Article

Licensing mechanisms during RNA-guided DNA integration by type I-F3 CAST

https://doi.org/10.1038/s41586-026-11287-9

Received: 8 June 2026

Accepted: 9 October 2026

Published online: 25 November 2026

Open access

Check for updates

Shirin Fatma[1,4], Shubham Dubey[1,4], Vinh H. Truong[1,4], Seong Guk Park[1], Hailey M. Wallace[1,2], Alfredo Jose Florez Ariza[1], Isabella King[3] & Elizabeth H. Kellogg[1]✉

Precise genomic integration of large DNA cargo remains a central challenge in genome engineering. CRISPR-associated transposases (CASTs)[1,2] offer a promising solution by coupling RNA-guided DNA recognition to transposon-mediated integration. Because of their exceptional targeting fidelity, type I-F3 CASTs have been a major research focus[1], yet the structural basis of their RNA-guided DNA integration has remained unclear. Here, using cryo-electron microscopy, computational protein design and functional assays, we report the licensing mechanisms that control DNA integration activity of *Vibrio cholerae* CAST (VchCAST). Structural determination was enabled by zero-shot computational redesign of TnsA, which enhanced integration activity by nearly two orders of magnitude without compromising specificity. The structure of the holo integration complex reveals the conformational licensing events that govern activation of the integration machinery at the on-target site. Multiple components (Cascade–TniQ, TnsC and TnsA–TnsB) undergo conformational changes upon assembly into the holo integration complex, and the catalytically competent state is stabilized only within the full assembly. Comparison with existing CAST integration complex structures reveals conserved architectural features, suggesting a common molecular basis for RNA-guided transposition across CASTs. Our findings establish a general mechanistic framework for RNA-guided DNA integration and demonstrate that protein design is a powerful strategy for engineering highly active CAST systems.

A long-standing challenge in genome engineering is how to achieve precise, programmable integration of large (≥1 kb) DNA cargo. CRISPR–Cas nucleases are capable of efficient DNA cleavage[3], but their reliance on host DNA repair pathways limits both the efficiency and fidelity of large insertions[4]. CASTs offer an alternative strategy by coupling RNA-guided DNA recognition to transposon-mediated integration, without the need for DNA double-strand breaks or homologous recombination[1,2]. CASTs evolved from Tn7-like transposons through the acquisition of a CRISPR effector, with distinct effectors defining each of the major subfamilies: type V-K, I-F3, I-B and I-D[5]. All CASTs encode a conserved set of proteins: TniQ (bridging protein), TnsC (AAA+ regulator) and TnsA–TnsB (TnsA/B, transposase complex) (Fig. 1a). The CRISPR effector and TniQ form the targeting module, and TnsA/B and TnsC form the transposition machinery (Fig. 1a). Although all CASTs perform RNA-guided transposition, they can display a remarkable diversity of targeting behaviours. For example, in type I-B2 CASTs, the transposition machinery can associate with either the sequence-specific DNA-binding protein TnsD or an RNA-guided targeting complex[6]. Type V-K CASTs exhibit a robust off-target pathway, driven by the propensity of TniQ and TnsC to load at non-cognate sites[7]. Type I-F3 CASTs have a Cascade effector and exhibit near-exclusive on-target integration[8], which makes them of particular interest for clinical[9] and microbial genome-editing applications[10].

Recent structural studies of CAST assemblies have begun to reveal how this conserved transposition machinery accommodates architecturally diverse CRISPR effectors. The V-K system from *Scytonema hofmannii*, ShCAST, is the best characterized to date; its holo integration complex (HIC) structure shows that the host factor S15 and TniQ associate with Cas12k and together strongly distort target DNA, an event that is required to recruit two helical turns of hexameric TnsC[11,12], which in turn engages the transposase TnsB via its C-terminal hook[13]. A complete TnsABCD integration complex structure exists for type I-B2 CAST[14], but structural information on other CAST assemblies is limited to isolated components or subcomplexes. TnsC organization is not fixed across CAST systems, as it forms a heptameric ring in other CAST systems, including I-F3[15] and I-B2[16], as well as in Tn7[17]. Whether mechanistic insights from one system can be extended to other CAST systems is unclear.

Despite extensive biochemical, genetic and structural studies of individual components, the mechanisms by which type I-F3 CASTs achieve precise on-target integration remain unknown. Chromatin immunoprecipitation with sequencing (ChIP–seq) experiments have revealed a hierarchical assembly pathway in which Cascade–TniQ occupancy precedes TnsC recruitment, and stable TnsC accumulation depends on the transposase TnsA/B[15]. Cascade–TniQ binds DNA promiscuously and forms partial R-loops at many candidate sites[15], but they cannot

[1]Structural Biology Department, St Jude Children's Research Hospital, Memphis, TN, USA. [2]Center for Protein Design, Center of Excellence for Data Driven Discovery, St Jude Children's Research Hospital, Memphis, TN, USA. [3]San Diego State University, San Diego, CA, USA. [4]These authors contributed equally: Shirin Fatma, Shubham Dubey, Vinh H. Truong. ✉e-mail: elizabeth.kellogg@stjude.org

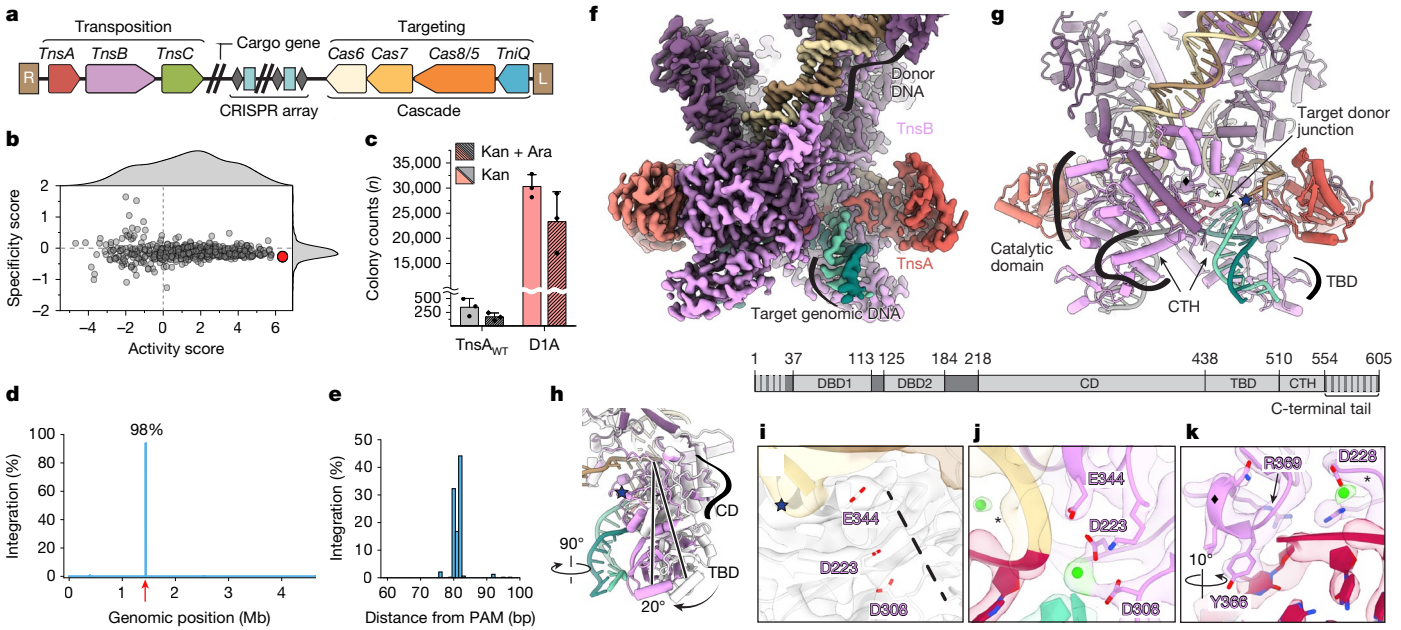


Fig. 1 | TnsA design enhances VchCAST integration activity and reveals TnsA-dependent conformational changes that promote target DNA engagement. a, Transposon organization. R and L, right and left transposon ends. Parallel lines indicate discontinuity. **b**, Dual screen of TnsA designs: activity and specificity scores are plotted relative to wild type (normalized to 0,0), with density plots on the top and right. The red circle represents the top design, D1A. **c**, In vivo integration activity of VchCAST. The graph shows individual data points. Data are mean ± s.d. from biological triplicates. Ara, arabinose selection; Kan, kanamycin selection. **d**, In vivo genome-wide insertion profile. Red arrow indicates target site. **e**, Close-up view of target site in **d** with insertion window. **f**, Cryo-EM density map (2.9 Å) of the class 1 STC at 0.129 threshold. Proteins coloured as in **a**. Target-site duplication is shown in red. **g**, Top, atomic model of the STC oriented as in **f**. Black symbols (asterisk, star and diamond) indicate regions shown in **h**–**k**. Bottom, domain organization of TnsB. Striped regions represent disordered N- and C-terminal tails. CD, catalytic domain; DBD, DNA-binding domain. **h**, Overlay of the catalytic TnsB in STC assembly class 2 (white) and class 1 (purple) shows conformational changes involved in transition from class 2 to class 1, with the catalytic domain and TBD rotating as indicated. View rotated 90° relative to **g**. **i**, Class 2 model shown in the same orientation as in **g**. Catalytic residues D223, D308 and E344 are shown as sticks; dashed lines indicate unresolved positions (331–341). **j**, Same as **i**, for the class 1 model. Magnesium ions are shown as green spheres. **k**, Close-up view of the region marked by the diamond and asterisk in **g**, rotated by 10°.

complete R-loop formation on their own[18], suggesting that additional steps are required to license TnsC recruitment. Structures of TnsC[15] have established the molecular details of its homo-oligomerization on DNA. Nonetheless, how all these components come together to form an integration-competent complex at the RNA-guided target site remains unknown. Thus, a structure of the fully assembled HIC is critical to establish the mechanistic basis for RNA-guided DNA integration and for rational engineering efforts to improve type I-F3 CAST integration efficiency for translational applications[9,19].

Here, using a combination of computational protein design, single-particle cryo-electron microscopy (cryo-EM) and directed in vivo mutagenesis, we uncover the mechanisms that regulate HIC assembly in the type I-F3 VchCAST system. The architectures of the strand-transfer complex (STC) and the full HIC reveal how RNA-guided target recognition, TnsC oligomerization and TnsA/B recruitment converge to organize a stable integration complex. Despite large structural differences across type V-K, I-F3 and I-B2 CASTs, conserved local features suggest universal mechanistic principles of RNA-guided DNA integration across CASTs. Our results reveal a context-dependent structural mechanism, whereby each of the aforementioned components undergoes significant conformational rearrangements relative to their isolated structures[15,18,20] as a means to license RNA-guided integration complex assembly. Our work also demonstrates how structure-guided computational protein design can be leveraged as a highly effective strategy for improving integration activity without compromising specificity.

Computational redesign of TnsA

Because TnsA was refractory to purification (Methods), we hypothesized that stabilizing it in a product-bound conformation could both enhance activity and facilitate structure determination. We therefore modelled the VchCAST STC (comprising TnsA, TnsB and a strand-transfer DNA substrate) using AlphaFold 3[21] (Extended Data Fig. 1). We then redesigned the TnsA sequence using LigandMPNN[22], with functionally important residues held fixed (32%; 74 out of 230 residues; Supplementary Fig. 1) and applied per-residue sampling biased towards TnsA homologues. Designs were grouped by sequence divergence from wild-type TnsA (TnsA$_{WT}$), then selected for experimental screening on the basis of predicted confidence scores (Methods).

We screened 500 top-ranking designs, each containing between 2 and 87 substitutions (Supplementary Table 1), using our previously developed high-throughput *Escherichia coli* screen[23]. In brief, a kanamycin resistance cassette serves as the donor DNA, such that kanamycin selection reports overall integration activity. On-target genomic integration is enriched by dual kanamycin and arabinose selection, which selects for disruption of an arabinose-inducible *ccdB* toxin gene. Activity and specificity scores are calculated as the $\log_2$-transformed fold change in variant frequency compared to wild type[23]. Of the 500 designs tested, 67% showed higher activity than TnsA$_{WT}$ (Fig. 1b). The top-performing design (D1A; Fig. 1b) incorporates 27 mutations (Extended Data Fig. 2) and has a $\log_2$-transformed activity score of 6.37, with no significant change in specificity (specificity score = −0.27; Fig. 1b). This improvement translates in vivo to an approximately 100-fold increase in integration activity relative to TnsA$_{WT}$ (Fig. 1c) while maintaining high fidelity, with 98% of insertions occurring on target (Fig. 1d) and at the expected distance, 82 base pairs (bp) from the protospacer adjacent motif (PAM)[1] (Fig. 1e).

TnsA is required for transposition, but its catalytic activity determines product identity: cleavage of the 5′ donor end generates simple insertions, whereas a failure to cleave generates co-integrates[24].

Long-read sequencing distinguishes these two products, establishing the extent of TnsA activity[8]. Using long-read sequencing, we found that D1A produced more co-integrates than wild-type TnsA (42% simple insertions versus 96.2%; Extended Data Fig. 3). This phenotype was attributable primarily to the Q73K substitution: reverting Q73 to the wild-type residue restored simple insertion frequency to near wild-type levels (92.4%) without compromising the transposition activity gains conferred by the broader D1A redesign (Extended Data Fig. 3), indicating that Q73K governs the simple insertion/co-integrate balance independently of those gains. This also does not appear to be a general consequence of the design campaign, as other top-scoring designs retain full catalytic activity (97–100% simple insertions; Extended Data Fig. 3). Finally, we examined the effect of redesign on the folded state of TnsA. Circular dichroism analysis revealed that $TnsA_{WT}$ has substantial random coil content, consistent with an unfolded protein, whereas D1A has a spectrum consistent with a well-folded, mostly α-helical protein (Supplementary Fig. 2). Together, these results demonstrate that zero-shot structure-based computational design can both improve CAST activity and facilitate structure determination, creating a framework for iterative structure-guided optimization.

TnsB active site reorganization

We next determined the cryo-EM structure of the STC containing integration host factor (IHF), D1A (hereafter referred to as TnsA), TnsB and a strand-transfer DNA substrate (2.9 Å overall resolution; Fig. 1f,g and Supplementary Figs. 3 and 4). We found that K73 occupies a second-shell position relative to the TnsA active site, sufficiently close to form a hydrogen bond with catalytic residues, providing a likely explanation for why the Q73K substitution reduces the catalytic activity of TnsA (Supplementary Fig. 5). As this perturbation is confined to the local active site environment, the D1A structure presented here should therefore represent an otherwise catalytically competent state. IHF promotes CAST transposition 10-fold, although it is not absolutely required for activity[25]. In our structure, IHF binds LE DNA at a site between LE1 and LE2 (Supplementary Fig. 3), bending the DNA to permit occupancy of additional TnsB binding sites. This binding mode has been described in molecular detail in a related study that lacked TnsA and other components of the HIC[20]. Here, we instead focus on the novel insights we observe in the presence of the partner protein TnsA.

TnsA localizes to the 5′ end of the donor DNA, where it binds the flanking host DNA and forms a continuous β-sheet with TnsB, a feature highly similar to previous TnsA structures from the type I-B2 PmcCAST system (Supplementary Fig. 6). We also identified conformational flexibility in the STC ensemble, which was separable by 3D classification. The highest-resolution class, containing the majority of particles, corresponds to class 1 (Supplementary Fig. 7a), whereas a second, minor 3D class (27% of particles, 3.2 Å overall resolution) lacks observable density for target DNA and for TnsB beyond the C-terminal helix (CTH) domain (class 2; Supplementary Fig. 7b,c and Extended Data Table 1). Closer inspection of class 2 reveals that the catalytic DDE triad (D223, D308, and E344) is too distant (6.2–6.9 Å; Supplementary Fig. 8) to form a functional active site, and that the loop preceding E344 (residues 331–341) is disordered (Fig. 1i). By contrast, class 1 has the expected arrangement of catalytic residues, as well as two magnesium ions: a catalytic magnesium with the expected coordination geometry and a second structural magnesium positioned nearby that stabilizes a sharp bend at the target–donor DNA junction (Fig. 1j). The conformational transition from class 2 to class 1 would entail a ~20° rotation of the catalytic domain and target DNA-binding domain (TBD) of TnsB at the TnsA–TnsB interface (blue star, Fig. 1h), reorienting these domains to engage target DNA. Concurrently, a target DNA-binding loop in the catalytic domain of TnsB (residues 359–378) becomes structured, further priming the complex for target DNA binding (Fig. 1k). These structural observations show that the STC exhibits conformational heterogeneity; one conformation (class 2) appearing catalytically inactive, and the other (class 1) showing rearrangements consistent with active site organization and target DNA engagement. As TnsA/B requires additional transposon-encoded partner proteins (that is, TnsC) for integration activity[1], these conformational changes may reflect a mechanism of activation for TnsA/B.

Architecture of the VchCAST HIC

We next determined the cryo-EM structure of the HIC, comprising Cascade–TniQ, TnsA, TnsB and TnsC, assembled on a strand-transfer DNA substrate that mimics the in vivo integration product (Supplementary Fig. 3). Our cryo-EM data analysis reveals substantial compositional heterogeneity, with three identifiable assemblies (Extended Data Fig. 4): Cascade–TniQ (target-recognition complex (TRC); 3 Å overall resolution), STC–TnsC (4.3 Å overall resolution) and the HIC (4.5 Å overall resolution; Fig. 2a). In the HIC, local resolution is highest at the Cascade core (3.7 Å) and decreases towards the STC periphery (about 7 Å; Extended Data Table 1 and Supplementary Fig. 9). Although IHF was included in the reconstitution, we did not observe IHF bound in the resulting structure, which may be attributable to the sample preparation method (Methods). Notably, the overall map was of sufficient quality to interpret different structural features without the need of a composite map stitched from subtracted focused-refined regions.

The HIC contains a Cascade effector, a TniQ dimer, a heptameric TnsC ring, a TnsB tetramer and two copies of TnsA (Fig. 2b), consistent with the stoichiometry established by structures of individual subcomplexes[15,18,20]. However, density for the seventh TnsC protomer (TnsC7) is noticeably weaker than for the other subunits (Supplementary Fig. 10), indicating substoichiometric occupancy or flexibility. The DNA-binding footprint accounts for the characteristic 82 bp integration distance from the PAM: 32 nucleotides occupied by the guide RNA spacer, followed by a 21, 14 and 15 bp footprint corresponding to TniQ, TnsC and TnsB, respectively, and terminating in a 5-nucleotide target-site duplication (Fig. 2c). Mutagenesis of the Cas8/5 PAM-interacting residues S127, R243, N246 and D461 substantially reduces integration activity (Supplementary Fig. 11), consistent with an important role for these residues in PAM recognition and transposition. Our structure of the VchCAST HIC reveals an arrangement wherein TnsC acts as a bifunctional adaptor, interacting with TniQ via its N-terminal face and with the STC via its C-terminal face (Fig. 2b). This architectural role of TnsC is similar to previous observations in ShCAST (type V-K)[12] and PmcCAST (type I-B2) systems[14,16]. This modularity explains how CAST systems have diversified to co-opt a wide range of CRISPR effectors while preserving the core transposition mechanism.

Role of the conserved TniQ–TnsC interface

A central question is how the targeting module is physically coupled to the core transposition machinery. In the HIC, TniQ2 serves as the primary physical bridge between the target recognition and transposition machinery (Fig. 3a), forming a predominantly hydrophobic interface with TnsC1 (Fig. 3b). Consistent with the functional importance of this interface, alanine substitutions at F2, L3 and F40 substantially reduced integration activity (Fig. 3c). TniQ forms a constitutive homodimer, but only the TniQ2 monomer interacts with TnsC, mirroring observations in the evolutionarily distant ShCAST and PmcCAST, in which a single TniQ subunit recruits TnsC[12,16]. Remarkably, despite the divergent overall architectures of the ShCAST (Protein Data Bank (PDB): 8EA3), PmcCAST (PDB: 8FF4) and VchCAST integration complexes, the TniQ–TnsC interface is notably conserved (Extended Data Fig. 5). In particular, the TnsC finger loop, previously shown to mediate critical

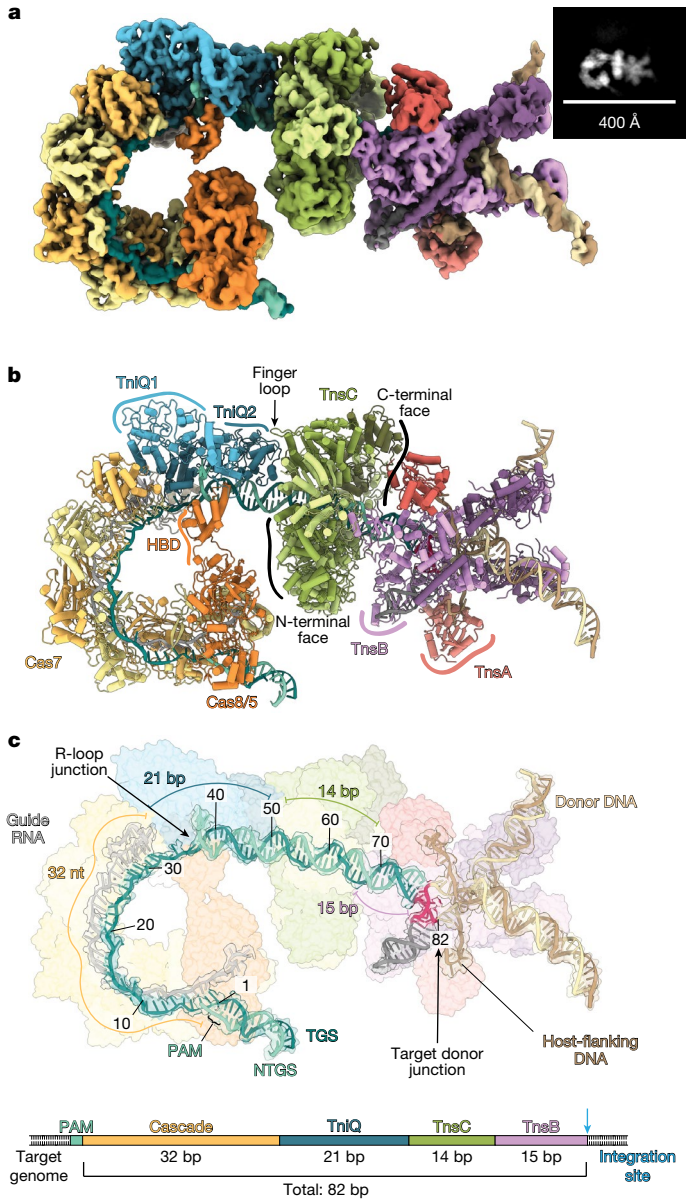


Fig. 2 | Overall architecture of the I-F3 VchCAST HIC. a, Cryo-EM density map of the HIC at 4.5 Å resolution, displayed at a threshold of 0.077, showing all components required for integration complex assembly. Inset, a representative 2D class average. Components are coloured as follows: Cas8/5, orange; Cas7, yellow; TniQ, blue; TnsC ring, green; TnsB, purple; and TnsA, red. Donor DNA is in beige and brown, guide RNA is in grey, and target DNA is in teal. Different shades distinguish individual protomers. **b**, Atomic model in cartoon representation. Components are labelled by protein name and structural features. The arrow points to the TnsC finger loop, and the black outline highlights the N- and C-terminal faces of TnsC. **c**, Top, surface representation of the protein components with transparency enabled. Nucleic acids are shown in cartoon representation, with segmented cryo-EM density (transparent, grey) displayed around DNA. The nucleic acid model depicts all key sequence features: PAM, target donor junction (that is, integration site), target-site duplication (red), genomic DNA (grey) and host flanking DNA. Nucleotide positions are numbered relative to the PAM. Coloured lines indicate protein footprints, and coloured numbers indicate the corresponding nucleotide footprints. Cascade, TniQ, TnsC and TnsB collectively occupy 82 base pairs, consistent with the observed insertion site. Bottom, genomic DNA-binding footprint. The blue arrow indicates the integration site.

TniQ interactions in ShCAST[12], engages TniQ2 similarly in VchCAST (Extended Data Fig. 5). Consistent with this observation, truncations within the TnsC finger loop severely reduce transposition activity[15]. These findings identify the TniQ–TnsC interface as a shared docking point that may enable architecturally diverse CRISPR effectors to couple to common transposition machinery.

Critical interactions between TnsB and TnsC

TnsB activity is strictly dependent on TnsC-mediated recruitment to the target site[15], making these interactions central to understanding integration. We next examined how TnsB engages TnsC within the HIC, revealing multiple points of contact along the C-terminal face of TnsC, specifically at protomers TnsC2, TnsC3 and TnsC6 (Fig. 3d). Notably, the TBD of TnsB localizes to the TnsC2–TnsC3 interface (Fig. 3e), consistent with previous ShCAST and PmcCAST structures[12,14] (Extended Data Fig. 6). Across Tn7-like transposons, ATP binding, but not hydrolysis, is required for integration activity[26,27]. AAA+ initiator clade regulators require ATP for oligomerization, as the ATP-binding pocket is formed at the inter-protomer interface[28]. This implies that transposition requires TnsC oligomerization on target DNA, although the mechanistic basis for this requirement has remained unclear. The conserved interaction between the TBD of TnsB and the TnsC inter-protomer interface now suggests that TnsB directly senses the oligomeric state of TnsC, providing a structural explanation for this long-standing observation.

Another conserved feature across CAST systems is the interaction between TnsC and the extreme C terminus of TnsB, known as the hook, which is critical for transposition[13]. The interchangeability of TnsB hooks between type I-F3 CASTs[29] underscores the importance of these structured interactions within otherwise disordered regions. Consistent with this role, we observe continuous density connecting the CTH (residues 563–573) of TnsB to the surface of TnsC6 (Fig. 3f), although additional contacts are likely to remain undetected owing to the length and extensive disorder of this region. The TnsB hook is predicted by AlphaFold 3[21] to correspond to the last eight residues of TnsB, an assignment supported by cryo-EM density observed in the HIC (Extended Data Fig. 7). The hook binding site in VchTnsC aligns well with ShTnsC[13] and PmcTnsC[14] (Extended Data Fig. 7), indicating that although structural details may differ, the binding location is conserved.

Intermolecular interactions via disordered termini

Having identified the TnsB hook interaction with TnsC, we reasoned that the disordered N- and C-terminal regions in all components of the core transposition machinery (TnsABC; Supplementary Fig. 12) could mediate additional interactions. Consistent with this idea, we observed contacts between the N terminus of TnsB and TnsA, and between the C-terminal tail of TnsC and TnsB (Fig. 3g). Residues 3–10 of the disordered N terminus of TnsB occupy a hydrophobic pocket on TnsA (Fig. 3h) close to aromatic residues F6 and F9, consistent with AlphaFold 3 predictions[21] (Extended Data Fig. 1) and sequence conservation analysis (Extended Data Fig. 8a). Although N-terminal truncation (Δ1–10) caused a significant reduction in transposition activity, the TnsB F9A mutation reproduced almost none of this effect, suggesting that residues other than F9 mediate the relevant contacts within the truncated region (Extended Data Fig. 8c). The C-terminal tail of TnsC (residues 321–330) is positioned against the CTH of TnsB (Fig. 3i and Extended Data Fig. 8b) at a hydrophobic interface, also supported by AlphaFold 3 predictions[21] (Supplementary Fig. 13) and conservation analysis (Extended Data Fig. 8b). The TnsC F329A mutation almost fully abrogated transposition activity, similar to the effect of C-terminal truncation of TnsC (Extended Data Fig. 8c). Analogous interactions between the C terminus of TnsC and the STC occur in PmcCAST, albeit involving different subunits and interfaces[14] (Extended Data Fig. 9). These findings uncover a previously unrecognized network of interactions involving the unstructured termini of TnsA, TnsB and TnsC, revealing how the core transposition machinery may maintain associations to facilitate integration-competent assembly.

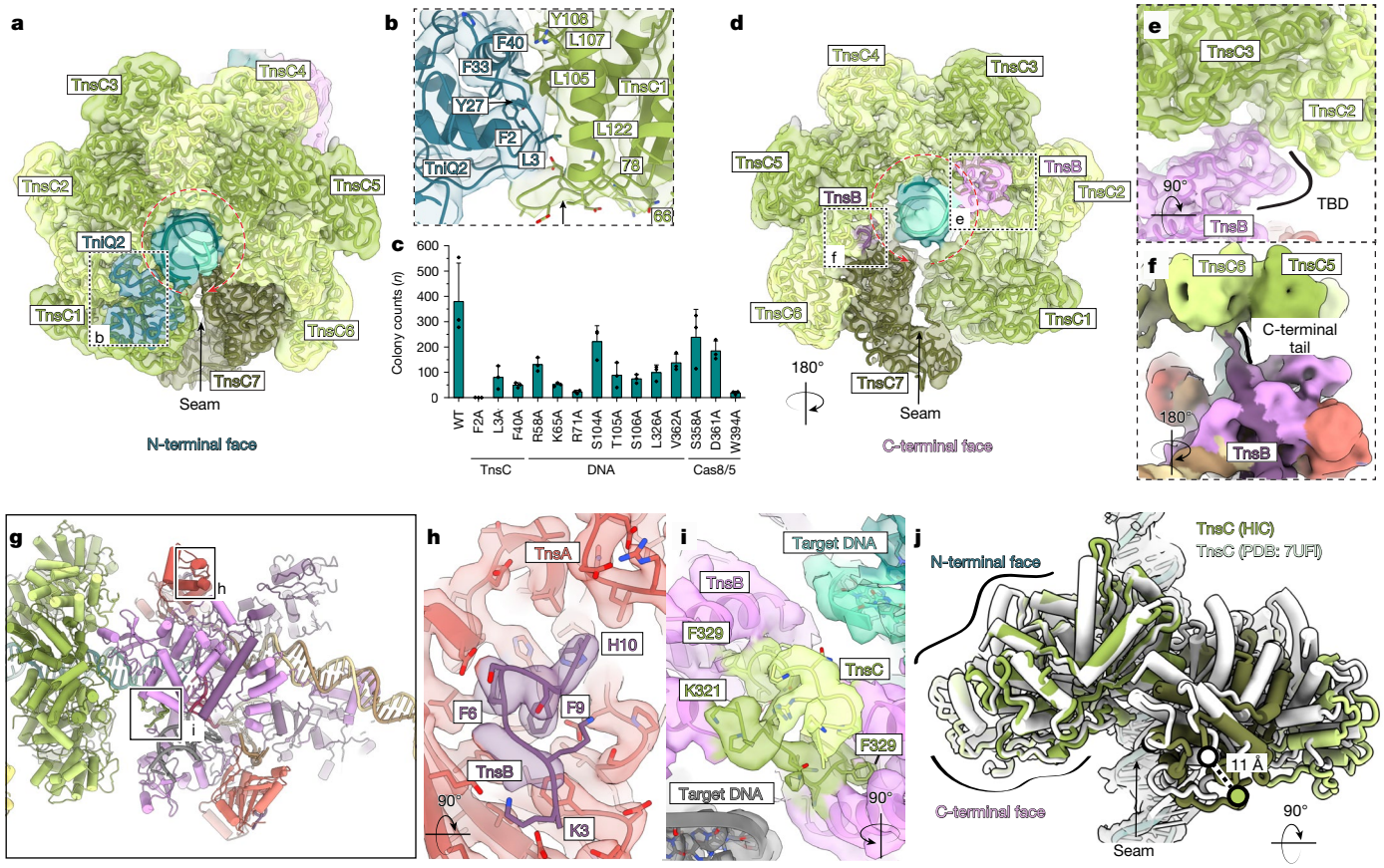


Fig. 3 | Critical protein–protein interactions within the HIC. a, N-terminal face of the TnsC heptamer, shown as an atomic model within the cryo-EM map of HIC (threshold = 0.077). TnsC protomers are numbered sequentially from the TniQ2-interacting protomer (red arrow); the black arrow marks the seam. The boxed region highlights the TniQ2–TnsC1 interface shown in **b**. **b**, Molecular details of the TniQ2–TnsC1 interface. The arrow shows the TnsC finger loop (residues 66–78). Key residues are shown as sticks. **c**, In vivo transposition assays with wild-type (WT) and indicated mutants of TniQ. The *y* axis presents colony counts after kanamycin selection. The graph shows individual data points and mean ± s.d. for $n = 3$ biological replicates. Mutated residues are grouped by category: TniQ2–TnsC1 interface (TnsC); DNA-interacting (DNA); and TniQ–Cas8/5 interface (Cas8/5). **d**, C-terminal face of the TnsC heptamer, shown as an atomic model within the cryo-EM map of HIC (threshold = 0.077). The red arrow indicates TnsC protomer order as in **a**. The boxed regions are shown in **e**,**f**. **e**, Detailed view of TnsB interactions spanning the TnsC inter-protomer interface, shown as in **d**. View is rotated as indicated. **f**, Low-pass-filtered (10 Å) cryo-EM density map of HIC, revealing the flexible C-terminal tail of TnsB. Rotation is relative to **e**. **g**, Model of HIC with docked C and N termini of TnsC and TnsB from the STC–TnsC and STC structures, respectively. The boxed regions are shown in **h** (TnsA–TnsB) and **i** (TnsB–TnsC). **h**, Interactions between the N terminus of TnsB and TnsA. The atomic model is docked into the cryo-EM density of STC (class 1, threshold = 0.11). View is rotated 90° relative to **g**. **i**, Interactions between the C-terminal tail of TnsC and TnsB. The atomic model is docked within the transparent 4.3 Å cryo-EM density of STC–TnsC (threshold = 0.06). View rotated 90° relative to **g**. **j**, Comparison of the TnsC oligomeric structures. View is rotated 90° relative to **a**. Displacement of TnsC7 indicated by circles.

Context dependence of the TnsC heptamer structure

A previous cryo-EM structure of VchTnsC solved in the presence of DNA revealed an open heptameric ring-like arrangement in which the bound DNA was poorly resolved (PDB: 7UFI)[15]. In our HIC structure, although we clearly resolved the path of target DNA, no specific TnsC–DNA interactions could be precisely mapped. Thus, it is likely that the TnsC ring primarily associates with target DNA via the complementary basic surface formed by its central pore (Supplementary Fig. 14). This is consistent with previous data showing that TnsC binds DNA relatively weakly[15], suggesting that TnsC requires partner proteins such as TniQ to initiate oligomerization on target DNA. Notably, comparison with PDB 7UFI reveals that the heptameric ring is significantly more open in the HIC. Aligning on TnsC1 results in an 11 Å displacement at TnsC7 relative to the equivalent protomer in PDB 7UFI (Fig. 3j), the result of an average 0.8 Å greater rise per protomer (Supplementary Fig. 15). This indicates that partner proteins induce larger conformational changes than those observed in TnsC with DNA alone. Our observations suggest that the oligomeric state of TnsC is sensitive to the presence of the partner proteins TniQ and TnsB and is mediated through distinct interaction interfaces located on specific TnsC protomers.

Cascade–TniQ remodelling drives TnsC recruitment

We next investigated how TnsC is selectively recruited to the target site. The compositional heterogeneity of our reconstitution enabled us to resolve a free Cascade–TniQ complex (Extended Data Fig. 4) (TRC; 3.0 Å overall resolution; Fig. 4a) that closely resembles a prior structure of the target DNA-bound VchCAST Cascade–TniQ complex[18] (Supplementary Fig. 16). This previously reported structure featured a partial R-loop and a disordered Cas8/5 helical bundle domain (HBD)[18]. By contrast, the Cascade–TniQ complex in the HIC (Fig. 4b) adopts a fully engaged conformation characterized by a complete R-loop and a structured Cas8/5 HBD. Comparison of Cascade–TniQ in the TRC and HIC structures reveals that R-loop completion is accompanied by a conformational change in TniQ2 to engage both target DNA and TnsC (Extended Data Fig. 10).

The non-target genomic strand (NTGS) is threaded through an electropositive channel at the tripartite TniQ1–TniQ2–Cas8/5 HBD

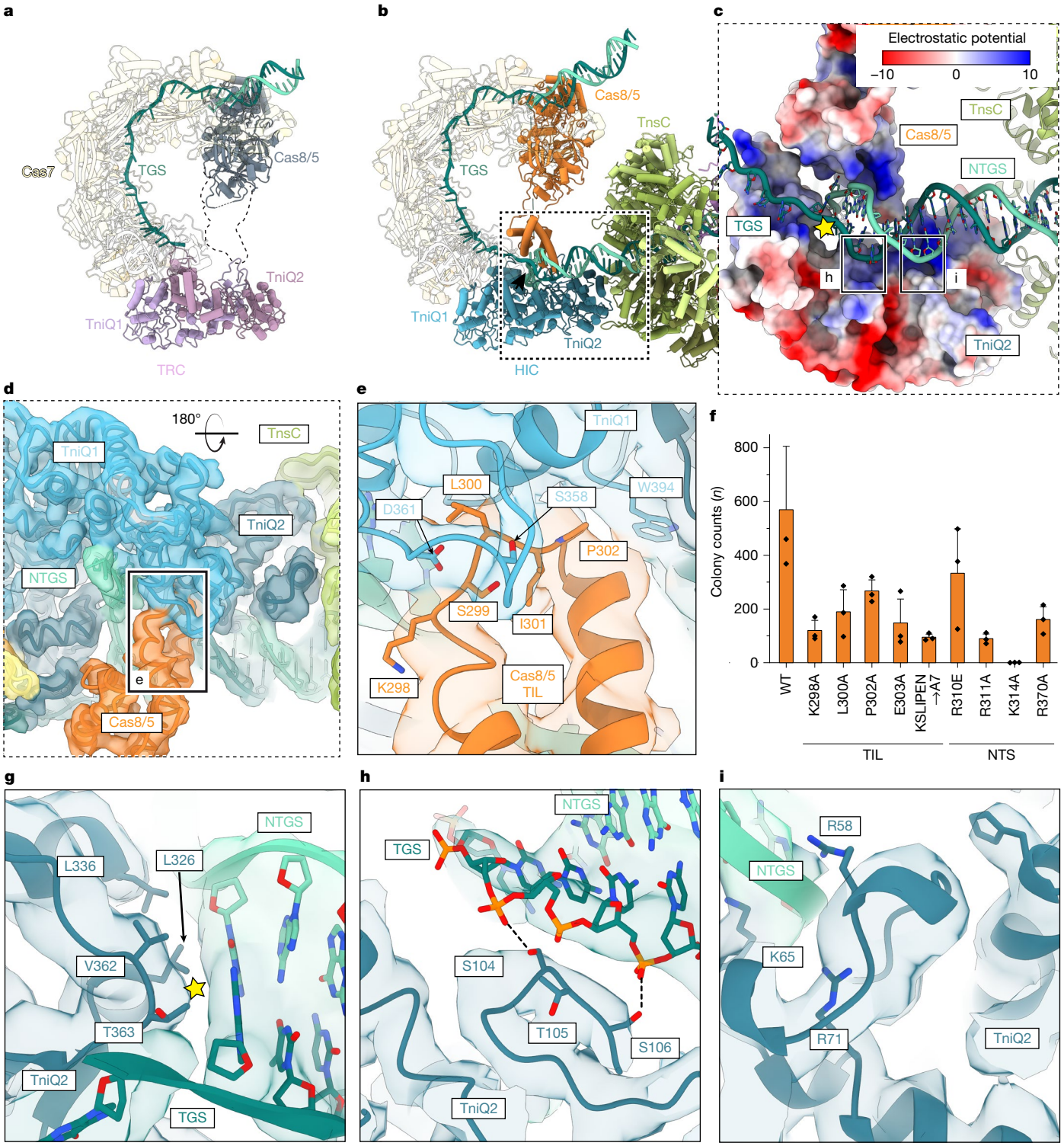


Fig. 4 | Cascade remodelling establishes a Cas8/5–TniQ junction that recognizes the R-loop and licenses recruitment of TnsC. a, Atomic model of the TRC, revealing unstructured Cas8/5 HBD (dotted lines). **b**, Atomic model of the HIC, focusing on Cascade–TniQ. Transition from TRC to HIC orders the target DNA duplex and facilitates recruitment of the TnsC ring. The box indicates areas that are enlarged in **c**–**e**. **c**, Positively charged surface at the Cas8/5–TniQ1 junction forms the binding interface for both the TGS and NTGS. Surfaces are coloured according to electrostatic potential. The yellow star indicates the R-loop junction. Boxes indicate the views shown in **h**,**i**. **d**, A narrow channel formed by TniQ1, TniQ2 and Cas8/5 accommodates threading of the NTGS. The atomic model is fitted into the cryo-EM density map (threshold = 0.077). The box indicates the view shown in **e**. Rotation relative to **b**. **e**, Cas8/5 helical bundle domain loop (the TIL; residues 298–304) interacts with TniQ1. Residues mediating Cas8/5–TniQ interactions shown as sticks. Same view as in **d**. **f**, In vivo transposition assays of wild-type and indicated mutants of Cas8/5. The *y* axis shows colony counts after kanamycin selection (overall activity). The graph shows individual points and mean ± s.d. for $n = 3$ biological replicates. Mutated residues are grouped by category: TniQ-interacting (TIL); and interacting with the NTGS. **g**, Selected TniQ2 residues interact with the target DNA duplex at the R-loop junction (yellow star, as in **c**). **h**,**i**, Molecular details of the view in **c**, showing interactions between TniQ2 and target DNA. Black dashed lines indicate putative hydrogen bonds between TniQ2 residues S104 and S106 and the target DNA backbone.

interface, where it rejoins the target genomic strand (TGS) to reform the target DNA duplex downstream of the R-loop junction (Fig. 4c,d). The Cas8/5 HBD contributes to this channel through electrostatic contacts with the NTGS via R310, R311, K314 and R370 (Supplementary Fig. 17). Additionally, a previously uncharacterized loop within the Cas8/5 HBD (residues KSLIPEN), which we term the TniQ-interacting loop (TIL), interacts directly with TniQ1 (Fig. 4e). The functional importance of this interface is confirmed by alanine substitutions in the Cas8/5 HBD, TIL and at TniQ1 residues D361, S358 and W394, all of which substantially reduce integration activity (Figs. 3c and 4f).

Separately, TniQ2 engages target DNA through two distinct modes: hydrophobic packing against the R-loop junction (L326, L336, V362 and T363; Fig. 4g) and electrostatic interactions with the DNA backbone via basic and polar residues (R58, K65, R71, S104 and S106; Fig. 4h,i). Mutagenesis of residues at both interfaces substantially reduces integration activity, consistent with their functional importance (Fig. 3c). These observations establish that R-loop completion drives a conformational transition in the target recognition module that repositions TniQ to engage target DNA and license recruitment of the core transposition machinery.

Discussion

We propose the following working model for RNA-guided transposition, based on our findings and incorporating prior observations on CASTs and other Tn7-like systems. Cascade–TniQ promiscuously samples genomic locations[15], consistent with its ability to stably bind target DNA with only partial R-loop formation (Fig. 5). A cooperative conformational rearrangement in Cas8/5 and TniQ is required to stabilize the full R-loop and enable TnsC recruitment, with direct contacts of TniQ at the R-loop junction providing a structural readout of R-loop completion. This coupling is further supported by molecular dynamics simulations demonstrating correlated motions between Cas8/5 and TniQ[30]. We observe no other DNA distortions consistent with specific recognition of the target site by the transposon-encoded machinery (Supplementary Fig. 18), contrary to that observed in other Tn7-like systems[14,31]. Further insight into target-site selectivity comes from the observation that PAM-distal mismatches permit Cascade binding but abolish transposition[15]. The conformational rearrangements in the HIC occur near spacer nucleotide positions 25–32, raising the possibility that mismatches in this region could introduce geometric incompatibility that prevents stable docking of the HBD and subsequent TnsC recruitment (Fig. 5). TnsC oligomerization is required to recruit TnsA/B to the target site and may be sensed directly by TnsB through conserved interactions spanning TnsC protomers. Whether full heptamer formation is required prior to TnsA/B recruitment remains unknown and is not resolvable using our reconstitution approach; however, previous work on Tn7 suggests that oligomerization is completed prior to TnsA/B recruitment[32].

TnsB also undergoes multiple layers of activation through association with partner proteins TnsA and TnsC. In isolation, the DDE catalytic triad is improperly arranged and too distant to coordinate magnesium ion binding, and target DNA remains disordered even when a strand-transfer substrate is provided[20]. Association with TnsA is insufficient to fully stabilize the catalytically competent conformation (Fig. 5). Only in the context of the HIC, in which both TnsA and TnsC are present, are the catalytic residues locked into an integration-competent state and target DNA stably accommodated. Although our reliance on a strand-transfer substrate does not allow us to establish the stepwise order of these conformations directly, the inactive-to-active transition that we observe is, to our knowledge, unreported in CAST systems despite extensive structural work from our group and others[13,14,33]. Notably, a conceptually related precedent exists in IS21, where the transposase IstA transitions from an inactive to an active conformation upon association with its obligate AAA+ partner IstB[34], suggesting that

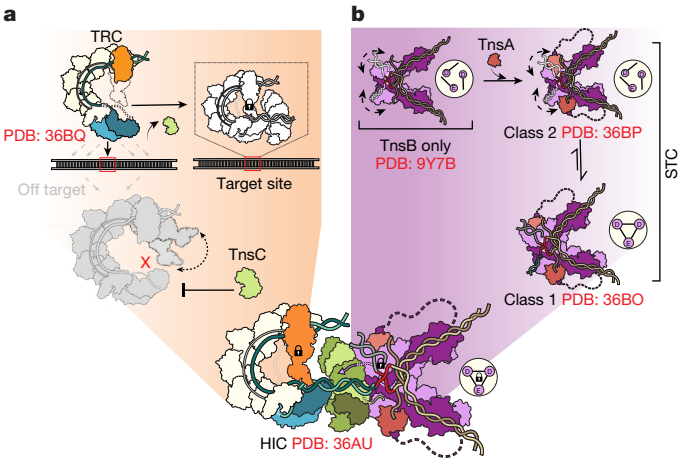


Fig. 5 | Proposed licensing mechanisms that regulate VchCAST assembly. Major licensing events during RNA-guided DNA integration directly at the CRISPR effector (**a**) and at the transposase (**b**). **a**, Top left, the TRC (PDB: 36BQ) displays an unstructured Cas8/5 helical bundle domain, a partial R-loop and an outward-open TniQ conformation. Unstructured regions are indicated with dotted lines. This state can sample many genomic locations (dotted arrows) in addition to the programmed target site (red box). Top right, upon recognition of the target, the TRC forms stabilizing interactions between Cas8/5 and TniQ, promoting R-loop completion and enabling recruitment of TnsC to complete this licensing step (lock symbol). Middle, off-target complexes (grey) are unable to form a complete R-loop (red X) and the Cas8/5 helical bundle domain remains flexible, and TnsC cannot be recruited. **b**, Top left, the TnsB-only STC (PDB: 9Y7B) contains flexible target-binding domains and an incompletely arranged active site, shown inside the white circle as a disconnected DDE catalytic triad. Top right and middle, upon addition of TnsA, the TnsA/B STC adopts two conformational states that exist in equilibrium: class 2 (PDB: 36BP) and class 1 (PDB: 36BO). The class 1 state exhibits the catalytically competent arrangement of the DDE triad (triangle within the circle). Dotted lines indicate flexible connections between components, including the interface between TnsA and TnsB. The targeting module (orange) and transposition module (purple) assemble with TnsC to form the HIC (PDB: 36AU). Both the targeting and the transposition modules adopt stable conformations and are locked into an integration-competent conformation. The DDE triad forms a catalytically competent active site. The C-terminal tails of TnsB and TnsC function to loosely couple components, with the hook of TnsB inserting into a TnsC pocket, forming β-sheet interactions (purple arrow), while the TnsC C-terminal tail contacts TnsB near target DNA. Dotted lines indicate disordered regions.

conformational licensing may be an effective mechanism for ensuring selective activation of integration across Tn7 and related transposons. In summary, we propose that each component of the integration machinery undergoes conformational changes upon incorporation into the HIC, providing a structural basis for the exceptionally high on-target specificity characteristic of type I-F3 CAST systems. Both the targeting module Cascade–TniQ and the transposase TnsA/B serve to license RNA-guided DNA integration, ensuring that catalytic competence is only achieved upon assembly of the complete machinery at the RNA-guided target site.

Structured interactions embedded within intrinsically disordered regions of CAST proteins have critical roles in transposition, most notably through the C-terminal hook of TnsB[12,29]. Here, combining our cryo-EM structure with AlphaFold 3 predictions, we localize the C-terminal hook of TnsB and identify its binding site on TnsC. We also identify two previously unrecognized interactions: the N terminus of TnsB engages a pocket on TnsA, and the C terminus of TnsC contacts TnsB in close proximity to target DNA. These findings reveal an extensive network of structured interactions within otherwise disordered termini that physically connect the core transposition machinery prior to integration complex assembly. The mapping of these contacts has

direct implications for the rational engineering of chimeric CAST systems and the design of improved variants.

The structural characterization of CAST systems has been largely motivated by the goal of engineering more efficient tools for genome editing; however, structure-guided efforts to improve CAST activity have had limited success[29,35,36]. Here we demonstrate that a high-throughput structure-based design approach can overcome these limitations, presenting a compelling and scalable alternative to directed evolution. We propose that thermodynamic stabilization of the integration-competent conformation is key to improve activity. This is consistent with the thermodynamic requirements of cut-and-paste transposition: the chemically isoenergetic reaction proceeds through states of increasing stability, and the product complex can be so stable that unfoldases (for example, ClpX) are required for its disassembly in vivo[37]. Consistent with this hypothesis, zero-shot structure-based protein design of a single protein, TnsA, enhanced integration activity by nearly two orders of magnitude in *E. coli*, approaching the gains achieved through seven rounds of directed evolution applied across the entire PseCAST system[19]. Notably, the redesigned variant not only improved integration activity but also enabled structural characterization of the HIC, illustrating how computational design and structural biology can be synergistic approaches.

The compromised catalytic activity of the top design variant D1A is attributable primarily to a lysine substitution at position 73, which is within range to form hydrogen-bonding interactions with the catalytic residue D91. We hypothesized that this substitution directly perturbs active site chemistry, which motivated our prioritization of the reversion mutant for testing; consistent with this hypothesis, reversion almost fully restored catalytic activity while maintaining substantial increases in activity. This trade-off reflects a known limitation of structure-based design methods: approaches such as ProteinMPNN can inadvertently stabilize protein structure at the expense of catalytic function[38], as functional residues are often locally destabilizing to the protein fold[39]. Therefore, success of the design strategy proposed here depends on mechanistic knowledge and the availability of a high-quality model of the integration-competent complex, precisely what the structures presented here provide, and what has been lacking in previous engineering efforts.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-026-11287-9.

1. 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).
2. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
5. Hsieh, S.-C. & Peters, J. E. Natural and engineered guide RNA-directed transposition with CRISPR-associated Tn7-like transposons. *Annu. Rev. Biochem.* **93**, 139–161 (2024).
6. Saito, M. et al. Dual modes of CRISPR-associated transposon homing. *Cell* **184**, 2441–2453.e18 (2021).
7. George, J. T. et al. Mechanism of target site selection by type V-K CRISPR-associated transposases. *Science* **382**, eadj8543 (2023).
8. Vo, P. L. H., Acree, C., Smith, M. L. & Sternberg, S. H. Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. *Mob. DNA* **12**, 13 (2021).
9. Lampe, G. D. et al. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat. Biotechnol.* **42**, 87–98 (2024).
10. Rubin, B. E. et al. Species- and site-specific genome editing in complex bacterial communities. *Nat. Microbiol.* **7**, 34–47 (2022).
11. Schmitz, M., Querques, I., Oberli, S., Chanez, C. & Jinek, M. Structural basis for the assembly of the type V CRISPR-associated transposon complex. *Cell* **185**, 4999–5010.e17 (2022).
12. Park, J.-U. et al. Structures of the holo CRISPR RNA-guided transposon integration complex. *Nature* **613**, 775–782 (2023).
13. Park, J.-U., Tsai, A. W.-L., Chen, T. H., Peters, J. E. & Kellogg, E. H. Mechanistic details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM. *Proc. Natl Acad. Sci. USA* **119**, e2202590119 (2022).
14. Wang, S., Siddique, R., Hall, M. C., Rice, P. A. & Chang, L. Structure of TnsABCD transpososome reveals mechanisms of targeted DNA transposition. *Cell* **187**, 6865–6881.e16 (2024).
15. Hoffmann, F. T. et al. Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. *Nature* **609**, 384–393 (2022).
16. Wang, S., Gabel, C., Siddique, R., Klose, T. & Chang, L. Molecular mechanism for Tn7-like transposon recruitment by a type I-B CRISPR effector. *Cell* **186**, 4204–4215.e19 (2023).
17. Shen, Y. et al. Structural basis for DNA targeting by the Tn7 transposon. *Nat. Struct. Mol. Biol.* **29**, 143–151 (2022).
18. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernández, I. S. Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system. *Nature* **577**, 271–274 (2020).
19. Witte, I. P. et al. Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase. *Science* **388**, eadt5199 (2025).
20. Truong, V. H. et al. Transposon end recognition and pairing by I-F3 CRISPR-associated transposase. *Sci. Adv.* **12**, eaee1374 https://doi.org/10.1126/sciadv.aee1374 (2026).
21. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
22. Dauparas, J. et al. Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **22**, 717–723 (2025).
23. Park, S. G. et al. Comprehensive profiling of activity and specificity of RNA-guided transposons reveals opportunities to engineer improved variants. *Nucleic Acids Res.* **53**, gkaf917 (2025).
24. May, E. W. & Craig, N. L. Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996).
25. Walker, M. W. G., Klompe, S. E., Zhang, D. J. & Sternberg, S. H. Novel molecular requirements for CRISPR RNA-guided transposition. *Nucleic Acids Res.* **51**, 4519–4535 (2023).
26. Stellwagen, A. E. & Craig, N. L. Analysis of gain-of-function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.* **305**, 633–642 (2001).
27. Park, J.-U. et al. Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021).
28. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
29. Lampe, G. D., Liang, A. R., Zhang, D. J., Fernández, I. S. & Sternberg, S. H. Structure-guided engineering of type I-F CASTs for targeted gene insertion in human cells. *Nat. Commun.* **16**, 7891 (2025).
30. Patel, A. C., Sinha, S., Arantes, P. R. & Palermo, G. Unveiling Cas8 dynamics and regulation within a transposon-encoded Cascade–TniQ complex. *Proc. Natl Acad. Sci. USA* **122**, e2422895122 (2025).
31. Kuduvalli, P. N., Rao, J. E. & Craig, N. L. Target DNA structure plays a critical role in Tn7 transposition. *EMBO J.* **20**, 924–932 (2001).
32. Shen, Y. et al. Assembly of the Tn7 targeting complex by a regulated stepwise process. *Mol. Cell* **84**, 2368–2381.e6 (2024).
33. Tenjo-Castaño, F. et al. Structure of the TnsB transposase-DNA complex of type V-K CRISPR-associated transposon. *Nat. Commun.* **13**, 5792 (2022).
34. de la Gándara, Á et al. Molecular basis for transposase activation by a dedicated AAA+ ATPase. *Nature* **630**, 1003–1011 (2024).
35. Liu, J. et al. Integration of therapeutic cargo into the human genome with programmable type V-K CAST. *Nat. Commun.* **16**, 2427 (2025).
36. Schargel, R. D. et al. De novo engineered guide RNA-directed transposition with TnpB-family proteins reveals features of naturally evolved systems. Preprint at *bioRxiv* https://doi.org/10.1101/2025.08.27.672191 (2025).
37. Abdelhakim, A. H., Oakes, E. C., Sauer, R. T. & Baker, T. A. Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA+ unfoldase ClpX. *Mol. Cell* **30**, 39–50 (2008).
38. Sumida, K. H. et al. Improving protein expression, stability, and function with ProteinMPNN. *J. Am. Chem. Soc.* **146**, 2054–2061 (2024).
39. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA* **92**, 452–456 (1995).

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

Methods

VchTnsB purification

The codon-optimized gene encoding *VchTnsB* from *V. cholerae* Tn6677 was subcloned into the pARZ4 expression vector carrying a carbenicillin resistance marker, which was engineered to append an N-terminal hexahistidine (His_6) tag followed by a maltose-binding protein (MBP) tag and a TEV protease recognition site. The sequence-verified construct was introduced into One Shot BL21(DE3) chemically competent *E. coli* cells (Invitrogen) for recombinant production.

Expression cultures were grown in 2× YT medium supplemented with 100 μg ml^{-1} carbenicillin at 37 °C with shaking until OD_{600} reached 0.4–0.6. Protein expression was induced by the addition of 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and cultures were subsequently incubated overnight at 16 °C. Cells were collected by centrifugation at 5,000 rpm for 20 min at 4 °C and resuspended in lysis buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM EDTA, 0.1% NP-40, 1 mM DTT, and protease inhibitor cocktail (Thermo Fisher Scientific; one tablet per 30 ml). Cells were lysed by probe sonication on ice using alternating 5 s on–10 s off cycles, and insoluble debris was removed by centrifugation at 16,000 rpm for 30 min at 4 °C.

The clarified lysate was applied to an amylose affinity column (New England Biolabs) pre-equilibrated with wash buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT). The resin was washed with 15–20 column volumes of wash buffer to remove non-specifically bound proteins, and the MBP–VchTnsB fusion was eluted by supplementing the wash buffer with 10 mM maltose. Peak fractions were pooled and diluted to 200 mM NaCl using heparin dilution buffer (50 mM Tris-HCl pH 7.4, 50 mM NaCl, 5% glycerol) before loading onto a HiTrap Heparin HP column (Cytiva) on an ÄKTA Pure FPLC system. Bound protein was eluted over a linear NaCl gradient from 200 mM to 1.1 M in 50 mM Tris-HCl (pH 7.4) and 5% glycerol buffer. Peak fractions were resolved by size-exclusion chromatography on a Superdex 200 column (GE Healthcare) equilibrated in 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, and 5% glycerol. Purity was assessed by Coomassie-stained SDS–PAGE, and purified fractions were aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

For applications requiring tag-free VchTnsB, an alternative construct was generated by engineering a SUMO–Ulp protease cleavage cassette between the His_6–MBP tag and the *VchTnsB* coding sequence. Expression and amylose affinity purification were performed as detailed above. Post elution, the pooled fractions were subjected to overnight Ulp1 protease digestion at 4 °C (1 mg protease per 50–100 mg of target protein) to remove the affinity tag, followed by the same downstream heparin affinity and size-exclusion chromatography workflow under identical buffer conditions.

$VchTnsA_{WT}$ purification

The *VchTnsA* gene, codon-optimized for expression in *E. coli*, was synthesized commercially (Integrated DNA Technologies, IDT) and cloned into a pARZ4 bacterial expression vector carrying a carbenicillin resistance marker. The construct encoded an N-terminal His_6 tag followed by a Strep–SUMO fusion tag and a Ulp1 protease cleavage site to facilitate purification and subsequent tag removal.

The recombinant plasmid was transformed into BL21 competent cells. Transformed cells were cultured in 2× YT medium supplemented with 100 μg ml^{-1} carbenicillin at 37 °C with shaking until the culture reached OD_{600} of approximately 0.4–0.6. Protein expression was induced by the addition of IPTG to a final concentration of 0.25 mM, and cultures were subsequently incubated overnight at 16 °C.

Cells were collected by centrifugation at 5,000 rpm for 30 min at 4 °C and resuspended in Strep lysis buffer containing 50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 5% glycerol, 0.1% NP-40, 1 mM DTT, and protease inhibitor cocktail (1 tablet per 30 ml). Cell lysis was performed by sonication using repeated 5 s on–10 s off cycles on ice. Insoluble material was removed by centrifugation at 16,000 rpm for 30 min at 4 °C, and the clarified supernatant was applied to Strep-Tactin Superflow Plus affinity resin (Qiagen) by gravity flow. The resin was washed with 15–20 column volumes of Strep wash buffer (50 mM Tris-HCl pH 7.4, 500 mM NaCl, 5% glycerol, 1 mM DTT), and bound His_6–SUMO-tagged $TnsA_{WT}$ was eluted in the same buffer supplemented with 0.1% (w/v) desthiobiotin.

Eluted fractions were pooled and subjected to overnight digestion at 4 °C with Ulp1 protease at an approximate 1:100 enzyme-to-protein mass ratio to remove the N-terminal His_6–SUMO tag. Protein recovery following the Strep affinity purification step was modest, and SDS–PAGE analysis indicated the presence of contaminating species as well as minor degradation products. Consequently, tag-free $TnsA_{WT}$ was further purified by heparin affinity chromatography using the protocol described for TnsB purification. Fractions containing purified $TnsA_{WT}$ were identified by SDS–PAGE, pooled, concentrated, and assessed for purity. Despite the additional purification step, overall yields remained low and the protein exhibited a tendency towards degradation during storage and downstream reconstitution experiments. Therefore, freshly purified protein was used whenever possible for biochemical and structural studies. Final aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

Designed VchTnsA purification

The gene encoding the designed VchTnsA variant D1A was cloned into the pARZ4 expression vector encoding an N-terminal His_6 tag and carrying a chloramphenicol resistance marker. The plasmid was transformed into BL21 cells, and protein expression was carried out under the same growth and induction conditions described for VchTnsB. Collected cell pellets were resuspended in His lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT, and protease inhibitor cocktail), lysed by probe sonication on ice, and clarified by centrifugation at 16,000 rpm for 30 min at 4 °C.

The clarified lysate was applied to Ni^{2+}-NTA affinity resin pre-equilibrated with lysis buffer. The resin was washed with His wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 50 mM imidazole, 1 mM DTT), and the His_6–SUMO–D1A fusion was eluted in buffer containing 50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 500 mM imidazole, and 1 mM DTT. Tag removal was accomplished by overnight Ulp1 protease digestion at 4 °C (1 mg protease per 50–100 mg target protein), and complete cleavage was confirmed by SDS–PAGE. The cleaved protein was subsequently polished by the HiTrap Heparin HP column and the Superdex 200 column under the buffer conditions described for VchTnsB. Compared to $VchTnsA_{WT}$, D1A exhibited improved expression, higher overall recovery throughout purification, and reduced susceptibility to degradation during purification, subsequent reconstitutions and storage. Purity was evaluated by Coomassie-stained SDS–PAGE, after which samples were aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

VchTnsAB purification

The *VchTnsAB* fusion construct was expressed as an N-terminal His_6–MBP fusion protein containing a TEV protease cleavage site between the affinity tag and *TnsAB* in a pET vector. The expression plasmid carried carbenicillin resistance and was transformed into BL21 cells.

A single colony was used to inoculate an overnight starter culture in 8–10 ml LB medium supplemented with 100 μg ml^{-1} carbenicillin. Glucose (0.2%) was also added to suppress expression of endogenous maltose-regulated genes, including amylases that can compromise amylose resin performance during purification. The overnight culture was used to inoculate 1 l of 2× YT medium containing carbenicillin, and cells were grown at 37 °C with shaking until OD_{600} reached approximately 0.4–0.6. Cultures were cooled to 18 °C for 30 min before induction with 0.25 mM IPTG, and protein expression was carried out overnight (~16 h) at 18 °C with shaking. Relative to other VchCAST protein components purified in this study, the TnsAB fusion construct

exhibited comparatively low expression levels, resulting in limited recovery of purified protein throughout the purification workflow. Cells were collected by centrifugation at 5,000 rpm for 20 min at 4 °C and resuspended in MBP lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM PMSF, 0.1% NP-40, 1 mM DTT, and protease inhibitor cocktail) at approximately 5 ml per gram of wet cell pellet. Resuspended cells were flash-frozen or stored overnight at −80 °C prior to lysis. Frozen cells were thawed on ice, supplemented with fresh protease inhibitors and PMSF, and lysed by sonication using 5 s on–10 s off cycles on ice. Lysates were clarified by centrifugation at 16,000 rpm for 30 min at 4 °C.

For MBP-affinity purification, approximately 2 ml amylose resin was equilibrated with 5 column volumes of MBP lysis buffer and loaded into a gravity flow column. The clarified lysate was applied to the resin, which was subsequently washed with 15–20 column volumes of MBP wash buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT). Bound His_6–MBP–TnsAB was eluted with MBP elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 10 mM maltose) in 5 sequential 5 ml elutions with approximately 5 min incubation per step. Fractions containing TnsAB were identified by SDS–PAGE and pooled.

Prior to heparin chromatography, pooled amylose fractions were diluted with heparin dilution buffer to reduce the NaCl concentration below 200 mM. The diluted sample was loaded onto a 1 ml heparin column equilibrated in heparin binding buffer (50 mM Tris-HCl pH 7.4, 200 mM NaCl, 5% glycerol), washed with 15–20 column volumes of binding buffer, and eluted with a linear salt gradient to heparin elution buffer (50 mM Tris-HCl pH 7.4, 1.1 M NaCl, 5% glycerol). Size-exclusion chromatography was performed using the procedure and buffer conditions described for TnsB purification. Peak fractions containing purified TnsAB were analysed by SDS–PAGE, pooled, and concentrated. Owing to the comparatively low expression of the TnsAB fusion construct, overall recovery of purified protein was low. During size-exclusion chromatography, a substantial fraction of the protein eluted in the column void volume, consistent with aggregation, whereas only a small proportion eluted as the expected monodisperse species. Consequently, recovery of purified TnsAB following size-exclusion chromatography was limited. Protein concentration was determined by Bradford assay. Final samples were aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

E. coli IHF purification

Recombinant IHF heterodimer was produced using the pACYCDuet-1-ihfαβ expression plasmid (a gift from the Nuñez laboratory), in which the *ihfa* and *ihfb* genes, amplified from *E. coli* BL21 genomic DNA (gDNA), were cloned into the pACYCDuet-1 backbone (Novagen). The ihfα subunit carried an N-terminal His_6 tag followed by a TEV protease recognition site. The construct was introduced into BL21 cells, and expression was carried out under the growth and induction conditions described above.

Cell pellets were collected by centrifugation (5,000 rpm, 20 min, 4 °C) and resuspended in Ni-NTA affinity lysis buffer (20 mM HEPES-NaOH pH 7.4, 500 mM KCl, 10% glycerol, 0.1% Triton X-100, 2 mM DTT, 0.5 mM PMSF, and protease inhibitor cocktail). Cells were disrupted by sonication using 7 cycles of 5 s on–10 s off, and insoluble material was removed by centrifugation at 16,000 rpm for 30 min at 4 °C. The clarified lysate was applied to a HisTrap HP column (Cytiva) on an ÄKTA Pure FPLC system, and bound proteins were resolved using a linear imidazole gradient from buffer A (20 mM HEPES-NaOH pH 7.4, 500 mM KCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) to buffer B (500 mM imidazole). IHF-containing fractions were pooled and incubated overnight at 4 °C with TEV protease to remove the His_6 tag. The heterodimer was subsequently polished by heparin affinity chromatography using a linear KCl gradient from 0.25 M to 1 M, generated by mixing low-salt and high-salt buffers sharing the same base composition (20 mM HEPES-NaOH pH 7.4, 10% glycerol, 1 mM DTT) with 250 mM and 1 M KCl, respectively. Peak fractions were pooled, concentrated using a 3 kDa molecular weight cut-off concentrator, and resolved by size-exclusion chromatography on a Superdex 75 16/60 column (GE Healthcare) equilibrated in 20 mM HEPES-NaOH (pH 7.4), 500 mM KCl, 5% glycerol, and 1 mM DTT. Purity was confirmed by Coomassie-stained SDS–PAGE, and purified IHF was aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

VchTnsC purification

The codon-optimized *VchTnsC* gene was synthesized (IDT) and subcloned into the pARZ4 expression plasmid carrying carbenicillin resistance. The construct encoded an N-terminal His_6 tag followed by a Strep–SUMO solubility tag and a Ulp1 protease recognition site. The plasmid was introduced into BL21 cells, and expression was performed under the growth and induction conditions described for VchTnsB. Cells were pelleted by centrifugation at 5,000 rpm for 20 min at 4 °C and resuspended in high-salt Strep lysis buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 5% glycerol, 0.1% NP-40, 1 mM DTT, and protease inhibitor cocktail). Cells were disrupted by sonication using alternating 5 s on–10 s off cycles, and debris was pelleted by centrifugation at 16,000 rpm for 30 min at 4 °C.

The clarified lysate was passed over Strep-Tactin Superflow Plus resin (Qiagen) by gravity flow. Non-specifically bound proteins were removed by washing with 15–20 column volumes of high-salt Strep wash buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 5% glycerol, 1 mM DTT), and the tagged protein was eluted in the same buffer supplemented with 0.1% (w/v) desthiobiotin. Eluted fractions were pooled and treated overnight at 4 °C with Ulp1 protease at a 1:100 (w/w) enzyme-to-substrate ratio. Tag-free TnsC was resolved by size-exclusion chromatography on a Superdex 200 column equilibrated in high-salt gel filtration buffer (20 mM HEPES pH 7.4, 1 M NaCl, 5% glycerol, 1 mM DTT). Fractions corresponding to the monomeric TnsC peak were pooled, concentrated, verified for purity by SDS–PAGE, aliquoted, and stored at −80 °C.

VchCascade–TniQ purification

The VchCascade–TniQ complex was produced using a pHelper-based plasmid encoding Cas8/5, Cas7, Cas6, His_6-tagged TniQ, and a CRISPR RNA (crRNA)–ffsx6 array encoding a *ccdB*-targeting guide RNA. The plasmid was transformed into *E. coli* T7 Express cells (New England Biolabs), and large-scale cultures were propagated in 2× YT medium with chloramphenicol at 37 °C to OD_{600} 0.6, induced with 0.25 mM IPTG, and expressed overnight at 16 °C. Cell pellets were resuspended in His lysis buffer and lysed as described for VchTnsA. The Cascade–TniQ complex was initially captured via the His_6 tag on TniQ using Ni-NTA affinity chromatography following the same purification protocol as performed for designed VchTnsA. Pooled elution fractions were subjected to anion-exchange chromatography on a Mono Q column (Cytiva) pre-equilibrated in binding buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM DTT). Proteins were eluted over a linear NaCl gradient to 1 M in elution buffer (20 mM Tris-HCl pH 8.0, 1,000 mM NaCl, 5% glycerol, 1 mM DTT), and fractions containing the Cascade–TniQ complex were identified by SDS–PAGE, pooled and concentrated.

To generate tag-free Cascade–TniQ, a pHelper construct incorporating a SUMO-Ulp1 cleavage cassette between the His_6 tag and TniQ was additionally cloned. Expression and purification were carried out identically to the procedure described above, with tag-free Cascade–TniQ obtained following overnight Ulp1 digestion after Ni-NTA and Mono Q chromatography. Protein integrity and purity were verified by SDS–PAGE, and samples were aliquoted and stored at −80 °C after flash-freezing in liquid nitrogen.

In vivo transposition assay

In vivo transposition assays were performed in *E. coli* cJP003 recipient cells, which carry an arabinose-inducible *ccdB* toxin cassette integrated at the intended genomic target site[23]. Two compatible pHelpers were

co-transformed into the recipient strain: one encoding the VchCAST transposition machinery (TnsB, TnsC and either TnsA or D1A) with carbenicillin resistance (adapted from Addgene #130633), and a second encoding the VchCAST targeting machinery (TniQ, Cas8/5, Cas7, Cas6 and *ccdB*-targeting single guide RNA (sgRNA)) with chloramphenicol resistance (adapted from Addgene #130635). For the assays of mutants, point mutations were introduced into pHelpers by site-directed mutagenesis using KLD Enzyme Mix (NEB, M0554S). Primers used for mutagenesis are listed in Supplementary Table 2.

Freshly transformed cells were first grown in 5 ml LB medium, inoculated into 50 ml LB medium containing carbenicillin, chloramphenicol, and 0.1 mM IPTG. Cultures were grown at 37 °C with shaking until the OD_{600} reached 0.6–0.8. Cells were collected, washed 3 times with ice-cold 10% glycerol, and normalized to equivalent cell densities. Recipient cells were electroporated with a non-replicating VchCAST pDonor carrying an R6K origin of replication at 5:1 plasmid-to-cell ratio. Cells were immediately resuspended in 1 ml SOC medium, recovered at 37 °C with shaking for 1 h, and incubated overnight at 30 °C without shaking to promote transposition. Cells were then diluted to obtain countable colonies and plated on LB agar plates containing 50 μg ml^{-1} kanamycin, 100 μg ml^{-1} carbenicillin, and 30 μg ml^{-1} chloramphenicol, with or without 0.1% (w/v) L-arabinose to assess overall activity (kanamycin) or on-target activity (kanamycin plus arabinose), respectively.

AlphaFold 3 modelling of VchCAST STC

A structural model of the TnsA/B STC with target and donor DNA was generated using AlphaFold 3[21] and used as input for structure-based sequence design. Protein and DNA substrate sequences are provided in Supplementary Table 3.

Computational design of TnsA

The AlphaFold 3-generated TnsA–TnsB STC model was used as input for structure-based protein design. A total of 74 TnsA residue positions were fixed during design based on five structural criteria: (1) residues predicted to participate in primary or secondary interactions with the active site; (2) residues within contact distance of the DNA; (3) residues predicted to form protein–protein interfaces with TnsB; (4) residues predicted to contribute to a binding pocket engaging the TnsB or TnsC hook regions; and (5) N- and C-terminal residues with low predicted confidence (pLDDT), indicative of structural disorder. The full set of designable positions is listed in Supplementary Table 4.

Sequences were generated using three graph neural network-based protein design methods that sample amino acid probabilities from local structural context: LigandMPNN[22], SolubleMPNN[40] and HyperMPNN[41]. Because these models are trained on structural rather than functional information, and because design was performed on an AlphaFold 3 model rather than an experimentally determined structure, a bias towards TnsA sequences was introduced to restrict sampling to structurally favourable substitutions and improve the likelihood of recovering functional variants. Bias was applied by modifying per-residue logits prior to softmax normalization, such that increasing bias values progressively shifted the amino acid probability distribution towards the native identity without fully enforcing it. Bias strengths of 0, 1, 2, 3 and 4 were applied to each of the three models using the flag bias_AA_per_residue. In a parallel strategy, an additional sampling bias was introduced towards residues conserved in a multiple sequence alignment of TnsA homologues.

Structure prediction and design filtering

All designed sequences were folded as monomers and as part of the STC using Chai-1[42]. Designed sequences were clustered using MMseqs2[43], and final designs were selected from each strategy based on threshold criteria across six metrics: backbone root mean square deviation (RMSD) of monomeric TnsA relative to input structure (<0.9–1.2 Å, depending on design strategy), backbone RMSD of predicted STC complex (<1.5–1.7 Å), predicted TM-score of the monomer (>0.918–0.928), predicted interface TM-score of the STC (>0.839–0.845), predicted local distance difference test (pLDDT) of the monomer (>0.893–0.910), and pLDDT of the STC (>0.834–0.840).

In total, 499 designed TnsA variants were selected for experimental screening. One additional sequence was independently generated using PROSS[44], yielding a final library of 500 sequences (Supplementary Table 1).

Generation of a VchTnsA library

A TnsA entry vector was constructed by replacing the $TnsA_{WT}$ region of the VchCAST TnsABC pHelper (Addgene #130633) with BsaI flanking sites. Nucleotide sequences of TnsA designs (693 bp) were fragmented into 300 bp oligonucleotides containing BsaI flanking sites, unique primer-binding sites, and unique per-design barcodes using the OMEGA program[45] and purchased from Twist Bioscience. The oligonucleotide pools were PCR-amplified and inserted into the entry vector by Golden Gate assembly at a 20:1 insert-to-vector molar ratio. The assembled TnsA library was purified and electroporated into NEB 10-beta electrocompetent *E. coli* (NEB, #C3020K). Cells were plated on agar plates containing 100 μg ml^{-1} carbenicillin after 1 h recovery at 37 °C. Colonies were collected at greater than 100× coverage per variant to ensure adequate library representation, and the TnsA library plasmid pool was extracted using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research, D4201).

High-throughput screening of VchTnsA designs

cJP003 cells were transformed with a Cascade–TniQ pHelper plasmid encoding TniQ, Cascade, a *ccdB*-targeting guide RNA, and a chloramphenicol resistance marker. Transformed cells were rendered electrocompetent by three successive washes with ice-cold 10% glycerol. The TnsA library was electroporated into the prepared cells at a 0.5:1 plasmid-to-cell ratio to ensure that each cell received a unique TnsABC pHelper variant. Cells were recovered at 37 °C for 1 h and plated on agar containing 100 μg ml^{-1} carbenicillin and 30 μg ml^{-1} chloramphenicol. Colonies were collected at greater than 100× coverage per variant and cultured at 37 °C in LB medium supplemented with 100 μg ml^{-1} carbenicillin and 30 μg ml^{-1} chloramphenicol.

Upon reaching $OD_{600} \approx 0.6$, cells were collected and washed 3 times with 10% glycerol to prepare electrocompetent recipient cells. At this step, pHelper plasmids from a subset of recipient cells were extracted as input reference samples using the Monarch Spin Plasmid Miniprep Kit (NEB, T1110L). Recipient cells were electroporated with the non-replicable VchCAST pDonor at a 5:1 plasmid-to-cell ratio, immediately resuspended in 1 ml SOC medium, and recovered at 37 °C with shaking for 1 h. Cells were subsequently incubated overnight at 30 °C without shaking. The following day, cells were plated on agar containing 100 μg ml^{-1} carbenicillin, 50 μg ml^{-1} kanamycin, and 30 μg ml^{-1} chloramphenicol for the activity screen, or on the same agar additionally supplemented with 0.1% (w/v) L-arabinose for the specificity screen. pHelper plasmids were extracted from selected colonies collected at greater than 100× coverage per variant using the ZymoPURE II Plasmid Midiprep Kit. All steps were performed in biological duplicate.

Barcode regions from the input reference and screened plasmid pools were PCR-amplified and indexed using the Nextera XT Index Kit v2 Set A (Illumina, FC-131-2001). Primers used for NGS are provided in Supplementary Table 5. Indexed amplicons were sequenced on an in-house MiSeq instrument. Activity and specificity scores for each design were calculated using an Enrich2-based pipeline[23,46] on a $\log_2$-fold change basis and normalized to the score of $TnsA_{WT}$.

Genomic profiling

E. coli cells that had undergone in vivo transposition assays were plated on agar containing 100 μg ml^{-1} carbenicillin, 50 μg ml^{-1} kanamycin and 30 μg ml^{-1} chloramphenicol to select for all integration

events. More than 1,000 colonies were scraped for gDNA extraction using the Wizard Genomic DNA Purification Kit (Promega, A1125). Tagmentation-based transposon insertion sequencing (TagTn-seq)[11] was performed on the extracted gDNA. Libraries were prepared by fragmenting gDNA with the Nextera XT DNA Library Kit (Illumina, FC-131-1024), followed by PCR amplification using an i7 primer and VchCAST right-end (RE)-specific primers, and indexing prior to sequencing on the in-house MiSeq. Reads mapping to the intact VchCAST pDonor were filtered from the resulting sequencing data using BBMap (v39.80), and RE-adjacent sequences were mapped to the cJP003 genome to profile integration site distribution.

Co-integrate analysis

TnsA activity is assessed through the formation of simple insertion products, as TnsA-mediated DNA cleavage results in cleavage of the 5′ end of donor DNA and lack of cleavage results in integration of the donor DNA backbone, otherwise referred to as co-integrates. We used long-read sequencing to distinguish between these two products, as previously established[8].

E. coli cells that had undergone in vivo transposition assays were plated on agar containing 100 μg ml^{-1} carbenicillin, 50 μg ml^{-1} kanamycin, 30 μg ml^{-1} chloramphenicol, and 0.1% (w/v) L-arabinose to select specifically for on-target integration events. More than 1,000 colonies were scraped and gDNA was extracted using the Wizard Genomic DNA Purification Kit (ProMega, A1125). Whole-genome sequencing was performed by Plasmidsaurus. Sequencing reads were analysed to identify integration products by detecting reads containing both target-site and left-end (LE) sequences, which define the boundaries of the integrated region. Reads containing pDonor backbone sequence were classified as co-integrates, while reads lacking pDonor backbone but containing the downstream target-site sequence were classified as simple insertions. Reads lacking both features were classified as ambiguous.

Circular dichroism spectroscopy

Far-UV circular dichroism spectra were collected using a JASCO CD spectropolarimeter. WT-TnsA and D1A were buffer exchanged into 10 mM potassium phosphate (pH 7.4) and analysed at a final concentration of 15 μM using a 0.1 cm path length quartz cuvette. Spectra were recorded from 195 to 280 nm at room temperature in continuous scanning mode with a 1.0 nm bandwidth and a scan speed of 200 nm min^{-1}. Eight accumulations were collected for each sample and averaged to improve the signal-to-noise ratio. A buffer-only spectrum was recorded under identical conditions and subtracted from each protein spectrum.

The corrected circular dichroism signal was converted to mean residue ellipticity according to the manufacturer's instructions:

$$[\theta] = \frac{\theta_{obs}}{10 \times n \times c \times l}$$

where θ_{obs} is the buffer-subtracted ellipticity, n is the number of amino acid residues, c is the protein concentration (M), and l is the cuvette path length (cm). WT-TnsA and D1A were normalized using 229 and 236 amino acid residues, respectively. For plotting, mean residue ellipticity values were divided by 1,000 and are reported as $[\theta] \times 10^{-3}$ (deg cm^2 dmol^{-1}).

Preparation of DNA substrates for complex reconstitution

The DNA substrate used was designed based on in vivo genomic profiling experiments, synthesized as single-stranded DNA, and annealed to form a strand-transfer product substrate which was used for complex reconstitution, following our previously described strategy[12,13]. The left end spans LE1 through LE3 (positions 1–109) and the right end spans RE1 through RE3 (1–72), as shown in Supplementary Fig. 4. The native arrangement of left and right ends are identical up to position 30 (encompassing the 8 bp terminal sequence and LE1/ RE1). Past the first binding site, the placement and spacing of LE2–LE3 and RE2–RE3 differ, additionally there is a IHF binding site between LE1 and LE2 (Supplementary Fig. 4).

Two double-stranded DNA substrates were independently assembled for the STC and transpososome reconstitution: a LE fragment and a RE fragment, each designed to recapitulate the relevant structural elements of the VchCAST RNA-guided transposition product, including the transposon terminal repeat sequences, the 5 bp target site duplication, and flanking target and gDNA regions (Supplementary Table 6). A diagram of the substrates can be found in Supplementary Fig. 4.

The LE fragment was assembled from three oligonucleotides: LE–TGS, which contains the transposon LE and flanking gDNA, Target-RC, which completes the double-stranded target genomic region, and LE–HFR, which provides the complementary LE region to LE–TGS along with a host flanking region for TnsA binding.

The RE fragment was assembled from four oligonucleotides: RE–TGS, which contains the RE and the target strand for guide RNA binding, NTGS–Scramble, a non-target strand oligonucleotide contained a scrambled sequence to facilitate bubble formation, RE–HFR, an oligonucleotide that complements the RE sequence of RE–TGS and contains a flanking HFR region, and DTB–LUEGO, a desthiobiotinylated oligonucleotide incorporated at the 5′ end of the RE fragment to enable streptavidin-based affinity capture of the assembled complexes during pull-down experiments[47].

Prior to annealing, all oligonucleotides were purified by HPLC using an Agilent 1260 Infinity II system equipped with an AdvanceBio Oligonucleotide column (Agilent Technologies) to ensure incorporation of only full-length species. The LE and RE fragments were each assembled independently by combining their respective oligonucleotides at equimolar concentrations in annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA), heating to 95 °C for 5 min, and gradually cooling to 50 °C at 1 °C per min using a thermal cycler (Bio-Rad). The annealed fragments were stored separately until used in STC and transpososome reconstitutions.

STC assembly

STC assembly reactions were prepared using an annealed DNA substrate (0.5 μM in 100 μl) and containing TnsB together with D1A using the same 1:5:10 stoichiometry. Complex assembly was initially carried out under high-salt conditions in a buffer containing 26 mM HEPES (pH 7.5), 5 mM Tris-HCl (pH 7.5), 20 mM KCl, 750 mM NaCl, 0.2 mM $MgCl_2$, 15 mM magnesium acetate, 2% glycerol, 1 mM DTT and 0.1% Tween-20. Following assembly, samples were dialysed overnight at 4 °C from high-salt buffer into an identical low-salt buffer in which the NaCl concentration was reduced to 150 mM to promote stable complex formation.

For affinity capture of assembled complexes, Streptavidin Mag Sepharose magnetic beads (Cytiva) were washed 3 times with 500 μl of low-salt buffer prior to use. Dialysed samples were incubated with the equilibrated streptavidin beads for 1 h at room temperature with continuous rotation to allow binding of biotinylated DNA substrates and associated protein complexes.

Following incubation, the beads were washed 3 times with 500 μl of low-salt buffer to remove unbound proteins and nonspecific material. Bound complexes were subsequently eluted using low-salt buffer supplemented with 10 mM biotin. Elution was performed in 2 sequential rounds using 50 μl buffer per step, with each incubation carried out for 30 min at room temperature under rotation before collection of the eluate. Elution fractions were analysed by SDS–PAGE to confirm the presence of protein components within the assembled complexes. Sample quality, homogeneity and the presence of assembled STCs were further evaluated by negative-stain electron microscopy prior to cryo-EM grid preparation and vitrification.

HIC on-bead assembly

Sequential on-bead assembly of the HIC was performed by first capturing the desthiobiotinylated RE substrate (0.5 μM in 100 μl) on

Streptavidin Mag Sepharose magnetic beads (Cytiva). Binding was carried out for 1 h at room temperature with gentle rotation in binding buffer comprising 26 mM HEPES (pH 7.5), 5 mM Tris-HCl (pH 7.5), 20 mM KCl, 750 mM NaCl, 0.2 mM $MgCl_2$, 15 mM MgOAc, 2% glycerol, 1 mM ATP, 1 mM DTT, and 0.1% Tween-20, followed by 3 washes with 500 µl of the same buffer.

Cascade–TniQ was subsequently introduced at a 1:2 DNA:protein molar ratio and incubated for 1 h at room temperature, followed by 3 washes. The LE substrate (0.5 µM), pre-complexed with *E. coli* IHF at a 1:8 DNA-to-protein molar ratio, was then added alongside TnsB (1:5 DNA:protein) and designed TnsA (1:10 DNA:protein), and the assembly was allowed to proceed for 1 h at room temperature. Beads were washed 3 times with 500 µl binding buffer. In the final step, TnsC was introduced at a 1:10 DNA-to-protein ratio in binding buffer supplemented with equimolar concentrations of ATP and ATPγS (1 mM each) and incubated for 1 h at room temperature. Unbound proteins were removed by 3 washes with 500 µl of ATP/ATPγS-containing binding buffer.

The assembled HIC was released from the beads by two sequential elution steps, each using 50 µl of elution buffer (binding buffer supplemented with 10 mM biotin) incubated under rotation for 30 min. Eluted fractions were resolved by SDS–PAGE to verify the presence of all HIC components. Prior to cryo-EM grid preparation, samples were evaluated by negative-stain electron microscopy to confirm the presence of assembled HIC particles and to assess sample homogeneity.

Preparation of graphene-coated cryo-EM grids

Graphene sheets with a poly(methyl methacrylate) (PMMA) support layer (ACS Material, TTG100BB,25 ×25 mm) were stored at 4 °C until use. Graphene transfer onto cryo-EM grids was performed based on a previously described protocol with minor modifications[48]. Ultrafoil grids (Au R1.2/1.3, Quantifoil) were placed on a 3D-printed grid transfer tool immersed in deionized water. Excess support material from the graphene pad was trimmed, and the PMMA/graphene pad was placed gently onto the water surface to avoid overflow onto the graphene layer. After positioning, the water level was gradually lowered by syringe withdrawal, allowing the graphene sheet to settle onto the grids. Grids were left overnight at room temperature to anneal, then baked at 150 °C for 30 min to promote firm adhesion. Grids were transferred to a Petri dish, and PMMA was removed by three successive acetone washes. Graphene-coated grids were individually transferred to a grid box and stored at room temperature until use.

Cryo-EM sample preparation

STC samples were prepared for cryo-EM either directly from pull-down elution fractions or following two- to four-fold dilution in low-salt buffer. Graphene-coated grids were prepared as described above. Prior to sample application, grids were plasma-cleaned using a Solarus II Plasma Cleaner (Gatan) under an Ar/O_2 atmosphere at 10 W for 10 s (s), with the graphene-coated surface oriented upward.

For vitrification, 4 µl of STC sample was applied to the carbon side of a graphene-coated grid and incubated for 60 s inside a Vitrobot Mark IV chamber (Thermo Fisher Scientific) maintained at 4 °C and 100% relative humidity. Grids were blotted for 3.5 s at a blot force of −7 and immediately plunge-frozen into liquid ethane cooled by liquid nitrogen.

HIC samples were prepared for cryo-EM from either undiluted pull-down eluates or two- to four-fold dilutions thereof. Graphene-coated grids were prepared as described above and plasma-cleaned using a Solarus II Plasma Cleaner (Gatan) under an Ar/O_2 atmosphere at 10 W for 10 s with the graphene-coated face oriented upward.

For vitrification, 4 µl of HIC sample was applied to the carbon side of a graphene-coated grid and incubated for 60 s inside a Vitrobot Mark IV chamber at 4 °C and 100% relative humidity. Grids were blotted for 3.5 s at a blot force of −7 before plunge-freezing into liquid ethane cooled by liquid nitrogen.

Ni-NTA affinity grids were prepared as previously described[49]. Quantifoil Au R1.2/1.3 grids were coated with a Ni-NTA lipid monolayer composed of 20% 18:1 DGS-NTA(Ni) (Avanti Polar Lipids) diluted with DOPC (18:1; Avanti Polar Lipids). Grids were incubated with 3 µl of His-tagged HIC diluted in 2 µl of elution buffer (26 mM HEPES pH 7.5, 5 mM Tris-HCl pH 7.5, 20 mM KCl, 750 mM NaCl, 0.2 mM $MgCl_2$, 15 mM MgOAc, 2% glycerol, 1 mM ATP, 1 mM DTT, 0.1% Tween-20) for 1 h at 4 °C in a humidity chamber to allow affinity capture. Unbound material was removed by three brief washes with elution buffer, after which grids were blotted and plunge-frozen in liquid ethane using a Vitrobot operated at 4 °C and 100% relative humidity with a blot force of −7 and blot time of 3.5 s.

Cryo-EM imaging and image processing of the STC

Vitrified STC samples were initially screened on a Talos Arctica transmission electron microscope (Thermo Fisher Scientific) operated with fringe-free illumination at 130,000× magnification (0.63 Å per pixel) and 200 kV, equipped with a Gatan K3 direct electron detector and a BioQuantum imaging filter (20 eV slit width). Data were collected using the EPU software package (Thermo Fisher Scientific). This screening step was used to assess grid quality and particle distribution prior to large-scale data collection. Grids of sufficient quality were subsequently imaged on a Titan Krios G3i transmission electron microscope (Thermo Fisher Scientific) operated with fringe-free illumination at 130,000× magnification (0.745 Å per pixel) and 300 kV, equipped with a Falcon 4i (4k × 4k) direct electron detector, a BioQuantum imaging filter, and a 20 eV energy-selecting slit. Data collection employed the aberration-free image shift and multi-shot-per-hole acquisition modes of EPU. For the STC dataset, 7,724 movies were collected using EPU with image shift, with nominal defocus ranging from −1.5 to −2.5 µm at 0.2 µm intervals. Each movie was acquired with a 3.1 s exposure and a total dose of 60 $e^- Å^{-2}$, fractionated into 60 frames. Frames were aligned using MotionCor2[50], and CTF estimation was performed in cryoSPARC v4.7.1[51]. The full data processing workflow is presented in Supplementary Fig. 4.

Particles were initially picked using blob picker, and approximately 290,000 extracted particles were subjected to 2D classification. Classes representing the STC were identified and used as templates for a second round of template-based particle picking, yielding approximately 1,000,000 particles that were re-subjected to 2D classification. The resulting 22,000 particles were used to train a Topaz neural network picking model (v0.2.5)[52]. After confirming model performance, additional particles were extracted from all micrographs using the trained model. The new particle set underwent 2D classification, and selection of high-quality classes yielded 97,972 particles that were used for ab initio reconstruction. Heterogeneous refinement separated the dataset into two classes: 70,006 particles representing the STC complex and the remainder classified as non-particles. Non-uniform refinement of the STC class produced an initial map at 2.9 Å resolution. The particle stack was then subjected to 3D classification using principal component analysis with a 13 Å resolution cut-off, separating 18,831 particles lacking target DNA from 51,169 particles with target DNA bound. Both subsets were independently subjected to non-uniform refinement, yielding a 3.2 Å map for the STC without target DNA and a 2.9 Å map for the STC with target DNA. Both maps were further processed with DeepEMhancer v121322[53] prior to model building.

Cryo-EM imaging and image processing of HIC

Initial cryo-EM data collection on graphene grids revealed a strong preferred orientation bias, limiting the angular coverage of the dataset. To improve particle orientation distribution and enable higher-quality three-dimensional reconstructions, Ni-NTA lipid monolayer grids were subsequently prepared[49].

A total of 144,294 movies were collected at a pixel size of 0.848 Å and processed using Patch Motion Correction and Patch CTF Estimation with a CTF-fit cut-off of less than 10 Å. Initial particle picking was

performed by blob picking on approximately 2,600 micrographs, yielding ~570,000 extracted particles (binned 3×), which were subjected to two rounds of 2D classification to generate ~5,000 representative particles for training a Topaz neural network picking model (v0.2.5)[52]. The trained model was applied across all ~22,000 micrographs, extracting 2,722,447 particles (binned 3×). Following two rounds of 2D classification, 897,848 particles were retained for further analysis. The complete data processing workflow is described in Extended Data Fig. 4.

Single-particle analysis revealed substantial compositional heterogeneity within the reconstituted sample. Two-dimensional class averages resolved density for all expected components in multiple orientations (Extended Data Fig. 4). Heterogeneous refinement (n = 4) partitioned the dataset into 4 populations: 2 HIC classes (131,242 particles, 14.6%; and 462,255 particles, 54.8%), 1 STC–TnsC class (127,845 particles, 14.2%) and 1 Cascade class (146,705 particles, 16.4%). These populations enabled classification of the following assemblies: a Cascade–TniQ sub-complex (3.3 Å), a STC (2.9 Å) an STC–TnsC assembly (4.3 Å) and the complete HIC (4.5 Å) (Extended Data Fig. 4). Each population was further processed by sub-classification and homogeneous refinement. For the HIC, signal subtraction of the STC–TnsC region followed by 2D slice projection-guided particle sorting enabled focused local refinement, yielding a map at 4.5 Å resolution (Fourier shell correlation (FSC) = 0.143). Non-uniform refinement of the Cascade region within the full complex achieved 3.5 Å (FSC = 0.143). Free Cascade and Cascade–TniQ populations were resolved to 2.8 Å, 3.0 Å (closed) and 3.1 Å (open), respectively (FSC = 0.143). The HIC exhibits variable local resolution (3.7–9.0 Å, Supplementary Fig. 9), and focused refinements were combined into a composite map.

Model building and refinement

For the STC maps, PDB 9Y7B was used as a starting model. The structure was initially docked into the STC-without-target-DNA map and manually rebuilt in Coot (v0.9.8.96)[54] where required. DNA sequences were modified manually and fitted into the density using ISOLDE as a ChimeraX plugin (v1.11.1). The TnsA subunit was modelled using AlphaFold 3 and docked into the corresponding density. This assembled STC model was subsequently used as the starting model for the STC-with-target-DNA map. Structural rearrangements in TnsB required individual domains to be docked independently into the density before rejoining in Coot.

For the Cascade–TniQ complex structures, individual Cas subunits were modelled using AlphaFold 3 and docked into the map. The DNA scaffold was derived from PDB 6PIJ and mutated to match the designed substrate sequence. The target recognition complex was assembled, and manual adjustments were performed in Coot.

Both the Cascade–TniQ and STC complex structures were used as starting models for modelling the HIC. TnsC protomers were individually modelled using AlphaFold 3 and docked into the density. The STC–TnsC portion of the HIC was used as the starting model for the off-pathway complex, with minor adjustments applied to the TnsC ring.

For all structures, real-space refinement of DNA substrates was performed in Phenix (v2.0rc1-5690)[55] with base-pair and secondary structure restraints enforced. Model validation was carried out by computing map-model FSC at a 0.5 cut-off. The masked cross-correlation coefficient (CCmask) from refinement output was reported as the representative model-map cross-correlation. All validation statistics are summarized in Extended Data Table 1.

Helical symmetry analysis

Helical rise and twist were measured from docked protomers in the HIC (PDB: 36AU) and VchTnsC structures (PDB: 7UFI) using a script from the Rosetta software suite: make_symmdef_file.pl[56]. Rise values (in Å) were extracted from the output and plotted using Python.

Reporting summary

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

Data availability

The cryo-EM map and atomic coordinates reported in this study have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB) under accession codes EMD-77364 (STC class 1), EMD-77365 (STC class 2), 36BO (STC class 1), 36BP (STC class 2), EMD-77366 and 36BQ (target recognition complex), EMD-77335 and 36AU (HIC), EMD-77313 and 36AB (HIC, focused on Cascade-Q), and EMD-77370 and 36BV (STC–TnsC). Previously published structural datasets used in this study are available from the Protein Data Bank under accession codes 9Y7B, 6PIJ, 7UFI, 9BW1, 8EA3 and 8FF4. The raw sequencing data are available from NCBI Sequence Read Archive (accession PRJNA1472095). All other data supporting the findings of this study can be found in the Supplementary materials.

40. Goverde, C. A. et al. Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
41. Ertelt, M. et al. HyperMPNN-a general strategy to design thermostable proteins learned from hyperthermophiles. Preprint at *bioRxiv* https://doi.org/10.1101/2024.11.26.625397 (2024).
42. Boitreaud, J. et al. Chai-1: decoding the molecular interactions of life. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.10.615955 (2024).
43. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
44. Goldenzweig, A. et al. Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* **70**, 380 https://doi.org/10.1016/j.molcel.2016.06.012 (2016).
45. Freschlin, C. R., Yang, K. K. & Romero, P. A. Scalable and cost-efficient custom gene library assembly from oligopools. *Sci. Adv.* **12**, eady2279 (2026).
46. Rubin, A. F. et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).
47. Jullien, N. & Herman, J.-P. LUEGO: a cost and time saving gel shift procedure. *Biotechniques* **51**, 267–269 (2011).
48. Ahn, E. et al. Batch production of high-quality graphene grids for cryo-EM: cryo-EM structure of *Methylococcus capsulatus* soluble methane monooxygenase hydroxylase. *ACS Nano* **17**, 6011–6022 (2023).
49. Skrajna, A. et al. Nickel-NTA lipid-monolayer affinity grids allow for high-resolution structure determination by cryo-EM. *J. Struct. Biol.* **217**, 108253 (2025).
50. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
51. 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).
52. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
53. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
54. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078 (2020).
55. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
56. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

Acknowledgements We thank I. Chen, A. Zanghellini, T. Şevik and the other members of the Kellogg laboratory for helpful discussions related to cryo-EM imaging, analysis and manuscript feedback; and A. Haque for support with cryo-EM imaging. We also thank the St Jude Hartwell Center for Bioinformatics and Biotechnology which is supported in part by ALSAC and the National Cancer Institute grant P30 CA021765 for their help conducting next-generation sequencing, the St Jude Cryo-EM Center, and St Jude Research Information Services for providing high-performance scientific computing systems and computational resources that contributed to the research in this manuscript. The authors used Claude (Anthropic) to assist with manuscript editing and language refinement. All scientific content, ideas, and interpretations are solely the work of the authors. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions E.H.K. conceived the project, supervised the research and contributed to manuscript writing. S.F. purified recombinant proteins, prepared DNA substrates, conducted substrate annealing, performed reconstitutions, prepared cryo-EM samples, performed Ni-NTA grid preparation and freezing, and conducted site-directed mutagenesis, in vivo transposition assays and biophysical characterization, contributed to structural interpretation, and assisted with manuscript writing and assembly. S.D. prepared graphene grids, vitrified cryo-EM samples, collected cryo-EM data, performed structure determination and model building, prepared figures, contributed to structural interpretation, and assisted with manuscript assembly. V.H.T. designed DNA substrates, assisted with cryo-EM data collection, assisted with

atomic model refinement and contributed to figures. S.G.P. conducted high-throughput screening of TnsA designs, co-integrate analysis and in vivo transposition assays. H.M.W. and I.K. developed TnsA stabilization pipelines and generated sequences for TnsA designs. H.M.W. conducted bioinformatic analyses of TnsA design hits. A.J.F.A. performed cryo-EM data analysis, advised on atomic model building and refinement, and contributed to structural interpretation. All authors contributed to manuscript writing and revisions.

Funding This research is supported by the Pew Charitable Trusts (E.H.K.), NIH NIGMS 7R01GM144566 (E.H.K.), NIH NIGMS F31GM151863 (V.H.T.), the Cystic Fibrosis Foundation (E.H.K.), and ALSAC (E.H.K.).

Competing interests St Jude Children's Research Hospital intends to file a US provisional patent application based on this work.

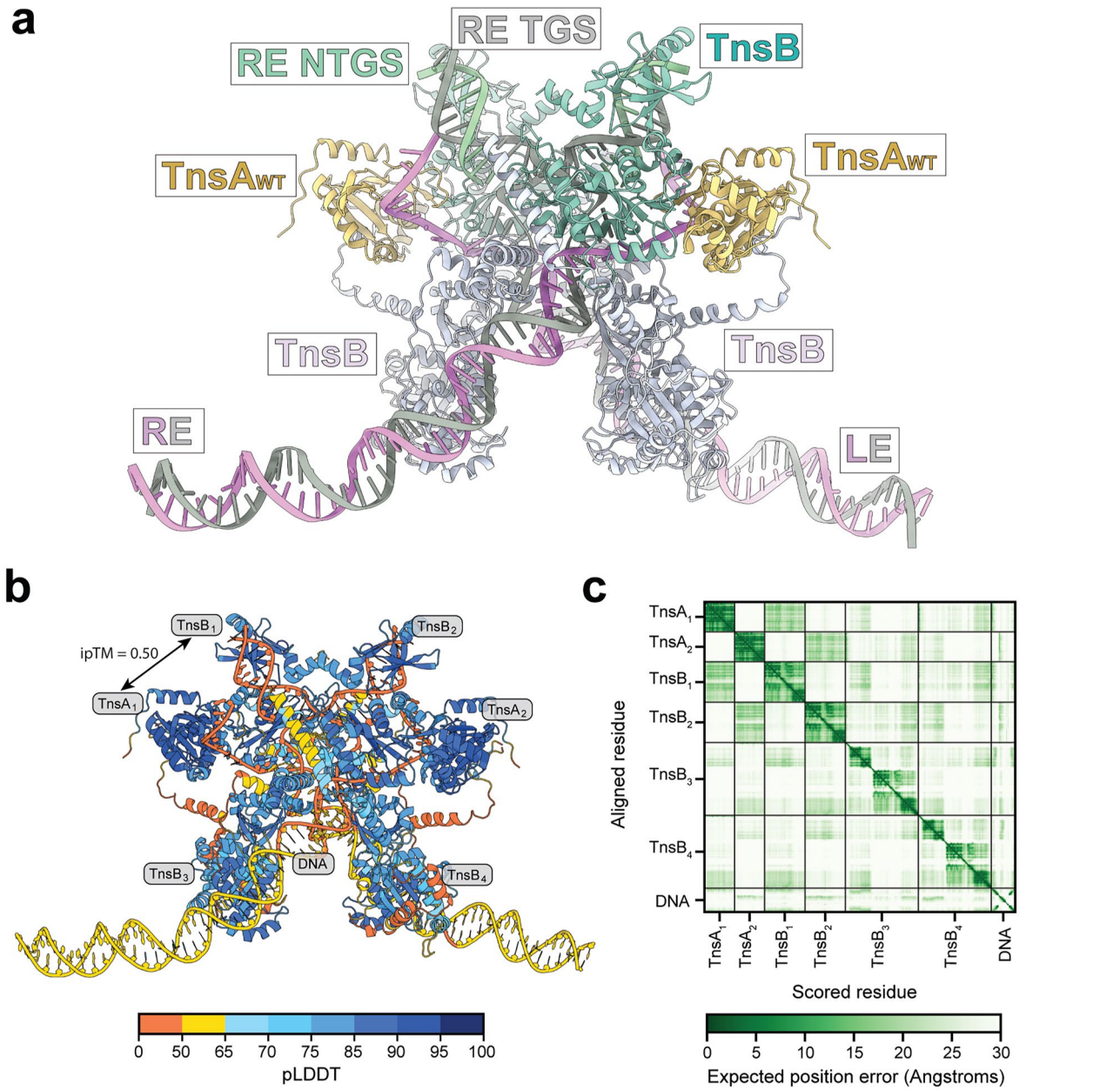
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-11287-9.

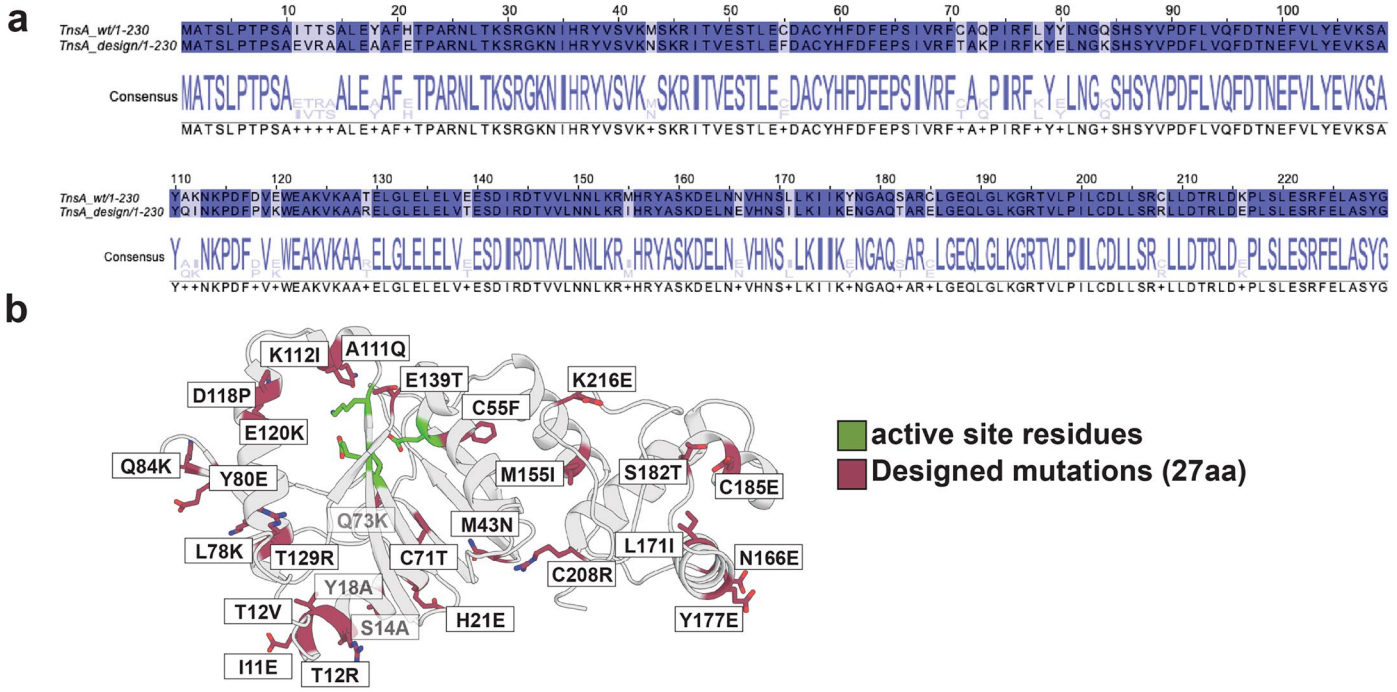
Correspondence and requests for materials should be addressed to Elizabeth H. Kellogg.

Peer review information *Nature* thanks Jonathan Strecker and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

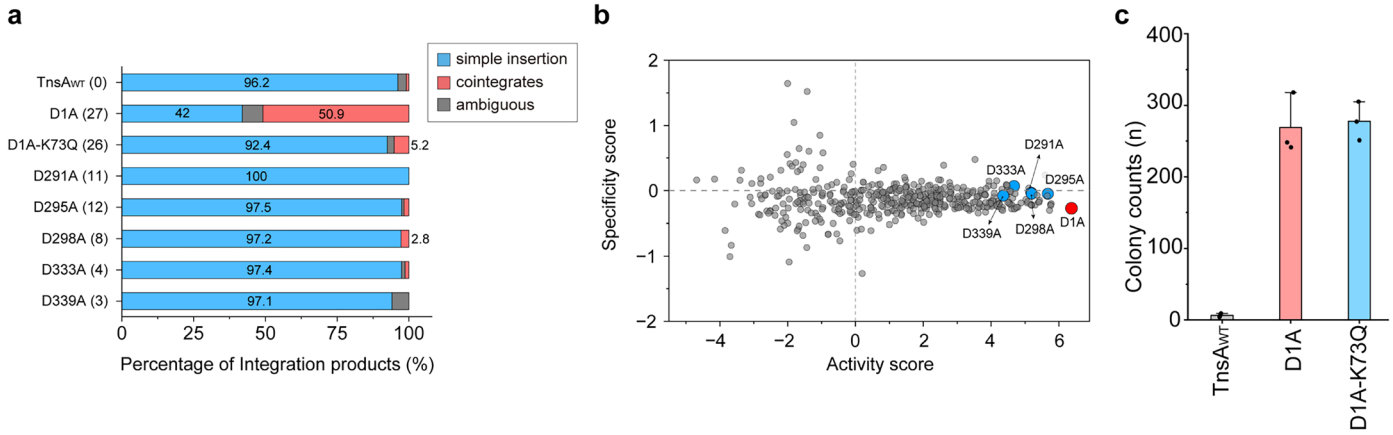
Reprints and permissions information is available at http://www.nature.com/reprints.



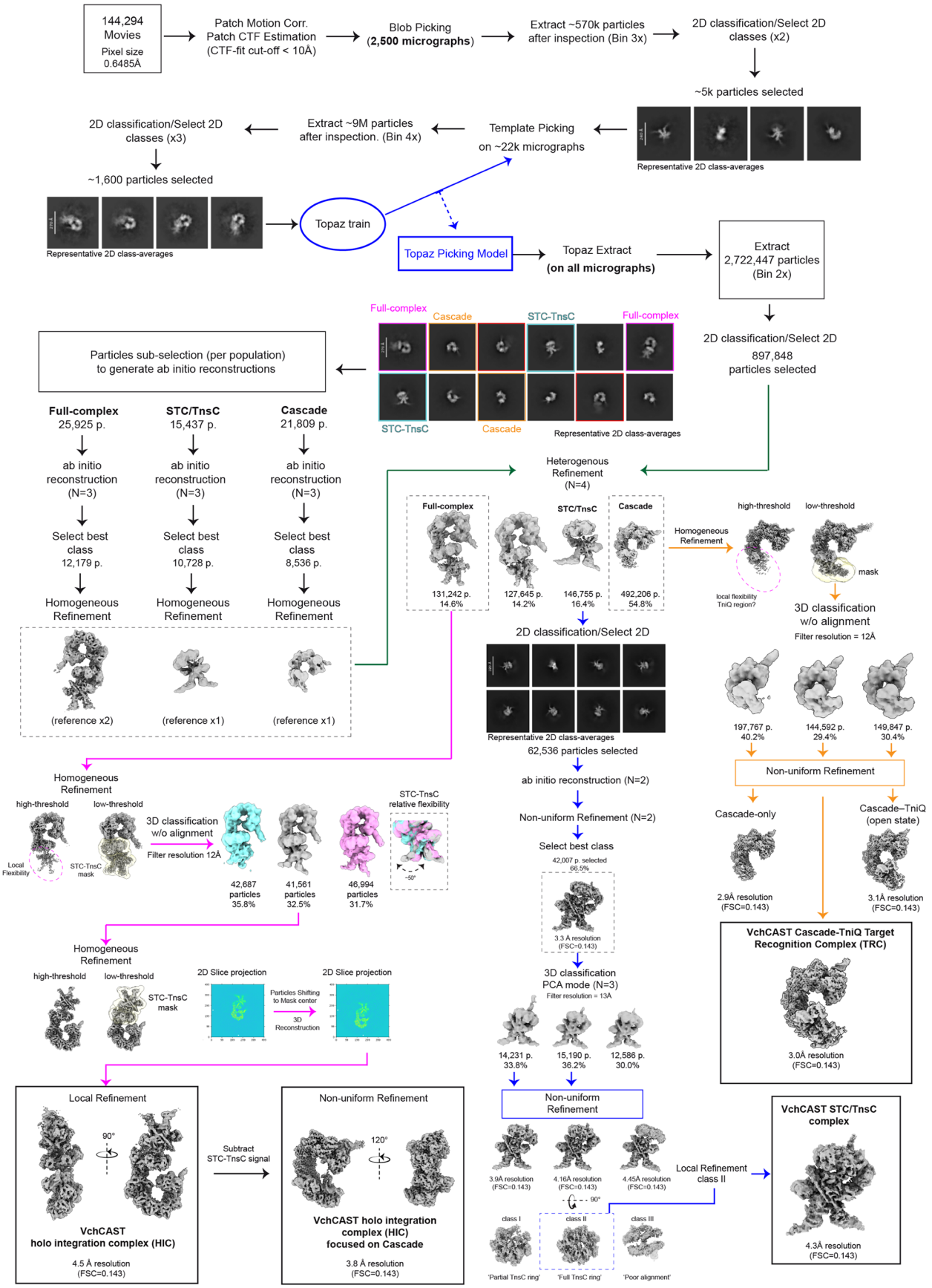
Extended Data Fig. 1 | Predicted model of the STC complex using AlphaFold 3 and validation scores. a. Proteins are shown as cartoons. Labels correspond to the different protein and nucleic acid chains. WT, wild-type; RE, right end; LE, left end; NTGS, non-target genomic strand; TGS, target genomic strand. $TnsA_{WT}$ denotes the wild-type TnsA sequence, distinguishing it from the designed TnsA construct used elsewhere in this study. **b**. Predicted STC complex colored by per-residue pLDDT score, with the scale shown at bottom. The black arrow indicates the TnsA-TnsB interface for which the interface predicted TM-score (ipTM) is reported. **c**. Predicted aligned error (PAE) plot for the STC complex model, showing the expected position error (Å) for every pair of aligned residues across all chains ($TnsA_1$, $TnsA_2$, $TnsB_1$ to $TnsB_4$, and DNA), as indicated by the color scale at bottom. Low error (dark green) along and near the diagonal within each chain block reflects high-confidence intra-chain folds, while the block structure between chains reflects the relative confidence of their predicted inter-chain positioning.



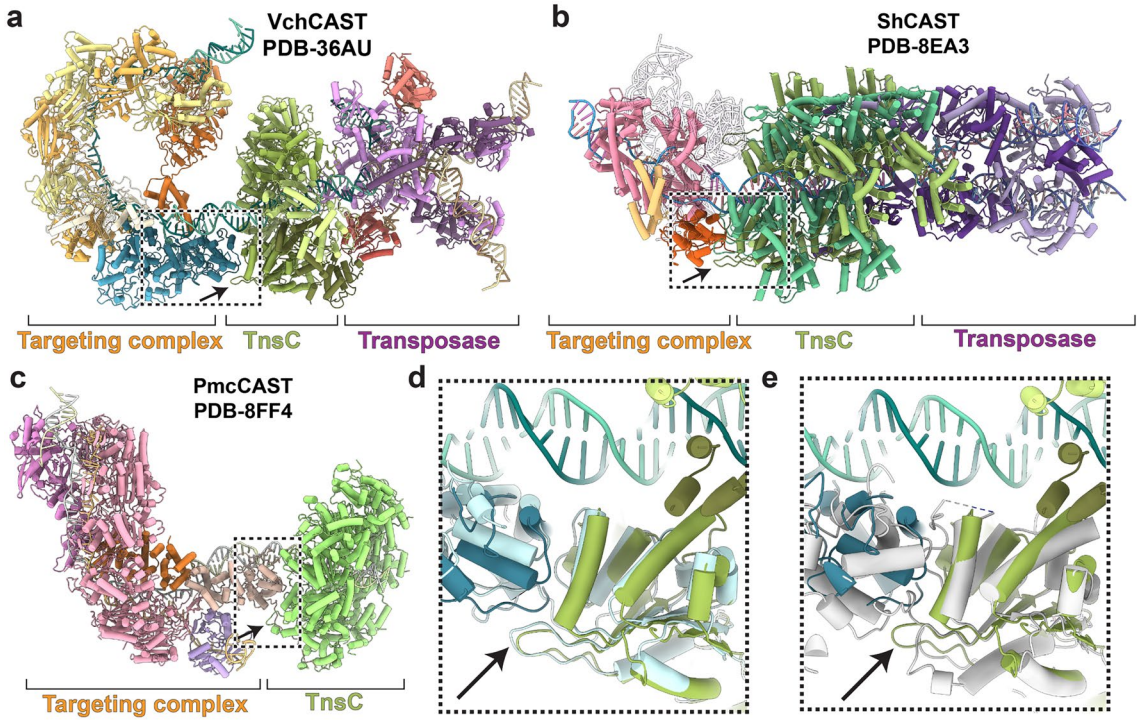
Extended Data Fig. 2 | Computational protein design introduces 27 amino acid substitutions into TnsA$_{WT}$. a. Sequence alignment of wild-type TnsA (TnsA$_{WT}$) and the top-scoring TnsA design (D1A). Conserved residues are shown in dark blue and substituted residues in light blue. **b**. Structural model of D1A highlighting the 27 substitutions introduced through protein design (red). Wild-type residue listed in one-letter amino acid code, followed by residue position, and mutant residue. Fixed active-site residues are shown in green.



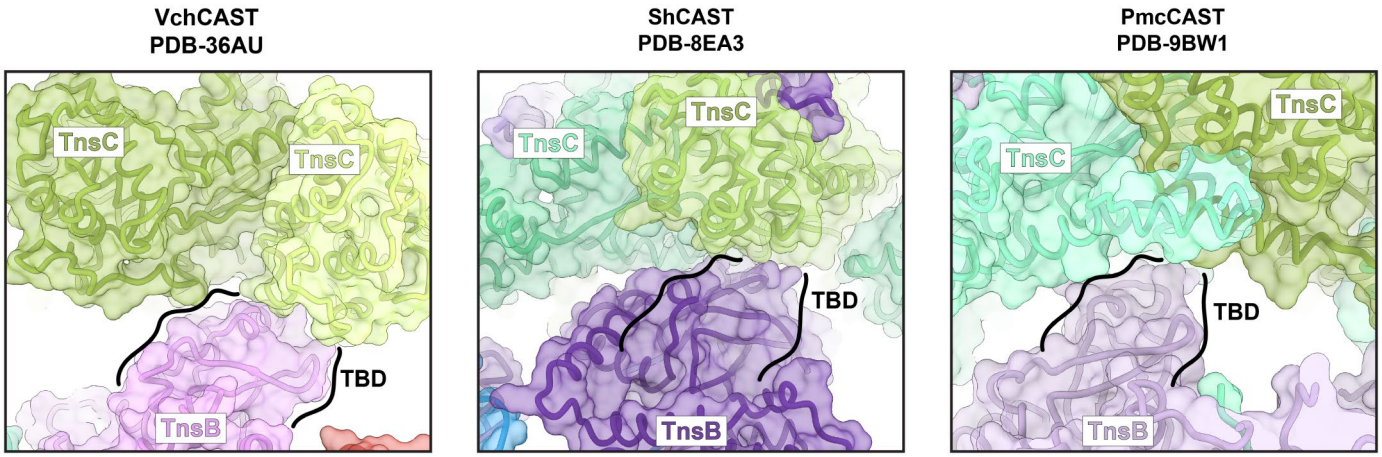
Extended Data Fig. 3 | Co-integrate analysis of TnsA designs using long-read sequencing. a. Distribution of simple insertion and co-integrate products determined by long-read (Oxford Nanopore) sequencing. Percentages are indicated within each bar. $TnsA_{WT}$ = wild-type, DXXXA refers to a particular designed variant of TnsA. The number of mutations in each design variant is shown in parentheses. D1A-K73Q corresponds to the top-scoring design, D1A, with a reversion to wild-type residue (Q73). **b.** Activity vs specificity plot for all TnsA designs, with the designs shown in panel a highlighted in blue. The top-scoring design, D1A, used for subsequent structural studies is shown in red. **c.** In vivo integration activity of VchCAST with $TnsA_{WT}$, D1A, and D1A-K73Q. Integration activity is shown as the average number of kanamycin-resistant colonies. Bars represent the mean ± s.d. of three independent biological replicates.



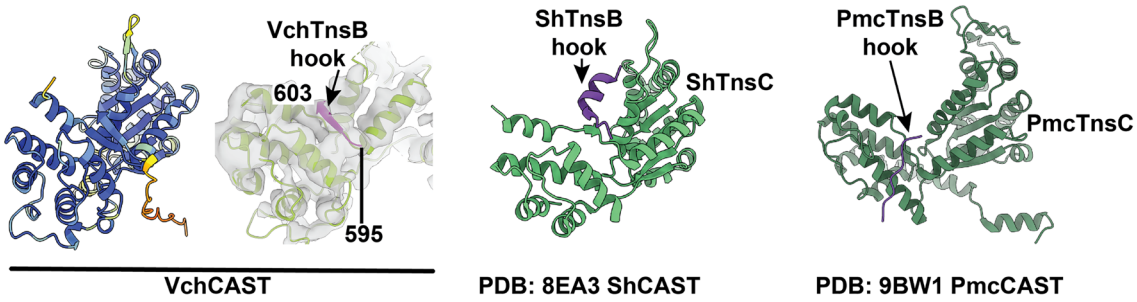
Extended Data Fig. 4 | Cryo-EM data analysis workflow for the VchCAST assemblies. Cryo-EM image processing and classification workflow leading to reconstruction of the full VchCAST holo integration complex (HIC), the Cascade-focused refined structure, the VchCAST Cascade-TniQ target recognition complex (TRC), and the VchCAST STC/TnsC complex.



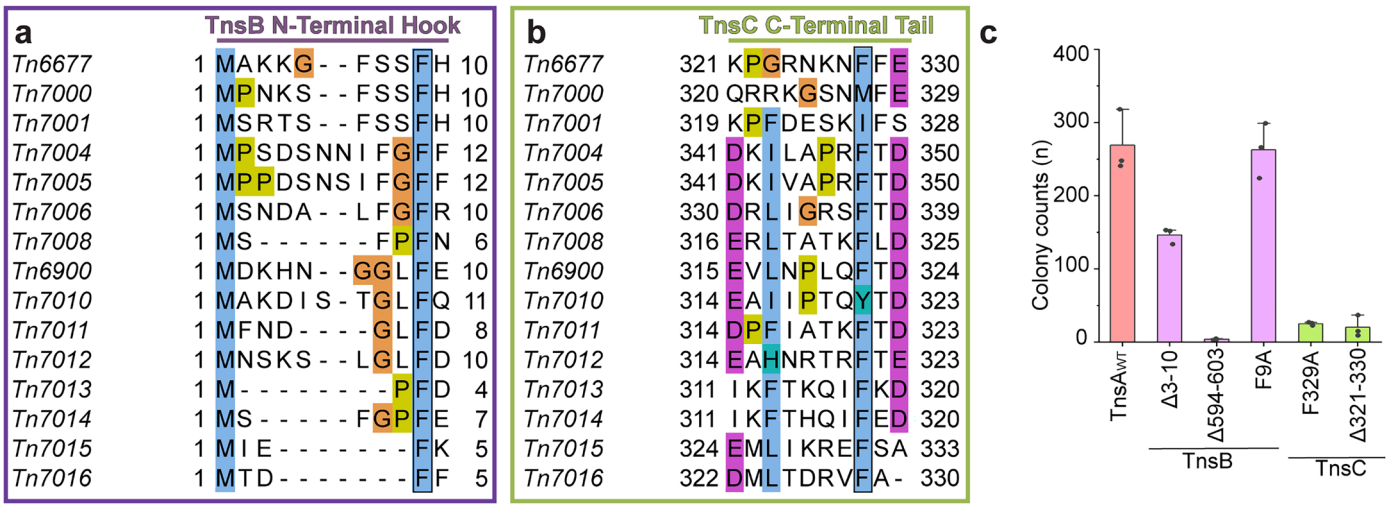
Extended Data Fig. 5 | The TniQ–TnsC interface is conserved across diverse CAST systems despite global architectural divergence. a-c. Overall architectures of three representative CAST systems: type I-F3 VchCAST (PDB:36AU), type V-K ShCAST (PDB:8EA3), and type I-B2 PmcCAST (PDB:8FF4). All depict targeting complexes (CRISPR effector + TniQ) that interact with TnsC followed by the transposase complex (TnsB in ShCAST and TnsA/B in VchCAST). The transposase component is not shown for PmcCAST because the reported PDB is a Cascade-TniQ/TnsC structure. Dotted boxes indicate the regions enlarged in panels d and e. Black arrows indicate the conserved TnsC finger loop observed in all three systems. **d-e.** Structural alignment of the TniQ-TnsC interface from VchCAST with the corresponding interfaces from ShCAST (panel **d**, RMSD = 1.2 Å) and PmcCAST (panel e, RMSD = 1.3 Å). Structural superposition was performed using TnsC1 as the reference. ShCAST TniQ and TnsC are shown in sky blue (panel d), PmcCAST TniQ and TnsC are shown in grey (panel e), and VchCAST is colored as used throughout this study. Color scheme: VchCAST colors are as in Fig. 2. ShCAST colors: Cas12k, pink; S15, tan; TniQ, orange; TnsC, green; and TnsB, purple. PmcCAST colors: Cas7 subunits, salmon; Cas5, pink; Cas11, wheat; Cas6, purple.



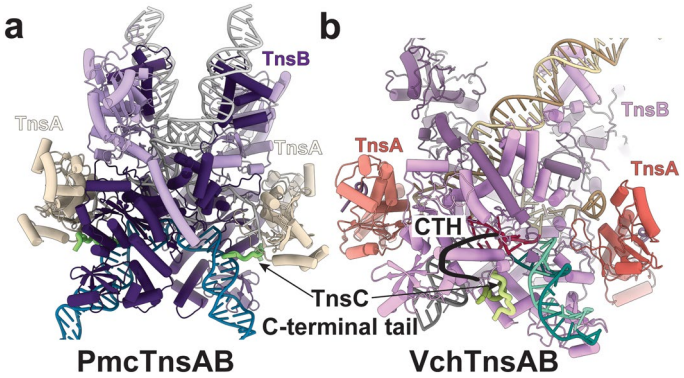
Extended Data Fig. 6 | The TnsB target-binding domain adopts a conserved positioning at the TnsC protomer interface. Representative views of the VchCAST, ShCAST, and PmcCAST holo integration complexes showing the target-binding domain (TBD) of TnsB positioned at the interface between two adjacent TnsC protomers. Proteins are shown as cartoons with transparent surface rendering enabled. TnsB and TnsC are colored in shades of purple and green, respectively. PDB ID is shown at the top of each panel. Black lines outline the TBD region.



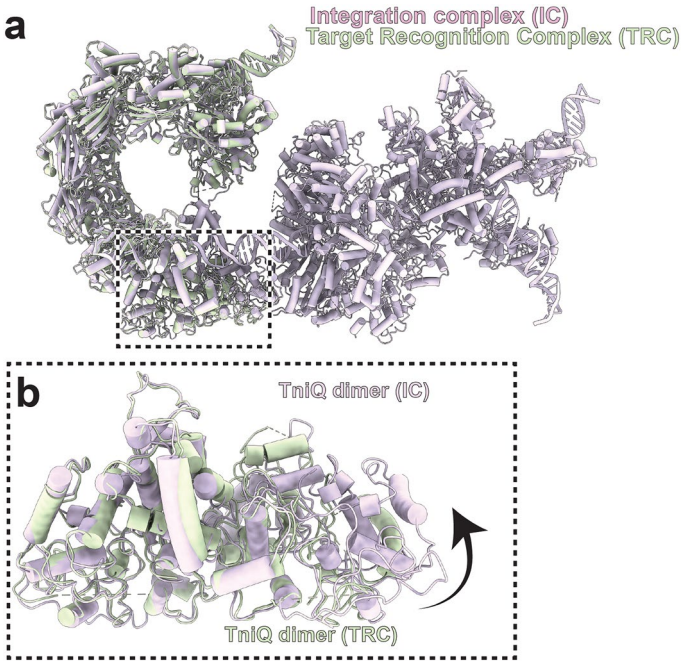
Extended Data Fig. 7 | The TnsB hook docks at the same binding site on TnsC in divergent CAST systems. Left: AlphaFold 3 prediction of a VchTnsC protomer co-folded with the C-terminal region of VchTnsB (residues PASIQKHEDEIVSADNDDWDIDYV). The model is colored according to the pLDDT confidence score using the standard AlphaFold color palette. Also shown is the VchTnsC with bound TnsB hook (residues 595–603) in the holo integration complex, together with the corresponding cryo-EM density (threshold = 0.077). VchTnsB hook is colored purple, and VchTnsC is colored green. Right: Comparison of TnsB hook-TnsC interactions observed in ShCAST and PmcCAST complexes with their respective TnsC protomers. TnsB hooks are colored purple, and TnsC protomers are colored in shades of green. Structures correspond to PDB entries 8EA3 and 9BW1, respectively.



Extended Data Fig. 8 | Sequence conservation of the TnsB N-terminal hook and TnsC C-terminal tail. a. Multiple sequence alignments of the TnsB N-terminal tail and **b**. TnsC C-terminal tail from representative I-F3 transposons. Residues are colored according to physicochemical properties and conservation. Residue positions are indicated at the beginning and end of each sequence. Residues conserved in VchCAST (Tn6677) and related systems are highlighted with a black box. The TnsB N-terminus contains a highly conserved N-terminal phenylalanine (F9), whereas the TnsC tail has a highly conserved phenylalanine (F329). **c**. In vivo integration activity of VchCAST carrying TnsB N-terminal hook and TnsC C-terminal tail variants. Integration activity was measured as the average number of kanamycin-resistant colonies obtained in the in vivo transposition assay. Bars represent the mean ± s.d. of three independent biological replicates.



Extended Data Fig. 9 | Position of TnsC C-terminal contacts within the strand-transfer complex of PmcCAST and VchCAST. a. The PmcCAST TnsAB STC structure (PDB:-9BW1) is shown in cartoon representation, in the same viewing direction as Fig. 1g. The TnsC C-terminal tail is labeled and colored green, and the C-terminal helix (CTH) domain of TnsB is indicated. **b.** The VchCAST STC/TnsC structure is shown in the same viewing direction, colored according to the scheme established in Fig. 2. Only TnsA, TnsB, DNA, and the TnsC C-terminal tail are shown for clarity.



Extended Data Fig. 10 | Structural comparison of the target recognition complex (TRC) and holo integration complex (HIC) reveals repositioning of TniQ. a. Structural superposition of the TRC onto the HIC (aligned on Cascade) shows overall agreement between the two complexes (RMSD = 0.9 Å), except for TniQ (RMSD = 4.4 Å). Structures are colored as indicated. **b**. Enlarged view of the boxed region in panel a, highlighting inward repositioning of TniQ in the HIC relative to the TRC. The black arrow indicates the direction of TniQ movement.

Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics

	VchCAST STC		VchCAST TRC (EMD-77366) (PDB-36BQ)	VchCAST HIC (EMD-77335) (PDB-36AU)	VchCAST HIC- focused on Cascade-Q (EMD-77313) (PDB-36AB)	VchCAST STC/TnsC (EMD-77370) (PDB-36BV)
	Class 1 (EMD-77364) (PDB-36BO)	Class 2 (EMD-77365) (PDB-36BP)				
Data collection and processing						
Magnification	130kx	130kx	130kx	130kx	130kx	130kx
Voltage (kV)	300	300	300	300	300	300
Electron exposure (e–/Å^2)	60	60	60	60	60	60
Defocus range (μm)	-1 to -2.5	-1.2 to -2.5	-1.2 to -2.5	-1.2 to -2.5	-1.2 to -2.5	-1.2 to -2.5
Pixel size (Å)	0.74	0.74	0.65	0.65	0.65	0.65
Symmetry imposed	C1	C1	C1	C1	C1	C1
Initial particle images (no.)	290,327	290,327	492,206	258,887	258,887	146,755
Final particle images (no.)	51,169	18,831	144,592	41,561	41,561	15,190
Map resolution (Å) FSC (0.143)	2.9 Å	3.2 Å	3.0 Å	4.5 Å	3.8 Å	4.3 Å
Map resolution range (Å)	2.6-7.5 Å	3.1-8.5 Å	2.9-5.9 Å	3.7-9.0 Å	2.8-7.8 Å	3.5-9.0 Å
Refinement						
Model resolution (Å) FSC (0.5)	3.2	4.2	3.1	6.9	4.2	7.8
Map sharpening *B* factor (Å^2)	N/A	N/A	-50	-50	-30	N/A
Model composition						
Non-hydrogen atoms	27250	24917	30842	72105	33126	40153
Protein residues	2577	2434	3523	7891	3672	4329
Nucleotides	302	243	124	396	175	238
Ligands	4	0	0	14	0	14
B factors (Å^2)						
Protein	20.8	154.9	46.0	163.6	71.8	227.5
Nucleotides	177.1	319.8	55.5	206.7	86.0	252.2
Ligand	1.7	N/A	N/A	153.9	N/A	193.5
R.m.s. deviations						
Bond lengths (Å)	0.005	0.003	0.003	0.006	0.003	0.004
Bond angles (°)	0.571	0.595	0.745	0.913	0.757	0.816
Validation						
MolProbity score	1.87	1.58	2.12	2.82	2.52	2.65
Clash score	6.22	8.67	12.17	16.99	13.3	17.43
Poor rotamers (%)	2.35	0.96	0.06	5.40	1.7	3.48
Ramachandran plot						
Favored (%)	96.3	97.5	91.0	90.7	92.2	91.6
Allowed (%)	3.4	2.4	8.4	8.6	7.5	7.5
Disallowed (%)	0.3	0.1	0.6	0.7	0.3	0.9

nature portfolio

Corresponding author(s): Elizabeth H. Kellogg

Last updated by author(s): September 14, 2026

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	Illumina NGS: MiSeq and NovaSeq X; Oxford Nanopore sequencing: Plasmidsaurus; cryo-EM data collection: EPU v3.13.0.10277.
Data analysis	BBMap v39.80; Bowtie 2 v2.5.5; Python v3.10.13; cryoSPARC v4.7.1; Topaz v0.2.5; DeepEMhancer v121322; Phenix v2.0rc1-5690; Coot v0.9.8.96; UCSF ChimeraX v1.11.1; PyMOL v3.1.7.2; OriginPro 2026. High throughput screening data performed using previously published scripts: DOI: 10.5281/zenodo.16915585

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

The cryo-EM maps and atomic coordinates reported in this study have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) under accession codes EMDB: 77364, 77365 and PDB: 36BO, 36BP (strand-transfer complex, Classes 1 and 2, respectively); EMDB: 77366 and PDB: 36BQ (target

recognition complex); EMDB: 77335 and PDB: 36AU (holo integration complex); EMDB: 77313 and PDB: 36AB (holo integration complex, focused on Cascade-TniQ); and EMDB: 77370 and PDB: 36BV (STC-TnsC). Previously published structural datasets used in this study are available from the Protein Data Bank under accession codes 9Y7B, 6PIJ, 7UFI, 9BW1, 8EA3, and 8FF4. The raw sequencing data are available from the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1472095. All other data supporting the findings of this study are provided in the manuscript and Supplementary Information.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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	No statistical methods were used to predetermine sample sizes. Sample sizes were chosen based on standard practices for the respective experimental approaches and to provide sufficient independent measurements to assess reproducibility. In vivo transposition assays were performed with three biological replicates (n = 3). High-throughput bacterial screens were performed with two biological replicates (n = 2), with >100-fold colony coverage per variant in each replicate.
Data exclusions	No experimental samples or biological replicates were excluded from the analyses. For sequencing-based analyses, read-level filtering was performed according to predefined criteria described in the Methods to retain reads appropriate for the intended downstream analysis.
Replication	Reproducibility was assessed using independent biological replicates where applicable and by repeating experimental measurements under the same conditions. In vivo transposition assays were performed in biological triplicate, and high-throughput bacterial screens were performed in biological duplicate. For structural studies, sample composition and quality were assessed by SDS-PAGE and negative-stain electron microscopy prior to cryo-EM analysis. All findings reported in the study were reproducible across the experiments performed.
Randomization	Experimental groups were defined by plasmid identity (e.g., wild-type or mutant constructs); therefore, random allocation between groups was not applicable. For each group, independent colonies were used to establish biological replicates, and experimental conditions were kept consistent across groups. Cultures were normalized to equivalent cell densities and received the donor plasmid at the same plasmid-to-cell ratio.
Blinding	For the in vivo transposition assays, colony counting was performed by investigators blinded to sample identity; the resulting colony counts were subsequently assigned to the corresponding experimental groups for analysis. For other biochemical, structural, and sequencing experiments, blinding was not performed because sample identity or composition was required during experimental setup and/or analysis. Consistent analysis procedures were applied across experimental groups.

Behavioural & social sciences study design

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

Study description	N/A
Research sample	N/A
Sampling strategy	N/A

Data collection	N/A
Timing	N/A
Data exclusions	N/A
Non-participation	N/A
Randomization	N/A

Ecological, evolutionary & environmental sciences study design

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

Study description	N/A
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing and spatial scale	N/A
Data exclusions	N/A
Reproducibility	N/A
Randomization	N/A
Blinding	N/A

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	N/A
Validation	N/A

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	N/A

Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See ICLAC register)	N/A

Palaeontology and Archaeology

Specimen provenance	N/A
Specimen deposition	N/A
Dating methods	N/A

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

Ethics oversight	N/A

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	N/A
Wild animals	N/A
Reporting on sex	N/A
Field-collected samples	N/A
Ethics oversight	N/A

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	N/A
Study protocol	N/A
Data collection	N/A
Outcomes	N/A

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

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A

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

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

(See Eklund et al. 2016)

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