

# **Supplementary Materials**

## **Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon**

Francisco Tenjo-Castaño<sup>1</sup>, Nicholas Sofos<sup>1</sup>, Blanca López-Méndez<sup>2</sup>, Luisa S. Stutzke<sup>1</sup>, Anders Fuglsang<sup>1</sup>, Stefano Stella<sup>1‡</sup> and Guillermo Montoya<sup>1\*</sup>

<sup>1</sup>Structural Molecular Biology Group, Novo Nordisk Foundation Centre for Protein Research, Faculty of Health and Medical Sciences University of Copenhagen; Copenhagen, 2200, Denmark.

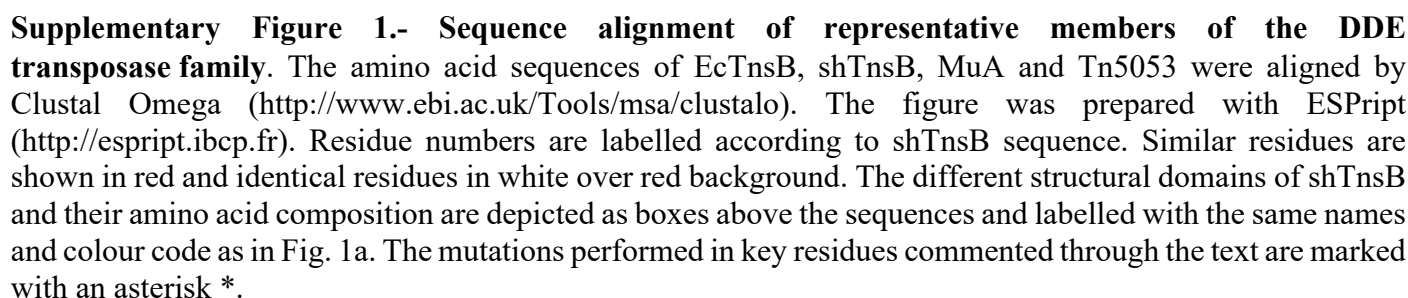
<sup>2</sup>Protein Purification and Characterisation Facility, Novo Nordisk Foundation Centre for Protein Research, Faculty of Health and Medical Sciences University of Copenhagen; Copenhagen, 2200, Denmark.

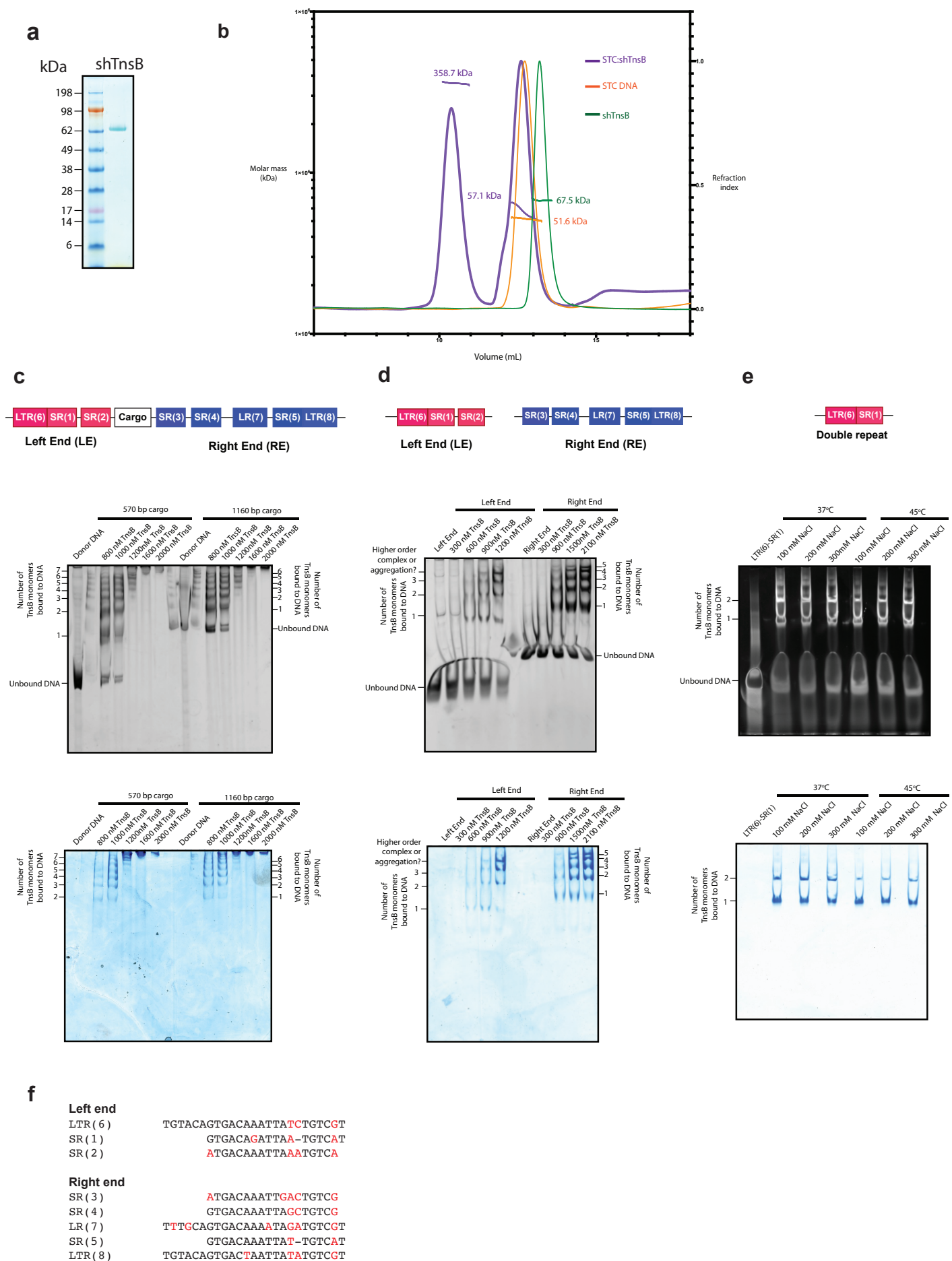
### **INDEX**

**Supplementary Figures 1-8**

**Supplementary Movie Legend**

**Supplementary Tables 1-2**

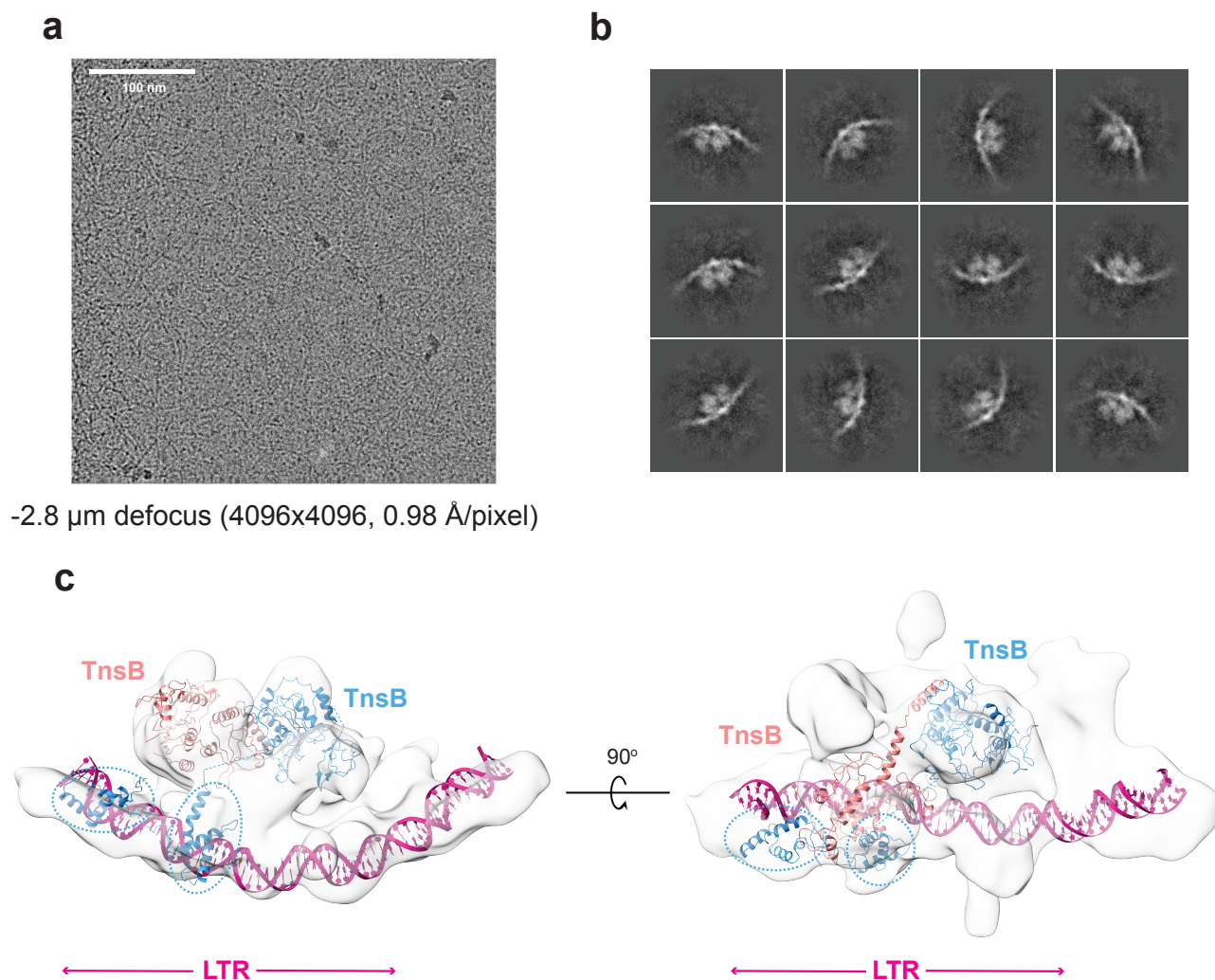




**Supplementary Figure 2.- Purification and characterisation of shTnsB and the Strand Transfer Complex.** **a)** SDS-PAGE showing the purified shTnsB **b)** SEC-MALS of purified TnsB (green), STC DNA (orange), and shTnsB:STC (purple). The results agree with the theoretical molecular weight of a TnsB

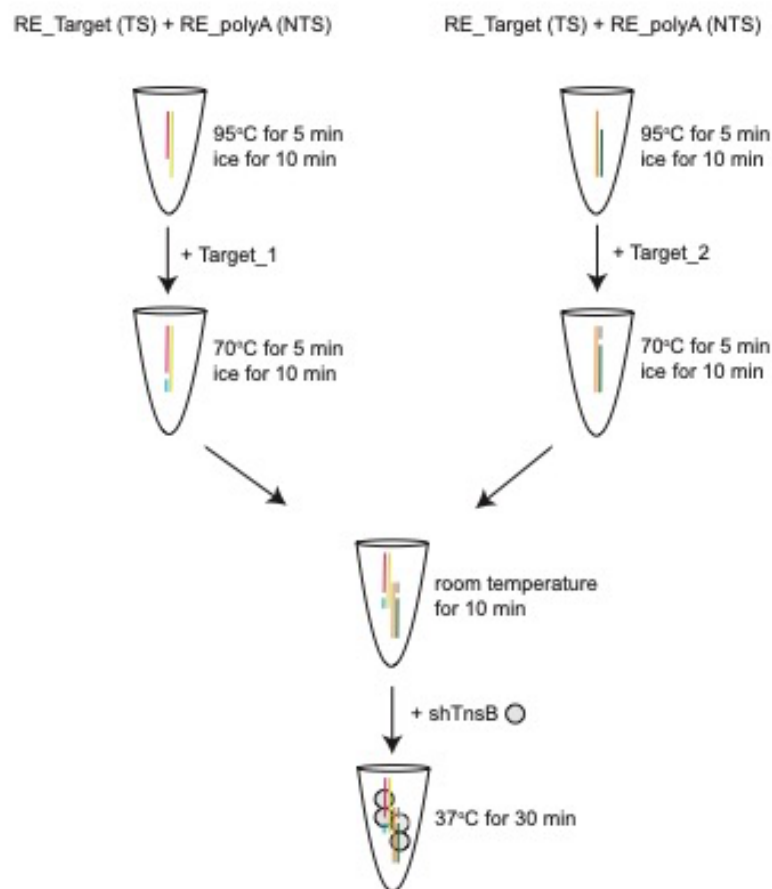
monomer (68 kDa) and the separate target bound RE and LE DNA (52.6 kDa and 52 kDa). As for the complex, the observed molecular weight (358.7 kDa) corresponds to a complex containing the target bound RE and LE and a TnsB tetramer, which has a theoretical molecular weight of 376.6 kDa. **c) Electromobility shift assay (EMSA) from** complexes containing donor DNA with two different cargos. Donor DNA concentration was 100 nM in all samples. TnsB concentration in each sample is indicated in the figure. **d) EMSA from** complexes containing only RE or LE. Donor DNA concentration was 300 nM in all samples. TnsB concentration in each sample is indicated in the figure. **e) EMSA from** complex containing only LTR(6) and SR(1) tested at different temperatures and NaCl concentrations. DNA concentration was 2  $\mu$ M and TnsB concentration was 4  $\mu$ M. All EMSA experiments were performed twice independently producing similar results. **f)** Alignment of the different long terminal repeats (LTR), short repeats (SR) and long repeat (LR) in the RE and LE.



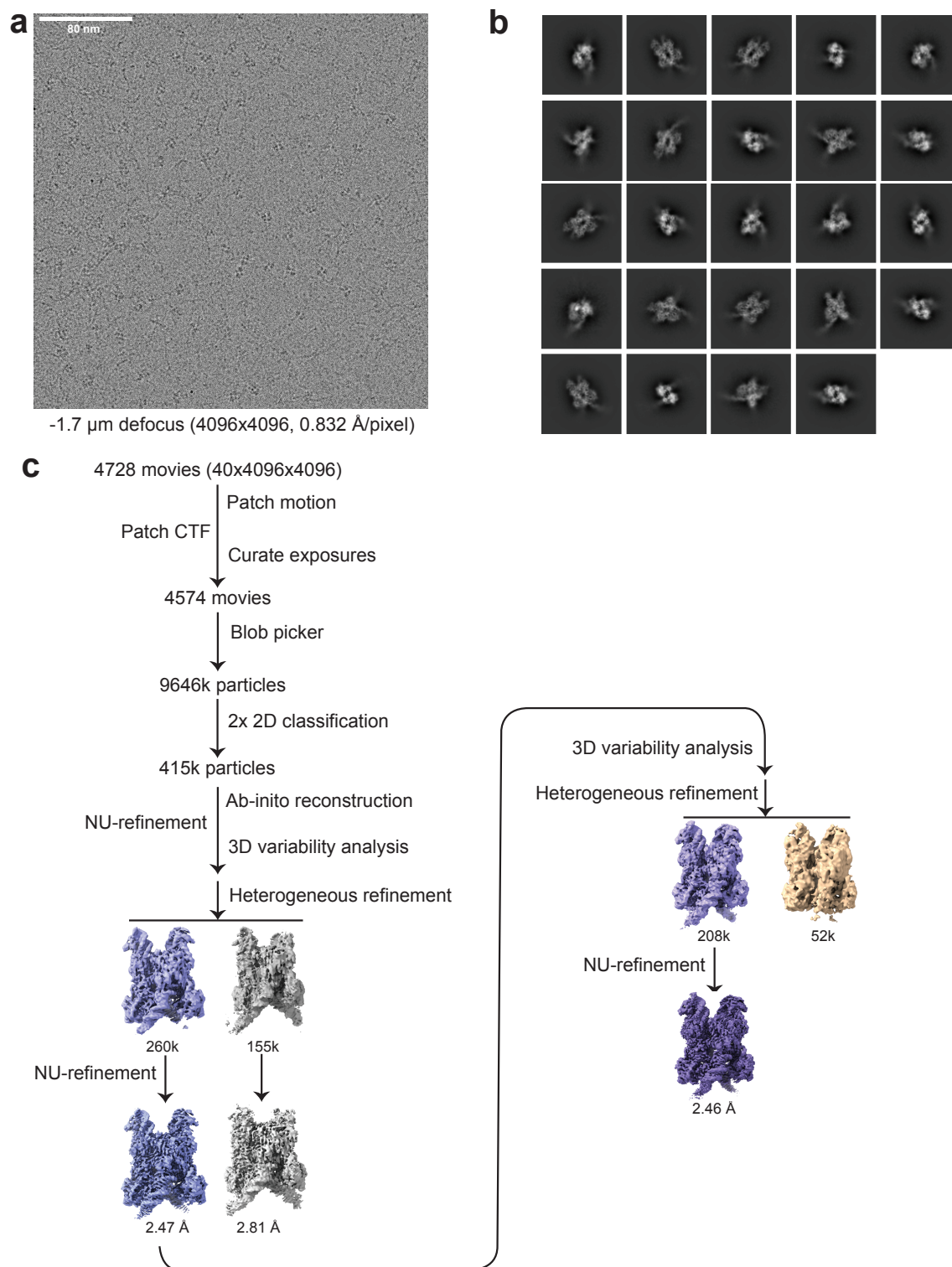


**Supplementary Figure 3.- Low resolution reconstruction of the shTnsB bound to RE/LE in the precatalytic state.** **a)** Representative cryo-EM micrograph (out of a total of 738 micrographs) of this protein-DNA complex in vitreous ice on an UltrAuFoil 1.2/1.3 grid at -2.8  $\mu\text{m}$  defocus. **b)** Reference-free 2D class averages selected for *ab initio* reconstruction displaying a bent DNA molecule bound to two protomers of shTnsB. **c)** 3D reconstruction based on selected 2D classes. Two protomers of shTnsB were fitted into the density based on the STC structure and following a similar pattern to the precatalytic structure of ecTnsB<sup>19</sup>. NTD1 and NTD2 are shown in dashed circles indicating where the LTR would be located.

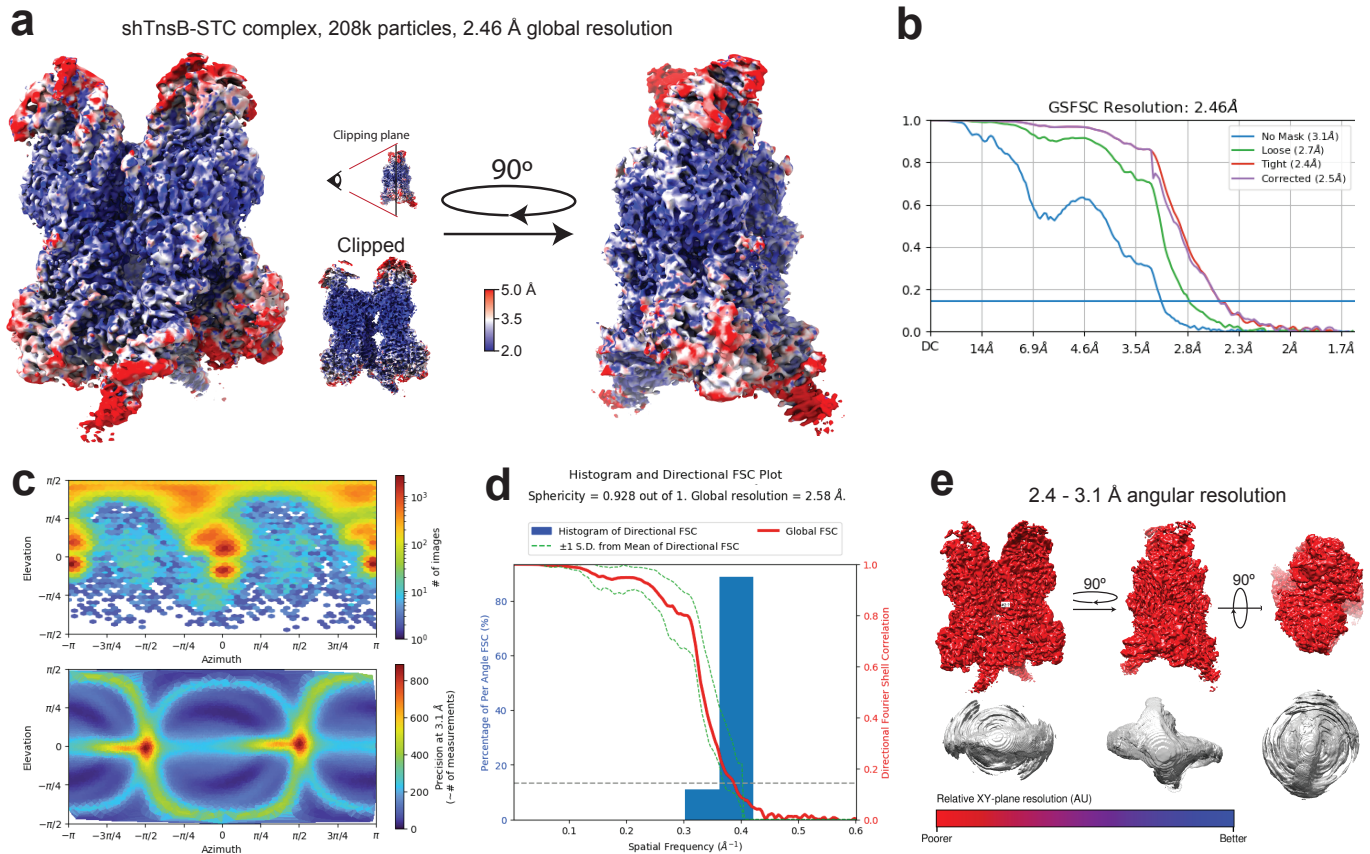
| Oligonucleotide          | Sequence (5'→3')                                                                         |
|--------------------------|------------------------------------------------------------------------------------------|
| RE_Target - Chain I (TS) | ATAAGGATTTTACTGATGACAATAATTTGTCA<br>CAACGACATATAATTAGTCACTGTACACGTAG<br>AGACGTAGCAATGCT  |
| RE_polyA - Chain J (NTS) | AAAAAAAAAAAAAAAAATGTACAGTGACTAATTA<br>TATGTCGTTGTGACAAATTATTGTCATCAGTA<br>AAATCCTTAT     |
| Target_1 - Chain K       | AGCATTGCTACGTCT                                                                          |
| LE_Target - Chain L (TS) | AATTAAATAGTCACAATGACATTAATCTGTCA<br>CCGACGACAGATAATTTGTCACTGTACACTAC<br>GCCTTTTGTGGAGATG |
| LE_polyA - Chain M (NTS) | AAAAAAAAAAAAAAAAATGTACAGTGACAAATTA<br>TCTGTCGTCGGTGACAGATTAATGTCATTGTG<br>ACTATTTAATT    |
| Target_2 - Chain N       | CATCTCCACAAAAGG                                                                          |



**Supplementary Figure 4.- Overview of the shTnsB-STC DNA reconstitution process.** Sequential mixing of the ssDNA corresponding to transferred strands (TS) and non-transferred strands (NTS) from RE and LE was performed in three steps before adding shTnsB. The upper table shows the sequence of each ssDNA and their specific name following the nomenclature in Supplementary Table 1. Colors follow the same code used in Figure 1 and Supplementary Figure 7.



**Supplementary Figure 5.- Single Particle Cryo-EM Analysis of the TnsB-STC.** (a) Representative cryo-EM micrograph (out of a total of 4728 micrographs) of the TnsB-STC in vitreous ice on an UltrAuFoil 1.2/1.3 grid at -1.7  $\mu\text{m}$  defocus. (b) Reference-free 2D class averages selected for *ab initio* reconstruction. (c) Overview of the cryo-EM data processing workflow as performed in cryoSPARC.



**Supplementary Figure 6.- Resolution Assessment and Validation of the TnsB-STC Cryo-EM map (a)** Local resolution maps in two orientation related by a 90° rotation around the y-axis. The inset shows the high resolution obtained in the central cut through the map, the clipping plane indicated above. **(b)** Gold standard Fourier shell correlation (GSFSC) curve after FSC-ask auto-tightening. **(c)** Orientation distribution plot (top) and posterior position directional distribution plot (bottom). **(d)** Histogram and directional FSC plot. Histogram (blue) shows the percentage of per angle FSC at the given spatial frequency. Global FSC (red) and  $\pm 1$  S.D. from mean directional FSC (green). Sphericity equals 0.928 out of 1. **(e)** Angular resolution map, and corresponding directional FSC volume (grey), in three perpendicular orientations. The relative angular resolution is indicated by the color bar.

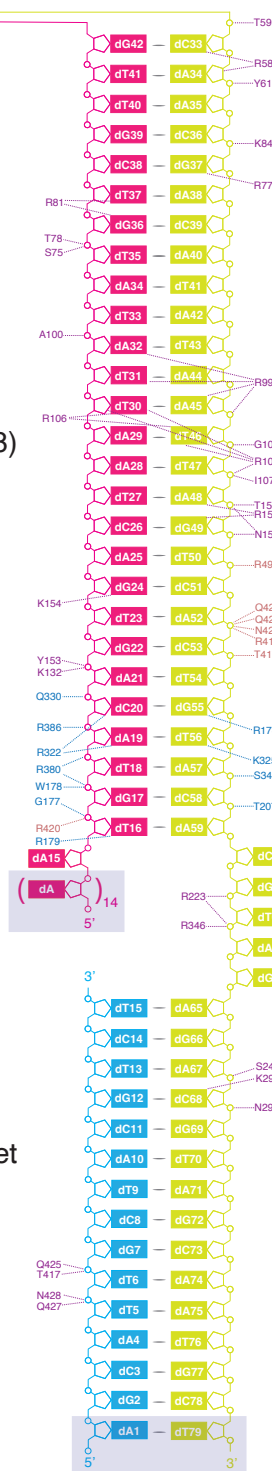


SR(5)



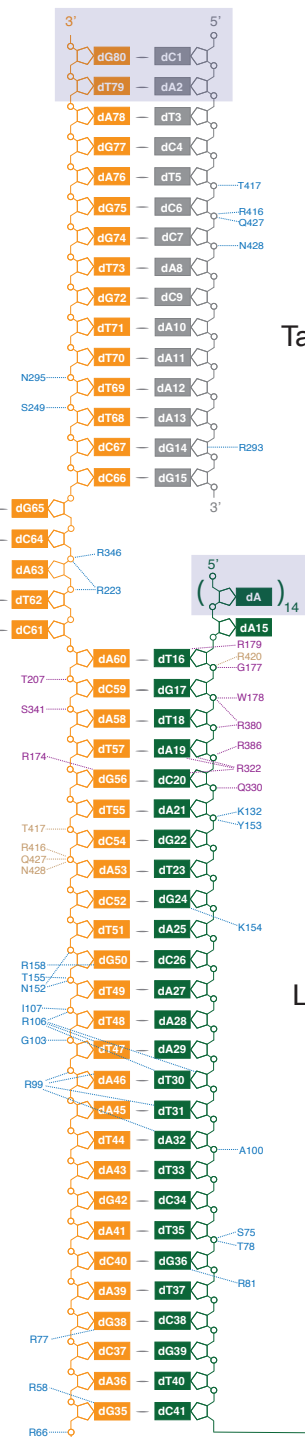
LTR(8)

Target

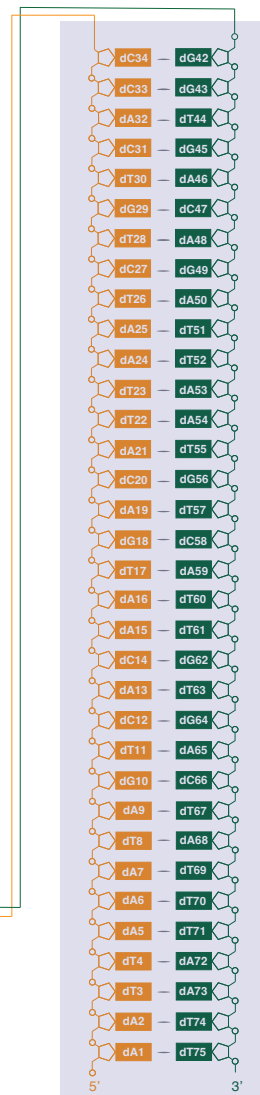


Target

LTR(6)

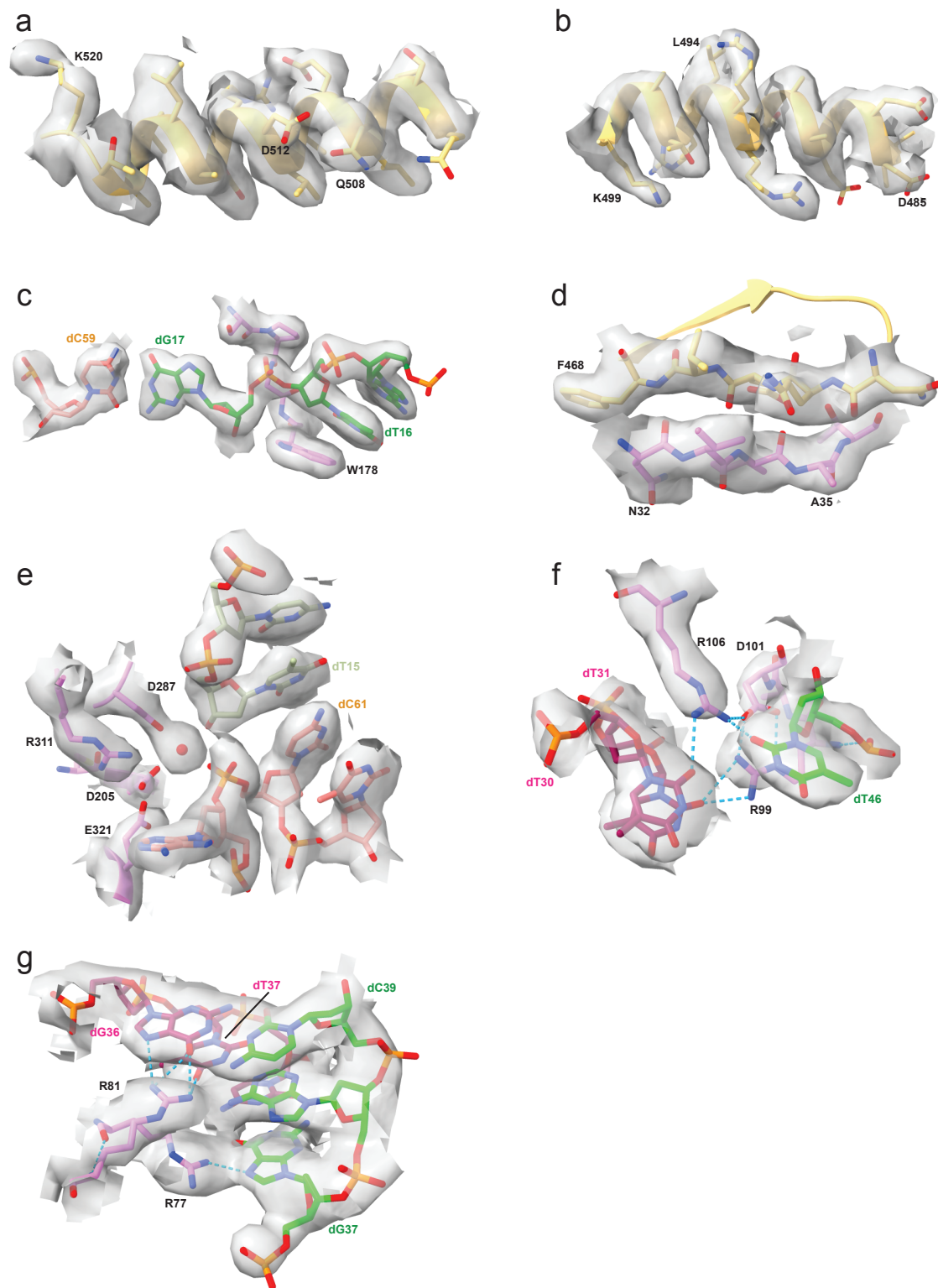


SR(1)



**Supplementary Figure 7.- Protein-nucleic acids interactions in the shTnsB-STC complex.**

Polar contacts of the nucleic acids with the protein side and main chain are indicated. The colors of each chain correspond to the same colours used in Fig. 1c-1e and Supplementary Table 1. Lines between nucleotide bases indicate pairing. Note that the central nucleotides in the attachment site (dT62 and dA63) are not paired. Given that the attachment site contains 5 base pairs (an odd number), the DNA distortion caused by STC formation forces the central base pair to melt. The regions inside the blue squares are not visible in the cryo-EM map.



**Supplementary Figure 8.- Fit of ShTnsB-STC atomic model to the cryo-EM map.** Map regions are displayed at a 2 Å range within the displayed protein residues and nucleotides, at contour level 0.04 for protein-only panels, and level 0.1 for protein/DNA panels. Key residues and nucleotides are labelled. **(a)** OD  $\alpha$ 2, residues 504-520. **(b)** OD  $\alpha$ 1, residues 484-499. **(c)** Stacking interaction between W178, from the NTD2-DDM loop region, and dT16, from the overhand of the LE\_polyA strand. **(d)**  $\beta$ -strands interaction, residues 32-36 and 464-468 of two neighboring chains, connecting the N-terminus with the MD domain of respective chains. **(e)** The DDE catalytic pocket and DNA, including the putative water molecule. **(f)** Polar interactions

of R99 and R106 from the NTD1-NTD2 linker region with DNA. **(g)** Polar interactions of the NTD1 domain R77 and R81 with DNA.



## **Supplementary Movie Legends**

**Supplementary Movie 1.- Conformational variability of the shTnsB-STC complex.** The 3D variability analysis of the structure using cryoSPARC, shows that the core of the protein-DNA complex is quite stable while the arms of the X shaped DNA are highly flexible.

## Supplementary Tables

**Supplementary Table 1. Oligonucleotides used in this study.** The bold nucleotides in the DNAs used to reconstitute the shTnsB-STC complex for cryo-EM denote the sequences visualized in the structure (see also Supp. Fig. 7)

| Oligonucleotide        | ssDNA or dsDNA | Sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Experiment | Synthesis |
|------------------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----------|
| RE_Target – Chain I TS | ssDNA          | ATAAGGATTTTACTGATGACAATAATTTGT<br><b>CACAACGACATATAATTAGTCACTGTACAC</b><br>GTAGAGACGTAGCAATGCT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Cryo-EM    | IDT       |
| RE_polyA – Chain J NTS | ssDNA          | AAAAAAAAAAAAAAAA <b>ATGTACAGTGACTAAT</b><br><b>TATATGTCGTTGTGACAAATTATTGTCATC</b><br>AGTAAAATCCTTAT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Cryo-EM    | IDT       |
| Target_1 – Chain K     | ssDNA          | <b>AGCATTGCTACGTCT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Cryo-EM    | IDT       |
| LE_Target – Chain L TS | ssDNA          | AATTAAATAGTCACAATGACATTAATCTGT<br><b>CACCGACGACAGATAATTTGTCACGTGACA</b><br><b>CTACGCCTTTTGTGGAGATG</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Cryo-EM    | IDT       |
| LE_polyA – Chain M NTS | ssDNA          | AAAAAAAAAAAAAAAA <b>ATGTACAGTGACAAAT</b><br><b>TATCTGTCGTCGGTGACAGATTAATGTCAT</b><br>TGTGACTATTTAATT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Cryo-EM    | IDT       |
| Target_2 – Chain N     | ssDNA          | <b>CATCTCCACAAAAGG</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Cryo-EM    | IDT       |
| LTR(6)-SR(1)           | dsDNA          | AATAGTCACAATGACATTAATCTGTCA<br>CCGACGACAGATAATTTGTCACGTGACACT<br>ACGCCTTT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | EMSA       | IDT       |
| LE                     | dsDNA          | ATTATATTGATGACATTTAATTTGTCATCA<br>ATTAATTAAGCAACGCTGATGGGTCACGAC<br>GACAATTAAATAGTCACAATGACATTAATC<br>TGTCACCGACGACAGATAATTTGTCACGTG<br>ACACTACGCCCTTTTGTGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | EMSA       | PCR       |
| RE                     | dsDNA          | TGCTACGTCTCTACGTGTACAGTGACTAAT<br>TATATGTCGTTGTGACAAATTATTGTCATC<br>AGTAAAATCCTTATACAGTATAGATTATAG<br>CGCTTTGGCAGTTTTAGCATAACCTCTTTG<br>CAGTGACAAAATAGATGTCGTTGTCCGTGA<br>TTGTGACAAATTAGCTGTCGCTTTGCAAGA<br>TAGGAAAAAGCTTTTGTGTATTTTCATAAT<br>GACAAATTGACTGTGCGCAGGAGGTAA                                                                                                                                                                                                                                                                                                                                                                                                                                                           | EMSA       | PCR       |
| Donor                  | dsDNA          | GTA AACGACGCCAGTGAATTGACGCGTA<br>TTGGGATGGAACAGCATTGCTACGTCTCTA<br>CGTGACAGTGACTAATTATATGTCGTTGT<br>GACAAATTATTGTCATCAGTAAAATCCTTA<br>TACAGTATAGATTATAGCGCTTTGGCAGTT<br>TTAGCATAACCTCTTTGCAGTGACAAAATA<br>GATGTCGTTGTCCGTGATTGTGACAAATTA<br>GCTGTCGCTTTGCAAGATAGGAAAAAGCTT<br>TTGTGTATTTTCATAATGACAAATTGACTG<br>TCGCAGGAGGTAAATAATGGAGAAAAAAT<br>CACTGGATATACCACCGTTGATATATCCCA<br>ATGGCATCGTAAAGAACATTTTGAGGCATT<br>TCAGTCAGTTGCTCAATGTACCTATAACCA<br>GACCGTTCAGCTGGATATTACGGCCTTTTT<br>AAAGACCGTAAAGAAAAATAAGCACAAAGTT<br>TTATCCGGCCTTTATTACATTCTTGCCCG<br>CCTGATGAATGCTCATCCGGAGTTCCGTAT<br>GGCAATGAAAGACGGTGAGCTGGTGATATG<br>GGATAGTGTTACCCCTTGTTACACCGTTTT<br>CCATGAGCAAACGAAACGTTTTTCATCGCT<br>CTGGAGTGAATACCACGACGATTTCCGGCA | EMSA       | PCR       |

|  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |  |
|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|
|  |  | GTTTCTACACATATATTGCAAGATGTGGC<br>GTGTTACGGTGAAAACCTGGCCTATTTCCC<br>TAAAGGGTTTATTGAGAATATGTTTTTCGT<br>CTCAGCCAATCCCTGGGTGAGTTTCACCAG<br>TTTTGATTTAAACGTGGCCAATATGGACAA<br>CTTCTTCGCCCCCGTTTTCACTATGGGCAA<br>ATATTATACGCAAGGCGACAAGGTGCTGAT<br>GCCGCTGGCGATTTCAGGTTTCATCATGCCGT<br>TTGTGATGGCTTCCATGTCGGCAGAATGCT<br>TAATGAATTACAACAGTACTGCGATGAGTG<br>GCAGGGCGGGGCGTAATTTTTTTTAAGGCAG<br>TTATTGGTGCCCTTCTAGAGTTTCAGTAAC<br>TATTAACTTAGGGGTGGGTTGAAAGCAAG<br>TCCTTTTATCCGCTTGTTTTAATTGCTTTG<br>TATAATAATTGCAGAGCATATTATATTGAT<br>GACATTTAATTTGTCATCAATTAATTAAGC<br>AACGCTGATGGGTCACGACGACAATTAAAT<br>AGTCACAATGACATTAATCTGTCACCGACG<br>ACAGATAATTTGTCAGTGTACACTACGCCT<br>TTTGTGGAGATGTCTAAGTTCCATCCCAAT<br>GGCGCGCCGAGCTTGGCGTAATCATGGTCA<br>TAGCTGTTTCCTG |  |  |
|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|

**Supplementary Table2. Cryo-EM processing and model refinement**

|                                           |                |
|-------------------------------------------|----------------|
|                                           | TnsB-STC       |
| EMD-15294                                 | EMDB:          |
| PDB 8AA5                                  | PDB:           |
| <b>Data collection</b>                    |                |
| Electron Microscope                       | Titan Krios G2 |
| Voltage (kV)                              | 300            |
| Electron detector                         | Falcon III     |
| Electron dose (e-/Å <sup>2</sup> )        | 40 (40 frms)   |
| Nominal defocus range (μm)                | -0.8 to -2.0   |
| Pixel size (Å)                            | 0.832          |
| <b>3D Reconstruction</b>                  |                |
| Raw images                                | 4728           |
| Initial particles                         | 9646 K         |
| Final particles                           | 260 K          |
| Map resolution (Å), 0.143 FSC             | 2.46 Å         |
| Model resolution (Å), 0.5 FSC             | 2.8            |
| Map sharpening B-factor (Å <sup>2</sup> ) | -74.5          |
| Map sharpening method                     | DeepEMhancer   |
| 3DFSC sphericity                          | 0.928          |
| Local resolution range (Å)                | ~1.8 - 10      |
| <b>Model composition</b>                  |                |
| Protein residues                          | 1486           |
| Nucleotides                               | 172            |
| Non-hydrogen atoms                        | 15509          |
| Water                                     | 2              |
| <b>Model refinement</b>                   |                |
| <i>ADP (Å<sup>2</sup>)</i>                |                |
| Protein                                   | 51.64          |
| Nucleotide                                | 47.93          |
| Water                                     | 25.59          |
| <i>R.m.s. deviations</i>                  |                |
| Bond length (Å)                           | 0.004          |
| Bond angles (°)                           | 0.580          |
| <i>Ramachandran statistics (%)</i>        |                |
| Favored                                   | 95.91          |
| Allowed                                   | 4.09           |
| Outlier                                   | 0.00           |
| Rotamer outlier (%)                       | 4.62           |
| MolProbity score                          | 2.32           |
| Clash score                               | 12.42          |
| Model vs Data CC (mask)                   | 0.76           |