ARTICLE

https://doi.org/10.1038/s41586-019-0908-x

CasX enzymes comprise a distinct family of RNA-guided genome editors

Jun-Jie Liu[1,2,3,12], Natalia Orlova[2,12], Benjamin L. Oakes[4,12], Enbo Ma[1], Hannah B. Spinner[4], Katherine L. M. Baney[4], Jonathan Chuck[1], Dan Tan[5], Gavin J. Knott[1], Lucas B. Harrington[1], Basem Al-Shayeb[6], Alexander Wagner[7], Julian Brötzmann[8], Brett T. Staahl[1,4], Kian L. Taylor[4], John Desmarais[4], Eva Nogales[1,2,3,9]* & Jennifer A. Doudna[1,3,4,9,10,11]*

The RNA-guided CRISPR-associated (Cas) proteins Cas9 and Cas12a provide adaptive immunity against invading nucleic acids, and function as powerful tools for genome editing in a wide range of organisms. Here we reveal the underlying mechanisms of a third, fundamentally distinct RNA-guided genome-editing platform named CRISPR–CasX, which uses unique structures for programmable double-stranded DNA binding and cleavage. Biochemical and in vivo data demonstrate that CasX is active for *Escherichia coli* and human genome modification. Eight cryo-electron microscopy structures of CasX in different states of assembly with its guide RNA and double-stranded DNA substrates reveal an extensive RNA scaffold and a domain required for DNA unwinding. These data demonstrate how CasX activity arose through convergent evolution to establish an enzyme family that is functionally separate from both Cas9 and Cas12a.

Archaea and bacteria use CRISPR–Cas systems for adaptive immunity against invading nucleic acids[1,2]. CRISPR arrays, which consist of repeated sequences interleaved with sequences acquired from foreign DNA, are templates for CRISPR RNAs (crRNAs) that guide a Cas nuclease to cleave complementary DNA sequences. In addition to their microbial functions, RNA-guided DNA binding and cutting proteins have proven to be transformative tools for genome and epigenome editing across a wide range of cell types and organisms[3–5]. Despite extensive effort, only two kinds of CRISPR–Cas nuclease—Cas9 and Cas12a (also known as Cpf1)—provide the foundation for this revolutionary technology[6,7].

Metagenomic analysis of microbial DNA from groundwater samples revealed a new protein—which we refer to here as CasX (a placeholder name, pending re-analysis of the class 2 CRISPR–Cas phylogeny; CasX is also known as Cas12e[5])—that prevented bacterial transformation by plasmid DNA when expressed with cognate crRNAs that target the plasmid[8]. Sequence analysis of CasX revealed no similarity to other CRISPR–Cas enzymes, except for the presence of a RuvC nuclease domain that is also found in both Cas9 and Cas12a enzyme families, transposases, and recombinases[8]. Phylogenetic analysis suggests that CasX arose from a TnpB-type transposase by an independent insertion event into ancestral CRISPR loci, distinct from Cas12a and the remaining type V effectors (Extended Data Fig. 1a). Consistent with this hypothesis, the only conserved part of CasX, the RuvC domain, shares less than 16% identity with RuvC domains in either Cas9 or Cas12a (Extended Data Fig. 1b). The evolutionary ambiguity of CasX hinted that this enzyme may have a structure and molecular mechanism distinct from that of other CRISPR–Cas enzymes. However, without full reconstitution of the CasX enzyme, it has not been possible to determine the basis of the previously reported plasmid interference activity.

Here we demonstrate that CasX is an RNA-guided DNA endonuclease that generates a staggered double-stranded break in DNA at sequences complementary to a 20-nucleotide segment of its guide RNA. We further find that CasX induces programmable, site-specific genome repression in *E. coli* and genome editing in human cells. Eight molecular models of CasX in different states (Supplementary Table 1), determined by cryo-electron microscopy (cryo-EM), reveal an unanticipated quaternary structure in which the RNA scaffold dominates the architecture and organization of the enzyme. Phylogenetic, biochemical and structural data show that CasX contains domains distinct from—but analogous to—those found in Cas9 and Cas12a, as well as novel RNA and protein folds; thus establishing the CasX enzyme family as the third CRISPR–Cas platform that is effective for genetic manipulation. Finally, distinct conformational states observed for CasX suggest an ordered non-target- and target-strand cleavage mechanism that may explain how CRISPR–Cas enzymes with a single active site, such as Cas12a, achieve double-stranded DNA (dsDNA) cleavage[5,9,10]. The small size of CasX (<1,000 amino acids), its DNA cleavage characteristics, and its derivation from non-pathogenic microorganisms offer important advantages over other CRISPR–Cas genome-editing enzymes.

Reconstituting RNA-guided CasX cutting of dsDNA

It has previously been demonstrated that CasX proteins can perform RNA-dependent plasmid interference in bacteria, and that the two natural RNAs necessary for this activity (crRNA and *trans*-activating CRISPR RNA (tracrRNA)) can be combined into a single-guide RNA (sgRNA) format[8] (Fig. 1a). To determine the precise molecular activity of this enzyme, we undertook biochemical studies of the wild-type CasX from Deltaproteobacteria (hereafter referred to as DpbCasX). We found that purified DpbCasX with sgRNA is capable of cleaving dsDNA that bears a sequence complementary to the 20-nucleotide guide RNA segment and adjacent to a TTCN protospacer adjacent motif (PAM) (Fig. 1b). Mapping the cut sites for the target and non-target strands of the DNA showed that DpbCasX generates products with staggered ends about 10-nucleotides in length, owing to cleavage at 12–14 nucleotides

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [2]California Institute for Quantitative Biosciences, University of California, Berkeley, CA, USA. [3]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Innovative Genomics Institute, University of California, Berkeley, CA, USA. [5]Clayton Foundation Laboratories of Peptide Biology, Salk Institute for Biological Studies, La Jolla, CA, USA. [6]Department of Plant and Microbiology, University of California, Berkeley, CA, USA. [7]Max-Planck-Institute for Biochemistry, Planegg, Germany. [8]Faculty of Chemistry and Pharmacy, Ludwig-Maximilians-University, Munich, Germany. [9]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA. [10]Department of Chemistry, University of California, Berkeley, CA, USA. [11]Gladstone Institutes, San Francisco, CA, USA. [12]These authors contributed equally: Jun-Jie Liu, Natalia Orlova, Benjamin L. Oakes. *e-mail: enogales@lbl.gov; doudna@berkeley.edu

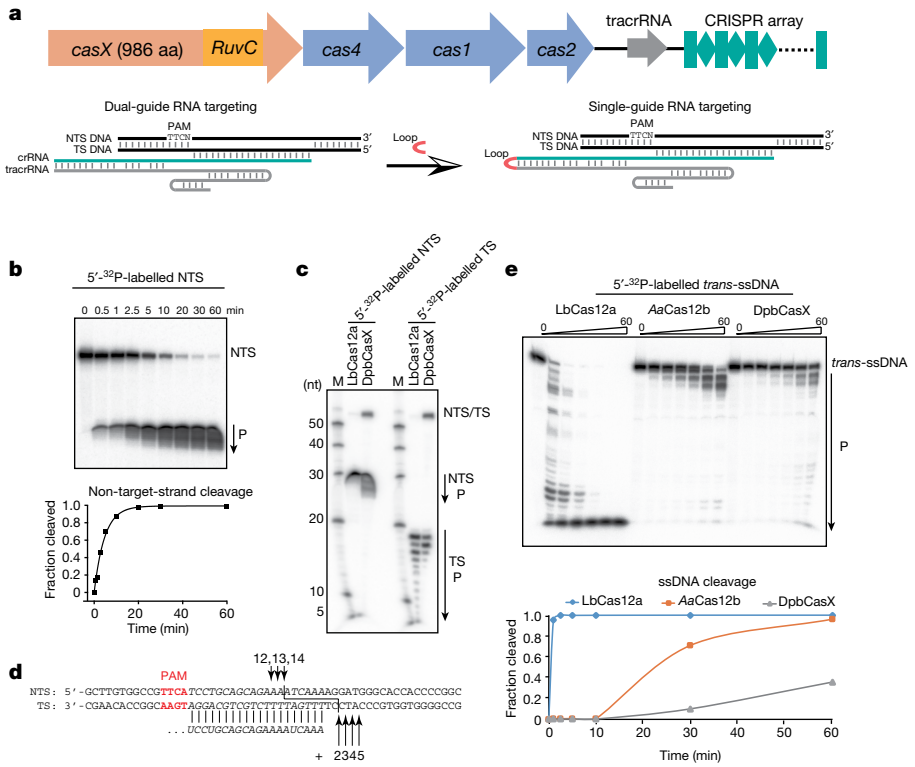


Fig. 1 | CasX cuts dsDNA with sgRNA in vitro. a, Top, schematic of CRISPR–CasX locus with *casX* gene (*RuvC* domain highlighted) in orange, *cas4*, *cas1* and *cas2* in blue, tracrRNA in grey and CRISPR array in teal. Cartoons are scaled according to the gene size. Bottom, schematic of the CasX dual-guide RNA and single guide; tracrRNA in grey, crRNA in teal and the target DNA in black. TS and NTS denote target strand and non-target strand, respectively. The RNA loop that connects the tracrRNA and crRNA is depicted in red. aa, amino acids. **b**, DNA cleavage efficiency by DpbCasX. P indicates the cleavage product. The cleavage fraction is calculated on the basis of the NTS band density compared to input NTS band density at reaction time of 0 min. **c**, Cleavage site mapping of DpbCasX and LbCas12a. Lane M shows labelled ladders. **d**, The cleavage sites for NTS and TS (marked with black arrows). **e**, Cleavage activity of DpbCasX on *trans*-ssDNA. The cleavage fraction is calculated based on the *trans*-ssDNA band density compared to input *trans*-ssDNA band density at reaction time of 0 min. *Aa*Cas12b, *Alicyclobacillus acidoterrestris* Cas12b. Four biological repeats for all the assays showed consistent results.

after the PAM on the non-target strand and at 22–25 nucleotides after the PAM on the target strand (Fig. 1c, d, Extended Data Fig. 1c). This mode of dsDNA cleavage is consistent with the staggered cuts to DNA observed for Cas12a and Cas12b (also known as C2c1), which represent other CRISPR–Cas enzymes that use a single RuvC active site for DNA cleavage[5,10,11].

Unlike Cas9, Cas12a becomes a highly active single-stranded DNA (ssDNA) nuclease after binding to target DNA, triggering non-specific ssDNA degradation[12,13]. To test whether CasX displays similar non-specific activity, single-stranded phage DNA was incubated with DpbCasX–guide RNA complexes that target a separate, unrelated dsDNA substrate. We found that *trans*-ssDNA cutting activity was minimal compared to that observed for Lachnospiraceae bacterium Cas12a (LbCas12a) and another related Cas12b enzyme (Fig. 1e). These results indicate that the presence of a single active site for dsDNA cleavage does not necessarily correspond to target-dependent *trans*-cleavage activity, hinting at structural and mechanistic differences between these enzyme families.

CasX enacts genome silencing and editing

To determine whether the RNA-guided DNA cutting activity of CasX can be harnessed for programmed genome targeting, DpbCasX and its sgRNA were expressed in *E. coli* using a guide sequence that is complementary to an integrated reporter in the genome of bacterial strain MG1655[14,15]. We found that DpbCasX reduced cell viability in a genome cleavage assay at close to—but slightly less than—the activity levels of *Streptococcus pyogenes* Cas9 (*Sp*Cas9) (Fig. 2a). We next tested whether CasX can function as an effector of CRISPR interference (CRISPRi) in *E. coli*. By aligning the RuvC domain among CasX and other Cas proteins, we found three potential active site residues: Asp672, Glu769 and Asp935. Biochemical screening indicated that mutating these residues to alanine—individually or in combination—creates a deactivated DpbCasX (Extended Data Fig. 1d) that is competent for RNA-guided DNA binding and gene repression rather than cutting (Fig. 2b). Silencing of green fluorescent protein (GFP) expression was observed with this deactivated DpbCasX construct using different guide RNAs that target multiple sites within the gene that encodes GFP (Extended Data Fig. 1e). We found that mutation of all three DpbCasX residues was required for maximal gene repression activity (Extended Data Fig. 1f). CasX-based bacterial CRISPRi thus provides an ideal system for rapid, visual and quantitative in vivo characterization of CasX constructs (Fig. 2c, d, Extended Data Fig. 1e).

We next tested whether CasX is capable of inducing cleavage and gene editing of mammalian genomes. Using a previously reported destabilized-GFP disruption assay[15] (Fig. 2e), we found that DpbCasX can induce targeted GFP gene disruption in HEK293T cells with limited efficiency, using guide RNAs complementary to either the template or the coding strand (guide 2 or guide 3, respectively) (Fig. 2f). We also explored the effectiveness of the CasX molecule from Planctomycetes (PlmCasX), which bears approximately 70% sequence identity to DpbCasX and can use the same sgRNA[8]. We find that PlmCasX enacts destabilized-GFP gene editing with higher efficiency than DpbCasX. Because the guide RNA that recognizes the coding strand functioned more robustly in each case, we wondered whether the additional GFP gene disruption observed for guide 3 could be explained by RNA targeting[4,16]. Nevertheless, there was no recovery of GFP expression in these cells over time (Fig. 2g), which is consistent with genome editing rather than transcript targeting. Furthermore, analysis of DNA derived from the PlmCasX-targeted GFP locus using a T7E1-based assay[17] revealed levels of genome editing consistent with

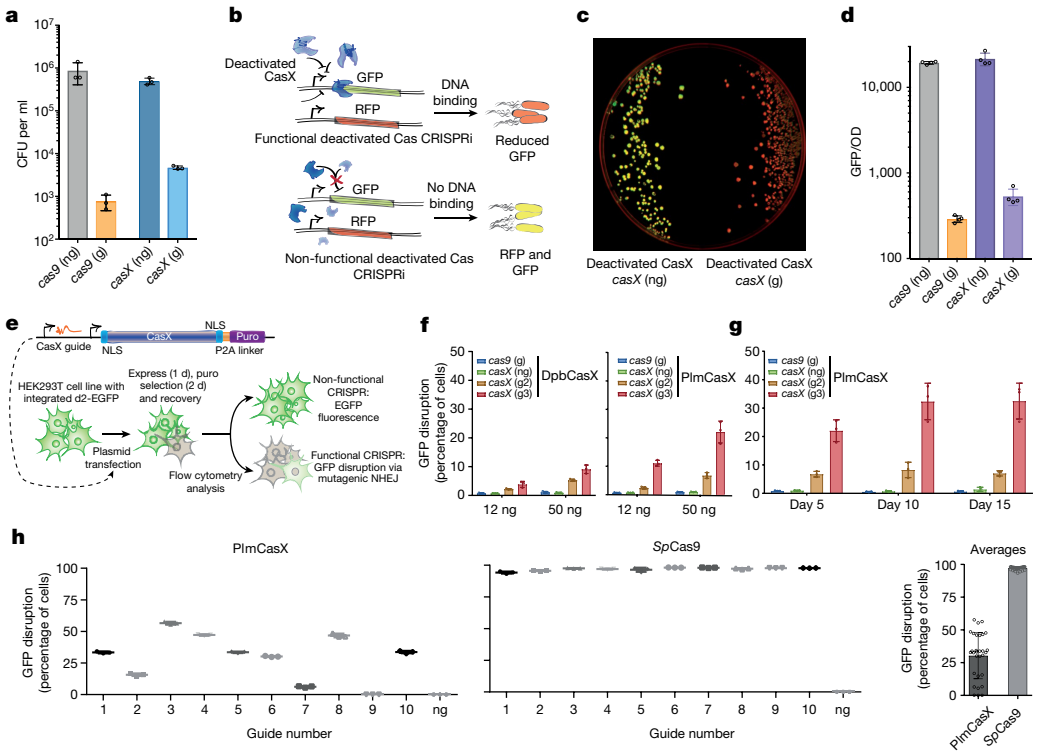


Fig. 2 | CasX effectively manipulates genomes in vivo. a, Genomic cleavage assay in *E. coli* ($n = 3$, mean ± s.d.). CFU, colony-forming units. **b**, Schematic of *E. coli* CRISPRi. **c**, *E. coli* GFP repression as visualized on plates on a dark reader. This assay was repeated more than three times with consistent results. **d**, Quantitative analysis (GFP/OD; GFP signal divided by optical density at 600 nm) of *E. coli* CRISPRi-based GFP repression at 12 h ($n = 4$, mean ± s.d.). **e**, Schematic of CasX human cell assay and readout. **f**, DpbCasX and PlmCasX GFP disruption in mammalian cell (HEK293T) assays at two doses of plasmids. **g**, Sustained GFP disruption of the high-dosage mammalian cell GFP-disruption assay from **f**. **h**, PlmCasX and *Sp*Cas9 GFP disruption at ten guide sites throughout EGFP ($n = 3$, mean ± s.d.). The average GFP disruption across all EGFP guides for CasX and Cas9 is shown. *cas9* (g), targeting guide RNA of *Sp*Cas9; *cas9* (ng), non-targeting RNA guide of *Sp*Cas9; *casX* (g), targeting RNA guide of DpbCasX; *casX* (ng), non-targeting RNA guide of DpbCasX. *Sp*Cas9 or inactive *Sp*Cas9 (d*Sp*Cas9) were used as positive controls.

the observed GFP disruption (Extended Data Fig. 2a, b). Next, we explored the effect of the concentration of the plasmid that encodes CasX sgRNA on the extent of genome editing. The highest amounts of transfected PlmCasX plasmid produced GFP-locus editing at levels comparable to the levels of genome editing observed in initial reports for CRISPR–Cas9 and CRISPR–Cas12a (about 34%)[6,7,11,18] (Extended Data Fig. 2c). Additionally, we developed two distinct clonal EGFP HEK293T reporter lines (labelled 1 and 2) using lentiviral integration. PlmCasX-mediated EGFP disruption in these individual cell lines was comparable to results in the destabilized-GFP reporter cell line, with guide 3 producing higher levels of editing than guide 2 (Extended Data Fig. 2d). Sub-cloning PCR-amplified segments of the GFP locus from treated cells revealed a wide variety of indels, many of which map to the cut sites identified in vitro (Extended Data Fig. 2e). Finally, we compared the editing efficiency of PlmCasX to that of *Sp*Cas9, which has been optimized over the past six years for high efficiency. In most cases, guide RNAs for both Cas9 and CasX recognize the EGFP locus at similar locations but with offsets to accommodate differences in PAM specificity (Extended Data Fig. 2f). By using an optimized transfection protocol, we observe that 9 out of 10 CasX guide RNAs disrupt GFP and that editing efficacy is on average about 30% that of *Sp*Cas9—and in some cases individual guides edit as high as 55% of cells (guide 3) (Fig. 2i, Methods). These results demonstrate that CasX belongs to a third, distinct class of CRISPR systems that is capable of targeted genomic regulation and editing, and motivated experiments aimed at determining the structural and mechanistic basis for these activities.

CasX has a unique domain composition

To understand how CasX (all structural characterizations were performed using DpbCasX) binds to DNA, a ternary complex containing deactivated CasX (D672A/E769A/D935A), sgRNA (122 nucleotides) and a complementary DNA substrate (30 base pairs (bp)) was analysed by single-particle cryo-EM (Extended Data Fig. 1g). Three-dimensional particle classification and refinement revealed two conformational populations of the ternary complex, at resolutions of 3.7 Å (state I) and 4.2 Å (state II) (Extended Data Fig. 3). These two conformational states were also observed by cryo-EM analysis of a CasX complex containing a full R-loop (45-bp DNA substrate), refined at resolutions of 3.2 Å (state I) and 5.2 Å (state II) (Extended Data Fig. 4). With the cryo-EM maps, atomic models of CasX ternary complexes in state I and state II were built ab initio (Extended Data Fig. 5). Although structural alignment of the entire modelled polypeptide chain revealed some similarity between CasX and Cas12a (Lachnospiraceae bacterium Cpf1 (LbCpf1), RCSB Protein Data Bank (PDB) code 5XUU); *z*-score 15.1, with a root mean square deviation (r.m.s.d.) value of 5.3 Å for 671 aligned residues)[19], a more detailed analysis of the domains showed that—as expected—this similarity results from the RuvC domain and oligonucleotide-binding domain (OBD); alignment of RuvC with LbCpf1 PDB 5XUT gives a *z*-score of 13.8 and an r.m.s.d. value of 2.5 Å for 173 aligned residues, and alignment of the OBD with LbCpf1 PDB 5XH6 gives a *z*-score of 9.6 and an r.m.s.d. value of 3.5 Å for 153 aligned residues[10,20]. Although CasX contains additional structural elements that appear analogous to those identified in other Cas proteins (including the helical I and helical II domains, and the REC1 and REC2 domains), these domains have highly distinct folds (Extended Data Fig. 6a, b).

Two domains were identified adjacent to the separated DNA strands in the CasX complex, which we refer as the non-target-strand binding (NTSB, residues 101–191 (red in Fig. 3a, b)) and the target-strand loading (TSL, residues 825–934 (pink in Fig. 3a, b)) domains. The NTSB domain contains a four-stranded β-sheet and sits next to the non-target

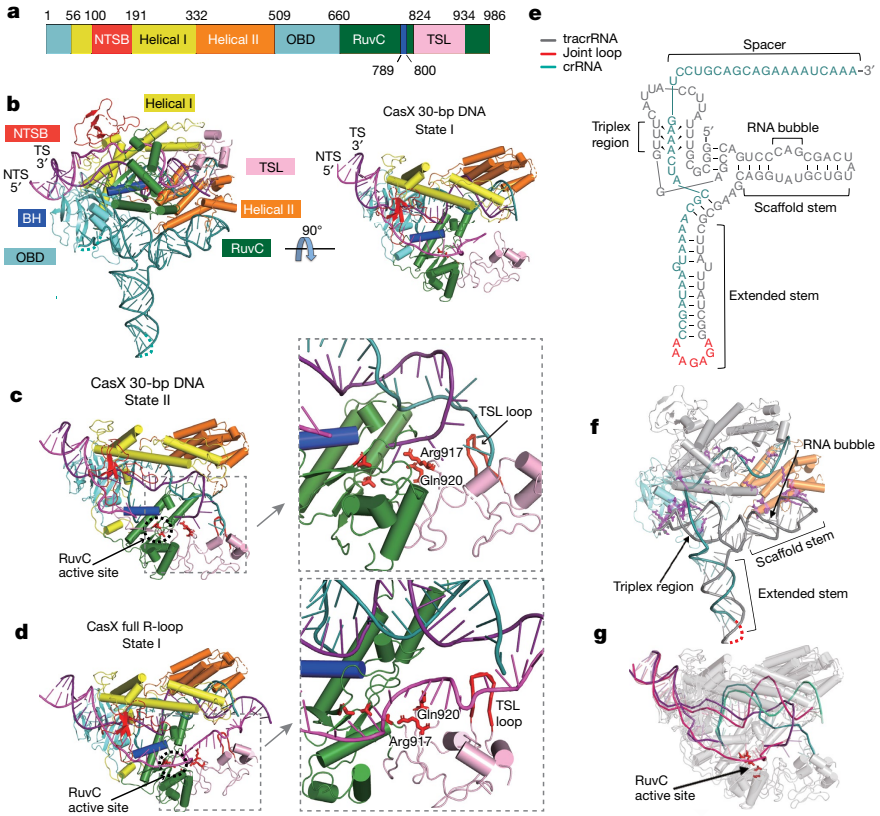


Fig. 3 | Overall structure of the CasX ternary complex. a, Domain composition of CasX. CasX contains: NTSB (red), helical I (yellow), helical II (orange), OBD (aquamarine), RuvC (green) and TSL (pink) domains, and a bridge helix (blue). **b**, Model of CasX ternary complex with 30-bp target DNA in state I, shown in side and top views. Domains are coloured as in **a**, and sgRNA is in teal. For the target DNA, the NTS is in magenta and the TS is in purple. **c**, **d**, Models of the CasX ternary complex with 30-bp target DNA in state II and state I, shown in top view. Residues Arg917 and Gln920 are shown as red sticks. The TSL loop is shown as a red ribbon. The positions of the RuvC active-site residues are shown as red sticks to illustrate the distance to the active site from the elements of the TSL domain. Right panels show zoomed-in views of the TSL domain. **e**, Schematic of the sgRNA fold with tracrRNA sequence in grey, crRNA sequence in teal and the joint loop in red. **f**, Molecular interactions between CasX and guide RNA. Protein residues interacting with guide RNA recognition are shown as magenta sticks. Helical II and OBD are coloured in orange and aquamarine, respectively. **g**, Models of CasX ternary complex in state I and II are aligned and superimposed. CasX is shown as a transparent grey cartoon, and the residues responsible for cleavage activity are shown in red. The nucleic acids are shown as ribbons to emphasize the rotation of the RNA–DNA duplex required for the transition between the two states.

strand of the DNA (Fig. 3b). We discuss this domain and its function in depth below. The TSL domain is located in a position analogous to that of the so-called 'Nuc', a domain that was incorrectly hypothesized to be a second nuclease domain responsible for DNA cleavage, in other type V CRISPR–Cas enzymes[9,13]. We propose that Nuc and similar domains should be renamed TSL to better explain the activity that we postulate to be target-strand placement in the RuvC active site[10]. In the *Acidaminococcus* sp. Cas12a Nuc domain, amino acids Arg1226 and Asp1235 aid target-strand cleavage, and an Arg1226Ala mutation produced an *Acidaminococcus* sp. Cas12a nickase by abolishing the ability of Cas12a to cut the target strand[9]. In the TSL domain of CasX, residues Arg917 and Gln920 interact with DNA (non-target strand in state I and target strand in state II) that is adjacent to the active site (Fig. 3c). Within the CasX full R-loop structure in state I, a TSL loop that contains three tyrosines (Tyr867, Tyr868 and Tyr870) and three positively charged residues (Arg869, Lys871 and Arg872) interacts with the migration point at which the RNA–DNA duplex ends and the DNA–DNA duplex reforms (Fig. 3d, Extended Data Fig. 6c). In other enzymes, similar loops or hairpin elements that contain a large hydrophobic amino acid (tyrosine or phenylalanine) are thought to be involved in DNA strand separation[21–23]. Moreover, the TSL domain also contains two CXXC motifs (residues 824–827 and 926–929) that form a tetra-Cys zinc-finger or zinc ribbon motif (Extended Data Fig. 6c, d), akin to those found in phage primases, transcription factors and the purported transposase ancestor for type V CRISPR proteins TnpB5[5,24–26]. Together, it is clear that CasX possesses domains analogous to other CRISPR–Cas proteins as well as completely novel domains, as anticipated by its complete sequence dissimilarity from other CRISPR proteins.

A prominent guide RNA scaffold for CasX

Structural modelling shows that the guide RNA accounts for about 26% of the mass in the CasX–sgRNA binary complex, which is a value that is greater than those observed for other type II or type V CRISPR–Cas effector complexes (about 8% in LbCas12a, about 20% in *Alicyclobacillus acidoterrestris* Cas12b and about 16% in *Sp*Cas9; Extended Data Fig. 6e). Dominating the CasX protein complex, the sgRNA includes three elements: a triplex stem–loop that contacts the OBD, a 'scaffold' stem that interacts with the helical II domain and a perpendicular stem–loop that projects away from the centre of mass of the structure (Fig. 3e, f, Supplementary Video 1). Alterations to the triplex or the scaffold stem diminished CasX activity in vivo, whereas truncated versions of the perpendicular stem–loop retained activity (Extended Data Fig. 6f).

In the absence of guide RNA, the CasX protein is poorly resolved by cryo-EM (Extended Data Fig. 7a, d). Consistent with the importance of RNA in CasX protein architecture, CasX cross-linking before and after addition of the sgRNA followed by analytical mass spectrometry revealed large RNA-induced CasX domain rearrangements (Extended Data Fig. 7e, g). In line with this analysis, a cryo-EM-derived model of the CasX–sgRNA complex (a map at about 7.5 Å resolution) shows the OBD, RuvC and helical II domains assembled along the RNA scaffold,

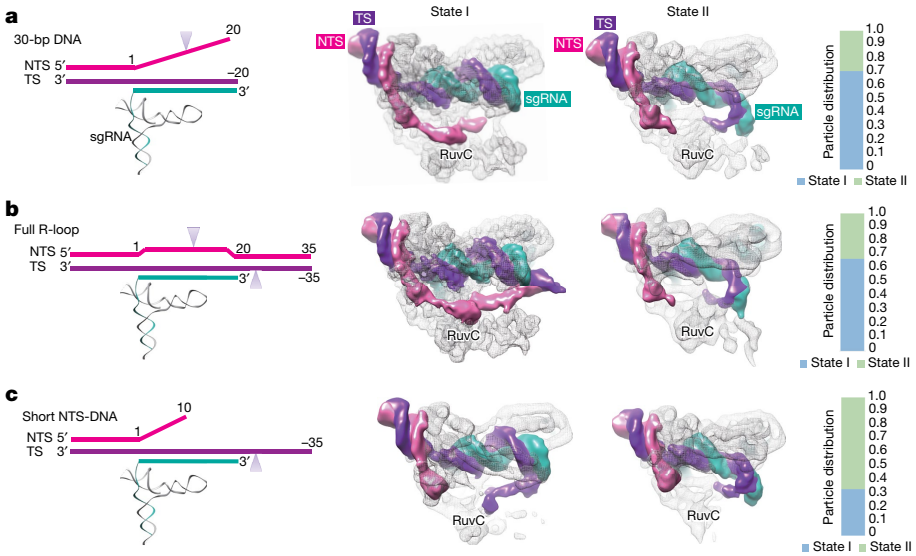


Fig. 4 | Distinct CasX conformational states. a–c, Conformational states with 30-bp target DNA (**a**), a DNA target forming the full R-loop (**b**) and the short NTS (20 nucleotides) and the 45-nucleotide TS (**c**). The schematic of the DNA probe used for each data collection is shown on the left, with cleavage sites shown by arrowheads. The top views of the cryo-EM maps for the CasX ternary complex in states I and II are shown in the centre panels. The TS density is coloured purple, the NTS is coloured magenta and the sgRNA density is coloured teal. The RuvC domain is indicated in each map. All the electron microscopy maps are low-pass-filtered to 4.5 Å. The relative percentages of particles belonging to each state revealed by cryo-EM analysis are shown in the right panels.

whereas the NTSB domain associates with the RuvC, helical I and helical II domains near the nuclease active site (Extended Data Fig. 6 b, f). Comparison with the DNA-bound structure shows that, upon DNA binding, the NTSB domain moves away from the centre of mass of the complex (Extended Data Fig. 7c, h, i).

Two conformations enable CasX dsDNA cleavage

Comparison of the two conformational states (states I and II) of the CasX ternary complex revealed a large structural change that alters target DNA strand accessibility to the RuvC domain (Fig. 3c, d, Supplementary Video 2). In state I, the non-target-strand DNA sits in the RuvC active site while the target-strand DNA–guide RNA duplex engages with the helical I and helical II domains (Fig. 3b, d). In state II, the target-strand DNA–guide RNA duplex is sharply bent, which allows RuvC access to the target-strand DNA (Fig. 3c). State I is compatible with non-target-strand DNA cleavage, whereas state II is compatible with cleavage of the target-strand DNA (Fig. 3g, Extended Data Fig. 7j, k).

Statistical analysis by single-particle sorting showed that the majority of particles (about 71%) in the 30-bp target DNA ternary complex adopted the state I conformation, and that the remaining 29% of particles were in state II (Fig. 4a, Extended Data Fig. 3). This preference suggests that non-target-strand DNA is cleaved by the RuvC domain first, followed by displacement and target-strand cleavage. Similar to the 30-bp DNA-containing sample, 67% of full R-loop (45-bp) DNA particles adopted state I (Fig. 4b, Extended Data Fig. 4).

In our sequential model of CasX-mediated DNA cleavage, a substrate-bound complex mimicking the intermediate state that occurs after non-target-strand cleavage should preferentially adopt state II. To test this idea, we performed cryo-EM analysis on a CasX ternary complex that contained sgRNA and a DNA substrate comprising a 45-nucleotide target strand and a post-cleavage-like 20-nucleotide-non-target strand (Extended Data Fig. 8). In this intermediate-state sample, most particles (about 66.4%) adopted the state II conformation, and the target strand was located near the RuvC active site (Fig. 4c). Reconstruction of state I showed the 5′-end overhang of the target-strand DNA folded back into the RuvC domain. This conformation is incompatible with dsDNA cleavage at position 22 and is unlikely to occur natively (Figs. 1d, 4c).

The CasX NTSB domain is required for DNA unwinding

The distinct and smaller architecture of CasX relative to other dsDNA-targeting CRISPR enzymes implies a unique mechanism of substrate recognition, which requires guide RNA strand invasion into duplex DNA. The observation that the NTSB domain interacts directly with non-target DNA strand both in state I and state II (Fig. 5a)

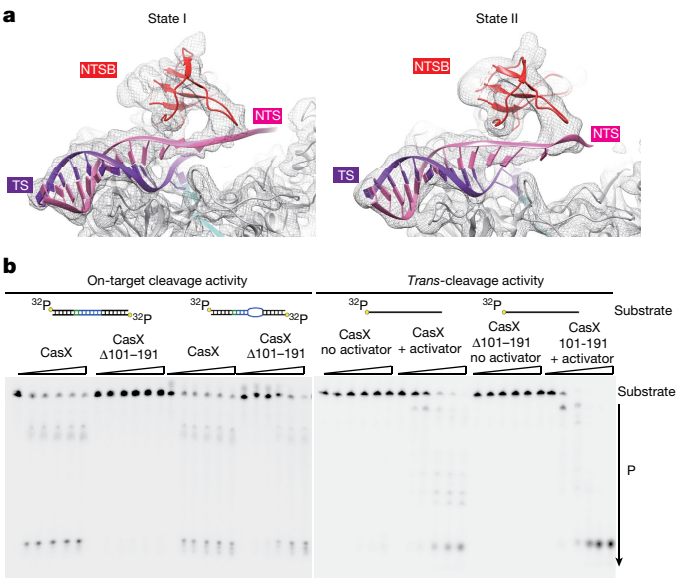


Fig. 5 | Novel domains for target DNA unwinding and loading. **a**, Electron density map showing the presence of a domain that directly interacts with the NTS, with models for the CasX ternary complex in state I and II within the cryo-EM map (shown as mesh surface, low-pass-filtered to 4.5 Å). CasX is shown in grey with the NTSB domain highlighted in red, the TS in purple and the NTS in magenta. **b**, Comparison of the cleavage activity of the wild-type CasX and NTSB domain deletion (CasXΔ101–191). The reactions were analysed at time points from 0 to 120 min. Completely base-paired probe and a bubbled probe were used to test the on-target activity, and a random 50-nucleotide oligonucleotide was used to test the *trans*-cleavage activity. P indicates the cleavage product. Three biological repeats for the assays showed consistent results.

raised the possibility that this structure contributes fundamentally to the mechanism of DNA unwinding. To test this hypothesis, we analysed the behaviour and activity of a protein construct that lacks the NTSB domain (CasXΔ101–191). Although it showed similar physical behaviour to that observed for the wild-type CasX on a size-exclusion column (Extended Data Fig. 9a), CasXΔ101–191 was incapable of cleaving a dsDNA substrate (Fig. 5b). Nevertheless, CasXΔ101–191 retains robust targeted ssDNA cleavage activity, including on mismatched duplex DNA substrates (Fig. 5b, Extended Data Fig. 9b). Of note, CasXΔ101–191 has slightly higher *trans*-cleavage activity than does wild type, suggesting that the NTSB domain may contribute to shielding the RuvC from accessing the *trans*-DNA substrates (Fig. 5b). Together, these results suggest that the NTSB domain is responsible for initiating and/or stabilizing DNA duplex unwinding by CasX. This finding also hints at the interesting possibility that the self-contained NTSB domain could be introduced into or acquired by other enzymes to assist with or stabilize dsDNA binding.

Conclusions

On the basis of our functional and structural data, we propose a model of CasX activation and DNA cleavage that includes the following steps: (1) guide RNA binding-induced CasX structural stabilization and DNA search; (2) NTSB-assisted DNA unwinding, R-loop formation and non-target-strand loading into the RuvC active site; (3) RNA–DNA hybrid duplex bending with the aid of the proposed TSL domain to position the target DNA strand for cleavage; and (4) product release after the cleavage of both DNA strands (Extended Data Fig. 10). Two distinct target DNA-bound states indicate that CasX coordinates sequential dsDNA cleavage by its single RuvC nuclease using the zinc-finger-containing TSL domain (Extended Data Fig. 10c, d). The TSL domain appears to confer a convergent mechanism of acute target-strand DNA bending that is central to all type V single-nuclease CRISPR–Cas enzymes.

These functional insights will enable the continued development of CasX as a third platform for RNA-programmed genome editing. The compact size, dominant RNA content and minimal *trans*-cleavage activity of CasX differentiate this enzyme family from Cas9 and Cas12a, and provide opportunities for therapeutic delivery and safety that may offer important advantages relative to existing genome-editing technologies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-0908-x.

Received: 31 July 2018; Accepted: 31 December 2018;
Published online: 04 February 2019

1. Marraffini, L. A. CRISPR–Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
2. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
3. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
4. Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A. & Doudna, J. A. RNA-dependent RNA targeting by CRISPR–Cas9. *eLife* **7**, e32724 (2018).
5. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
6. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
7. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
8. Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
9. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
10. Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR–Cas endonuclease. *Cell* **167**, 1814–1828.e1812 (2016).
11. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
12. Chen, J. S. et al. CRISPR–Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
13. Swarts, D. & Jinek, M. Mechanistic insights into the *cis*-and *trans*-acting deoxyribonuclease activities of Cas12a. Preprint at https://www.biorxiv.org/content/early/2018/06/22/353748 (2018).
14. Oakes, B. L., Nadler, D. C. & Savage, D. F. Protein engineering of Cas9 for enhanced function. *Methods Enzymol.* **546**, 491–511 (2014).
15. Oakes, B. L. et al. Profiling of engineering hotspots identifies an allosteric CRISPR–Cas9 switch. *Nat. Biotechnol.* **34**, 646–651 (2016).
16. O'Connell, M. R. et al. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263–266 (2014).
17. Zhu, X. et al. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.* **4**, 6420 (2014).
18. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
19. Yamano, T. et al. Structural basis for the canonical and non-canonical PAM recognition by CRISPR–Cpf1. *Mol. Cell* **67**, 633–645.e633 (2017).
20. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
21. Moolenaar, G. F., Höglund, L. & Goosen, N. Clue to damage recognition by UvrB: residues in the β-hairpin structure prevent binding to non-damaged DNA. *EMBO J.* **20**, 6140–6149 (2001).
22. Shen, J., Gai, D., Patrick, A., Greenleaf, W. B. & Chen, X. S. The roles of the residues on the channel β-hairpin and loop structures of simian virus 40 hexameric helicase. *Proc. Natl Acad. Sci. USA* **102**, 11248–11253 (2005).
23. Castella, S., Bingham, G. & Sanders, C. M. Common determinants in DNA melting and helicase-catalysed DNA unwinding by papillomavirus replication protein E1. *Nucleic Acids Res.* **34**, 3008–3019 (2006).
24. Hahn, S. & Roberts, S. The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. *Genes Dev.* **14**, 719–730 (2000).
25. Okuda, M. et al. A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. *J. Biol. Chem.* **279**, 51395–51403 (2004).
26. Pan, H. & Wigley, D. B. Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase. *Structure* **8**, 231–239 (2000).

Acknowledgements Electron microscopy data were collected at the Bay Area Cryo-EM (BACEM) facility located at UC Berkeley. We thank D. B. Toso and P. Grob for expert electron microscopy assistance; A. Chintangal and P. Tobias for computational support; and D. Savage, J. Cofsky and A. V. Wright for comments on the manuscript. Research in this publication was supported by the National Science Foundation under award number 1244557 (J.A.D.); by the National Institutes of Health under award number P50GM082250 (HARC Center, J.A.D.); and by the National Institute of General Medical Sciences of the National Institutes of Health under award number P01GM051487 (J.A.D. and E.N.). J.A.D. and E.N. are Howard Hughes Medical Institute Investigators.

Author contributions J.-J.L., N.O., B.L.O., E.N. and J.A.D. designed the experiments. J-.J.L. and N.O. prepared the CasX and RNP complexes. N.O., J.-J.L., E.M., J.C., L.B.H., A.W., G.J.K., J.B. and J.D. carried out the biochemical assays. B.L.O., H.B.S., K.L.M.B., B.T.S. and K.L.T. performed the in vivo experiments. B.A.-S. did the phylogenetic analysis. J.-J.L. did the cryo-EM analysis. J.-J.L. and D.T. did the cross-linking mass spectrometry analysis. J.-J.L and N.O. built the atomic structures. J.-J.L., N.O., B.L.O., E.N. and J.A.D. wrote the manuscript.

Competing interests J.A.D., B.L.O., L.B.H., J.-J.L. and N.O. have filed a related patent on CasX mutations and guide RNAs described herein with the US Patent and Trademark Office. J.A.D. is a co-founder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics and Mammoth Biosciences, and a Director of Johnson & Johnson. J.A.D. is a scientific advisor to Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Synthego, Metagenomi, Mammoth Biosciences and Inari. J.A.D. has research projects sponsored by Biogen and Pfizer. L.B.H. is a co-founder of Mammoth Biosciences. B.L.O. is a co-founder of Scribe Therapeutics.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-0908-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-0908-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.N. or J.A.D.
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 and investigators were not blinded to allocation during experiments and outcome assessment.

Strains and media. The in vivo CRISPRi[27] and cleavage assays described below use *E. coli* MG1655 containing genomically integrated and constitutively expressed GFP and red fluorescent protein (RFP). Standard cloning techniques were used to create all plasmids. Plasmid construction and retention was ensured with AmpR and CmR as selectable markers[15]. EZ-rich defined growth medium (EZ-RDM, Teknova) was used in all CRISPRi assay fluorescent measurements; 2×YT (LB) with the addition of 1.5% Bacto Agar (BD) was used for all plating assays.

***E. coli* assays.** CRISPRi assays were performed in a similar manner to previous work[15]. In brief, to test the ability of CasX to bind genomic DNA and repress transcription, electrocompetent *E. coli* were co-transformed with a plasmid encoding the guide RNA and a plasmid encoding the CasX protein as described. They were grown on medium containing two antibiotics to ensure selection for both plasmids. Colonies were picked in triplicate from these plates into EZ-RDM liquid medium and grown for 12 h. These saturated cultures were diluted 1:1,000 into EZ-RDM medium containing 2 nM anhydrotetracycline inducer and 150 μl of this mixture was followed for OD_{600} and GFP (a.u.) via a 96-well microplate reader (Tecan m1000) every 10 min over the course of 12 h at 37 °C, unless otherwise noted.

To perform the bacterial genome targeting assay, 100 ng of the CRISPR–Cas protein-encoding plasmid was electroporated into electrocompetent MG1655 *E. coli* expressing the GFP-targeting sgRNA plasmid using a BTX Harvard apparatus ECM 630 High Throughput Electroporation System, in biological triplicate. The guide sequence was moved onto the protein-encoding plasmid and 200 ng of this was used in the transformation. The cells were recovered for one hour in 300 μl SOC medium at 37 °C, unless otherwise noted. Two technical replicates of tenfold serial dilutions were spotted onto plates containing antibiotics for the plasmid(s) used in the transformation. These grew at either 37 °C for 12 h or 30 °C for 16 h and were used to calculate CFU per ml.

Human-cell GFP disruption. The EGFP reporter construct was created in a modified lentivirus backbone with EF1-alpha promoter driving the EGFP gene of interest, and a second PGK promoter driving production of hygromycin. Transduced 293T cells, validated and mycoplasma-free (tested by the UC Berkeley Cell Culture Facility), were selected with hygromycin (used at 250 μg/ml). EGFP clones were isolated by sorting single cells into 96-well plates and characterized by intensity of EGFP. Lentivirus was produced by PEI (Polysciences, 24765) transfection of 293T cells with gene delivery vector co-transfected with packaging vectors pspax2 and pMD2.G, essentially as previously described[28]. HEK293T EGFP experiments were conducted in a similar manner to previous assays. In brief, the EGFP HEK293T reporter cells were seeded into 96-well plates and transfected according to the manufacturer's protocol with lipofectamine 2000 (Life Technologies) and the described amount of plasmid DNA encoding the CasX, sgRNA and CasX, P2A–puromycin fusion. The next day cells were selected with 1.5 μg/ml puromycin for 3 days and analysed by fluorescence-activated cell sorting 8 days after selection to allow for clearance of EGFP protein from the cells. Cells were passaged once to maintain sub-confluent conditions. EGFP expression was traced using an Attune NxT Flow Cytometer and high-throughput autosampler. For extended assays, cells were passaged 1:10 and reanalysed on the date notes. The PlmCasX versus *Sp*Cas9 EGFP line 1 disruption assays were done as described above, but the protocol was improved by using lipofectamine 3000 instead of 2000, according to the manufacturer's protocols, selecting with 1.5 μg of puromycin for 48 h and analysing 11 days later.

T7EI assay. T7EI assays were performed as previously described with slight modification[15]. In brief, cells were suspended 1:1 in QuickExtract (lucigen) buffer and DNA was extracted using the manufacturer's protocol. This mixture was used directly in a PCR reaction designed to amplify the GFP locus and ~200 ng of PCR product was used for denaturing, annealing and digestion with T7E1 (NEB), according to the manufacturer's protocol. Samples were analysed on a 2% agarose gel with SYBRsafe (Thermo Fisher).

EGFP fragment sub-cloning and sequencing. The edited GFP locus was amplified via PCR from HEK cells treated with Quick Extract buffer. The PCR fragments were then T4 blunt-ligated into a small ColEI plasmid that was digested with PmeI and treated with CIP. Ligation products were transformed into cells and colonies were grown overnight before being picked, Miniprepped (Qiagen) and Sanger-sequenced at the UC Berkeley DNA Sequencing Facility. Sequencing results were aligned to the target and visualized via Snapgene.

Protein expression, purification and complex reconstitution. The gene encoding CasX was sub-cloned into the 2CT-10 expression vector. CasX(D672A/E769A/D935A) and CasXΔ101–191 were obtained by amplifying the CasX plasmid using mutagenetic PCR primers. All the proteins were expressed using Rosetta *E. coli* cells. Main culture (Terrific broth, containing 100 mg/l ampicillin) was inoculated with 3% of overnight culture grown in Luria broth. The main culture was grown to an OD of 0.5–0.6, cooled down and protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, and expression was allowed to proceed overnight at 16 °C. Cells were collected, re-suspended in Ni buffer A (500 mM sodium chloride, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP) and frozen at −80 °C For wild-type CasX protein preparation, cells were thawed, diluted twice with Ni buffer A, followed by addition of PMSF (final concentration 0.5 mM), and 3 tablets of Roche protease inhibitor cocktail per 100 ml of cell suspension. Cells were lysed by sonication and pelleted at 35,000*g* for 30 min. Clarified lysate was purified using Ni-NTA agarose beads, using step-gradient elution with imidazole-containing buffer (Ni buffer B (highest imidazole concentration): 500 mM sodium chloride, 500 mM imidazole, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP). The pure fractions were pooled and TEV protease was added (1 mg protease/20 mg purified protein in final concentration). The protein with TEV protease was dialysed overnight against the following buffer: 500 mM sodium chloride, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP. Then, protein was applied to a maltose binding protein (MBP) column and the MBP flow-through was applied to a heparin column. Protein was eluted from the heparin column using a sodium chloride gradient up to 1 M sodium chloride. For the wild-type protein, there were two peaks containing CasX. The peak that eluted at lower salt concentration was found to contain inactive and aggregated protein and was not pooled; only the second peak contained active protein and only that protein was used for the assays. The active protein from the heparin column was concentrated and applied to a Superdex200 10/300 column in the following buffer: 500 mM potassium chloride, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP. Pure protein was concentrated and flash-frozen. CasXΔ101–191 purification was purified as the same way as wild-type CasX. The overall expression yield was similar, but the amount of the well-folded protein (second peak) was lower than in case of wild-type protein. For CasX(D672A/E769A/D935A), the purification was similar, except that dialysis buffer was: 300 mM sodium chloride, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP and size-exclusion buffer was 300 mM potassium chloride, 50 mM HEPES, pH 7.5, 10% glycerol, 0.5 mM TCEP, and all the protein eluted as a single well-folded protein peak on heparin column.

sgRNA was in vitro-transcribed using T7 RNA polymerase and purified using 10% urea–PAGE. The in vitro transcription template was GAAATTAATACGAC TCACTATAggCGCGTTTATTCCATTACTTTGGAGCCAGTCCCAGCGACT ATGTCGTATGGACGAAGCGCTTATTTATCGGAGAGAAACCGATAAGTA AAACGCATCAAAGTCCTGCAGCAGAAAATCAAA.

The CasX–sgRNA complex was assembled by incubating protein with 1.6-fold stoichiometric excess of sgRNA for 30 min at room temperature. The ternary complexes were assembled by incubating CasX–sgRNA with 1.8-fold stoichiometric excess of annealed DNA target for 30 min at room temperature. After the complexes were assembled, they were purified by size-exclusion chromatography using Superdex200 10/300 column.

DNA cleavage assays. DNA substrates were 5′-end-labelled with T4 PNK (NEB) in the presence of gamma ^{32}P-ATP. Unless otherwise noted, the following conditions were used: proteins were diluted to 4 μM with dilution buffer (500 mM NaCl, 10% glycerol, 20 mM Tris-HCl, pH 7.5, 1 mM magnesium chloride, 0.5 mM TCEP). sgRNA was diluted to 6 μM with reaction buffer (20 mM HEPES, pH 7.5, 10 mM magnesium chloride, 150 mM potassium chloride, 1% glycerol, 0.5 mM TCEP). Resulting stocks of protein and sgRNA were mixed in 1:1 molar ratio and incubated for 10 min at room temperature to produce active complex. Cleavage reactions were conducted in 1× reaction buffer; the radiolabelled probe concentration was 2 nM. Reactions were initiated by addition of CasX–sgRNA to a final concentration of 200 nM. The reactions were conducted at 37 °C, and aliquots were taken at the following time points: 0, 2, 5, 30, 60 and 120 min. The aliquots were immediately mixed with formamide loading buffer (final concentration 45% formamide and 50 mM EDTA, with trace amount of bromophenol blue) and heated for 10 min at 90 °C for quenching. Samples were separated by 10% or 12% urea–PAGE, gels were dried and the results were visualized using a phosphoimager (Amersham Typhoon (GE Healthcare)).

In the cleavage assays used to determine the DNA cut sites (Fig. 1b, c), the following concentrations were used: 100 nM Cas protein and 120 nM guide RNA. In the experiment in which *trans*-cleavage activity was compared between different CRISPR–Cas proteins (Fig. 1e) the following concentrations were used: 100 nM Cas proteins, 120 nM guide RNA, 150 nM activator and M13mp18 ssDNA (New England Biolabs). In the experiments in which *trans*-cleavage activity was compared between CasX and CasXΔ101–191, a random 50-nucleotide oligonucleotide substrate was used. All the nucleotide sequences and plasmids used in this study have been summarized in Supplementary Table 2.

Electron microscopy sample preparation and data collection. CasX complexes in a buffer containing 20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM DTT and 0.25% glycerol were used for cryo-EM sample preparation. Immediately after glow-discharging the grid for 14 s using a Solaris plasma cleaner, 3.6-μl droplets of the sample (~3 μM) were placed onto C-flat grids with 2-μm holes and

2-μm spacing between holes (Protochips). The grids were rapidly plunged into liquid ethane using a FEI Vitrobot MarkIV maintained at 8 °C and 100% humidity, after being blotted for 4 s with a blot force of 8. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 keV with a GIF energy filter, at a nominal magnification of 135,000× (0.9-Å pixel size) for ternary complexes and 105,000× (1.15-Å pixel size) for binary complex, with defocus ranging from −0.5 to −2 μm. Micrographs were recorded using SerialEM on a Gatan K2 Summit direct electron detector operated in super-resolution mode[29]. We collected a 4.8-s exposure fractionated into 32, 150-ms frames with a dose of 9.58 e^{-} $Å^{-2}$ s^{-1}.

Apo-CasX in a buffer containing 20 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM DTT and 5% glycerol was used for cryo-EM sample preparation following the same sample protocol used for CasX complexes. Data were acquired using a FEI Titan Krios transmission electron microscope operated at 300 keV with energy filter and Volta phase plate, at a nominal magnification 105,000× (1.15-Å pixel size) with defocus of about −0.5μm.

Electron microscopy data analysis. For CasX binary and ternary complexes, the 28 frames (we skipped the first 2 and last 2 frames) of each image stack in super-resolution model were aligned, decimated, summed and dose-weighted using Motioncor2[30]. CTF values of the summed-micrographs were determined using Gctf[31]. Initial particle picking to generate template images was performed using EMAN2[32]. About 10,000 particles were selected and then imported into Relion2.0 for reference-free 2D classification[33]. Particle picking for the complete data set was carried out using Gautomatch (by K. Zhang, unpublished) with templates generated in the previous 2D classification. Local CTF was re-calculated by Gctf with the determined box files. Particles were extracted from the dose-weighted, summed micrographs in Relion2.0 and then imported into CryoSparc[34] for 2D classification, ab initio modelling, heterogeneous refinement, homogenous refinement and local resolution calculation.

For images obtained with a Volta phase plate, following preprocessing the CTF and phase-shift values of the summed-micrographs were determined using Gctf and then applied to dose-weighted, summed micrographs for further processing. Parameters in the electron microscopy analysis for the five CasX samples have been summarized in Supplementary Table 1.

Cross-linking and mass spectrometry. CasX samples in HEPES buffer were crossed-linked using 1 mM bis-sulfosuccinimidyl-suberate (BS3) at 30 °C for 30 min. The reactions were stopped by adding 50 mM Tris (final concentration). Cross-linked samples were then digested by trypsin and purified for mass spectrometry analysis. Cross-linked peptides were identified using an upgraded version of pLink[35]. In pLink, the parameter of cross-linker was set to BS3. The parameter of enzyme was set to trypsin with up to three missed cleavages. Precursor mass tolerance and fragment mass tolerance were both set to 20 ppm. At least six amino acids were required for each peptide chain. Carbamidomethylation on cysteine was searched as a fixed modification. Oxidation on methionine was searched as a variable modification. Search results were filtered by requiring false discovery rate $< 5\%$ at the spectral level. Further inspection of tandem mass spectrometry spectra were performed using pLabel[36].

Atomic model building. For the CasX ternary complex containing a 30-bp target DNA, the cryo-EM density of state I at 3.7 Å resolution was used for secondary structure search in PHENIX with the 'Find Helices and Strands' program[37]. The protein main chain was manually traced in Coot[38]. After main chain building, side chains were assigned manually based on the electron microscopy map in Coot and were then further improved using the cryo-EM map of state I with the full R-loop at a resolution of 3.2 Å. The DNA substrates and gRNA were manually built ab initio in Coot based on the cryo-EM density. To improve backbone geometry, the atomic model was subjected to PHENIX real-space refinement (global minimization and ADP refinement) with secondary structure, Ramachandran, rotamer and nucleic-acid restraints. The final model was validated using Molprobity[39] and cross-linking mass spectrometry data. The atomic models of state II were obtained by running flexible fitting on the state I atomic model against the state II cryo-EM map (4.2 Å resolution) with secondary structure restraints in MDFF[40]. PHENIX real-space refinement was further used to improve backbone geometry. This state II atomic model was directly adopted for structural interpretation of the CasX ternary-complex state II with full R-loop DNA and shortened non-target strand DNA.

For the CasX ternary complex containing a full R-loop DNA, the atomic model of CasX ternary-complex state I with 30-bp target DNA was fitted into the state I cryo-EM map of CasX ternary complex containing full R-loop DNA (resolution of 3.2 Å) using UCSF Chimera[41]. Additional DNA nucleotides were manually built in Coot. The atomic model was subjected to PHENIX real-space refinement against the cryo-EM map and validated using Molprobity.

For the CasX ternary complex containing a shortened non-target-strand DNA, the atomic model of CasX ternary-complex state I with 30-bp target DNA was fitted into the state I electron microscopy map of CasX ternary complex containing the shortened non-target-strand DNA (resolution 4.5 Å) using Chimera. DNA nucleotides were manually modified in Coot. The atomic model was subjected to PHENIX real-space refinement against the cryo-EM map and validated using Molprobity.

X-ray fluorescence elemental analysis. Targeted proteins at a concentration of around 1 μg/μl were mixed with 4 volumes of acetone and incubated at −20 °C for 1 h. Precipitated proteins were collected by centrifugation. The protein pellets were smashed in the buffer containing 50 mM Tris, 150 mM NaCl and 0.03% dodecyl maltoside at pH 7.5, loaded on to a nylon loop and then flash-frozen in place on a goniometer using a nitrogen stream. The samples were excited with a 14-keV X-ray beam and fluorescence spectra were collected. Elements in the sample were identified on the basis of characteristic emission energies.

RaxML maximum-likelihood phylogenetic tree. Cas12 proteins[8,11,42] were aligned with TnpB representatives using MAFFT[43]. Alignment columns were trimmed using the trimAl -gappyout method, and a maximum-likelihood phylogenetic tree was constructed using RAxML[44] with PROTGAMMALG as the substitution model and 100 bootstrap samplings. The family tree was visualized using iTOL v.3[45].

Sequence identity pairwise comparison. A non-redundant set of Cas9 orthologues (type II-A) was compiled by clustering proteins with $\geq 90\%$ identity using CD-HIT[46]. Cas9, Cas12a and CasX proteins were each aligned separately using MAFFT and the RuvC domains for each protein in the alignment were inferred from their known crystal structures. A multiple sequence alignment of the resulting RuvC domains was used to extract the per cent identity of each pair of orthologues and generate a heat map that illustrates the pairwise comparisons of the RuvC domains. Histograms of the frequency of occurrence of each identity value in pairwise comparisons of Cas12a, Cas9 and CasX were plotted from the heat map.

Reagent, software and protocol availability. All the reagents used in this work are commercialized. All the protocols have been described in detail above. All the software used in this work is noted with references, and is available for academic usage. Further requests for information regarding reagent, software and protocol availability should be addressed to the corresponding authors.

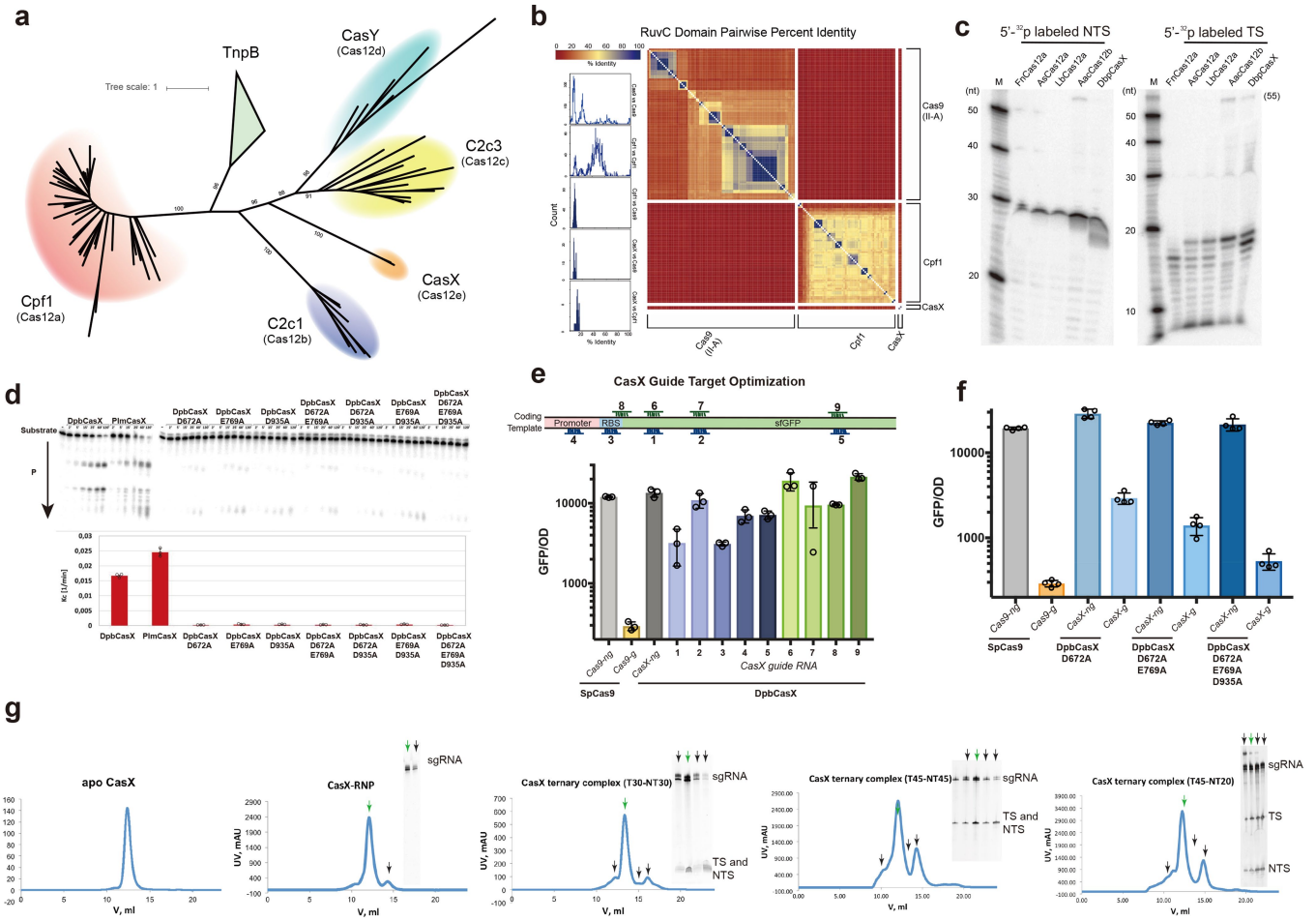
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

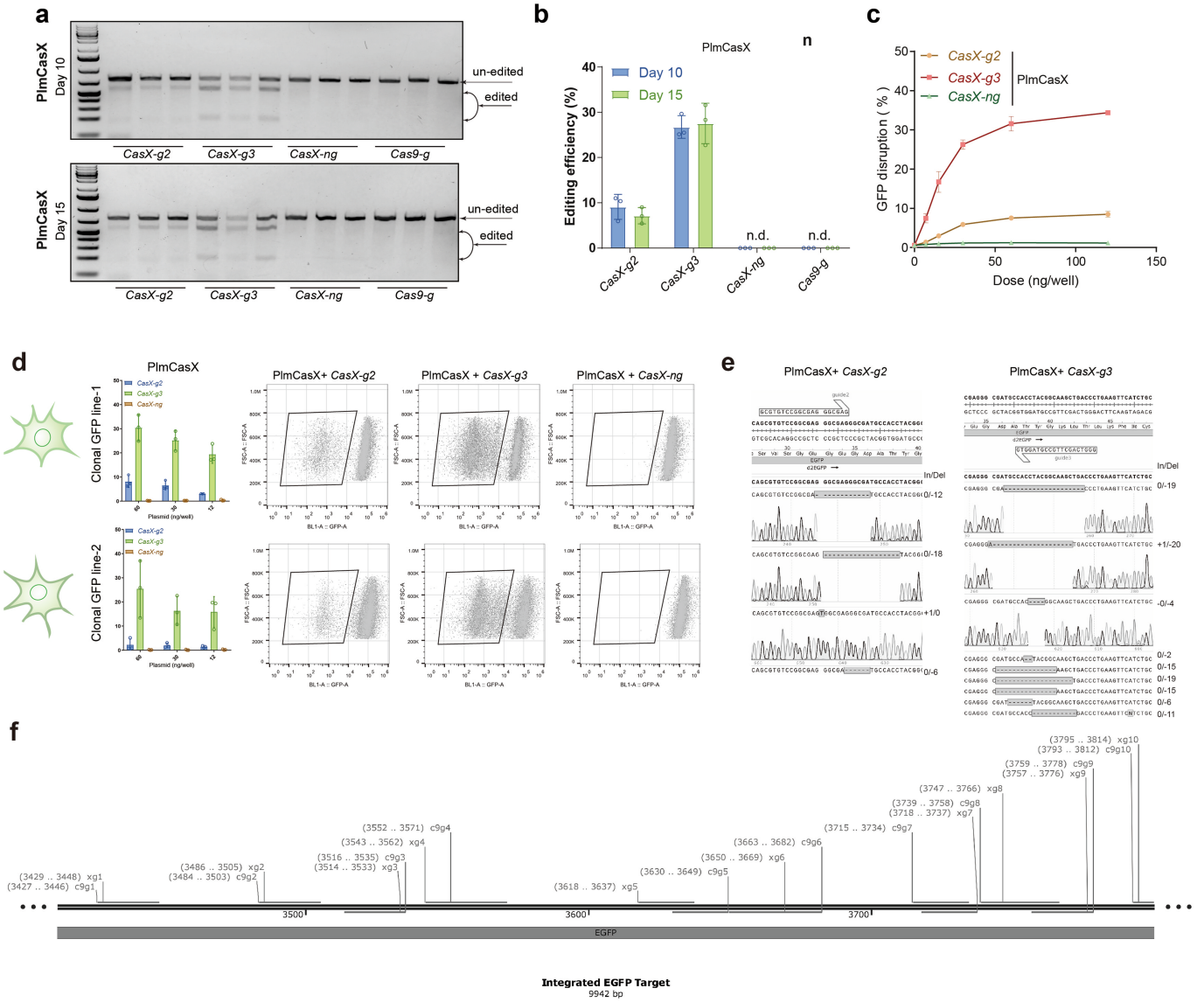
All data that support the conclusions of this Article can be found in the figures and the Source Data. The cryo-EM structural models and electron density maps have been deposited in the Protein Data Bank under the codes 6E5O, 6E7A and 6E79 and the Electron Microscopy Data Bank under the codes EMD-8987, EMD-8988, EMD-8980, EMD-8994, EMD-8996, EMD-8991, EMD-8989 and EMD-8990. More information is summarized in Supplementary Table 1. All the plasmids and oligonucleotide sequences used in this study are summarized in Supplementary Table 2. Any other relevant data are available from the corresponding authors upon reasonable request.

27. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
28. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nat. Protoc.* **1**, 241–245 (2006).
29. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
30. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
31. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
32. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
33. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
34. 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).
35. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
36. Asara, J. M., Christofk, H. R., Freimark, L. M. & Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **8**, 994–999 (2008).
37. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
39. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
40. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).

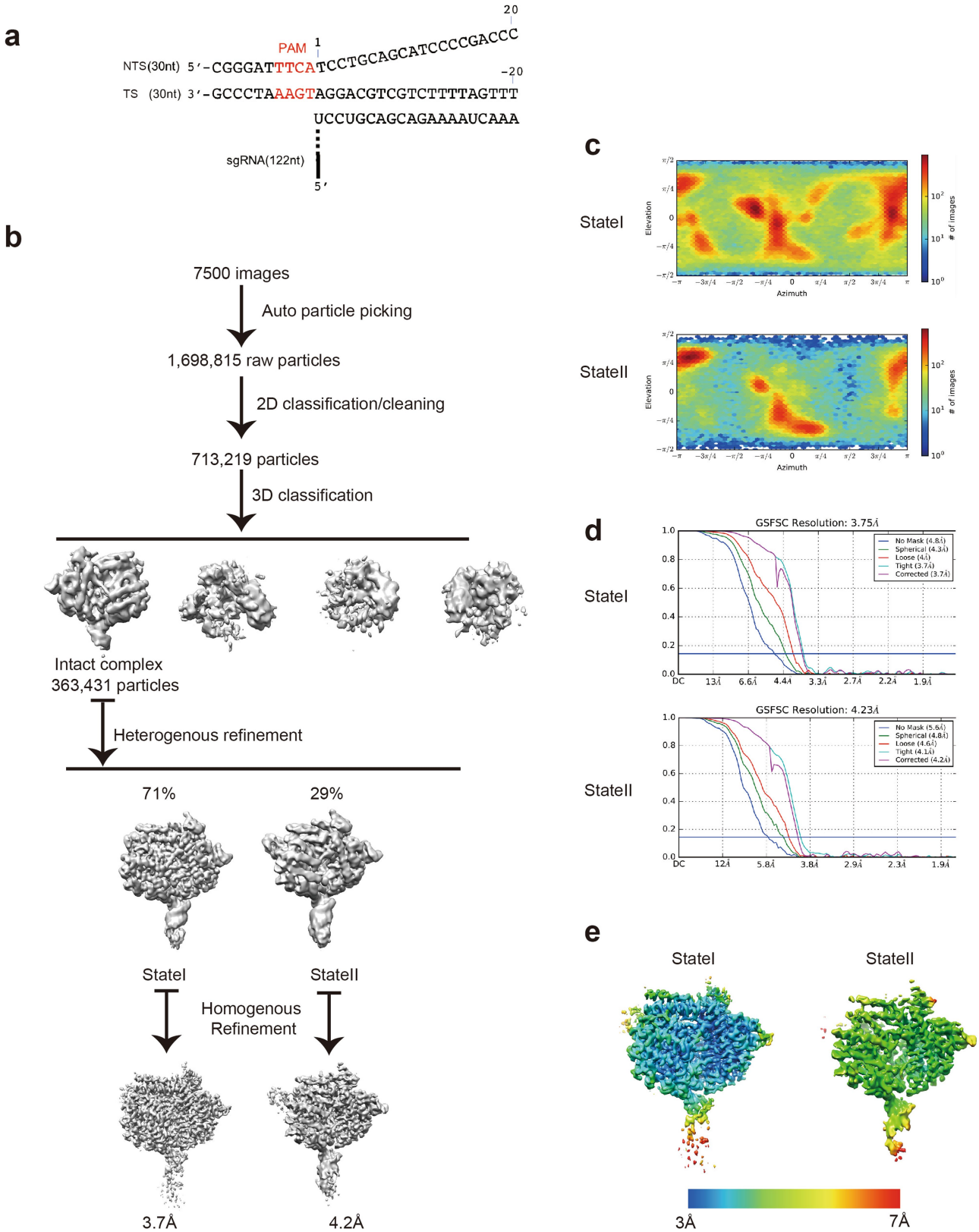
41. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
42. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell* **60**, 385–397 (2015).
43. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
44. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
45. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
46. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).



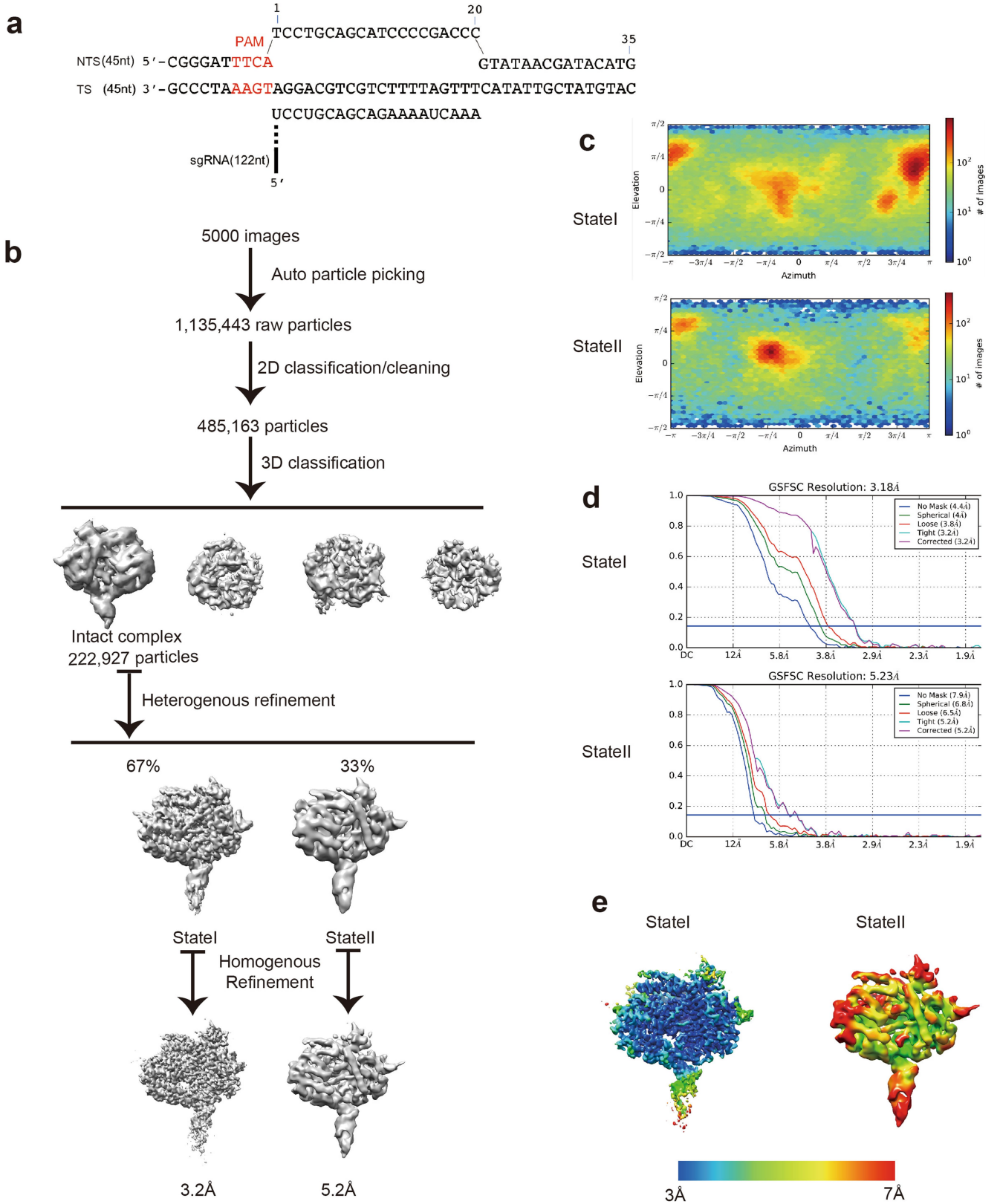
Extended Data Fig. 1 | CasX purification and substrate cleavage. **a**, RaxML maximum-likelihood phylogenetic tree of type V effector proteins with TnpB nucleases. Triangle denotes collapsed branches. Bootstrap values are indicated as percentage points; values above 88 are shown between the major branches. **b**, Per cent sequence-identity pairwise comparisons between the conserved RuvC domains of the class 2 effectors Cas9 (type II-A), Cas12a (type V-A) and CasX (type V-E) inferred from MAFFT alignment, depicted in an all-versus-all fashion. High sequence identity is shown in blue, and low sequence identity is shown in red. Histograms representing interfamily and intrafamily sequence-identity value distributions are shown along the edge. **c**, DNA cleavage site comparison among Cas12a, Cas12b and CasX. Five repeats with consistent results. **d**, DNA cleavage activity of DpbCasX mutations ($n = 3$, mean ± s.d.). **e**, Schematic of *gfp* gene. Target regions for guides 1 to 9 are marked along the gene. CasX guide screening by GFP disruption ($n > 2$, mean ± s.d.). **f**, CRISPRi efficiency for CasX active-site mutations. The Cas proteins and guide RNAs used in each assay are marked. *Cas9-ng*, non-targeting RNA guide of *Sp*Cas9; *Cas9-g*, targeting RNA guide of *Sp*Cas9; *CasX-ng*, non-targeting RNA guide of DpbCasX; *CasX-g*, targeting RNA guide of DpbCasX. GFP-disruption efficiency of targeting guide is shown by GFP signal divided by optical density at 600 nm, compared to the non-targeting guide control ($n = 4$, mean ± s.d.). **g**, Purification of apo-CasX, CasX–guide RNA binary complex and CasX–guide RNA–DNA ternary complex with three DNA designs, by size-exclusion chromatography. The representative S200 size-exclusion traces by absorbance at 280 nm are shown. Samples were taken from the labelled peaks and analysed with urea-PAGE with sybrGold. NTS, non-target strand from target DNA. TS, target strand from target DNA. All the reconstitutions have been repeated more than three times with consistent results.



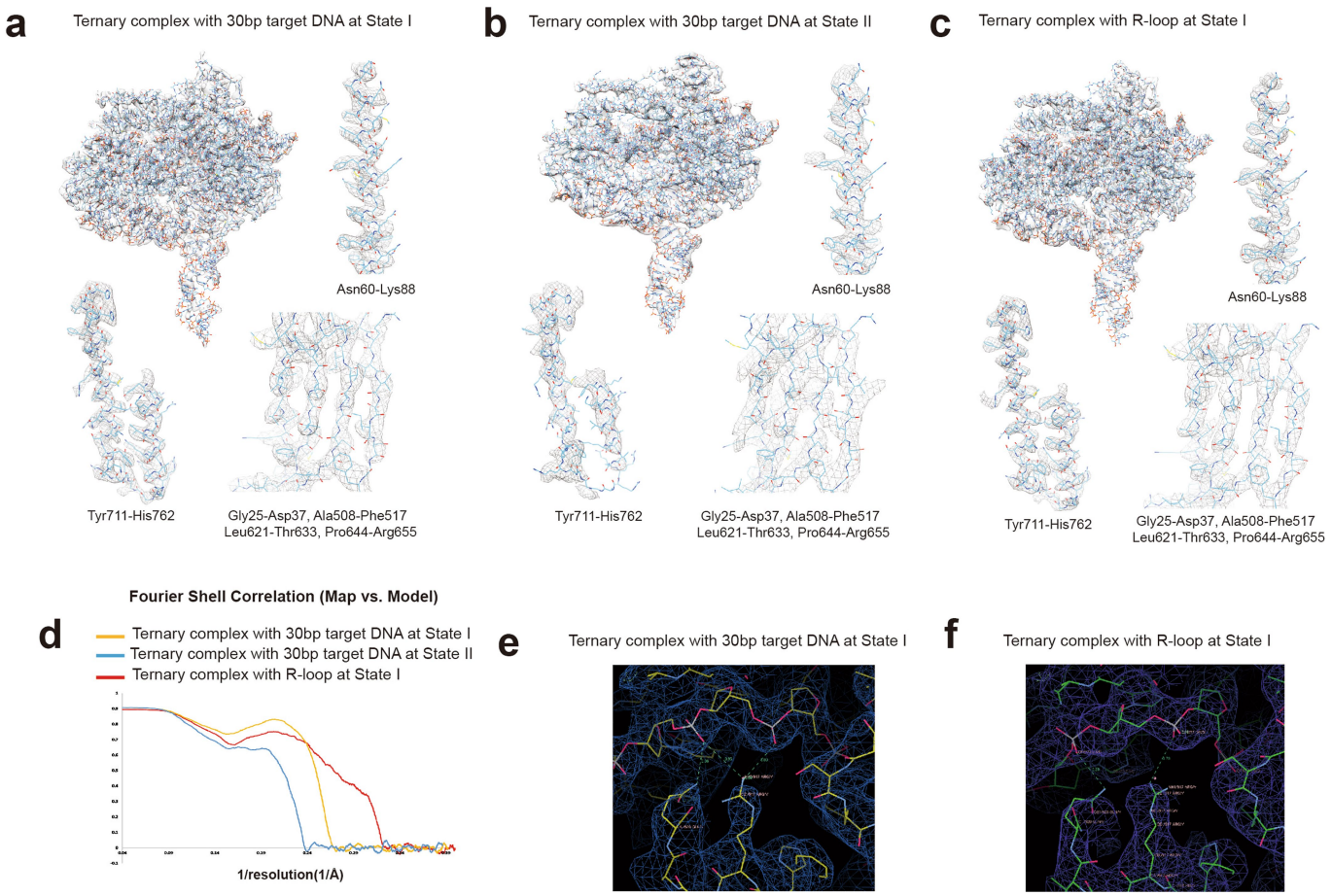
Extended Data Fig. 2 | Mammalian cell editing by CasX. a, PlmCasX T7E1 gene-editing validation of the mammalian cell GFP-disruption assay from Fig. 2g. **b**, PlmCasX T7EI quantification of **a** ($n = 3$, mean ± s.d.). **c**, PlmCasX GFP-disruption dose response ($n = 3$, mean ± s.d.). The Cas proteins and guide RNAs used in each assay are marked. *Cas9-ng*, non-targeting RNA guide of *Sp*Cas9; *Cas9-g*, targeting RNA guide of *Sp*Cas9; *CasX-ng*, non-targeting RNA guide of DpbCasX; *CasX-g*, targeting RNA guide of DpbCasX. In the assays of human cells, *CasX-g2* and *CasX-g3* denote GFP targeting guides to the template and non-template strand, respectively, and the GFP targeting guide of Cas9 (*Cas9-g*)—which is not expected to direct CasX activity—is used as the negative control. **d**, EGFP disruption of clonal EGFP HEK293T cell lines with PlmCasX and various doses of plasmid ($n = 3$, mean ± s.d.). Raw fluorescence-activated cell-sorting data are plotted with GFP on the *x* axis and forward scatter on the *y* axis, with gates drawn to demonstrate how GFP-negative cells are gated. **e**, Indels of GFP generated by PlmCasX cleavage as analysed by sub-cloning and Sanger sequencing of 20 clones. Three repeats with consistent results. **f**, Map depicting the target sites for each of the CasX and Cas9 guides on the EGFP coding sequence for Fig. 2h.



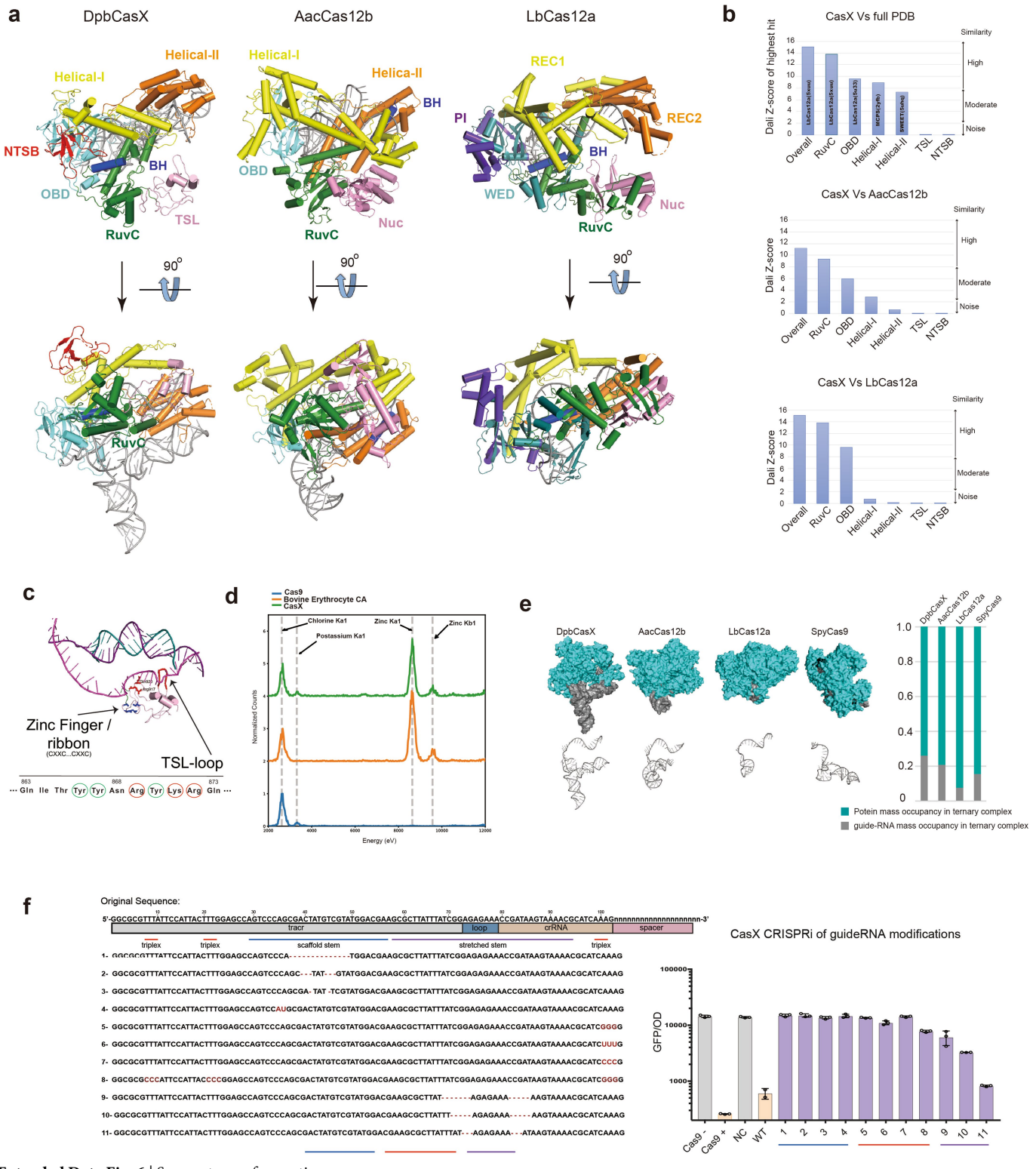
Extended Data Fig. 3 | Electron microscopy analysis of CasX–guide RNA–DNA ternary complex with a 30-bp target DNA. a, Target DNA sequence in this ternary complex. **b**, Electron microscopy analysis pipeline. From 7,500 drift-corrected micrographs, 1,698,815 particles were picked and used for 2D classification. By 2D-based manual screening, 713,219 good particles were selected for 3D classification into 4 classes. From the class that shows the most intact architecture, 363,431 particles were further used for heterogeneous refinement. This generated two reconstructions, state I and state II, which contained 71% and 29% of the particles, respectively. State I and state II were then independently refined to 3.8 Å and 4.2 Å, respectively. **c**, Euler angle distribution of the refined particles belonging to state I and state II. **d**, Gold-standard Fourier shell correlation (GSFSC) curve calculated using two independent half-maps. **e**, The density maps for both states, coloured by local resolution as calculated in Cryopsarc. The resolution ranges from 3 Å to 7 Å. **c** and **d** are taken directly from the standard output of Cryosparc.



Extended Data Fig. 4 | Electron microscopy analysis of CasX–guide RNA–DNA ternary complex with full R-loop (45-bp target DNA). **a**, Target DNA sequence in this ternary complex. **b**, Cryo-EM analysis pipeline. From 5,000 drift-corrected micrographs, 1,135,443 particles were picked and used for 2D classification. By 2D-based manual screening, 485,163 good particles were selected for 3D classification into 4 classes. From the class showing better structure preservation, 222,927 particles were further used for heterogeneous refinement; this generated two models, state I and state II, which contained 67% and 33% of the particles, respectively. State I and state II were then independently refined to 3.2 Å and 5.2 Å, respectively. **c**, The Euler angle distribution for state I and state II. **d**, GSFSC curve calculated using two independent half-maps. **e**, Cryo-EM structures of state I and state II coloured by local resolution as calculated in Cryopsarc. The resolution ranges from 3 Å to 7 Å. **c** and **d** are standard outputs of Cryosparc.

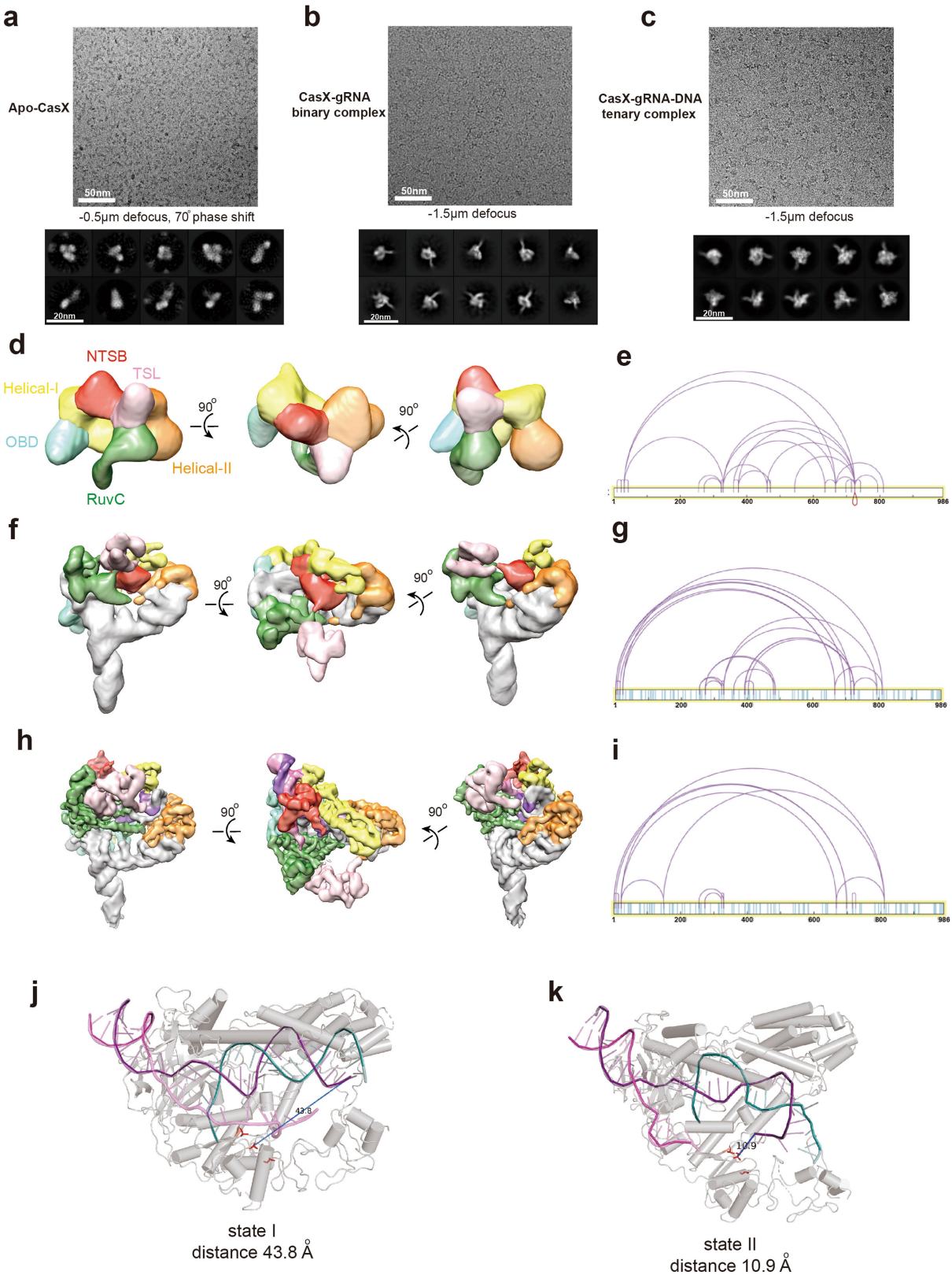


Extended Data Fig. 5 | Atomic model building of CasX ternary complexes for state I and state II. a–c, Atomic models and cryo-EM maps (shown with a threshold of 8σ or 9σ) for the CasX ternary complex with 30-bp DNA in state I (**a**) and state II (**b**), and for state I of the CasX ternary complex with full R-loop (45-bp DNA) (**c**). Representative regions of the cryo-EM density for different secondary structure regions are shown. **d**, Map against model FSCs. **e**, **f**, Zoomed-in views of atomic models fitted in electron microscopy densities. Arg917 or Gln920 and the DNA residues within 4 Å distance are linked by dashed lines.

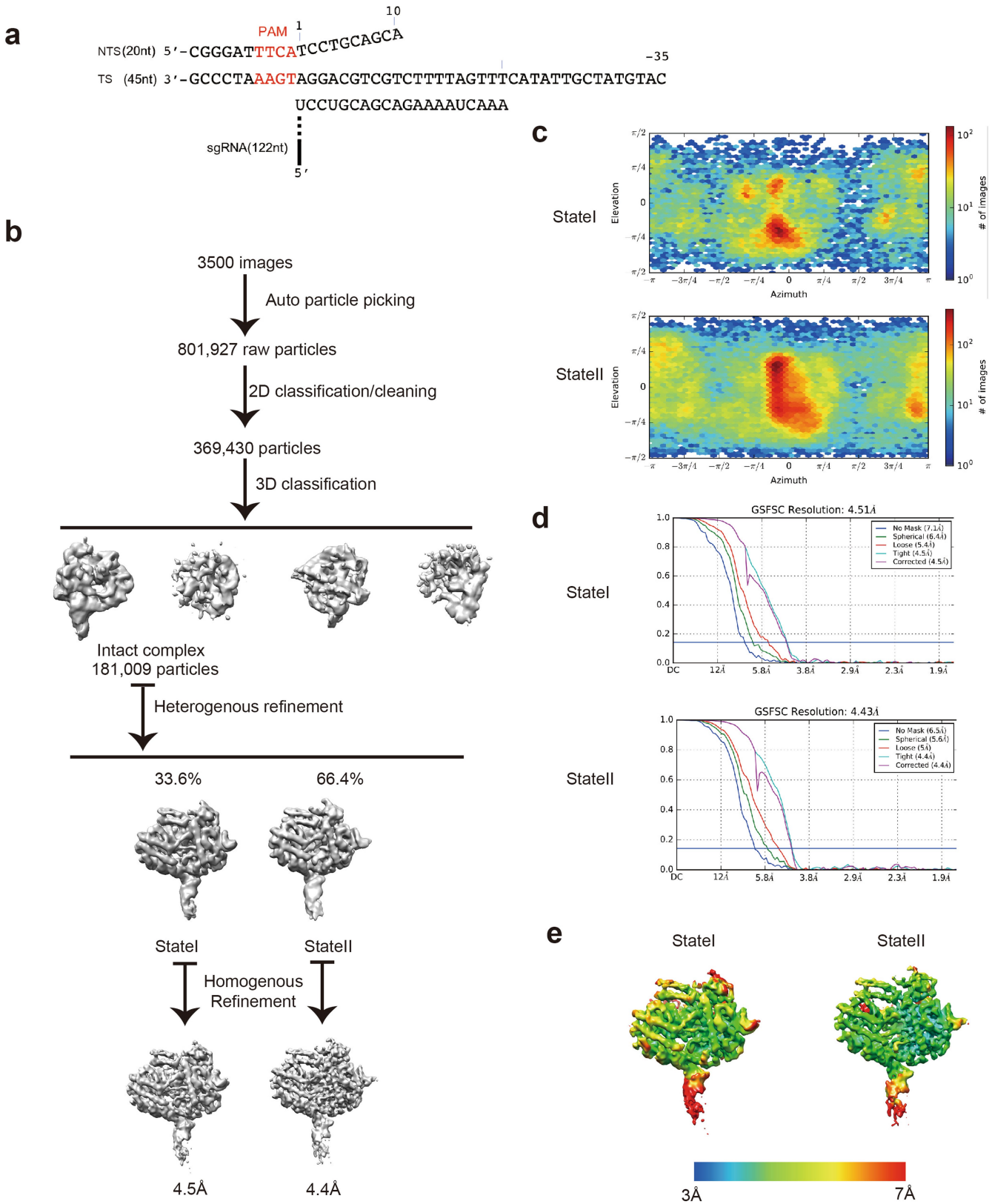


Extended Data Fig. 6 | See next page for caption.

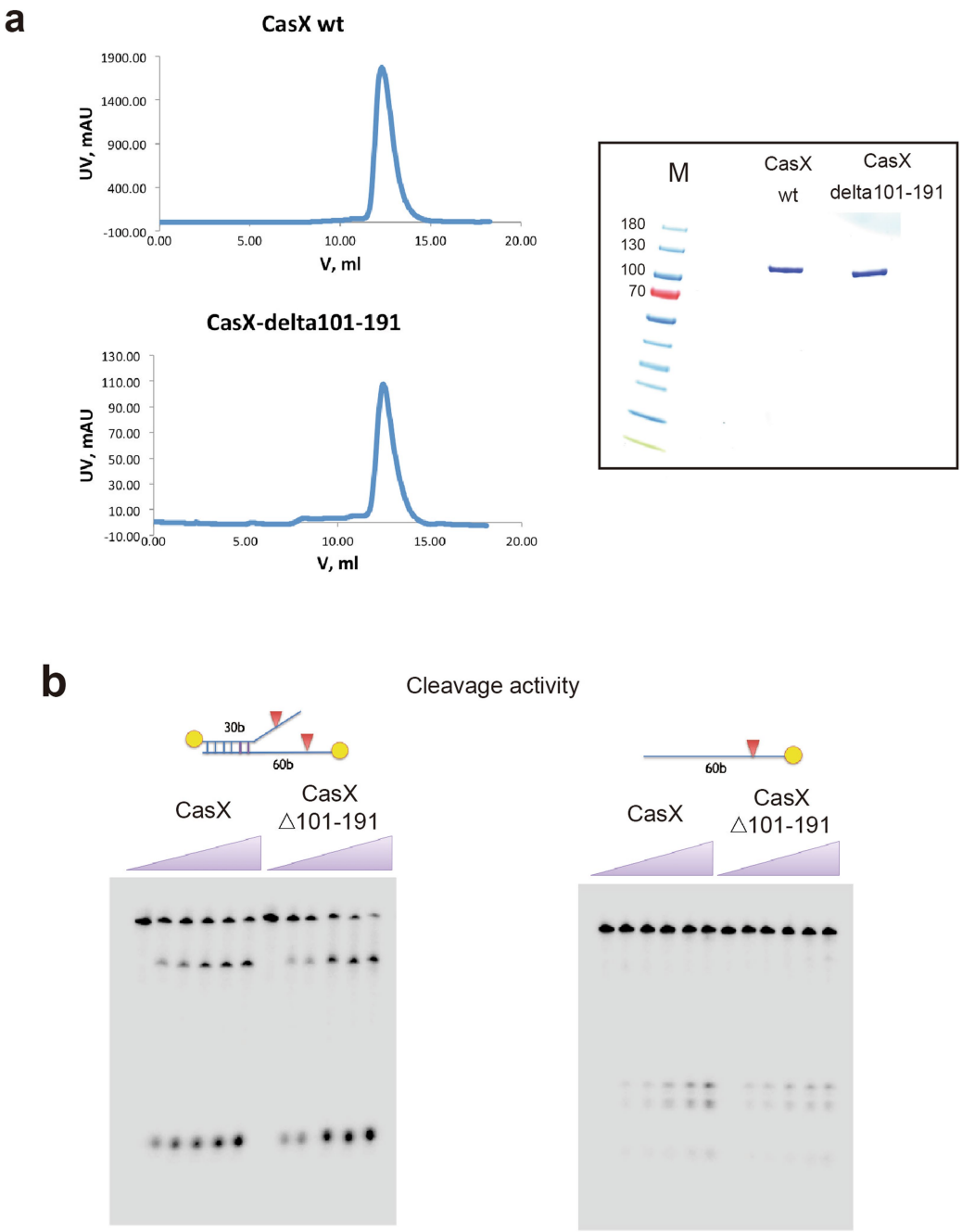
Extended Data Fig. 6 | Structural comparison of CRISPR effectors. **a**, OBD (or wedge domain, WED) shown in aquamarine, helical I (or REC1) domains are shown in yellow, helical II (or REC2) domains are shown in orange, RuvC domains are shown in green, TSL (or Nuc) domains are shown in pink and bridge helices are shown in blue. The NTSB domain in CasX is shown in red, and PAM interaction (or PI) domain of LbCas12a is shown in purple. Guide RNA and target DNA are shown in grey. Two orientations are presented for each model. **b**, Overall structure and individual domains of CasX were analysed using the Dali server against the full PDB database. The protein hit with highest *Z*-score for each target is shown in the top panel. The hits are marked with protein name and PDB code. The similarity scores between CasX overall structure or domains and *Alicyclobacillus acidoterrestris* Cas12b (here given as AacCas12b) are pulled out from the Dali full PDB analysis and shown in the middle panel. The similarity scores between CasX overall structure/domains and LbCas12a are pulled out from the Dali full PDB analysis and shown in the bottom panel. A *z*-score of above eight indicates a high degree of similarity; a *z*-score below eight but above two indicates moderate similarity (usually an irrelevant random match; and a *z*-score below two indicates noise. **c**, The TSL domain and full R-loop structures are subtracted from the ternary complex. Zinc ribbon residues are coloured in blue. The primary sequence across the TSL loop is shown. Tyrosines are marked with teal circles. Positively charged residues are marked with red circles. **d**, Zinc-finger validation by X-ray fluorescence elemental analysis. Bovine erythrocyte carbonic anhydrase that contains zinc in the active site was used as a positive control. Representative zinc peaks appeared in the purified CasX sample but not in the purified Cas9 sample. **e**, Atomic models of DpbCasX, *Alicyclobacillus acidoterrestris* Cas12b, LbCas12a and *Sp*Cas9 (here given as SpyCas9) binary complexes are shown in surface representation. Protein parts are coloured in cyan, and nucleic acids in dark grey. CasX, *Alicyclobacillus acidoterrestris* Cas12b and *Sp*Cas9 require both crRNA and tracrRNA (or a fused sgRNA), whereas LbCas12a uses only crRNA. Guide RNAs are subtracted from the complexes and shown as ribbons in bottom panels independently. The mass ratio of protein to guide RNA is shown on the right. Values of relative mass occupancy for protein and guide RNA within the three binary complexes (protein + guide RNA) are shown. Protein mass occupancies are coloured in cyan, and guide RNA in dark grey. **f**, CRISPRi efficiency by guide RNA mutation ($n = 3$, mean ± s.d.). The sequence for the fused sgRNA is shown. tracrRNA, the joint loop, crRNA and spacer region are marked. The sequences for mutated guide RNA are aligned with the original guide RNA sequence and shown. Cas9 is used as a positive control. +, targeting guide; −, non-targeting guide as a negative control. NC, non-complementary CasX guide. WT, complementary wild-type guide for CasX. GFP-disruption efficiency of targeting guide is shown by the GFP signal divided by optical density at 600 nm, compared to the non-targeting guide control.



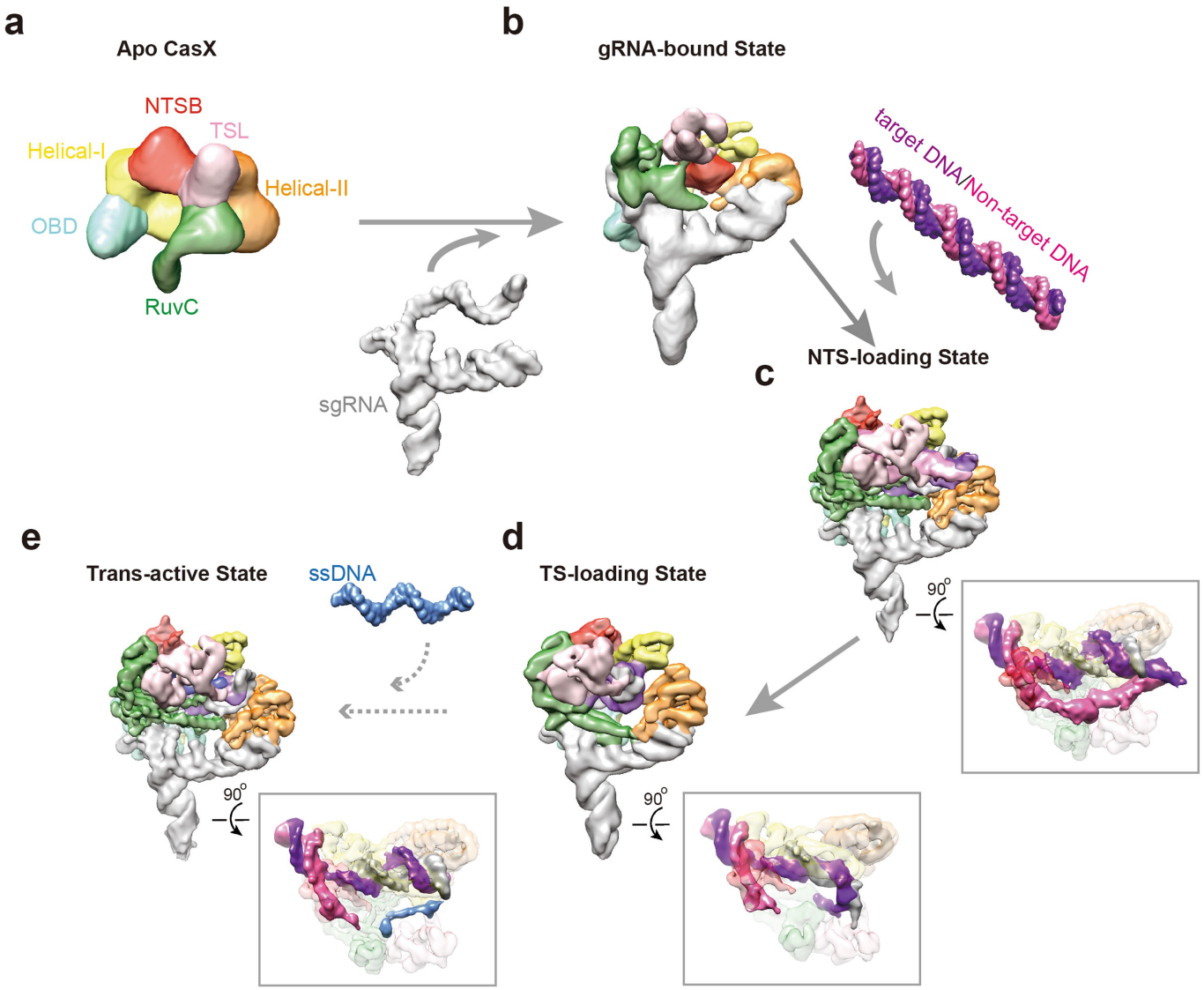
Extended Data Fig. 7 | Structural comparison of apo, binary and ternary CasX samples. a, Drift-corrected image of apoCasX obtained with a 70° phase shift and defocus of 0.5 μm. Scale bar, 50 nm. **b**, Drift-corrected image of CasX–guide RNA complex with a defocus of −1.5 μm. **c**, Drift-corrected image of CasX–guide RNA–DNA complex with a defocus of −1.5 μm. Representative reference-free 2D class averages are shown in the bottom panels for the three samples (scale bar for bottom panels, 20 nm). **d**, Cryo-EM reconstruction of apo-CasX. Three representative orientations are shown with coloured domains. OBD, aquamarine; NTSB, red; helical I, yellow; helical II, orange; RuvC, dark green; TSL, light pink; and bridge helix, blue. **e**, BS3 cross-linking signals revealed by mass spectrometry for the apo-CasX sample. The two lysines within a cross-linked pair are connected with a purple curve. **f**, **g**, As for **d** and **e**, for the CasX–guide RNA binary complex. **h**, **i**, As for **d** and **e**, for the CasX–guide RNA–DNA ternary complex. **j**, **k**, Accessibility of target-strand DNA by the RuvC domain in state I and state II. Distance between the TS DNA cleavage region and RuvC active site as calculated using Pymol is 43.8 Å for state I (**j**) and 10.9 Å for state II (**k**).



Extended Data Fig. 8 | Electron microscopy analysis of CasX–guide RNA–DNA ternary complex with shortened NTS (20-nucleotide NTS and 45-nucleotide TS). a, Target DNA sequence in this ternary complex. **b**, Cryo-EM analysis pipeline. From 3,500 drift-corrected micrographs, 801,927 particles were picked and used for 2D classification. By 2D-based manual screening, 369,430 good particles were selected for 3D classification into 4 classes. From the class showing better structure preservation, 181,009 particles were further used for heterogeneous refinement; thish generated two models, state I and state II, with 33.6% and 66.4% of the particles, respectively. State I and state II were then independently refined to 4.5 Å and 4.4 Å, respectively, by homogenous reconstruction. **c**, The Euler angle distribution of refined particles belonging to state I and state II. **d**, GSFSC curve calculated using two independent half-maps, indicating an overall resolution of 4.5 Å for state I and 4.4 Å for state II. **e**, Cryo-EM structures of state I and state II, coloured by local resolution as calculated in Cryopsarc. The resolution ranges from 3 Å to 7 Å. **c** and **d** are directly adopted from the standard outputs of Cryosparc.



Extended Data Fig. 9 | Purification and substrate cleavage of CasX with a NTSB deletion. a, The representative S200 size-exclusion traces by absorbance at 280 nm for wild-type CasX and for CasX with a deletion of the NTSB domain. SDS–PAGE of wild-type CasX protein and CasX protein with a deletion of the NTSB domain by Coomassie brilliant blue staining is shown in the right panel. **b**, Comparison of the cleavage activities of wild-type CasX and CasX with a truncation of the NTSB domain on an unwound probe (only the PAM region is base-paired, the rest of the probe is mismatched) and on just a single target DNA strand. All the assays have been repeated three times with consistent results.



Extended Data Fig. 10 | Proposed model for sequential CasX activation for DNA cleavage. a, Proposed overall architecture of apo-CasX. The different protein domains are coloured as in Fig. 3. **b**, Cryo-EM map of the guide RNA-bound CasX. Upon guide RNA binding, CasX undergoes a domain rearrangement (guide RNA is shown as a grey solid surface). **c**, Cryo-EM map of the CasX ternary complex in the NTS loading state (state I). Upon target dsDNA recognition and unwinding by the CasX–guide RNA complex, the non-target strand is preferentially positioned into the RuvC active site for cleavage. **d**, Cryo-EM map of the CasX ternary complex in the TS loading state (state II). After non-target-strand cleavage, the entire RNA–DNA duplex is bent by the TSL domain, thus positioning the target strand into RuvC active site. **e**, Cryo-EM of the CasX ternary complex mimicking a hypothetical *trans*-active state. After target-strand DNA cleavage, the tension within the bent RNA–DNA duplex favours the return of the CasX ternary complex to state I, thus enabling the RuvC domain to cut any accessible ssDNA. The model shown here corresponds to the CasX ternary complex with a short NTS DNA in state I to mimic the *trans*-ssDNA cleavage state (the 5′ overhang of TS DNA, which folds back to RuvC domain, is coloured in blue).

nature research

Corresponding author(s): Jennifer A. Doudna, Eva Nogales

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	SerialEM V3.6
Data analysis	Relion2.0, Cryosparc v1, Spider, EMAN2, Coot, Phenix2400, VMD-MDFF v0.4, UCSF-Chimera v1, Pymol v2, Snapgene V4, GraphPad Prism 7,Motioncor2, Gctf V1, Gautomatch V1, pLink1.9, pLabel2.8, Molprobity, MAFFT v7, RAxML v8, iTOL v3, CD-HIT

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 upon request. 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

The EM maps have been deposited to EMDB with the following accession codes: EMD-8987, EMD-8988, EMD-8980, EMD-8994, EMD-8996, EMD-8991, EMD-8989, EMD-8890. The refined atomic models have been deposited to PDB with the following accession codes: 6E50, 6E7A, 6E79

All the unctted gels have been included in the source Data. CryoEM raw data and crosslinking data(> 10TB) are available upon email request.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Reported in figures
Data exclusions	No data excluded
Replication	All the data presented in this work have been replicated three or more times in biological replicate with consistent results including, E. coli. assays, Mammalian cell gene editing assays, protein purification, DNA cleavage assays, cryo-EM sample preparations, the single particle analysis of CasX with different substrates serve as technical replication for each other.
Randomization	For cryo-EM data collection, the exposure squares are randomly selected across the grid with good ice distribution. In cryoEM analysis, all the particles are randomly assigned to different classes in initial round, then generally converged in later rounds based on the cross correlation value with the class averages.
Blinding	No blinding was undertaken. All the exprements in this story are eitehr in vitro or cell based in vivo. Blindling is not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Oakes et al. 2016 & The UC Berkeley Cell Culture Facility
Authentication	Cell lines were not re-authenticated for this study
Mycoplasma contamination	Cells were not re-tested for mycoplasma for this study, they were originally tested in the 2016 study, and The UC Berkeley Cell Culture Facility maintains verified mycoplasma free stocks.
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	Cells were dissociated from plates with trypsin, quenched with FBS & spun down, the media was decanted and the cells were resuspended in PBS
Instrument	All samples were run on an Attune NXT cytometer with a HTS
Software	FlowJo was used to analyze samples
Cell population abundance	All samples represent at least 10,000 events
Gating strategy	FSC/SSC and then on FSC hight vs area to select for single cells, following this cells were gated on FSC vs GFP for GFP disruption

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