

 Open access • Posted Content • DOI:10.1101/706143

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

— [Source link](#) 

Tyler S. Halpin-Healy, Sanne E. Klompe, Samuel H. Sternberg, Israel S. Fernández

Institutions: Columbia University

Published on: 18 Jul 2019 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Protospacer adjacent motif, Trans-activating crRNA, CRISPR, Protein engineering and Transposable element

Related papers:

- [Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration](#)
- [CRISPR-Cas9 Structures and Mechanisms.](#)
- [One more piece down to solve the III-A CRISPR puzzle.](#)
- [Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins.](#)
- [Casposase structure and the mechanistic link between DNA transposition and spacer acquisition by CRISPR-Cas.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/structural-basis-of-dna-targeting-by-a-transposon-encoded-5fzrtoykhh>

Abstract

Bacteria have evolved adaptive immune systems encoded by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR-associated (Cas) genes to maintain genomic integrity in the face of relentless assault from pathogens and mobile genetic elements [1–3]. Type I CRISPR-Cas systems canonically target foreign DNA for degradation via the 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 Tn7-like transposons [6, 7]. How CRISPR- and transposon-associated machineries collaborate during DNA targeting and insertion has remained elusive. Here we determined structures of a novel TniQ-Cascade complex encoded by the *Vibrio cholerae* Tn6677 transposon using single particle electron cryo-microscopy (cryo-EM), revealing the mechanistic basis of this functional coupling. The quality of the cryo-EM maps allowed for de novo modeling 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 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 (PAM) recognition and R-loop formation. The present work lays the foundation for a structural understanding of how DNA targeting by TniQ-Cascade

28 **leads to downstream recruitment of additional transposon-associated proteins, and**
29 **will guide protein engineering efforts to leverage this system for programmable**
30 **DNA insertions in genome engineering applications.**

31

32 We previously demonstrated that a transposon derived from *Vibrio cholerae* Tn6677 undergoes
33 programmable transposition in *E. coli* directed by a CRISPR RNA (crRNA), and that this activ-
34 ity requires four transposon- and three CRISPR-associated genes in addition to a CRISPR array
35 (**Fig. 1a** [7]). Whereas TnsA, TnsB, and TnsC exhibit functions that are consistent with their
36 homologs from a related and well-studied cut-and-paste DNA transposon, *E. coli* Tn7 (reviewed in
37 citePeters:2014aa), we showed that TniQ, a homolog of *E. coli* TnsD, forms a co-complex with the
38 Cascade ribonucleoprotein complex encoded by the Type I-F variant CRISPR-Cas system. This
39 finding suggested an alternative role for TniQ, as compared to the role of *Eco*TnsD in identifying
40 target sites during Tn7 transposition. Rather, we proposed that RNA-guided DNA targeting by
41 Cascade could deliver TniQ to DNA in a manner compatible with downstream transpososome for-
42 mation, and that TniQ might interact with Cascade near the 3' end of the crRNA, consistent with
43 RNA-guided DNA insertion occurring approx. 49-bp downstream from the PAM-distal edge of the
44 target site. To determine this unambiguously, we purified the *V. cholerae* TniQ-Cascade complex
45 loaded with a native crRNA and determined its structure by cryo-EM. The overall complex adopts
46 a helical architecture with protuberances at both ends (**Fig. 1** and **Extended Data Fig. 1**
47 and **2**). The global architecture is similar to previously determined structures of Cascade from I-E
48 and I-F systems (Extended Data Fig. 3) [8–11] with the exception of a large mass of additional
49 density attributable to TniQ (see below). Maximum likelihood classification methods implemented
50 in Relion3 [12] allowed us to identify significant dynamics in the entire complex, which appears to

51 "breathe", widening and narrowing the distance between the two protuberances (**Extended Data**
52 **Fig. 1d** and **Supplementary Movie 1**). The large subunit encoded by a natural Cas8-Cas5
53 fusion protein (hereafter referred to simply as Cas8) forms one protuberance and recognizes the
54 5' end of the crRNA via base- and backbone-specific contacts (**Extended Data Fig. 4, 5a-**
55 **c, 6a**), akin to the canonical roles played by Cas8 and Cas5 (**Extended Data Fig. 3**). Cas8
56 exhibits two primary subdomains formed mainly by α -helices, along with a third domain of ap-
57 proximately 100 residues (residues 277 to 385) that is predicted to form three α -helices but could
58 not be built in our maps due to its intrinsic flexibility (**Fig. 1c**). However, low-pass filtered maps
59 revealed that this flexible domain connects with the TniQ protuberance at the opposite end of the
60 crescent-shaped complex (**Extended Data Fig. 2e**). Additionally, there seemed to be a loose
61 coupling between the Cas8 flexible domain and overall "breathing" of the complex, as stronger
62 density for that domain could be observed in the closed state (**Extended Data Fig. 1d** and
63 **Supplementary Movie 1**). Six Cas7 subunits protect much of the crRNA by forming a helical
64 filament along its length (**Fig. 1b** and **d**), similar to other Type I Cascade complexes (**Extended**
65 **Data Fig. 3** [8–11]). A "finger" motif in Cas7 clamps the crRNA in regular intervals, causing
66 every sixth nucleotide (nt) of the 32-nt spacer to flip out while leaving the flanking nucleotides
67 available for DNA recognition (**Extended Data Fig. 4f**). These bases are pre-ordered in short
68 helical segments, with a conserved phenylalanine stacking below the first base of every segment.
69 Cas7.1, the monomer furthest away from Cas8, interacts with Cas6 (also known as Csy4), which
70 is the ribonuclease responsible for processing of the precursor RNA transcript derived from the

71 CRISPR locus. The Cas6-Cas7.1 interaction is mediated by a β -sheet formed by the contribution
72 of a β -strands from Cas6 and the two β -strands that form the "finger" of Cas7.1 (**Extended Data**
73 **Fig. 5f**). Cas6 also forms extensive interactions with the conserved stem-loop in the repeat-derived
74 3' crRNA handle (**Fig. 1 and Extended Data Fig. 5d and e**), with an arginine-rich α -helix
75 (residues 110 to 128) docked in the major groove, positioning multiple basic residues within interac-
76 tion distance of the negatively charged RNA backbone. The interaction established between Cas6
77 and Cas7.1 forms a continuous surface where TniQ is docked, forming the other protuberance of
78 the crescent. The intrinsic flexibility of the complex rendered lower local resolutions in this area
79 of the maps, which we overcame using local alignments masking the area comprising TniQ, Cas6,
80 Cas7.1 and the crRNA handle (**Extended Data Fig. 7**). The enhanced maps allowed for de novo
81 modeling and refinement of TniQ, for which no previous structure or homology model has been re-
82 ported (**Fig. 2**). Notably, TniQ binds to Cascade as a dimer with head-to-tail configuration (**Fig.**
83 **2**), a surprising result given the expectation that *EcoTnsD* functions as a monomer during Tn7
84 transposition [13]. TniQ is composed of two domains: an N-terminal domain of approximately 100
85 residues formed by three short α -helices and a second, larger domain of approximately 300 residues
86 with signature sequence for the TniQ family. A DALI search [14] using the refined TniQ model
87 as a probe yielded significant structural similarity of the N-terminal domain to proteins containing
88 Helix-Turn-Helix (HTH) domains (**Extended Data Fig. 8**). This domain is often involved in
89 nucleic acid recognition, however there are reported examples where it has been re-purposed for
90 protein-protein interactions [15]. The remaining C-terminal TniQ-domain is formed by 10 α -helices

91 of variable length and is predicted to contain two tandem zinc finger motifs, though this region
92 was poorly defined in the maps (**Fig. 2**). Overall, the double domain composition of TniQ results
93 in an elongated structure, bent at the junction of the HTH and the TniQ-domain (**Fig. 2**). The
94 HTH domain of one monomer engages the TniQ-domain of the other monomer via interactions be-
95 tween α -helix 3 (H3) and α -helix 11 (H11), respectively, in a tight protein-protein interaction (**Fig.**
96 **2c**). This reciprocal interaction is complemented by multiple interactions established between the
97 TniQ-domains from both monomers (up to 45 non-covalent interactions as reported by PISA [16]).
98 Tethering of the TniQ dimer to Cascade is accomplished by specific interactions established with
99 both Cas6 and Cas7.1 (**Fig. 3**). One monomer of TniQ interacts with Cas6 via its C-terminal
100 TniQ-domain, while the other TniQ monomer contacts Cas7.1 through its N-terminal HTH domain
101 (**Fig. 2b, 3**). The loop connecting alpha-helices H6 and H7 of the TniQ-domain of the first TniQ
102 monomer is inserted in a hydrophobic cavity formed at the interface of two α -helices of Cas6 (**Fig.**
103 **3b, d**). The TniQ histidine residue 265 is involved in rearranging the hydrophobic loop connecting
104 H6 and H7 (**Fig. 3d**), which is inserted in the hydrophobic pocket of Cas6 formed by residues
105 L20, Y74, M78, Y83 and F84. The HTH domain of the other TniQ monomer interacts with Cas7.1
106 through a network of interactions established mainly by α -helix H2 and the linker connecting H2
107 and H3 (**Fig. 3c, e**). Thus, both the HTH domain and the TniQ-domain exert dual roles to drive
108 TniQ dimerization and dock onto Cascade. In order to explore the structural determinants of DNA
109 recognition by the TniQ-Cascade complex, we determined the structure of the complex bound to
110 a double-stranded DNA (dsDNA) substrate containing the 32-bp target sequence, 5'-CC-3' PAM,

111 and 20-bp of flanking dsDNA on both ends (**Fig. 4** and **Extended Data Fig. 9**). Density for 28
112 nucleotides of the target strand (TS) and 8 nucleotides for the non-target strand (NTS) could be
113 confidently assigned in the reconstructed maps (**Fig.4c**). As with previous I-F Cascade structures,
114 Cas8 recognizes the double-stranded PAM within the minor groove (**Extended Data Fig. 10** [10]),
115 and an arginine residue (R246) establishes a stacking interaction with a guanine nucleotide on the
116 TS, which acts like a wedge to separate the double-stranded PAM from the neighboring unwound
117 DNA where base-pairing with the crRNA begins (**Fig. 4b**). Twenty-two nucleotides of the TS
118 within the 32-bp target showed clear density, but surprisingly, the terminal nine nucleotides were
119 not ordered. The TS base-pairs with the spacer region of the crRNA in short, discontinuous, helical
120 segments, as observed previously for I-E and I-F DNA-bound Cascade complexes [10,11] with every
121 6th base flipped out of the heteroduplex by the insertion of a Cas7 finger (**Extended Data Fig.**
122 **6b**). The observed 22-bp heteroduplex is stabilized by the four Cas7 monomers proximal to the
123 PAM (Cas7.6-7.3), but even after local masked refinements, no density could be observed for any
124 TS nucleotides that would base-pair with the 3' end of the crRNA spacer bound by Cas7.2 and
125 Cas7.1. These two Cas7 monomers are proximal to Cas6 and in the region previously described
126 to exhibit dynamics due to the interaction of the Cas8 flexible domain with the inner face of the
127 TniQ-dimer. In addition, the disordered nucleotides also correspond to positions 25-28 of the target
128 site where RNA-DNA mismatches are detrimental for RNA-guided DNA integration [7]. Thus, we
129 propose the possibility that the partial R-loop structure we observed may represent an intermediate
130 conformation refractory to integration, and that further structural rearrangements may be critical

131 for further stabilization of an open conformation, possibly driven by recruitment of the TnsC AT-
132 Pase. Here we present the first cryo-EM structures of a CRISPR-Cas effector complex bound to the
133 transposition protein TniQ, with and without target DNA. These structures reveal the unexpected
134 presence of TniQ as a dimer that forms bipartite interactions with Cas6 and Cas7.1 within the Cas-
135 cade complex, forming a likely recruitment platform for downstream-acting transposition proteins18
136 (**Fig. 4d**). Our structures furthermore reveal a possible fidelity checkpoint, whereby formation of a
137 complete R-loop requires conformational rearrangements that may depend on extensive RNA-DNA
138 complementarity and/or downstream factor recruitment; this proofreading step could account for
139 the highly specific RNA-guided DNA integration we previously reported for the *V. cholerae* trans-
140 poson [7]. In light of recent work demonstrating exaptation of Type V-K CRISPR-Cas systems
141 by similar Tn7-like transposons that also encode TniQ [17, 18], it will be interesting to determine
142 whether tethering of TniQ to evolutionarily distinct CRISPR RNA effector complexes - Cascade or
143 Cas12k - is a general theme of RNA-guided transposition.

144 **Methods**

145 **TniQ-Cascade purification.**

146 Protein components of TniQ-Cascade were expressed from a pET-derivative vector containing the
147 native *V. cholerae* tniQ-cas8-cas7-cas6 operon with an N-terminal His10-MBP-TEVsite fusion on
148 TniQ. The crRNA was expressed separately from a pACYC-derivative vector containing a minimal

149 repeat-spacer-repeat CRISPR array encoding a spacer from the endogenous *V. cholerae* CRISPR
150 array. The TniQ-Cascade complex was overexpressed and purified as described previously [7], and
151 was stored in Cascade Storage Buffer (20 mM Tris-Cl, pH 7.5, 200 mM NaCl, 1 mM DTT, 5%
152 glycerol).

153 **Sample preparation for electron microscopy.**

154 For negative staining, 3 μ L of purified TniQ-Cascade ranging from 100 nM to 2 μ M was incubated
155 with plasma treated (H₂/O₂ gas mix, Gatan Solarus) CF400 carbon-coated grids (EMS) for 1
156 minute. Excess solution was blotted and 3 μ L of 0.75% uranyl formate was added for an additional
157 minute. Excess stain was blotted away and grids were air-dried overnight. Grid screening for both
158 negative staining and cryo conditions was performed on a Tecnai-F20 microscope (FEI) operated
159 at 200 KeV and equipped with a Gatan K2-Summit direct detector. Microscope operation and
160 data collection were carried out using the Leginon/Appion software. Initial negative staining grid
161 screening allowed determination of a suitable concentration range for cryo conditions. Several grid
162 geometries were tested in the 1-4 μ M concentration range for cryo conditions using a Vitrobot Mark-
163 II operated at 4 C, 100% humidity, blot force 3, drain time 0, waiting time 15 seconds, and blotting
164 times ranging from 3-5 seconds. The best ice distribution and particle density was obtained with
165 0.6/1 UltrAuFoil grids (Quantifoil).

166 **Electron microscopy.**

167 A preliminary dataset of 300 images in cryo was collected with the Tecnai-F20 microscope using a
168 pixel size of 1.22 Å/pixel with illumination conditions adjusted to 8 e-/pixel/second with a frame
169 window of 200 ms. Preprocessing and image processing were integrally done in Relion3 [12] with
170 ctf estimation integrated via a wrapper to Gctf [19]. An initial model computed using the SGD
171 algorithm [20] implemented in Relion3 was used as initial reference for a refine 3D job that gener-
172 ated a sub-nanometric reconstruction with approximately 10,000 selected particles. Clear secondary
173 structure features in the 2D averages and the 3D reconstruction could be identified. For the DNA-
174 bound TniQ-Cascade complex containing DNA, we pre-incubated two complementary 74-nt oligonu-
175 cleotides (NTS: 5' TTCATCAAGCCATTGGACCGCCTTACAGGACGCTTTGGCTTCATTGCTTTTCAC
176 3' , TS: 5' TTTTGGCCGTCAAGGCGAAGCTGAAAAGCAATGAAGCCAAAGCGTCCTGTAAGGCGGT
177 3') for 5 minutes at 95° C in hybridization buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 5 mM
178 MgCl₂) to form dsDNA, which was subsequently aliquoted and flash frozen. Complex formation
179 was performed by incubating a 3x molar excess of dsDNA with TniQ-Cascade at 37° C for 5 min-
180 utes prior to vitrification, which followed the conditions optimized for the apo complex (defined
181 as TniQ-Cascade with crRNA but no DNA ligand). High resolution data for the apo complex
182 were collected in a Tecnai-Polaris-F30 microscope operated at 300 KeV equipped with a K3 direct
183 detector (Gatan). A 30 μm C2 aperture was used with a pixel size of 0.95 Å/pixel and illumination
184 conditions in microprobe mode adjusted to a fluence of 16e-/pixel/second. Four-second images with
185 a frame width of 100 ms (1.77 e-/2/frame) were collected in counting mode. For the DNA-bound

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

191 **Image processing.**

192 Motion correction was performed for every micrograph applying the algorithm described for Mo-
193 tioncor2 [21] implemented in Relion3 with 5 by 5 patches for the K2 data and 7 by 5 patches for
194 the K3 data. Parameters of the contrast transfer function for each motion-corrected micrograph
195 were obtained using Gctf integrated in Relion3. Initial particle picking of a subset of 200 images
196 randomly chosen was performed with the Laplacian tool of the Auto-picking module of Relion3,
197 using an estimated size for the complex of 200 \AA 15,000 particles were extracted in a 300 pixels
198 box size and binned 3 times for an initial 2D classification job. Selected 2D averages from this job
199 were used as templates for Auto-picking of the full dataset. The full dataset of binned particles
200 was subjected to a 2D classification job to identify particles able to generate averages with clear
201 secondary structure features. The selected subgroup of binned particles after the 2D classification
202 selection was refined against a 3D volume obtained by SGD with the F20 data. This "consensus"
203 volume was inspected to localize areas of heterogeneity which were clearly identified at both ends
204 of the crescent shape characteristic of this complex. Both ends were then individually masked using

205 soft masks of around 20 pixels that were subsequently used in classification jobs without alignments
206 in Relion3. The T parameter used for this classification job was 6 and the total number of classes
207 was 10. This strategy allowed us to identify two main population of particles which correspond to
208 an "open" and "closed" state of the complex. Particles from both subgroups were separately re-
209 extracted to obtain unbinned datasets for further refinement. New features implemented in Relion3,
210 namely Bayesian polishing and ctf parameters refinement, allowed the extension of the resolution
211 to 3.4, 3.5 and 2.9 Å for the two apo and the DNA-bound complexes, respectively. Post processing
212 was performed with a soft-mask of 5 pixels being the B-factor estimated automatically in Relion3
213 following standard practice. A final set of local refinements was performed with the masks used
214 for classification. The locally aligned maps exhibit very good quality for the ends of the C-shape.
215 These maps were used for de novo modeling and initial model refinement.

216 **Model building and refinement.**

217 For the Cas7 and Cas6 monomers, the *E. coli* homologs (PDB accession code 4TVX) were initially
218 docked with Chimera [22] and transformed to poly-alanine models. Substantial rearrangement
219 of the finger region of Cas7 monomers, as well as other secondary structure elements of Cas6,
220 were performed manually in COOT [23] before amino acid substitution of the poly-alanine model.
221 Well-defined bulky side chains of aromatic residues allowed a confident assignment of the register.
222 The crRNA was also well defined in the maps and was traced de novo with COOT. For Cas8
223 and TniQ in particular, no structural similarity was found in the published structures able to

224 explain our densities. Locally refined maps using soft masks at both ends of the crescent-shaped
225 complex rendered well-defined maps below 3.5 Å resolution. These maps were used for manual de
226 novo tracing of a poly-alanine model in COOT that was subsequently mutated to the *V. cholerae*
227 sequences. Bulky side chains for aromatic residues showed excellent density and were used as
228 landmarks to adjust the register of the sequence. For refinement, an initial step of real space
229 refinement against the cryo-EM maps was performed with the phenix real space refinement tool of
230 the Phenix package [24], with secondary structure restraints activated. A second step of reciprocal
231 space refinement was performed in Refmac5 [25], with secondary restraints calculated with Prosmart
232 [26] and LibG [27]. Weight of the geometry term versus the experimental term was adjusted to
233 avoid overfitting of the model into cryo-EM map, as previously reported³⁰. Model validation was
234 performed in Molprobit [28].

References

- [1] Ron L Dy, Corinna Richter, George P C Salmond, and Peter C Fineran. Remarkable mechanisms in microbes to resist phage infections. *Annu Rev Virol*, 1(1):307–31, Nov 2014.
- [2] Frank Hille, Hagen Richter, Shi Pey Wong, Majda Bratovič, Sarah Ressel, and Emmanuelle Charpentier. The biology of crispr-cas: Backward and forward. *Cell*, 172(6):1239–1259, 03 2018.
- [3] Shany Doron, Sarah Melamed, Gal Ofir, Azita Leavitt, Anna Lopatina, Mai Keren, Gil Amitai, and Rotem Sorek. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, 359(6379), 03 2018.
- [4] Tomas Sinkunas, Giedrius Gasiunas, Saktham P Waghmare, Mark J Dickman, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. In vitro reconstitution of cascade-mediated crispr immunity in streptococcus thermophilus. *EMBO J*, 32(3):385–94, Feb 2013.
- [5] Sy Redding, Samuel H Sternberg, Myles Marshall, Bryan Gibb, Prashant Bhat, Chantal K Guegler, Blake Wiedenheft, Jennifer A Doudna, and Eric C Greene. Surveillance and processing of foreign dna by the escherichia coli crispr-cas system. *Cell*, 163(4):854–65, Nov 2015.
- [6] Joseph E Peters, Kira S Makarova, Sergey Shmakov, and Eugene V Koonin. Recruitment of crispr-cas systems by tn7-like transposons. *Proc Natl Acad Sci U S A*, 114(35):E7358–E7366, 08 2017.

- 253 [7] Sanne E Klompe, Phuc L H Vo, Tyler S Halpin-Healy, and Samuel H Sternberg. Transposon-
254 encoded crispr-cas systems direct rna-guided dna integration. *Nature*, 571(7764):219–225, Jul
255 2019.
- 256 [8] Ryan N Jackson, Sarah M Golden, Paul B G van Erp, Joshua Carter, Edze R Westra, Stan
257 J J Brouns, John van der Oost, Thomas C Terwilliger, Randy J Read, and Blake Wieden-
258 heft. Structural biology. crystal structure of the crispr rna-guided surveillance complex from
259 escherichia coli. *Science*, 345(6203):1473–9, Sep 2014.
- 260 [9] Saikat Chowdhury, Joshua Carter, MaryClare F Rollins, Sarah M Golden, Ryan N Jack-
261 son, Connor Hoffmann, Lyn’Al Nosaka, Joseph Bondy-Denomy, Karen L Maxwell, Alan R
262 Davidson, Elizabeth R Fischer, Gabriel C Lander, and Blake Wiedenheft. Structure reveals
263 mechanisms of viral suppressors that intercept a crispr rna-guided surveillance complex. *Cell*,
264 169(1):47–57.e11, 03 2017.
- 265 [10] Tai Wei Guo, Alberto Bartesaghi, Hui Yang, Veronica Falconieri, Prashant Rao, Alan Merk,
266 Edward T Eng, Ashleigh M Raczkowski, Tara Fox, Lesley A Earl, Dinshaw J Patel, and
267 Sriram Subramaniam. Cryo-em structures reveal mechanism and inhibition of dna targeting
268 by a crispr-cas surveillance complex. *Cell*, 171(2):414–426.e12, Oct 2017.
- 269 [11] Sabin Mulepati, Annie Héroux, and Scott Bailey. Structural biology. crystal structure of a
270 crispr rna-guided surveillance complex bound to a ssdna target. *Science*, 345(6203):1479–84,
271 Sep 2014.

- 272 [12] Jasenko Zivanov, Takanori Nakane, Björn O Forsberg, Dari Kimanius, Wim Jh Hagen, Erik
273 Lindahl, and Sjors Hw Scheres. New tools for automated high-resolution cryo-em structure
274 determination in relion-3. *Elife*, 7, 11 2018.
- 275 [13] Jason W Holder and Nancy L Craig. Architecture of the tn7 posttransposition complex: an
276 elaborate nucleoprotein structure. *J Mol Biol*, 401(2):167–81, Aug 2010.
- 277 [14] Liisa Holm and Laura M Laakso. Dali server update. *Nucleic Acids Res*, 44(W1):W351–5, 07
278 2016.
- 279 [15] L Aravind, Vivek Anantharaman, Santhanam Balaji, M Mohan Babu, and Lakshminarayan M
280 Iyer. The many faces of the helix-turn-helix domain: transcription regulation and beyond.
281 *FEMS Microbiol Rev*, 29(2):231–62, Apr 2005.
- 282 [16] Eugene Krissinel. Stock-based detection of protein oligomeric states in jspisa. *Nucleic Acids
283 Res*, 43(W1):W314–9, Jul 2015.
- 284 [17] Guilhem Faure, Sergey A Shmakov, Winston X Yan, David R Cheng, David A Scott, Joseph E
285 Peters, Kira S Makarova, and Eugene V Koonin. Crispr-cas in mobile genetic elements: counter-
286 defence and beyond. *Nat Rev Microbiol*, 17(8):513–525, Aug 2019.
- 287 [18] Jonathan Strecker, Alim Ladha, Zachary Gardner, Jonathan L Schmid-Burgk, Kira S
288 Makarova, Eugene V Koonin, and Feng Zhang. Rna-guided dna insertion with crispr-associated
289 transposases. *Science*, 365(6448):48–53, 07 2019.

- 290 [19] Kai Zhang. Gctf: Real-time ctf determination and correction. *J Struct Biol*, 193(1):1–12, Jan
291 2016.
- 292 [20] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryosparc: algorithms
293 for rapid unsupervised cryo-em structure determination. *Nat Methods*, 14(3):290–296, 03 2017.
- 294 [21] Shawn Q Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A Verba, Yifan Cheng, and
295 David A Agard. Motioncor2: anisotropic correction of beam-induced motion for improved
296 cryo-electron microscopy. *Nat Methods*, 14(4):331–332, 04 2017.
- 297 [22] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Green-
298 blatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for ex-
299 ploratory research and analysis. *J Comput Chem*, 25(13):1605–12, Oct 2004.
- 300 [23] P Emsley, B Lohkamp, W G Scott, and K Cowtan. Features and development of coot. *Acta*
301 *Crystallogr D Biol Crystallogr*, 66(Pt 4):486–501, Apr 2010.
- 302 [24] Pavel V Afonine, Billy K Poon, Randy J Read, Oleg V Sobolev, Thomas C Terwilliger, Alexan-
303 dre Urzhumtsev, and Paul D Adams. Real-space refinement in phenix for cryo-em and crys-
304 tallography. *Acta Crystallogr D Struct Biol*, 74(Pt 6):531–544, 06 2018.
- 305 [25] G N Murshudov, A A Vagin, and E J Dodson. Refinement of macromolecular structures by
306 the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, 53(Pt 3):240–55, May
307 1997.

- 308 [26] Robert A Nicholls, Marcus Fischer, Stuart McNicholas, and Garib N Murshudov.
309 Conformation-independent structural comparison of macromolecules with prosmart. *Acta Crys-*
310 *tallogr D Biol Crystallogr*, 70(Pt 9):2487–99, Sep 2014.
- 311 [27] Alan Brown, Fei Long, Robert A Nicholls, Jaan Toots, Paul Emsley, and Garib Murshudov.
312 Tools for macromolecular model building and refinement into electron cryo-microscopy recon-
313 structions. *Acta Crystallogr D Biol Crystallogr*, 71(Pt 1):136–53, Jan 2015.
- 314 [28] Christopher J Williams, Jeffrey J Headd, Nigel W Moriarty, Michael G Prisant, Lizbeth L
315 Videau, Lindsay N Deis, Vishal Verma, Daniel A Keedy, Bradley J Hintze, Vincent B Chen,
316 Swati Jain, Steven M Lewis, W Bryan Arendall, 3rd, Jack Snoeyink, Paul D Adams, Simon C
317 Lovell, Jane S Richardson, and David C Richardson. Molprobity: More and better reference
318 data for improved all-atom structure validation. *Protein Sci*, 27(1):293–315, 01 2018.
- 319 [29] Israel S Fernández, Xiao-Chen Bai, Garib Murshudov, Sjors H W Scheres, and V Ramakrish-
320 nan. Initiation of translation by cricket paralysis virus ires requires its translocation in the
321 ribosome. *Cell*, 157(4):823–31, May 2014.
- 322 [30] Alp Kucukelbir, Fred J Sigworth, and Hemant D Tagare. Quantifying the local resolution of
323 cryo-em density maps. *Nat Methods*, 11(1):63–5, Jan 2014.

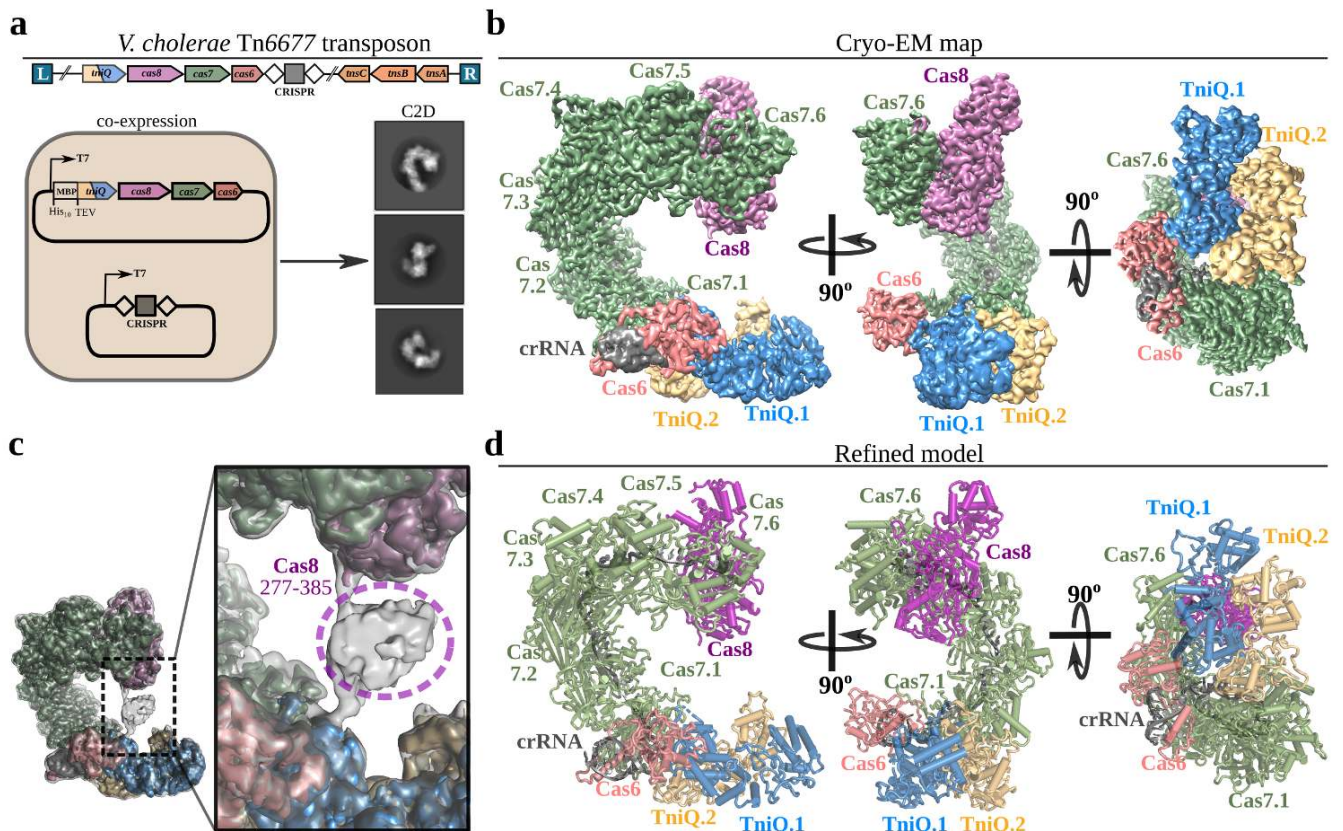
324 **Data availability.**

325 Maps and models have been deposited in the EMDB (accession codes 20349, 20350 and 20351) and
326 the PDB (accession codes 6PIF, 6PIG and 6PIJ).

327 **Acknowledgements.**

328 We acknowledge Bob Grassucci and Zhening Zhang for technical assistance in cryo-EM data acqui-
329 sition. Part of this work was performed at the Simons Electron Microscopy Center and National
330 Resource for Automated Molecular Microscopy located at the New York Structural Biology Center,
331 supported by grants from the Simons Foundation (SF349247), NYSTAR, and the NIH National
332 Institute of General Medical Sciences (GM103310).

Figure 1



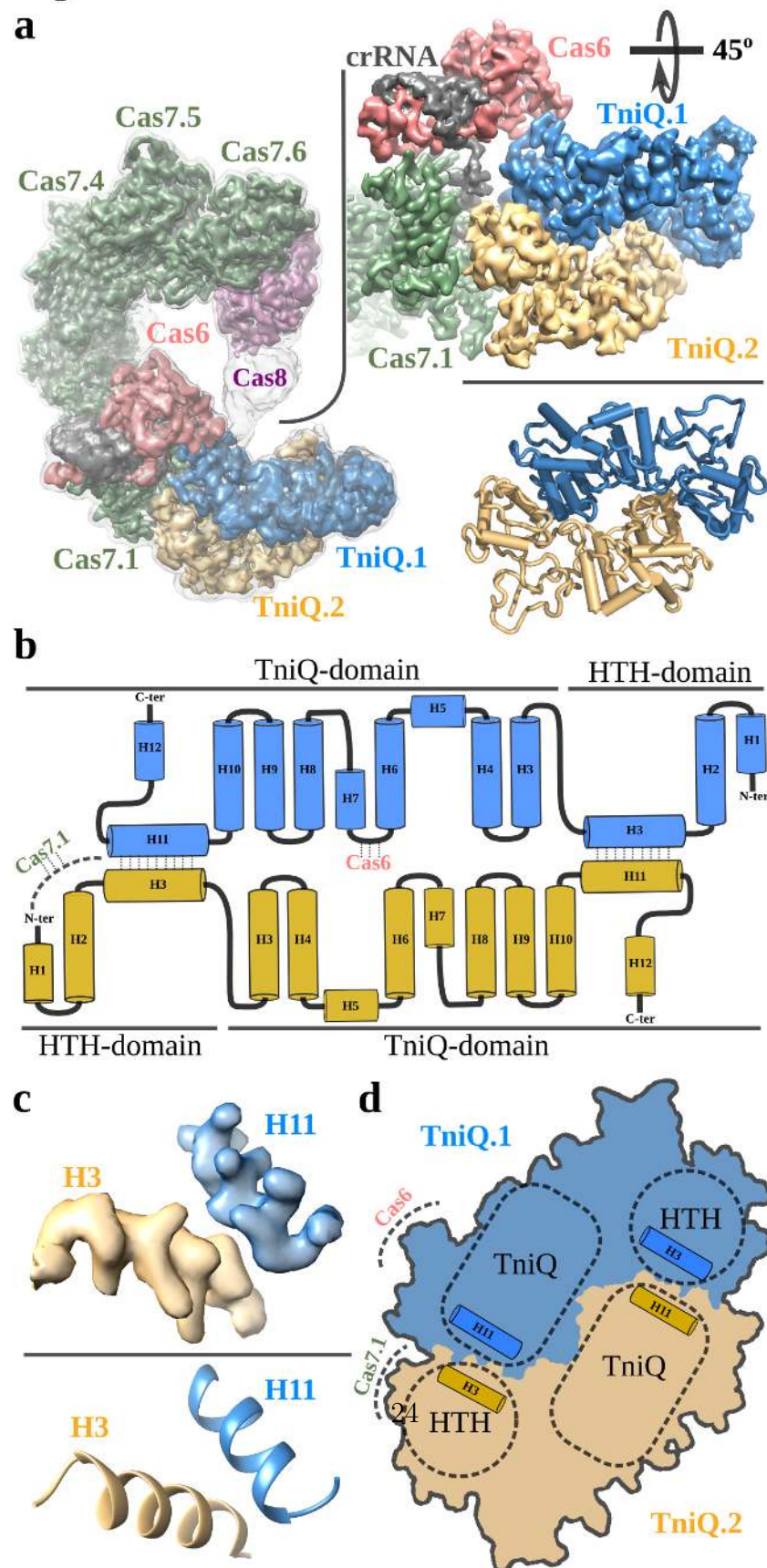
333

334 **Fig.1. Overall architecture of the *V. cholerae* TniQ-Cascade complex.**

335 **a**, Genetic architecture of the Tn6677 transposon (top), and plasmid constructs used to express
 336 and purify the TniQ-Cascade co-complex. Selected cryo-EM reference-free 2D classes in multiple
 337 orientations are shown on the right. **b**, Orthogonal views of the cryo-EM map for the TniQ-Cascade
 338 complex, showing Cas8 (pink), six Cas7 monomers (green), Cas6 (salmon), crRNA (grey), and TniQ
 339 monomers (blue, yellow). The complex adopts a helical architecture with protuberances at both
 340 ends. **c**, A flexible domain in Cas8 comprising residues 277-385 (grey) could only be visualized in

341 low-pass filtered maps. The unsharpened map is shown as semi-transparent, grey map overlaid on
342 the post-processed map segmented and colored according to a. **d**, Refined model for the TniQ-
343 Cascade complex derived from the cryo-EM maps shown in **b**.

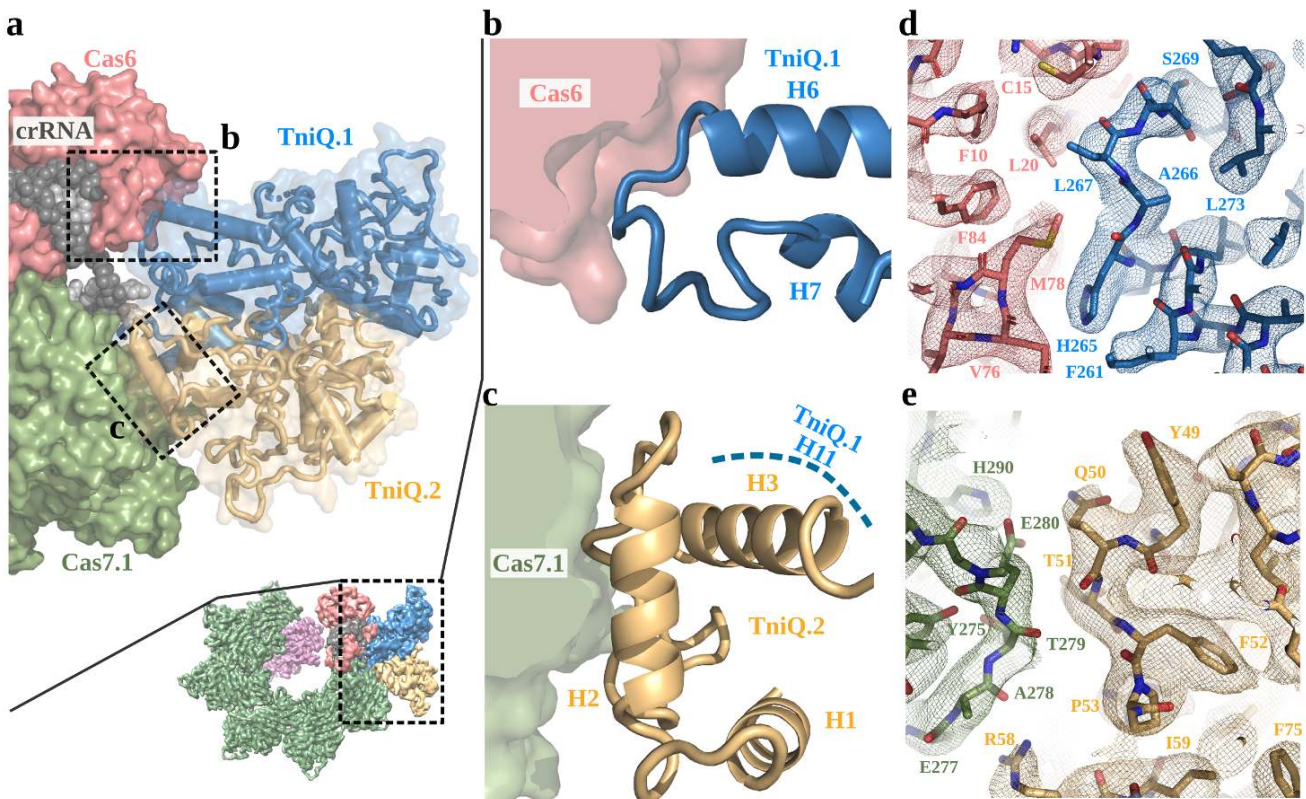
Figure 2



345 **Fig.2. TniQ binds Cascade in a dimeric, head-to-tail configuration.**

346 **a**, Left, overall view of the TniQ-Cascade cryo-EM unsharpened map (grey) overlaid on the post-
347 processed map segmented and colored as in **Fig.1**. Right, cryo-EM map (top) and refined model
348 (bottom) of the TniQ dimer. The two monomers interact with each other in a head-to-tail configura-
349 tion and are anchored to Cascade via Cas6 and Cas7.1. **b**, Secondary structure diagram of the TniQ
350 dimer: eleven α -helices are organized into an N-terminal Helix-Turn-Helix (HTH) domain and a C-
351 terminal TniQ-domain. Dimer interactions between H3 and H11 are indicated, as are interaction
352 sites with Cas6 and Cas7.1. **c**, Cryo-EM density for the H3-H11 interaction shows clear side-chain
353 features (top), allowing accurate modeling of the interaction (bottom). **d**, Schematic of the dimer
354 interaction, showing the important dimerization interface between the HTH and TniQ-domain.

Figure 3

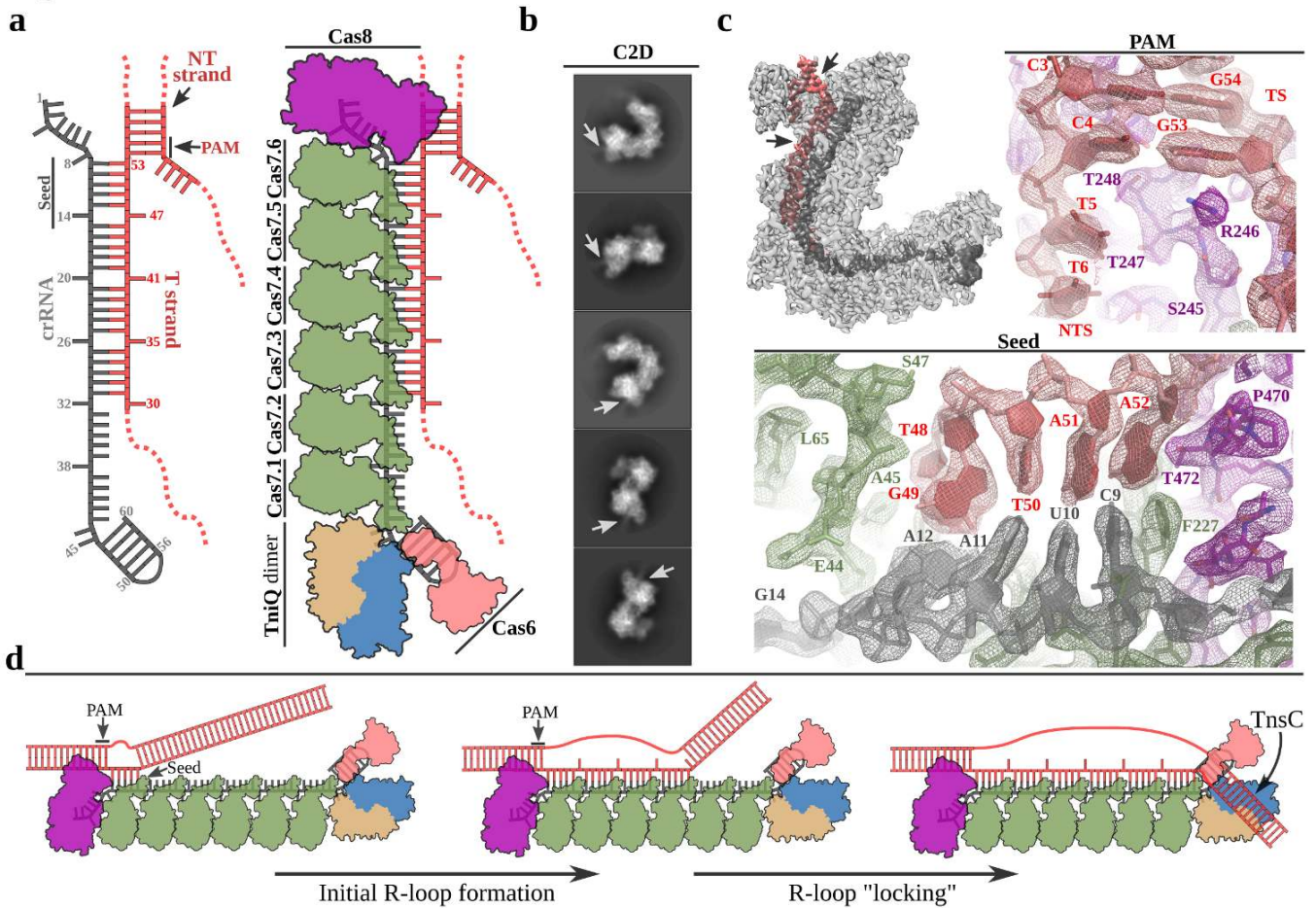


355

356 **Fig.3. Cas6 and Cas7.1 form a binding platform for TniQ.**

357 **a**, Top, zoomed area showing the interaction site of Cascade and the TniQ dimer. Cas6 and Cas7.1
358 are displayed as molecular Van der Waals surfaces, the crRNA is shown as grey spheres, and the
359 TniQ monomers as ribbons. **b**, The loop connecting TniQ.1 α -helices H6 and H7 (blue) binds
360 within a hydrophobic cavity of Cas6. **c**, Cas7.1 interacts via with the HTH domain of the TniQ.2
361 monomer (yellow), mainly through H2 and the loop connecting H2 and H3. **d-e**, Experimental
362 cryo-EM densities observed for the TniQ-Cas6 (textbfd) and TniQ-Cas7.1 (textbfe) interaction.

Figure 4



363

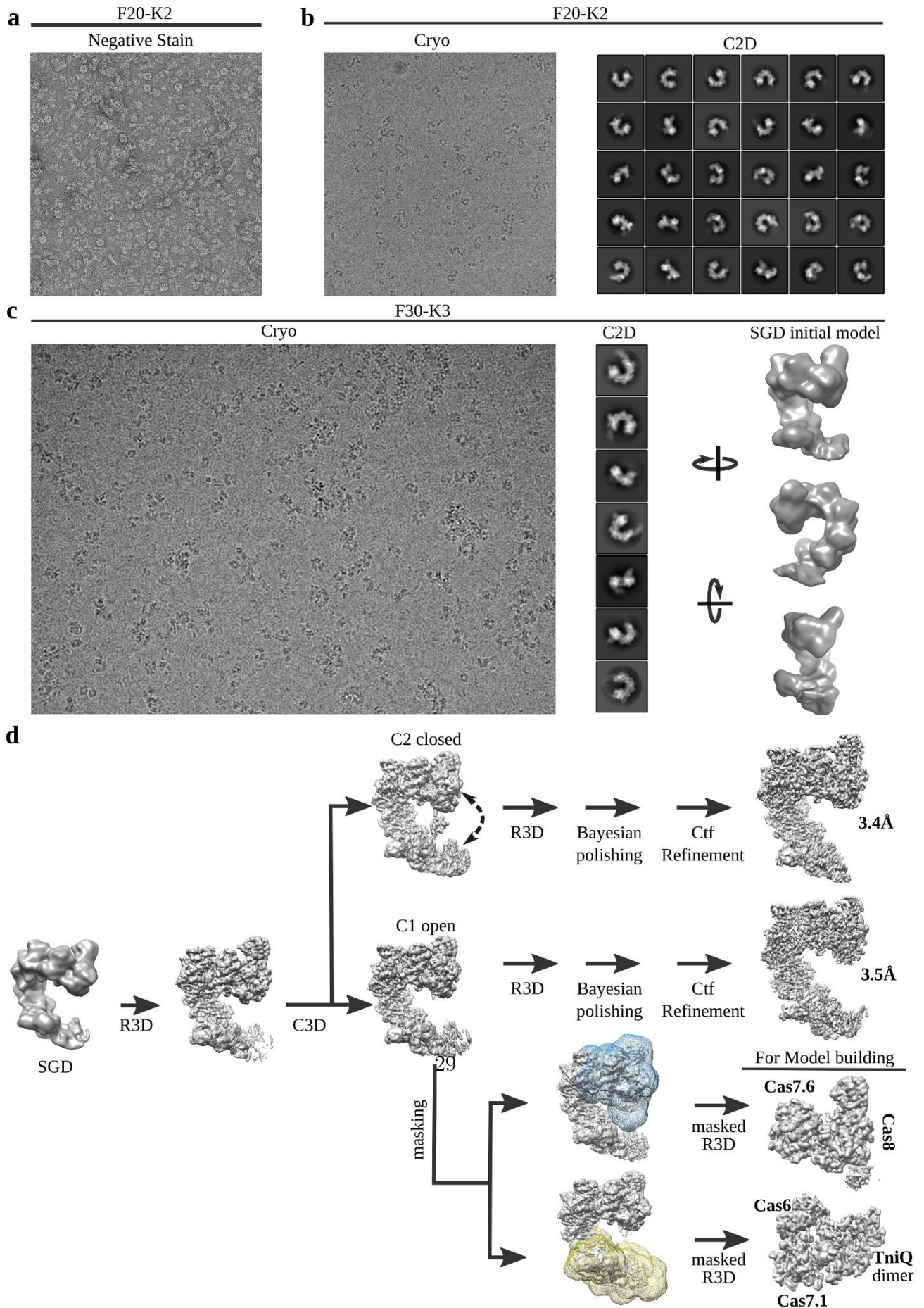
364 **Fig.4. DNA-bound structure of the TniQ-Cascade complex.**

365 **a**, Schematic of crRNA and the portion of the dsDNA substrate that was experimentally observed
 366 within the electron density map for DNA-bound TniQ-Cascade. Target Strand (TS), non-target
 367 strand (NTS), as well as the PAM and seed regions are indicated. **b**, Selected cryo-EM reference-
 368 free 2D classes for DNA-bound TniQ-Cascade; density corresponding to dsDNA could be directly
 369 observed protruding from the Cas8 component in the 2D averages (white arrows). **c**, Cryo-EM map

370 for DNA-bound TniQ-Cascade. The crRNA is in dark grey and the DNA is in red. On the right and
371 bottom, detailed views for the PAM and seed recognition regions of the map, with refined models
372 represented as sticks within the electron density. Cas8 is shown in pink, Cas7 in green, crRNA in
373 grey, and DNA in red. **d**, The *V. cholerae* transposon encodes a TniQ-Cascade co-complex that
374 utilizes the sequence content of the crRNA to bind complementary DNA target sites (left). We
375 propose that the incomplete R-loop observed in our structure (middle) represents an intermediate
376 state that may precede a downstream "locking" step involving proofreading of the RNA-DNA
377 complementarity. TniQ is positioned at the PAM-distal end of the DNA-bound Cascade complex,
378 where it likely interacts with TnsC during downstream steps of RNA-guided DNA insertion.

379 **EXTENDED DATA FIGURES**

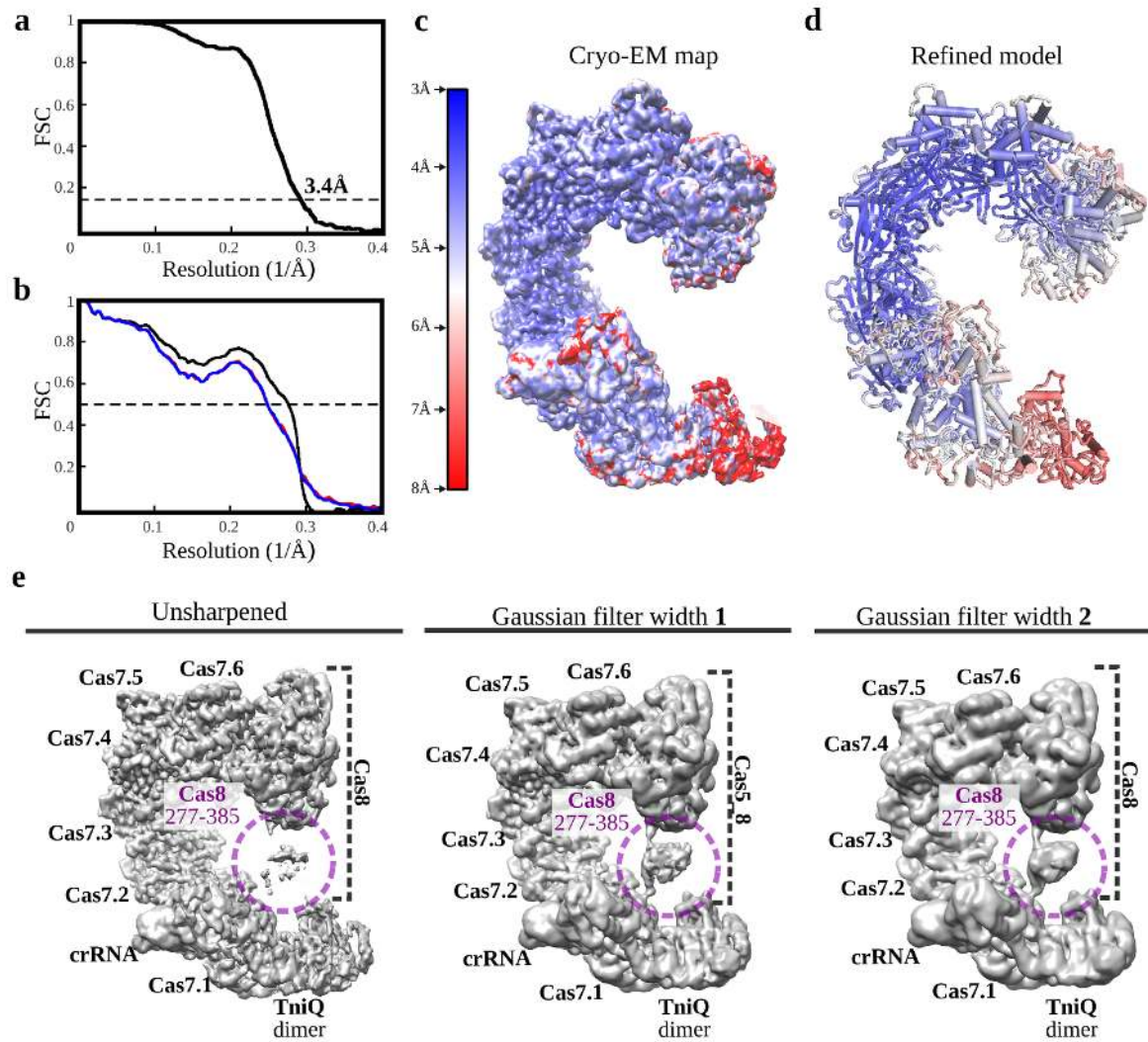
Extended Data Figure 1



381 **Extended Data Fig. 1. Cryo-EM sample optimization and image pro-**
382 **cessing workflow.**

383 **a**, Representative negatively stained micrograph for 500 nM TniQ-Cascade. **b**, Left, representative
384 cryo-EM image for 2 μ M TniQ-Cascade. A small dataset of 200 images was collected in a Tecnai-F20
385 microscope equipped with a Gatan K2 camera. Right, reference-free 2D class averages for this initial
386 cryo-EM dataset. **c**, Left, representative image from a large dataset collected in a Tecnai Polara
387 microscope equipped with a Gatan K3 detector. Middle, detailed 2D class averages were obtained
388 that were used for initial model generation using the SGD algorithm implemented in Relion3 (right).
389 **d**, Image processing workflow used to identify the two main classes of the TniQ-cascade complex in
390 open and closed conformations. Local refinements with soft masks were used to improve the quality
391 of the map within the terminal protuberances of the complex. These maps were instrumental for
392 *de novo* modeling and initial model refinement.

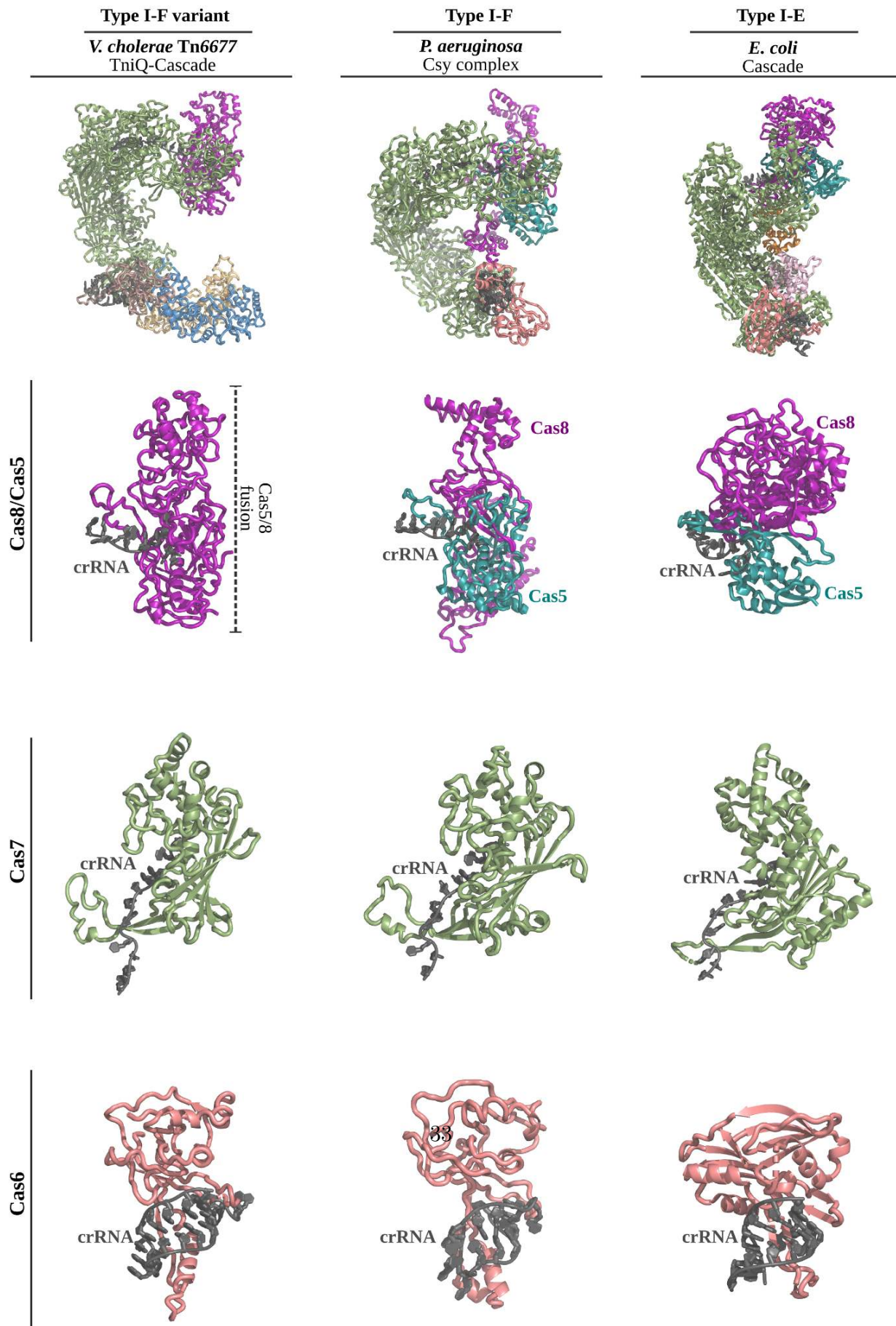
Extended Data Figure 2



394 **Extended Data Fig. 2. Fourier Shell Correlation (FSC) curves, local**
395 **resolution, and unsharpened filter maps for the TniQ-Cascade complex**
396 **in closed conformation.**

397 **a**, Gold-standard FSC curve using half maps; the global resolution estimation is 3.4 Å by the FSC
398 0.143 criterion. **b**, Cross-validation model-vs-map FSC. Blue curve, FSC between the shacked model
399 refined against half map 1; red curve, FSC against half map 2, not included in the refinement; black
400 curve, FSC between final model against the final map. The overlap observed between the blue
401 and red curves guarantees a non-overfitted model [29]. **c**, Unsharpened map colored according to
402 local resolutions, as reported by RESMAP [30]. **d**, Final model colored according to B-factors
403 calculated by REFMAC. **e**, A flexible Cas8 domain encompassing residues 277-385 contacts the
404 TniQ dimer at the other side of the crescent shape. Applying a Gaussian filter of increasing width
405 to the unsharpened map allows for a better visualization of this flexible region.

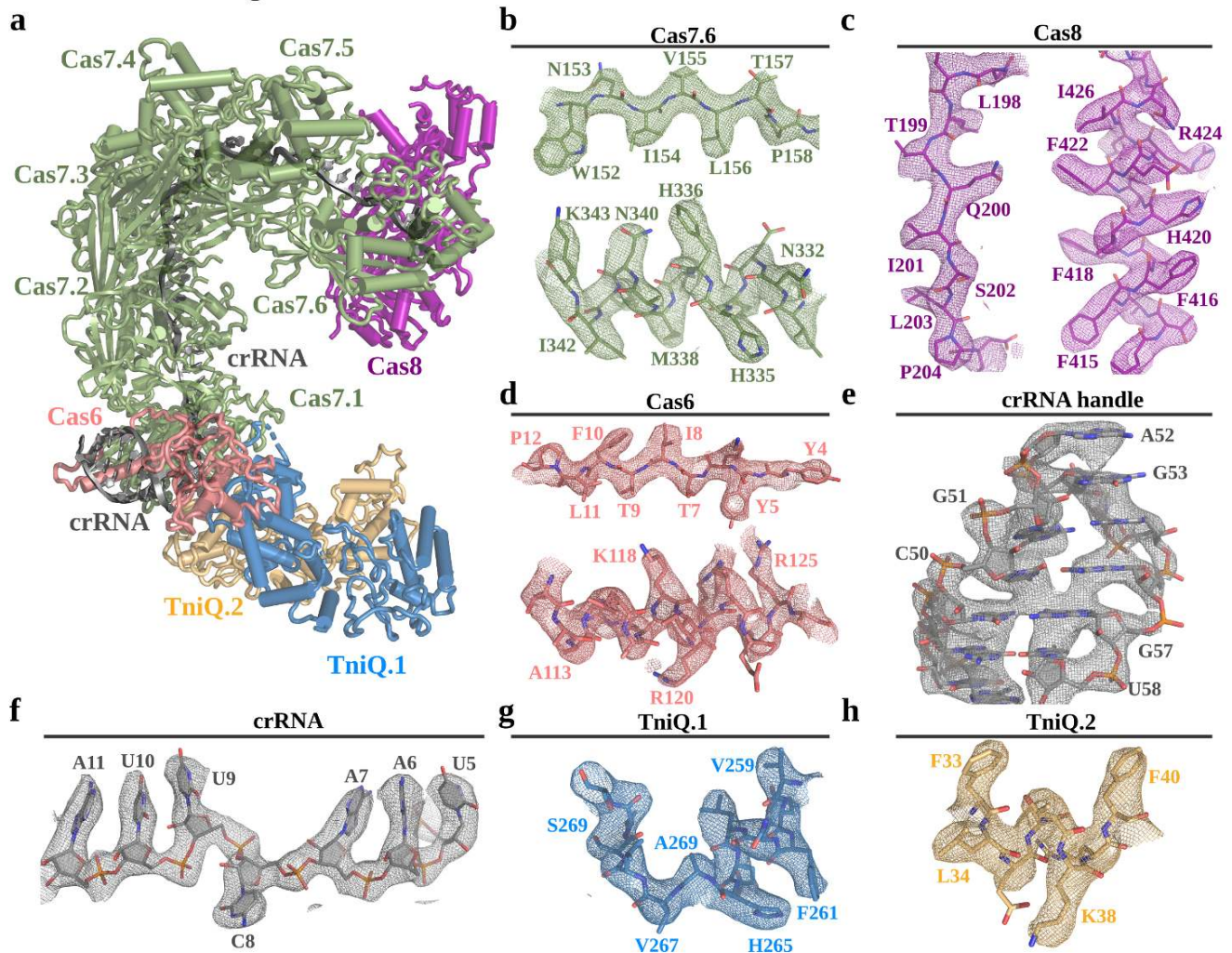
Extended Data Figure 3



407 **Extended Data Fig. 3. Superposition of TniQ-Cascade with structurally**
408 **similar Cascade complexes.**

409 The *V. cholerae* I-F variant TniQ-Cascade complex (left) was superposed with *Pseudomonas aerug-*
410 *inosa* I-F Cascade11 (also known as Csy complex; middle, PDB ID: 6B45) and *Escherichia coli* I-E
411 Cascade9 (right, PDB ID: 4TVX). Shown are superpositions of the entire complex (top), the Cas8
412 and Cas5 subunits with the 5' crRNA handle (middle top), the Cas7 subunit with a fragment of
413 crRNA (middle bottom), and the Cas6 subunit with the 3' crRNA handle (bottom).

Extended Data Figure 4



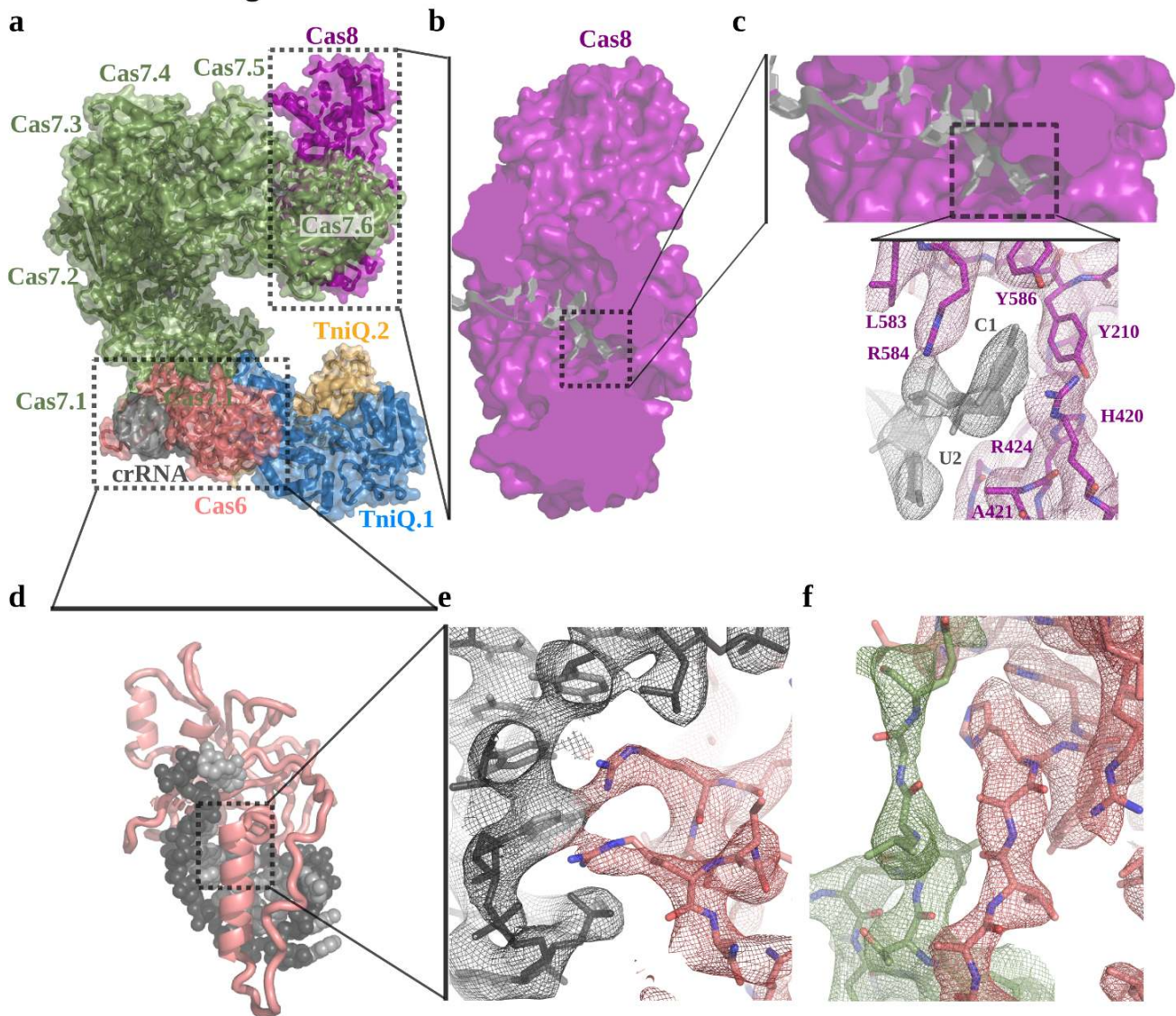
414

415 Extended Data Fig. 4. Representative cryo-EM densities for all the
416 components of the TniQ-Cascade complex in closed conformation.

417 a, Final refined model of TniQ-Cascade, with Cas8 in purple, Cas7 monomers in green, Cas6 in red,
418 the TniQ monomers in blue and yellow, and the crRNA in grey. b-h, Final refined model inserted

419 in the final cryo-EM density for select regions of all the molecular components of the TniQ-Cascade
420 complex. Residues are numbered.

Extended Data Figure 5



421

422 **Extended Data Fig. 5. Cas8 and Cas6 interaction with the crRNA.**

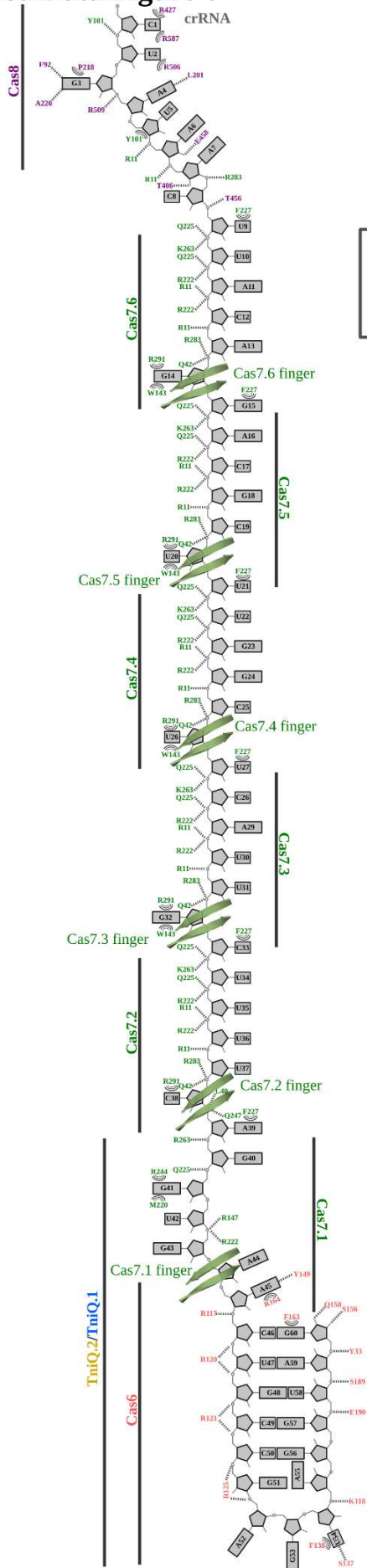
423 **a**, Refined model for the TniQ-Cascade shown as ribbons inserted in the semitransparent Van der

424 Walls surface, colored as in Fig 1. **b** and **c**, Zoomed view of Cas8, which interacts with the 5'

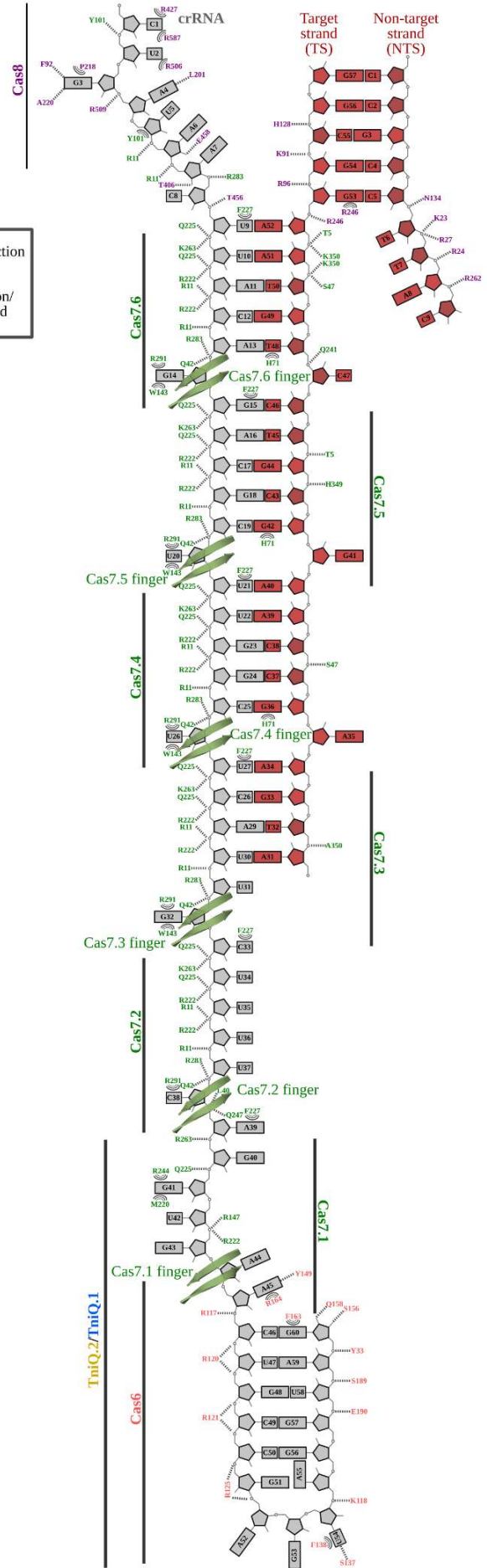
425 end of the crRNA. The inset shows electron density for the highlighted region, where the base of
426 nucleotide C1 is stabilized by stacking interactions with arginine residues R584 and R424. **d**, Cas6
427 interacts with the 3' end of the crRNA "handle" (nucleotides 45-60). **e**, An arginine-rich α -helix
428 is deeply inserted within the major groove of the terminal stem-loop. This interaction is mediated
429 by electrostatic interactions between basic residues of Cas6 and the negatively charged phosphate
430 backbone of the crRNA. **f**, Cas6 (red) also interacts with Cas7.1 (green), establishing a β -sheet
431 formed by β -strands contributed from both proteins.

Extended Data Figure 6

a



b



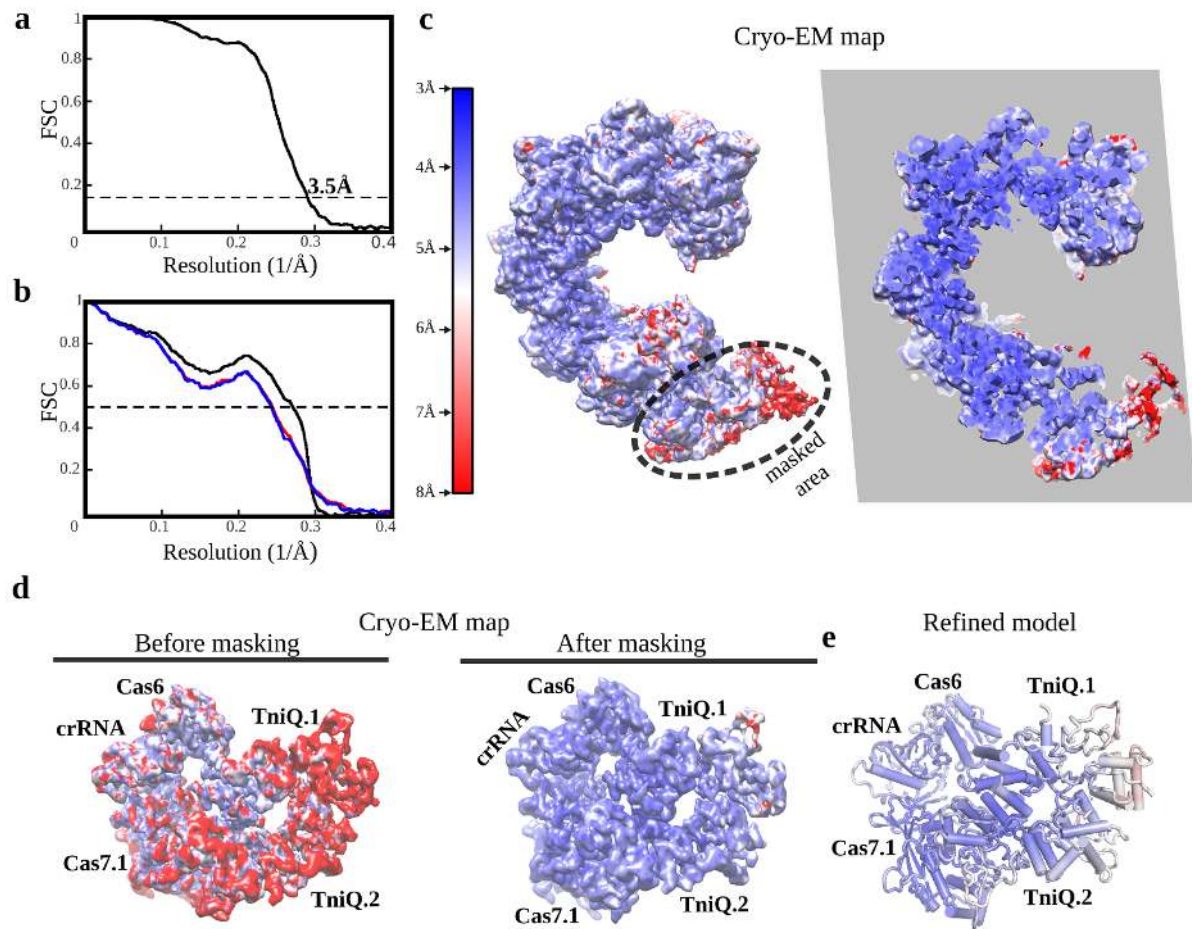
433 **Extended Data Fig. 6. Schematic representation of crRNA and target**

434 **DNA recognition by TniQ-Cascade.**

435 **a**, TniQ-Cascade residues that interact with the crRNA are indicated. Approximate location for all
436 protein components of the complex are also shown, as well as the position of each Cas7 "finger".?

437 **b**, TniQ-Cascade residues that interact with crRNA and target DNA, shown as in **a**.

Extended Data Figure 7



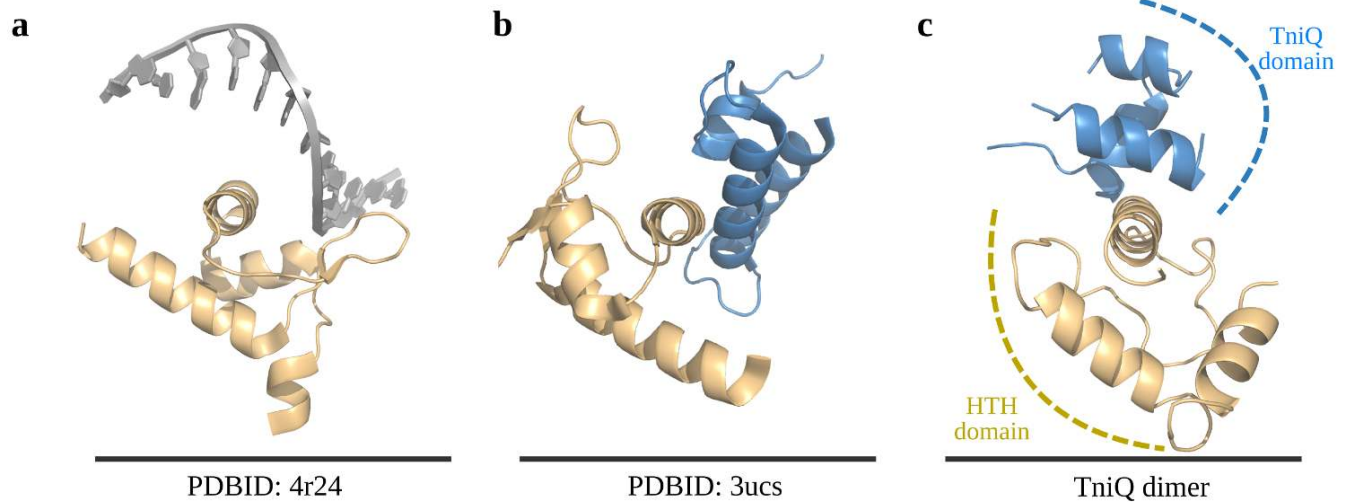
438

439 **Extended Data Fig. 7. Fourier Shell Correlation (FSC) curves, local**
440 **resolution, and local refined maps for the TniQ-Cascade complex in open**
441 **conformation.**

442 **a**, Gold-standard FSC curve using half maps; the global resolution estimation is 3.5 Å by the FSC
443 0.143 criterion. **b**, Cross-validation model-vs-map FSC. Blue curve, FSC between shacked model

444 refined against half map 1; red curve, FSC against half map 2, not included in the refinement;
445 black curve, FSC between final model against the final map. The overlapping between the blue
446 and red curves guarantees a non-overfitted model. **c**, Unsharpened map colored according to local
447 resolutions, as reported by RESMAP. Right, slice through the map shown on the left. **d**, Local
448 refinements with soft masks improved the maps in flexible regions. Shown the region of the map
449 corresponding to the TniQ dimer. Unsharpened maps colored according to the local resolution
450 estimations are shown before (left) and after (right) masked refinements. **e**, Final model for the
451 TniQ dimer region, colored according to the local B-factors calculated by REFMAC.

Extended Data Figure 8

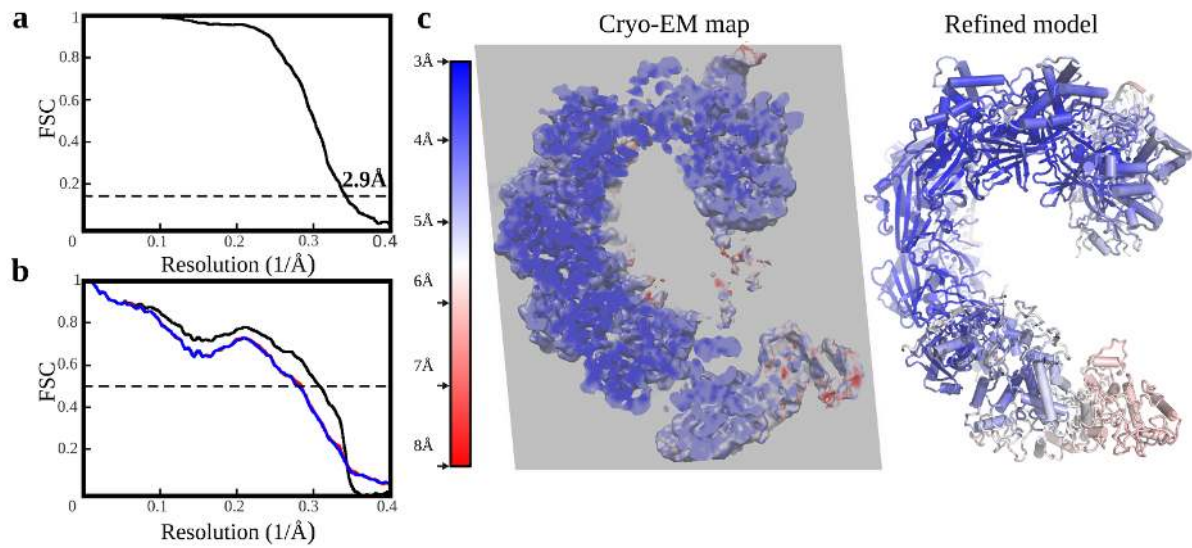


452

453 **Extended Data Fig. 8. TniQ harbors a HTH domain involved in protein-**
454 **protein interactions within the TniQ dimer.**

455 A DALI search¹⁵ using the refined TniQ model as probe found significant similarity between the
456 N-terminal domain of TniQ with PDB entries 4r24 (**a**) and 3ucs (**b**) (Z score 4.1/4.1, r.m.s.d.
457 3.8/5.1). Both proteins contain Helix-Turn-Helix (HTH) domains and HTH domains are often
458 involved in nucleic acid recognition and mediate protein-protein interactions [15]. **c**, The TniQ
459 dimer is stabilized in a head-to-tail configuration by reciprocal interactions mediated by the HTH
460 domain and the TniQ-domains from both monomers.

Extended Data Figure 9

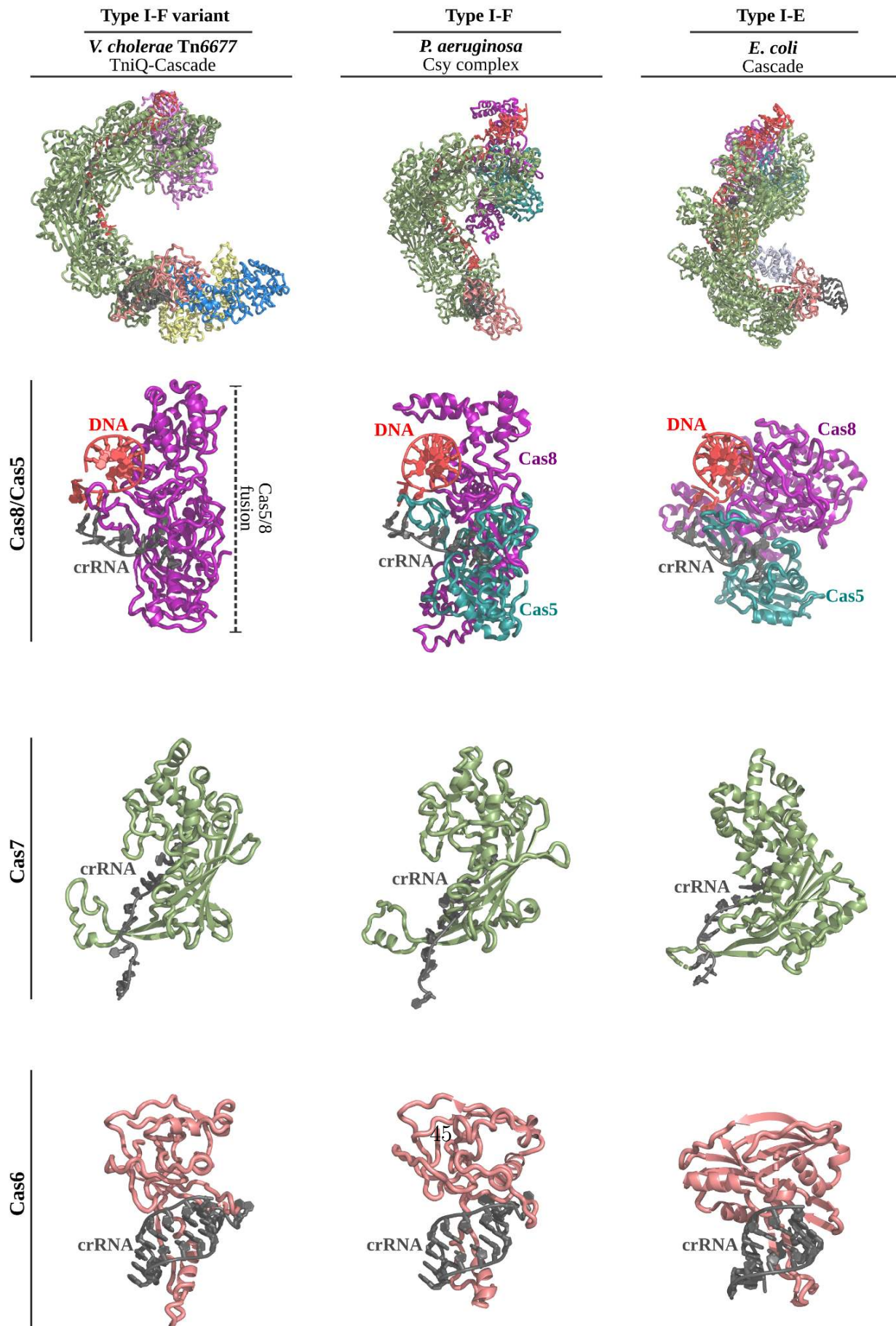


461

462 **Extended Data Fig. 9. Fourier Shell Correlation (FSC) curves, local res-**
463 **olution, and unsharpened filter maps for the DNA-bound TniQ-Cascade**
464 **complex complex.**

465 **a**, Gold-standard FSC curve using half maps; the global resolution estimation is 2.9 Å by the FSC
466 0.143 criterion. **b**, Cross-validation model-vs-map FSC. Blue curve, FSC between the sharpened model
467 refined against half map 1; red curve, FSC against half map 2, not included in the refinement; black
468 curve, FSC between final model against the final map. The overlap observed between the blue and
469 red curves guarantees a non-overfitted model. **c**, Left, unsharpened map colored according to local
470 resolutions, as reported by RESMAP. dsDNA is visible at the top right projecting outside of the
471 complex. Right, final model colored according to B-factors calculated by REFMAC.

Extended Data Figure 10



473 **Extended Data Fig. 10. Superposition of DNA-bound TniQ-Cascade**
474 **with structurally similar Cascade complexes.**

475 The DNA-bound structure of *V. cholerae* I-F variant TniQ-Cascade complex (left) was superposed
476 with DNA-bound structures of *Pseudomonas aeruginosa* I-F Cascade11 (also known as Csy complex;
477 middle, PDB ID: 6B44) and *Escherichia coli* I-E Cascade9 (right, PDB ID: 5H9F). Shown are
478 superpositions of the entire complex (top), the Cas8 and Cas5 subunits with the 5' crRNA handle
479 and double-stranded PAM DNA (middle top), the Cas7 subunit with a fragment of crRNA (middle
480 bottom), and the Cas6 subunit with the 3' crRNA handle (bottom).