify the molecular requirements for the assembly of the different Tn7 target-site selection complexes and understand how their assembly predetermines the orientation bias of Tn7 insertions.

Acknowledgements

We are grateful to members of the Guarné and Thibodeaux laboratories for helpful discussions. Cryo-EM screening and data collection were performed at the McGill Facility for Electron Microscopy Research (FEMR). We thank Dr Kaustuv Basu for assistance during cryo-EM data collection and Dr Joaquin Ortega for helpful discussion about cryo-EM data processing.

Author contributions: Conceptualization: A.G. and S.S.K. Investigation: S.S.K., Y.S., T.B.OH., L.A.M., N.W.W., and R.G. Formal analysis: S.S.K. and A.G. Visualization: S.S.K. and A.G. Writing—original draft: S.S.K. and A.G. Writing—review & editing: A.G. with contributions from all authors. Supervision: A.G. and C.J.T.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was funded by the Canadian Institutes of Health Research (PJT-155941 and PJT-189946 to A.G.) and the Intramural Research Program of the National Institutes of Health (to R.G.). S.S.K. and T.B.O'H. were funded in part by a Fonds de Recherche du Québec—Santé doctoral awards. The CMCF-BM beamline at the Canadian Light Source is a national research facility of the University of Saskatchewan supported by the Canada Foundation for Innovation, the Natural

Sciences and Engineering Research Council, the National Research Council, the Canadian Institutes of Health Research, the Government of Saskatchewan, and the University of Saskatchewan. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Funding to pay the Open Access publication charges for this article was provided by Canadian Institutes of Health Research (PJT-189946).

Data availability

Coordinates and structure factors generated from this work have been deposited in the Protein Data Bank (PDB ID 9D5L). Cryo-EM maps are available from the Electron Microscopy Data Bank (EMD-46582). All other data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Peters JE, Craig NL. Tn7: smarter than we thought. *Nat Rev Mol Cell Biol* 2001;**2**:806–14. https://doi.org/10.1038/35099006
2. Kaczmarska Z, Czarnocki-Cieciura M, Gorecka-Minakowska KM *et al.* Structural basis of transposon end recognition explains central features of Tn7 transposition systems. *Mol Cell* 2022;**82**:2618–32. https://doi.org/10.1016/j.molcel.2022.05.005
3. Turlan C, Chandler M. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol* 2000;**8**:268–74. https://doi.org/10.1016/S0966-842X(00)01757-1
4. Shen Y, Gomez-Blanco J, Petassi MT *et al.* Structural basis for DNA targeting by the Tn7 transposon. *Nat Struct Mol Biol* 2022;**29**:143–51. https://doi.org/10.1038/s41594-022-00724-8
5. Waddell CS, Craig NL. Tn7 transposition: recognition of the attTn7 target sequence. *Proc Natl Acad Sci USA* 1989;**86**:3958–62. https://doi.org/10.1073/pnas.86.11.3958
6. Peters JE. Gene transfer in Gram-negative bacteria. In: *Methods for General and Molecular Microbiology*. Reddy CA, Washington DC: ASM Press, 2007, 735–755.
7. Rybarski JR, Hu K, Hill AM *et al.* Metagenomic discovery of CRISPR-associated transposons. *Proc Natl Acad Sci USA* 2021;**118**:e2112279118.
8. Hsieh SC, Peters JE. Natural and engineered guide RNA-directed transposition with CRISPR-associated Tn7-like transposons. *Annu Rev Biochem* 2024;**93**:139–61. https://doi.org/10.1146/annurev-biochem-030122-041908
9. Peters JE. Targeted transposition with Tn7 elements: safe sites, mobile plasmids, CRISPR/Cas and beyond. *Mol Microbiol* 2019;**112**:1635–44. https://doi.org/10.1111/mmi.14383
10. Bainton RJ, Kubo KM, Feng JN *et al.* Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. *Cell* 1993;**72**:931–43. https://doi.org/10.1016/0092-8674(93)90581-A
11. Kuduvalli PN, Rao JE, Craig NL. Target DNA structure plays a critical role in Tn7 transposition. *EMBO J* 2001;**20**:924–32. https://doi.org/10.1093/emboj/20.4.924
12. Rao JE, Miller PS, Craig NL. Recognition of triple-helical DNA structures by transposon Tn7. *Proc Natl Acad Sci USA* 2000;**97**:3936–41. https://doi.org/10.1073/pnas.080061497
13. Shen Y, Krishnan SS, Petassi MT *et al.* Assembly of the Tn7 targeting complex by a regulated stepwise process. *Mol Cell* 2024;**84**:2368–81. https://doi.org/10.1016/j.molcel.2024.05.012
14. Peters JE, Craig NL. Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev* 2001;**15**:737–47. https://doi.org/10.1101/gad.870201
15. Wolkow CA, DeBoy RT, Craig NL. Conjugating plasmids are preferred targets for Tn7. *Genes Dev* 1996;**10**:2145–57. https://doi.org/10.1101/gad.10.17.2145
16. Peters JE, Craig NL. Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol Cell* 2000;**6**:573–82. https://doi.org/10.1016/S1097-2765(00)00056-3
17. Shi Q, Parks AR, Potter BD *et al.* DNA damage differentially activates regional chromosomal loci for Tn7 transposition in *Escherichia coli*. *Genetics* 2008;**179**:1237–50. https://doi.org/10.1534/genetics.108.088161
18. Parks AR, Li Z, Shi Q *et al.* Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* 2009;**138**:685–95. https://doi.org/10.1016/j.cell.2009.06.011
19. Shi Q, Straus MR, Caron JJ *et al.* Conformational toggling controls target site choice for the heteromeric transposase element Tn7. *Nucleic Acids Res* 2015;**43**:10734–45. https://doi.org/10.1093/nar/gkv913
20. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:133–44. https://doi.org/10.1107/S0907444909047374
21. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:125–32. https://doi.org/10.1107/S0907444909047337
22. Adams PD, Afonine PV, Bunkoczi G *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:213–21. https://doi.org/10.1107/S0907444909052925
23. Emsley P, Lohkamp B, Scott WG *et al.* Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:486–501. https://doi.org/10.1107/S0907444910007493
24. Schindelin J, Arganda-Carreras I, Frise E *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**:676–82. https://doi.org/10.1038/nmeth.2019
25. Manalastas-Cantos K, Konarev PV, Hajizadeh NR *et al.* ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr* 2021;**54**:343–55. https://doi.org/10.1107/S1600576720013412
26. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* 2018;**34**:1944–6. https://doi.org/10.1093/bioinformatics/btx846
27. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 1992;**25**:495–503. https://doi.org/10.1107/S0021889892001663
28. Franke D, Svergun DI. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr* 2009;**42**:342–6. https://doi.org/10.1107/S0021889809000338
29. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999;**76**:2879–86. https://doi.org/10.1016/S0006-3495(99)77443-6
30. Zhao H, Brautigam CA, Ghirlando R *et al.* Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation. *CurrProtoc Protein Sci* 2013;**Chapter 20**:Unit 20.12. https://doi.org/10.1002/0471140864.ps2012s71
31. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 2000;**78**:1606–19. https://doi.org/10.1016/S0006-3495(00)76713-0
32. Cole JL, Lary JW, Moody TP *et al.* Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* 2008;**84**:143–79. https://doi.org/10.1016/S0091-679X(07)84006-4

33. Schorb M, Haberbosch I, Hagen WJH *et al.* Software tools for automated transmission electron microscopy. *Nat Methods* 2019;**16**:471–7. https://doi.org/10.1038/s41592-019-0396-9
34. Punjani A, Rubinstein JL, Fleet DJ *et al.* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 2017;**14**:290–6. https://doi.org/10.1038/nmeth.4169
35. Meng EC, Goddard TD, Pettersen EF *et al.* UCSF ChimeraX: tools for structure building and analysis. *Protein Sci* 2023;**32**:e4792. https://doi.org/10.1002/pro.4792
36. Varadi M, Bertoni D, Magana P *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* 2024;**52**:D368–75. https://doi.org/10.1093/nar/gkad1011
37. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
38. Nanji T, Gehrke EJ, Shen Y *et al.* *Streptomyces* IHF uses multiple interfaces to bind DNA. *Biochim Biophys Acta Gen Subj* 2019;**1863**:129405. https://doi.org/10.1016/j.bbagen.2019.07.014
39. Swiercz JP, Nanji T, Gloyd M *et al.* A novel nucleoid-associated protein specific to the actinobacteria. *Nucleic Acids Res* 2013;**41**:4171–84. https://doi.org/10.1093/nar/gkt095
40. Park JU, Tsai AW, Rizo AN *et al.* Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature* 2023;**613**:775–82. https://doi.org/10.1038/s41586-022-05573-5
41. Schmitz M, Querques I, Oberli S *et al.* Structural basis for the assembly of the type V CRISPR-associated transposon complex. *Cell* 2022;**185**:4999–5010. https://doi.org/10.1016/j.cell.2022.11.009
42. Wang S, Gabel C, Siddique R *et al.* Molecular mechanism for Tn7-like transposon recruitment by a type I-B CRISPR effector. *Cell* 2023;**186**:4204–15. https://doi.org/10.1016/j.cell.2023.07.010
43. Sharpe PL, Craig NL. Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. *EMBO J* 1998;**17**:5822–31. https://doi.org/10.1093/emboj/17.19.5822

Received: August 21, 2024. **Revised:** April 25, 2025. **Editorial Decision:** April 30, 2025. **Accepted:** May 20, 2025