nature structural & molecular biology

SUPPLEMENTARY INFORMATION

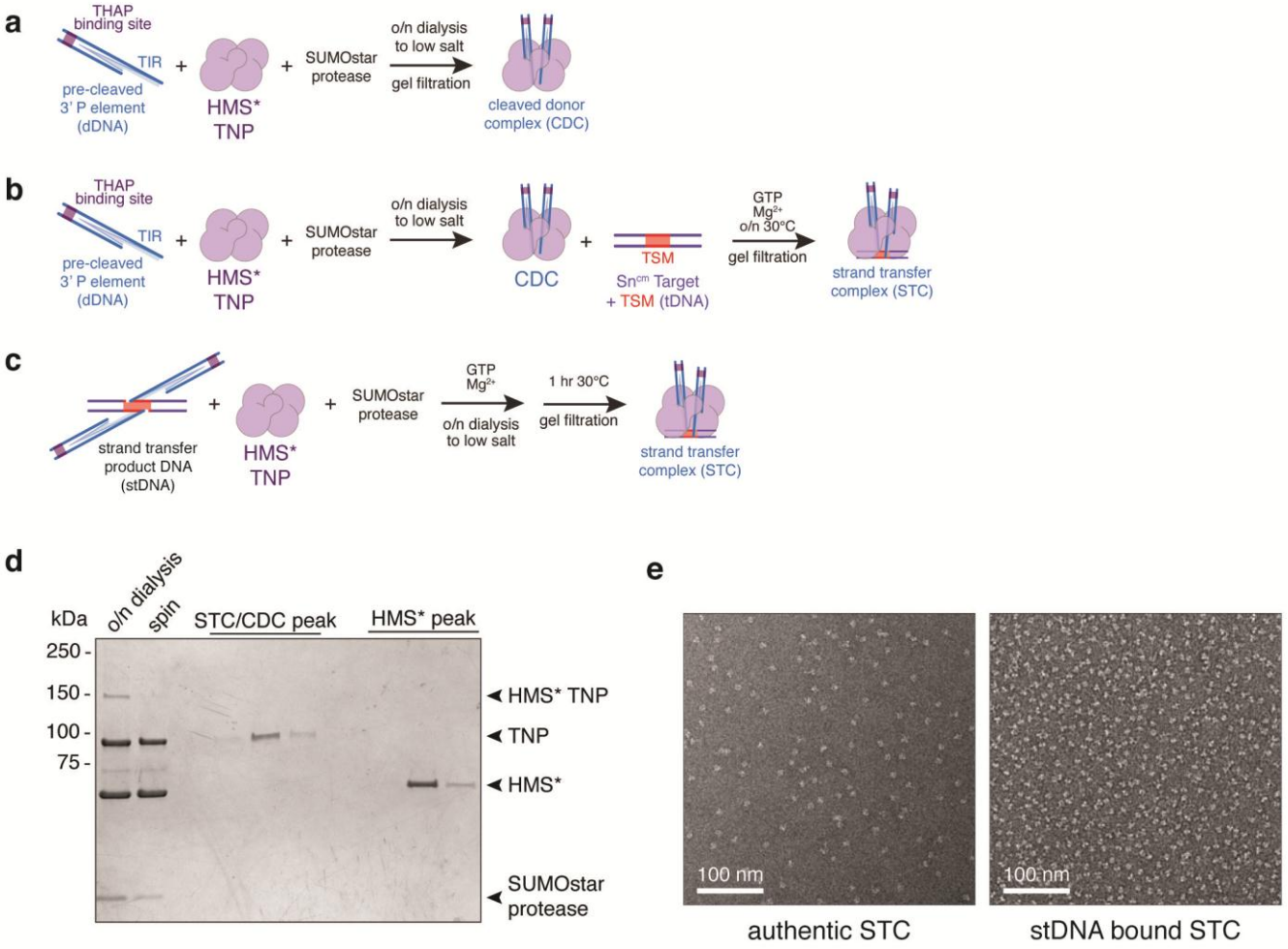
https://doi.org/10.1038/s41594-019-0319-6

In the format provided by the authors and unedited.

# Structure of a P element transposase–DNA complex reveals unusual DNA structures and GTP-DNA contacts

**George E. Ghanim[1,2,6], Elizabeth H. Kellogg[2,5,6]*, Eva Nogales[1,3,4] and Donald C. Rio[1,2]***

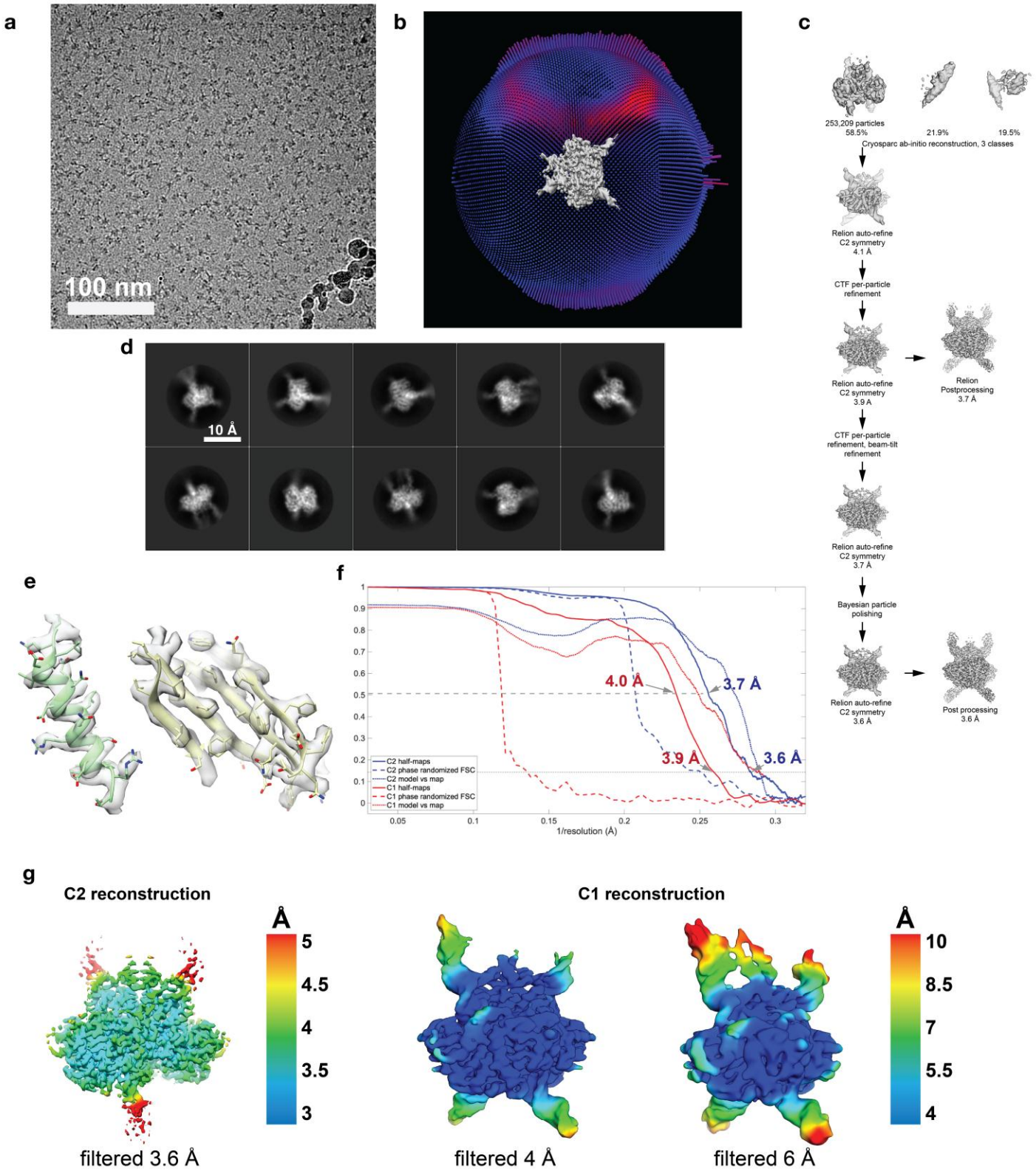
[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [2]California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA, USA. [3]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. [4]Molecular Biophysics and Integrative Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Present address: Molecular Biology and Genetics Department, Cornell University, Ithaca, NY, USA. [6]These authors contributed equally: George E. Ghanim, Elizabeth H. Kellogg.
*e-mail: ehk68@cornell.edu; don_rio@berkeley.edu




**Supplementary Figure 1**

Assembly of the CDC, authentic STC, and STC bound to stDNA

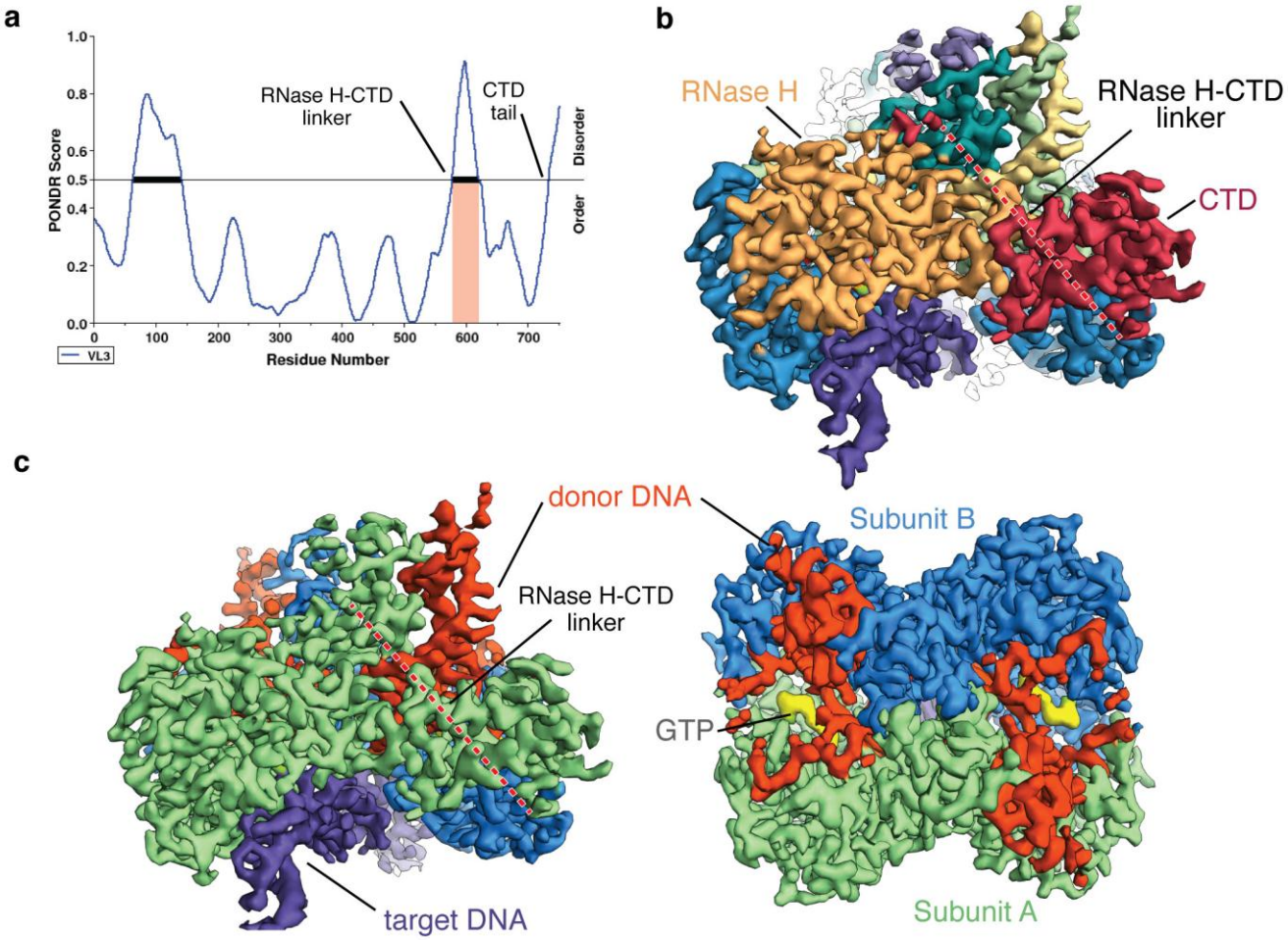
**a**, Diagram of the cleaved donor complex (CDC) assembly pathway. (TIR, terminal inverted repeat). **b**, Diagram of the authentic strand transfer complex (STC) assembly pathway. CDCs were assembled as in Supplementary Fig. 1a, then provided with an idealized hotspot target DNA from the *Drosophila singed* locus (tDNA). (GTP, guanosine triphosphate; TSM, target sequence motif). **c**, Diagram of the strand transfer complexes assembled on strand transfer product DNA (stDNA). **d**, Representative Coomassie-stained SDS-PAGE of the overnight dialysis and gel filtration fractions. 'o/n dialysis' and 'spin' lanes were diluted 1:10 before loading. (HMS*, 6xHis-Maltose binding protein-SUMO* tandem solubility tag). **e**, Negative stain electron micrographs of authentic STC and stDNA bound STC.




**Supplementary Figure 2**

Image processing of tilted dataset leading to a 3.6 Å resolution cryo-EM reconstruction

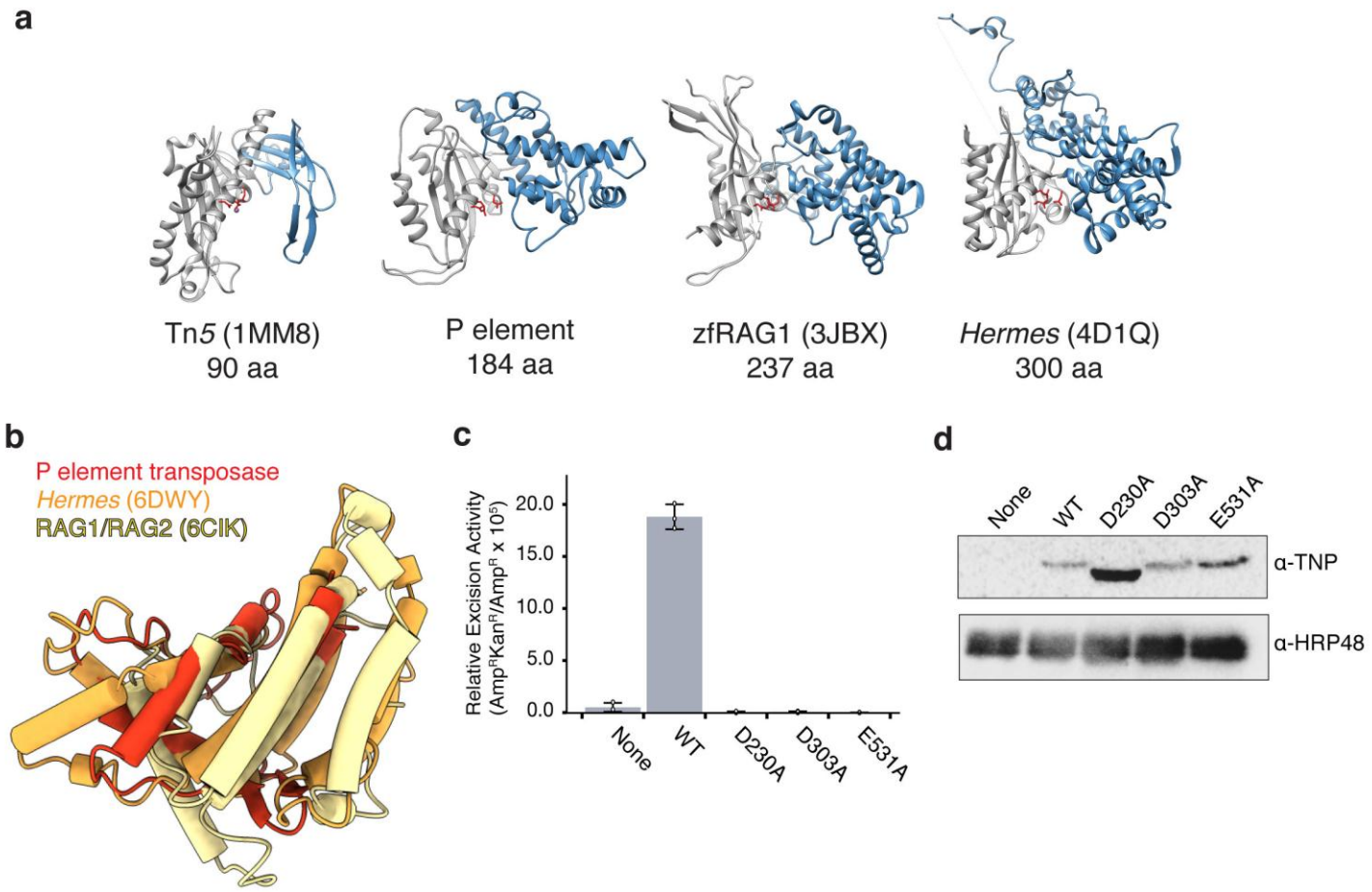
**a**, Representative cryo-EM image collected with a 40° tilt showing well-defined, monodispersed particles (scale bar represents 100 nm). **b**, Angular distribution of particles from the tilted dataset is cone-like, corresponding to a majority of top-views. **c**, A single, well-defined reconstruction was produced using cryoSPARC and subsequently refined to high resolution using RELION-3.0 (see Methods for details). **d**, Reference-free 2D classes of the tilted data reveal secondary structure features. **e**, The secondary structure features are consistent with the estimated resolution of the map, with well-defined secondary structure and distinctive densities for large side-chain. **f**, The overall resolution (based on the Fourier shell correlation (FSC) 0.143 criterion) for the symmetrized reconstruction is 3.5 Å (3.6 Å if using randomized phases), and 3.9 Å for the unsymmetrized reconstruction. The map versus model resolution is 3.7 and 4 Å, respectively, for the symmetrized and unsymmetrized maps. **g**, 3D map for the C2 (symmetrized, left) and C1 (unsymmetrized, right) reconstructions colored by local resolution showing the core of the structure to be around 3.5 Å. To show some of the most disordered regions, the C1 map is shown low-pass filtered to both 4 Å and 6 Å.




**Supplementary Figure 3**

The STC is dimeric and contains disordered regions

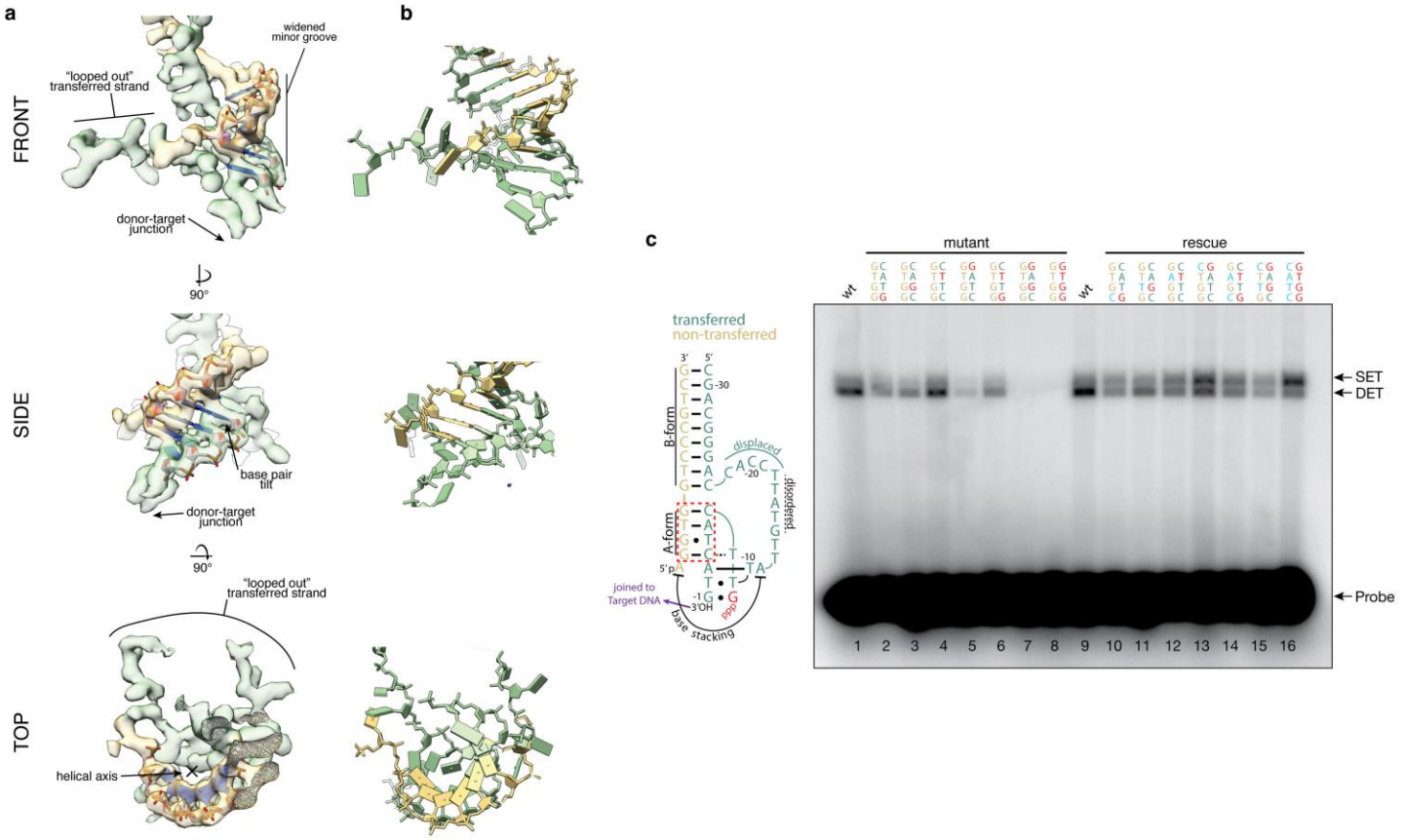
**a**, PONDR scores of predicted disordered regions. Disordered regions predicted with high confidence (indicated by black bars), are within the leucine zipper dimerization domain and the RNase H – CTD linker region. Contrary to the prediction results, we observe the dimerization domain to be largely ordered in the C1 reconstruction, in spite of the prediction suggesting that this region may undergo a disorder-to-order transition upon dimerization (Dunker, A. K. *et al., J. Mol. Graph. Model.* **19,** 26–59, 2001). **b**, The disordered linker spanning the RNase H and CTD domains is represented by a red dashed line, with density for ordered regions colored by domain as in Fig. 2c. **c**, STC subunit organization. Densities are shown as in Fig. 2c but are colored by subunit (blue and green). Donor DNAs are colored in red, and target DNA in purple. The density corresponding to GTP is indicated in yellow.




**Supplementary Figure 4**

A comparison of RNase H insertion domains among different transposases and RNase H catalytic mutants are inactive

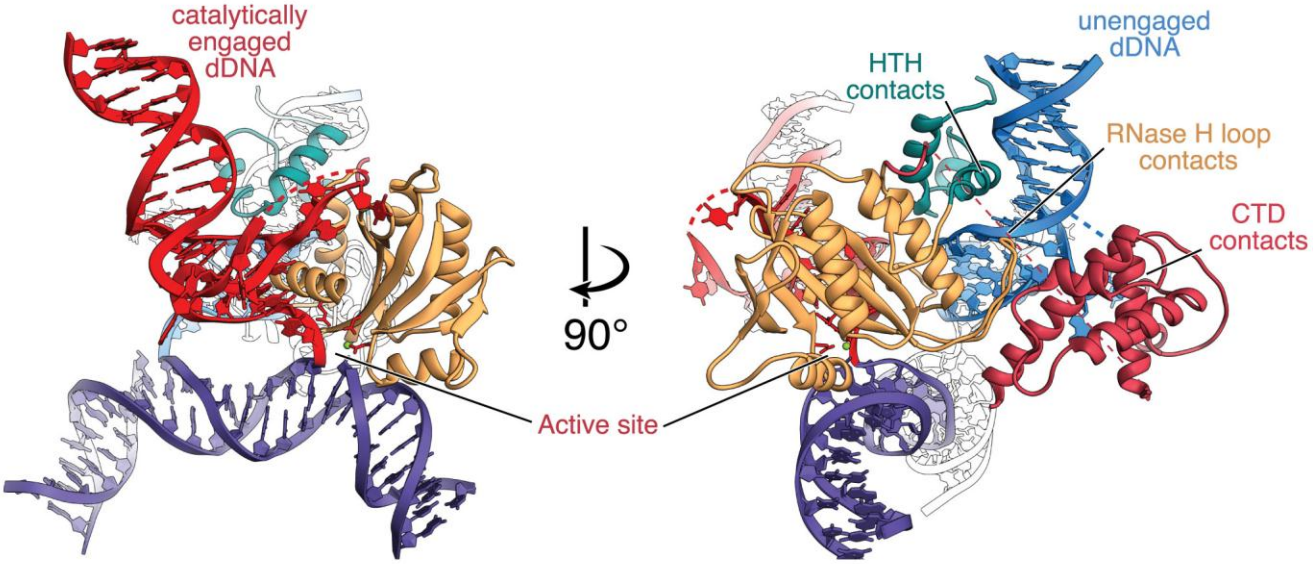
**a**, Architectures of insertion domains found in other DNA transposases. The RNase H domains (grey) of other structurally characterized DNA transposases (or the transposase-related RAG1 protein) were aligned by their respective catalytic residues (indicated in red) and ordered by increasing insertion domain size (blue). Insertion domain sizes (indicated below) were determined by approximate start and end insertion positions. The PDB numbers from which these structures were derived are in parentheses. **b**, Structural alignment of the P element transposase insertion domain, the *Hermes* insertion domain (1DWY) and the RAG1–RAG2 insertion domain (6CIK) reveals structural similarities at the fold level. **c**, Bar graph of relative *in vivo* P element excision activity of alanine-substituted catalytic mutants (D230, D303 and E531). Cell-based excision assays were performed as previously described (Rio, D. C. *et al., Cell.* **44,** 21–32, 1986)(Beall, E. L. *et al., Genes Dev.* **10,** 921–933, 1996). Single alanine mutants were generated by site-directed mutagenesis of pPBSKS (+) pAc-TNP and verified by sequencing over the entire coding sequence. The assay was conducted in triplicate (n = 3). Error bars indicate standard deviations. (WT, wild type). **d**, Representative immunoblot of wild type transposase and catalytic mutant protein expression levels. Cells were harvested 24hr after transfection and lysates were normalized to cell number. Membrane was cut and then immunoblotted with anti-transposase antibodies (α-TNP) or a loading control (α-HRP48).




**Supplementary Figure 5**

Characteristics of A-form DNA are well resolved and base pairing between distant donor DNA regions is required for strand transfer activity

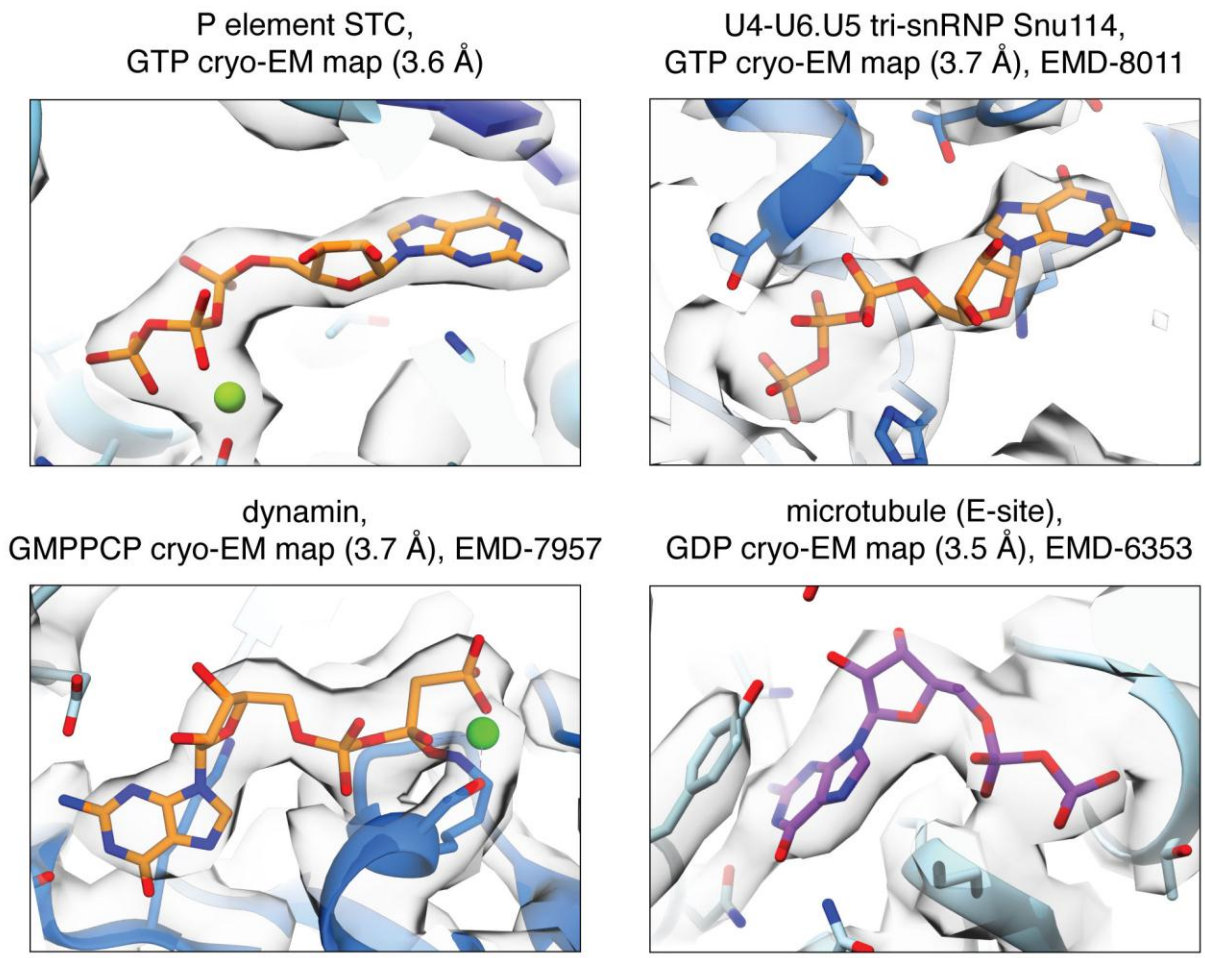
**a**, Ideal A-form DNA fitted into donor DNA reconstruction, depicting widened minor groove, base pair tilt, axial rise and helical axis dislocation relative to base pairs. A single donor DNA is depicted for clarity. The reconstruction is colored green and yellow, for transferred strand and non-transferred strand, respectively. Relevant regions of DNA are indicated. **b**, Atomic model of donor DNA depicting A-form DNA characteristics. Views are as in **a**, except only relevant regions the donor DNA atomic model are depicted. **c**, Schematic of the secondary structure of a donor DNA terminal inverted repeat (left). Watson-Crick base pairing is indicated by solid lines. Non-canonical base pairing is indicated by dots, or dotted lines. Nucleotides of the transferred strand are numbered -1 to -31, starting at the 3' terminal guanosine. Distant noncanonical A-form helical base pairing between the transferred and non-transferred strand is highlighted (dashed red box). Agarose gel of a strand transfer assay with 5'-radiolabeled mutant and rescue donor DNAs (right). Assays were largely performed as previously described (Beall, E. L., *et al., EMBO J.* **17,** 2122–2136, 1998). The base pairs are shown above each lane, with the substituted bases highlighted in red (mutant, lanes 2 - 8). Compensatory substitutions in the non-transferred strand are shown above each lane, with substitutions to restore base pairing highlighted in blue (rescue, lanes 10 - 16). The expected positions of single-ended integration (SET) and double-ended integration (DET), as well as free donor DNA (Probe), are indicated. wt, wild type donor DNA.




**Supplementary Figure 6**

A single transposase subunit engages both P element donor DNAs

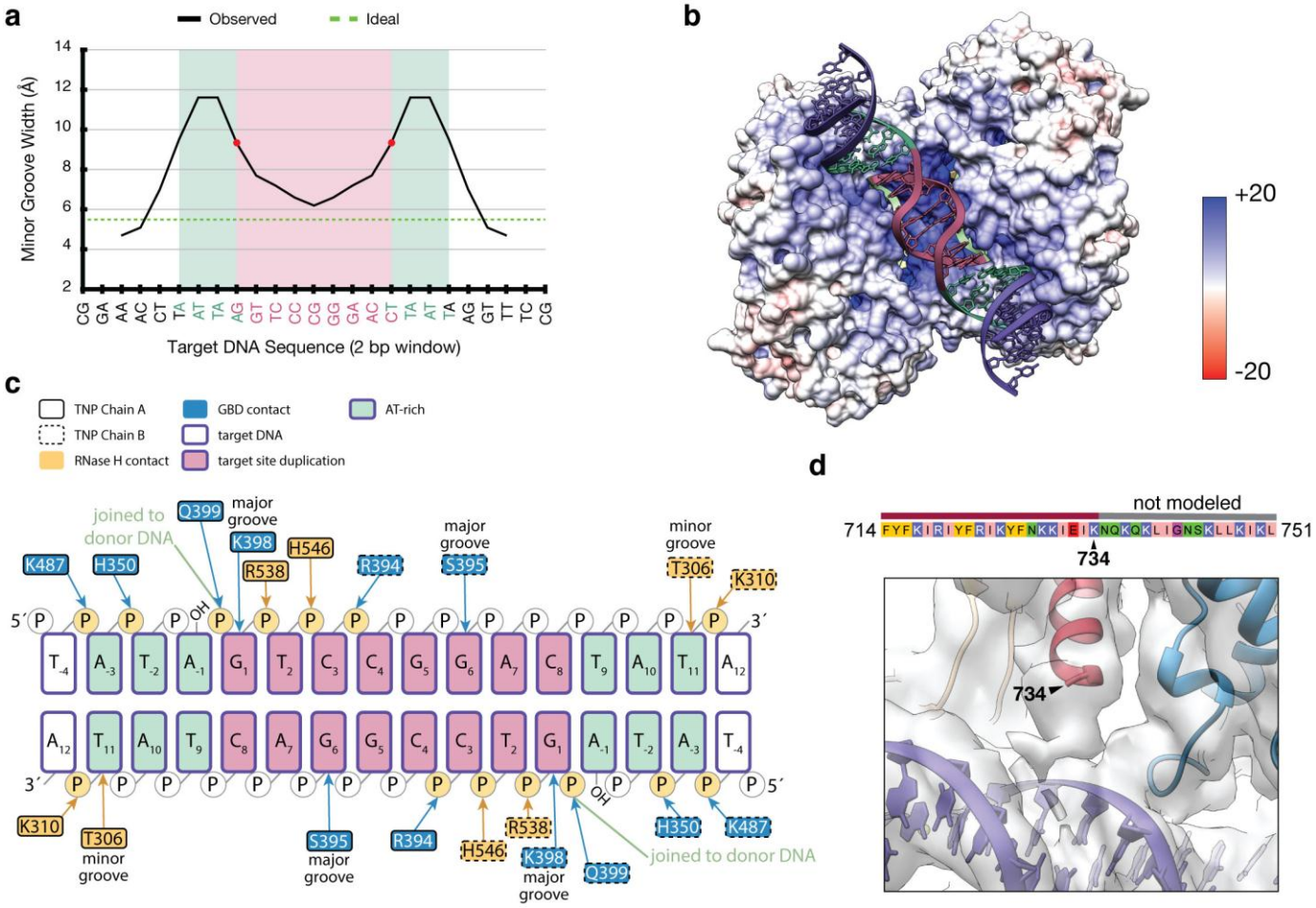
Left, the RNase H domain of one subunit of transposase is catalytically engaged with one P element donor DNA (red DNA). The domains are colored as in Fig. 2c. For clarity, the dimerization domain, the GBD, and the other transposase subunit are not shown. Catalytic residues are depicted in red. Right, a 90° rotated view shows the same subunit contacting the other P element donor DNA (blue DNA), through the HTH domain, a long loop in the RNase H domain, and through the CTD. This mode of engagement likely acts as a regulatory step to ensure proper assembly with both P element ends before proceeding to catalysis.




**Supplementary Figure 7**

TNP nucleotide density is consistent with GTP and distinguishable from GDP within the reported resolution regime

The GTP density in our cryo-EM map (top-left) is consistent with the GTP density observed the cryo-EM reconstruction of the U4–U6.U5 tri-snRNP Snu114 (3.7Å, top-right) and the non-hydrolyzable GTP analog (GMPPCP) of dynamin (3.7Å, bottom-left) and inconsistent with GDP in the β-tubulin subunit (3.5 Å, bottom-right), for which GTP is hydrolyzed during microtubule assembly.




**Supplementary Figure 8**

Target DNA binds in a positively charged groove

**a**, Plot of target DNA minor groove width. The minor groove width was calculated from the target DNA model using the 3DNA webserver (Li, S., *et al.*, *Nucleic Acids Research.* **47,** W26–W34, 2019), with a 2 bp sliding window, accounting for phosphate van der Waals radii. The target DNA sequence is depicted on the x-axis and colored as in Fig. 5**a**. Red dots indicate transposition sites, on either the top or bottom strand of the target DNA. **b**, Electrostatic surface potential of the STC as viewed from below the target DNA binding site. Calculations were performed in UCSF Chimera (Pettersen, E. F., *et al., J Comput. Chem.* **25,** 1605–1612, 2004). Blue denotes a positive charge and red denotes a negative charge. Target DNA is shown as in Fig. 5a. **c**, Schematic representation of observed base-specific and backbone contacts between transposase and the target DNA. Target DNA (purple border) is numbered as in Fig. 5e (target site duplication, pink fill; AT-rich flanks, green fill). Residue numbers are indicated and outlined in a solid or dashed border to indicate transposase subunit A, or transposase subunit B, respectively. Residues are colored according to domain (RNase H, orange; GBD, blue). Direct contacts are shown as solid lines; aromatic base stacking interactions are shown as dashed lines; major groove, minor groove and main chain contacts are indicated; interacting phosphates are highlighted yellow. **d**, Unmodeled density at the C-terminus is oriented towards the target DNA. The map is low-pass filtered to 4 Å to more clearly show the presence of additional, poorly-ordered density. While we could not confidently build into the density beyond position 734, the highly basic nature and positioning of the weak density near DNA suggests that this region likely plays a role in target DNA binding. Consistent with this, C-terminal tags on transposase decrease the overall excision and strand transfer activity (unpublished results, D. Rio).

## Supplementary Table 1. DNA substrates used in this study

| ID # | Description | Sequence |
|---|---|---|
| | **Figure 1b-e** | |
| **399** | 3' P element pre-cleaved strand transfer oligo, extends 10 bp past THAP binding site | CAAGCATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCATCATG |
| **401** | 5' phosphorylated non-transferred strand 3' P element strand transfer oligo, extends 9 bp past THAP binding site | /5PHOS/AGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTT |
| **409** | Blunt Singed locus replaced with TSM TOP | CAACGGGTTTCATATAGTCCGGACTATAGTTCGTGAGCGG |
| **410** | Blunt Singed locus replaced with TSM BOTTOM | CCGCTCACGAACTATAGTCCGGACTATATGAAACCCGTTG |
| | **Figure 1b, e & Figure 2** | |
| **623** | 3' strand transfer (55) half site integration into Sn60 target, extends 13 bp of "right side" of Sn60 | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCATCATGGTCCG GACTATAGTTCGTGAGCGG |
| **624** | bottom strand of Sn60 target for half site | CCGCTCACGAACTATA |
| **625** | 5' phosphorylated non-transferred 3' strand transfer (55), extends 5 bp past THAP binding site | /5PHOS/AGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| | **Figure 5e** | |
| | 5' TAMRA labeled 3' P element pre-cleaved strand transfer oligo, extends 2 bp past TBS | /TAMRA/CGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCATCATG |
| **441** | Blunt Targets for high resolution STA, singed locus w/ TSM | CGCTCGCAACGGGTTTCATATAGTCCGGACTATAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **442** | Blunt Targets for high resolution STA, singed locus w/ TSM | AGAGGAGAGAACGACCGCTCACGAACTATAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **476** | Targets for oligonucleotide strand transfer assay, singed locus w/ TSM, GGMM at 4 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGTCGGGACTATATGAAACCCGTTGCGAGCG |
| **477** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 5 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGTGCGGACTATATGAAACCCGTTGCGAGCG |
| **478** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 6 (2009 NAR) | CGCTCGCAACGGGTTTCATATAGTCCGGGCTATAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **479** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 6 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGGCCGGACTATATGAAACCCGTTGCGAGCG |
| **480** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 7 (2009 NAR) | CGCTCGCAACGGGTTTCATATAGTCCGGAGTATAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **481** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 7 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **482** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 8 (2009 NAR) | CGCTCGCAACGGGTTTCATATAGTCCGGACGATAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **483** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 8 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATGGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **484** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 9 (2009 NAR) | CGCTCGCAACGGGTTTCATATAGTCCGGACTGTAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **485** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 9 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTAGAGTCCGGACTATATGAAACCCGTTGCGAGCG |

| | | |
|---|---|---|
| **486** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 10 (2009 NAR) | CGCTCGCAACGGGTTTCATATAGTCCGGACTAGAGTTCGTGAGCGGTCGTTCTCTCCTCT |
| **487** | Targets for high resolution STA, singed locus w/ TSM, GGMM at 10 (2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTGTAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **488** | Targets for high resolution STA, singed locus w/ TSM, Nick between 3/4(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGTCC |
| **489** | Targets for high resolution STA, singed locus w/ TSM, Nick between 3/4(2009 NAR) | GGACTATATGAAACCCGTTGCGAGCG |
| **490** | Targets for high resolution STA, singed locus w/ TSM, Nick between 4/5(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGTC |
| **491** | Targets for high resolution STA, singed locus w/ TSM, Nick between 4/5(2009 NAR) | CGGACTATATGAAACCCGTTGCGAGCG |
| **492** | Targets for high resolution STA, singed locus w/ TSM, Nick between 5/6(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAGT |
| **493** | Targets for high resolution STA, singed locus w/ TSM, Nick between 5/6(2009 NAR) | CCGGACTATATGAAACCCGTTGCGAGCG |
| **494** | Targets for high resolution STA, singed locus w/ TSM, Nick between 6/7(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATAG |
| **495** | Targets for high resolution STA, singed locus w/ TSM, Nick between 6/7(2009 NAR) | TCCGGACTATATGAAACCCGTTGCGAGCG |
| **496** | Targets for high resolution STA, singed locus w/ TSM, Nick between 7/8(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTATA |
| **497** | Targets for high resolution STA, singed locus w/ TSM, Nick between 7/8(2009 NAR) | GTCCGGACTATATGAAACCCGTTGCGAGCG |
| **497** | Targets for high resolution STA, singed locus w/ TSM, Nick between 8/9(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTAT |
| **499** | Targets for high resolution STA, singed locus w/ TSM, Nick between 8/9(2009 NAR) | AGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **500** | Targets for high resolution STA, singed locus w/ TSM, Nick between 9/10(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACTA |
| **501** | Targets for high resolution STA, singed locus w/ TSM, Nick between 9/10(2009 NAR) | TAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **502** | Targets for high resolution STA, singed locus w/ TSM, Nick between 10/11(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAACT |
| **503** | Targets for high resolution STA, singed locus w/ TSM, Nick between 10/11(2009 NAR) | ATAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **504** | Targets for high resolution STA, singed locus w/ TSM, Nick between 11/12(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAAC |
| **505** | Targets for high resolution STA, singed locus w/ TSM, Nick between 11/12(2009 NAR) | TATAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| **506** | Targets for high resolution STA, singed locus w/ TSM, Nick between 12/13(2009 NAR) | AGAGGAGAGAACGACCGCTCACGAA |
| **507** | Targets for high resolution STA, singed locus w/ TSM, Nick between 12/13(2009 NAR) | CTATAGTCCGGACTATATGAAACCCGTTGCGAGCG |
| | **Supplementary Figure 5c** | |

| | | |
|---|---|---|
| **727** | 3' P element pre-cleaved strand transfer oligo, wild type transferred strand | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCATCATG |
| **728** | 3' P element pre-cleaved strand transfer oligo, double swap at nt 26 and 28 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCTTGATG |
| **729** | 3' P element pre-cleaved strand transfer oligo, double swap at nt 25 and 27 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTGAGCATG |
| **730** | 3' P element pre-cleaved strand transfer oligo, complete swap at nt 25-28 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTGTGGATG |
| **731** | 3' P element pre-cleaved strand transfer oligo, wild type non-transferred strand | AGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **732** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#728** | ACGAGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **733** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#729** | AGTTCGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **734** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#730** | ACTACGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **736** | 3' P element pre-cleaved strand transfer oligo, single swap at nt 28 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCATGATG |
| **737** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#736** | ACGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **738** | 3' P element pre-cleaved strand transfer oligo, single swap at nt 27 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCAGCATG |
| **739** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#738** | AGTTGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **740** | 3' P element pre-cleaved strand transfer oligo, single swap at nt 26 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTCTTCATG |
| **741** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#740** | AGGAGGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |
| **742** | 3' P element pre-cleaved strand transfer oligo, single swap at nt 25 of 31 bp TIR | ATACGTTAAGTGGATGTCTCTTGCCGACGGGACCACCTTATGTTATTTGATCATG |
| **743** | 3' P element pre-cleaved non-transferred strand oligo, compensatory swap for **#742** | AGGTCGTCCCGTCGGCAAGAGACATCCACTTAACGTAT |

**Supplementary Table 2. Cryo-EM data collection, refinement and validation statistics**

| | STC-C2 (EMD-20254, PDB 6P5A) | STC-C1 (EMD-20321 PDB 6PE2) |
|---|---|---|
| **Data collection and processing** | | |
| Magnification | 35,000 | 35,000 |
| Voltage (kV) | 200 | 200 |
| Electron exposure (e–/Å$^2$) | 60 | 60 |
| Defocus range (μm) | -1 to -3 μm | -1 to -3 μm |
| Pixel size (Å) | 1.16 | 1.16 |
| Symmetry imposed | C2 | C1 |
| Initial particle images (no.) | 547,929 | 547,929 |
| Final particle images (no.) | 252,574 | 252,574 |
| Map resolution (Å) | 3.6 | 3.9 |
| FSC threshold | 0.143 | 0.143 |
| Map resolution range (Å) | 3 - 5 | 4 - 10 |
| | | |
| **Refinement** | | |
| Initial model used | - | 6P5A |
| Model resolution (Å) | 3.7 | 4 |
| FSC threshold | 0.5 | 0.5 |
| Model resolution range (Å) | - | - |
| Map sharpening *B* factor (Å$^2$) | 100 | 100 |
| Model composition | | |
| Nonhydrogen atoms | 11,956 | 12,753 |
| Protein residues | 1,120 | 1,148 |
| Ligands | 6 | 6 |
| *B* factors (Å$^2$) | | |
| Protein | 47.13 | 185.97 |
| Ligand | 38.87 | 171.81 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.008 | 0.003 |
| Bond angles (°) | 0.52 | 0.532 |
| Validation | | |
| MolProbity score | 1.22 | 1.56 |
| Clashscore | 4.46 | 6.4 |
| Poor rotamers (%) | 0% | 0% |
| Ramachandran plot | | |
| Favored (%) | 98% | 96.75% |
| Allowed (%) | 2% | 3.25% |
| Disallowed (%) | 0% | 0% |