Article

Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system

https://doi.org/10.1038/s41586-019-1849-0

Received: 16 July 2019

Accepted: 25 October 2019

Published online: 18 December 2019

Tyler S. Halpin-Healy[1], Sanne E. Klompe[1], Samuel H. Sternberg[1]* & Israel S. Fernández[1]*

Bacteria use adaptive immune systems encoded by CRISPR and Cas genes to maintain genomic integrity when challenged by pathogens and mobile genetic elements[1–3]. Type I CRISPR–Cas systems typically target foreign DNA for degradation via joint action of the ribonucleoprotein complex Cascade and the helicase–nuclease Cas3[4,5], but nuclease-deficient type I systems lacking Cas3 have been repurposed for RNA-guided transposition by bacterial Tn*7*-like transposons[6,7]. How CRISPR- and transposon-associated machineries collaborate during DNA targeting and insertion remains unknown. Here we describe structures of a TniQ–Cascade complex encoded by the *Vibrio cholerae* Tn*6677* transposon using cryo-electron microscopy, revealing the mechanistic basis of this functional coupling. The cryo-electron microscopy maps enabled de novo modelling and refinement of the transposition protein TniQ, which binds to the Cascade complex as a dimer in a head-to-tail configuration, at the interface formed by Cas6 and Cas7 near the 3′ end of the CRISPR RNA (crRNA). The natural Cas8–Cas5 fusion protein binds the 5′ crRNA handle and contacts the TniQ dimer via a flexible insertion domain. A target DNA-bound structure reveals critical interactions necessary for protospacer-adjacent motif recognition and R-loop formation. This work lays the foundation for a structural understanding of how DNA targeting by TniQ–Cascade leads to downstream recruitment of additional transposase proteins, and will guide protein engineering efforts to leverage this system for programmable DNA insertions in genome-engineering applications.

We previously demonstrated that a transposon derived from *V. cholerae* Tn*6677* undergoes programmable transposition in *Escherichia coli* directed by a crRNA, and that this activity requires four transposon- and three CRISPR-associated genes in addition to a CRISPR array[7] (Fig. 1a). Whereas TnsA, TnsB and TnsC exhibit functions that are consistent with their homologues from the related and well-studied cut-and-paste DNA transposon *E. coli* Tn*7*[8], we showed that TniQ, a homologue of *E. coli* TnsD, forms a co-complex with the Cascade ribonucleoprotein complex encoded by the type I-F variant CRISPR–Cas system. This finding suggested an alternative role for TniQ, compared with the role of *E. coli* TnsD in identifying target sites during Tn*7* transposition. We proposed that RNA-guided DNA targeting by Cascade could deliver TniQ to DNA in a manner compatible with downstream transpososome formation, and that TniQ might interact with Cascade near the 3′ end of the crRNA, consistent with RNA-guided DNA insertion occurring around 49 bp downstream of the protospacer-adjacent motif (PAM)-distal edge of the target site. To determine this unambiguously, we purified the *V. cholerae* TniQ–Cascade complex loaded with a native crRNA and determined its structure by cryo-electron microscopy (cryo-EM) (Supplementary Table 1).

Cryo-EM structure of TniQ–Cascade complex

The overall complex adopts a helical architecture with protuberances at both ends (Fig. 1, Extended Data Figs. 1, 2). The global architecture is similar to previously determined structures of Cascade from I-E and I-F systems[9–12] (Extended Data Fig. 3), with the exception of a large mass of additional density attributable to TniQ (see below). Maximum likelihood classification methods implemented in Relion3[13] enabled us to identify marked dynamics in the entire complex, which appears to 'breathe', widening and narrowing the distance between the two protuberances (Extended Data Fig. 1d, Supplementary Video 1). The large subunit encoded by a natural Cas8–Cas5 fusion protein (hereafter referred to simply as Cas8) forms one protuberance and recognizes the 5′ end of the crRNA via base- and backbone-specific contacts (Extended Data Figs. 4, 5a–c, 6a), similar to the canonical roles of Cas8 and Cas5 (Extended Data Fig. 3). Cas8 contains two primary subdomains formed mainly by α-helices and a third domain of approximately 100 residues (residues 277 to 385) that is predicted to form three α-helices but could not be built in our maps owing to its intrinsic flexibility (Fig. 1c). However, low-pass-filtered maps revealed that this flexible domain connects with the TniQ protuberance at the opposite end of the crescent-shaped complex (Extended Data Fig. 2e). Additionally, there seemed to be a loose coupling between the Cas8 flexible domain and overall breathing of the complex, as stronger density for that domain could be observed in the closed state (Extended Data Fig. 1d, Supplementary Video 1).

Six Cas7 subunits protect much of the crRNA by forming a helical filament along its length (Fig. 1b, d), similar to other type I Cascade complexes[9–12] (Extended Data Fig. 3). A 'finger' motif in Cas7 clamps the crRNA at regular intervals, causing every sixth nucleotide (nt) of the 32-nt spacer to flip out while leaving the flanking nucleotides

[1]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. *e-mail: shsternberg@gmail.com; isf2106@cumc.columbia.edu

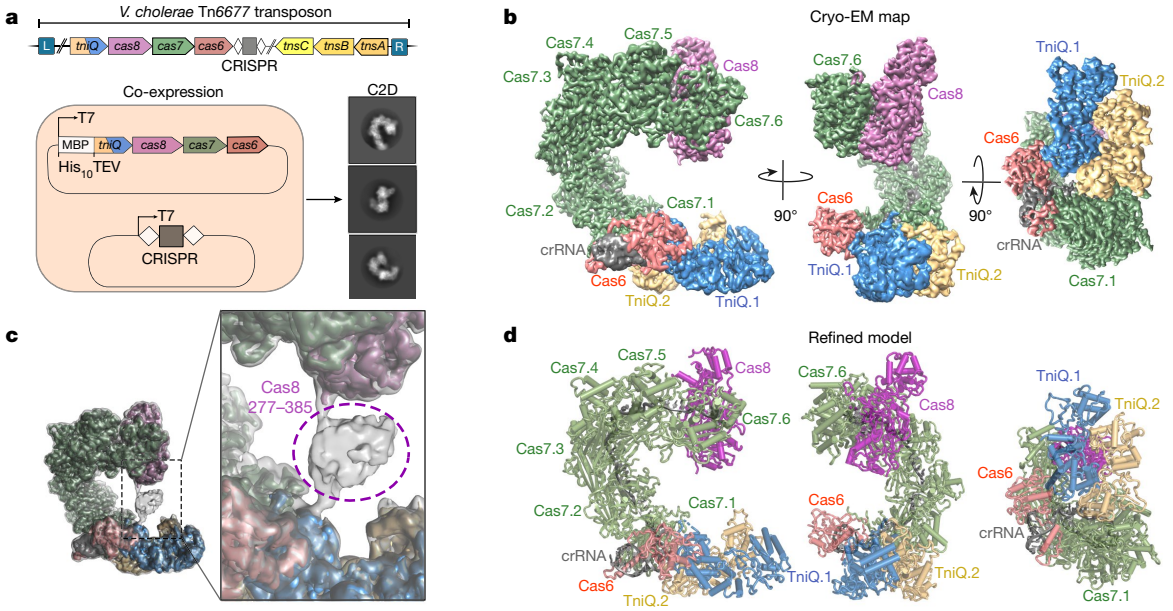


Fig. 1 | Overall architecture of the *V. cholerae* TniQ–Cascade complex. **a**, Genetic architecture of the Tn*6677* transposon (top), and plasmid constructs used to express and purify the TniQ–Cascade complex. Right, selected cryo-EM reference-free two-dimensional class averages in multiple orientations. **b**, Orthogonal views of the cryo-EM map of the TniQ–Cascade complex, showing Cas8 (purple), six Cas7 monomers (green), Cas6 (salmon), crRNA (grey) and TniQ monomers (blue, yellow). The complex adopts a helical architecture with protuberances at both ends. **c**, A flexible domain in Cas8 comprising residues 277–385 (grey) could only be visualized in low-pass-filtered maps. The unsharpened map is shown as semi-transparent, grey map overlaid on the post-processed map segmented and coloured as in **a**. **d**, Refined model for the TniQ–Cascade complex derived from the cryo-EM maps shown in **b**.

available for DNA recognition (Extended Data Figs. 4f, 6). These bases are pre-ordered in short helical segments, with a conserved phenylalanine stacking below the first base of every segment. Cas7.1, the monomer furthest away from Cas8, interacts with Cas6 (also known as Csy4), which is the RNase responsible for processing of the precursor RNA transcript derived from the CRISPR locus. The Cas6–Cas7.1 interaction is mediated by a β-sheet formed by the contribution of a β-strand from Cas6 and the two β-strands that form the finger of Cas7.1 (Extended Data Fig. 5f). Cas6 also forms extensive interactions with the conserved stem-loop in the repeat-derived 3′ crRNA handle (Fig. 1, Extended Data Fig. 5d, e), with an arginine-rich α-helix (residues 110 to 128) docked in the major groove, positioning multiple basic residues within interaction distance of the negatively charged RNA backbone.

The interaction established between Cas6 and Cas7.1 forms a continuous surface on which TniQ is docked, forming the other protuberance of the crescent. The intrinsic flexibility of the complex resulted in lower local resolutions in this area of the maps, which we overcame using local alignments masking the area comprising TniQ, Cas6, Cas7.1 and the crRNA handle (Extended Data Fig. 7). The enhanced maps enabled de novo modelling and refinement of TniQ, for which no previous structure or homology model has been reported, to our knowledge (Fig. 2). Notably, TniQ binds to Cascade as a dimer with head-to-tail configuration (Fig. 2), a surprising result given the expectation that *E. coli* TnsD functions as a monomer during Tn*7* transposition[14].

TniQ binds to Cascade as a dimer

TniQ is composed of two domains: an N-terminal domain of approximately 100 residues formed by three short α-helices and a second, larger domain of approximately 300 residues with a signature sequence for the TniQ family. A DALI search[15] using the refined TniQ model as a probe yielded marked structural similarity of the N-terminal domain to proteins containing helix–turn–helix (HTH) domains. This domain is often involved in nucleic acid recognition; however, there are examples where it has been re-purposed for protein–protein interactions[16]. The remaining C-terminal TniQ domain is formed by ten α-helices of variable length and is predicted to contain two tandem zinc finger motifs, although this region was poorly defined in the maps (Fig. 2). Overall, the double domain composition of TniQ results in an elongated structure, bent at the junction of the HTH and the TniQ domain (Fig. 2). The HTH domain of one monomer engages the TniQ domain of the other monomer via interactions between α-helix 3 (H3) and α-helix 12 (H12), respectively, in a tight protein–protein interaction (Fig. 2c). This reciprocal interaction is complemented by multiple interactions established between the TniQ domains from both monomers (up to 45 non-covalent interactions as reported by PISA[17]).

Tethering of the TniQ dimer to Cascade is accomplished by specific interactions established with both Cas6 and Cas7.1 (Fig. 3). One monomer of TniQ interacts with Cas6 via its C-terminal TniQ domain, whereas the other TniQ monomer contacts Cas7.1 through its N-terminal HTH domain (Figs. 2b, 3). The loop connecting α-helices H7 and H8 of the TniQ domain of the first TniQ monomer is inserted in a hydrophobic cavity formed at the interface of two α-helices of Cas6 (Fig. 3b, d). The TniQ histidine residue 265 is involved in rearranging the hydrophobic loop connecting H7 and H8 (Fig. 3d), which is inserted in the hydrophobic pocket of Cas6 formed by residues L20, Y74, M78, Y83 and F84. The buried surface in the Cas6–TniQ.1 interaction interface has an area of 420 Å^2. The HTH domain of the other TniQ monomer interacts with Cas7.1 through a network of interactions established mainly by α-helix H2 and the linker connecting H2 and H3, burying a surface area of 595 Å^2 (Fig. 3c, e). Thus, the HTH domain and the TniQ domain exert dual roles to drive TniQ dimerization and dock onto Cascade. The aggregate buried surface area for the TniQ–Cascade interaction is 1,015 Å^2, significantly smaller than other Cascade–effector interactions such as with the nuclease Cas3, in which 2,433 Å^2 is buried[18]. This difference is not surprising given the flexibility observed for the TniQ dimer in its association with Cascade (Supplementary Video 1).

Structure of the DNA-bound TniQ–Cascade complex

To investigate the structural determinants of DNA recognition by the TniQ–Cascade complex, we determined the structure of the complex

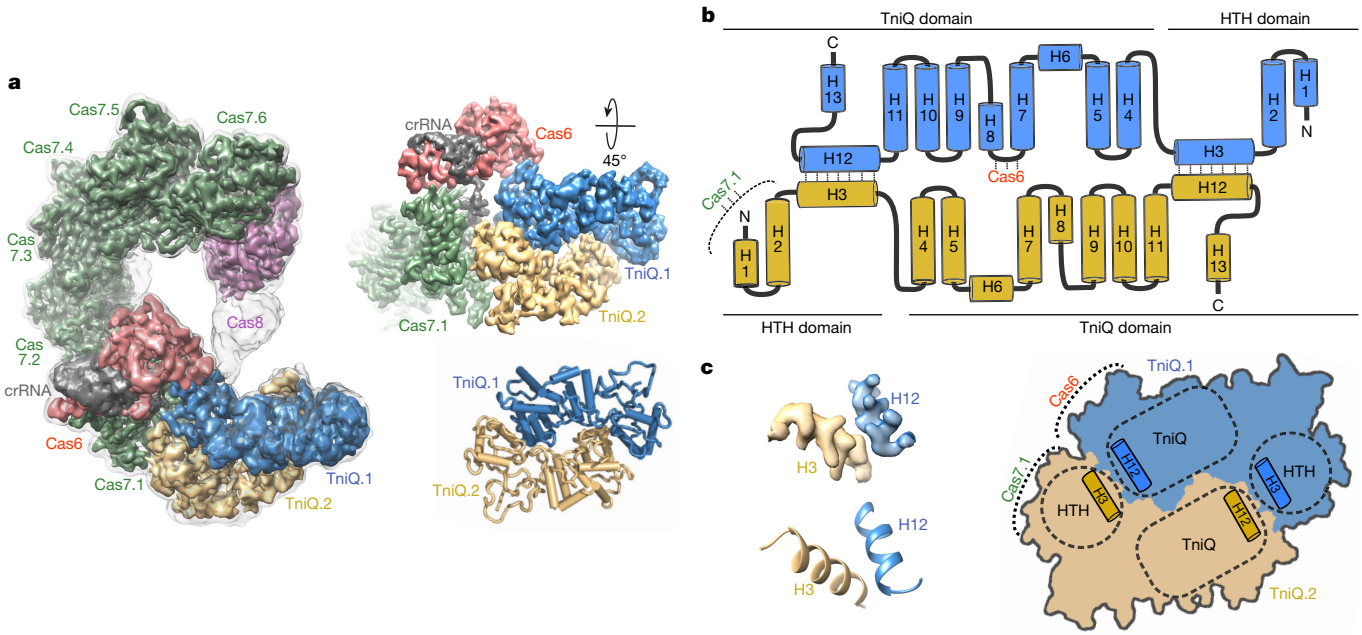


Fig. 2 | TniQ binds Cascade in a dimeric, head-to-tail configuration. a, Left, overall view of the TniQ–Cascade cryo-EM unsharpened map (grey) overlaid on the post-processed map segmented and coloured as in Fig. 1. Right, cryo-EM map (top) and refined model (bottom) of the TniQ dimer. The two monomers interact with each other in a head-to-tail configuration and are anchored to Cascade by Cas6 and Cas7.1. **b**, Secondary structure diagram of the TniQ dimer: thirteen α-helices are organized into an N-terminal HTH domain and a C-terminal TniQ domain. Dimer interactions between H3 and H12 are indicated, as are interaction sites with Cas6 and Cas7.1. **c**, Cryo-EM density for the H3–H12 interaction shows clear side-chain features (top), allowing accurate modelling of the interaction (bottom). **d**, Schematic of the dimer interaction, showing the important dimerization interface between the HTH and TniQ domain.

bound to a double-stranded DNA (dsDNA) substrate containing the 32-bp target sequence, 5′-CC-3′ PAM, and 20 bp of flanking dsDNA on both ends (Fig. 4, Extended Data Fig. 8). Density for 28 nucleotides of the target strand and 8 nucleotides of the non-target strand could be confidently assigned in the reconstructed maps (Fig. 4c). As with previous I-F Cascade structures, Cas8 recognizes the double-stranded PAM within the minor groove[11] (Extended Data Fig. 9), and an arginine residue (R246) establishes a stacking interaction with a guanine nucleotide on the target strand, which acts as a wedge to separate the double-stranded PAM from the neighbouring unwound DNA where base-pairing with the crRNA begins (Fig. 4c).

Twenty-two nucleotides of the target strand within the 32-bp target showed clear density, but surprisingly, the terminal nine nucleotides were not ordered. The target-strand base pairs with the spacer region of the crRNA in short, discontinuous helical segments, as observed previously for I-E and I-F DNA-bound Cascade complexes[11,12], with every sixth base flipped out of the heteroduplex by the insertion of a Cas7 finger (Extended Data Fig. 6b). The observed 22-bp heteroduplex is stabilized by the four Cas7 monomers proximal to the PAM (Cas7.6–Cas7.3), but even after local masked refinements, no density could be observed for any target strand nucleotides that would base-pair with the 3′ end of the crRNA spacer bound by Cas7.2 and Cas7.1. These two Cas7 monomers are proximal to Cas6 and in the region previously described to exhibit dynamics owing to the interaction of the Cas8 flexible domain with the inner face of the TniQ dimer. In addition, the disordered nucleotides also correspond to positions 25–28 of the target site where RNA–DNA mismatches

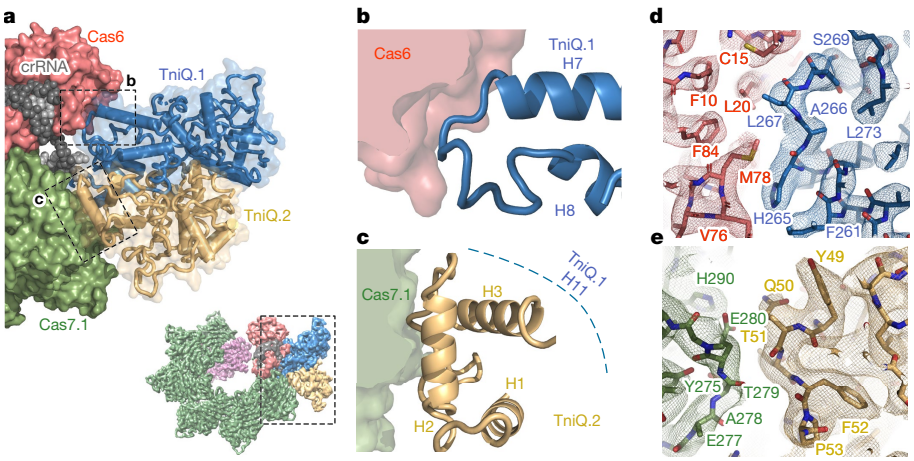


Fig. 3 | Cas6 and Cas7.1 form a binding platform for TniQ. a, Top, magnified area showing the interaction site of Cascade and the TniQ dimer. Cas6 and Cas7.1 are displayed as molecular Van der Waals surfaces, the crRNA is shown as grey spheres and the TniQ monomers are shown as ribbons. **b**, The loop connecting TniQ.1 α-helices H7 and H8 (blue) binds within a hydrophobic cavity of Cas6. **c**, Cas7.1 interacts with the HTH domain of the TniQ.2 monomer (yellow), mainly through H2 and the loop connecting H2 and H3. **d**, **e**, Experimental cryo-EM densities observed for the TniQ–Cas6 (**d**) and TniQ–Cas7.1 (**e**) interactions.

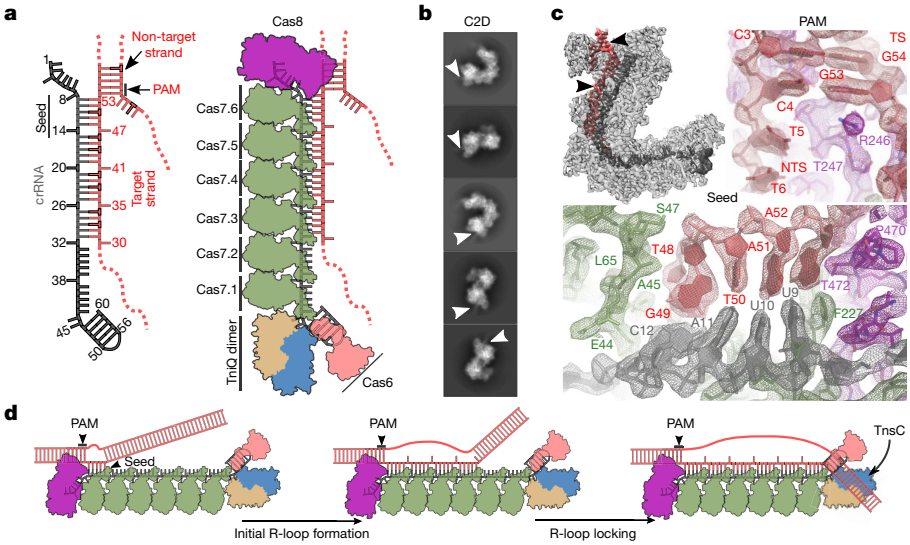


Fig. 4 | DNA-bound structure of the TniQ–Cascade complex. a, Schematic of crRNA and the portion of the dsDNA substrate that was experimentally observed within the electron density map for DNA-bound TniQ–Cascade. The target strand, non-target strand, PAM and seed regions are indicated (left); protein components are shown on the right. **b**, Selected cryo-EM reference-free two-dimensional class averages for DNA-bound TniQ–Cascade; density corresponding to dsDNA could be directly observed protruding from the Cas8 component in the two-dimensional class averages (white arrows). **c**, Cryo-EM map for DNA-bound TniQ–Cascade. The crRNA is shown in dark grey and the DNA is shown in red. Right bottom, detailed views of the PAM and seed-recognition regions of the map, with refined models represented as sticks within the electron density. Cas8 is shown in purple, Cas7 is shown in green, the crRNA is in grey and DNA is shown in red. NTS, non-target strand. **d**, The *V. cholerae* transposon encodes a TniQ–Cascade co-complex that uses the sequence content of the crRNA to bind complementary DNA target sites. We propose that the incomplete R-loop observed in our structure (middle) represents an intermediate state that may precede a downstream 'locking' step involving proofreading of the RNA–DNA complementarity. TniQ is positioned at the PAM-distal end of the DNA-bound Cascade complex, where it probably interacts with TnsC during downstream steps of RNA-guided DNA insertion.

are detrimental for RNA-guided DNA integration[7]. Thus, we propose that the partial R-loop structure that we observed could represent an intermediate conformation refractory to integration, and that further structural rearrangements may be critical for further stabilization of an open conformation, possibly driven by recruitment of the TnsC ATPase.

Here we present cryo-EM structures of a CRISPR–Cas effector complex bound to the transposition protein TniQ, with and without target DNA. These structures reveal the unexpected presence of TniQ as a dimer that forms bipartite interactions with Cas6 and Cas7.1 within the Cascade complex, forming a probable recruitment platform for downstream-acting transposition proteins[19] (Fig. 4d). Our structures further reveal a possible fidelity checkpoint, whereby formation of a complete R-loop requires conformational rearrangements that may depend on extensive RNA–DNA complementarity and/or downstream factor recruitment; this proofreading step could account for the highly specific RNA-guided DNA integration that we previously reported for the *V. cholerae* transposon[7]. In light of recent work demonstrating exaptation of type V-K CRISPR–Cas systems by similar Tn*7*-like transposons that also encode TniQ[20,21], it will be informative to determine whether tethering of TniQ to evolutionarily distinct crRNA effector complexes—Cascade or Cas12k—is a general theme of RNA-guided transposition.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1849-0.

1. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable mechanisms in microbes to resist phage infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).
2. Hille, F. et al. The biology of CRISPR–Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
3. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
4. Sinkunas, T. et al. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* **32**, 385–394 (2013).
5. Redding, S. et al. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR–Cas system. *Cell* **163**, 854–865 (2015).
6. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR–Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).
7. 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).
8. Peters, J. E. Tn7. *Microbiol. Spectr.* **2**, https://doi.org/10.1128/microbiolspec.MDNA3-0010-2014 (2014).
9. Jackson, R. N. et al. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479 (2014).
10. Chowdhury, S. et al. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47–57 (2017).
11. Guo, T. W. et al. Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR–Cas surveillance complex. *Cell* **171**, 414–426 (2017).
12. Mulepati, S., Héroux, A. & Bailey, S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**, 1479–1484 (2014).
13. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 163 (2018).
14. Holder, J. W. & Craig, N. L. Architecture of the Tn7 posttransposition complex: an elaborate nucleoprotein structure. *J. Mol. Biol.* **401**, 167–181 (2010).
15. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44** (W1), W351–W355 (2016).
16. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 (2005).
17. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43** (W1), W314–W319 (2015).
18. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* **361**, eaat0839 (2018).
19. Choi, K. Y., Spencer, J. M. & Craig, N. L. The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD. *Proc. Natl Acad. Sci. USA* **111**, E2858–E2865 (2014).
20. Faure, G. et al. CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
21. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

TniQ–Cascade purification

Protein components of TniQ–Cascade were expressed from a pET-derivative vector containing the native *V. cholerae tniQ-cas8-cas7-cas6* operon with an N-terminal His_{10}-MBP-TEV site fusion on TniQ. The crRNA was expressed separately from a pACYC-derivative vector containing a minimal repeat–spacer–repeat CRISPR array encoding a spacer from the endogenous *V. cholerae* CRISPR array. The TniQ–Cascade complex was overexpressed and purified as described previously[7], and was stored in Cascade storage buffer (20 mM Tris-Cl, pH 7.5, 200 mM NaCl, 1 mM DTT, 5% glycerol).

Sample preparation for electron microscopy

For negative staining, 3 µl of purified TniQ–Cascade ranging from 100 nM to 2 µM was incubated with plasma treated (H_2/O_2 gas mix, Gatan Solarus) CF400 carbon-coated grids (EMS) for 1 min. Excess solution was blotted and 3 µl of 0.75% uranyl formate was added for an additional minute. Excess stain was blotted away and grids were air-dried overnight. Grid screening for both negative staining and cryo conditions was performed on a Tecnai-F20 microscope (FEI) operated at 200 KeV and equipped with a Gatan K2-Summit direct detector. Microscope operation and data collection were carried out using the Leginon/Appion software. Initial negative staining grid screening allowed determination of a suitable concentration range for cryo conditions. Several grid geometries were tested in the 1–4 µM concentration range for cryo conditions using a Vitrobot Mark-II operated at 4 °C, 100% humidity, blot force 3, drain time 0, waiting time 15 s, and blotting times ranging from 3–5 s. The best ice distribution and particle density was obtained with 0.6/1 UltrAuFoil grids (Quantifoil).

Electron microscopy

A preliminary dataset of 300 images in cryo was collected with the Tecnai-F20 microscope using a pixel size of 1.22 Å/pixel with illumination conditions adjusted to 8 e^-/pixel/second with a frame window of 200 ms. Preprocessing and image processing were integrally done in Relion3[13] with ctf estimation integrated via a wrapper to Gctf[22]. An initial model computed using the SGD algorithm[23] implemented in Relion3 was used as initial reference for a refine three-dimensional job that generated a sub-nanometric reconstruction with approximately 10,000 selected particles. Clear secondary structure features in the two-dimensional averages and the three-dimensional reconstruction could be identified.

For the DNA-bound TniQ–Cascade complex containing DNA, we pre-incubated two complementary 74-nt oligonucleotides: (NTS: 5′-TTCATCAAGCCATTGGACCGCCTTACAGGACGCTTTGG CTTCATTGCTTTTCAGCTTCGCCTTGACGGCCAAAA-3′, TS: 5′-TTTTGG CCGTCAAGGCGAAGCTGAAAAGCAATGAAGCCAAAGCGTCCTGTAAGG CGGTCCAATGGCTTGATGAA-3′) for 5 min at 95 °C in hybridization buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$) to form dsDNA, which was subsequently aliquoted and flash-frozen. Complex formation was performed by incubating a 3× molar excess of dsDNA with TniQ–Cascade at 37 °C for 5 min before vitrification, which followed the conditions optimized for the apo complex (defined as TniQ–Cascade with crRNA but no DNA ligand).

High-resolution data for the apo complex were collected in a Tecnai-Polara-F30 microscope operated at 300 KeV equipped with a K3 direct detector (Gatan). A 30-µm C2 aperture was used with a pixel size of 0.95 Å/pixel and illumination conditions in microprobe mode adjusted to a fluence of 16 e^-/pixel/second. Four-second images with a frame width of 100 ms (1.77 $e^-/Å^2$/frame) were collected in counting mode.

For the DNA-bound complex, high-resolution data were collected in a Titan Krios microscope (FEI) equipped with an energy filter (20 eV slit width) and a K2 direct detector (Gatan) operated at 300 KeV. A 50-µm C2 aperture was used with a pixel size of 1.06 Å/pixel and illumination conditions adjusted in nanoprobe mode to a fluence of 8 e^-/pixel/second. Eight-second images with a frame width of 200 ms (1.42 $e^-/Å^2$/frame) were collected in counting mode.

Image processing

Motion correction was performed for every micrograph applying the algorithm described for Motioncor2[24] implemented in Relion3 with 5 by 5 patches for the K2 data and 7 by 5 patches for the K3 data. Parameters of the contrast transfer function for each motion-corrected micrograph were obtained using Gctf[22] integrated in Relion3. Initial particle picking of a subset of 200 images randomly chosen was performed with the Laplacian tool of the Auto-picking module of Relion3, using an estimated size for the complex of 200 Å. Then, 15,000 particles were extracted in a 300-pixel box size and binned 3 times for an initial two-dimensional classification job. Selected two-dimensional averages from this job were used as templates for Auto-picking of the full dataset. The full dataset of binned particles was subjected to a two-dimensional classification job to identify particles able to generate averages with clear secondary structure features. The selected subgroup of binned particles after the two-dimensional classification selection was refined against a three-dimensional volume obtained by SGD with the F20 data. This consensus volume was inspected to localize areas of heterogeneity that were clearly identified at both ends of the crescent shape characteristic of this complex. Both ends were then individually masked using soft masks of around 20 pixels that were subsequently used in classification jobs without alignments in Relion3. The T parameter used for this classification job was 6 and the total number of classes was 10. This strategy allowed us to identify two main population of particles which correspond to an open and closed state of the complex. Particles from both subgroups were separately re-extracted to obtain unbinned data sets for further refinement. New features implemented in Relion3, namely Bayesian polishing and ctf parameters refinement, allowed the extension of the resolution to 3.4, 3.5 and 2.9 Å for the two apo and the DNA-bound complexes, respectively. Post-processing was performed with a soft-mask of 5 pixels being the B-factor estimated automatically in Relion3 following standard practice. A final set of local refinements was performed with the masks used for classification. The locally aligned maps exhibit very good quality for the ends of the C-shape. These maps were used for de novo modelling and initial model refinement.

Model building and refinement

For the Cas7 and Cas6 monomers, the *E. coli* homologues (PDB accession code 4TVX) were initially docked with Chimera[25] and transformed to poly-alanine models. Substantial rearrangement of the finger region of Cas7 monomers, as well as other secondary structure elements of Cas6, were performed manually in COOT[26] before amino acid substitution of the poly-alanine model. Well-defined bulky side chains of aromatic residues allowed a confident assignment of the register. The crRNA was also well defined in the maps and was traced de novo with COOT. For Cas8 and TniQ in particular, no structural similarity was found in the published structures that was able to explain our densities. Locally refined maps using soft masks at both ends of the crescent-shaped complex rendered well-defined maps below 3.5 Å resolution. These maps were used for manual de novo tracing of a poly-alanine model in COOT that was subsequently mutated to the *V. cholerae* sequences. Bulky side chains for aromatic residues showed excellent density and were used as landmarks to adjust the register of the sequence.

For refinement, an initial step of real-space refinement against the cryo-EM maps was performed with the phenix.real_space refinement

tool of the Phenix package[27], with secondary structure restraints activated. A second step of reciprocal space refinement was performed in Refmac5[28], with secondary restraints calculated with Prosmart[29] and LibG[30]. Weight of the geometry term versus the experimental term was adjusted to avoid overfitting of the model into cryo-EM map, as previously reported[31]. Model validation was performed in Molprobity[32].

Reporting summary

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

Data availability

Maps and models have been deposited in the Electron Microscopy Data Bank with accession codes EMD-20349, EMD-20350 and EMD-20351 and the Protein Data Bank with accession codes 6PIF, 6PIG and 6PIJ.

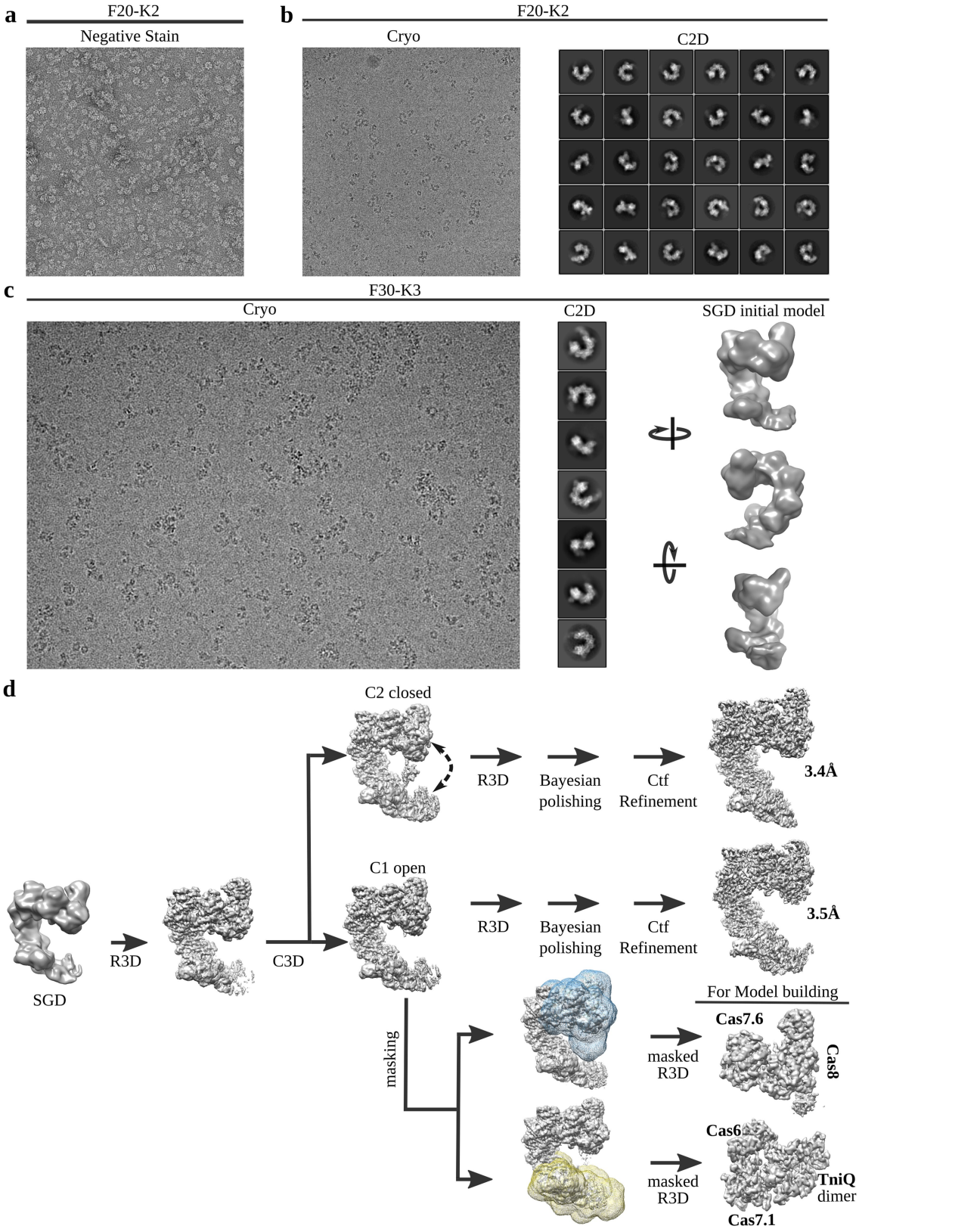
22. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
23. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
24. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
25. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
27. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
29. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
30. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
31. Fernández, I. S., Bai, X.-C., Murshudov, G., Scheres, S. H. W. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831 (2014).
32. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
33. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

Acknowledgements We thank B. Grassucci and Z. Zhang for technical assistance with cryo-EM data acquisition. Part of this work was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy located at the New York Structural Biology Center, supported by grants from the Simons Foundation (SF349247), NYSTAR and the NIH National Institute of General Medical Sciences (GM103310).

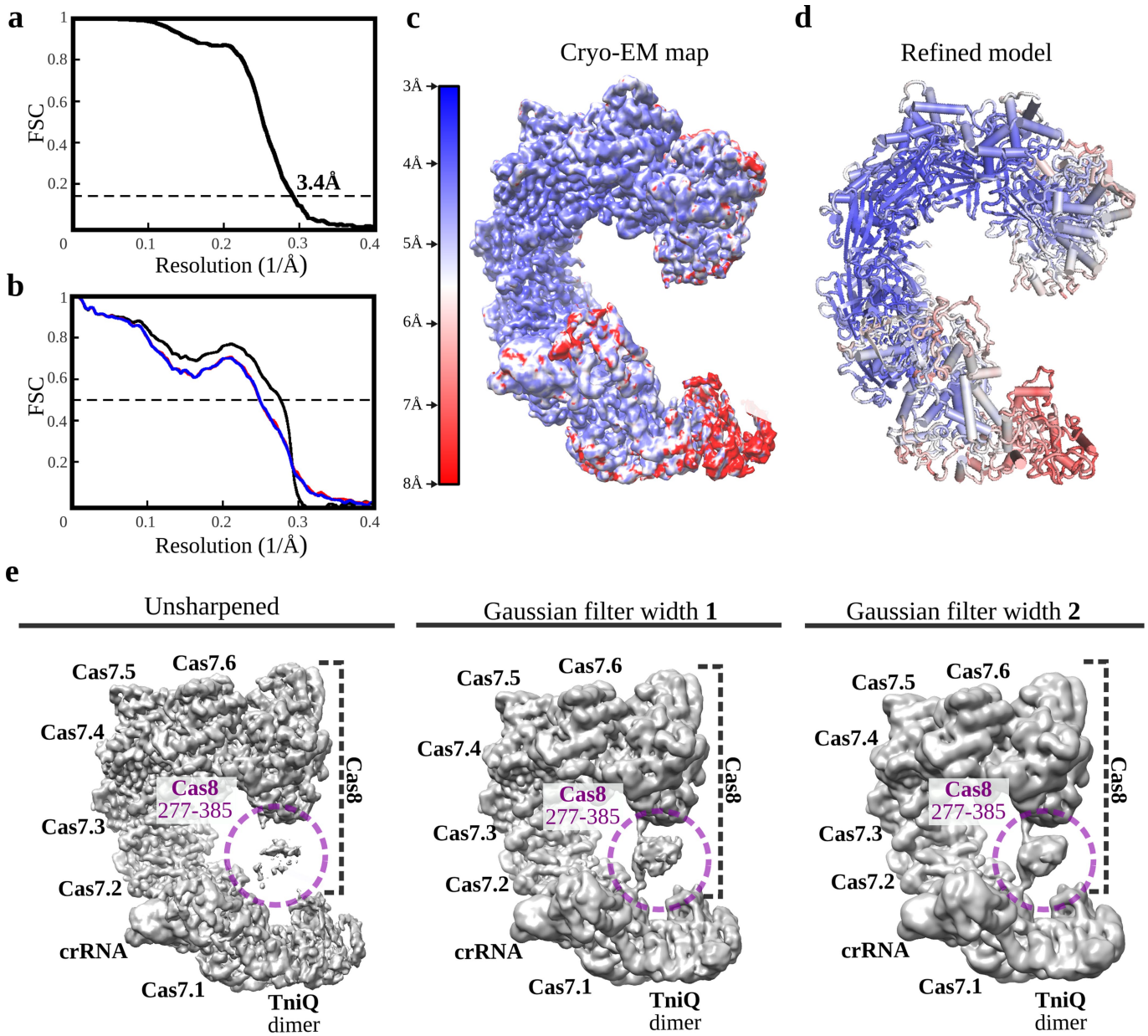
Author contributions All authors conceived and designed the project. T.S.H.-H. purified ribonucleoprotein complexes and assisted in cryo-EM data acquisition. I.S.F. collected EM data and solved the structures. I.S.F., S.H.S. and the other authors discussed the data and wrote the manuscript.

Competing interests Columbia University has filed a patent application related to this work. S.E.K. and S.H.S. are inventors on other patents and patent applications related to CRISPR–Cas systems and uses thereof. S.H.S. is a co-founder and scientific advisor to Dahlia Biosciences and an equity holder in Dahlia Biosciences and Caribou Biosciences.

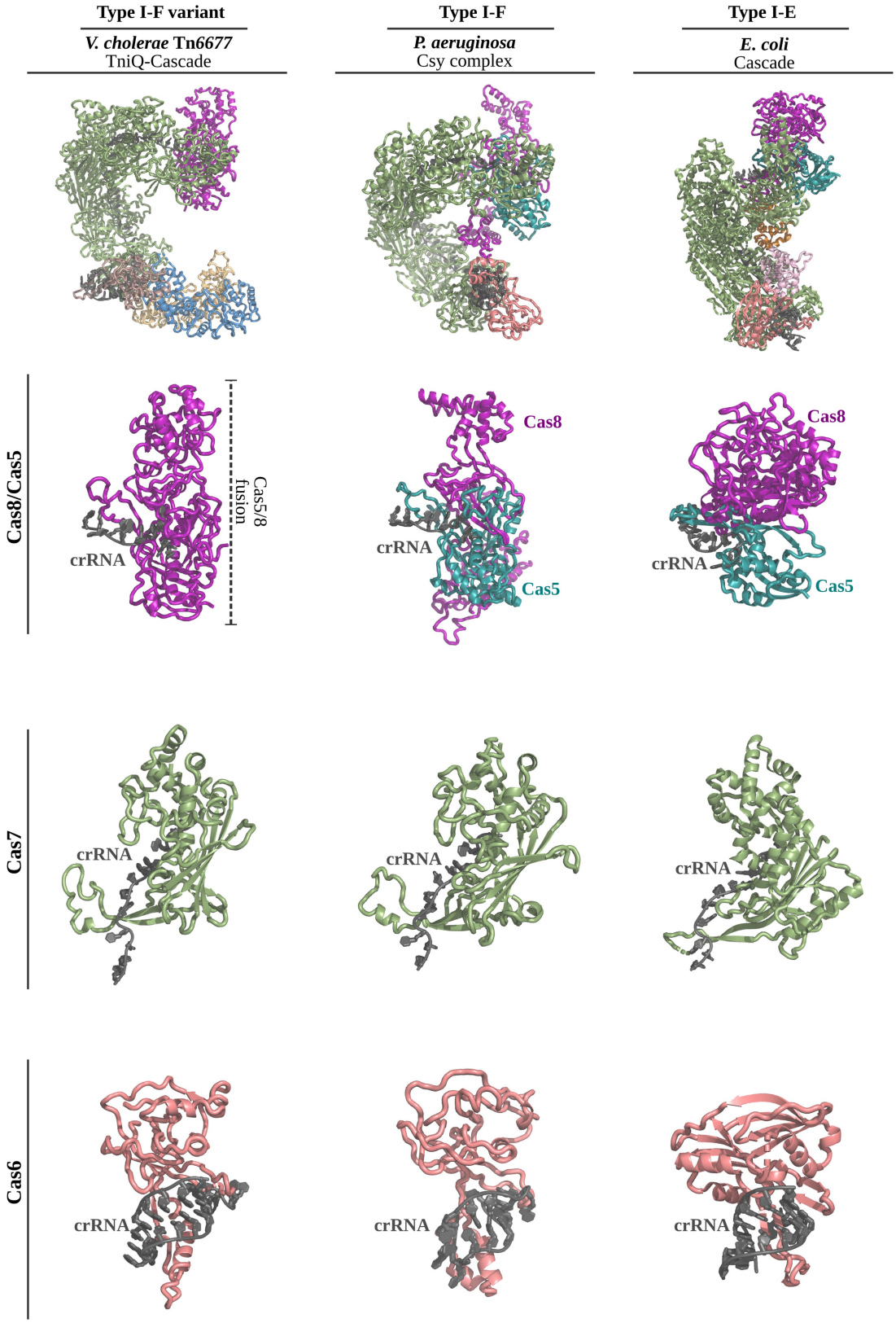
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1849-0.
Correspondence and requests for materials should be addressed to S.H.S. or I.S.F.
Peer review information *Nature* thanks Scott Bailey, Ronald Chalmers, David Taylor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



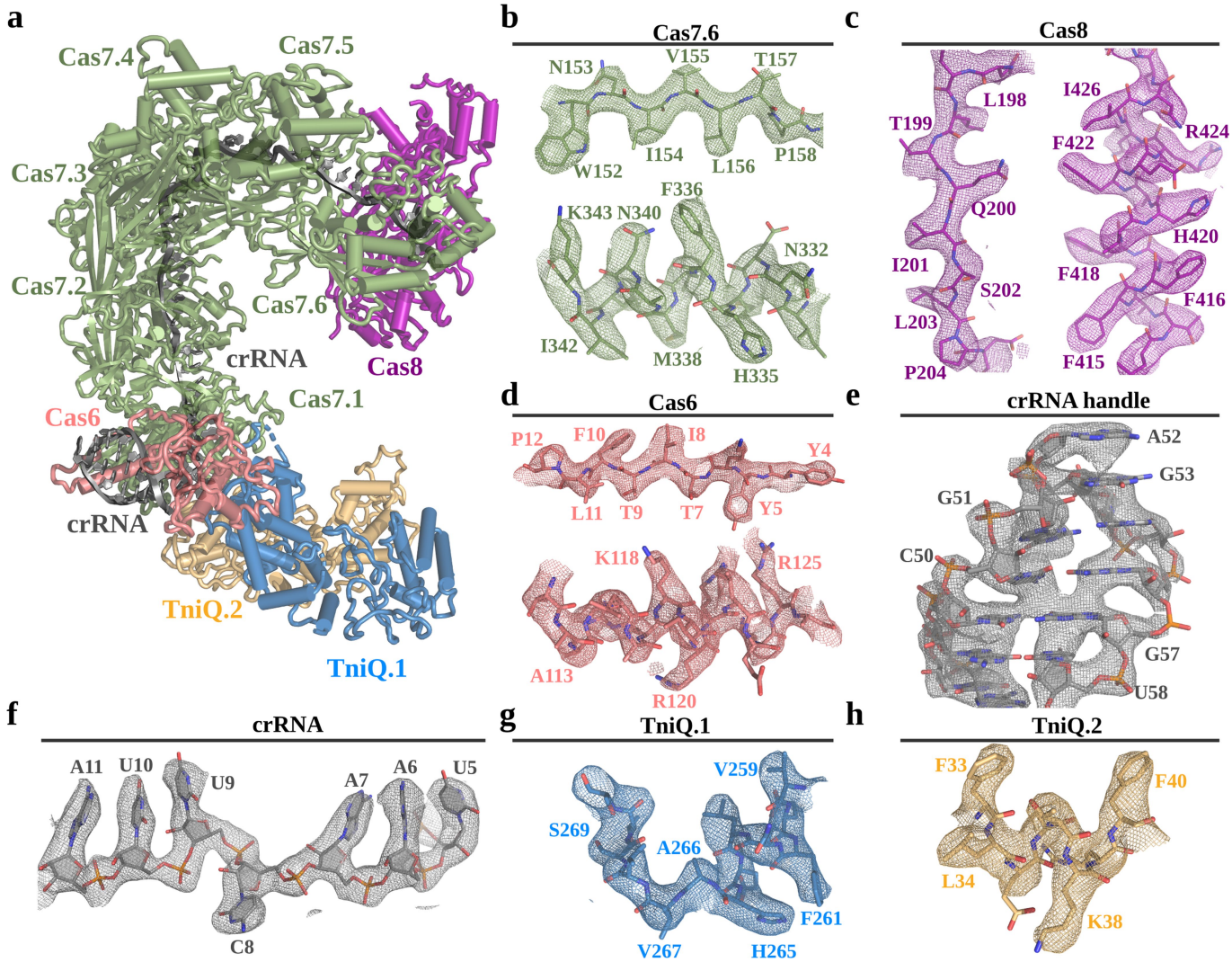
Extended Data Fig. 1 | Cryo-EM sample optimization and image processing workflow. a, Representative negatively stained micrograph for 500 nM TniQ–Cascade. **b**, Left, representative cryo-EM image for 2 µM TniQ–Cascade. A small dataset of 200 images was collected in a Tecnai-F20 microscope equipped with a Gatan K2 camera. Right, reference-free two-dimensional class averages for this initial cryo-EM dataset. **c**, Left, representative image from a large dataset collected in a Tecnai Polara microscope equipped with a Gatan K3 detector. Middle, detailed two-dimensional class averages were obtained that were used for initial model generation using the SGD algorithm[23] implemented in Relion3[13] (right). **d**, Image processing workflow used to identify the two main classes of the TniQ–cascade complex in open and closed conformations. Local refinements with soft masks were used to improve the quality of the map within the terminal protuberances of the complex. These maps were instrumental for de novo modelling and initial model refinement.



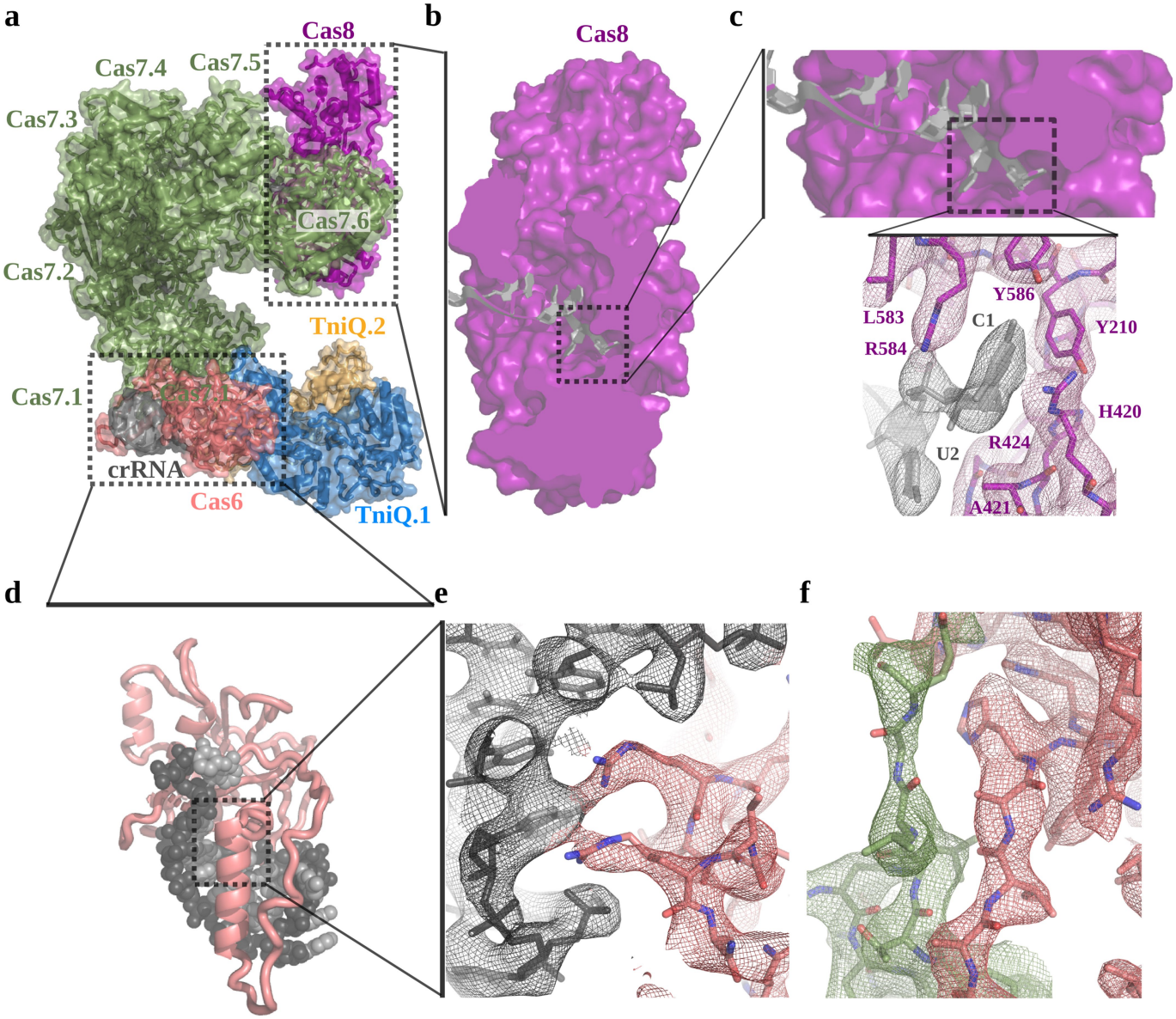
Extended Data Fig. 2 | Fourier shell correlation curves, local resolution, and unsharpened filter maps for the TniQ–Cascade complex in closed conformation. a, Gold-standard Fourier shell correlation (FSC) curve using half maps; the global resolution estimation is 3.4 Å by the FSC 0.143 criterion. **b**, Cross-validation model-versus-map FSC. Blue curve, FSC between the shacked model refined against half map 1; red curve, FSC against half map 2, not included in the refinement; black curve, FSC between final model against the final map. The overlap observed between the blue and red curves guarantees a non-overfitted model. **c**, Unsharpened map coloured according to local resolutions, as reported by RESMAP[33]. **d**, Final model coloured according to B-factors calculated by REFMAC[28]. **e**, A flexible Cas8 domain encompassing residues 277–385 contacts the TniQ dimer at the other side of the crescent shape. Applying a Gaussian filter of increasing width to the unsharpened map allows for a better visualization of this flexible region.



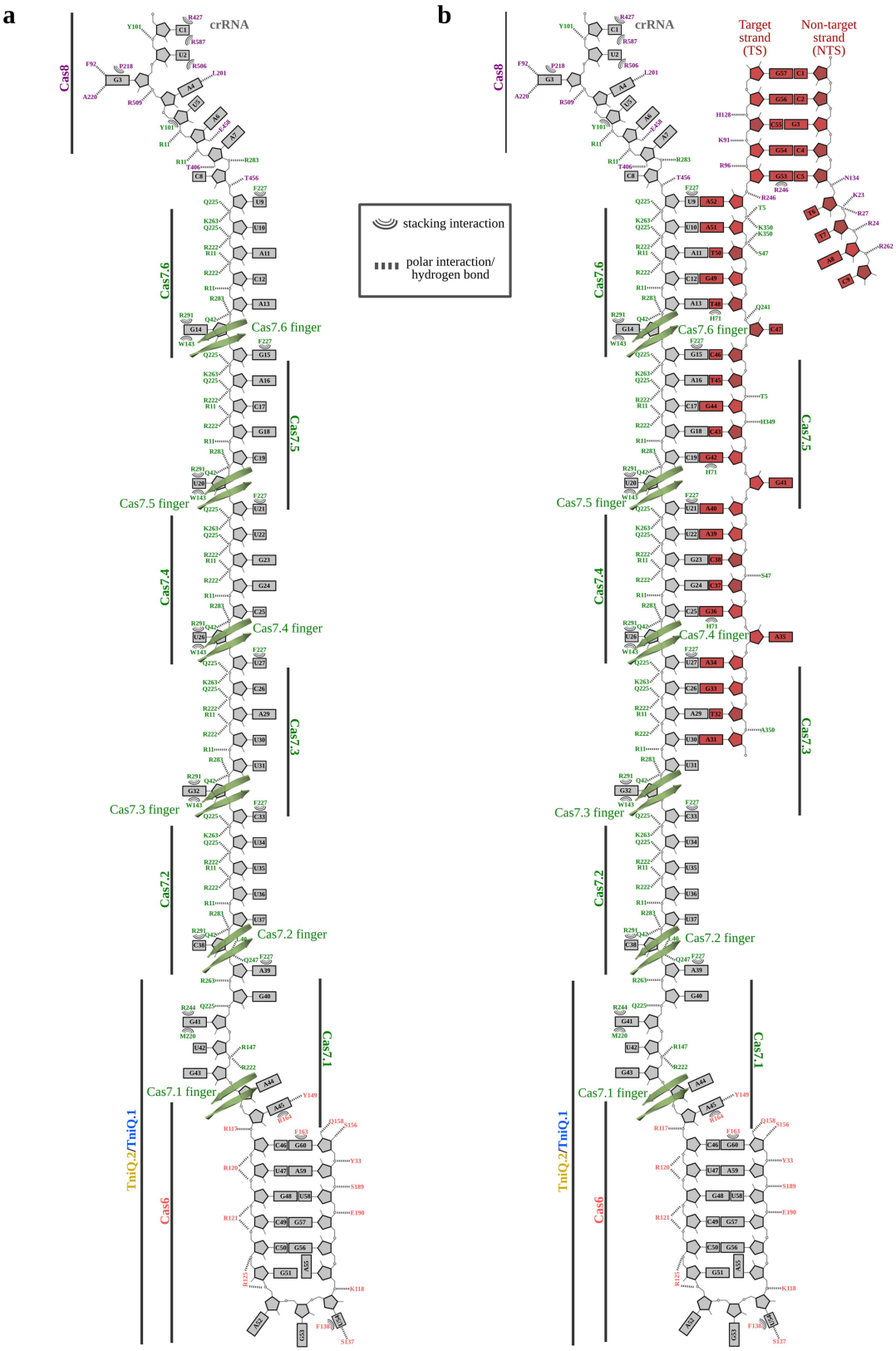
Extended Data Fig. 3 | Alignment of TniQ–Cascade with structurally similar Cascade complexes. The *V. cholerae* I-F variant TniQ–Cascade complex (left) was superposed with *Pseudomonas aeruginosa* I-F Cascade[11] (also known as Csy complex; middle, PDB ID: 6B45) and *E. coli* I-E Cascade[9] (right, PDB ID: 4TVX). Shown are alignments of the entire complex (top), the Cas8 and Cas5 subunits with the 5′ crRNA handle (second from top), the Cas7 subunit with a fragment of crRNA (second from bottom) and the Cas6 subunit with the 3′ crRNA handle (bottom).



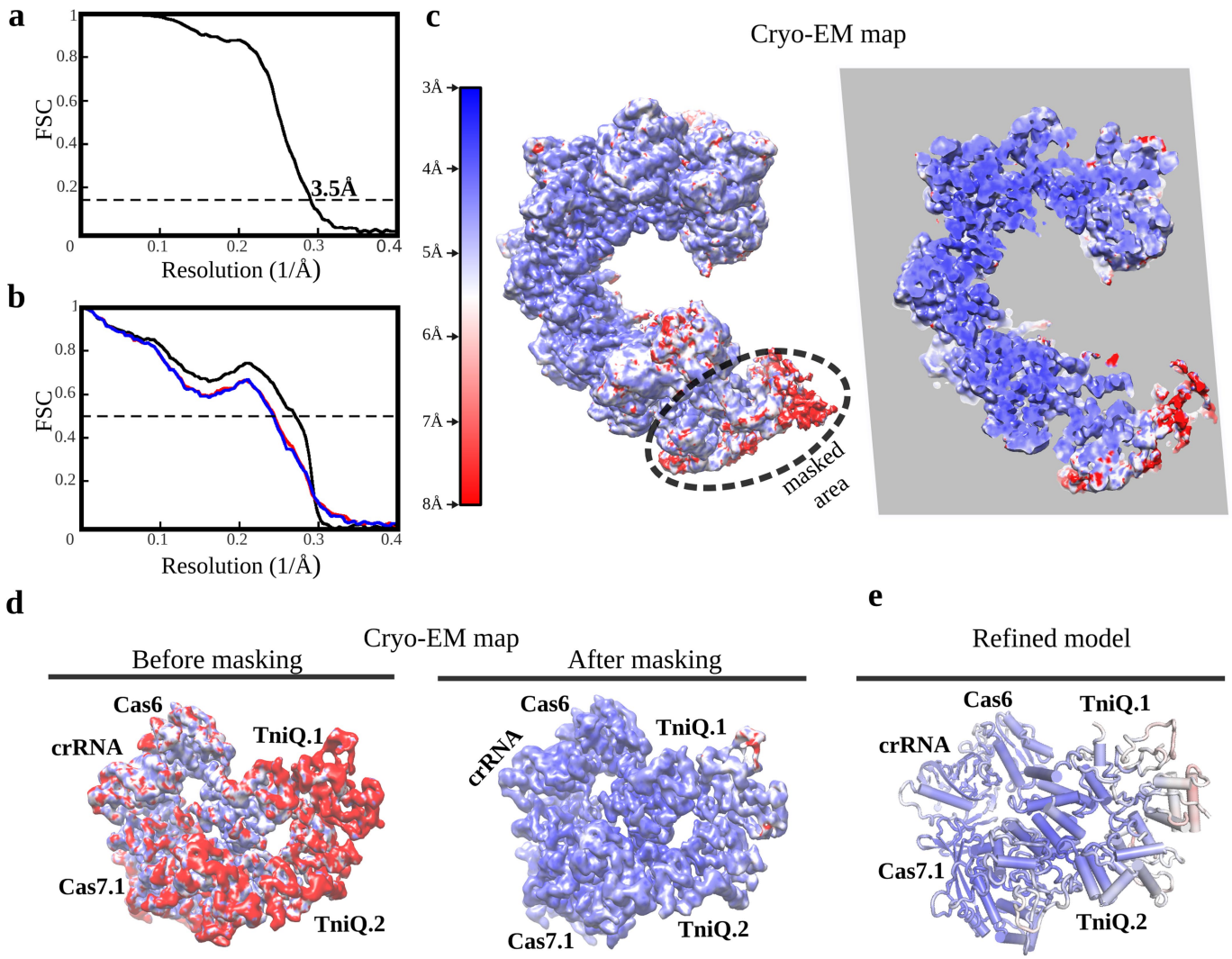
Extended Data Fig. 4 | Representative cryo-EM densities for all the components of the TniQ–Cascade complex in closed conformation. a, Final refined model of TniQ–Cascade, with Cas8 in purple, Cas7 monomers in green, Cas6 in salmon, the TniQ monomers in blue and yellow, and the crRNA in grey. **b**–**h**, Final refined model inserted in the final cryo-EM density for select regions of all the molecular components of the TniQ–Cascade complex. Residues are numbered.



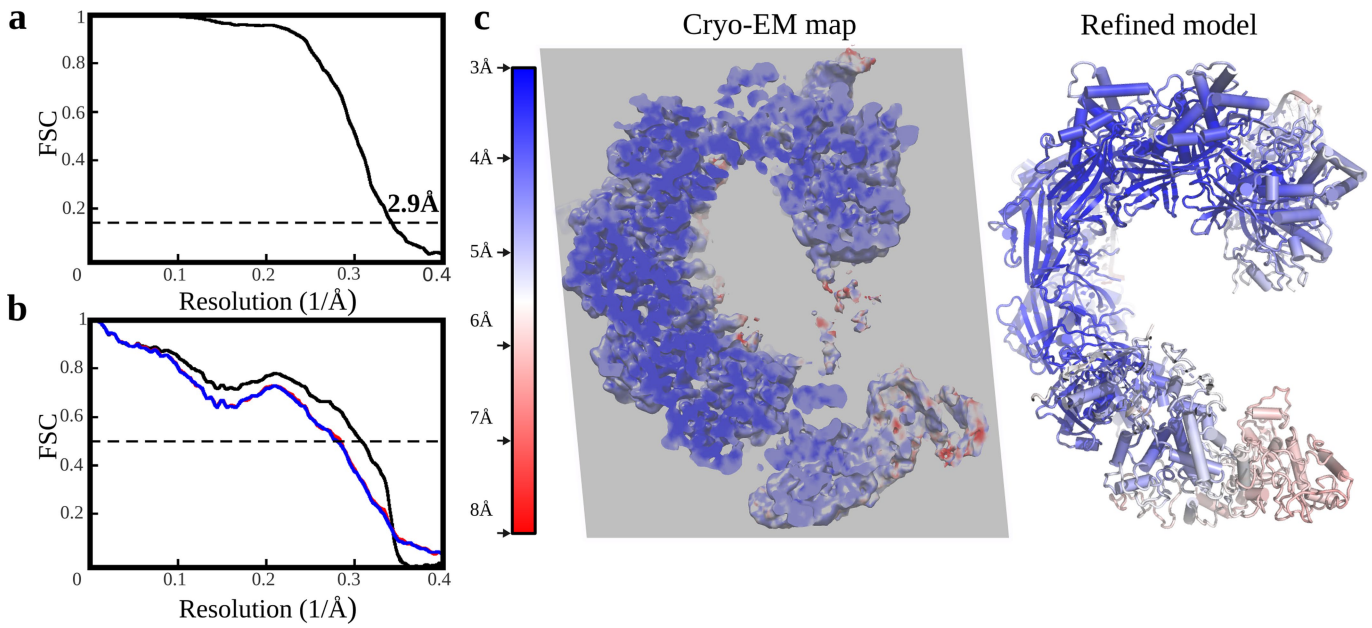
Extended Data Fig. 5 | Cas8 and Cas6 interaction with the crRNA. a, Refined model for the TniQ–Cascade shown as ribbons inserted in the semi-transparent Van der Waals surface, coloured as in Fig. 1. **b**, **c**, Magnified view of Cas8, which interacts with the 5′ end of the crRNA. The inset shows electron density for the highlighted region, where the base of nucleotide C1 is stabilized by stacking interactions with arginine residues R584 and R424. **d**, Cas6 interacts with the 3′ end of the crRNA 'handle' (nucleotides 45–60). **e**, An arginine-rich α-helix is deeply inserted within the major groove of the terminal stem–loop. This interaction is mediated by electrostatic interactions between basic residues of Cas6 and the negatively charged phosphate backbone of the crRNA. **f**, Cas6 (salmon) also interacts with Cas7.1 (green), establishing a β-sheet formed by β-strands contributed from both proteins.



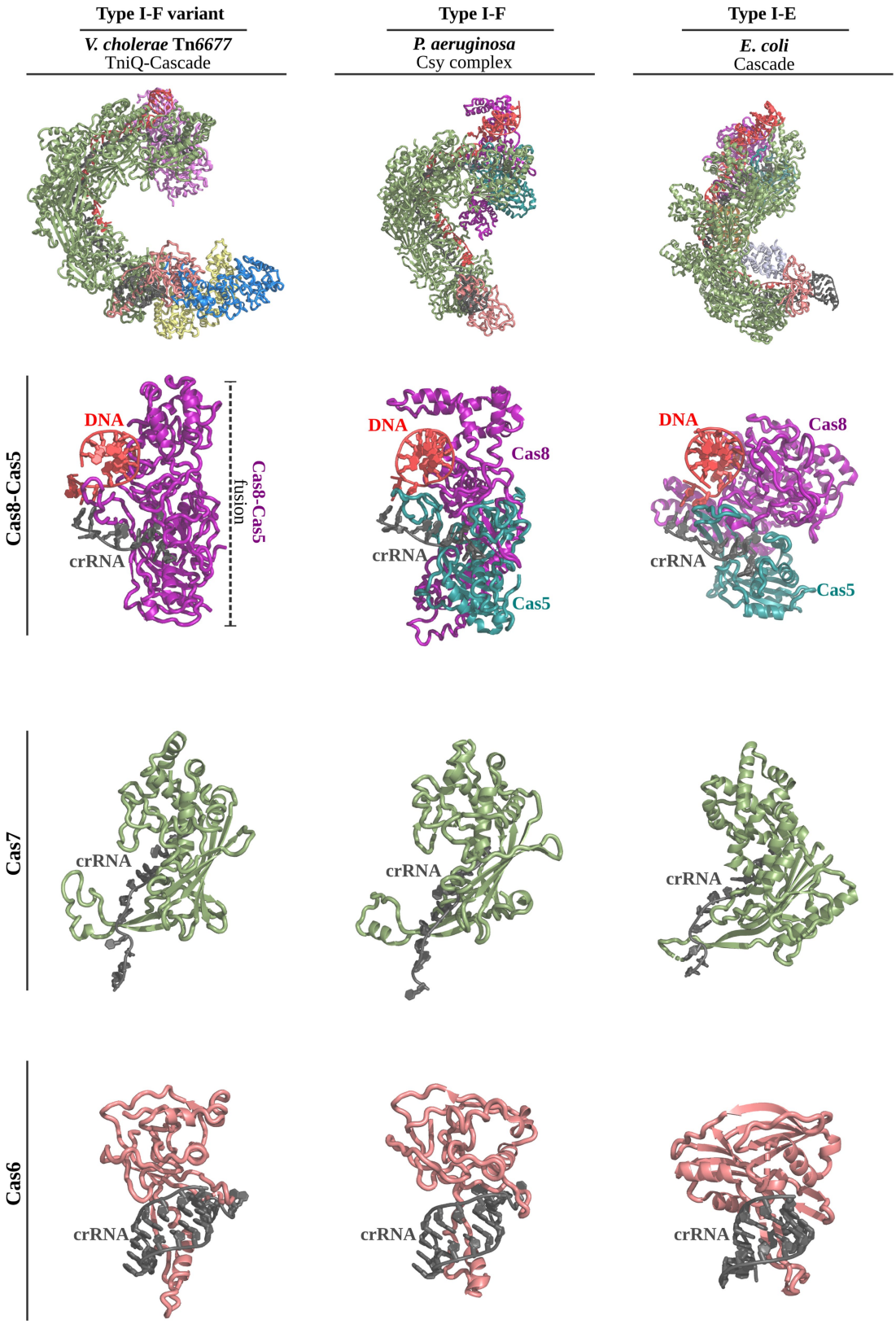
Extended Data Fig. 6 | Schematic representation of crRNA and target DNA recognition by TniQ–Cascade. a, TniQ–Cascade residues that interact with the crRNA are indicated. Approximate location for all protein components of the complex are also shown, as well as the position of each Cas7 finger. **b**, TniQ–Cascade residues that interact with crRNA and target DNA, shown as in **a**.



Extended Data Fig. 7 | FSC curves, local resolution, and local refined maps for the TniQ–Cascade complex in open conformation. a, Gold-standard FSC curve using half maps; the global resolution estimation is 3.5 Å by the FSC 0.143 criterion. **b**, Cross-validation model-versus-map FSC. Blue curve, FSC between shacked model refined against half map 1; red curve, FSC against half map 2, not included in the refinement; black curve, FSC between final model against the final map. The overlapping between the blue and red curves guarantees a non-overfitted model. **c**, Unsharpened map coloured according to local resolutions, as reported by RESMAP[33]. Right, slice through the map shown on the left. **d**, Local refinements with soft masks improved the maps in flexible regions. Shown is the region of the map corresponding to the TniQ dimer. Unsharpened maps coloured according to the local resolution estimations are shown before (left) and after (right) masked refinements. **e**, Final model for the TniQ dimer region, coloured according to the local B-factors calculated by REFMAC[28].



Extended Data Fig. 8 | FSC curves, local resolution, and unsharpened filter maps for the DNA-bound TniQ–Cascade complex complex. a, Gold-standard FSC curve using half maps; the global resolution estimation is 2.9 Å by the FSC 0.143 criterion. **b**, Cross-validation model-versus-map FSC. Blue curve, FSC between the shacked model refined against half map 1; red curve, FSC against half map 2, not included in the refinement; black curve, FSC between final model against the final map. The overlap observed between the blue and red curves guarantees a non-overfitted model. **c**, Left, unsharpened map coloured according to local resolutions, as reported by RESMAP[33]. dsDNA is visible at the top right projecting outside of the complex. Right, final model coloured according to B-factors calculated by REFMAC[28].



Extended Data Fig. 9 | Alignment of DNA-bound TniQ–Cascade with structurally similar Cascade complexes. The DNA-bound structure of *V. cholerae* I-F variant TniQ–Cascade complex (left) was superposed with DNA-bound structures of *P. aeruginosa* I-F Cascade[11] (also known as Csy complex; middle, PDB ID: 6B44) and *E. coli* I-E Cascade[9] (right, PDB ID: 5H9F). Shown are alignments of the entire complex (top), the Cas8 and Cas5 subunits with the 5′ crRNA handle and double-stranded PAM DNA (middle top), the Cas7 subunit with a fragment of crRNA (middle bottom), and the Cas6 subunit with the 3′ crRNA handle (bottom).

nature research

Corresponding author(s): Israel S. Fernandez

Last updated by author(s): Aug 13, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	For cryoEM data collection we have used the Leginon/Appion package.
Data analysis	Ctf estimation was performed with Gctf v1.06 and Image processin with Relion v3.0-beta. Figures were generated with Chimera and PyMol.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

PDBs and cryoEM maps are deposited in the PDB and EMDB public databases with accession codes: 6PIF, 6PIG and 6PIJ.

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

Life sciences study design

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

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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