Article

Target site selection and remodelling by type V CRISPR-transposon systems

https://doi.org/10.1038/s41586-021-04030-z

Received: 22 June 2021

Accepted: 16 September 2021

Published online: 10 November 2021

Check for updates

Irma Querques[1,2], Michael Schmitz[1,2], Seraina Oberli[1], Christelle Chanez[1] & Martin Jinek[1]✉

Canonical CRISPR–Cas systems provide adaptive immunity against mobile genetic elements[1]. However, type I-F, I-B and V-K systems have been adopted by Tn7-like transposons to direct RNA-guided transposon insertion[2–7]. Type V-K CRISPR-associated transposons rely on the pseudonuclease Cas12k, the transposase TnsB, the AAA+ ATPase TnsC and the zinc-finger protein TniQ[7], but the molecular mechanism of RNA-directed DNA transposition has remained elusive. Here we report cryo-electron microscopic structures of a Cas12k-guide RNA–target DNA complex and a DNA-bound, polymeric TnsC filament from the CRISPR-associated transposon system of the photosynthetic cyanobacterium *Scytonema hofmanni*. The Cas12k complex structure reveals an intricate guide RNA architecture and critical interactions mediating RNA-guided target DNA recognition. TnsC helical filament assembly is ATP-dependent and accompanied by structural remodelling of the bound DNA duplex. In vivo transposition assays corroborate key features of the structures, and biochemical experiments show that TniQ restricts TnsC polymerization, while TnsB interacts directly with TnsC filaments to trigger their disassembly upon ATP hydrolysis. Together, these results suggest that RNA-directed target selection by Cas12k primes TnsC polymerization and DNA remodelling, generating a recruitment platform for TnsB to catalyse site-specific transposon insertion. Insights from this work will inform the development of CRISPR-associated transposons as programmable site-specific gene insertion tools.

In prokaryotes, CRISPR–Cas adaptive immune systems target mobile genetic elements, including transposons, for nucleolytic degradation using Cas proteins and CRISPR RNA (crRNA) guides[1,8]. However, several nuclease-deficient type I-F, I-B and V-K CRISPR–Cas systems have been co-opted by Tn7-like transposons to direct transposon DNA insertion into specific target sites[2–7]. In these systems, DNA targeting involves transposon-encoded Cas effectors, either a multisubunit complex lacking Cas3 (termed Cascade) in type I systems[3] or a single catalytically inactive Cas12k protein in type V-K systems[7], and results in transposon insertion at a fixed distance from the target site specified by the crRNA. Cas12k-mediated DNA targeting requires a crRNA guide and a *trans*-activating CRISPR RNA (tracrRNA), and is coupled to the concerted activities of three transposon proteins—the transposase TnsB, the AAA+ ATPase TnsC and the zinc-finger protein TniQ (Fig. 1a)—to direct transposon insertion downstream of the target site. In the prototypical *Escherichia coli* Tn7 transposon[9], the DDE-type TnsB transposase catalyses the 5′-DNA strand breakage and joining reactions required for transposon DNA excision from a donor locus and integration into target DNA[10,11], while TnsC helps recruit TnsB to the target site and stimulates its activity[12–14]. The Tn7 core transposition machinery relies on DNA-binding cofactors TnsD and TnsE to direct transposition to specific DNA sequences[15] and structures[16,17], respectively. In CRISPR-associated transposons, TnsD and TnsE are replaced by the crRNA-guided effectors and the putative DNA binding protein TniQ. DNA targeting by type I-F3 CRISPR-transposon systems has been elucidated through structural studies of Cascade-TniQ complexes[18–21]. However, the mechanistic basis of RNA-guided DNA targeting in type V-K systems remains unknown, hindering the development of crRNA-directed transposition for programmable, homology-independent DNA insertion. Here, to understand target DNA recognition and transpososome assembly, we carried out structural studies of the *S. hofmanni* type V-K CRISPR-associated transposon (ShCAST) system[7], determining cryo-electron microscopy (cryo-EM) structures of a ternary Cas12k-guide RNA–target DNA complex and DNA-bound TnsC.

Target DNA recognition by Cas12k

To understand how Cas12k recognizes its DNA target, we analysed a ternary complex comprising *S. hofmanni* Cas12k, 254-nucleotide single-guide RNA (sgRNA, a crRNA–tracrRNA fusion containing a 24-nucleotide guide segment) and 50-base pair (bp) target double-stranded DNA (dsDNA) using single-particle cryo-EM, obtaining a reconstruction at an overall resolution of 3.0 Å (Fig. 1b–e, Extended Data Figs. 1, 2, Extended Data Table 1). Three-dimensional classification and refinement revealed that the Cas12k complexes exhibit considerable conformational heterogeneity (Extended Data Fig. 2). Whereas the majority of sgRNA nucleotides could be modelled, only a part of the R-loop structure formed by the sgRNA and target dsDNA could be resolved, corresponding to a 9-bp duplex of the target DNA strand (TS) and the non-target DNA strand (NTS) containing the protospacer-adjacent

[1]Department of Biochemistry, University of Zurich, Zurich, Switzerland. [2]These authors contributed equally: Irma Querques and Michael Schmitz. ✉e-mail: jinek@bioc.uzh.ch

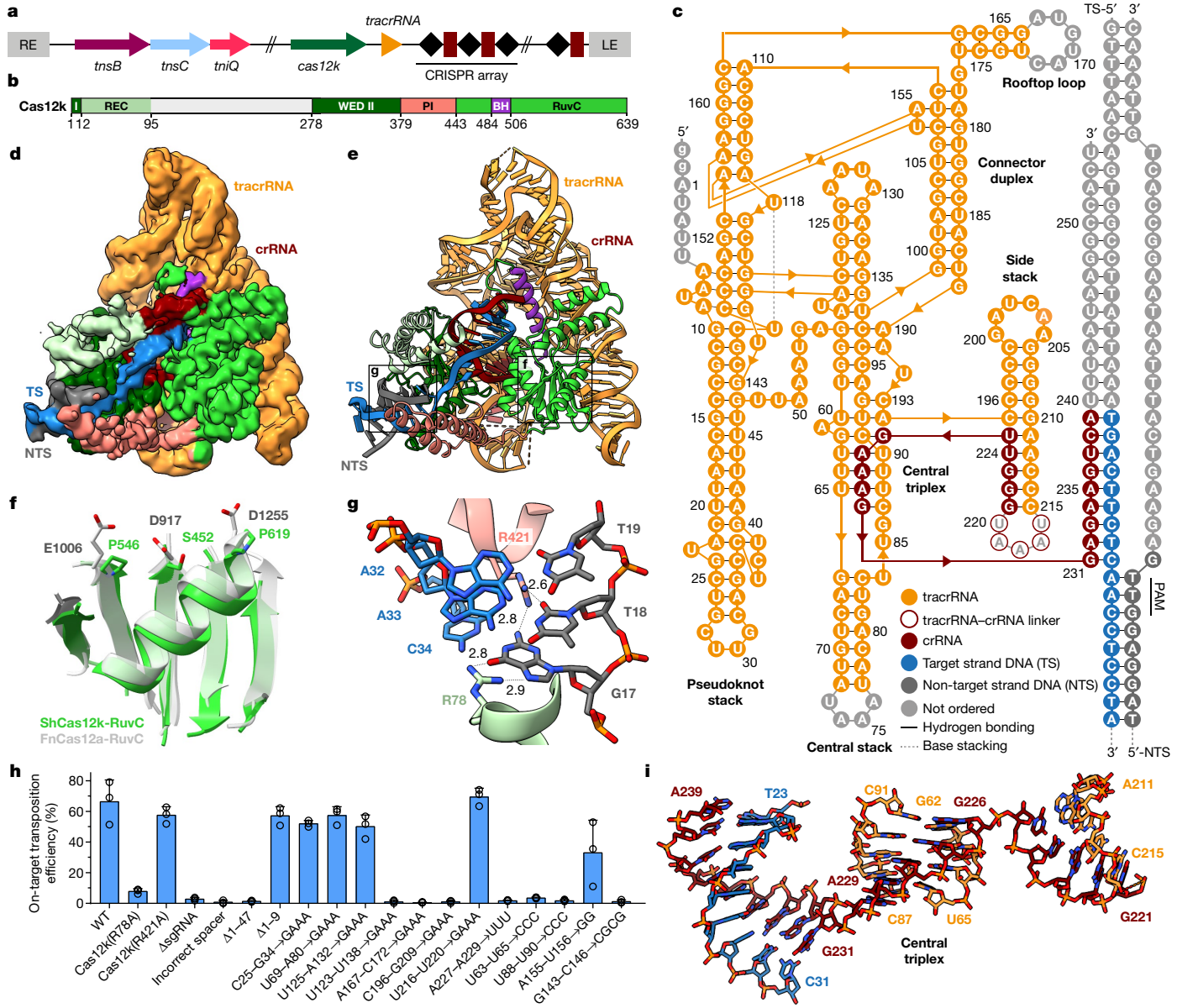


Fig. 1 | Cryo-EM structure of the *S. hofmanni* Cas12k–sgRNA–target DNA complex. a, Schematic of the type V-K CRISPR-associated transposon system from *S. hofmanni*. LE, left transposon end; RE, right transposon end. **b**, Domain organization of Cas12k. BH, bridge helix. **c**, Schematic of the sgRNA structure. **d**, **e**, Electron density map (**d**) and structural model (**e**) of the Cas12k-sgRNA-target DNA complex. **f**, Detail of the inactivated *S. hofmanni* (Sh)Cas12k RuvC domain overlaid with the *Francisella tularensis* subsp. *novicida* U112 (Fn) Cas12a RuvC domain (PDB: 6I1K)[30]. **g**, Detailed view of PAM recognition. Hydrogen bonds are depicted with dotted lines. Distances are indicated in Å. **h**, Site-specific transposition activity in *E. coli* of ShCAST systems containing mutations in Cas12k PAM-interacting residues or sgRNA, as determined by ddPCR analysis. Data are presented as mean ± s.d. ($n = 3$ biologically independent replicates). WT, wild type. **i**, Details of the sgRNA central triplex.

motif (PAM) and a downstream 9-bp heteroduplex formed by the sgRNA and PAM-proximal region of the TS DNA (Fig. 1c).

The molecular architecture of Cas12k comprises the wedge domain (WED), recognition (REC), PAM-interacting (PI) and RuvC domains, similar to other Cas12-family effectors (Fig. 1b, e). The RuvC domain adopts a canonical fold and superimposes closely with the corresponding domains of other Cas12-family proteins (Fig. 1f). However, the Cas12k RuvC active site lacks the catalytic Asp-Glu-Asp motif involved in Mg^{2+} binding, instead featuring Cas12k Ser452, Pro546 and Pro619 (Fig. 1f), confirming that Cas12k is a pseudonuclease that is incapable of cleaving target DNA. The sgRNA–TS DNA heteroduplex is nested in a cleft formed by the REC and RuvC domains and terminates after 9 bp as it contacts the Cas12k bridge helix (Extended Data Fig. 3a, b), suggesting that full sgRNA–TS DNA hybridization requires additional conformational rearrangements in Cas12k.

crRNA-guided DNA recognition in the ShCAST system depends on a 5′-GTN-3′ PAM upstream of the target site[7]. In the ternary complex structure, the PAM-containing DNA duplex is contacted by the WED, REC and PI domains. The G–C base pair at the −3 PAM position is recognized via a bidentate hydrogen bond between the side chain of Cas12k Arg78 and the major groove edge of the guanine base (Fig. 1g). Cas12k Arg421 contributes to PAM recognition through hydrogen bonds to the minor groove edges of the guanine and thymine bases at the −3 and −2 PAM positions, respectively (Fig. 1g). To quantify the effect of Cas12k mutations on transposition efficiency in *E. coli*, we used a droplet digital PCR (ddPCR) assay. The R78A mutation nearly abolished transposition in vivo (Fig. 1h), whereas the R421A mutation did not affect transposition frequency, indicating that Cas12k Arg78 is the major structural determinant of PAM recognition in ShCAST.

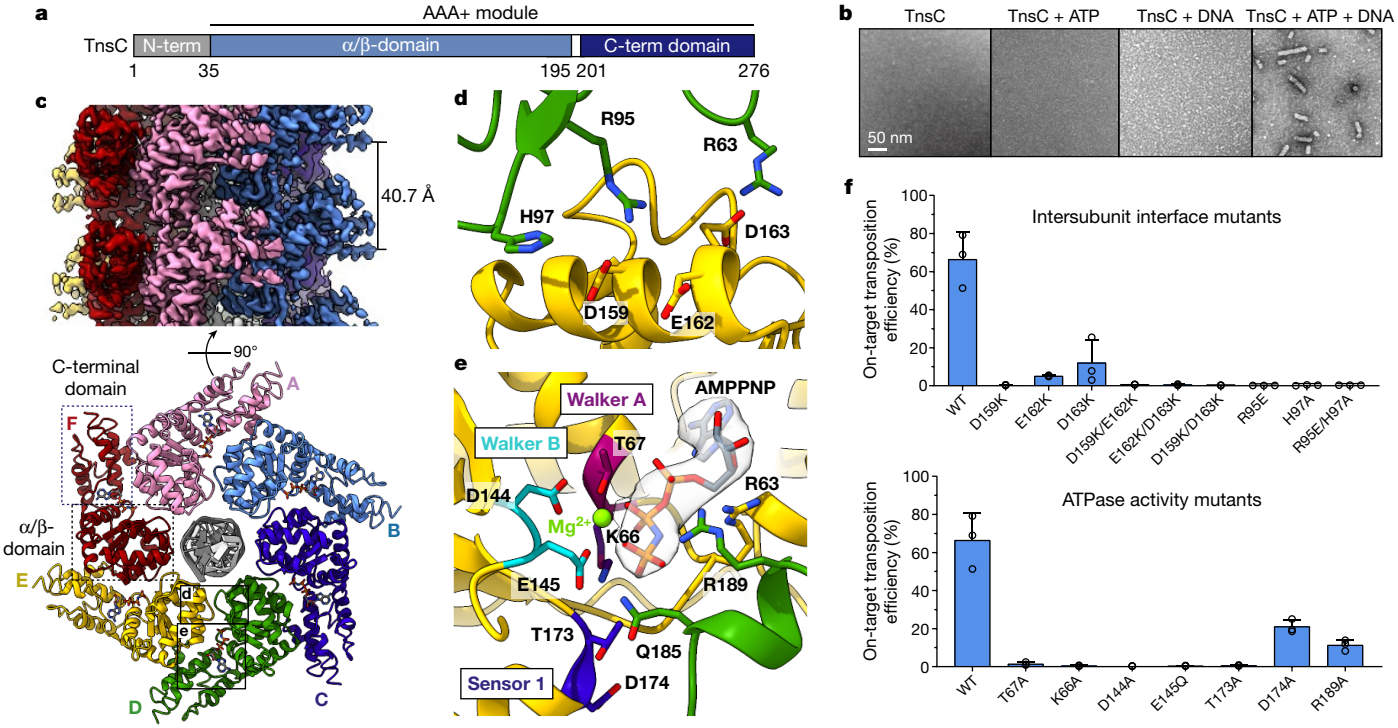


Fig. 2 | Cryo-EM structure of the TnsC–dsDNA filaments. a, Domain organization of TnsC. **b**, Representative negative-stain electron micrographs of TnsC in the presence or absence of dsDNA and/or ATP. Magnification, 68,000×. The experiment was repeated two times independently with similar results. **c**, Cryo-EM density map and structural model of the dsDNA- and AMPPMP-bound TnsC filament at 3.6 Å overall resolution (side and top views). Protein subunits are differentiated by colour. dsDNA is shown in grey. **d**, **e**, Detailed views of the inter-protomer interface and of the ATPase catalytic site. Density corresponding to AMPPNP is shown (contour level of 4.0σ). **f**, Site-specific transposition activity in *E. coli* of ShCAST systems containing mutations at the inter-protomer interface and in the catalytic pocket of TnsC, as determined by ddPCR analysis. Data are presented as mean ± s.d. (n = 3 biologically independent replicates).

Key functional features of guide RNA

The Cas12k sgRNA has an intricate pseudoknot topology comprising two coaxial duplex stacks that clamp the Cas12k RuvC domain and a smaller side stack containing the crRNA–tracrRNA repeat–anti-repeat duplex (Fig. 1c, Extended Data Fig. 3c). The central element of the sgRNA is a triplex region (nucleotides G226–A229) derived from the crRNA repeat sequence, which contacts the WED domain and connects the repeat–anti-repeat duplex of the sgRNA with the sgRNA–TS DNA duplex (Fig. 1i). The central triplex is conserved in the Cas12e (CasX) sgRNA[22], which is otherwise structurally divergent from the Cas12k sgRNA scaffold (Extended Data Fig. 4). The sgRNA also includes: a triplex region involving the 5′-terminal segment of the tracrRNA (nucleotides A6–A9) forming A-minor interactions with the pseudoknot duplex stack; a pseudoknot duplex between nucleotides G10–C13 and G143–C146; a 'rooftop' stem–loop and a 'connector' duplex that bridge the pseudoknot and central stacks; and a triplex junction featuring ribose-zipper interactions between sgRNA nucleotides C96–C97 and A190–A191 (Fig. 1c, Extended Data Fig. 3c). The tracrRNA provides a scaffold that extensively interacts with the Cas12k WED and RuvC domains to ensure their correct orientation and presents the guide segment of the crRNA for base-pairing interactions with the target DNA.

We tested the effects of structure-based sgRNA mutations on transposition efficiency in vivo using ddPCR. Mutations of the central triplex (nucleotides U88–U90), rooftop loop (A167–C172), pseudoknot duplex (G143–C146) and truncation of the sgRNA 5′-terminal region (nucleotides 1–47) all resulted in substantial reduction or complete loss of transposition (Fig. 1h, Extended Data Fig. 3d). By contrast, substitution of the peripheral stem–loops of the tracrRNA (nucleotides C25–G34, U69–A80, U125–A132 and U216–U220) with GAAA tetraloops did not perturb transposition efficiency (Fig. 1h). These results confirm the role of key sgRNA structural elements in the transposition activity of ShCAST and suggest ways to minimize the sgRNA scaffold for genome-engineering applications.

TnsC forms helical filaments

The TnsC component of type V CRISPR-associated transposons is a putative AAA+ ATPase (Fig. 2a) postulated to link Cas12-mediated target site recognition to transpososome recruitment[7,23]. Transposon-associated AAA+ ATPases such as MuB and IstB are known to multimerize in the presence of ATP[24,25]. To determine whether TnsC behaves similarly, we used negative-stain electron microscopy to directly observe multimerization of purified *S. hofmanni* TnsC. Upon incubation with a linear dsDNA in the presence of ATP (Fig. 2b) or the non-hydrolysable analogue ATPγS (Extended Data Fig. 5a), TnsC formed polymeric filaments with an average length of about 70 nm. Filament formation did not occur in the absence of dsDNA or ATP. Thus, TnsC polymerization is a DNA-dependent process that requires ATP binding but not hydrolysis, similar to the multimerization of other transposon-associated AAA+ ATPases[24,25].

We used cryo-EM to determine the structure of a DNA-bound TnsC filament in the presence of the non-hydrolysable analogue AMPPNP and Mg^{2+} at an overall resolution of 3.6 Å (Fig. 2c–e, Extended Data Fig. 5b–f, Extended Data Table 1). The structure reveals that TnsC forms a right-handed helix with a pitch of 40.7 Å and rise of 6.8 Å, accommodating six TnsC copies per turn (Fig. 2c). The TnsC protomers contain a canonical bipartite AAA+ ATPase module, composed of an α/β-domain comprising a five-stranded β-sheet flanked by α-helices on both sides, and an α-helical C-terminal domain (Fig. 2c, Extended Data Fig. 5d). The

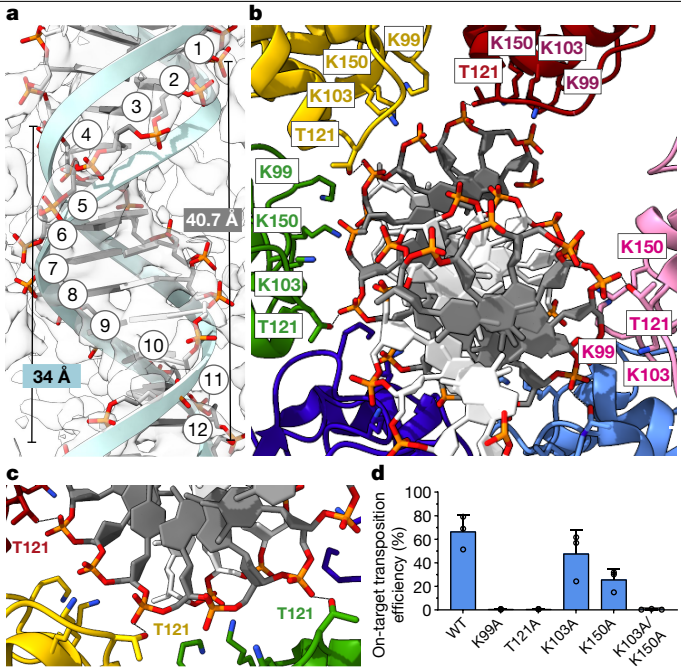


Fig. 3 | DNA binding and remodelling by the TnsC filament. a, Overview of the bound DNA duplex (backbone in stick representation; tracking strand, dark grey; opposite strand, white) and comparison with an ideal B form dsDNA (light blue cartoon backbone). Backbone phosphates of the tracking strand are numbered. **b**, **c**, Close-up views of key DNA-interacting residues. TnsC protomers are shown in distinct colours. Hydrogen bonds are indicated with grey dashed lines. **d**, Site-specific transposition activity in *E. coli* of ShCAST systems containing mutations in the DNA binding interface of TnsC, as determined by ddPCR analysis. Data are presented as mean ± s.d. ($n = 3$ biologically independent replicates).

ATPase active site is located in a cleft between the α/β- and C-terminal domains and contains the canonical AAA+ structural motifs involved in ATP binding and hydrolysis, including Walker A and Walker B (Fig. 2e). The active site is completed by the insertion of the arginine finger TnsC Arg189 and Gln185 from the adjacent TnsC protomer, thereby coupling ATP binding to TnsC polymerization. Helical assembly of TnsC protomers is further enabled by additional salt-bridge interactions involving TnsC Arg63, Arg95 and His97 from one protomer and TnsC Asp159, Asp163 and Glu162 from another (Fig. 2d).

To validate our structural findings, we analysed the polymerization behaviour of TnsC mutant proteins by negative-stain electron microscopy (Extended Data Fig. 6a, b). The E145Q Walker B mutant protein, predicted to be deficient in ATP hydrolysis but not ATP binding, supported filament formation in the presence of ATP or AMPPNP. By contrast, TnsC containing the D174A mutation in the sensor 1 motif was incapable of filament formation, confirming that ATP binding, but not hydrolysis, is required for TnsC polymerization. Mutations of key residues in the ATP binding pocket or at the inter-protomer interface resulted in a substantial reduction or complete loss of transposition in vivo (Fig. 2f). Thus, both TnsC polymerization and ATP hydrolysis by TnsC are critical for the activity of ShCAST.

TnsC filaments assemble on remodelled DNA

The TnsC filament encloses the bound DNA duplex, which is structurally distorted to match the filament helical symmetry. The duplex has 12 bp per helical turn, a pitch of 40.7 Å, a rise of 3.4 Å per bp along the helical axis and a narrower diameter (17.6 Å) compared with canonical B-form DNA (20 Å) (Fig. 3a). The right-handed TnsC filament makes continuous interactions with the deoxyribose-phosphate backbone of only one DNA strand (the tracking strand), with each protomer straddling the duplex minor groove and contacting two consecutive backbone phosphate groups on the tracking strand via a salt bridge with TnsC Lys103 and a hydrogen bond with TnsC Thr121 (Fig. 3b, c). The other DNA strand is not contacted directly by TnsC, however, its backbone phosphate groups are adjacent to TnsC Lys99 and Lys150 (Fig. 3b). T121A and K99A TnsC mutant proteins generated substantially shorter filaments in vitro, whereas filament formation was not perturbed by the K103A mutation (Extended Data Fig. 6c, d). Corroborating these results, the DNA-binding activities of T121A and K99A TnsC mutants were substantially reduced (Extended Data Fig. 6e). TnsC mutations resulting in defective filament formation also abolished or substantially reduced transposition in vivo (Fig. 3d), validating the functional importance of DNA-dependent TnsC polymerization and TnsC-induced DNA remodelling for the transposition activity of type V CRISPR-associated transposons.

TnsB triggers filament disassembly

In Tn7 and Mu transposons, the transposases TnsB and MuA interact with and stimulate the ATPase activity of their cognate AAA+ ATPases[26–28]. To test whether this also occurs in ShCAST, we performed pull-down experiments using affinity-tagged TnsB. In the presence of DNA, TnsB efficiently co-precipitated TnsC when ATP or AMPPNP was also present, but not in the presence of ADP (Fig. 4a). A truncated TnsB protein comprising the catalytic and C-terminal domains was sufficient to mediate the interaction; however, a TnsB construct lacking the C-terminal domain (TnsB residues Leu494–Phe584) was unable to co-precipitate TnsC. Thus, TnsB interacts via its C-terminal region with TnsC filaments but not with monomeric TnsC. Negative-stain electron microscopy revealed that ATP-bound TnsC filaments disassembled in the presence of TnsB (Fig. 4b, Extended Data Fig. 7a, b), whereas AMPPNP-bound filaments remained intact, indicating that TnsB-driven disassembly was dependent on ATP hydrolysis. These results suggest that TnsB interacts with polymerized TnsC and stimulates its ATPase activity, triggering filament disassembly.

TniQ restricts TnsC polymerization

TniQ is required for the transposition activity of ShCAST in vivo[7]. In type I-F CRISPR-transposon systems, a TniQ dimer interacts directly with the crRNA-guided Cascade[18–21], but whether TniQ plays an analogous role in type V systems is unknown. To investigate TniQ function in type V systems, we determined the crystal structure of TniQ at a resolution of 1.3 Å (Fig. 4c, Extended Data Table 2). The structure reveals a conserved domain architecture comprising an amino-terminal helix-turn-helix (HTH) motif followed by two zinc-finger domains (Fig. 4c), as previously observed in the structure of TniQ from the *Vibrio cholerae* type I-F CRISPR-associated transposon[18–21] (Extended Data Fig. 7c). However, unlike the dimeric *V. cholerae* TniQ, *S. hofmanni* TniQ lacks the winged HTH domain involved in dimerization, consistent with the observation that *S. hofmanni* TniQ is monomeric in solution (Extended Data Fig. 7d).

We next tested the interactions between TniQ, Cas12k and TnsC using pull-down assays. We assembled StrepII-tagged Cas12k with an sgRNA and a target DNA and immobilized the resulting complex on Strep-tactin beads. TnsC was efficiently co-precipitated by the Cas12k–sgRNA–target DNA complex in the presence of ATP, but did not bind in the absence of target DNA (Fig. 4d), indicating that the interaction is a result of ATP-dependent TnsC oligomerization on the target DNA. The interaction of TnsC with Cas12k was neither enhanced nor inhibited by the addition of TniQ (Fig. 4d), implying that TniQ is not required for Cas12k-dependent recruitment of TnsC to target DNA. Furthermore, TniQ was co-precipitated by Cas12k only in the presence of DNA-bound TnsC (Fig. 4d). Together, these results suggest a direct interaction between TniQ and TnsC and imply that unlike in

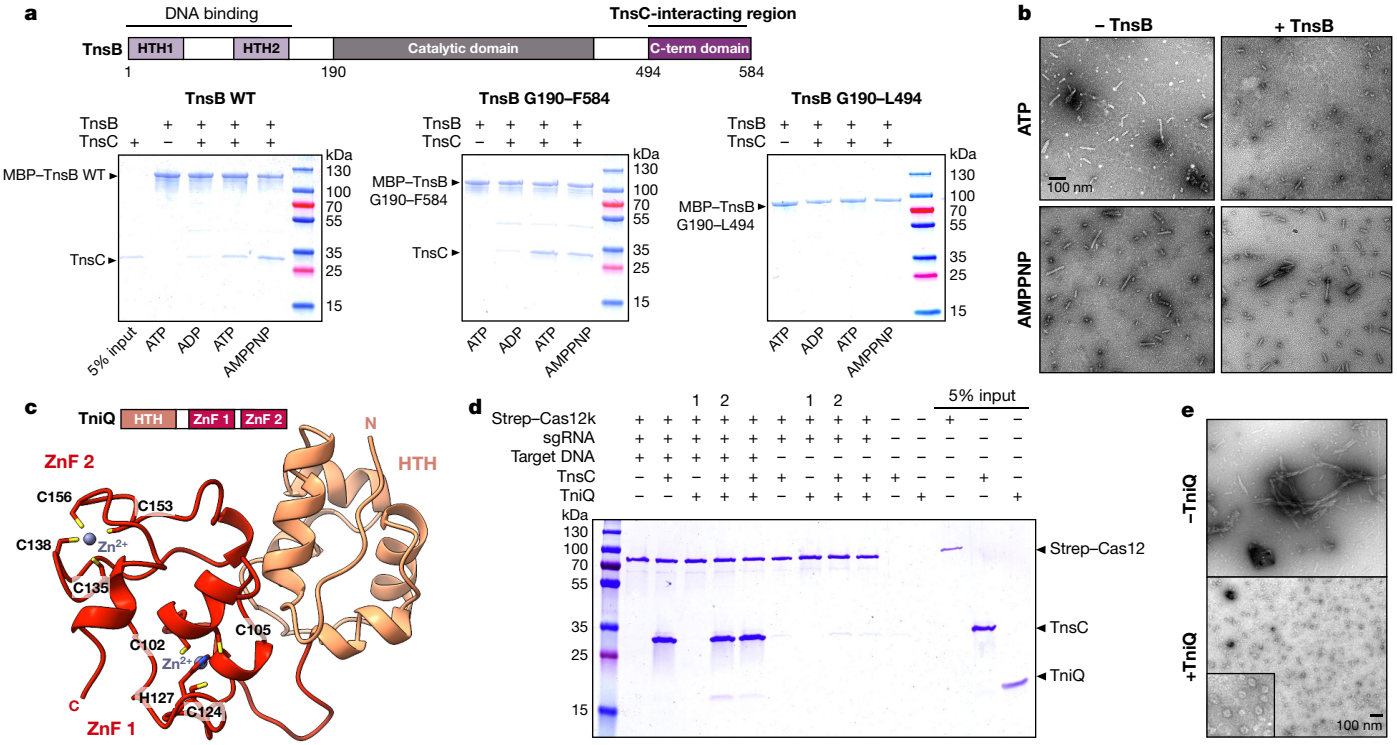


Fig. 4 | Functional interactions and roles of ShCAST components. a, Domain organization of TnsB (top). HTH, helix turn helix motif. Co-precipitation of TnsC by TnsB constructs fused to maltose binding protein (MBP) in the presence of different nucleotides (bottom). **b**, Representative negative-stain EM micrographs of TnsC–dsDNA complexes in the presence of TnsB and different nucleotides. Magnification, 98,000×. **c**, Top, domain architecture of TniQ. ZnF, zinc-finger motif. Bottom, crystal structure of TniQ. Side chains of zinc-coordinating residues are shown as sticks. **d**, Co-precipitation of TnsC and TniQ by immobilized StrepII-fused Cas12k–sgRNA–target DNA complex (1) when TnsC and TniQ components were added together or (2) after sequential addition of TnsC and TniQ components. **e**, Representative negative stain electron micrographs of TnsC in the presence of dsDNA, ATP and TniQ. Magnification of 68,000× for main micrographs, 120,000× for insert. Experiments in **a**, **b**, **d**, **e** were repeated twice independently with similar results. Gel source data are presented in Supplementary Fig. 1.

type I CRISPR-transposon systems, the interaction of TniQ with the CRISPR effector in type V-K systems is indirect. To probe the effect of TniQ on TnsC oligomerization, we examined TnsC filament formation in the presence of TniQ. Incubation of TniQ and TnsC (Fig. 4e) in the presence of ATP and dsDNA yielded substantially shorter TnsC filaments or ring-shaped assemblies, indicating that TniQ inhibits TnsC polymerization.

Cas12k-mediated transposition mechanism

Our structural and functional analysis of the DNA-bound Cas12k effector complex reveals an extensive tracrRNA scaffold that appears to compensate for the compact size of Cas12k, reminiscent of other small type V CRISPR-associated effectors[22]. The Cas12k complex forms an incomplete R-loop structure, consistent with the reported minimal requirements for guide RNA-target complementarity in ShCAST[7]. This draws parallels with the Cascade-TniQ complex from type I-F CRISPR-transposon systems[18] and suggests that recruitment of downstream transposon factors, including TnsC, might allosterically regulate Cas12k to promote R-loop completion, serving as an activation checkpoint. The structure of TnsC filaments shows that ATP-dependent TnsC polymerization occurs on remodelled, underwound DNA, suggesting that DNA unwinding and R-loop formation by Cas12k may help nucleate TnsC filament assembly, although R-loop formation alone is not sufficient for ShCAST activity[7]. On the basis of our findings, we propose a mechanistic model of Cas12k-dependent DNA transposition in type V-K CRISPR-transposon systems (Extended Data Fig. 8), in which target site recognition by Cas12k initiates local TnsC polymerization to generate an interaction platform for the transposase TnsB. TniQ inhibits TnsC polymerization, consistent with a recent study showing that TniQ caps one end of the TnsC filament[29]. In contrast to a previously proposed model[29], our interaction data, together with the observation that TniQ is not essential for the transposition activity of ShCAST in vitro[7], suggest that TniQ might restrict TnsC polymerization to the vicinity of the target site by capping the Cas12k-distal end of the filament. Upon recruitment, TnsB stimulates the ATPase activity of TnsC to trigger filament disassembly, which exposes the target site and enables transposon insertion at a fixed distance from Cas12k, as suggested by the structure of DNA-bound TnsC in the post-hydrolysis state[29]. In sum, this work provides mechanistic insights into type V CRISPR-associated transposons and provides a structural framework for their development as an RNA-guided insertion technology for genome engineering.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04030-z.

1. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237–266 (2013).
2. Faure, G. et al. CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
3. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).

4. Petassi, M. T., Hsieh, S. C. & Peters, J. E. Guide RNA categorization enables target site choice in Tn7-CRISPR–Cas transposons. *Cell* **183**, 1757–1771.e1718 (2020).
5. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR–Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).
6. Saito, M. et al. Dual modes of CRISPR-associated transposon homing. *Cell* **184**, 2441–2453.e2418 (2021).
7. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
8. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary genomics of defense systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
9. Peters, J. E. & Craig, N. L. Tn7: smarter than we thought. *Nat. Rev. Mol. Cell Biol.* **2**, 806–814 (2001).
10. May, E. W. & Craig, N. L. Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996).
11. Sarnovsky, R. J., May, E. W. & Craig, N. L. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996).
12. Choi, K. Y., Spencer, J. M. & Craig, N. L. The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. *Proc. Natl Acad. Sci. USA* **111**, E2858–E2865 (2014).
13. Ronning, D. R. et al. The carboxy-terminal portion of TnsC activates the Tn7 transposase through a specific interaction with TnsA. *EMBO J.* **23**, 2972–2981 (2004).
14. Stellwagen, A. E. & Craig, N. L. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. *Genetics* **145**, 573–585 (1997).
15. Mitra, R., McKenzie, G. J., Yi, L., Lee, C. A. & Craig, N. L. Characterization of the TnsD-attTn7 complex that promotes site-specific insertion of Tn7. *Mob. DNA* **1**, 18 (2010).
16. Parks, A. R. et al. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* **138**, 685–695 (2009).
17. Wolkow, C. A., DeBoy, R. T. & Craig, N. L. Conjugating plasmids are preferred targets for Tn7. *Genes Dev.* **10**, 2145–2157 (1996).
18. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernandez, I. S. Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system. *Nature* **577**, 271–274 (2020).
19. Jia, N., Xie, W., de la Cruz, M. J., Eng, E. T. & Patel, D. J. Structure–function insights into the initial step of DNA integration by a CRISPR–Cas–transposon complex. *Cell Res.* **30**, 182–184 (2020).
20. Li, Z., Zhang, H., Xiao, R. J. & Chang, L. F. Cryo-EM structure of a type I-F CRISPR RNA guided surveillance complex bound to transposition protein TniQ. *Cell Res.* **30**, 179–181 (2020).
21. Wang, B. B., Xu, W. H. & Yang, H. Structural basis of a Tn7-like transposase recruitment and DNA loading to CRISPR–Cas surveillance complex. *Cell Res.* **30**, 185–187 (2020).
22. Liu, J. J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
23. Li, Z. C., N. L. and Peters, J. E. in *Bacterial Integrative Mobile Genetic Elements. Austin* (eds Roberts, A. P. & Mullany, P.) 1–32 (Landes Bioscience, 2013).
24. Arias-Palomo, E. & Berger, J. M. An atypical AAA+ ATPase assembly controls efficient transposition through DNA remodeling and transposase recruitment. *Cell* **162**, 860–871 (2015).
25. Mizuno, N. et al. MuB is an AAA+ ATPase that forms helical filaments to control target selection for DNA transposition. *Proc. Natl Acad. Sci. USA* **110**, E2441–E2450 (2013).
26. Stellwagen, A. E. & Craig, N. L. Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* **16**, 6823–6834, (1997).
27. Greene, E. C. & Mizuuchi, K. Target immunity during Mu DNA transposition. Transpososome assembly and DNA looping enhance MuA-mediated disassembly of the MuB target complex. *Mol. Cell* **10**, 1367–1378 (2002).
28. Greene, E. C. & Mizuuchi, K. Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. *EMBO J.* **21**, 1477–1486 (2002).
29. Park, J. U. et al. Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021).
30. Swarts, D. C. & Jinek, M. Mechanistic insights into the *cis*- and *trans*-acting DNase activities of Cas12a. *Mol. Cell* **73**, 589–600.e584 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Plasmid DNA constructs

The DNA sequences of *S. hofmanni* Cas12k (WP_029636312.1), TnsC (WP_029636336.1), TniQ (WP_029636334.1) and TnsB (WP_084763316.1) proteins were codon optimized for heterologous expression in *E. coli* and synthesized by GeneArt (Thermo Fisher Scientific). The *ShCas12k* gene was inserted into the 1B (Addgene 29653) and 1R (Addgene 29664) plasmids using ligation-independent cloning (LIC), resulting in constructs carrying an N-terminal hexahistidine tag followed by a tobacco etch virus (TEV) protease cleavage site and a N-terminal hexahistidine-StrepII tag followed by a TEV cleavage site, respectively. The *ShTnsB* gene and truncated derivatives were inserted by LIC into 1B (Addgene 29653) and 1C (Addgene 29659) plasmids, generating constructs carrying an N-terminal hexahistidine tag and a hexahistidine–maltose-binding protein (MBP) tag followed by a TEV cleavage site. The *ShTnsC* gene was inserted into the 1S (Addgene 29659) plasmid to produce a construct carrying a N-terminal hexahistidine and hexahistidine-SUMO tag followed by a TEV cleavage site and *ShTniQ* was cloned into a 1C vector. Point mutations were introduced by Gibson assembly using gBlock gene fragments synthetized by IDT or annealed oligonucleotides provided by Sigma as inserts. The pDonor (Addgene 127921), pHelper (Addgene 127924) and pTarget (Addgene 127926) plasmids[7] used in droplet digital PCR experiments were sourced from Addgene. The PSP1-targeting spacer was cloned into pHelper by Gibson assembly, yielding pHelper-PSP1. Mutations in the sgRNA or in the sequence of the *ShCas12k* and *ShTnsC* genes were introduced into the pHelper plasmid by Gibson assembly. Plasmids were cloned and propagated in Mach I cells (Thermo Fisher Scientific) with the exception of pHelper, which was grown in One Shot PIR1 cells (Thermo Fisher Scientific). Plasmids were purified using the GeneJET plasmid miniprep kit (Thermo Fisher Scientific) and verified by Sanger sequencing. All sequences of primers, gBlocks and oligonucleotides used for cloning are provided in Supplementary Table 1.

Protein expression and purification

For expression of Cas12k constructs, hexahistidine-Strep II-tagged and hexahistidine-tagged Cas12k proteins were expressed in *E. coli* BL21 Star (DE3) cells. Cell cultures were grown at 37 °C with shaking at 100 rpm until they reached an optical density at 600 nm of 0.6, and protein expression was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and continued for 16 h at 18 °C. Cells were collected and resuspended in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 1 μg ml^{-1} pepstatin, 200 μg ml^{-1} AEBSF and lysed in a Maximator cell homogenizer at 2,000 bar and 4 °C. The lysate was cleared by centrifugation at 40,000*g* for 30 min at 4 °C and applied to two 5-ml Ni-NTA cartridges connected in tandem. The column was washed with 50 ml of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole before elution with 25 ml of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole. Elution fractions were pooled and dialysed overnight against 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM EDTA and 1 mM DTT in the presence of TEV protease. Dialysed proteins were loaded onto a 5-ml HiTrap Heparin HP column (GE Healthcare) and eluted with a linear gradient of 20 mM HEPES-KOH pH 7.5, 1 M KCl. Elution fractions were pooled, concentrated using 30,000 molecular weight cut-off centrifugal filters (Merck Millipore) and further purified by size-exclusion chromatography using a Superdex 200 (16/600) column (GE Healthcare) in 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT yielding pure, monodisperse proteins. Purified proteins were concentrated to 10–15 mg ml^{-1}, flash frozen in liquid nitrogen and stored at −80 °C until further use.

Expression of wild-type TnsC, TnsC mutants, TnsB, TnsB G190-F584, TnsB G190L494 and TniQ was performed in *E. coli* BL21 Rosetta2 (DE3) cells. Cells were grown in LB medium until reaching an optical density at 600 nm of 0.6 and expression was induced by addition of 0.4 mM IPTG. Proteins were expressed at 18 °C for 16 h. For TnsC, the cells were collected and resuspended in lysis buffer containing 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 10 mM imidazole supplemented with EDTA-free protease inhibitor (Roche). The cell suspension was lysed by ultrasonication and the lysate was cleared by centrifugation at 40,000*g* for 40 min. Cleared lysate was applied to a 5-ml Ni-NTA cartridge (Qiagen). The column was washed in two steps with lysis buffer supplemented with 25 mM and 100 mM imidazole, and bound proteins were eluted with 25 ml of same buffer supplemented with 500 mM imidazole pH 7.5. Eluted proteins were dialysed overnight against 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT in the presence of TEV protease. The proteins were further purified using a 5 ml HiTrap HP Heparin column (GE Healthcare) and eluted with a buffer containing 20 mM Tris-HCl pH 7.5, 700 mM NaCl, 5% glycerol and 1 mM DTT. The eluted fraction was concentrated and further purified by size-exclusion chromatography using an S200 increase (10/300 GL) column (GE Healthcare) in 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM DTT, yielding pure, monodisperse proteins. Purified TnsC was concentrated to 1–2 mg ml^{-1} using 30,000-kDa molecular weight cut-off centrifugal filters (Merck Millipore) and flash-frozen in liquid nitrogen. All mutant TnsC proteins were purified using the same protocol as for the wild-type protein.

For TnsB, cells were collected and resuspended in lysis buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol and 5 mM imidazole supplemented with EDTA-free protease inhibitor (Roche) and lysed by ultrasonication. Lysate was clarified by centrifugation at 40,000*g* for 40 min. The lysate was applied to a 5-ml Ni-NTA cartridge (Qiagen) and the column was washed with lysis buffer supplemented with 25 mM imidazole. The protein was eluted with 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 200 mM imidazole. Eluted proteins were dialysed overnight against 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM DTT in the presence of TEV protease and subsequently purified using a 5-ml HiTrap HP Heparin column (GE Healthcare), eluting with a buffer containing 20 mM Tris-HCl pH 8.0, 400 mM NaCl, and 1 mM DTT. The eluted fraction was concentrated and further purified by size-exclusion chromatography using a Superdex (16/600) column (GE Healthcare) in 20 mM Tris-HCl pH 8.0, 400 mM NaCl, and 1 mM DTT. Purified TnsB was concentrated to 7 mg ml^{-1}, and flash-frozen in liquid nitrogen. For the TnsB G190-F584 and TnsB G190-L494 constructs, cells were collected and resuspended in 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 5 mM imidazole supplemented with EDTA-free protease inhibitor (Roche) and lysed by ultrasonication. Cleared lysate was applied to a 5-ml Ni-NTA cartridge (Qiagen) and the column was washed in two steps with lysis buffer supplemented with 25 and 50 mM imidazole. The proteins were eluted with 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 125 mM imidazole and dialysed overnight against 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 4 mM β-mercaptoethanol in the presence of TEV. Dialysed proteins were passed through a 5-ml Ni-NTA cartridge and washed with buffer containing 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol and 25 mM imidazole, and further purified by size-exclusion chromatography using an Superdex (16/600) column (GE Healthcare) in 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM DTT. Purified TnsB proteins were concentrated to 13–24 mg ml^{-1} and flash-frozen in liquid nitrogen. For the purification of the MBP-tagged constructs used in pull-down experiments (MBP–TnsB, MBP–TnsB G190F584 and MBP–TnsB G190L494), eluted fractions were set aside after the first step of Ni affinity chromatography and directly loaded on a Superdex 200 Increase (10/300 GL) column (GE Healthcare) and eluted in 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 1 mM DTT. Appropriate fractions were concentrated to 0.7–3.0 mg ml^{-1} using 30,000-kDa molecular weight cut-off centrifugal filters (Merck Millipore) and flash-frozen in liquid nitrogen.

For TniQ, cells were collected and resuspended in 20 mM Tris-HCl pH 7.8, 500 mM NaCl, 5% glycerol and 5 mM imidazole supplemented with EDTA-free protease inhibitor (Roche) and lysed by ultrasonication. The

cleared lysate was applied to a 5-ml Ni-NTA cartridge (Qiagen) and the column was washed in 2 steps with lysis buffer supplemented with 25 and 50 mM imidazole. The protein was eluted with lysis buffer supplemented with 300 mM imidazole. Eluted protein was dialysed overnight against 20 mM Tris-HCl pH 7.8, 500 mM NaCl, 1 mM DTT in the presence of TEV protease. Dialysed protein was passed through a 5-ml MBPTrap column (GE Healthcare). The flow-through fraction was concentrated and further purified by size-exclusion chromatography using a Superdex 200 (16/600) column (GE Healthcare) in 20 mM Tris-HCl pH 7.8, 250 mM NaCl, 1 mM DTT. Size-exclusion chromatography allowed to estimate the oligomeric state of the protein (monomer), as shown in Extended Data Fig. 7d. Purified TniQ was concentrated to 10 mg ml^{-1} using 10,000-kDa molecular weight cut-off centrifugal filters (Merck Millipore) and flash-frozen in liquid nitrogen.

sgRNA preparation

DNA sequence encoding the T7 RNA polymerase promoter upstream of the *ShCas12k* sgRNA was sourced as a gBlock (IDT), cloned into a pUC19 plasmid using restriction digest with BamHI and EcoRI, and confirmed by Sanger sequencing. The sequence encoding the T7 RNA polymerase promoter and sgRNA was amplified by PCR and purified by ethanol precipitation for use as template for in vitro transcription with T7 RNA polymerase as described previously[31]. The transcribed RNA was gel purified, precipitated with 70% (v/v) ethanol, dried and dissolved in nuclease-free water.

TnsC filament formation

For the reconstitution of TnsC filaments, wild-type TnsC protein was diluted to a final concentration of 0.5 μM in a buffer containing 20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and mixed at a ratio of 1:5 with a double-stranded 92-bp duplex oligonucleotide (Supplementary Table 1) in a 24-μl reaction volume. Protein and DNA were incubated for 10 min at 25 °C. Upon addition of nucleotide (ATP, ATPγS or AMPPNP) at a final concentration of 1 mM, reactions were further incubated for 20 min at 37 °C. To analyse filament assembly in the presence of TnsB, TnsC filaments were formed using the same procedure described above, followed by addition of TnsB at a final concentration of 10 μM (1:2 ratio TnsC:TnsB) in a 60-μl final reaction volume. To probe filament formation in the presence of TniQ, TnsC was diluted to a final concentration of 3 μM in a buffer containing 20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and 1 mM ATP and mixed with a double-stranded 69-bp duplex oligonucleotide (Supplementary Table 1) at a ratio of 1:25 (DNA:TnsC) in a 20-μl reaction volume. TniQ was added to the reactions at a final concentration of 6 μM and incubated for 1 h at 37 °C. Prepared samples were then analysed by negative-stain electron microscopy.

Negative-stain electron microscopy

Samples (4 μl) were applied to glow-discharged continuous carbon film supported copper grids (CF300-CU-50 grids, Electron Microscopy Sciences). After 1 min incubation, the grids were washed once with 4 μl of 2% (w/v) uranyl formate and stained in a 4-μl drop of 2% uranyl formate for 1 min prior to blotting. Negatively-stained specimens were imaged using a FEI Tecnai G2 Spirit transmission electron microscope operated at an acceleration voltage of 120 kV. Data were acquired at a nominal magnification of 68,000×, 98,000×, 120,000× or 180,000× as indicated in figure legends. The length of the filaments was measured manually with the Gatan Digital micrograph software.

Transposition assay and droplet digital PCR analysis

Transposition assay were conducted as in ref. [6], with minor modifications. All transposition experiments were performed in One Shot PIR1 *E. coli* cells (Thermo Fisher Scientific). The strain was first co-transformed with 20 ng each of pDonor and pTarget, and transformants were isolated by selective plating on double antibiotic LB-agar plates. Liquid cultures were then inoculated from single colonies, and the resulting strains were made chemically competent using standard methods, aliquoted and snap frozen. pHelper plasmids (20 ng) were then introduced in a new transformation reaction by heat shock, and after recovering cells in fresh LB medium at 37 °C for 1 h, cells were plated on triple antibiotic LB-agar plates containing 100 μg ml^{-1} carbenicillin, 50 μg ml^{-1} kanamycin and 33 μg ml^{-1} chloramphenicol. After overnight growth at 37 °C for 16 h, colonies were collected from the plates, resuspended in 15 μl lysis buffer (TE with 0.1% Triton X 100) and heated for 5 min at 95 °C. Water (60 μl) was added to the samples before centrifugation for 10 min at 16,000*g*. The supernatant was then transferred and the nucleic acid concentration adjusted to 0.3 ng μl^{-1}. The isolated DNA (0.75 ng) was used for subsequent analysis by droplet digital PCR (ddPCR). A mixture of five primers (900 nM final concentration each), two probes (250 nM final each) (Supplementary Table 1) and template DNA were combined with ddPCR Supermix for Probes (No dUTP) (BioRad) in a final 20-μl reaction volume. Droplets were generated with 70 μl Droplet Generation Oil for Probes (BioRad) using the QX200 Droplet Generator (BioRad). The final sample (40 μl) was transferred to 96-well plates for amplification by PCR (1 cycle, 95 °C, 10 min; 40 cycles, 94 °C, 30 s, 58 °C, 1 min; 1 cycle, 98 °C, 10 min; 4 °C hold). Samples were read with a QX200 Droplet Reader, and data analysed using the QuantaSoft to determine the concentration of inserts and template in each reaction (Abs counting mode). On-target transposition efficiencies were calculated as inserts/(inserts + targets). All ddPCR measurements presented in the text and figures were determined from three independent biological replicates and measured in technical duplicates.

Pull-down experiments

For pull-down experiments using Cas12k as bait, sgRNA was first mixed with hexahistidine-StrepII-tagged Cas12k (Strep-Cas12k) in assembly buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 10 mM $MgCl_2$, 1 mM DTT), and incubated for 20 min at 37 °C to allow complex formation. A dsDNA target (Supplementary Table 1) was then added to the reaction in a Strep-Cas12k:sgRNA:dsDNA molar ratio of 1:1.2:2 and incubated for 20 min at 37 °C. The final 20-μl reaction contained 2.5 μg (at a concentration of 1.6 μM) Strep-Cas12k, 1.9 μM sgRNA, and 3.2 μM DNA in assembly buffer. Samples were mixed with 12.5 μl Strep-Tactin beads (iba) equilibrated in pull-down wash buffer 1 (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.05% Tween20) and incubated 20 min at 4 °C on a rotating wheel. The beads were washed three times with pull-down wash buffer 1 to remove excess nucleic acids. The beads were resuspended in 250 μl of pull-down wash buffer 1 and TniQ and/or TnsC was added in tenfold molar excess together with AMPPNP (1 mM final concentration).

Samples were incubated 20 min at 37 °C, then washed twice with pull-down wash buffer 2 (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.05% Tween20, 1 mM AMPPNP) and eluted with pull-down elution buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 1 mM AMPPNP, 5 mM desthiobiotin). Eluted samples were analysed by SDS–PAGE using Any kDa gradient polyacrylamide gels (Bio-Rad) and stained with Coomassie Brillian Blue.

For pull-downs using TnsB as bait, MBP–TnsB and the truncated derivatives TnsB G190–F584 and TnsB G190–L494 (62.5 pmol) were immobilized on 25 μl amylose agarose beads (New England Biolabs) in pull-down buffer (20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 1 mM DTT) by incubation for 30 min at 4 °C. The beads were resuspended and washed three times with 200 μl of pull-down buffer. In parallel, to allow filament formation, TnsC was first diluted to a final concentration of 8 μM in pull-down buffer and mixed at a molar ratio of 5:1 (TnsC:DNA) with a double-stranded 92-bp DNA duplex (Supplementary Table 1) in a 24-μl reaction volume and incubated for 10 min at 25 °C. Upon addition of nucleotide (ADP, ATP, or AMPPNP) at a final concentration of 1 mM, the 25-μl reactions were further incubated for 20 min at 37 °C. The 25-μl mixture containing TnsC was then added to the amylose beads containing TnsB as bait, followed by incubation

for 15 min at 37 °C. Beads were then washed three times with 200 µl of pull-down buffer. Proteins were eluted from the beads by adding SDS loading buffer, and analysed on a 4–15% gradient polyacrylamide gel (Bio-Rad).

DNA-binding assays

Increasing concentrations of wild-type or mutant TnsC (0 µM, 1.3 µM, 2.5 µM, 8 µM) were incubated with a fluorescently labelled double-stranded 27-bp duplex oligonucletide (20 nM final concentration) (Supplementary Table 1) for 30 min at 37 °C in a buffer containing 20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and 1 mM AMPPNP, in a final volume of 20 µl. Reactions were resolved on a 4–20% native polyacrylamide gel (Bio-Rad) and imaged using a Typhoon scanner (GE Helthcare).

Crystallization and structure determination of TniQ

Crystals of TniQ were grown using the hanging drop vapor diffusion method. TniQ at a concentration of 6.3 mg ml^{-1} was mixed in a 1:1 ratio (1 µl + 1 µl) with reservoir solution containing 100 mM HEPES-KOH pH 7.5, 17–25% PEG6000 and 250 mM NaCl. Crystals were cryoprotected with reservoir solution supplemented with 25% (v/v) ethylene glycol and flash-cooled in liquid nitrogen. Data were collected at a temperature of 100 K at the beamline PXIII of the Swiss Light Source (Paul Scherrer Institute) using the DA+ data collection software and processed with XDS[32]. The crystals belonged to space group $P2_12_12_1$ and contained four copies of TniQ per asymmetric unit. Experimental phases were obtained by the single-wavelength anomalous diffraction method from data measured at a wavelength of 1.0000 Å, utilizing the anomalous signal of Zn^{+2} scatters present in the protein. Initial phases were calculated using Phenix[33] (Autosol[34]) and density modification of experimental phases in AutoSHARP[35] resulted in an interpretable map which was used for initial model building using ARP/wARP[36]. The atomic model was completed by iterative building in Coot[37] and refined with Phenix.refine[38] (Extended Data Table 2).

Cas12k–sgRNA–DNA ternary complex sample preparation and cryo-EM data collection

The TS and NTS DNAs were sourced as synthetic, HPLC-purified oligonucleotides (IDT) and annealed in a 1:1 molar ratio. sgRNA was pre-incubated with Cas12k for 20 min at 37 °C in assembly buffer (20 mM HEPES-KOH pH 7.5, 50 mM KCl, 10 $MgCl_2$, 1 mM DTT). dsDNA was added and the sample was further incubated for 20 min at 37 °C. The final sample contained 47.8 µM Cas12k, 57.4 µM sgRNA and 334.6 µM dsDNA (1:1.2:7 molar ratio) in 20 mM HEPES-KOH pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT. The sample was purified by size-exclusion chromatography using a Superdex 200 (5/150) column. Eluted fractions were analysed by SDS–PAGE on an Any kDa polyacrylamide gel (Bio-Rad, stained with Coomassie) and by denaturing PAGE on a 10% polyacrylamide–7 M urea gel upon proteinase K digestion for 15 min at 37 °C. 2.5 µl of selected elution fractions were applied to a 300-mesh holey carbon grid (Au R1.2/1.3, Quantifoil Micro Tools), blotted for 2 s at 75% humidity, 4 °C, plunge frozen in liquid ethane (using a Vitrobot Mark IV plunger, FEI) and stored in liquid nitrogen. Cryo-EM data collection was performed on a FEI Titan Krios microscope operated at 300 kV equipped with a Gatan K3 direct electron detector using counting mode (EMBL). A total of 10,952 micrographs were recorded at a calibrated magnification of 130,000× with a pixel size of 0.64 Å. Each movie comprises 40 subframes with a total dose of 51.81 e^- $Å^{-2}$. Data acquisition was performed automatically using SerialEM[39] with four shots per hole at −0.9 µm to −1.9 µm defocus (0.1-µm steps).

Image processing and model building of the Cas12k–sgRNA–DNA complex

Cas12k–sgRNA–dsDNA data were processed using cryoSPARC[40]. A total of 10,952 movies were imported and motion-corrected (patch motion correction (multi)). Upon patch CTF estimation (multi), exposures were selected by estimated resolution (better than 3.5 Å) and defocus (<2 µm), yielding 10,866 movies. Particles were picked on denoised micrographs using Topaz[41] with the pretrained model ResNet16 (64 units). Particle picks were inspected and selected particles (3.3 million) extracted in a box size of 300 × 300 pixel (Fourier cropped to 100 × 100 pixel) and subjected to 2D classification. Particle-containing classes were selected (101 classes, 2.2 million particles) and rebalanced to 6 super-classes each containing a maximum of 200,000 particles, yielding a final set of 1.2 million particles. The particle stack was used to calculate ab initio models. Two ab initio models matching the appearance of 2D classes were selected as volumes for heterogeneous refinement with the particle stack, resulting in reconstructions of eight 3D classes. Classes were inspected visually, respective particles re-extracted in unbinned form and subjected to non-uniform refinement using the respective size-corrected volume from the heterogeneous refinement as input. The local resolution was calculated based on the resulting map using the local resolution functionality in cryoSPARC and plotted on the map using UCSF Chimera[42]. The structure model for the Cas12k–sgRNA–target DNA complex was built in Coot[37]. The sgRNA and target DNA were built de novo exclusively. The model was refined in Coot using restraints for the nucleic acids calculated with LibG[43] (base pair, stacking plane and sugar pucker restraints) in ccp4 and refined using Phenix[33,38]. Real-space refinement was performed with the global minimization and atomic displacement parameter (ADP) refinement options selected. Secondary structure restraints, side chain rotamer restraints, and Ramachandran restraints were used. Key refinement statistics are listed in Extended Data Table 1. The final atomic model includes Cas12k residues 1–43, 56–70, 76–94 and 278–636, sgRNA nucleotides 5–72, 77–164, 174–215 and 221–239, 17 nucleotides of TS DNA and 10 nucleotides of NTS DNA. The quality of the atomic model, including basic protein and DNA geometry, Ramachandran plots, clash analysis and model cross-validation, was assessed with MolProbity[44,45] and the validation tools in Phenix. Structural superposition was performed in Coot using the secondary structure matching[46] (SSM) function. Figure preparation for maps and models and calculation of map contour levels was performed using UCSF ChimeraX[47].

TnsC–dsDNA complex cryo-EM sample preparation and data collection

Purified TnsC was diluted to a final concentration of 6.4 µM in a buffer containing 20 mM HEPES-KOH pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and mixed with a 92-nucleotide dsDNA oligonucleotide containing 2 mismatched regions (Supplementary Table 1) at a ratio of 5:1 (TnsC:DNA) in a 21.4-µl reaction volume and incubated for 10 min at 25 °C. Upon addition of AMPPNP at a final concentration of 1 mM, the sample was further incubated for 20 min at 37 °C. Sample vitrification was performed with a Leica EM GP2 plunge freezer at 10 °C and at 70% humidity. A 3.5 µl volume of sample was applied onto a freshly glow-discharged 200-mesh copper 2 nm C R1.2/1.3 grids (Quantifoil Micro Tools). The grids were blotted for 1 s and plunge-vitrified in liquid ethane. Cryo-EM data collection was performed on a FEI Titan Krios G3i microscope (University of Zurich) operated at 300 kV equipped with a Gatan K3 direct electron detector in super-resolution counting mode. A total of 5,055 micrographs were recorded at a calibrated magnification of 130,000× resulting in super-resolution pixel size of 0.325 Å. Each movie comprised 36 subframes with a total dose of 66.39 e^- $Å^{-2}$. Data acquisition was performed with EPU Automated Data Acquisition Software for Single Particle Analysis (ThermoFisher) with three shots per hole at −1.0 µm to −2.4 µm defocus (0.2-µm steps).

Image processing, model building and validation of the TnsC–dsDNA complex

All movies were motion-corrected and dose-weighted with MotionCor2[48]. Aligned, non-dose-weighted micrographs were then used to

estimate the contrast transfer function (CTF) with Gctf[49]. All subsequent image processing steps were performed using helical reconstruction in RELION[50,51]. Approximately 1,100 segments were manually picked in RELION. One round of reference-free 2D classification was performed to produce templates for reference-dependent auto-picking. A limited resolution E-step (low-pass filter) of 10 Å was applied to prevent overfitting. Using the resulting 2D classes as templates, overlapping helical filament segments were automatically picked with an inter-box distance of 10 Å. A total of 1,733,884 segments were extracted with a box size of 352 pixels from the full dataset and subjected to three cycles of reference-free 2D classification to remove low-quality filaments, yielding 743,027 particles. Three iterative cycles of 3D auto-refinement in RELION using a cylinder with a 12-nm diameter as the initial model were conducted to produce initial reconstructions. A spherical mask with a diameter equal to 90% of the box size was applied. Selected 2D classes were then used with the best ab initio model to perform 3D auto-refinement with a local search of helical symmetry (twist 58–62° and rise 6–8 Å). The helical parameters were estimated based on a pitch of 41 Å calculated from the power spectra and assuming a number of 6 units/turns. The central *z*-height parameter in RELION was set to 30%. The output volume (3.94 Å resolution) was used as the reference model for 3D classification. The particles were divided into four classes in 3D classification. A 3D class with the most clearly discernible secondary structure features (59.44° twist and 6.65 Å rise and 77,544 particles) was selected and 3D auto-refined. A single round of Bayesian polishing, further helical refinement and post-processing with a soft mask applied around the filament produced a volume with final reported resolution of 3.57 Å, final twist 59.72° and rise 6.78 Å. The 3D reconstruction was sharpened with a B-factor of −79.50 $Å^{-2}$. The overall map resolution reported in Extended Data Table 1 was derived from Fourier shell correlation (FSC) calculations between reconstructions from two independently refined half-maps, and reported resolutions are based on the gold-standard criterion. Local resolution was estimated with RELION. For model building, a TnsC homology model based on the crystal structure of the DNA replication initiator factor ORC2 (PDB:1W5T)[52] was generated using the Phyre2 server[53] and the two main domains (α/β-domain and C-terminal domain) were manually docked separately as rigid bodies in the cryo-EM density map with UCSF Chimera[42], followed by real space fitting with the Fit in Map function. The atomic model of a TnsC protomer was manually rebuilt in Coot[37]. The coordinated AMPPNP molecules and magnesium ions were manually docked in Coot. The DNA duplex was manually built in Coot and then subjected to refinement in Coot using base pair, stacking plane and sugar pucker restraints generated using LibG[43] in ccp4. The sequence of the dsDNA was randomly selected and is purely representative as any DNA sequence information was lost during helical symmetry averaging. The adjacent subunits in the filament were then generated by applying the helical symmetry from RELION to the rebuilt atomic model. Real-space refinement of the resulting model was performed in Phenix[33,38] with the global minimization and atomic displacement parameter (ADP) refinement options selected. The following restraints were used in real-space refinement: secondary structure restraints, non-crystallographic symmetry (NCS) restraints between the protein subunits, side chain rotamer restraints, and Ramachandran restraints. Key refinement statistics are listed in Extended Data Table 1. The final atomic model includes 7 copies of TnsC (one helical filament turn plus an additional TnsC copy), each bound to a single Mg^{2+} ion and AMPPNP, and a 22-bp DNA duplex. The quality of the atomic model, including basic protein and DNA geometry, Ramachandran plots, and clash analysis, was assessed and validated with MolProbity[44,45] and validation tools as implemented in Phenix. Structural superposition was performed in Coot using the secondary structure matching[46] (SSM) function. Figure preparation for maps and models and calculation of map contour levels was performed using UCSF ChimeraX[47].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Maps and atomic coordinates of the reported cryo-EM structures have been deposited in the Electron Microscopy Data Bank under accession codes EMD-13486 (*S. hofmanni* Cas12k–sgRNA–DNA complex) and EMD-13489 (Tns-DNA filament), and the Protein Data Bank with accession codes 7PLA (Cas12k–sgRNA–DNA complex) and 7PLH (Tns–DNA filament). Structure factors and atomic coordinates of the reported X-ray crystallographic structure of TniQ has been deposited in the Protein Data Bank with accession code 7OXD.

31. Anders, C., Niewoehner, O. & Jinek, M. In vitro reconstitution and crystallization of Cas9 endonuclease bound to a guide RNA and a DNA target. *Methods Enzymol.* **558**, 515–537 (2015).
32. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
33. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
34. Terwilliger, T. C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D* **65**, 582–601 (2009).
35. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
36. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
38. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
39. de la Cruz, M. J., Martynowycz, M. W., Hattne, J. & Gonen, T. MicroED data collection with SerialEM. *Ultramicroscopy* **201**, 77–80 (2019).
40. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
41. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Res. Comput. Mol. Biol.* **10812**, 245–247 (2018).
42. Pettersen, E. F. et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
43. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
44. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
45. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020).
46. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
47. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
48. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
49. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
50. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
51. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
52. Singleton, M. R. et al. Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J. Mol. Biol.* **343**, 547–557 (2004).
53. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

Acknowledgements We thank M. Sawicka, L. Loeff and S. Sorrentino (University of Zurich Center for Microscopy and Image Analysis) for assistance with cryo-EM sample preparation and data collection; the Cryo-Electron Microscopy Service Platform at EMBL Heidelberg for instrument access and F. Weis for assistance with data collection; R. Ciuffa for advice on helical reconstruction; M. Pacesa for help with cryo-EM data processing; F. Boneberg and A. Walter for technical assistance; B. Blattmann at the Protein Crystallization Center (University of Zurich) for assistance with crystallization screening; M. Wang, V. Olieric and T. Tomizaki at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) for assistance with X-ray diffraction measurements; S. Kreutzer and the ETH Genome Engineering and Measurement Lab for assistance with ddPCR assays; G. Riddhough (Life Science Editors) for assistance with manuscript editing. This work was supported by Swiss National Science Foundation Project Grant 31003A_182567 and European Research Council (ERC) Consolidator Grant no. ERC-CoG-820152. I.Q. was supported by FEBS and EMBO (ALTF 296-2020) long-term postdoctoral fellowships. M.S. is a member of the Biomolecular Structure and Mechanism PhD Program of the Life Science Zurich Graduate School. M.J. is an International Research Scholar of the Howard Hughes Medical Institute, and Vallee Scholar of the Bert L. and N. Kuggie Vallee Foundation.

Author contributions I.Q., M.S. and M.J. conceived the study and designed experiments. I.Q. prepared cryo-EM samples, collected cryo-EM data and determined the structure of TnsC. I.Q. performed crystallization experiments and determined the structure of TniQ. M.S. prepared cryo-EM samples, conducted cryo-EM analysis and determined the structure of Cas12k. M.S. and I.Q. carried out biochemical and ddPCR functional experiments and negative-stain electron microscopy analysis. S.O. expressed and purified TnsC mutant proteins. C.C. assisted with sample preparation for biochemical and ddPCR assays. I.Q., M.S. and M.J. analysed the data and wrote the manuscript.

Competing interests The authors declare no competing interests.

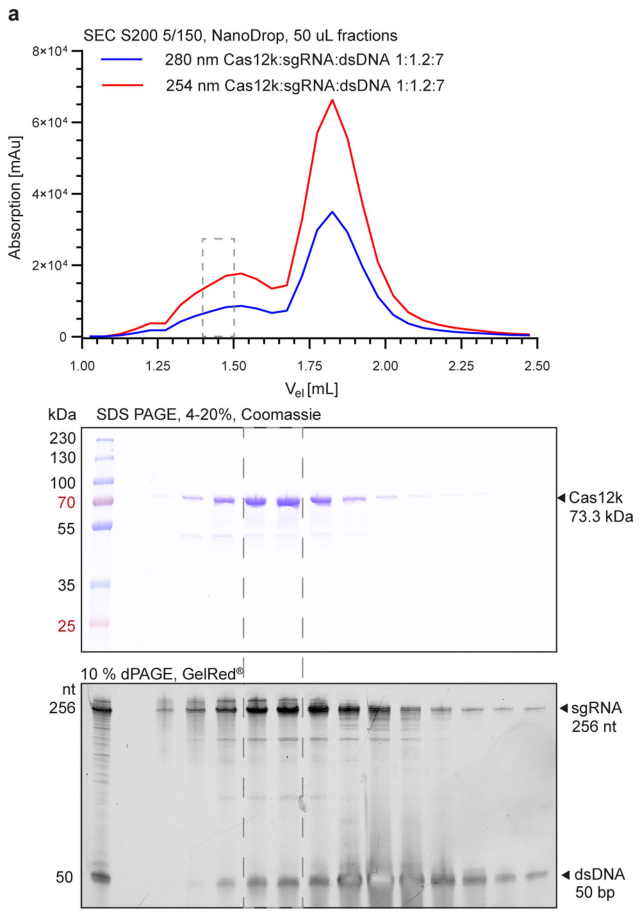
Additional information

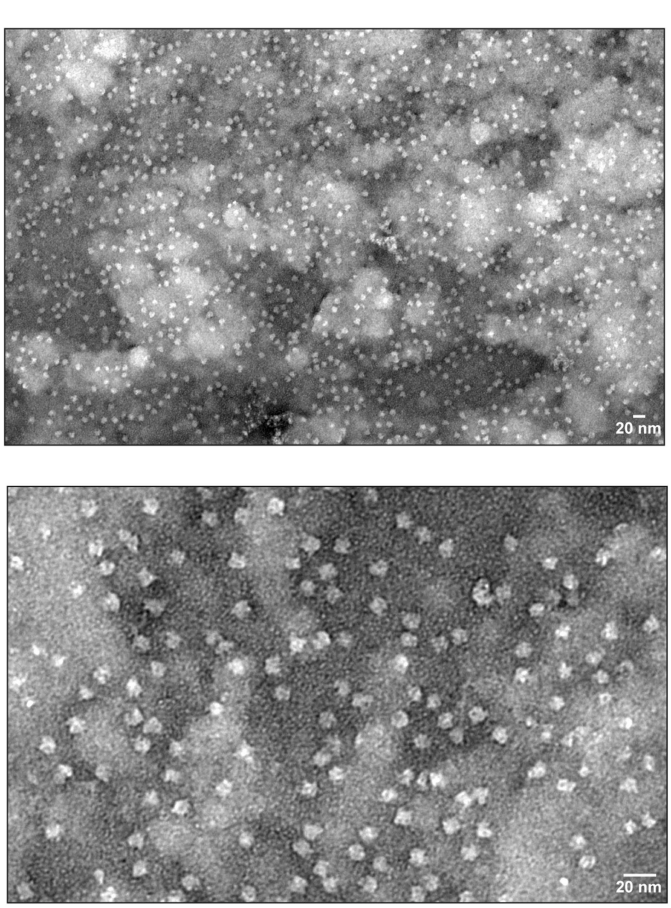
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04030-z.

Correspondence and requests for materials should be addressed to Martin Jinek.

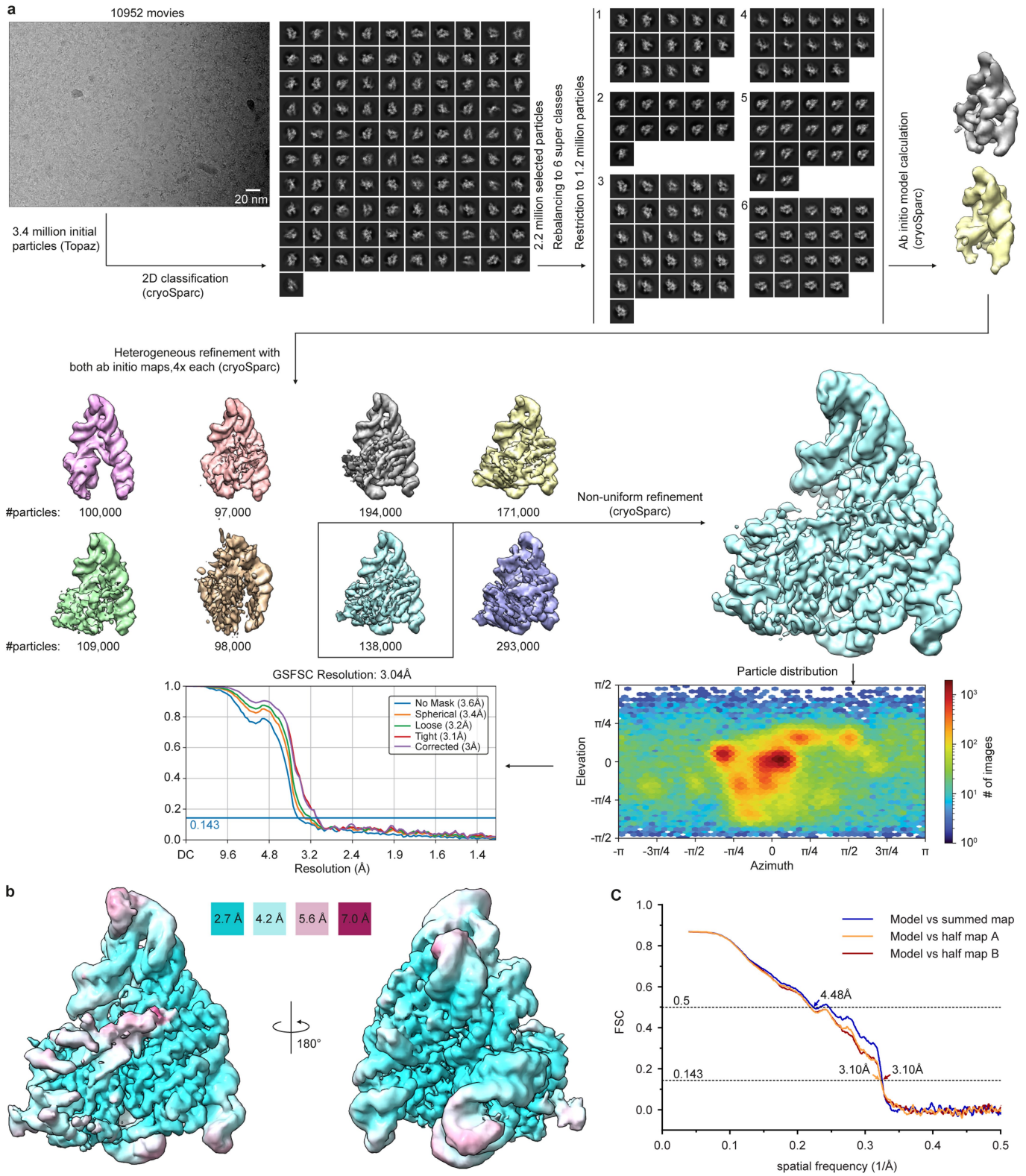
Peer review information *Nature* thanks Hiroshi Nishimasu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.

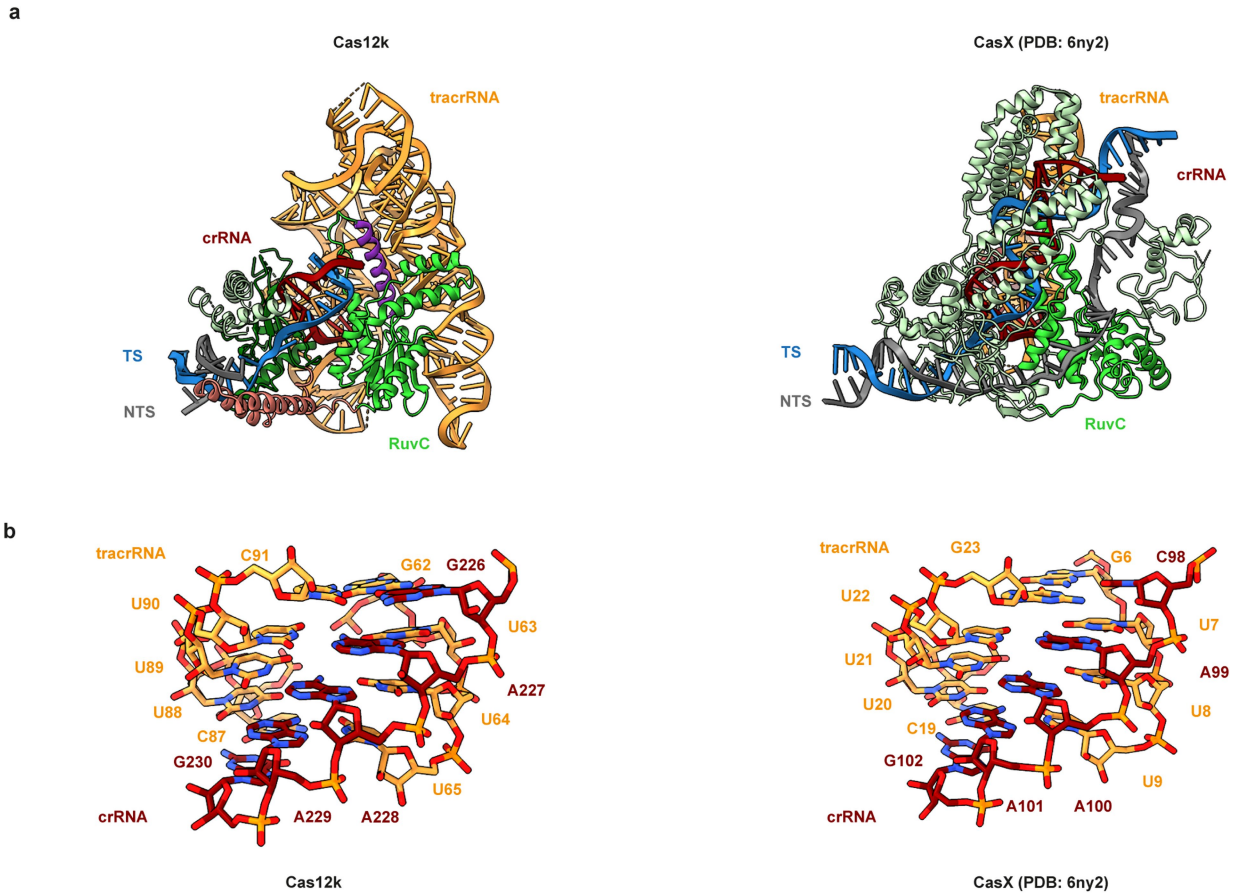




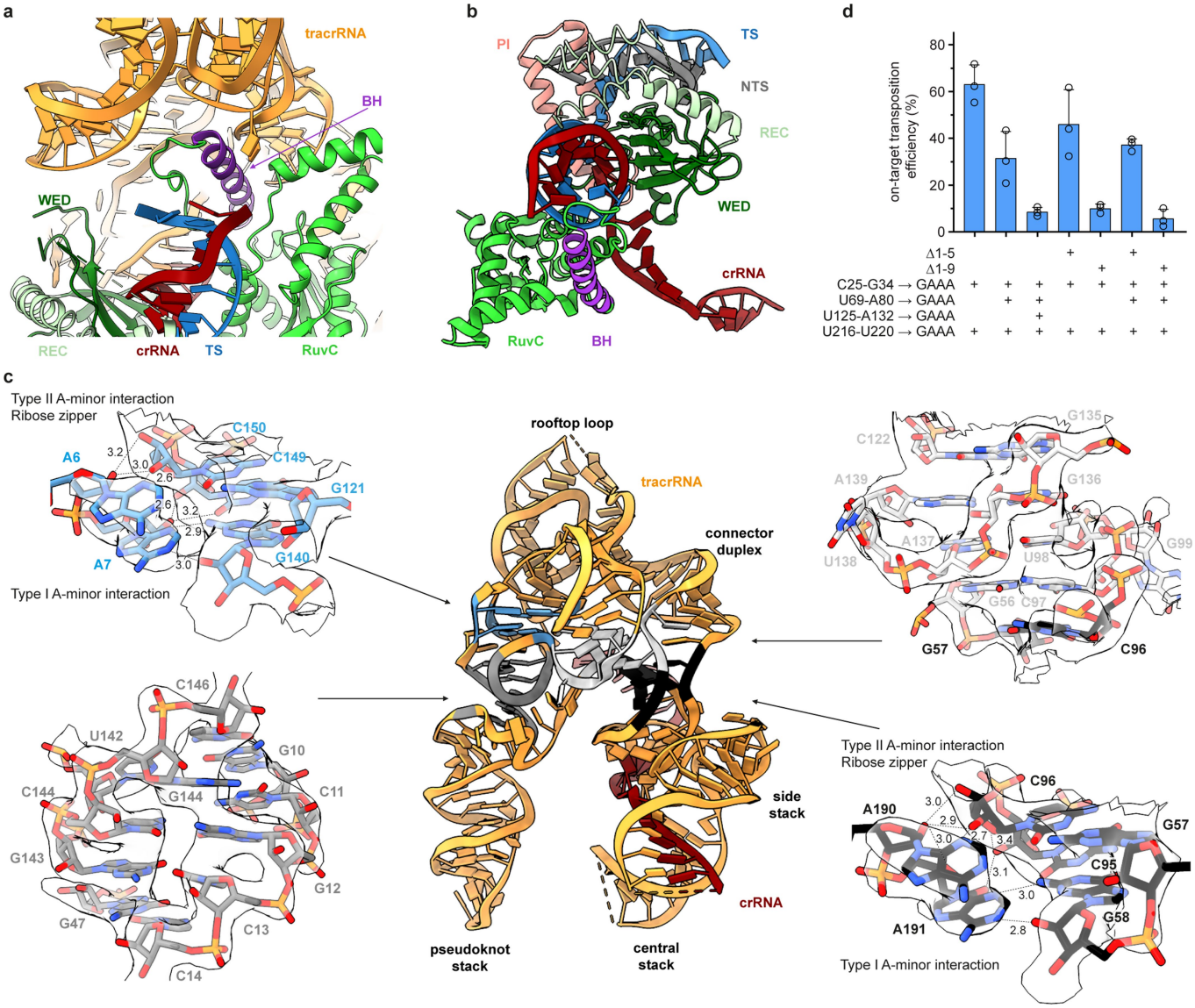
Extended Data Fig. 1 | Cas12k-sgRNA-target DNA complex assembly. a, Top: Size exclusion chromatography analysis of the Cas12k-sgRNA-target DNA complex. Middle: SDS-PAGE analysis of fractions from a. Proteins were visualised by Coomassie blue staining. Bottom: denaturing PAGE analysis of fractions from a. Nucleic acids were stained with a fluorescent dye (Gel Red). **b**, Representative negative stain EM micrographs of the ShCas12k-sgRNA-dsDNA complex at 68,000x magnification (top) and 180,000x magnification (bottom). Experiment was repeated three times independently with similar results. For gel source data, see Supplementary Fig. 1.



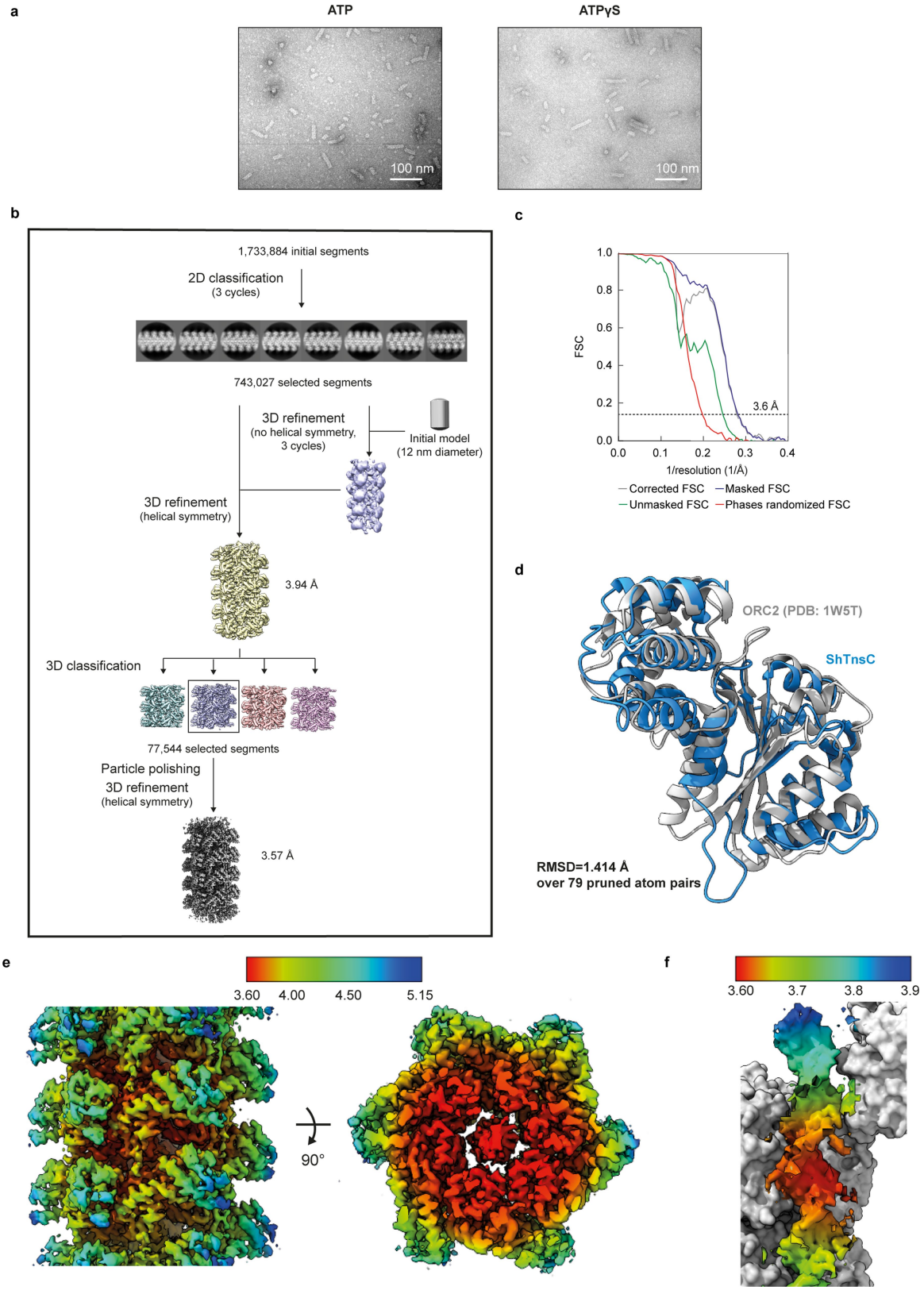
Extended Data Fig. 2 | Cas12k cryo-EM image processing workflow and model validation. a, Cryo-EM image processing workflow for the Cas12k-sgRNA-target DNA complex. Fourier Shell Correlations (FSC) of ShCas12k reconstruction from two independently refined half-maps. The gold-standard cutoff (FSC = 0.143) is marked with a blue line. **b**, Final electron density map colored according to the local resolution. **c**, Validation of Cas12k-sgRNA-target DNA structure model.



Extended Data Fig. 3 | Structural comparison between Cas12k and CasX. **a**, Structural model of Cas12k-sgRNA-target DNA (left) and CasX-sgRNA-target DNA (right, PDB: 6ny2[22]) complexes. The structures were superimposed using secondary structure matching[46] of the RuvC domains and are shown in identical orientations. **b**, Zoom-in views of the tracrRNA-crRNA triplexes.

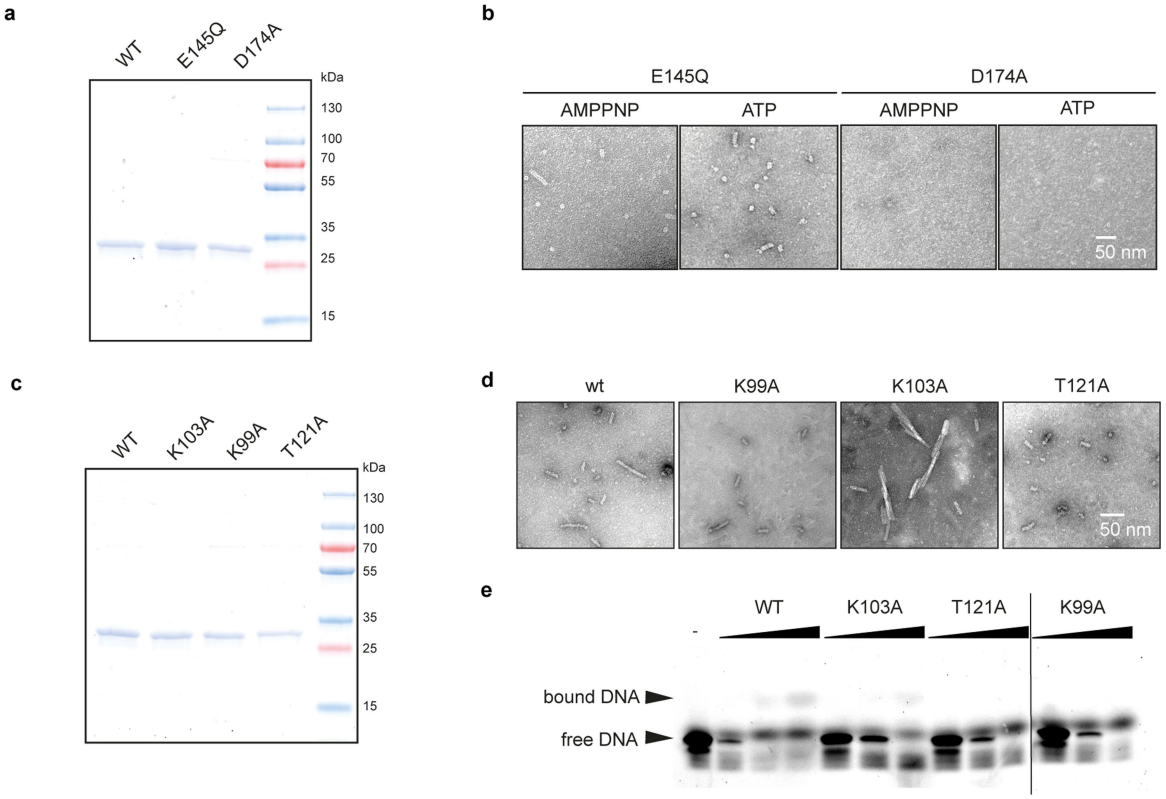


Extended Data Fig. 4 | Structural features and functional analysis of ShCAST sgRNA. a, Close-up view of the RNA:DNA duplex in proximity of the bridge helix (BH). **b**, Close-up view of the sgRNA-TS DNA heteroduplex in the Cas12k-sgRNA-target DNA complex. **c**, Structural model of the Cas12k sgRNA with detailed views of the 5′ terminal segment (top left, 6.6 σ contour level) and the triplex junction (bottom right, 9.0 σ contour level) of the tracrRNA forming ribose-zipper and A-minor interactions, the pseudoknot duplex (bottom left, 9.0 σ contour level) and the central stack junction (top right, 7.6 σ contour level). **d**, Droplet digital PCR (ddPCR)-based analysis of the transposition activity of structure-based sgRNA scaffold mutants in the ShCAST system. Data are presented as mean ± s.d. (n=3 biologically independent replicates).

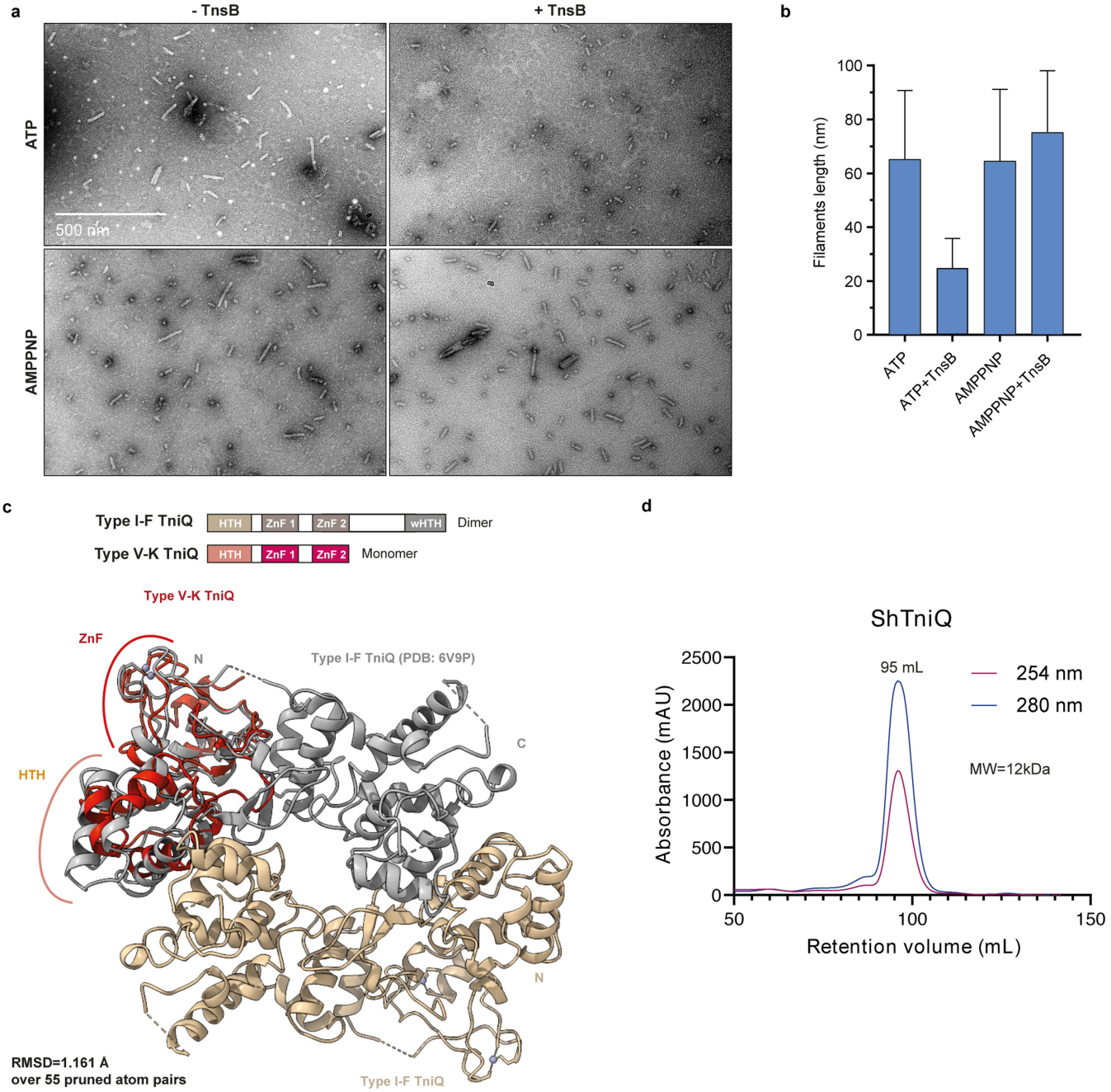


Extended Data Fig. 5 | See next page for caption.

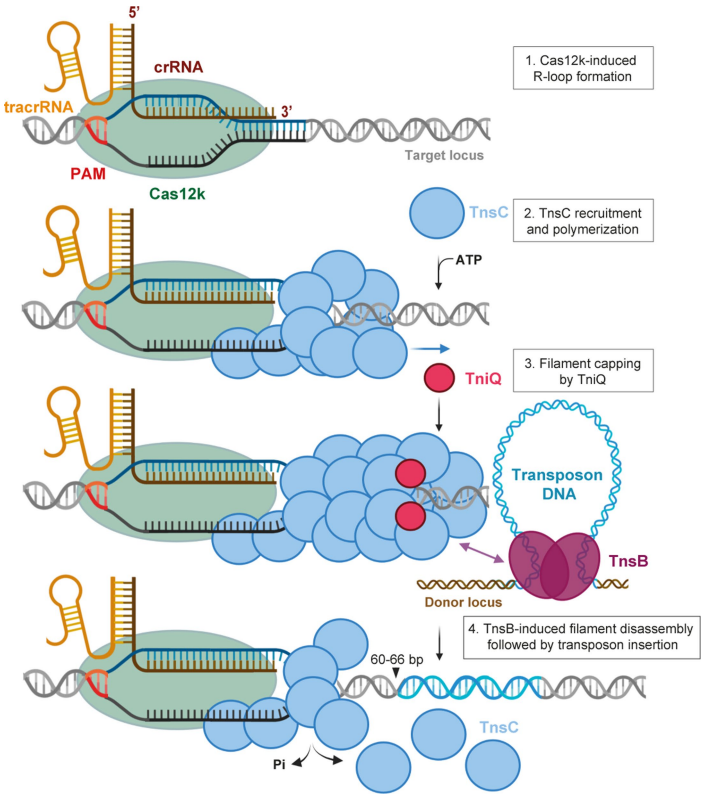
Extended Data Fig. 5 | ShTnsC filament formation and cryo-EM image processing of TnsC-dsDNA helical filaments. a, Representative negative-stain EM micrographs of TnsC in the presence of a 92-bp dsDNA and ATP or ATPγS. Scale bars, 100 nm. Magnification, 120,000x. Experiment was repeated twice independently with similar results. **b**, Cryo-EM data processing workflow for the TnsC-dsDNA complex. Selected 2D class averages used in image reconstruction and intermediate and final reconstructions are shown. **c**, Fourier Shell Correlations (FSC) of TnsC filament reconstruction from two independently refined half-maps. The gold-standard cutoff (FSC = 0.143) is marked with dashed lines. The resolution value of the FSC corrected curve at this level is indicated. **d**, Structural alignment of the cryo-EM structure of *S. hofmanni* TnsC (blue) and the crystal structure of the *Aeropyrum pernix* ORC2 protein (grey; PDB ID: 1W5T[52]) used as initial homology model for model building (single protomers). The root mean square deviation (RMSD) between the structures is 1.41 Å over 79 pruned atom pairs, as calculated in Chimera. **e**, Local resolution estimation (Å) for the cryo-EM volumes of the TnsC-dsDNA helical filaments. **f**, Local resolution estimation (Å) of the bound DNA duplex.



Extended Data Fig. 6 | Biochemical analysis of ShTnsC mutants. a, SDS-PAGE analysis of purified wild type and ATPase activity mutant TnsC proteins. **b**, Representative negative-stain EM micrographs of TnsC ATPase activity mutants incubated in the presence of dsDNA and ATP or AMPPNP. Magnification, 98,000x. Experiment was repeated twice independently with similar results. **c**, SDS-PAGE analysis of purified wild type and DNA binding mutant TnsC proteins. **d**, Representative negative stain EM micrographs of TnsC in the presence of AMPPNP and dsDNA and with mutations in the DNA binding interface. Magnification, 98,000x. Experiment was repeated twice independently with similar results. **e**, Electromobility shift assay using fluorophore-labeled 27-bp dsDNA in the presence of AMPPNP and either wild type or DNA binding mutant TnsC proteins. -, no protein control. Experiment was repeated twice independently with similar results. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | Role of ShTnsB in ShTnsC filament disassembly and crystal structure of ShTniQ. a, Representative negative stain EM micrographs of dsDNA-bound TnsC filaments in the presence of TnsB and AMPPNP or ATP. Magnification, 98,000x. **b**, Quantification of filament length in the indicated samples. Data represents the average length ± s.d. of 50 randomly selected filaments in each sample. Experiment in **a** and quantification in **b** were repeated twice independently with similar results. **c**, Top panel: Domain architecture of TniQ proteins. HTH, helix turn helix motif. wHTH, winged HTH. ZnF, zinc-finger motif. Bottom panel: Structural alignment of the crystal structures of type V *S. hofmanni* TniQ (red) and type I-F *Vibrio cholerae* TniQ (grey and beige protomers; PDB ID: 6V9P[19]) (bottom panel). The root mean square deviation (RMSD) between the structures is 1.16 Å over 55 pruned atom pairs, as calculated in Chimera. **d**, Size exclusion chromatography analysis of TniQ. The retention volume corresponds to that of a protein of 12 kDa, as calculated based on molecular weight standards. The theoretical molecular weight of *S. hofmanni* TniQ is 19 kDa.



Extended Data Fig. 8 | Mechanistic model of Cas12k-directed transposition. Schematic diagram depicting transposition of type V-K CRISPR-associated transposons. Cas12k in association with a crRNA-tracr RNA dual guide RNA recognizes target DNA sequences, forming a partial R-loop structure. Full R-loop is formed upon recruitment of TnsC by interactions with DNA-bound Cas12k, which nucleates ATP-dependent formation of a helical filament around structurally remodeled DNA. Filament growth is restricted by TniQ capping the Cas12k-distal end of the TnsC filament. The TnsC filament serves as a recruitment platform for TnsB, which interacts directly with TnsC and stimulates its ATPase activity. This leads to filament disassembly, making DNA downstream of the PAM accessible for transposon insertion. In the post-hydrolysis state, TnsC forms a single-turn helical hexamer around target DNA, possibly acting as a molecular ruler to define the spacing between the Cas12k binding and transposon insertion sites (60-66 nt downstream of the PAM). Image created with BioRender.com and adapted.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics for the *S. hofmanni* Cas12k-sgRNA-target DNA and TnsC-dsDNA complexes

	ShCas12k-sgRNA-dsDNA (EMDB-13486) (PDB 7PLA)	ShTnsC-AMPPNP-dsDNA (EMDB-13489) (PDB 7PLH)
Data collection and processing		
Magnification	130,000	130,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	51.82	66.39
Defocus range (μm)	-0.9 to -1.9 (0.1 steps)	-1.0 to -2.4 (0.2 steps)
Pixel size (Å)	0.64	0.325
Symmetry imposed	C1	C1
Initial particle images (no.)	10,952	5,055
Final particle images (no.)	10,866	4,988
Map resolution (Å)	3.04	3.58
FSC threshold	0.143	0.143
Map resolution range (Å)	2.7-7.0	3.58-5.05
Refinement		
Initial model used (PDB code)	n. a.	1W5T
Model resolution (Å)	4.5	3.9
FSC threshold	0.5	0.5
Map sharpening *B* factor (Å^2)	-113.9	-79.50
Model composition		
Non-hydrogen atoms	8701	15542
Protein residues	437	1820
Nucleotide residues	244	44
Ligands	0	MG :7 AMPPNP :7
B factors (Å^2) min/max/mean		
Protein	50.00/288.87/177.33	17.40/95.74/39.30
Nucleotide	20.00/30.00/29.96	20.00/20.00/20.00
Ligand		39.53/40.64/40.61
R.m.s. deviations		
Bond lengths (Å)	0.008	0.010
Bond angles (°)	1.152	1.659
Validation		
MolProbity score	1.71	2.36
Clashscore	9.53	13.30
Poor rotamers (%)	0.26	8.98
Ramachandran plot		
Favored (%)	96.74	98.06
Allowed (%)	3.26	1.94
Disallowed (%)	0.00	0.00
Ramachandran Z score		
whole	-0.03	-2.29
helix	1.11	-0.98
sheet	-1.19	-1.95
loop	-1.32	-2.19

Extended Data Table 2 | X-ray crystallographic data collection and refinement statistics for *S. hofmanni* TniQ. Values in parentheses are for highest-resolution shell

	ShTniQ (PDB 7OXD)
Data collection	
Space group	P 21 21 21
Cell dimensions	
a, *b*, *c* (Å)	41.636 49.875 86.8
α, β, γ (°)	90 90 90
Resolution (Å)	43.40 - 1.3 (1.346 - 1.3)
R_{merge}	0.03864 (0.9239)
$I / \sigma I$	23.26 (1.84)
Completeness (%)	99.75 (99.72)
Redundancy	6.7 (6.3)
Refinement	
Resolution (Å)	43.24 - 1.3
No. reflections	45126 (4426)
R_{work} / R_{free}	0.1881 (0.2866)/ 0.1952 (0.2872)
No. atoms	1470
Protein	1232
Ligand/ion	2
Water	236
B-factors	
Protein	18.97
Ligand/ion	17.19
Water	30.86
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.92

nature portfolio

Corresponding author(s): Martin Jinek

Last updated by author(s): Sep 2, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	EPU v2.9.0.1519REL, SerialEM v3.8, QuantaSOFT v1.7.4.0917, Gatan Digital micrograph v1.84.1282, DA+
Data analysis	cryoSPARC v3.2.0, Topaz v0.2.4, RELION v3.1.2, MotionCor2 v1.4.0, Gctf v1.06, Phenix v1.19.1-4122 (for refinement of TniQ structure), XDS v Mar 15, 2019 (BUILT=20190315), Autosol as implemented in Phenix 1.14-3260, autoSHARP v3.10.9, Phenix v1.19.2-4158 (for cryoEM structures), ARP/wARP v8.0, ccp4 v7.1.013, Coot v0.9.2-preEL in CCP-EM v1.5.0, MolProbity v4.5.1, UCSF Chimera 1.14, UCSF ChimeraX v1.2, QuantaSOFT v1.7.4.0917

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Maps and atomic coordinates of cryo-EM structures have been deposited in the Electron Microscopy Data Bank with accession codes EMD-13486 and EMD-13489 and the Protein Data Bank with accession codes 7PLH and 7PLA. Structure factors and atomic coordinates of the X-ray crystallographic structure of ShTniQ have been deposited in the Protein Data Bank with accession code 7OXD. The models used for structural comparison are available as PDB IDs 6I1K, 6NY2, 1W5T and 6V9P.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes are reported in the figure legends. ddPCR measurements were performed twice for three biological replicates. The selected sample size is a standard for the ddPCR method and has been reported for transposition assays in Strecker et al., 2019.
Data exclusions	No data was excluded.
Replication	Information on replication of the experiments is reported in the figure legends. ddPCR measurements were repeated twice for each of the three biological replicates. Negative stain experiments, EMSA and pull-down assays were repeated twice. All replication attempts showed similar results.
Randomization	For transposition assays, all colonies corresponding to each individual sample were grown on solid media and pooled together for further analysis. Each individual sample represents an entire heterogeneous population and thus further randomization is not applicable. ddPCR measurements were performed on different days and grouped in distinct batches independently of the assigned sample number (see below).
Blinding	Samples for ddPCR assays were labeled with numbers and prepared unblinded by one operator and measured by another operator who did not have prior knowledge of the sample labels and protein/RNA mutations tested. Quantification of filament length from micrographs was done by one operator without knowledge of the conditions tested.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested,*

Study description	*hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.