Science
AAAS

Supplementary Materials for

Mechanism of target site selection by type V-K CRISPR-associated transposases

Jerrin Thomas George *et al*.

Corresponding author: Samuel H. Sternberg, shsternberg@gmail.com

Science **382**, eadj8543 (2023)
DOI: 10.1126/science.adj8543

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods:

Plasmid construction

All constructs used in the experiments with ShCAST were cloned using the pHelper (pUC19) and pDonor (pCDFDuet1) described previously (*14*). These constructs were derived from a CAST system from *Scytonema hofmanni* UTEX B 2349 (*11*). Cloning was performed using a combination of Gibson assembly, around-the-horn (inverse) PCR, or restriction digestion and ligation. All DNA fragments used for cloning were PCR amplified using Q5 DNA polymerase (NEB). TnsC and Cas12k were cloned into the pCOLADuet1 vector backbone using Gibson assembly. The expression of TnsC and Cas12k was varied by cloning inducible (lac, T7) or constitutive promoters (J23114, J23105, J23119) using around-the-horn PCR. *S. hofmanni* S15 was cloned into *E. coli* expression constructs and as fusion constructs using a synthesized gene fragment (TWIST Biosciences). Varying spacers for guide RNA in WT pHelper or pHelper lacking TnsC were designed by around-the-horn PCR.

Constructs for CAST protein purification having an N-terminal hexahistidine fused to a SUMO tag and an in-frame TEV cleavage site (His_6-SUMO-TEV) were designed by restriction digestion of the p1S vector (QB3 MacroLab) with Ssp1-HF (NEB) followed by Gibson assembly of a fragment containing the protein open reading frame. Protein expression was carried out in *E. coli* BL21 (DE3) cells.

pTarget for biochemical transposition experiments was designed by cloning a target site from λ-DNA into the pUC19 vector. Modulating regions in pTarget (pT-1 to pT-6) to induce artificial untargeted transposition was achieved using around-the-horn PCR.

All primers and DNA oligos used in the study were ordered from IDT. Cloning was performed using NEB Turbo competent *E. coli* cells, and plasmids were extracted using Qiagen Miniprep columns. The cloned constructs were verified by Sanger sequencing (GENEWIZ) or whole plasmid sequencing by Oxford Nanopore (Plasmidsaurus). *E. coli* was grown in LB agar plates or liquid LB media. Constructs with pUC19 backbone were grown in 100 μg/mL carbenicillin, pCOLADuet1 and p1S backbone in 50 μg/mL kanamycin, and pCDFDuet1 backbone in 100 μg/mL spectinomycin. All plasmid sequences and their descriptions are available in table S2. The sequences of recombinantly expressed proteins used in this study are available in table S3.

E. coli transposition assay

Transposition assay was performed in *E. coli* BL21(DE3) cells based on a method described previously (*14*), using plasmid systems composed of pDonor, and pHelper. Chemically competent cells having pDonor were transformed with pHelper, plated, and grown at 37 °C for 16 h in the presence of appropriate antibiotics (100 μg/mL carbenicillin and 100 μg/mL spectinomycin). Cells

were scrapped off and resuspended in LB. 4 x 10^9 cells were re-plated, induced, and grown in the presence of 0.1 mM IPTG and antibiotics for 24 h at 37 °C. Cells were scrapped and resuspended in LB. 2 x 10^9 cells were used for genomic extraction using Wizard Genomic DNA Purification Kit (Promega). For transposition experiments requiring an additional plasmid expressing Cas12k, TnsC, or TnsB, competent cells having pDonor and, pCas12k, pTnsC, or pTnsB were transformed with the pHelper and grown at the same conditions in the presence of appropriate antibiotics (50 µg/mL kanamycin, 100 µg/mL carbenicillin and 100 µg/mL spectinomycin). All *E. coli* transposition experiments were performed in $N = 2$ biological replicates.

Recombinant expression of CAST proteins

All CAST proteins were cloned into a pET derivative vector with an N-terminal His_6-SUMO-TEV fusion and were purified similarly to previous protocols (*11*, *24*). The constructs were transformed in *E. coli* BL21(DE3) and grown in 2xYT in the presence of kanamycin (50 µg/mL) at 37 °C until an OD600 of 0.5-0.6 was reached. Protein expression was induced in the presence of 0.5 M isopropyl ß-D-1-thiogalactopyranoside (IPTG) for 16 h at 16 °C.

To purify TnsB and TniQ, cells were harvested by centrifuging at 6000 g for 15 min at 4 °C and were lysed by sonication in lysis buffer (20 mM Tris pH 7.5, 0.5 M NaCl, 5 mM $Mg(OAc)_2$, 10 mM imidazole, 0.1% Triton X-100, 1 mM DTT and 5% (v/v) glycerol) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.2 mg/mL lysozyme and 0.5x cOmplete protease inhibitor cocktail tablet (Roche). Soluble protein was isolated by centrifuging at 35,050 g for 50 min, and the supernatant was incubated with Ni-NTA agarose beads, pre-equilibrated with equilibration buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM PMSF, 1 mM DTT, 10 mM imidazole and 5% (v/v) glycerol) for 45 min at 4 °C. The Ni-NTA beads were further washed with 20 column volumes (CVs) of wash buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM PMSF, 1mM DTT, 30 mM imidazole and 5% glycerol) in a gravity column and protein was eluted in elution buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 0.5 mM PMSF, 1 mM DTT, 300 mM imidazole, and 5% glycerol). Next, the His_6-SUMO-tag was cleaved off using 5% (w/v) TEV protease and was simultaneously dialyzed in a Slide-a-Lyzer cassette (Thermo Fischer Scientific) against dialysis buffer (20 mM HEPES pH 7.5, 0.2 M NaCl, 0.5 mM PMSF, 1 mM DTT, and 1% (v/v) glycerol) for 12 h at 4 °C. Any precipitated protein was removed by centrifugation at 20,000 g for 20 min at 4 °C. The proteins were subjected to Heparin-affinity chromatography using a 0.2-1 M NaCl gradient spanning 10 CV in a buffer containing 20 mM HEPES pH 7.5, 0.2 M NaCl, 1 mM DTT and 1% glycerol. It should be noted that TniQ does not bind to Heparin and elutes in the flowthrough, whereas contaminant proteins were removed in this step. Further, both the proteins were purified by size-exclusion chromatography in SEC Buffer (20 mM HEPES pH 7.5, 0.2 M NaCl, 1mM DTT and 1% glycerol) using a Superdex 200pg 16/600 column. TniQ and TnsB were concentrated to 127 and 100 µM and stored in a buffer containing 20 mM HEPES pH 7.5, 0.25 M

NaCl, 5% glycerol, 1 mM DTT at -80 °C. We note no contamination of S15 in our TniQ potentially due the step involving Heparin (table S3)

For Cas12k, the lysis, equilibration, wash, and elution buffers contained Tris pH 8.0 instead of HEPES pH 7.5. The dialysis buffer used contained 20 mM HEPES pH 7.5, 0.25 M NaCl, 0.5 mM PMSF, 1 mM DTT, and 1% (v/v) glycerol. The SEC buffer contained 20 mM HEPES pH 7.5, 0.25 M NaCl, 1 mM DTT, and 1% (v/v) glycerol. The rest of the protocol and buffers used were the same as those for TnsB and TniQ. Cas12k was concentrated to 109 µM and stored under the same conditions as TnsB and TniQ.

For TnsC, the lysis, equilibration, wash, and elution buffers contained 0.5 M NaCl to increase protein solubility. TEV cleavage and dialysis were performed in 20 mM HEPES pH 7.5, 1 M NaCl, 0.5 mM PMSF, 1 mM DTT, and 1% (v/v) glycerol after which the protein was subject to a second round of Ni-NTA to separate His_6-SUMO tag from the cleaved TnsC. The column was washed with 20 mM HEPES pH 7.5, 500 mM NaCl, 30 mM imidazole, 1 mM DTT, 0.5 mM PMSF, 5% (v/v) glycerol, 1 mM DTT, 0.5 mM PMSF and the flowthrough was concentrated and further purified by size exclusion chromatography using Superdex 200pg 16/600 column in a buffer containing 20 mM HEPES pH 7.5, 1 M NaCl, 5% glycerol, 1 mM DTT. TnsC was concentrated to 142 µM and stored in a buffer containing 20 mM HEPES pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 1 mM DTT.

For S15, the lysis, equilibration, wash, and elution buffers contained Tris pH 8.0 instead of HEPES pH 7.5. The dialysis buffer contained 20 mM HEPES pH 7.5, 0.15 M NaCl, 0.5 mM PMSF, 1 mM DTT, and 1% (v/v) glycerol. S15 was eluted from Heparin using a gradient of 0.15 to 1 M NaCl in 20 mM HEPES pH 7.5, 1mM DTT and 15% (v/v) glycerol. The SEC buffer contained 20 mM HEPES pH 7.5, 0.25 M NaCl, 1 mM DTT, and 1% (v/v) glycerol. All the rest of the protocols and buffers used were the same as in case of TnsB and TniQ. S15 was concentrated to 168 µM and stored under the same conditions as TnsB and TniQ.

BCQ transpososome sample preparation

The BCQ transpososome was reconstituted as previously described for the Cas12k-bound transpososome (*26*), with some modifications noted below. Two separate DNA substrates were first prepared by annealing three synthetic oligonucleotides (IDT) respectively: LE1, LE2, and LE3 for target-LE DNA and RE1, RE2, and RE3 for target-RE DNA. The oligonucleotides were mixed in a 1:1:1 molar ratio and supplemented with 10X concentrated annealing buffer for the following final composition: 10 mM Tris pH 7.5, 50 mM NaCl, and 1 mM EDTA. These mixtures were then heated up to 95 °C for 5 minutes and slowly cooled to room temperature at the rate of 1 °C per minute using a thermal cycler (BioRad).

TnsB, TnsC, and TniQ proteins were purified identically as previously described (*11*, *26*). All the protein stocks were first buffer-exchanged into the dilution buffer (25 mM HEPES pH 7.5,

200 mM NaCl, 1 mM DTT, and 2% glycerol) using 0.5 mL centrifugal filters (Millipore). For the target pot, TnsC, TniQ, and the annealed target-LE DNA were first mixed and supplemented with ATP and $MgCl_2$, resulting in the final composition of the following: 45 µM TnsC, 30 µM TniQ, 3 µM target-LE DNA, 3 mM ATP, and 10 mM $MgCl_2$ in the dilution buffer. For the donor pot, TnsB and target-RE DNA were combined to make the following composition: 36 µM TnsB, 6 µM target-RE DNA, and 10 mM $MgCl_2$ in dilution buffer. Target and donor pots were incubated independently for 10 min at room temperature before being combined in a 2:1 volume ratio to reconstitute the transpososome. The composition of the final reaction condition was the following: 30 µM TnsC, 20 µM TniQ, 12 µM TnsB, 2 µM target-LE DNA, 2 µM target-RE DNA, 2 mM ATP, and 10 mM $MgCl_2$. The final reaction mixture was then incubated for 30 min at 37 °C.

Cryo-EM sample preparation and imaging

Homemade graphene-oxide (GO) coated grids were made as previously described (*4*, *46*) or cryo-EM sample preparation. The BCQ transpososome sample was prepared by diluting the final reconstitution mixture three-fold using the dilution buffer (described above). Four microliters of the diluted sample were loaded on the carbon side of the GO-coated grid, which was mounted on the Mark IV Vitrobot (ThermoFisher) set to 4 °C and 100% humidity. The grid was then incubated for 20 seconds in the Vitrobot chamber to let the proteins adhere to the GO surface of the grid. Then the grid was blotted using standard filter paper (Ted Pella) for 7 seconds with a blot force of 5 before being plunged into a slurry of liquid ethane cooled with liquid nitrogen.

The vitrified samples were imaged using 200 kV Talos Arctica (ThermoFisher) equipped with K3 direct electron detector (Gatan) and BioQuantum energy filter (Gatan). The electron beam was carefully aligned following the established protocol (*47*) or parallel illumination and coma-free alignment. 3,081 micrographs were recorded using SerialEM (*48*) at 63,000X magnification, corresponding to 1.33Å per pixel scale. A three-by-three image shift was used to accelerate the image collection with a nominal defocus from -1 µm to -2.5 µm. The total electron dose was 50 electrons per $1Å^2$ during 3.2 seconds of recording, fractionated into 50 frames.

Cryo-EM image analysis and visualization

Warp (*49*) as used for beam-induced motion correction, CTF estimation, and initial particle picking of the collected movies. The preprocessed micrographs were filtered based on the CTF-estimated resolution, resulting in 2,790 micrographs. These micrographs and corresponding particle stacks were imported to cryoSPARC (50) or downstream image analysis. 2D classification of the particle stack resulted in a subset of particles with a cryo-EM density of both TnsC oligomer and TnsB strand-transfer complex (STC). This set of particles was used to train the Topaz (*51*) neural network, which extracted 993,190 particles. 2D classification of this initial particle stack resulted in 577,628 particles with a cryo-EM density of TnsB STC and TnsC (table S1). These particles

were then subjected to heterogeneous refinement in cryoSPARC using the three initial references of the following: (1) TnsB STC only, (2) two turns of TnsC, and (3) TnsB STC bound to two turns of TnsC, which was generated using ab initio reconstruction. 237,487 particles classified into the third class were then subjected to focused refinement on the TnsC region using a mask covering the TnsC. The resulting particles were then exported to RELION (*52*) or 3D classification focusing on TniQ and adjacent two TnsC subunits. Since the molecular mass of this region was relatively small (~80 kDa), 3D classification was done without image re-alignment. One class (42,210 particles), corresponding to strong density of TniQ, was selected and re-imported into cryoSPARC for another round of 3D classification, focusing on the single subunit of TniQ. This final classification resulted in a final stack of 13,392 particles. This final particle stack was subjected to homogeneous refinement to generate a consensus map. Focused refinements were done through two separate non-uniform refinement jobs focusing on the TniQ-TnsC region or the TnsB STC region of the reconstruction. Local resolution and directional resolution were estimated using blocres (*53*) and 3DFSC (*54*), respectively. For Fig. 1F local-resolution filtered maps from each focused refinement were aligned onto the consensus map and then combined using UCSF Chimera (*55*) command 'vop maximum'. Atomic model was generated through rigid-body docking of the chains from the RNA-guided transpososome structure (PDB: 8EA3) (*23*). Figures describing the cryo-EM reconstruction or atomic model were generated using UCSF ChimeraX (*55*).

E. coli growth time course

E. coli Bl21(DE3) cells were transformed with a plasmid containing combinations of TnsC and TnsB on an inducible promoter (T7) or an empty plasmid and were grown for 16 h in the presence of appropriate antibiotics (50 μg/mL kanamycin and 100 μg/mL carbenicillin). Single colonies were picked for each biological replicate and were inoculated into a primary culture and grown for 16 h with antibiotics. 1:200 dilution of the primary culture was added to sterile 96-well black/clear bottom plates (Thermo Scientific) in a final volume of 200 μL with antibiotics. The culture was grown at 37 °C on a sterile 96-well plate with transparent bottom with continuous shaking, and OD600 was measured for 16 h using a Synergy Neo2 plate reader (BioTek). All readings were blank-corrected. The error bar represents the standard deviation measured across a minimum of two biological replicates. Although TnsC expression in the lag phase exhibited severe toxicity, we noted no observable influence when expression of an N-terminally tagged TnsC was induced in *E. coli* in the mid-log phase. This could be either due to decreased TnsC filamentation by N-terminal tagging or cellular toxicity being less critical to TnsC expression at the mid-log phase (OD ~ 0.5-0.6).

Fluorescence polarization

Fluorescence polarization was performed as reported earlier (*32*). A 55 bp dsDNA labeled with a 5' Fluorescein (6-FAM) tag on one of its strands was used for all fluorescence polarization experiments (table S4). Recombinantly expressed TnsC (0-15 μM) was titrated with the FAM-labeled dsDNA in 1X binding buffer (20 mM HEPES pH 7.5, 2 mM $MgCl_2$, 200 mM NaCl, 10 μM $ZnCl_2$, 1 mM DTT) in the presence of 2 mM ATP or its analogs. The samples were incubated at 37 °C for 30 min in a 384-well plate. The fluorescence gain was adjusted to a sample lacking TnsC while keeping a requested polarization value of 60 mP. Fluorescence polarization was measured in a 384-well black plate using a Synergy Neo2 plate reader (BioTek). The error bar represents the standard deviation measured across two technical replicates. For low salt measurements, the concentration of NaCl was maintained at 50 mM.

For fluorescence polarization measurements for TniQ alone or its interaction with DNA bound to TnsC, TniQ (0-1 μM) was titrated and incubated for 30 min at 37 °C with fluorescently labeled DNA alone or DNA preincubated with 0.3 μM TnsC (30 min at 37 °C). Experiments were conducted in a 1X binding buffer containing 50 mM NaCl in a 25 μL final volume.

***In vitro* ATP hydrolysis assay**

Relative ATP hydrolysis was measured using a Malachite Green Phosphate Assay Kit (Sigma-Aldrich) according to the manufacturer's protocol as previously described (*32*). TnsC (10 μM) alone or TnsC (10 μM) together with TnsB (20 μM) were incubated with 10 μM dsDNA (table S4) in the presence of 1X ATPase reaction buffer (20 mM HEPES pH 7.5, 2 mM $MgCl_2$, 180 mM NaCl, 10 μM $ZnCl_2$, 1 mM DTT) supplemented with 1 mM ATP in a final 5 μL volume for 120 min at 37 °C. The ATP hydrolysis reaction was diluted with 40 μL water and incubated with 10 μL of the working reagent for 5 min. The absorbance at 650 nm was measured in a transparent 384-well plate using a Synergy Neo2 plate reader (BioTek). All samples were blank corrected. The ATPase activity measured was normalized to the sample containing TnsC, TnsB and dsDNA which liberated 174 μM of phosphate. Error bar represents the s.d. measured across two technical replicates (N = 2).

Biochemical transposition assay

Guide RNA was transcribed *in vitro* using HiScribe T7 High Yield RNA Synthesis Kit (NEB) and using a linear dsDNA as a template (table S4) and purified by MEGAclear Transcription Clean-Up Kit (Invitrogen). Prior to setting the transposition reaction, the ribonucleoprotein (RNP) complex composed of Cas12k (0.75 μM) and guide RNA (sgRNA-6, 9 μM) was subjected to pre-incubation at R.T. for 5 min in a total volume of 1.4 μL. *In vitro* integration was performed in the presence of RNP (50 nM), TniQ (100 nM), TnsC (100 nM), TnsB (1 μM), pTarget (0.6 nM) and pDonor (1.75 nM) supplemented with ATP (2 mM), BSA (50 μg/mL), SUPERase•In Rnase

Inhibitor (Invitrogen) (0.25 U/µL) in a 1X IVI Buffer (25 mM HEPES pH 7.5, 5 mM Tris pH 8.0, 0.05 mM EDTA, 20 mM $MgCl_2$, 30 mM NaCl, 20 mM KCl, 1 mM DTT, 5% (v/v) glycerol). The reaction, having a final volume of 20 µL, was incubated at 37 °C for 2 h, quenched by heat denaturing at 95 °C for 3 min and flash frozen until library preparation for NGS. Transposition was performed in the presence of Mg^{2+} ion concentration of 2 mM throughout the reaction, in contrast to the conditions mentioned earlier (*11*, *56*). We noted that pre-incubation with low Mg^{2+} at 30 °C is not necessary for the biochemical reconstitution of CAST transposition. For titration involving individual proteins, the concentration was varied as indicated (Fig. 2, fig. S5), keeping all other components constant. S15 was supplemented for reactions involving the same. The products of *in vitro* transposition are Shapiro intermediates as opposed to resolved co-integrates in cellular experiments. The transposon-target DNA junction corresponding to the Shapiro intermediate was amplified by PCR and visualized on 1.5% agarose. qPCR, or short-read sequencing was also performed on the Shapiro intermediate. Due to linear amplification in the first cycle of PCR for the Shapiro intermediate compared to the reference used in qPCR, the efficiencies measured are only relative for biochemical experiments. Without S15, we observed a relative on-target efficiency of ~10% detected by qPCR whereas supplementing S15 resulted in 33% efficiency.

qPCR analysis for *in vitro* and genomic transpositions

qPCR for detecting on-target integration in the genome was performed with two sets of primers. A combination of – 1) genome-specific (targets for sgRNA 1-5) and left transposon-specific primers probing t-LR and 2) genome-specific primers for detecting an *E. coli* reference gene *rssA*. Primer pairs were designed to amplify products between 100 and 250 bp and showed amplification efficiency between 100 and 110%. Probes for Taqman qPCR were designed with a 5' FAM-label ZEN/3' IBFQ (IDT) as a universal probe which annealed to the transposon left-end and can be displaced only upon amplification. The control probe was designed to bind to *E. coli rssA* gene and contained a 5' SUN-label ZEN/3' IBFQ (IDT). qPCR for in vitro integration was detected using SYBR Green. qPCR for genomic samples was performed using 5 ng of purified genomic DNA in the presence of 1 µL each of 18 µM forward primer and reverse primer pairs (table S4), 5 µL of TaqMan Fast Advanced Master Mix (Thermo Fischer Scientific), 0.5 µL of each 5 µM probe (table S4) and 1 µL of water in a final volume of 10 µL. On-target integration efficiency is measured as 100 X (2^ΔCq), where ΔCq is the Cq (control rssA) – Cq (integration).

qPCR for detecting the Shapiro intermediate resulting from on-target integration was performed with two sets of primers. A combination of – 1) pTarget and left transposon-specific primers probing t-LR and 2) primers for detecting pTarget. qPCR measurements for in vitro integration were performed with 2 µL of the 20 µL *in vitro* integration reaction, 1 µL each of 10 µM forward and reverse primer (table S4), 5 µL of SsoAdvanced Universal SYBR Green 2X

Supermix (BioRad), and 2 µL of water in a final volume of 10 µL. The first cycle of qPCR only results in linear amplification in comparison to the control targeting pTarget and therefore we note that efficiency measurements are relative. On-target integration efficiency is measured as 100 X (2^ΔCq), where ΔCq is the Cq (pTarget) – Cq (integration).

All qPCR reactions were performed on 384-well clear/white plates (BioRad) on a CFX384 Real-Time PCR Detection System (BioRad) using the following thermal cycling conditions: DNA denaturing (DNA denaturation 98 °C for 2.5 min), 40 cycles of amplification (98 °C for 10 s, 62 °C for 20 s), and terminal melt-curve analysis was performed (65–95 °C in 0.5 °C per 5 s increments). All qPCR primers used in this study are mentioned in table S4.

TagTn-seq library preparation and sequencing

In order to achieve high sequencing coverage across genomic and plasmid targets, samples from biochemical and *E. coli* transposition assays were treated with AvrII (NEB) to reduce the contaminating pDonor. A single AvrII cut site was located 21-nt from the left transposon-end in pDonor whereas pTarget had no sites. The *E. coli* BL21-DE3 genome contains only 16 AvrII cut sites. For samples from biochemical experiments, 1.5X µL of Mag-Bind TotalPure NGS magnetic beads (Omega) were added to each sample and the DNA was purified using the manufacturer's protocol. The sample was eluted in 10 µL volume and digested with AvrII (NEB) (5 U) in rCutSmart Buffer (NEB) for 1 h at 37 °C. 1.5X magnetic beads were added to each sample and DNA was purified and eluted in a 10 µL volume. For genomic samples, 1 µg of isolated genomic DNA (having contaminating pDonor) was digested with AvrII (5 U) in a 50 µL volume for 2 h at 37 °C. Total genomic DNA was purified with magnetic beads, eluted in 10 µL volume and quantified using Qubit dsDNA High Sensitivity Kit (Invitrogen) and was used for tagmentation using Nextera XT DNA Library Preparation Kit (Illumina).

4 ng of DNA from either biochemical or *E. coli* transposition assays were mixed with 5 µL of tagmentation DNA buffer (Illumina) and 1 µL of amplicon tagmentation mix (Illumina) in a final volume of 10 µL and incubated at 55°C for 7 min. The tagmentation reaction was quenched and DNA was amplified by PCR-1 step, using CAST left transposon-end specific primer (Y3 mix, 0.42 µM) having a universal TrueSeq adaptor (i5) overhang, Nextera adaptor (i7) specific primer (0.42 µM) (table S4) and 30 µL KAPA HiFi HotStart ReadyMix in a 60 µL final volume. Y3 mix is a combination of three primers with variable length between CAST transposon-end specific region and universal TrueSeq adaptor to introduce diversity during Illumina flow cell clustering (table S4). After 20 cycles of amplification at an annealing temperature of 57 °C, the amplified DNA was purified using 1.5X of magnetic beads and eluted in 10 µL volume. This DNA was next subjected to PCR-2 using TrueSeq (i5) specific primers (0.42 µM) having an indexed universal p5 overhang, Nextera adaptor (i7) specific primer (0.42 µM) having an indexed universal p7 overhang, and 30 µL KAPA HiFi HotStart ReadyMix in a 60 µL final volume. After 13 cycles of

amplification at an annealing temperature of 54 °C, the barcoded amplicon was purified by magnetic beads and eluted in 10 µL volume. The DNA was resolved by 1.5 % agarose followed by excision and isolation within the size range of 300-600 bp using Gel Extraction Kit (Qiagen). NGS libraries were pooled and quantified by qPCR using NEBNext Library Quant Kit (NEB). Sequencing was performed on Illumina NextSeq platform with NextSeq high-output kit. 75/150-cycle single-end reads were obtained and automated adaptor trimming and demultiplexing (Illumina) was performed.

Analysis of NGS data

Analysis of TagTn-seq data was performed using a custom Python pipeline as described (*14*). Tag-Tn statistics are provided in table S6. Demultiplexed raw reads where half of the bases had a Phred quality score of less than 20, which corresponds to greater than 1% base miscalling were removed from the analysis. Reads containing the last 23 bases of the transposon-end sequence (5' GACAGATAATTTGTCACTGTACA 3') and a 22-bp adjacent flanking genomic sequences were extracted and noted as the total-transposon-end containing reads. This 22-bp adjacent region represents the fingerprint region used to identify transposon insertions and was used to align to the reference genome or plasmid using Bowtie2 (*57*). The reference genome for *E. coli* BL21(DE3) was based on published data from the National Center for Biotechnology Information (NCBI) genomes. Only reads which mapped perfectly and only once to the genome were chosen. Reads which did not get mapped to the genome were checked for sequences corresponding to the pDonor and noted as pDonor contamination. Alignments from Bowtie2 were used for generating genome/plasmid-wide coordinates for integration. If the read was mapped to the same strand as the input fasta file, then it was noted as on 'fwd' strand, and if mapped to the complementary strand as the input fasta file, then it is noted as on 'rev' strand. The read 'position' was indexed at the fifth position of target site duplication (TSD) for each event, with respect to the 'fwd strand'. The orientation of the integration relative to the fasta file was concluded based on whether the library was sequenced from the right end or the left end of the transposon for TagTn-sequencing. The orientation of the transposon insertion with respect to the protospacer at the on-target window (100 bp from the end of protospacer), was noted as target-left-right (t-LR) or target-right-left (tRL).

Untargeted reads and on-target reads were assigned using a custom Python script. For biochemical integration, events that mapped within a hundred base window after the end of the protospacer on pTarget were presumed as on-target events, whereas any other integration events on pDonor and pTarget were totaled and noted as untargeted. Reads due to contaminating pDonor and a potential PCR recombination artifact (at position 1198 and 2328 nt) were masked for all samples. For *E. coli* integration, reads that mapped within a hundred base window after the end of the protospacer were noted as on-targets, and reads elsewhere in the genome as untargeted. Raw reads for integration at on-target and untargeted sites for each sample were normalized with the

total transposon-end containing reads (including pDonor contamination) detected for that sample and further scaled with respect to the sample having the highest transposon-end containing reads (in the experimental set). In order to minimize amplification biases during NGS sequencing only unique insertion sites were used for reporting sequence preferences in transposition. For visualizing normalized plots comparing integration across the genome, raw reads for each coordinate for a sample were normalized and scaled in the same way as above, converted to bigwig files and visualized in IGV.

For the plot comparing untargeted integration sites across pTarget, reads at the on-target window were excluded. A continuous single-base position for the pTarget and the reads detected at that position was generated using a custom Python script. The read at each of these bases were compared between samples. Pearson correlation coefficients (r) between data sets were calculated in Prism 9.

Analysis of genetic neighborhood for integration events detected across the *E. coli* genome

A custom Python script was used to analyze whether the integration events detected elsewhere in the genome are RNA-dependent off-targets. The position of insertion determined by NGS analysis was used for the genetic neighborhood analysis. Events at the on-target window were removed from the analysis. For every integration site detected, the most likely RNA-dependent off-target was determined as follows. Off-target was assumed to be in the target-left-right orientation similar to an on-target event. A 7 bp window (3 bp upstream and downstream) of sequence, 63 bp upstream from the integration site in the LR orientation was checked for the enrichment of a PAM sequence (GTN). Similarly, N = 10,000 random regions in the *E. coli* genome were chosen, and the corresponding window upstream was checked to get the proportion of random sequences having a potential PAM.

For integration events, PAM-containing sequences were further checked for the presence of matches in the protospacer for sgRNA 1-5 respectively (23 bp downstream from the potential PAM detected). The maximum matching sequence detected was selected as the potential number of spacer matches representing that insertion event. Similarly, the random region of *E. coli* genome sampled previously with a PAM-containing sequence was checked for matches in protospacer for sgRNA 1-5 respectively. In both cases, when more than one potential protospacer had a valid PAM, the maximum number of matches to the spacer was determined as the potential number of spacer matches for that location.

AT-enrichment analysis

AT-enrichment analysis was performed with a custom Python script. The position of insertion determined by NGS analysis was used for analyzing the AT content. Events at the on-target location were removed from the analysis if a spacer was provided. For every unique insertion site

on pTarget, λ-DNA or the *E. coli* genome, a window of sequence either upstream or downstream was chosen (50 bp for pTarget or λ-DNA and 100 bp window for the *E. coli* genome). The AT content in both these regions (upstream/downstream) was calculated and the highest of the two was selected as the representative AT content of that unique insertion site. Unique events were binned across their AT content to create a distribution of events detected at each AT bin. As a control to calculate the AT content of pTarget, λ-DNA or the *E. coli* genome, the regions on the DNA were randomly sampled (N = 50,000) and similarly, the highest AT content of the adjacent bin to every unique sampled location was chosen as representative of that sampling event. The randomly sampled events were also binned across their AT content to create a distribution of random samplings events (counts) detected at each AT bin. A cumulative frequency distribution plot was used to compare the difference in AT content for unique integration events and randomly sampled regions on the pTarget, λ-DNA or the *E. coli* genome. The AT content of unique integration events and random sampling events were used to perform a Mann-Whitney U test which was used to test the significance of the distributions. pDonor was not used for AT-enrichment analysis due to confunding factors such as target immunity (*58*).

To generate plots to visualize the AT content across the pTarget and integration, the pTarget was divided into 59 bins corresponding to 46 bp and the AT percentage was calculated using a custom Python script. The reads detected on pTarget was plotted as an overlay with the AT content detected at that bin. For λ-DNA, the genome was divided into 45 bins of 1078 bp window and AT percentages were calculated and plotted similarly. Correlation between reads detected for integration and fluorescence intensity of mNG-TnsC binding across λ-DNA (from DNA curtains) were calculated by similarly binning the reads and fluorescence intensity detected across λ-DNA into 45 bins of 1078 bp window and Spearman correlation coefficient (r) was calculated using Prism 9.

Essential gene analysis

Essential gene analysis (fig. S6L) was performed with a custom Python script. The position of insertion determined by NGS analysis was used to determine if events landed on an essential gene or not. Events at the on-target location and at the T7RNAP locus were removed from the analysis. The remaining reads for these events were annotated with their CDS features in accordance with the NCBI-published genome. Essential genes were noted based on previous reports on *E. coli* K-12 and reads falling to these regions were classified as essential gene insertions (*59*). Reads landing outside these regions were noted as non-essential insertions. The percentage of the *E. coli* genome which was essential was calculated by summing the length of all essential genes and dividing by the genome length.

Sequence logo for untargeted events

To create sequence logos for untargeted events, we took every unique site of integration in the *E. coli* genome, and a window upstream and downstream of the position of integration on both the 'fwd' and 'rev' strand. Insertions in the on-target window were omitted for samples having Cas12k and a targeting guide RNA. Further, TSD correction was accounted on the strands, and insertions were oriented in the left-to-right orientation based on the strand of the input fasta file and the strand to which each read was mapped. The extracted sequence for both the strands was outputted and WebLogo v2.8.2 was used for plotting the sequence logo showing per residue conservation.

For ShCAST, a 75-base window on either side of the integration was chosen for building the sequence logo. The sequence logo was indexed with the third position of TSD set as position '0' and 70 bases on either side were plotted. WebLogo (*60*) in Fig. 3H was plotted for insertions on both the strands observed for two biological replicates of a sample having a pHelper lacking Cas12k and guide RNA. Sequence features for samples with Cas12k and guide RNA (sgRNA-1) was performed using previously reported data for ShCAST (*14*) generated by random fragmentation and insertion site amplified from the transposon right-end (fig. S7A). The plots revealed consistent sequence featured compared to samples subjected to TagTn-seq and sequencing from the transposon left end. No conservation of sequences with partial guide complementarity was noted upstream, consistent with the analysis (fig. S1C-E) that the majority of events in type V-K CASTs are RNA-independent.

For ShoCAST, previously reported data (*14*) with a pHelper having ShoCas12k and guide RNA (sgRNA-7) was used and analyzed to inform the position of each integration event. WebLogo (*60*) in fig. S7C was plotted for unique insertions. A 75-base window was chosen for building the sequence logo. The sequence logo was indexed with the third position of TSD set as position '0' and 70 bases on either side were plotted.

DNA curtains

Single-molecule double-tethered dsDNA curtain experiments were carried out as previously described (*35*). Custom flow cells were assembled using quartz slides, on which chromium patterns were deposited through nanofabrication (*61*). The dsDNA substrate was prepared by annealing and ligating two custom oligonucleotide handles (IDT) at the COS sites of bacteriophage λ-DNA (NEB), such that one end of the DNA contained a biotin modification while the other a digoxigenin (dig). Prior to assembly of DNA curtains, flow cells were first passivated with a lipid bilayer. The biotin ends of the DNA were tethered through a biotin-streptavidin linkage to biotin-modified lipids within the bilayer. These DNA molecules were then aligned as the chromium barriers and flow-stretched so that the dig-tagged ends were anchored to chromium pedestals coated with anti-dig antibody (Roche). These DNA molecules were then flow-stretched so that the dig-ends were anchored at chromium pedestals coated with anti-dig antibody (Roche). A pre-defined barrier-

pedestal distance of 12 μm and orthogonal attachment chemistry yielded uniform and unambiguous orientation of the double-tethered dsDNA curtain. Fluorescent mNeonGreen-TnsC (mNG-TnsC) was injected into the flow cell, from a 50 μL sample loop, using a syringe pump. Sample illumination was achieved using a continuous-wave 488 nm laser (Coherent Sapphire, 200 mW), shuttered externally (Uniblitz LS6). The emission signal was visualized through a custom-built prism-type total internal reflection fluorescence (TIRF) microscopy system, based on a Nikon Eclipse TE2000-U, equipped with a 60X water immersion objective and an EMCCD camera (Andor iXon X3). All single-molecule experiments were performed in reaction buffer (20 mM HEPES pH 7.5, 2 mM $MgCl_2$, 200 mM NaCl, 10 μM $ZnCl_2$, 1 mM DTT, 0.2 mg/mL BSA, and 2 mM ATP), at room temperature (25 °C). In equilibrium binding experiments, flow was stopped after mNG-TnsC (100 nM) had reached the flow cell and data were collected at the rate of 1 frame per 30 seconds for 30 min. mNG-TnsC was injected at a concentration of 500 nM for the experiment shown in fig. S6C and movie S2.

Data analysis for DNA curtains

Files saved in ND2 format were first converted to TIFF stacks in FIJI (*62*), where all subsequent image processing steps took place. For binding and disassembly experiments, a kymograph was first generated for each DNA molecule by making a single-pixel wide slice, drawn along the length of the DNA, through the time series. Fluorescence intensity over time was then extracted by plotting the profile of the kymograph of interest in FIJI, which produced intensity values, each averaged over the entire one-pixel-wide DNA, over time. Intensity time series from >50 individual filaments were combined, normalized, and plotted as mean ± s.d. over time (Fig. 3C and 3D). Apparent rates (k_{obs}; Fig. 3D) were obtained by nonlinear regression to a one-phase association model in Prism 10. Statistical significance of the difference in apparent rates (k_{obs}) measured on AT- and GC-rich regions was further tested by resampling (bootstrapping). Original data set was randomly divided into three subsets of equal size. Each subset was then fitted to a one phase association model to obtain k_{obs} for AT- and GC-rich regions. The triplicates of values for k_{obs} were compared between AT- and GC-rich binding, using the unpaired Student's t-test in Prism 10. For intensity versus binding position (Fig. 3C), the last 10 frames (5 min) of a 30-minute binding experiment were averaged to remove intensity fluctuations in any given frame due to random DNA fluctuations within the evanescent field. For each DNA of interest, a signal intensity profile was plotted over a 33-pixel long region, centered over the 45-pixel total length of the DNA (12 μm), to reduce signal contamination and bleed-over from nonspecific binding of mNG-TnsC to chromium features. Intensity profiles from > 60 individual DNA molecules were combined and plotted as mean ± standard deviation. The A/T percentage of the λ-sequence was obtained by first segregating the entire sequence (48,502 bp) into 45 bins of 1078 bp width, then calculating the A/T percentage for the sequence within each bin. The number of bins was chosen to reflect the

fact that λ-DNA double tethered at 12 μm end-to-end distance spanned 45 pixels in the camera of this imaging platform. Thus, fluorescence intensity measured at any pixel along the entire filament should represent the amount of mNG-TnsC bound to the DNA within the window of 1078 bp.

Western blot

An N-terminal Flag epitope tag was cloned into TnsC expressed with a T7 or lac promoter. Cells were grown at the same conditions as the *E. coli* transposition assay in the presence of 0.1 mM IPTG. 2 X 10^7 were incubated with 5 μL of 6X SDS loading dye and boiled for 10 min at 95 °C. The samples were resolved on Mini-PROTEAN TGX Stain-Free Precast Gels and imaged to confirm equal loading. The gel was transferred using iBlot 2 Gel Transfer Device using the manufacturer's protocol. The transferred blots were washed twice with wash buffer (1% PBS, 0.1% Tween) for 5 min each and then washed with blocking buffer (1% PBS, 0.1% Tween, 5% BSA) for another 5 min. The blot was then incubated for 30 min in 70 mL blocking buffer at RT with gentle shaking. After 30 min, the blot was again washed with block buffer and incubated for 1h with either Monoclonal ANTI-FLAG M2 (Sigma) (1:7500 dilution) in 15 mL volume or GAPDH Loading Control Monoclonal Antibody (Thermo Scientific) (1:2500 dilution) in 5 mL volume in blocking buffer with gentle shaking. The same wash protocol was repeated and incubated with a secondary HRP conjugated antibody, Goat Anti-Mouse IgG1 HRP (Abcam) (1:15000 dilution) in 15 mL volume and incubated for 2 h. The blot was washed twice in the wash buffer for 5 min each and imaged using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) after incubating for 2 min.

Statistics and reproducibility. All *E. coli* transposition assays and NGS were performed in two independent biological replicates. qPCR for the same was performed in two independent replicates. *E.* coli growth curves were measure in minimum of two independent replicates. Titration for *in vitro* transposition were performed once for each concentration. Fluorescence polarization and ATP hydrolysis assays were performed in two technical replicates.

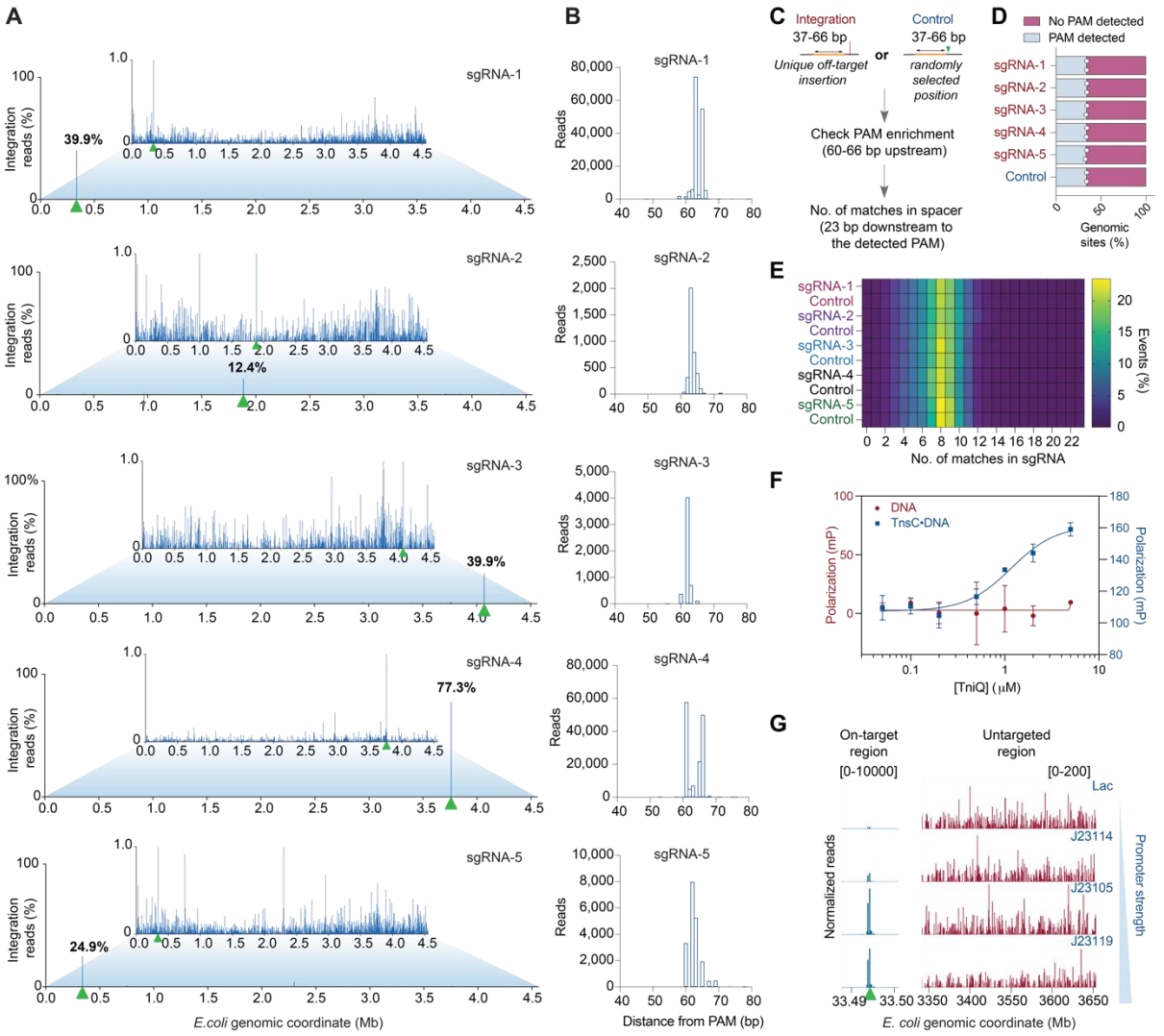

Fig. S1. Type V-K CASTs maintain a distinct, RNA-independent pathway. (**A**) Genome-wide view of *E. coli* genome-mapping reads for the original/WT ShCAST system encoding the indicated sgRNAs; the zoomed-in views visualize reads comprising 1% or less of genome-mapping reads. Target sites are marked with a green triangle. (**B**) Integration site distributions determined from the NGS data, plotted as the distance from the PAM sequence to the first transposon bp. (**C**) Schematic depicting the neighborhood analysis used to explore PAM (5'-GTN-3') and sgRNA complementarity enrichment flanking off-target ShCAST integration events. Insertions were presumed to occur in the (target-left-right) T-LR (*14*), and a 60-66 bp window upstream of the insertion site was analyzed. For sites with a cognate PAM, the adjacent 23 bp were analyzed for complementarity to the sgRNA spacer sequence. Controls analyses used random samplings of the *E. coli* genome. (**D**) Off-target ShCAST integration events from five distinct sgRNAs are not enriched in cognate PAMs, relative to a randomly sampled control dataset (Materials and Methods). Percentage of integration events (for $N = 2$ biological replicates) detected with or without a PAM when compared to random regions of the *E. coli* genome. (**E**) Off-target ShCAST integration events from five distinct sgRNAs are not enriched in the number of complementary matches to the sgRNA, relative to a randomly sampled control datasets (Materials and Methods). (**F**) TniQ selectively binds TnsC-DNA filaments, but not naked DNA, as observed using

fluorescence polarization experiments with a 55-bp FAM-labeled DNA substrate. Data shown represent mean ± s.d for $N = 2$ technical replicates. (**G**) Normalized integration reads detected at the on-target site (green arrow, left) and a representative untargeted site (right), with varying Cas12k promoter strength. Note the differing y-axis ranges.

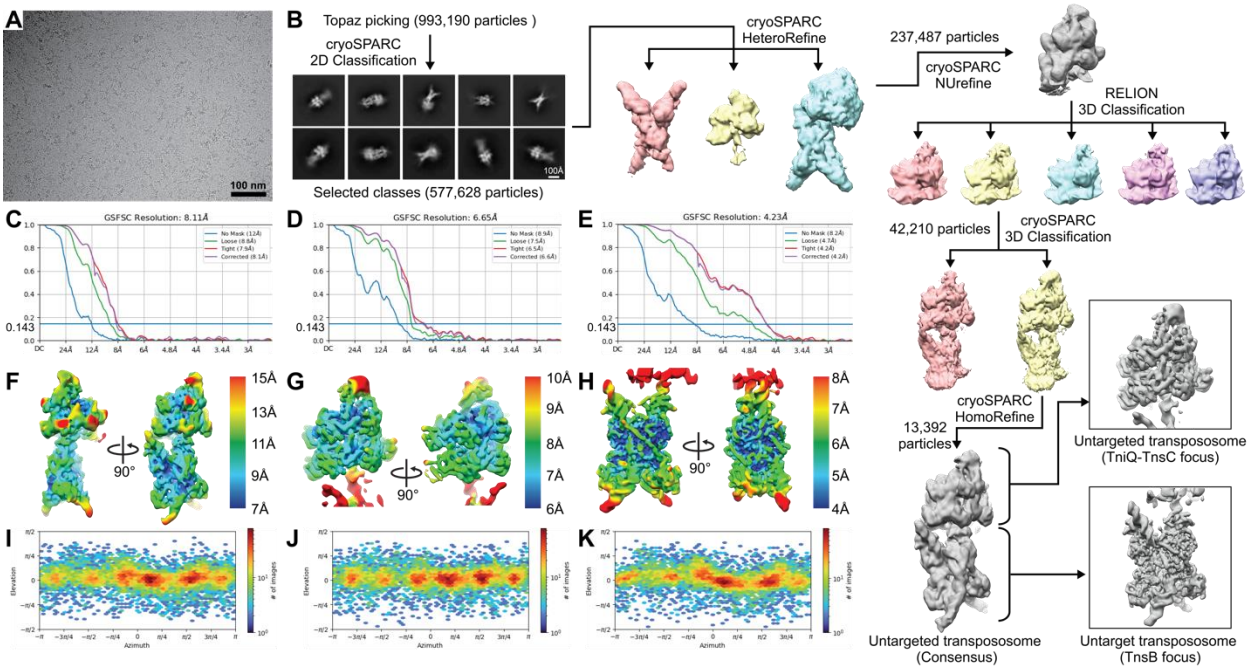

Fig. S2. Cryo-EM image analysis of the untargeted, RNA-independent BCQ transpososome. (**A**) Representative cryo-EM micrograph of the BCQ transpososome sample, generated by incubating TnsB, TnsC, TniQ with target-LE and target-RE DNA substrates. The black bar represents 100 nm. (**B**), Computational workflow used for analyzing the BCQ transpososome cryo-EM dataset. Topaz (*51*)-extracted particles were first subjected to 2D classification in cryoSPARC (*50*). 2D classes with cryo-EM density for the TnsB strand-transfer complex (STC) and TnsC oligomer were selected for downstream heterogeneous refinement in cryoSPARC. Particles classified as the TnsB-TnsC complex were then subjected to non-uniform refinement (NUrefine) in cryoSPARC using a mask that covers the TniQ-TnsC region. The aligned particle stacks were then further classified using 3D classification in RELION (*52*), focusing on TniQ and the adjacent two subunits of TnsC. One class showing the best resolved TniQ density was selected for another round of 3D classification in cryoSPARC focusing on TniQ. The final particle stack with stronger TniQ density was then subjected to homogeneous refinement. Focused refinement was done on the TniQ-TnsC complex or the TnsB-STC region of the reconstruction. See Materials and Methods for details. (**C**)-(**E**), Gold-standard Fourier shell correlation (GSFSC) curves from the consensus refinement (**C**) and the focused refinements of the TniQ-TnsC region (**D**) or the TnsB-STC region (**E**) of the reconstruction. The 0.143 GSFSC cutoff is indicated as a blue horizontal line. (**F**)-(**H**), Estimated local resolution from the consensus refinement (**F**) and the focused refinements of the TniQ-TnsC region (**G**) or the TnsB-STC region (**H**). Local-resolution filtered reconstructions were colored based on the estimated local resolution, as indicated in the legend. (**I**)-(**K**), Angular distribution heatmap of the particle stack in the final consensus refinement (**I**) and focused refinements of the TniQ-TnsC region (**J**) and the TnsB-STC region (**K**). Colors indicate particle count, as indicated in the legend.

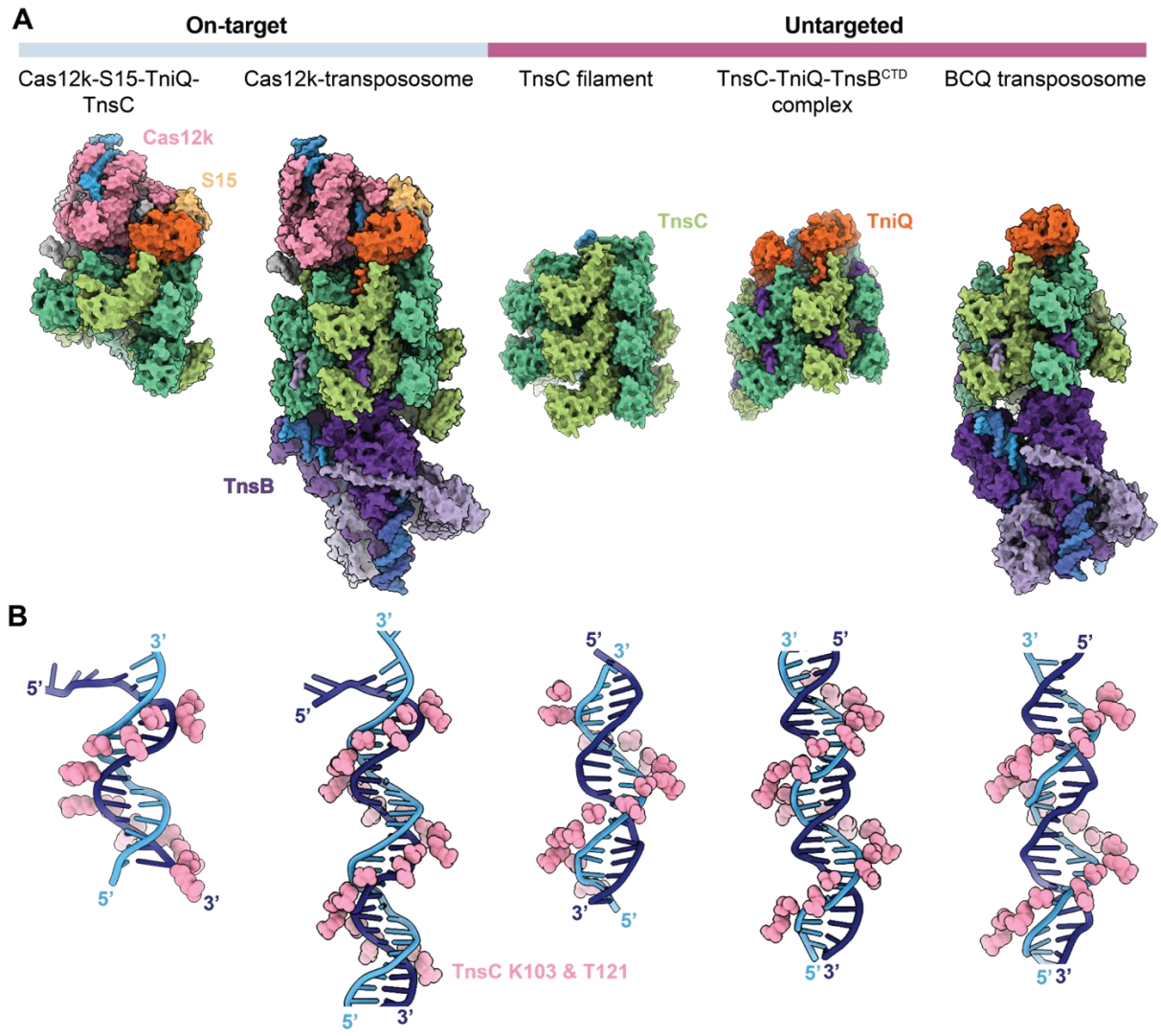

Fig. S3. TnsC adopts a similar overall architecture across various partial and complete transpososome complexes. (**A**) Structures of TnsC-associated complexes from ShCAST in various configurations. The structures being compared are, from left to right: Cas12k-S15-TniQ-TnsC recruitment complex (PDB: 8BD5) (*22*), Cas12k-transpososome (PDB: 8EA3) (*23*), TnsC filament (PDB: 7M99) (*26*), TnsC-TniQ-TnsB^{CTD} complex (PDB: 7SVU) (*23*), and BCQ transpososome (this study). Components associated with each structure include Cas12k (pink), S15 (tan), TnsC (green), TniQ (orange), and TnsB (purple). (**B**) Interactions between TnsC and DNA from each of the structures shown in **A**. TnsC residues K103 and T121 (pink) are highlighted alongside both DNA strands (blue).

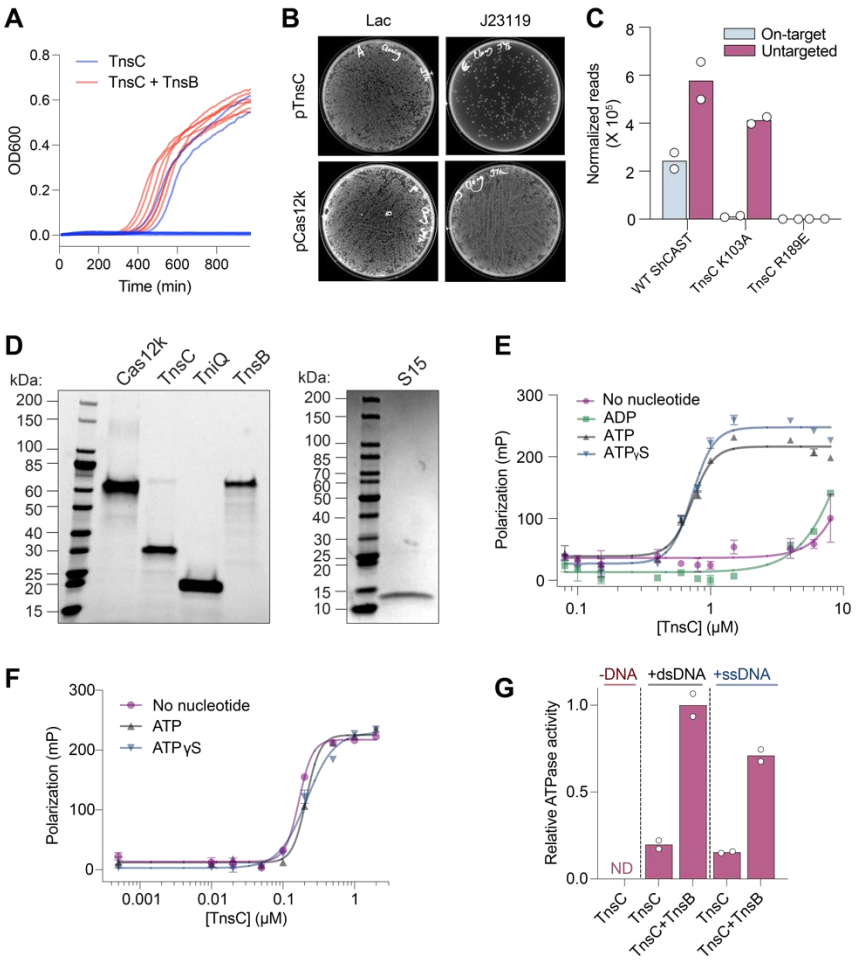

Fig. S4. Supporting experimental data for ShCAST biochemical reconstitution. (**A**) *E. coli* growth curves of seven biological replicate cultures of a TnsC over-expression strain under inducible conditions, with or without TnsB. The majority of cultures without TnsB fail to grow. **(B)** Pictures of transformation plates after attempts to clone TnsC or Cas12k expression plasmids with a Lac (weak) or J23119 (strong) promoter, providing extra evidence that strong TnsC over-expression leads to cellular toxicity. The same competent cells and input DNA amounts were used in all cases, and surviving colonies with a strong promoter driving TnsC frequently exhibited mutations in the promoter or ORF. (**C**) Total genome-mapping reads detected for the indicated expression construct, normalized, and scaled (Materials and Methods); the mean is shown from *N* = 2 independent biological replicates. (**D**) SDS-PAGE analysis of purified ShCAST components used in biochemical transposition assays. (**E**) DNA binding by TnsC is ATP dependent under 200 mM NaCl conditions, as observed using fluorescence polarization experiments with a 55-bp FAM-labeled DNA substrate. Data shown represent mean ± s.d for *N* = 3 technical replicates. (**F**) Under low-salt conditions (50 mM NaCl), DNA binding by TnsC is nucleotide-independent. Data are shown as in (**E**). (**G**) TnsB stimulates the ATP hydrolysis activity of TnsC in a DNA-dependent reaction, as determined using a Malachite Green assay. The mean is shown from *N* = 2 technical replicates; ND, not detected.

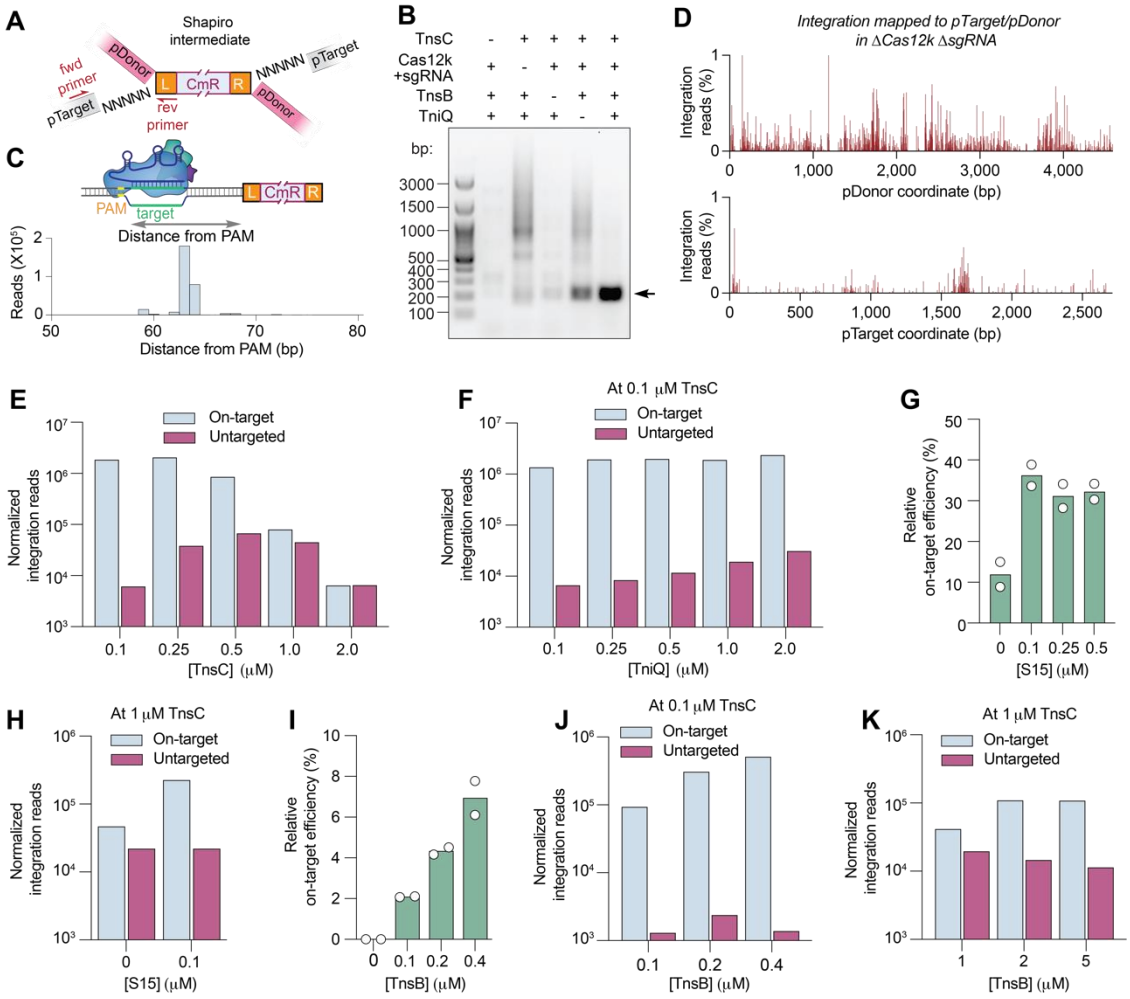

Fig. S5. The relative stoichiometry of CAST components controls pathway choice. (A) Biochemical transposition assays were performed *in vitro* with purified ShCAST components, followed by PCR-based amplification of the junction between pTarget and the transposon left end (LE) using the indicated primers (red). Note that ShCAST generates Shapiro intermediate products in these experiments (*18*); NNNNN denotes the single-stranded target-site duplication (TSD) region. (**B**) Transposition products were analyzed by PCR amplification and 1.5% agarose gel electrophoresis for the indicated conditions. The on-target product (black arrow) is only efficiently generated in the presence of all ShCAST components, though TniQ is partially dispensable for *in vitro* transposition under these conditions, as previously reported (*11*). (**C)** Integration site distribution for the on-target products in (**B**) determined by high-throughput amplicon sequencing. (**D**) Zoomed-in view of integration reads comprising 1% or less of pDonor- (top) and pTarget- (bottom) mapping reads, in a biochemical transposition assay performed without Cas12k or sgRNA. **e,** Normalized integration reads detected at on-target and untargeted sites from biochemical transposition assays, at varying TnsC concentrations. (**F**) Normalized integration reads detected at on-target and untargeted sites from biochemical transposition assays, with 0.1 μM TnsC and varying TniQ concentrations. (**G**) On-target transposition efficiency from biochemical transposition assays as detected by qPCR, with 0.1 μM TnsC and varying S15

concentrations. (**H**) Normalized integration reads detected at on-target and untargeted sites from biochemical transposition assays with 1 μM TnsC, with or without S15. (**I**) On-target transposition efficiency from biochemical transposition assays as detected by qPCR, with 0.1 μM TnsC and varying TnsB concentrations. (**J**) Normalized integration reads detected at on-target and untargeted sites from biochemical transposition assays, with 0.1 μM TnsC and varying TnsB concentrations. (**K**) Normalized integration reads detected at on-target and untargeted sites from biochemical transposition assays, with 1 μM TnsC and varying TnsB concentrations.

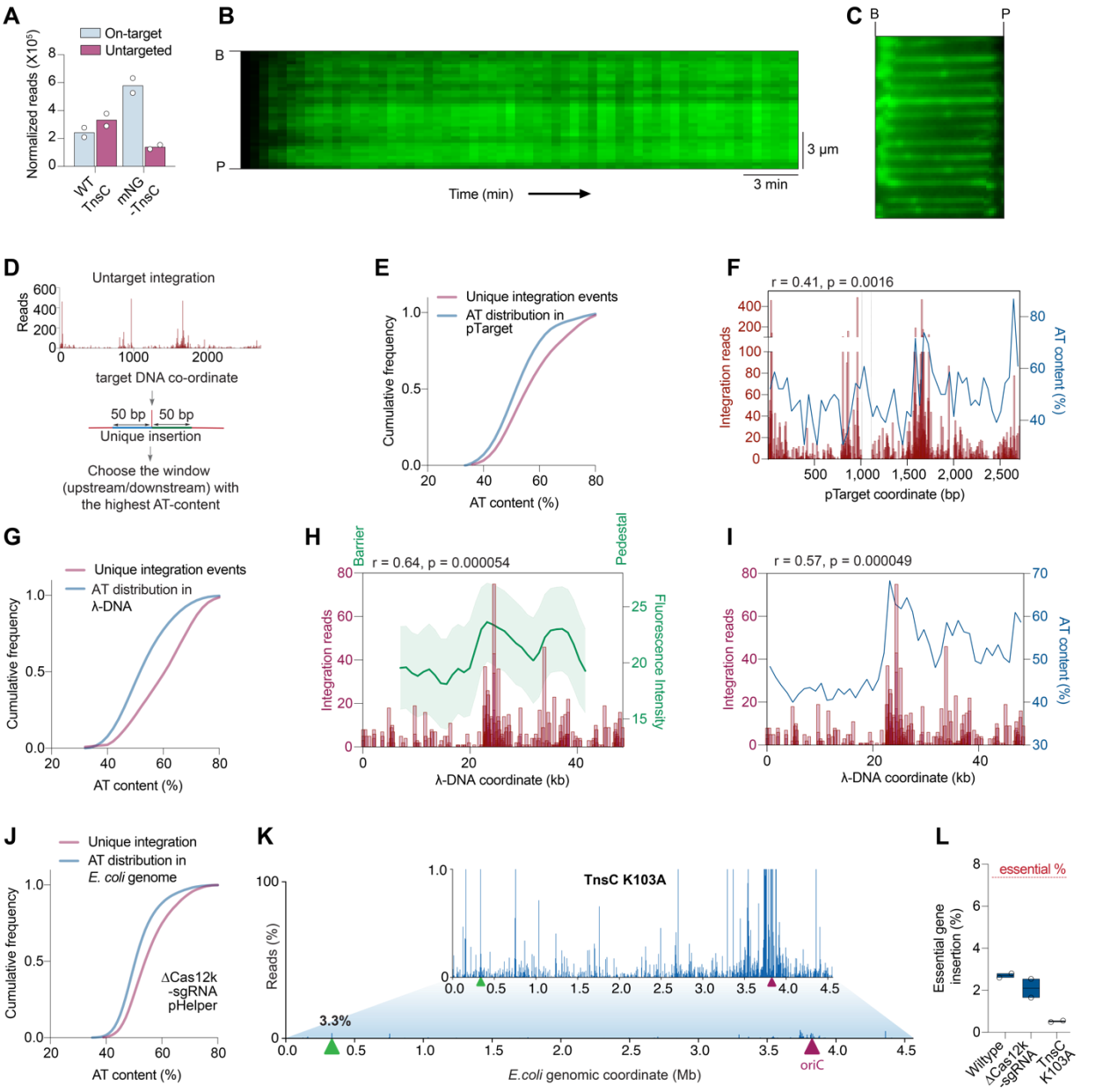

Fig. S6. TnsC binding and TnsC-mediated integration exhibit preferential AT bias. (**A**) Normalized integration reads detected at on-target and untargeted sites from cellular transposition assays, with either WT TnsC or an mNeonGreen (mNG)-TnsC fusion. (**B**) Representative kymograph from a DNA curtains assay with 100 nM mNG-TnsC, showing the time-dependent evolution λ-DNA binding. B, barrier; P, pedestal. (**C**) λ-DNA was completely coated with TnsC at high (500 nM) concentrations, as exemplified by a representative DNA curtains image (Movie S2). (**D**) Analytical workflow to investigate the AT content surrounding untargeted integration events in transposition assays. Two windows (50-bp for biochemical or 100-bp for genomic integration) flanking each unique event were analyzed, and the highest AT content was retained for subsequent analyses. A similar workflow was performed for random sampling of the same DNA substrate. (**E**) Cumulative frequency distributions for the AT content within a 50-bp window flanking unique integration events on pTarget (in biochemical integration assay), using ShCAST

with WT TnsC and sgRNA-6 (N = 504 unique integration events), compared to random sampling of pTarget (N = 50,000 counts; Materials and Methods). The distributions were significantly different, based on results of a Mann-Whitney U test ($P = 3.64 \times 10^{-11}$). (**F**), Graph comparing the distribution of AT content (46-bp bins) and integration reads on pTarget, with indicated Spearman correlation coefficient and results from a two-tailed significance test. (**G**)**,** Cumulative frequency distributions for the AT content within a 50-bp window flanking unique integration events on λ-DNA, using ShCAST with WT TnsC and sgRNA-6 (N = 190 unique integration events), compared to random sampling of λ-DNA (N = 50,000 counts; Materials and Methods). The distributions were significantly different, based on results of a Mann-Whitney U test ($P = 4.77 \times 10^{-11}$). (**H**), Graph comparing the distribution of mNG-TnsC fluorescence intensity (1,078-bp bins) and integration reads on λ-DNA, with indicated Spearman correlation coefficient and results from a two-tailed significance test. (**I**) Graph comparing the distribution of AT content (1,078-bp bins) and integration reads on λ-DNA, with indicated Spearman correlation coefficient and results from a two-tailed significance test. (**J**) Cumulative frequency distributions for the AT content within a 100-bp window flanking unique integration events in the *E. coli* genome, using ShCAST in the absence of Cas12k and sgRNA (N = 3,103 unique integration events), compared to random sampling of the *E. coli* genome (N = 50,000 counts; Materials and Methods). The distributions were significantly different, based on results of a Mann-Whitney U test ($P = 4.40 \text{ X } 10^{-83}$). (**K**) Genome-wide view of *E. coli* genome-mapping reads for ShCAST with a K103A TnsC mutation; the zoomed-in view visualizes reads comprising 1% or less of genome-mapping reads. The target site is marked with a green triangle, and the *E. coli* oriC is marked with a maroon triangle. (**L**) Untargeted genomic integration events are depleted in essential genes. The observed percentage of integration reads occurring within essential genes was quantified and plotted for each of the indicated datasets, from N = 2 independent biological replicates. The percentage of sequences in the *E. coli* genome denoted as essential is marked with a red line.

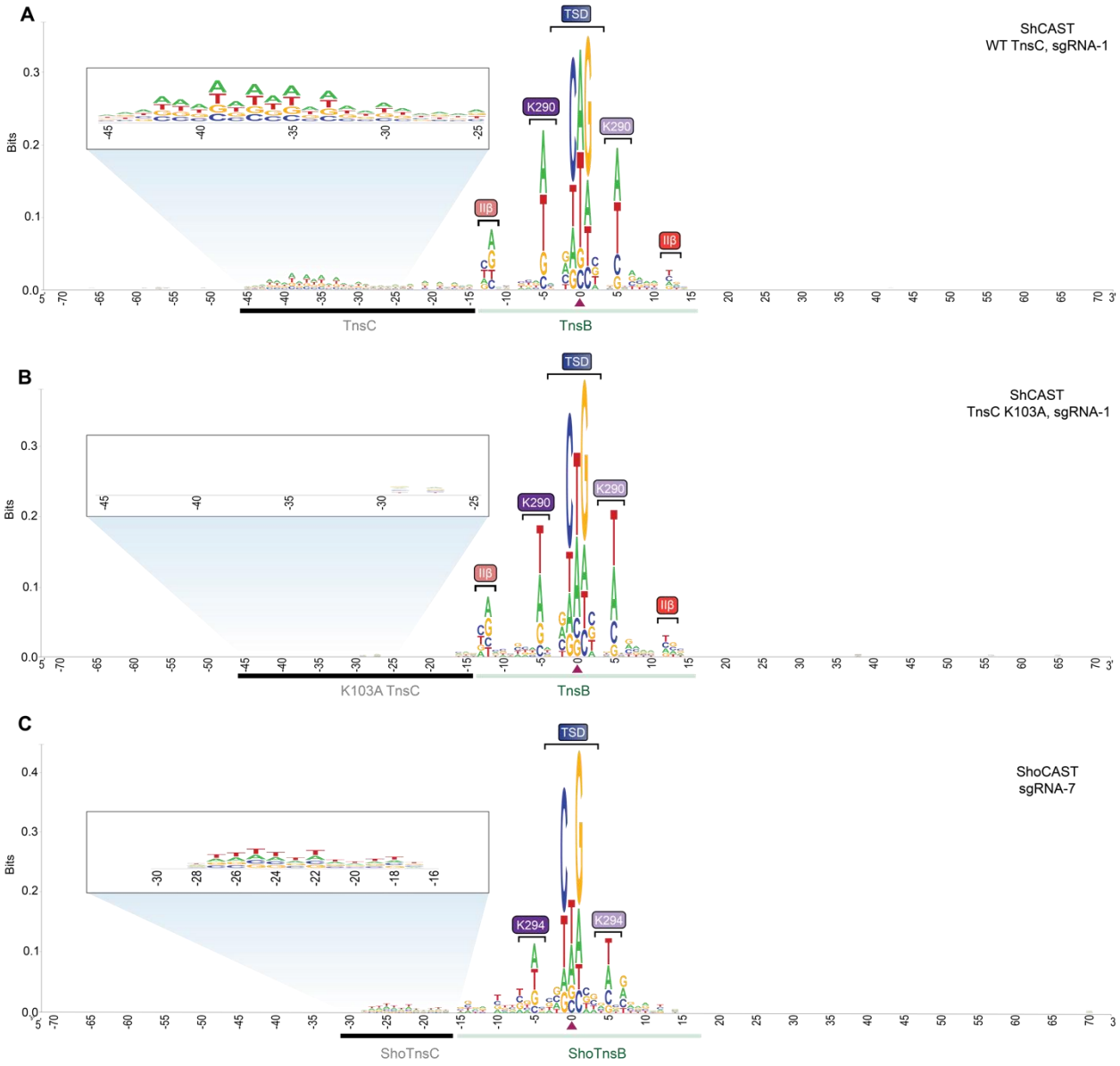

Fig. S7. TnsC selects AT-rich features upstream of the integration site. (**A**) WebLogo from a meta-analysis of untargeted genomic transposition (*N* = 6,800 unique integration events) from previously published data with pHelper and sgRNA-1 (*14*). The site of integration is noted with a maroon triangle. As with the dataset from this study (Fig. 3h), these results reveal an AT-rich sequence spanning ~25 bp that likely reflects the footprint of two turns of a TnsC filament (black), next to motifs within/near the target-site duplication (TSD) region that represent TnsB-specific sequence motifs (green). Specific TnsB residues/domains contacting the indicated nucleotides are shown. The zoomed-in inset highlights periodicity in the sequence bound by TnsC. (**B**) WebLogo from a meta-analysis of untargeted genomic transposition (*N* = 6,800 unique integration events) with a modified pHelper encoding a TnsC K103A mutation and sgRNA-1, visualized as in (**A**). The AT-rich motif is conspicuously absent, as compared to experiments with WT TnsC. (**C**) WebLogo from a meta-analysis of untargeted genomic transposition (*N* = 6,800 unique integration events) with the divergent ShoCAST system encoding sgRNA-7, visualized as in (**A**). The AT-

rich motif is conspicuously shorter in size with ShoCAST as compared to ShCAST, in agreement with the distinct integration distance distribution for both systems (*11*, *14*).

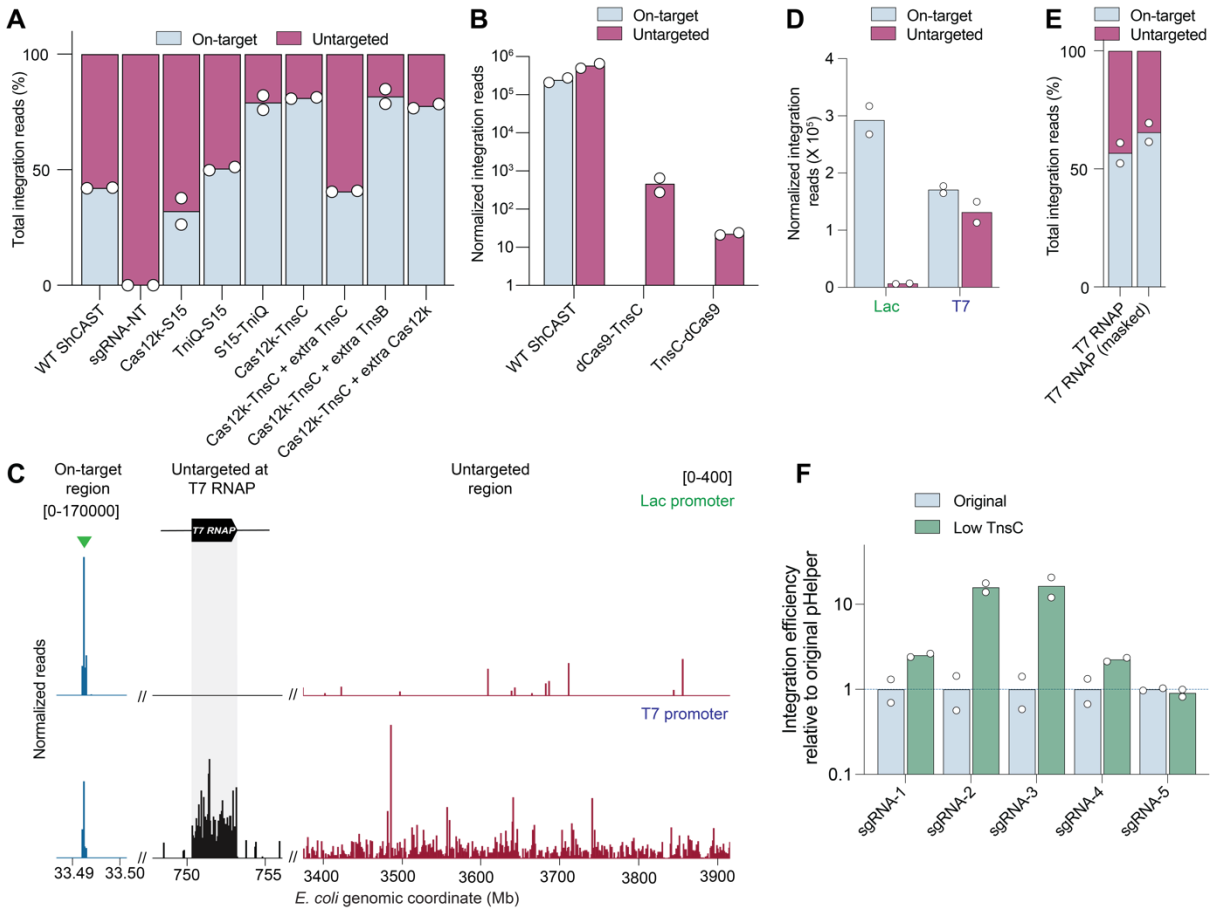

Fig. S8. Engineering strategies to rationally alter ShCAST specificity. (**A**) Fraction of total genome-mapping integration reads detected at on-target and untargeted sites, with WT ShCAST and sgRNA-1, ShCAST and a non-targeting sgRNA control (sgRNA-NT), or the listed ShCAST fusion constructs with sgRNA-1. In some cases, a fusion was supplemented with an additional plasmid encoding unfused protein components, as indicated. (**B**) Normalized genome-mapping integration reads detected at on-target and untargeted sites, for WT ShCAST and N- and C-terminal dCas9-TnsC fusions. No on-target reads were obtained for these fusion constructs. **C**, Normalized integration reads detected at the on-target site (left), the T7 RNAP locus (middle), and a representative untargeted region (right), using either a TnsC expression plasmid driven by a Lac or T7 promoter. A high density of untargeted insertions was observed within the T7 RNAP gene, likely due to positive selection for loss-of-function transposition events. Note the differing y-axis ranges. (**D**) Normalized genome-mapping integration reads detected at on-target and untargeted sites, when TnsC was expressed with a Lac or T7 promoter. (**E**) Fraction of total genome-mapping integration reads detected at on-target and untargeted sites from experiments with T7 promoter-driven TnsC, with or without computational masking of insertions within the T7 RNAP gene. (**F**) Relative on-target transposition efficiency measured for the modified design in which TnsC was expressed from a low-strength promoter compared to the original pHelper. The guide RNA are indicated sgRNA 1-5. Measurements were made by Taqman qPCR and normalized against the *rssA* locus. For (**A**), (**B**), (**D**)-(**F**), the mean is shown from $N = 2$ independent biological replicates.

Table S1. Cryo-EM data collection, refinement, and validation statistics

	BCQ transpososome (Consensus) (EMD-41280)	BCQ transpososome (TniQ-TnsC focus) (EMD-41280)	BCQ transpososome (Consensus) (EMD-41280)
Data collection and Processing (for each dataset)			
Microscope	Talos-Arctica	Talos-Arctica	Talos-Arctica
Voltage (keV)	200	200	200
Camera	K3	K3	K3
Magnification	63,000	63,000	63,000
Pixel size at detector (Å/pixel)	1.330	1.330	1.330
Total electron exposure (e^{-}/Å^2)	50	50	50
Exposure rate (e-/pixel/sec)	27.39	27.39	27.39
Number of frames collected during exposure	50	50	50
Defocus range (μm)	-1.0 – -2.5	-1.0 – -2.5	-1.0 – -2.5
Automation software (EPU, SerialEM or manual)	SerialEM	SerialEM	SerialEM
Energy filter slit width (if used)	20 eV	20 eV	20 eV
Micrographs collected (no.)	3,081	3,081	3,081
Micrographs used (no.)	2,790	2,790	2,790
Total extracted particles (no.)	993,190	993,190	993,190
For each reconstruction:			
Refined particles (no.)	577,628	577,628	577,628
Final particles (no.)	13,392	13,392	13,392
Point-group or helical symmetry parameters	C1	C1	C1
Resolution (global, Å)			
FSC 0.5 (unmasked/masked)	20.4/9.7	17.4/8.8	18.1/8.0
FSC 0.143 (unmasked/masked)	11.8/8.1	7.9/6.4	6.3/4.2
Resolution range (local, Å)	6.9 – 14.2	5.8 – 13.4	3.7 – 11.3
Resolution range due to anisotropy (Å)	7.6 – 11.2	6.7 – 8.3	4.2 – 6.9
Map sharpening *B* factor (Å^2) / (B factor Range)	-492.9	-354.5	-114.6
Map sharpening methods	cryoSPARC	cryoSPARC	cryoSPARC

Table S2.
Description and sequences of plasmids used in this study.

Table S3.
Recombinantly purified CAST proteins used in this study.

Table S4.
DNA oligos used in this study for biochemical experiments, PCR, qPCR, and NGS.

Table S5.
Guide RNAs and genomic target sites used in this study.

Table S6.
TagTn-seq statistics.

Movie S1.
Time-course experiments performed on single-molecule DNA curtains reveal preferential binding of mNG-labeled TnsC near the 3' (pedestal) end of λ-DNA substrate. Time (min) is indicated at the top and the scale bar (10 μm) at the bottom of the movie clip. Barrier and pedestal of DNA curtains are labeled with 'B' and 'P' respectively.

Movie S2.
Time-course experiments performed on single-molecule DNA curtains with TnsC at high concentrations (500 nM) led to the complete coating of λ-DNA. Labels are indicated as in Movie S1.

References and Notes

1. N. L. Craig, Target site selection in transposition. *Annu. Rev. Biochem.* **66**, 437–474 (1997). doi:10.1146/annurev.biochem.66.1.437 Medline

2. P. Siguier, E. Gourbeyre, M. Chandler, Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014). doi:10.1111/1574-6976.12067 Medline

3. P. Siguier, E. Gourbeyre, A. Varani, B. Ton-Hoang, M. Chandler, Everyman's guide to bacterial insertion sequences. *Microbiol. Spectr.* **3**, A3–A0030, 2014 (2015). doi:10.1128/microbiolspec.MDNA3-0030-2014 Medline

4. E. Arias-Palomo, J. M. Berger, An atypical AAA+ ATPase assembly controls efficient transposition through DNA remodeling and transposase recruitment. *Cell* **162**, 860–871 (2015). doi:10.1016/j.cell.2015.07.037 Medline

5. N. Mizuno, M. Dramićanin, M. Mizuuchi, J. Adam, Y. Wang, Y.-W. Han, W. Yang, A. C. Steven, K. Mizuuchi, S. Ramón-Maiques, MuB is an AAA+ ATPase that forms helical filaments to control target selection for DNA transposition. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2441–E2450 (2013). doi:10.1073/pnas.1309499110 Medline

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

7. J. E. Peters, N. L. Craig, Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev.* **15**, 737–747 (2001). doi:10.1101/gad.870201 Medline

8. J. E. Peters, N. L. Craig, Tn7: smarter than we thought. *Nat. Rev. Mol. Cell Biol.* **2**, 806–814 (2001). doi:10.1038/35099006 Medline

9. J. A. Finn, A. R. Parks, J. E. Peters, Transposon Tn7 directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. *J. Bacteriol.* **189**, 9122–9125 (2007). doi:10.1128/JB.01451-07 Medline

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

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

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

13. G. Faure, M. Saito, S. Benler, I. Peng, Y. I. Wolf, J. Strecker, H. Altae-Tran, E. Neumann, D. Li, K. S. Makarova, R. K. Macrae, E. V. Koonin, F. Zhang, Modularity and diversity of target selectors in Tn7 transposons. *Mol. Cell* **83**, 2122–2136.e10 (2023). doi:10.1016/j.molcel.2023.05.013 Medline

14. P. L. H. Vo, C. Ronda, S. E. Klompe, E. E. Chen, C. Acree, H. H. Wang, S. H. Sternberg, CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021). doi:10.1038/s41587-020-00745-y Medline

15. B. E. Rubin, S. Diamond, B. F. Cress, A. Crits-Christoph, Y. C. Lou, A. L. Borges, H. Shivram, C. He, M. Xu, Z. Zhou, S. J. Smith, R. Rovinsky, D. C. J. Smock, K. Tang, T. K. Owens, N. Krishnappa, R. Sachdeva, R. Barrangou, A. M. Deutschbauer, J. F. Banfield, J. A. Doudna, Species- and site-specific genome editing in complex bacterial communities. *Nat. Microbiol.* **7**, 34–47 (2022). doi:10.1038/s41564-021-01014-7 Medline

16. S.-C. Hsieh, J. E. Peters, Discovery and characterization of novel type I-D CRISPR-guided transposons identified among diverse Tn7-like elements in cyanobacteria. *Nucleic Acids Res.* **51**, 765–782 (2023). doi:10.1093/nar/gkac1216 Medline

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

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

19. P. A. Rice, N. L. Craig, F. Dyda, Comment on "RNA-guided DNA insertion with CRISPR-associated transposases". *Science* **368**, eabb2022 (2020). doi:10.1126/science.abb2022 Medline

20. C. J. Tou, B. Orr, B. P. Kleinstiver, Precise cut-and-paste DNA insertion using engineered type V-K CRISPR-associated transposases. *Nat. Biotechnol.* **41**, 968–979 (2023). doi:10.1038/s41587-022-01574-x Medline

21. G. D. Lampe, R. T. King, T. S. Halpin-Healy, S. E. Klompe, M. I. Hogan, P. L. H. Vo, S. Tang, A. Chavez, S. H. Sternberg, Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat. Biotechnol.*, 1–12 (2023). doi:10.1038/s41587-023-01748-1 Medline

22. M. Schmitz, I. Querques, S. Oberli, C. Chanez, M. Jinek, Structural basis for the assembly of the type V CRISPR-associated transposon complex. *Cell* **185**, 4999–5010.e17 (2022). doi:10.1016/j.cell.2022.11.009 Medline

23. J.-U. Park, A. W.-L. Tsai, A. N. Rizo, V. H. Truong, T. X. Wellner, R. D. Schargel, E. H. Kellogg, Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature* **613**, 775–782 (2023). doi:10.1038/s41586-022-05573-5 Medline

24. I. Querques, M. Schmitz, S. Oberli, C. Chanez, M. Jinek, Target site selection and remodelling by type V CRISPR-transposon systems. *Nature* **599**, 497–502 (2021). doi:10.1038/s41586-021-04030-z Medline

25. R. Xiao, S. Wang, R. Han, Z. Li, C. Gabel, I. A. Mukherjee, L. Chang, Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Mol. Cell* **81**, 4457–4466.e5 (2021). doi:10.1016/j.molcel.2021.07.043 Medline

26. J.-U. Park, A. W.-L. Tsai, E. Mehrotra, M. T. Petassi, S.-C. Hsieh, A. Ke, J. E. Peters, E. H. Kellogg, Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021). doi:10.1126/science.abi8976 Medline

27. R. M. Harshey, Transposable phage mu. *Microbiol. Spectr.* **2**, 2.5.31 (2014). doi:10.1128/microbiolspec.MDNA3-0007-2014 Medline

28. G. Ya. Kholodii, S. Z. Mindlin, I. A. Bass, O. V. Yurieva, S. V. Minakhina, V. G. Nikiforov, Four genes, two ends, and a res region are involved in transposition of Tn5053: A paradigm for a novel family of transposons carrying either a mer operon or an integron. *Mol. Microbiol.* **17**, 1189–1200 (1995). doi:10.1111/j.1365-2958.1995.mmi_17061189.x Medline

29. S. Hou, M. Brenes-Álvarez, V. Reimann, O. S. Alkhnbashi, R. Backofen, A. M. Muro-Pastor, W. R. Hess, CRISPR-Cas systems in multicellular cyanobacteria. *RNA Biol.* **16**, 518–529 (2019). doi:10.1080/15476286.2018.1493330 Medline

30. M. Ziemann, V. Reimann, Y. Liang, Y. Shi, H. Ma, Y. Xie, H. Li, T. Zhu, X. Lu, W. R. Hess, CvkR is a MerR-type transcriptional repressor of class 2 type V-K CRISPR-associated transposase systems. *Nat. Commun.* **14**, 924 (2023). doi:10.1038/s41467-023-36542-9 Medline

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

32. F. T. Hoffmann, M. Kim, L. Y. Beh, J. Wang, P. L. H. Vo, D. R. Gelsinger, J. T. George, C. Acree, J. T. Mohabir, I. S. Fernández, S. H. Sternberg, Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. *Nature* **609**, 384–393 (2022). doi:10.1038/s41586-022-05059-4 Medline

33. P. I. Hanson, S. W. Whiteheart, AAA+ proteins: Have engine, will work. *Nat. Rev. Mol. Cell Biol.* **6**, 519–529 (2005). doi:10.1038/nrm1684 Medline

34. J.-U. Park, A. W.-L. Tsai, T. H. Chen, J. E. Peters, E. H. Kellogg, Mechanistic details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202590119 (2022). doi:10.1073/pnas.2202590119 Medline

35. A. Meir, M. Kong, C. Xue, E. C. Greene, DNA llight on complex molecular systems during homologous recombination. *J. Vis. Exp.* (160): (2020). doi:10.3791/61320

36. A. Skalka, E. Burgi, A. D. Hershey, Segmental distribution of nucleotides in the DNA of bacteriophage lambda. *J. Mol. Biol.* **34**, 1–16 (1968). doi:10.1016/0022-2836(68)90230-1 Medline

37. E. C. Greene, K. Mizuuchi, Direct observation of single MuB polymers: Evidence for a DNA-dependent conformational change for generating an active target complex. *Mol. Cell* **9**, 1079–1089 (2002). doi:10.1016/S1097-2765(02)00514-2 Medline

38. J. Ge, Z. Lou, H. Cui, L. Shang, R. M. Harshey, Analysis of phage Mu DNA transposition by whole-genome *Escherichia coli* tiling arrays reveals a complex relationship to distribution of target selection protein B, transcription and chromosome architectural elements. *J. Biosci.* **36**, 587–601 (2011). doi:10.1007/s12038-011-9108-z Medline

39. M. W. G. Walker, S. E. Klompe, D. J. Zhang, S. H. Sternberg, Novel molecular requirements for CRISPR RNA-guided transposition. *Nucleic Acids Res.* **51**, 4519–4535 (2023). doi:10.1093/nar/gkad270 Medline

40. F. Tenjo-Castaño, N. Sofos, B. López-Méndez, L. S. Stutzke, A. Fuglsang, S. Stella, G. Montoya, Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon. *Nat. Commun.* **13**, 5792 (2022). doi:10.1038/s41467-022-33504-5 Medline

41. T. E. Haran, U. Mohanty, The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **42**, 41–81 (2009). doi:10.1017/S0033583509004752 Medline

42. A. E. Stellwagen, N. L. Craig, Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. *Genetics* **145**, 573–585 (1997). doi:10.1093/genetics/145.3.573 Medline

43. M. C. Biery, F. J. Stewart, A. E. Stellwagen, E. A. Raleigh, N. L. Craig, A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res.* **28**, 1067–1077 (2000). doi:10.1093/nar/28.5.1067 Medline

44. W. Chen, Z.-H. Ren, N. Tang, G. Chai, H. Zhang, Y. Zhang, J. Ma, Z. Wu, X. Shen, X. Huang, G.-Z. Luo, Q. Ji, Targeted genetic screening in bacteria with a Cas12k-guided transposase. *Cell Rep.* **36**, 109635 (2021). doi:10.1016/j.celrep.2021.109635 Medline

45. Custom Python scripts for: J. T. George, C. Acree, J.-U. Park, M. Kong, T. Wiegand, Y. L. Pignot, E. H. Kellogg, E. C. Greene, S. H. Sternberg, Mechanism of target site selection by type V-K CRISPR-associated transposases, Zenodo (2023); https://doi.org/10.1101/2023.07.14.548620.

46. F. Wang, Y. Liu, Z. Yu, S. Li, S. Feng, Y. Cheng, D. A. Agard, General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24269–24273 (2020). doi:10.1073/pnas.2009707117 Medline

47. M. A. Herzik Jr., Setting up parallel illumination on the Talos Arctica for high-resolution data collection. *Methods Mol. Biol.* **2215**, 125–144 (2021). doi:10.1007/978-1-0716-0966-8_6 Medline

48. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). doi:10.1016/j.jsb.2005.07.007 Medline

49. D. Tegunov, P. Cramer, Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019). doi:10.1038/s41592-019-0580-y Medline

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

51. T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, B. Berger, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019). doi:10.1038/s41592-019-0575-8 Medline

52. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi:10.1016/j.jsb.2012.09.006 Medline

53. G. Cardone, J. B. Heymann, A. C. Steven, One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013). doi:10.1016/j.jsb.2013.08.002 Medline

54. Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017). doi:10.1038/nmeth.4347 Medline

55. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi:10.1002/jcc.20084 Medline

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

57. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline

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

59. E. C. A. Goodall, A. Robinson, I. G. Johnston, S. Jabbari, K. A. Turner, A. F. Cunningham, P. A. Lund, J. A. Cole, I. R. Henderson, The essential genome of Escherichia coli K-12. *mBio* **9**, e02096-17 (2018). doi:10.1128/mBio.02096-17 Medline

60. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004). doi:10.1101/gr.849004 Medline

61. E. C. Greene, S. Wind, T. Fazio, J. Gorman, M.-L. Visnapuu, DNA curtains for high-throughput single-molecule optical imaging. *Methods Enzymol.* **472**, 293–315 (2010). doi:10.1016/S0076-6879(10)72006-1 Medline

62. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi:10.1038/nmeth.2019 Medline