nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-020-00721-2

Check for updates

Cryo-EM structure of the RNA-guided ribonuclease Cas12g

Zhuang Li[1,3], Heng Zhang[1,3], Renjian Xiao[1,3], Ruijie Han[1] and Leifu Chang[1,2] ✉

Cas12g, the type V-G CRISPR–Cas effector, is an RNA-guided ribonuclease that targets single-stranded RNA substrate. The CRISPR–Cas12g system offers a potential platform for transcriptome engineering and diagnostic applications. We determined the structures of Cas12g–guide RNA complexes in the absence and presence of target RNA by cryo-EM to a resolution of 3.1 Å and 4.8 Å, respectively. Cas12g adopts a bilobed structure with miniature REC2 and Nuc domains, whereas the guide RNAs fold into a flipped 'F' shape, which is primarily recognized by the REC lobe. Target RNA and the CRISPR RNA (crRNA) guide form a duplex that inserts into the central cavity between the REC and NUC lobes, inducing conformational changes in both lobes to activate Cas12g. The structural insights would facilitate the development of Cas12g-based applications.

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) systems endow prokaryotes with immunity against mobile genetic elements[1–4]. CRISPR systems are grouped into two classes comprising six types[5–7]. The class 1 system (types I, III and IV) employs multiprotein effector complexes to cleave foreign nucleic acid, while the class 2 system (types II, V and VI) utilizes a single multidomain effector endonuclease. The specific DNA targeting and cleavage activities of CRISPR–Cas systems have inspired their successful applications in genome editing, with CRISPR–Cas9 from the type II system most widely used[8,9]. However, RNA-targeting CRISPR–Cas effectors have been developed into various tools for manipulating specific RNA molecules, such as site-specific RNA editing[10], RNA knockdown[11] and molecular diagnostic detection of nucleic acids[10]. RNA editing offers a tool for temporary and reversible alteration in gene function, thus avoiding permanent changes to the genome as in the case of DNA editing, where off-targeting is a big concern when used for clinical treatment. RNA-targeting endonucleases have been found in type II, III and VI CRISPR–Cas systems[12–16]. Nevertheless, current RNA-targeting applications are mostly based on the RNA-dedicated type VI systems (Cas13a–13d), which exclusively catalyze RNA cleavage with HEPN domains[17–19].

Cas12g is a type V RNA-guided endonuclease recently identified through mining the metagenomic database[20]. Cas12g specifically recognizes RNA substrates, distinguishing it from other Cas12 proteins discovered so far, including the well-studied Cas12a (Cpf1), Cas12b (C2C1) and Cas12e (CasX) that target DNA and have been successfully harnessed for genome editing in mammalian cells[21–24]. Cas12g employs two guide RNAs, a crRNA and a *trans*-activating crRNA (tracrRNA), to target single-stranded RNA (ssRNA) for cleavage without the requirement of a protospacer adjacent motif (PAM) sequence[20]. After activation by target RNA, Cas12g displays collateral RNase and single-strand DNase activities[20], similar to the collateral activity of Cas12a and Cas13 proteins[25–28]. Distinct from Cas12a and Cas13, in vitro experiments showed that Cas12g is incapable of processing pre-crRNA[20].

Originating from a hot spring metagenome, Cas12g is thermostable while harboring comparable RNA detection sensitivity compared to the best-performing Cas13 effectors[20]. Another advantage of Cas12g is its compact size (only 767 amino acids), compared to Cas13 proteins (average sizes for Cas13a–13d are 1,250, 1,150, 1,120 and 930 amino acids, respectively[29]). The small size potentially facilitates flexible packaging into size-constrained therapeutic viral vectors, such as adeno-associated virus (AAV)[30]. Thus, Cas12g has the potential to enhance RNA detection and other in vivo transcriptome engineering applications.

Understanding the molecular basis underlying target RNA recognition and cleavage by Cas12g is critical for the development of Cas12g-based tools. Structural studies of type V effectors, including Cas12a[28,31–39], Cas12b[40–42], Cas12e[24] and Cas12i[43], showed that they display a canonical bilobed architecture encompassing a recognition lobe (REC) and a nuclease lobe (NUC). The REC lobe contains the REC1 and REC2 domains, which are responsible for substrate recognition, whereas the NUC lobe is composed of the RuvC, Wedge (WED), PAM-interacting (PI) and Nuc domains, with the nuclease site located in the RuvC domain. However, low sequence similarity between Cas12g and other Cas12 effectors and RNA-targeting function of Cas12g indicates differences in its mechanisms for target recognition and cleavage, motivating us to investigate the structure of Cas12g.

Results

Overall structure of Cas12g–sgRNA. We assembled a binary complex of wild-type Cas12g and a 136-nucleotide single-guide RNA (sgRNA, fusion of tracrRNA and crRNA) and determined the structure of this complex at an average resolution of 3.1 Å by cryo-EM (Extended Data Fig. 1 and Supplementary Table 1). The resultant map allowed ab-initio model building of most residues and nucleotides of the complex, with the exception of several segments at the periphery of the complex (Extended Data Fig. 2 and Supplementary Fig. 1).

The overall structure of Cas12g displays a bilobed architecture containing the REC and NUC lobes, enclosing a positively charged central channel between the two lobes (Fig. 1). sgRNA primarily engages Cas12g by contacting the REC lobe, with additional interactions with the WED domain from the NUC lobe (Fig. 1c). A DALI search revealed Cas12b as the closest structural match to Cas12g (*Z* score: ~14; root mean square deviation: 5.8 Å over 466 amino acids)

[1]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA. [2]Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN, USA. [3]These authors contributed equally: Zhuang Li, Heng Zhang, Renjian Xiao. ✉e-mail: lchang18@purdue.edu

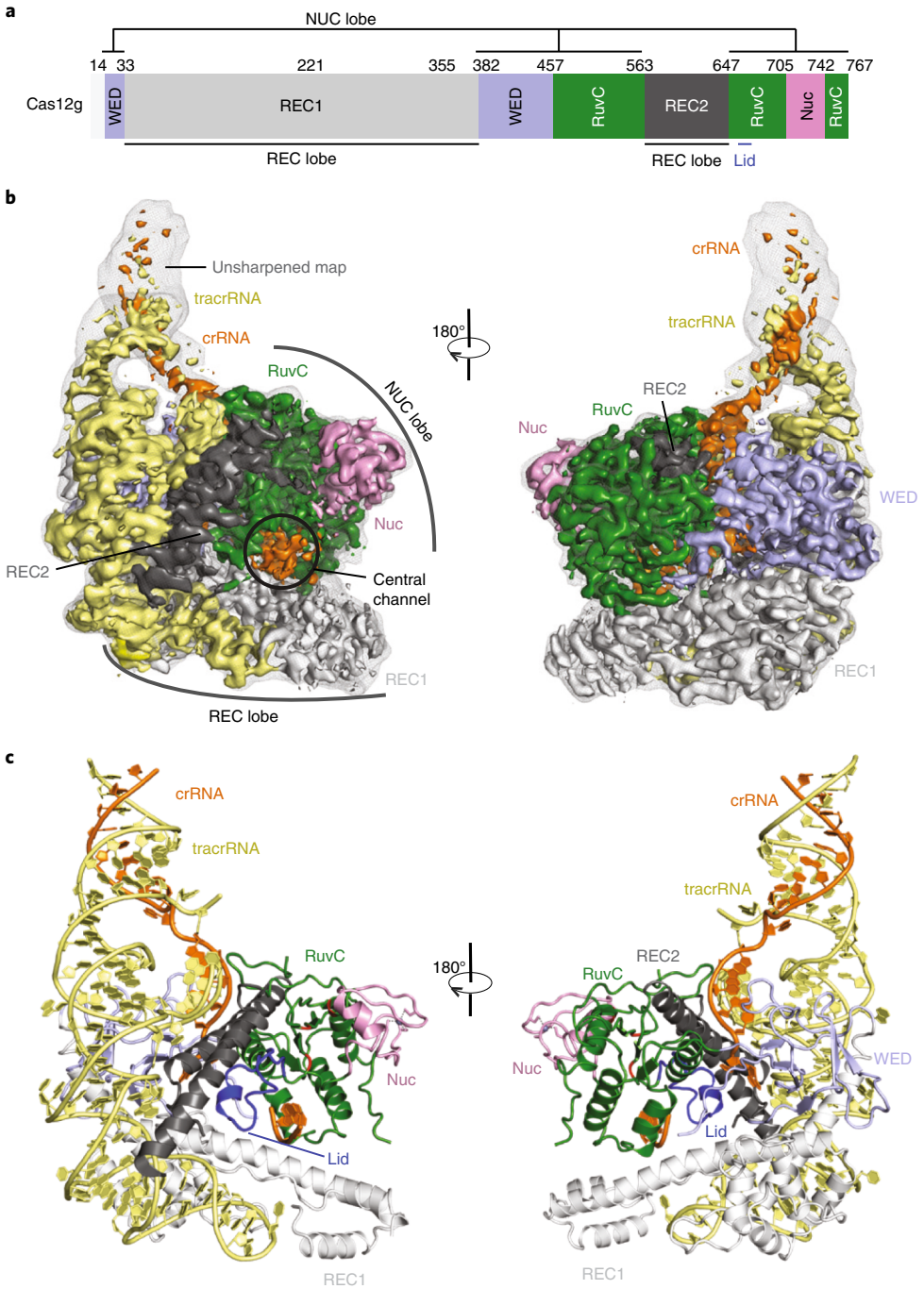


Fig. 1 | Overall structure of the Cas12g–sgRNA binary complex. a, Domain organization of Cas12g. **b**, Cryo-EM map of the Cas12g–sgRNA complex in two views with each domain of Cas12g color coded as in **a**. The crRNA and tracrRNA regions of sgRNA are colored in orange and light yellow, respectively. **c**, Atomic model of the Cas12g–sgRNA complex shown in cartoon.

(Extended Data Fig. 3 and Supplementary Fig. 1), whereas Cas12a and Cas12e are also close matches (*Z* scores: 9–10).

The dumbbell-shaped REC1 domain contains an N-terminal helical subdomain (α1–5) and two helices at the C terminus (α7–8), connected by a featured long helix (α6) (Fig. 2a). A large portion of the C terminus between α7 and α8 (amino acids 220–354, REC1$^{220–354}$) is barely visible from the map (Figs. 1b and 2a), indicating that this region is flexible. Three-dimensional (3D) classification of cryo-EM images focusing on this region resulted in a major class with this region determined at a lower resolution (Extended Data Fig. 1g), showing that REC1$^{220–354}$ is located on the exit of the central channel (Extended Data Fig. 1h). Secondary structure prediction suggests that this region is mainly composed of helices, similar to the structure of this region in Cas12b (Extended Data Figs. 2b and 3b). The REC2 domain is simplified into two helices, lacking a four-helix bundle as in Cas12b that is involved in the recognition of the crRNA–target DNA heteroduplex[40] (Fig. 2b and Extended Data Fig. 3c).

The NUC lobe contains the WED, RuvC and Nuc domains, but lacks the PI domain, consistent with the previous report that PAM is

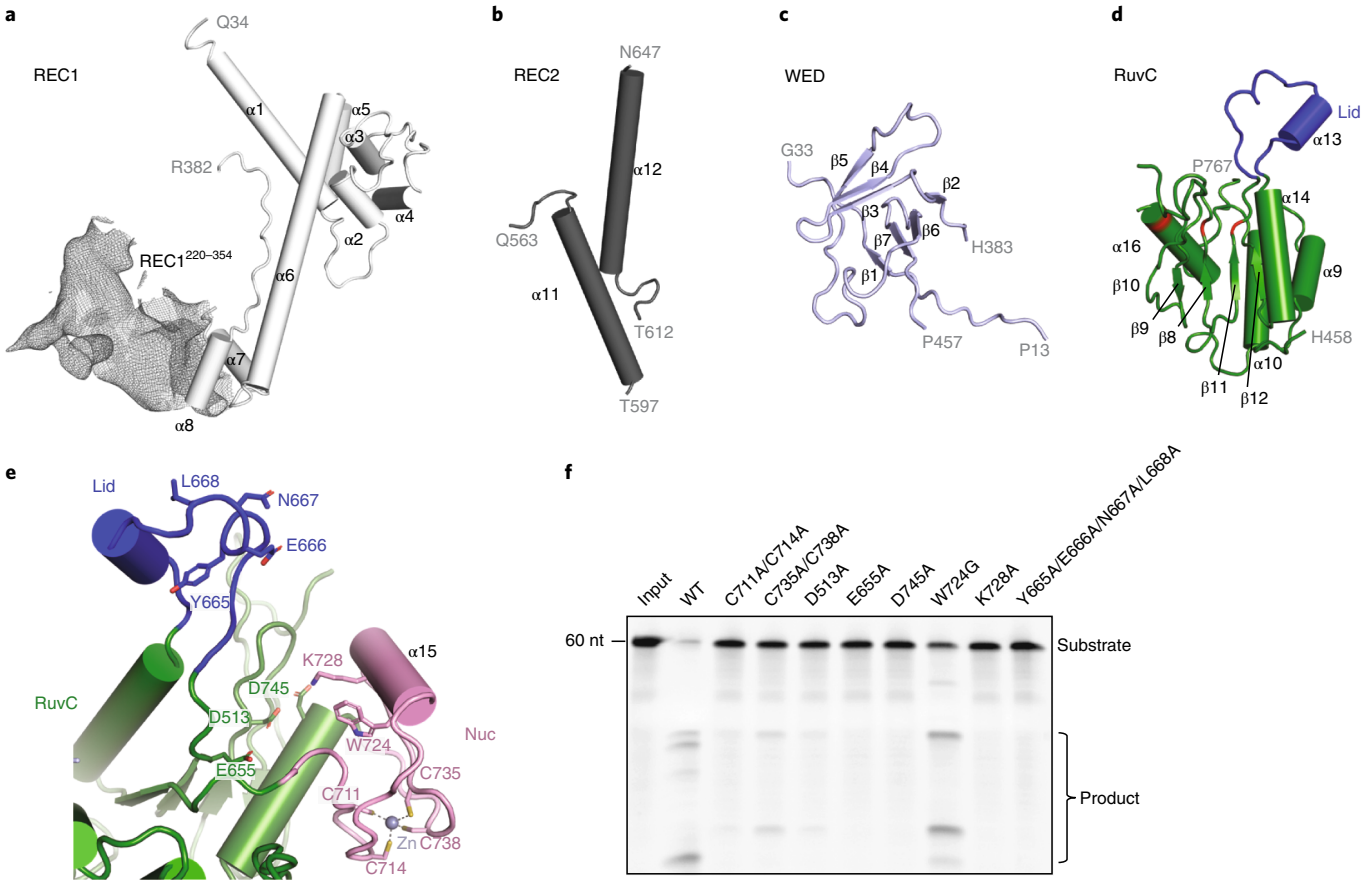


Fig. 2 | Domain structure and active site of Cas12g. a–d, Cartoon presentation of the REC1 domain (**a**), REC2 domain (**b**), WED domain (**c**), and RuvC domain (**d**). Density of REC1$^{220-354}$ observed by 3D-focused classification is shown in mesh. The triplet of acidic residues in the active site and the lid motif are colored in red and blue, respectively. **e**, The endonuclease center in the RuvC domain. The acidic residues D513, E655 and D745 are labeled. The Nuc domain is attached to the active center, with K728 and W724 pointing to the active site. A zinc ion chelated by four cysteines (C711, C714, C735 and C738) is present in the Nuc domain. **f**, Substrate RNA cleavage assay using wild-type Cas12g and Cas12g with mutations as indicated. The results shown are representative of three experiments. nt, nucleotide.

not required in substrate recognition of Cas12g[20]. The WED domain consists of seven β-strands, folding into two β-sheet layers (Fig. 2c). The conserved RuvC domain is composed of a five-β-strand (β8–12) core flanked by five α-helices, with α13 and surrounding loops (amino acids 659–679) equivalent to the lid motif of other Cas12 nucleases[36,43] (Fig. 2d). The lid motif in Cas12a is involved in substrate recognition, and conformational changes of the lid motif act as molecular checkpoints to activate catalysis[36,43]. Attached to the RuvC domain is the Nuc domain, which is only 37 amino acids in length and contains two CXXC zinc finger motifs, 711-CAEC-714 and 735-CEGC-738 (Fig. 2e). Alanine substitutions of these cysteine residues abolished substrate RNA cleavage, suggesting that the zinc finger motifs are critical for the nuclease activity of Cas12g (Fig. 2f). The zinc finger motifs were also observed in the Nuc domain of Cas12e[24], but not in Cas12a[28,31–39], Cas12b[40–42] and Cas12i[43] (Supplementary Table 2).

Nuclease site of Cas12g. Located in the interface between the RuvC and Nuc domains are a conserved triplet of acidic residues (D513, E655 and D745) from the RuvC domain and two bulky residues (W724 and K728) from the Nuc domain (Fig. 2e). Alanine substitutions of each of the triplet (D513A, E655A and D745A) and K728 abolished substrate cleavage activity of Cas12g, while mutation of W724 (W724G) showed compromised nuclease activity (Fig. 2f). The lid motif in Cas12g lies above the active site (Fig. 2e). Substitution of key residues within the lid motif (665-YENL-668) with alanine abolished the substrate cleavage activity (Fig. 2f) as well as collateral cleavage activity towards ssDNA and ssRNA (Extended Data Fig. 4a,b), indicating that the lid motif plays a critical role in the activation of Cas12g, as observed in Cas12a[36]. These results suggest that Cas12g utilizes a nuclease site at the interface between the RuvC and Nuc domains to perform multiple-turnover cleavage of RNA substrate (Extended Data Fig. 4c).

Structure of sgRNA. The sgRNA of Cas12g contains a 92-nucleotide tracrRNA at the 5′ end, and a 42-nucleotide crRNA at the 3′ end (18-nuncleotide repeat and 24-nuncleotide spacer-derived sequence) (Fig. 3a). The sgRNA folds into a flipped 'F'-shaped structure consisting of tracrRNA scaffolding duplexes 1–2 and anti-repeat:repeat (AR:R) duplexes 1–2 (Fig. 3a,b and Extended Data Fig. 2g). AR:R duplex 1 is extended from scaffolding duplex 2 with no observable contacts to Cas12g, consistent with its flexibility and compromised resolution in the cryo-EM map (Fig. 1b and Extended Data Fig. 1j). Lack of direct contact with Cas12g implies that the potential function of AR:R duplex 1 is to tether the tracrRNA and crRNA, rather than to directly contribute to the nuclease activity. Consistent with this hypothesis, truncation of the AR:R duplex 1 in the sgRNA presented comparable nuclease activity with the full-length sgRNA (Fig. 3c).

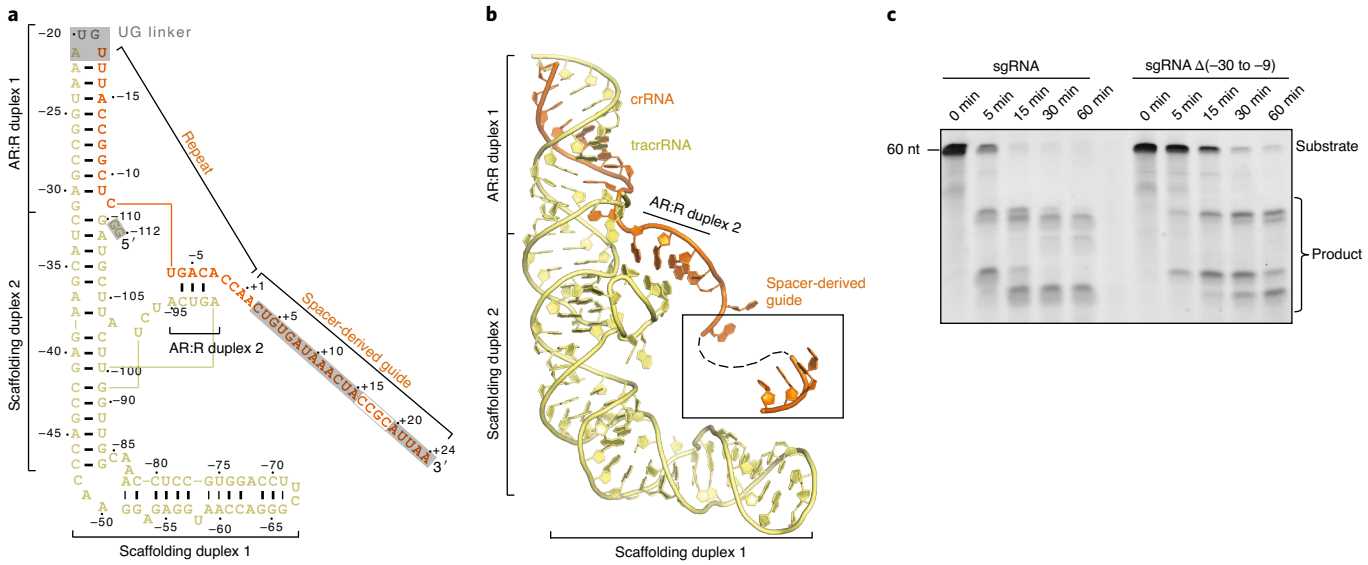


Fig. 3 | Overall structure of sgRNA. a, Schematic of the sgRNA. Unmodeled regions are indicated by a gray background. **b**, Structure of the sgRNA in the Cas12g–sgRNA complex in cartoon presentation. **c**, RNA substrate cleavage assay using full-length sgRNA and sgRNA with the AR:R duplex 1 deleted. The results shown are representative of three experiments.

Interactions between sgRNA and Cas12g. Scaffolding duplex 1, formed by tracrRNA (−82 to −51), is recognized by the concave surface of the REC1 domain and the REC2 domain, mainly through nonsequence-specific interactions (Fig. 4a,b and Extended Data Fig. 2h), occupying a binding surface in REC1 that is used for recognition of the crRNA–target DNA duplex in other Cas12 effectors, such as Cas12b (Extended Data Fig. 5a–c). Scaffolding duplex 2, almost perpendicular to scaffolding duplex 1, is bracketed by the REC1, REC2 and WED domains (Fig. 1c).

A stretch of tracrRNA (−99 to −92) protruding from the scaffolding duplex 2 base pairs with the crRNA repeat region (−6 to −4), forming the AR:R duplex 2 (Fig. 4c,d), which establishes extensive interactions with the REC2, RuvC and WED domains of Cas12g through charged interactions. For example, H572, R575, H576 and H624 of the REC2 domain, and K539 of the RuvC domain, establish charged interactions with multiple phosphate groups of the loop (Fig. 4c,d), stabilizing the AR:R duplex 2 and thereby facilitating the correct positioning of the crRNA guide sequence. Mutations of each of these residues (H572A, R575E, H576A, H624A and K539E), or introducing mismatches in the AR:R duplex 2 (U-98A/U-97A/C-96G) severely reduced the nuclease activity of Cas12g (Fig. 4e).

The spacer-derived guide inserts into the positively charged channel of Cas12g (Fig. 1b,c). However, the corresponding densities are very weak, indicating their flexibility (Extended Data Fig. 2g), with the exception of a four-nucleotide segment bound to the RuvC domain (Fig. 4f). This segment forms an A-form helical conformation at the exit of the central channel, reminiscent of Cas13a, which is reported to adopt a 'central seed' to initiate target recognition[44]; however, the resolution of this region did not allow for unambiguous assignment of the crRNA sequence.

Structure of Cas12g–sgRNA–target RNA. To gain insights into target RNA recognition, we assembled a Cas12g–sgRNA–target RNA ternary complex by incubating a catalytically inactive Cas12g (E655A) with the sgRNA and a 24-nucleotide target RNA. We determined the structure at a resolution of 4.8 Å by cryo-EM (Fig. 5a,b, Extended Data Fig. 6 and Supplementary Table 1). Unambiguous secondary-structure features of the map allow fitting of individual domain structure into the map (cross-correlation coefficient, 0.92). An A-form duplex is observed in the central channel of Cas12g, which is interpreted as the crRNA–target RNA heteroduplex (Fig. 5a,b). Modeling of the duplex showed a length of about 24 base pairs (bp), consistent with the length of target RNA used for complex assembly. The crRNA–target RNA duplex is primarily attached to the RuvC, REC1 and REC2 domains (Fig. 5b).

Structural comparison between the Cas12g–sgRNA binary complex and the Cas12g–sgRNA–target RNA ternary complex reveals notable conformational changes. The REC and NUC lobes move away from the central channel by approximately 10 Å and 3 Å, respectively, opening the central channel to accommodate the crRNA–target RNA duplex (Fig. 5c). The crRNA–target RNA duplex induces movement in the lid motif, which is presumably involved in the interactions with the duplex (Fig. 5c). Together with the mutagenesis study mentioned above (Fig. 2f), this evidence indicates that the lid motif plays a critical role in the activation of Cas12g. Different from the binary complex, in which the REC1$^{220–354}$ is located at the exit of the central channel, the REC1$^{220–354}$ in the ternary complex is attached to the crRNA–target RNA duplex (Fig. 5d and Extended Data Fig. 7a,b). Deletion of REC1$^{220–354}$ abolished the nuclease activity, indicating an important role in substrate cleavage (Fig. 5e). The REC1$^{220–354}$ region is rich in positively charged residues, which are probably involved in the recognition of the crRNA–target RNA duplex. Consistent with this hypothesis, mutation of six positively charged residues within the first 20 amino acids of this region (K221, K223, R224, R226, R232 and K237) showed a similar effect in RNA cleavage efficiency as deletion of REC1$^{220–354}$ (Fig. 5e).

REC1$^{220–354}$ engages the crRNA–target RNA duplex at positions 16–19. RNA substrate cleavage assay showed that mismatches between the crRNA and RNA substrate at positions 16–19 almost abolished substrate cleavage activity, while mismatches at flanking positions had little effect (Fig. 5f and Extended Data Fig. 5d), suggesting that recognition of the crRNA–target RNA duplex by REC1$^{220–354}$ is critical for Cas12g activation. Mismatches at positions 16–19 also severely reduced the collateral cleavage activity towards ssDNA and ssRNA (Extended Data Fig. 4a,b). Substrate RNAs with mismatches at positions 16–19 are capable of binding to Cas12g–sgRNA (Extended Data Fig. 8), indicating that the base pairing and stabilization of this region is vital for activation of Cas12g.

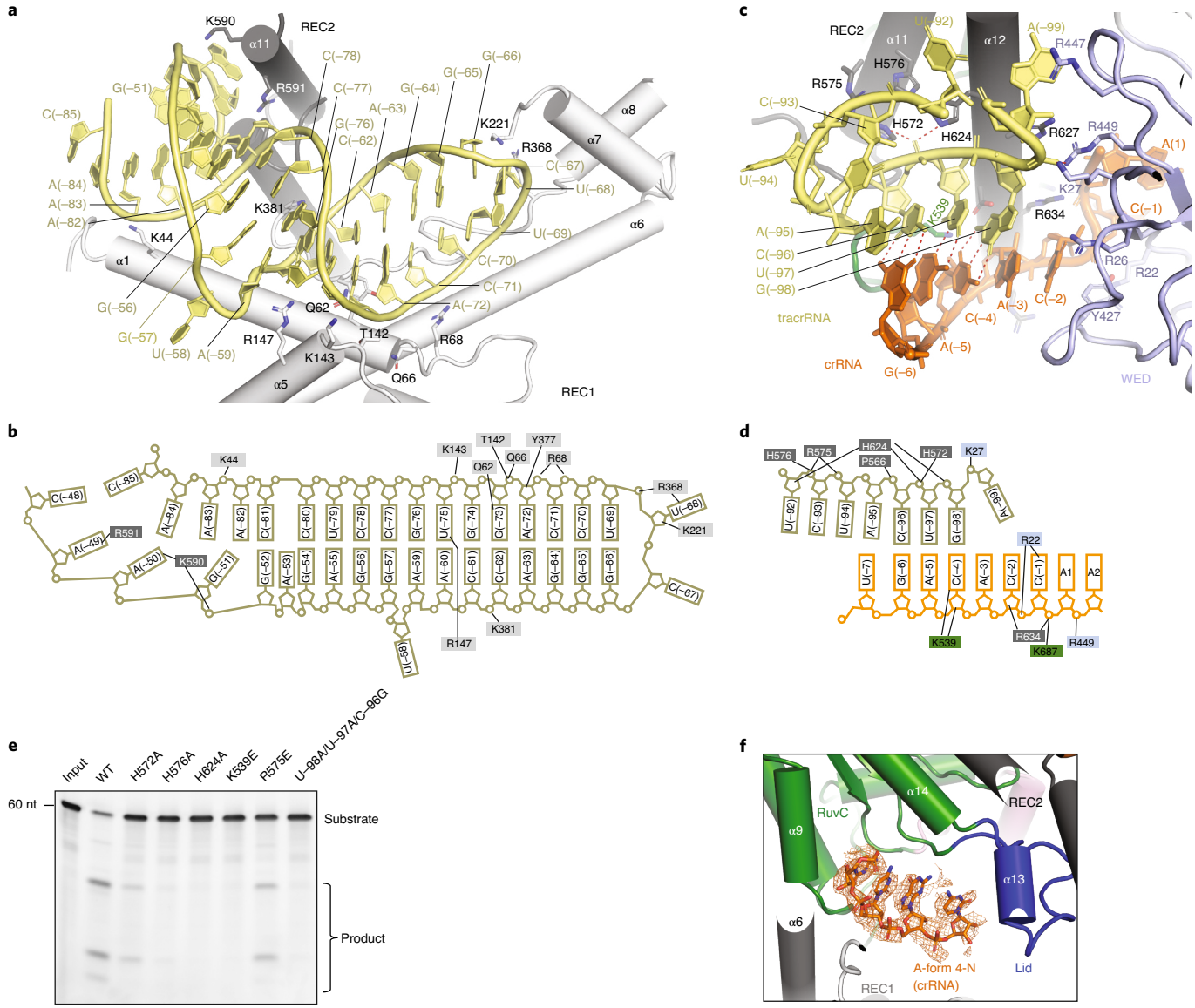


Fig. 4 | Interactions between Cas12g and sgRNA. a,**b**, Detailed interactions (**a**) between scaffolding duplex 1 and the REC lobe of Cas12g, and the schematic of the interactions (**b**). **c**,**d**, Detailed interactions (**c**) between the AR:R duplex 2 of sgRNA and Cas12g, and the schematic of the interactions (**d**). **e**, Substrate RNA cleavage assay using wild-type Cas12g and Cas12g with mutations in residues involved in interactions with the AR:R duplex 2, as well as mismatches in the AR:R duplex 2. The results shown are representative of three experiments. **f**, A four-nucleotide segment from the crRNA is pre-ordered in an A-form configuration, attached to the RuvC domain. EM density corresponding to the four-nucleotide segment is shown in mesh.

Interestingly, target ssDNA is also able to bind to Cas12g–sgRNA (Extended Data Fig. 8). However, cryo-EM analysis suggests that target ssDNA failed to induce conformational changes in Cas12g as observed by target ssRNA (Extended Data Fig. 9), which might explain why ssDNA is incapable of activating Cas12g (discussed below). Collectively, the results provide mechanistic insights into target RNA recognition and activation of Cas12g.

Discussion

To date, 11 subtypes have been reported (Cas12a–12k)[45,46] within type V CRISPR–Cas systems. These effectors are classified into three major branches on the basis of phylogenetic analysis[20]. Branches 1 (for example, Cas12b[40–42]) and 2 (for example, Cas12a[31–36,38,39] and Cas12e[24]) have been structurally studied previously. However, the Cas12g structure reported here represents the structure of a branch 3 effector, shedding light on other members in this branch[20].

The structures reported here suggest that Cas12g share similar domain architecture with several specialized features (Supplementary Table 2). The concave surface of the REC1 domain used for recognition of the crRNA–target DNA duplex in other Cas12 effectors is repurposed for recognition of tracrRNA scaffolding duplex 1. The C-terminal region of REC1, REC1$^{220–354}$, is structurally flexible, in contrast to the ordered structure of this region in other Cas12 effectors[24,31–36,38–42]. However, this region plays a critical regulatory role in the recognition of target RNA and activation of Cas12g.

Target recognition of type V and II systems is generally initiated by interactions between effectors and the PAM duplex of substrates. Recognition of PAM duplex triggers unwinding of DNA substrate, preparing the target strand for hybridization with the crRNA[36,47]. Because Cas12g specifically recognizes ssRNA, it is not surprising that the PI domain is not required. Target recognition

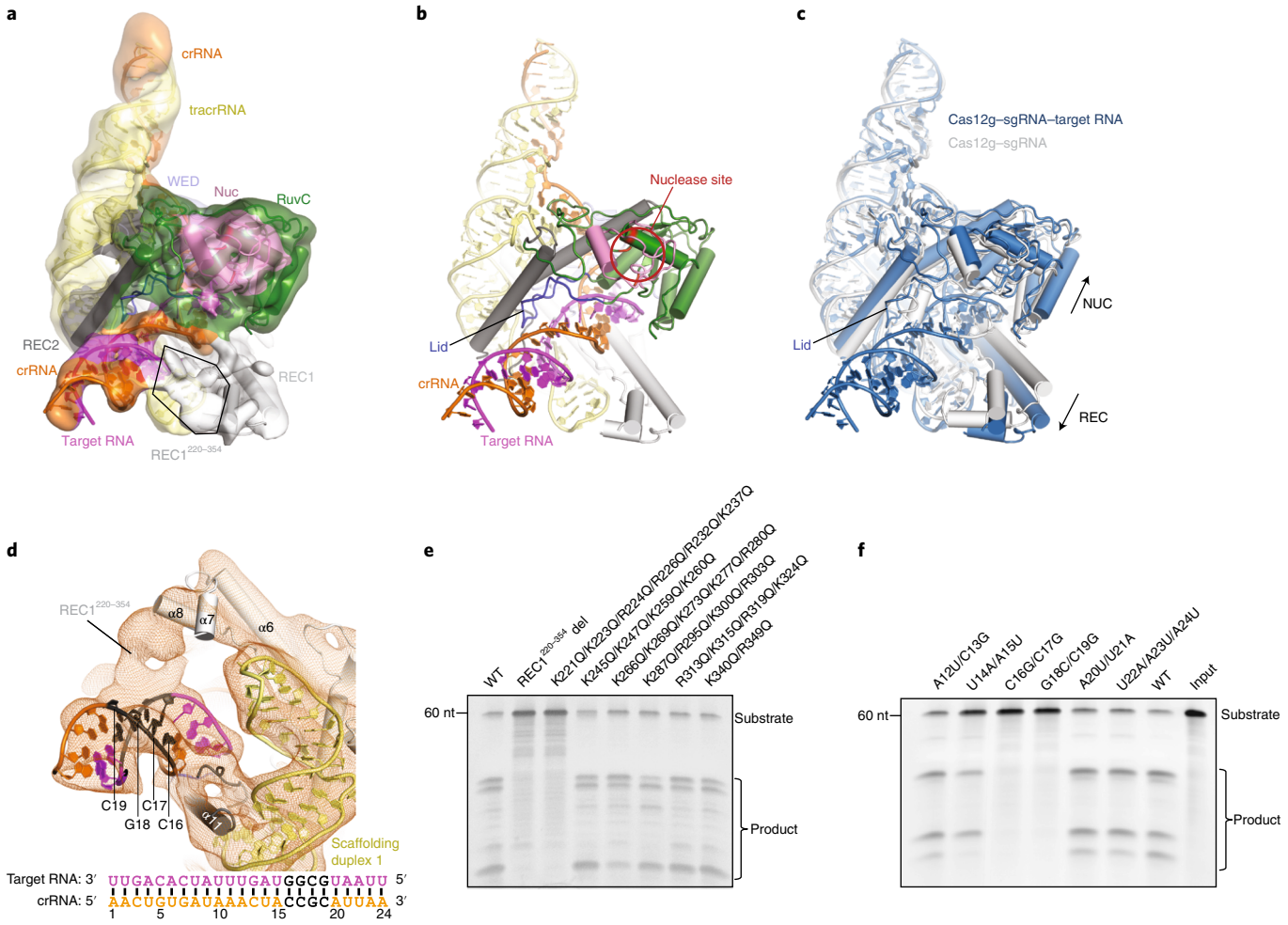


Fig. 5 | Target RNA recognition of Cas12g. a, Cryo-EM map and model of the Cas12g–sgRNA–target RNA complex with each domain of Cas12g color coded as in Fig. 1a. The crRNA, tracrRNA and target RNA are colored in orange, light yellow and magenta, respectively. **b**, Cartoon presentation of the Cas12g–sgRNA–target RNA complex. **c**, Structural superimposition of Cas12g–sgRNA (gray) and Cas12g–sgRNA–target RNA (sky blue) complexes. Conformational changes in Cas12g domains and sgRNA after recognition of target RNA are indicated by arrows. **d**, Contact between REC1$^{220–354}$ and positions 16–19 of the crRNA–target RNA duplex. Schematic of the crRNA spacer-derived guide and target RNA is shown below. **e**, Substrate RNA cleavage assay using wild-type Cas12g and Cas12g with REC1$^{220–354}$ deleted or with group mutations within REC1$^{220–354}$. **f**, Substrate RNA cleavage assay using wild-type and mutant target RNAs. The results in **e** and **f** are representatives of three experiments.

in type VI effectors is initiated by protein–protospacer flanking sequence (PFS) interactions[44]. However, bioinformatics analysis of target-flanking sequences of Cas12g revealed no such requirements for interference[20]. No restrictions of PAM or PFS are an advantage of Cas12g when used for RNA-targeting applications.

In Cas12a, Cas12b and Cas12i, a short region (5–7 nucleotides) at the 5′ end of the spacer-derived guide is pre-ordered in a helical conformation, serving as a seed for substrate recognition[37,40,41,43] (Supplementary Table 2). The seed sequence initiates substrate recognition and is sensitive to mismatches with the substrate. However, such a pre-ordered segment is not observed at the 5′ end of the spacer-derived guide in the Cas12g–sgRNA complex. Instead, a four-nucleotide segment in an A-form configuration is observed at the exit of the central channel (Fig. 4e).

Most Cas12 nucleases, such as Cas12a, Cas12b, Cas12e and Cas12i, specifically recognize target DNA. PAM recognition initiates unwinding of double-stranded DNA (dsDNA) and allows hybridization between the target strand and crRNA from the 5′ end of the crRNA guide. Activation of Cas12 nucleases requires the formation of the crRNA–target DNA heteroduplex at a certain length. For instance, activation of Cas12a requires the heteroduplex to be at least 16–17 bp (refs. [36,47]), whereas activation of Cas12i requires at least 14–15 bp (ref. [43]). Some anti-CRISPR proteins inhibit Cas12 nuclease by suppressing the progression of crRNA–target DNA heteroduplex, such as AcrVA4 which inhibits Cas12a by blocking the heteroduplex formation at 8 bp (ref. [39]).

Cas12g specifically recognizes ssRNA. ssDNA with the sequence complementary to the crRNA is incapable of activating Cas12g[20] (Extended Data Fig. 4a,b). In the Cas12g–sgRNA complex, the 5′ end of the crRNA guide is embedded, whereas the 3′ end of the crRNA guide is exposed to solvent and is a likely initiation site for hybridization between crRNA and target RNA (Fig. 1b,c). Cryo-EM analysis of the Cas12g–sgRNA–target DNA complex shows that the complex is in an inactivated conformation similar to the Cas12g–sgRNA binary complex (Extended Data Fig. 9g). A possible explanation is that hybridization between ssDNA and the 3′ end of the crRNA guide (presumably in a flexible conformation) failed to progress to a required length for activation of Cas12g (Extended Data Fig. 9h), although the possibility of ssDNA disassociation on cryo-EM grids cannot be completely ruled out. REC1$^{220–354}$ and a simplified REC2 domain are possible structural features associated with the RNA specificity of Cas12g (Extended Data Fig. 7).

It was recently shown that type V nucleases, including Cas12a, Cas12b, Cas12e and Cas12i, share a conserved activation mechanism[43]. Before target DNA binding, the lid motif covers the RuvC active site, preventing access to the substrate. Formation of the crRNA–target DNA heteroduplex is accompanied by displacement of the lid motif, which senses formation of the crRNA–target DNA duplex. Displacement of the lid motif opens the RuvC active site and contributes to recruitment of substrate[43]. Conformational changes observed in Cas12g upon target RNA binding, and the critical role of the lid motif, indicate that Cas12g may adopt a similar activation mechanism. Once activated by target RNA, both RNA and DNA substrates may be loaded to the RuvC active site in a favored geometry for catalysis. Future studies will be required to understand the catalytic mechanism of Cas12g towards both RNA and DNA substrates.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-020-00721-2.

Received: 1 July 2020; Accepted: 9 December 2020;

References

1. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82**, 237–266 (2013).
2. Marraffini, L. A. CRISPR–Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
3. Jiang, F. & Doudna, J. A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
4. Hille, F. et al. The biology of CRISPR–Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
5. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
6. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
7. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
8. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
9. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
10. Cox, D. B. T. et al. RNA editing with CRISPR–Cas13. *Science* **358**, 1019–1027 (2017).
11. Abudayyeh, O. O. et al. RNA targeting with CRISPR–Cas13. *Nature* **550**, 280–284 (2017).
12. Terns, M. P. CRISPR-based technologies: impact of RNA-targeting systems. *Mol. Cell* **72**, 404–412 (2018).
13. Dugar, G. et al. CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9. *Mol. Cell* **69**, 893–905 (2018).
14. 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).
15. Rousseau, B. A., Hou, Z., Gramelspacher, M. J. & Zhang, Y. Programmable RNA cleavage and recognition by a natural CRISPR–Cas9 system from *Neisseria meningitidis*. *Mol. Cell* **69**, 906–914 (2018).
16. O'Connell, M. R. et al. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263–266 (2014).
17. Yan, W. X. et al. Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell* **70**, 327–339 (2018).
18. Smargon, A. A. et al. Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell* **65**, 618–630 (2017).
19. Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
20. Yan, W. X. et al. Functionally diverse type V CRISPR–Cas systems. *Science* **363**, 88–91 (2019).
21. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
22. Strecker, J. et al. Engineering of CRISPR–Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).
23. Teng, F. et al. Repurposing CRISPR–Cas12b for mammalian genome engineering. *Cell Discov.* **4**, 63 (2018).
24. Liu, J. J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
25. Chen, J. S. et al. CRISPR–Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
26. Li, S.-Y. et al. CRISPR–Cas12a has both *cis*- and *trans*-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493 (2018).
27. Gootenberg, J. S. et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439–444 (2018).
28. Swarts, D. C. & Jinek, M. Mechanistic insights into the *cis*- and *trans*-acting DNase activities of Cas12a. *Mol. Cell* **73**, 589–600 (2019).
29. O'Connell, M. R. Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR–Cas systems. *J. Mol. Biol.* **431**, 66–87 (2019).
30. Konermann, S. et al. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* **173**, 665–676 (2018).
31. Dong, D. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
32. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type V CRISPR–Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* **26**, 901–913 (2016).
33. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
34. Yamano, T. et al. Structural basis for the canonical and non-canonical PAM recognition by CRISPR–Cpf1. *Mol. Cell* **67**, 633–645 (2017).
35. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).
36. Stella, S. et al. Conformational activation promotes CRISPR–Cas12a catalysis and resetting of the endonuclease activity. *Cell* **175**, 1856–1871 (2018).
37. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR–Cas12a. *Mol. Cell* **66**, 221–233 (2017).
38. Nishimasu, H. et al. Structural basis for the altered PAM recognition by engineered CRISPR–Cpf1. *Mol. Cell* **67**, 139–147 (2017).
39. Zhang, H. et al. Structural basis for the Inhibition of CRISPR–Cas12a by anti-CRISPR proteins. *Cell Host Microbe* **25**, 815–826 (2019).
40. 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 (2016).
41. Liu, L. et al. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol. Cell* **65**, 310–322 (2017).
42. Wu, D., Guan, X., Zhu, Y., Ren, K. & Huang, Z. Structural basis of stringent PAM recognition by CRISPR–C2c1 in complex with sgRNA. *Cell Res.* **27**, 705–708 (2017).
43. Zhang, H., Li, Z., Xiao, R. & Chang, L. Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.* **27**, 1069–1076 (2020).
44. Liu, L. et al. The molecular architecture for RNA-guided RNA cleavage by Cas13a. *Cell* **170**, 714–726 (2017).
45. Swarts, D. C. Stirring up the type V alphabet soup. *CRISPR J.* **2**, 14–16 (2019).
46. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
47. Jeon, Y. et al. Direct observation of DNA target searching and cleavage by CRISPR–Cas12a. *Nat. Commun.* **9**, 2777 (2018).

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 America, Inc. 2021

Methods

Protein expression and purification. The plasmid used for expressing full-length Cas12g with an N-terminal 6× His-tag was obtained from Addgene[20]. The plasmid was transformed into BL21-CodonPlus (DE3)-RIL cells for protein expression in Terrific Broth medium. The protein expression was induced by adding 0.5 mM IPTG at 18 °C overnight. The cells were collected and then lysed by sonication in buffer A (25 mM Tris–HCl, 500 mM NaCl, 5% glycerol, 1 mM PMSF and 5 mM β-mercaptoethanol, pH 8.0). After centrifugation, the supernatants were incubated with the Ni-NTA resin and washed extensively with buffer B (25 mM Tris–HCl, 500 mM NaCl, 30 mM imidazole and 5 mM β-mercaptoethanol, pH 8.0) and the His-tagged target proteins were eluted with buffer B supplemented with 300 mM imidazole. Subsequently, the target proteins were loaded into the heparin column (GE Healthcare) and eluted with a linear NaCl gradient, followed by size-exclusion chromatography (Superdex 200, GE Healthcare) in buffer C (25 mM Tris–HCl, 150 mM NaCl, 2 mM DTT and 0.1 mM $MgCl_2$, pH 8.0). Peak fractions were pooled, concentrated and stocked at −80 °C.

To assemble the Cas12g–sgRNA complex, the sgRNA was incubated with Cas12g proteins at a ratio of 1.3:1 at 37 °C for 30 min. To reconstitute the Cas12g–sgRNA–target RNA complex, the Cas12g E655A mutant protein was incubated with sgRNA and target RNA or target DNA at a ratio of 1:1.3:5 at 37 °C for 60 min. These complexes were loaded onto a Superdex 200 column equilibrated with buffer C for further purification.

Mutagenesis. Single-site mutations were introduced by the QuikChange site-directed mutagenesis method. Group mutations containing multiple sites were introduced by ligating inverse PCR-amplified backbone with mutations bearing DNA oligonucleotides via an In-Fusion Cloning Kit. Large-fragment deletion ($REC1^{220–354}$ deletion) was introduced by ligating inverse PCR-amplified backbone with DNA oligonucleotides encoding a flexible linker ('SGASGGAGS'). All mutants were confirmed by DNA sequencing. The mutant proteins were expressed and purified as described for wild type.

Circular dichroism spectroscopy. Circular dichroism spectra of wild type and mutants of Cas12g were measured at room temperature with a Chirascan (Applied Photophysics) spectropolarimeter using a quartz cuvette with a 1-mm path length. The samples (1.3 mg ml^{-1}) were measured in low-salt buffer (20 mM Tris–HCl, pH 8.0, 50 mM NaCl). The measurements were recorded at wavelengths from 260 to 185 nm (0.2 nm per step) with a total of three scans for each. The spectra were corrected by subtracting the spectrum from the buffer background.

In vitro transcription and purification of RNA. The DNA template was synthesized from IDT and amplified by PCR. In vitro transcription was carried out using the HiScribe T7 High Yield RNA synthesis kit (NEB) according to the manufacturer's protocol. The transcribed RNA was loaded onto to the Resource-Q column (GE Healthcare) and eluted with a linear sodium chloride gradient (50 mM–1,000 mM) in 25 mM Tris–HCl, pH 8.0. The eluted RNA was concentrated and stocked at −80 °C.

In vitro RNA cleavage assay. Cas12g proteins (290 ng) purified from gel filtration were premixed with sgRNA at 37 °C for 30 min at a molar ratio of 4:1 in buffer C. Target RNA (160 ng; sequence: CUCAUGUUUGACAGCUUAUCAUCGAUAAGCUUUAAUGCGGUAGUUUAUCACAGUUAAAUU) synthesized from IDT was then added into the mixture for another 30 min unless otherwise specified. The reaction was quenched by adding EDTA and proteinase K. The products were denatured with TBE–urea loading buffer (ThermoFisher Scientific) at 66 °C for 10 min, and analyzed on a 15% TBE–urea gel and stained using SYBR Gold (Invitrogen).

For stoichiometric titration assay for the turnover studies, the Cas12g–sgRNA complex was incubated with target RNA at different molar ratios at 37 °C for 30 min. The experiment was done in triplicate. The resulting gels were scanned in gray mode and band intensities were measured by ImageJ[48].

Collateral cleavage assay. Collateral cleavage reactions were initiated by mixing a 267-nucleotide collateral ssRNA (215 ng; sequence: GGAUAUUAAUAGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUAAAUGAGGGUUAGUUUGACUGUAUAAAUACAGUCUUGCUUUCUGACCCUGGUAGCUGCUCACCCUGAUGCUGCUGUCAAUAGACAGGAUAGGUGCGCUCCCAGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGCUACUGAAUCACCCCCGAUCAAGGGGGAACCCUCCAAAAGGUGGGUUGAAAGGAGAAGUCAUUUAAUAAGGCCACUGUUAAA) or M13 ssDNA (315 ng, New England Biolabs) and different targets with the binary complex (wild type or lid mutant) at 42 °C for 60 min in buffer C. The reaction was quenched by adding EDTA and proteinase K. The collateral products were resolved on urea–PAGE (15%) or 0.7% agarose gel and were visualized by SYBR Gold staining (Invitrogen).

Size-exclusion chromatography. The pre-assembled binary Cas12g–sgRNA complex was incubated with target DNA, target RNA and target RNA with mismatches with molar ratios of 1:1.15. The resulting mixture was incubated at 37 °C for 30 min before being applied to a Superdex 200 column, which was equilibrated with buffer C. Urea–PAGE (15%) was used to analyze and monitor the complex formation.

Electron microscopy. Aliquots of 3-μl samples (Cas12g–sgRNA and $Cas12g^{E655A}$–sgRNA–target RNA) at different concentrations (0.3 mg ml^{-1} and 0.5 mg ml^{-1}, respectively) were applied to glow-discharged Quantifoil holey carbon grids (Cu, R1.2/1.3, 400 mesh). The grids were blotted for 4 s and plunged into liquid ethane with a Gatan Cryoplunge 3 plunger. Cryo-EM data were collected with a Titan Krios microscope (FEI) running at 300 kV and images were collected using Leginon[49] at a nominal magnification of ×81,000 (with a calibrated pixel size of 1.05 Å per pixel) with a defocus range from 1.2 μm to 2.5 μm. The images were collected with a K3 summit electron direct detector in super-resolution mode together with a GIF-Quantum energy filter working at a slit width of 20 eV. A dose rate of 20 electrons per pixel per second and an exposure time of 3.12 s were used, generating 40 movie frames with a total dose of ~54 electrons per $Å^2$. Appion[50] was used to do quick motion correction and contrast transfer function (CTF) estimation to monitor the data quality during data collection. A total of 2,520 and 987 movie stacks were collected for Cas12g–sgRNA and $Cas12g^{E655A}$–sgRNA–target RNA, respectively (Supplementary Table 1).

Image processing. The movie frames were imported to RELION-3 (ref. [51]) and then aligned using MotionCor2 (ref. [52]) with a binning factor of 2. CTF parameters of the motion-corrected images were estimated using Gctf[53]. For particle picking, two-dimensional (2D) averages were first generated on the basis of a few thousand particles picked without a template, and then used for subsequent template-based autopicking.

For Cas12g–sgRNA complex, particles were first autopicked and extracted from the dose-weighted micrographs. The dataset was split into batches for 2D classifications, which were used to remove false and bad particles that were assigned into 2D averages with poor features. Particles from different 2D averages were used to create the initial model in cryoSPARC[54]. The dataset was split into four parts for 3D classification. From 3D classification, three major classes can be discerned. Class I adopts a rigid structure with clear structural features, while classes II and III have poor structural features (Extended Data Fig. 1e). Particles in Class I were selected for 3D refinement, converging at a resolution of 3.4 Å. After Bayesian polishing, another round of 3D classification was performed to exclude bad particles. The resulting dataset with 469,157 particles was used for CTF refinement and 3D refinement, converging at a resolution of 3.1 Å. Focused 3D classification around the REC1 domain was further performed without alignment resulting in a class with this region observed from ~10% of particles.

For the $Cas12g^{E655A}$–sgRNA–target RNA dataset, a similar strategy was applied. 3D refinement was performed against the final dataset containing 38,238 particles, converging at a resolution of 4.8 Å (Extended Data Fig. 6). Cryo-EM image processing is summarized in Supplementary Table 1. For the $Cas12g^{E655A}$–sgRNA–target DNA dataset, 3D refinement was performed against 53,612 particles, converging at a resolution of 5.8 Å (Extended Data Fig. 9).

Model building, refinement and validation. De novo model building of the Cas12g–sgRNA structure was performed manually in Coot[55]. Secondary-structure predictions by PSIPRED[56] and the structure of Cas12b (PDB: 5U34) were used to assist manual building. Refinement of the structure models against corresponding maps was performed using the phenix.real_space_refine tool in Phenix[57]. For the $Cas12g^{E655A}$–sgRNA–target RNA complex, the structure model of the Cas12g–sgRNA binary complex was fitted into the cryo-EM map, and each domain was manually adjusted in Coot, and the crRNA–target RNA duplex was built as a standard A-form double helix. The resultant model was refined against the corresponding cryo-EM map using the phenix.real_space_refine tool in Phenix. 3D FSC analysis for the presented maps was performed using the Remote 3D FSC Processing Server (https://3dfsc.salk.edu/upload/). Map-to-model FSC analysis was performed using the phenix.validation_cryoem tool in Phenix.

Sequence alignment and structural visualization. The PROMALS3D[58] program was used to align the Cas12g and Cas12b sequences on the basis of their structures. Figures were prepared using PyMOL and UCSF Chimera[59].

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

Data availability

Cryo-EM reconstructions of Cas12g–sgRNA and Cas12g–sgRNA-target RNA complexes have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-22257 and EMD-22258, respectively. Coordinates for atomic models of Cas12g–sgRNA and Cas12g–sgRNA-target RNA complexes have been deposited in the Protein Data Bank under accession numbers 6XMF and 6XMG, respectively. Source data are provided with this paper.

References

48. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
49. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
50. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
51. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
52. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
53. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
54. 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).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
56. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
57. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 531–544 (2018).
58. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
59. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

We thank T. Klose and V. Bowman for help with cryo-EM, and S. Wilson for computation. This work made use of the Purdue Cryo-EM Facility. This work was supported by the NIH grant R01GM138675 and a Showalter Trust Research Award to L.C.

Author contributions

L.C. supervised the study. H.Z., R.X. and Z.L. prepared samples. Z.L., H.Z. and L.C. collected and processed cryo-EM data. Z.L., H.Z., R.X. and R.H. performed biochemical analysis. Z.L., H.Z. and L.C. prepared figures. All authors analyzed the data. Z.L., H.Z. and L.C prepared the manuscript with input from R.X. and R.H.

Competing interests

The authors declare no competing interests.

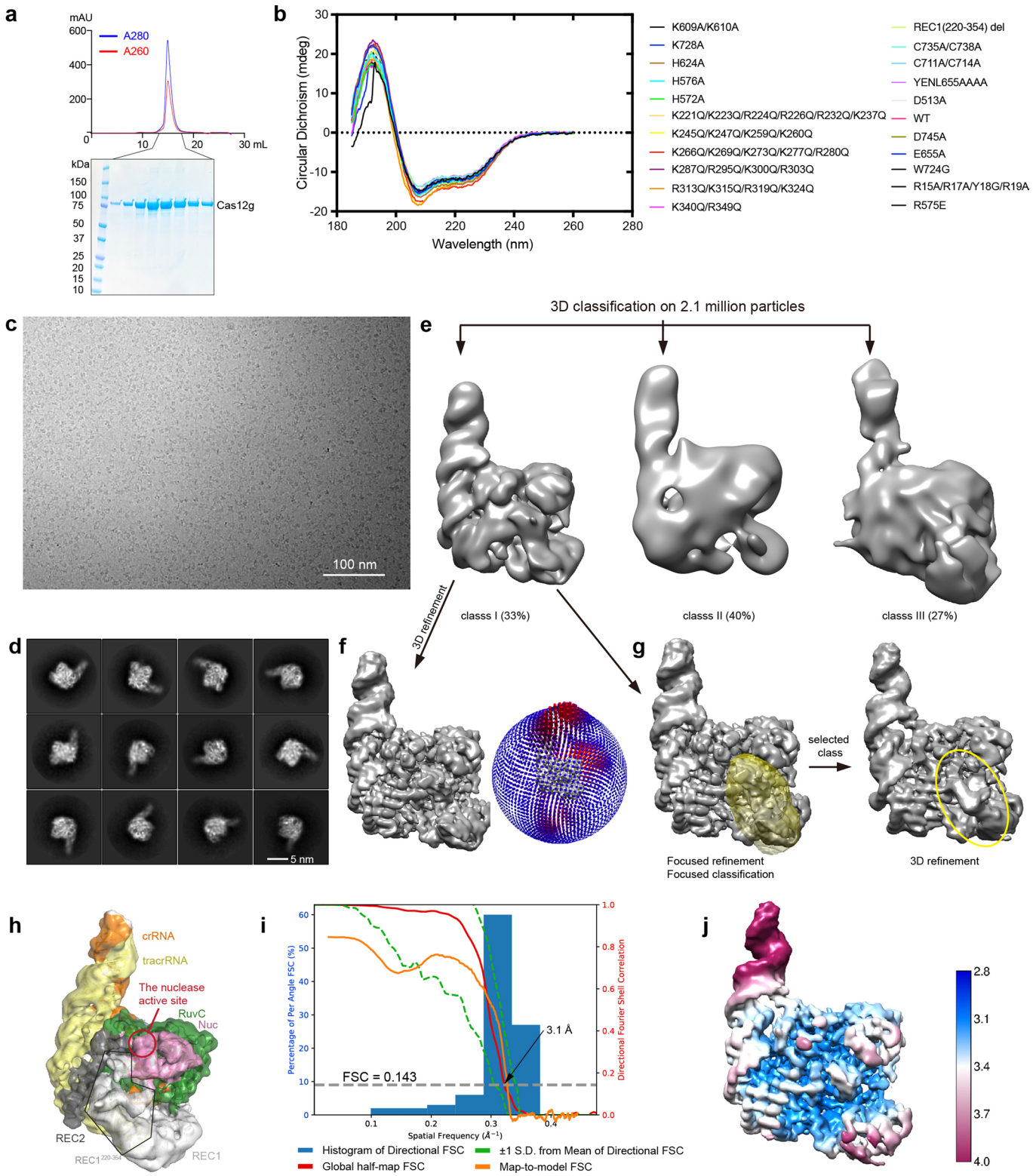
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-020-00721-2.

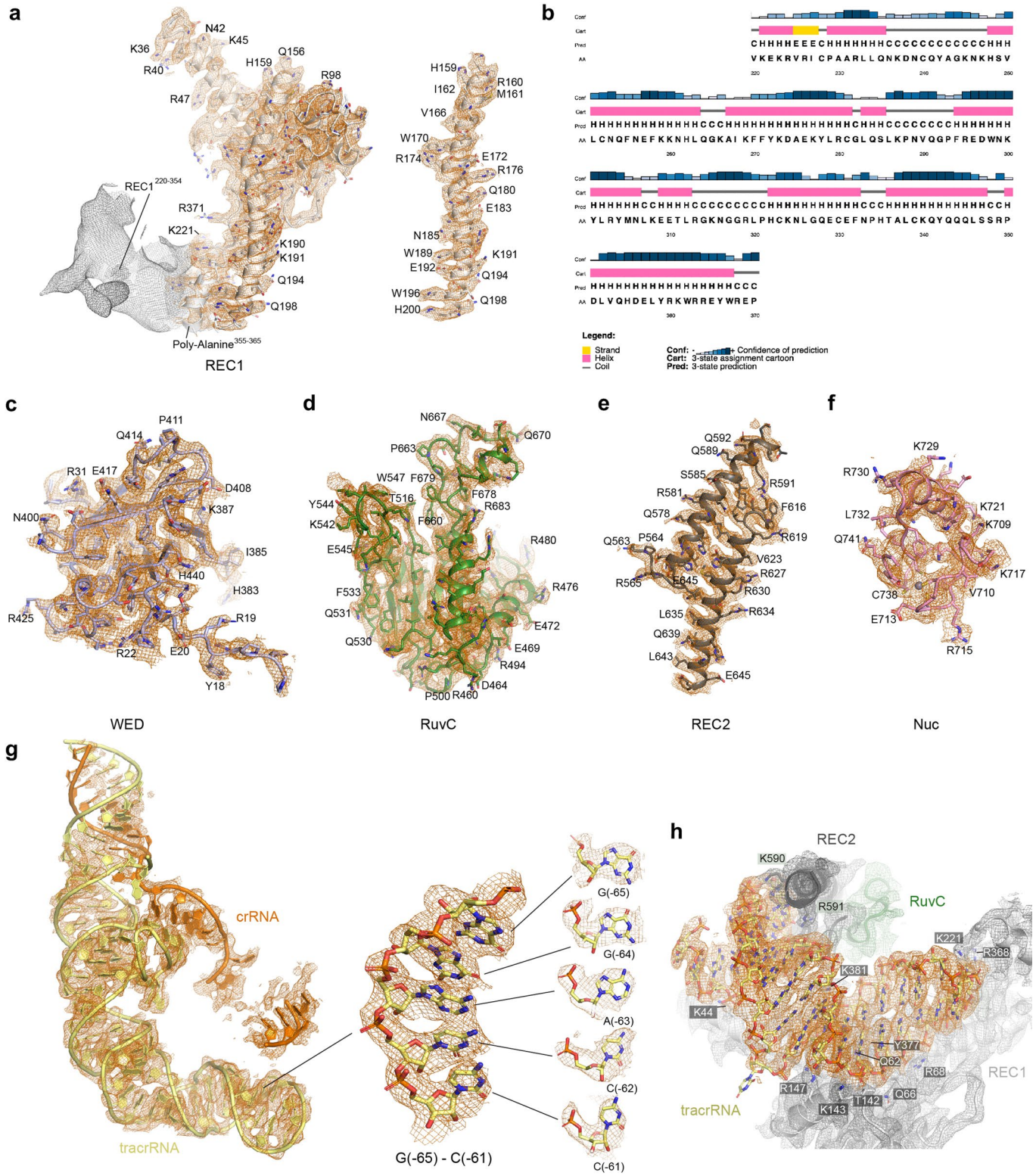
Supplementary information is available for this paper at https://doi.org/10.1038/s41589-020-00721-2.

Correspondence and requests for materials should be addressed to L.C.

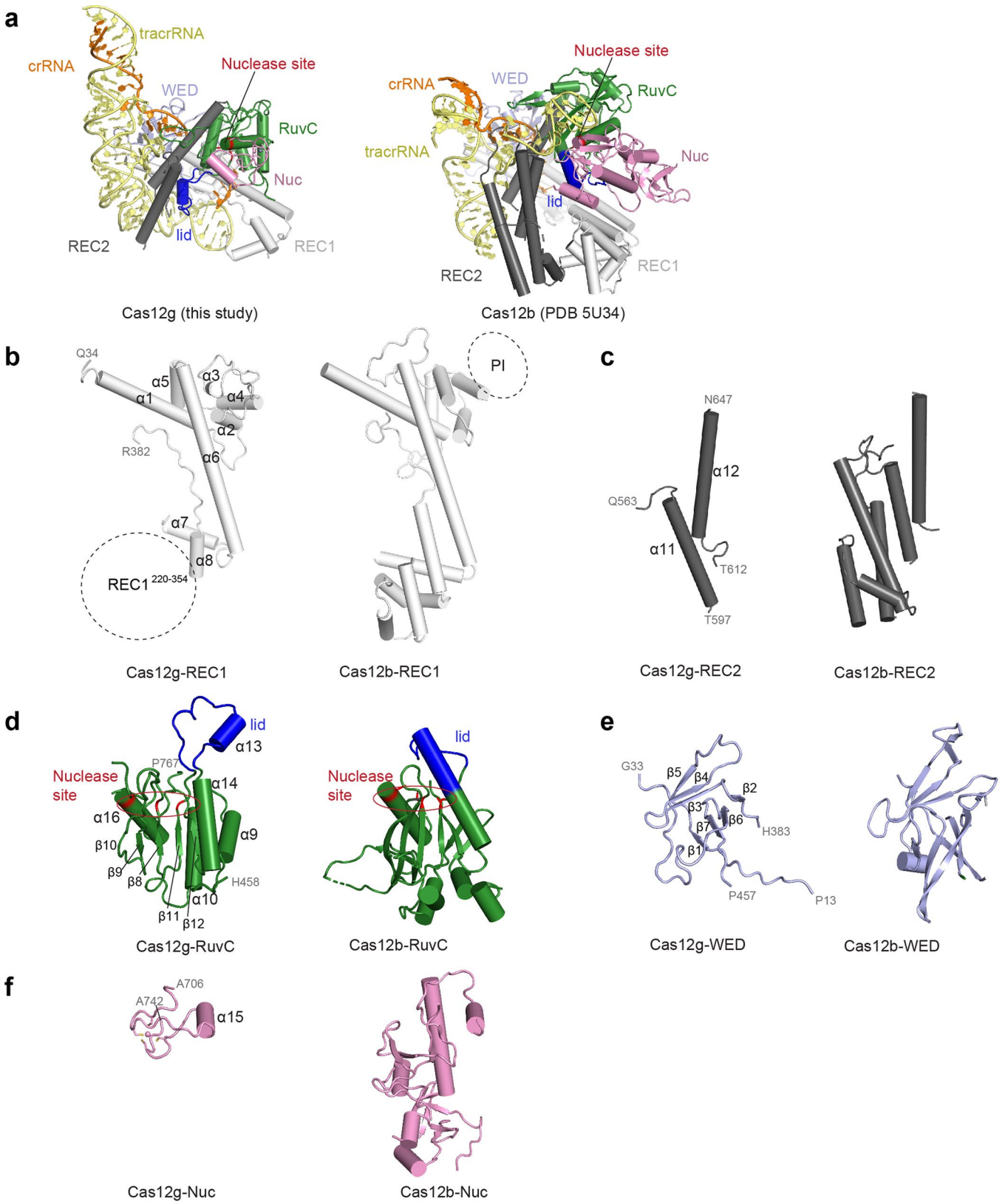
Reprints and permissions information is available at www.nature.com/reprints.



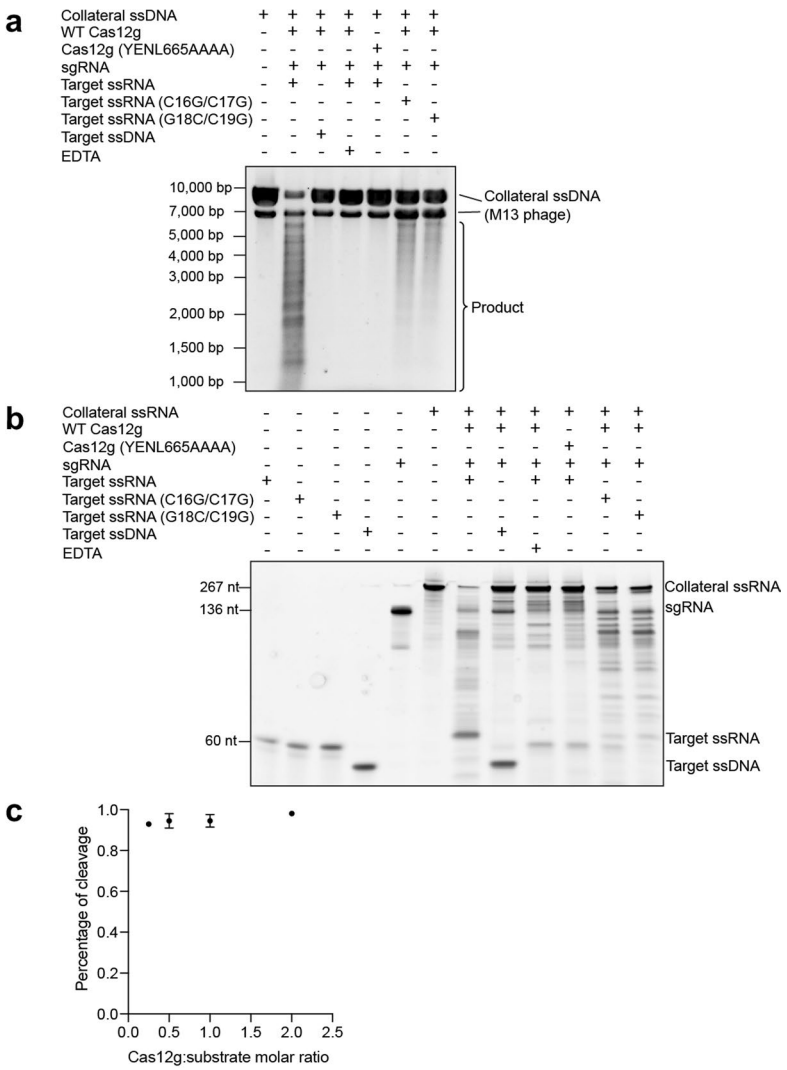
Extended Data Fig. 1 | Sample preparation and cryo-EM of the Cas12g-sgRNA binary complex. a, Purification of Cas12g. Upper: Size exclusion chromatography (SEC) profile. UV absorbance curves at 280 nm and 260 nm are shown in blue and red, respectively. Lower: SDS-PAGE analysis. **b**, Circular dichroism (CD) spectra of wild-type and mutants of Cas12g used in this study. **c**, A representative raw cryo-EM micrograph of the Cas12g-sgRNA complex from a total of 2520 micrographs. **d**, Representative 2D class averages from a total of 100 images. **e**, Three major classes from 3D classification. **f**, 3D refinement for Class I with angular distribution of particles. **g**, Focused classification around the flexible region of the REC1 domain (REC1$^{220-354}$). **h**, Cryo-EM map of class I with each domain of Cas12g and sgRNA colored coded. **i**, Plot of the global half-map FSC (solid red line), map-to-model FSC (solid orange line), and spread of directional resolution values (±1σ from mean, green dotted lines; the blue bars indicate a histogram of 100 such values evenly sampled over the 3D FSC). FSC plot for the reconstruction suggests an average resolution of 3.1 Å. **j**, Local resolution map for the reconstruction in **f**.



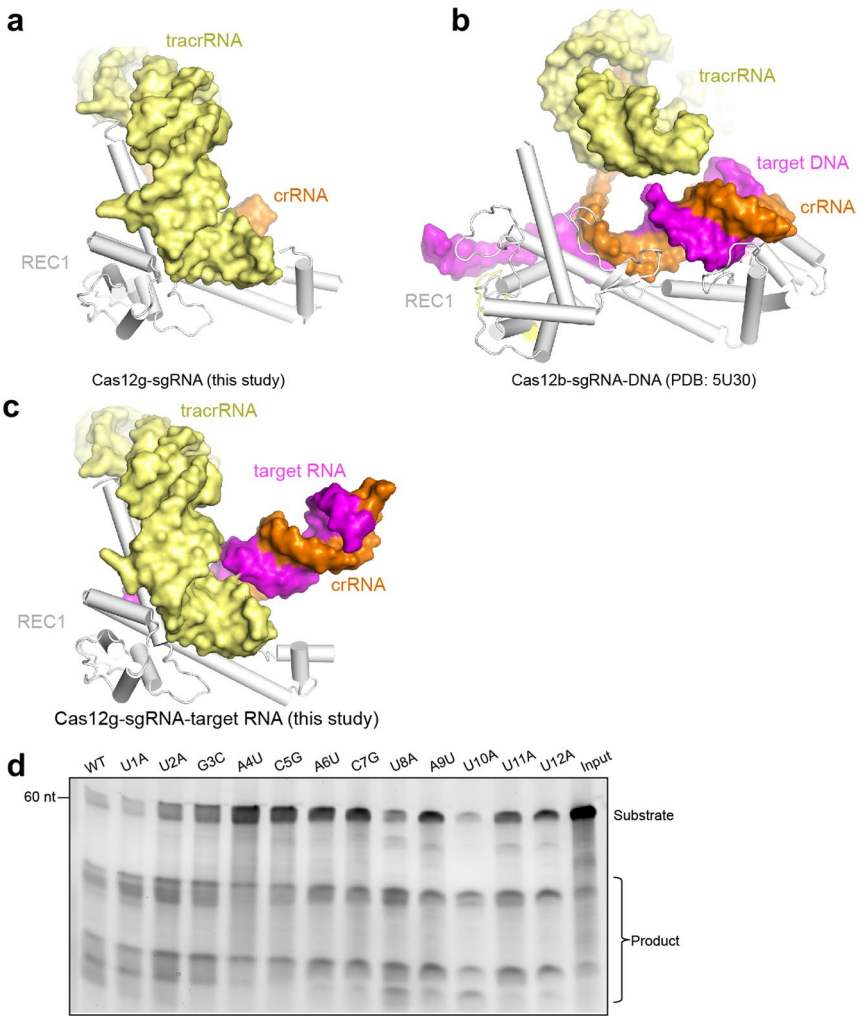
Extended Data Fig. 2 | Detailed cryo-EM density map of the Cas12g-sgRNA complex with atomic model fitted in. a, Fitting of the REC1 domain, with density of REC1$^{220-354}$ from focused classification shown in grey mesh. A representative α-helix from the REC1 domain is shown in detail on the right. **b**, Secondary structure prediction of REC1$^{220-354}$ using PSIPRED. **c**, Fitting of the WED domain. **d**, Fitting of the RuvC domain. **e**, Fitting of the REC2 domain. **f**, Fitting of the Nuc domain. **g**, Fitting of nucleic acids to the corresponding cryo-EM map, with details of a segment shown as indicated on the right. The atomic models are shown in stick with crRNA and tracrRNA regions colored in orange and yellow, respectively. **h**, Detailed interactions between scaffolding duplex 1 and Cas12g. Cryo-EM density map is shown in mesh.



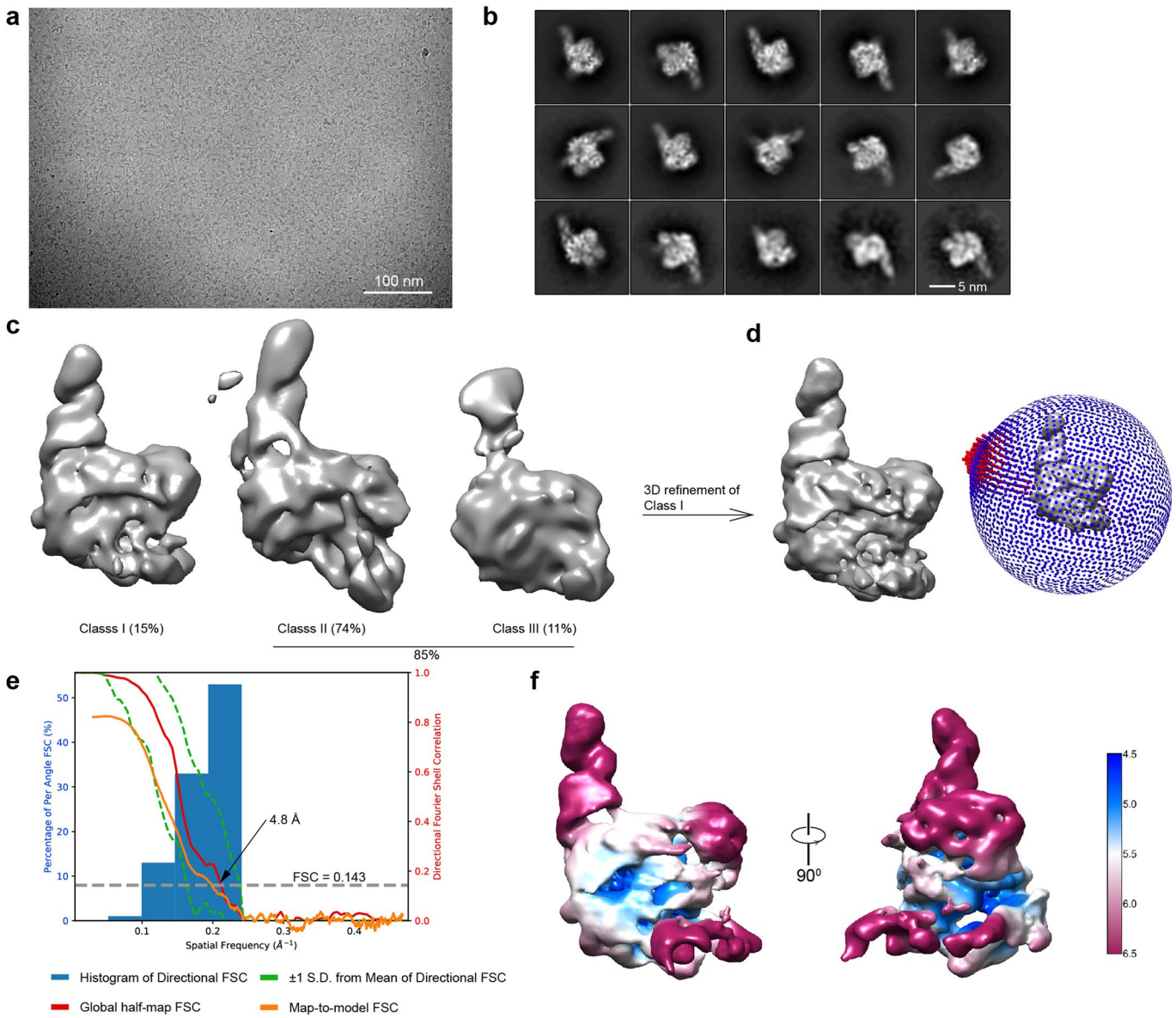
Extended Data Fig. 3 | Structural comparison between Cas12g and Cas12b. a, Overall structures of the Cas12g-sgRNA and the Cas12b-sgRNA complexes. **b**-**f**, Structural comparison of each domain. The structures are aligned by secondary-structure matching (SSM) in COOT.



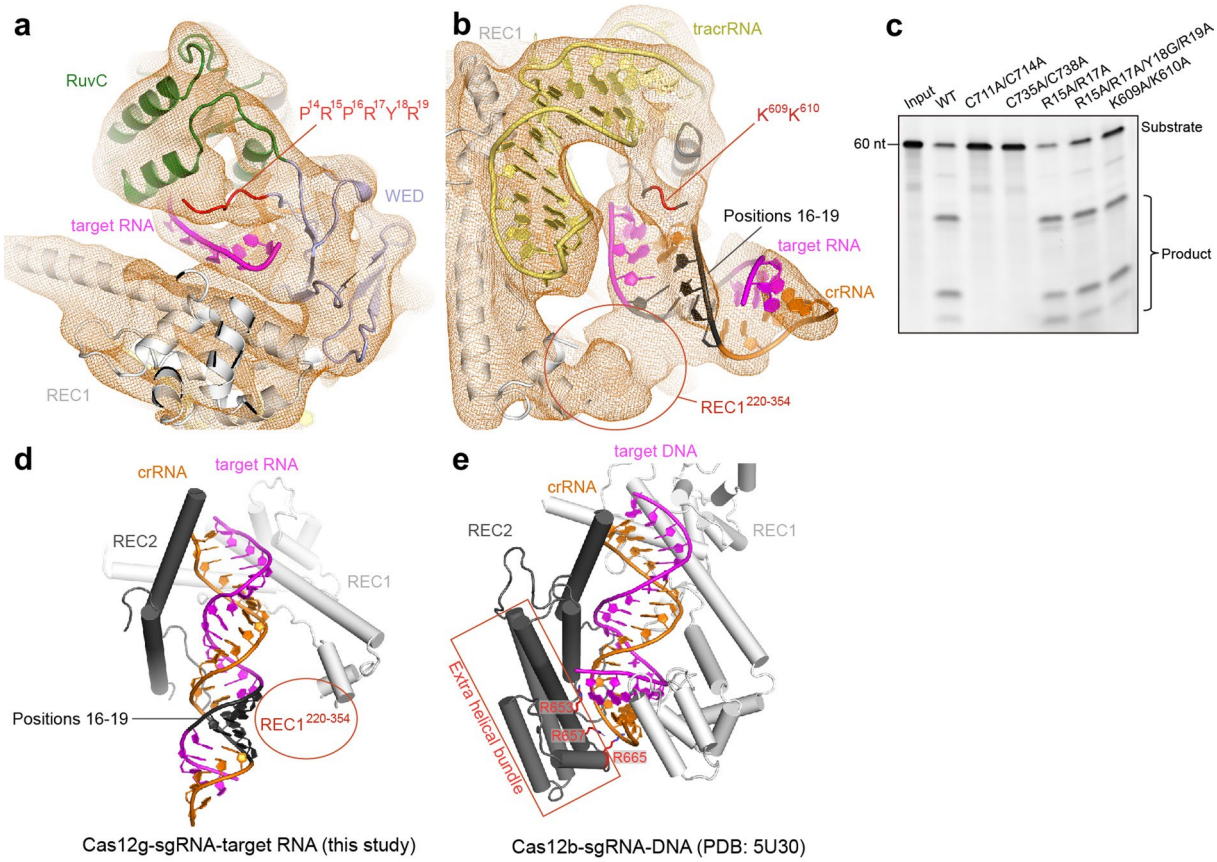
Extended Data Fig. 4 | Collateral cleavage and stoichiometric titration assays for Cas12g. a, Collateral cleavage of unrelated ssDNA by Cas12g. The results shown are representative of three experiments. **b**, Collateral cleavage of unrelated ssRNA by Cas12g. The results shown are representative of three experiments. **c**, Stoichiometric titration assay indicates that Cas12g is a multiple turnover enzyme towards RNA substrate. Each point represents the cleavage ratio after 30 minutes at 37 °C. Error bars represent mean ± SD, where n = 3 replicates.



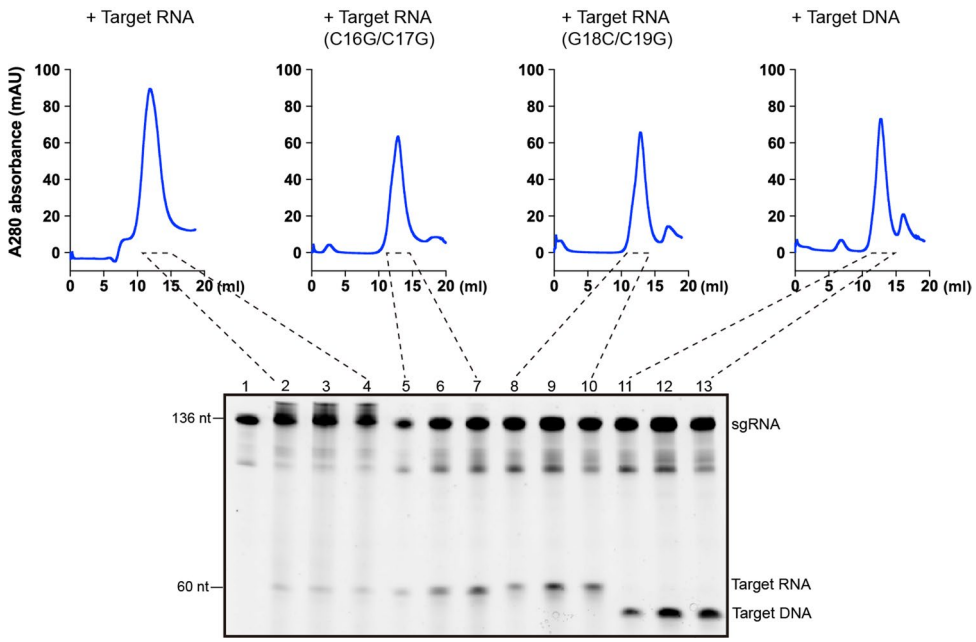
Extended Data Fig. 5 | Comparison of the REC1 domain in Cas12g and Cas12b, and crRNA guide mismatch assay for Cas12g. a, The tracrRNA scaffolding duplex 1 is bound to the concave surface of the REC1 domain in the Cas12g-sgRNA binary complex. **b**, The crRNA-target DNA duplex is bound to the concave surface of the REC1 domain in the Cas12b-sgRNA-DNA complex. **c**, The tracrRNA scaffolding duplex 1 is bound to the concave surface of the REC1 domain in the Cas12g-sgRNA-target RNA complex. **d**, RNA cleavage assay using wild-type and mutant target RNAs. The results shown are representative of three experiments.



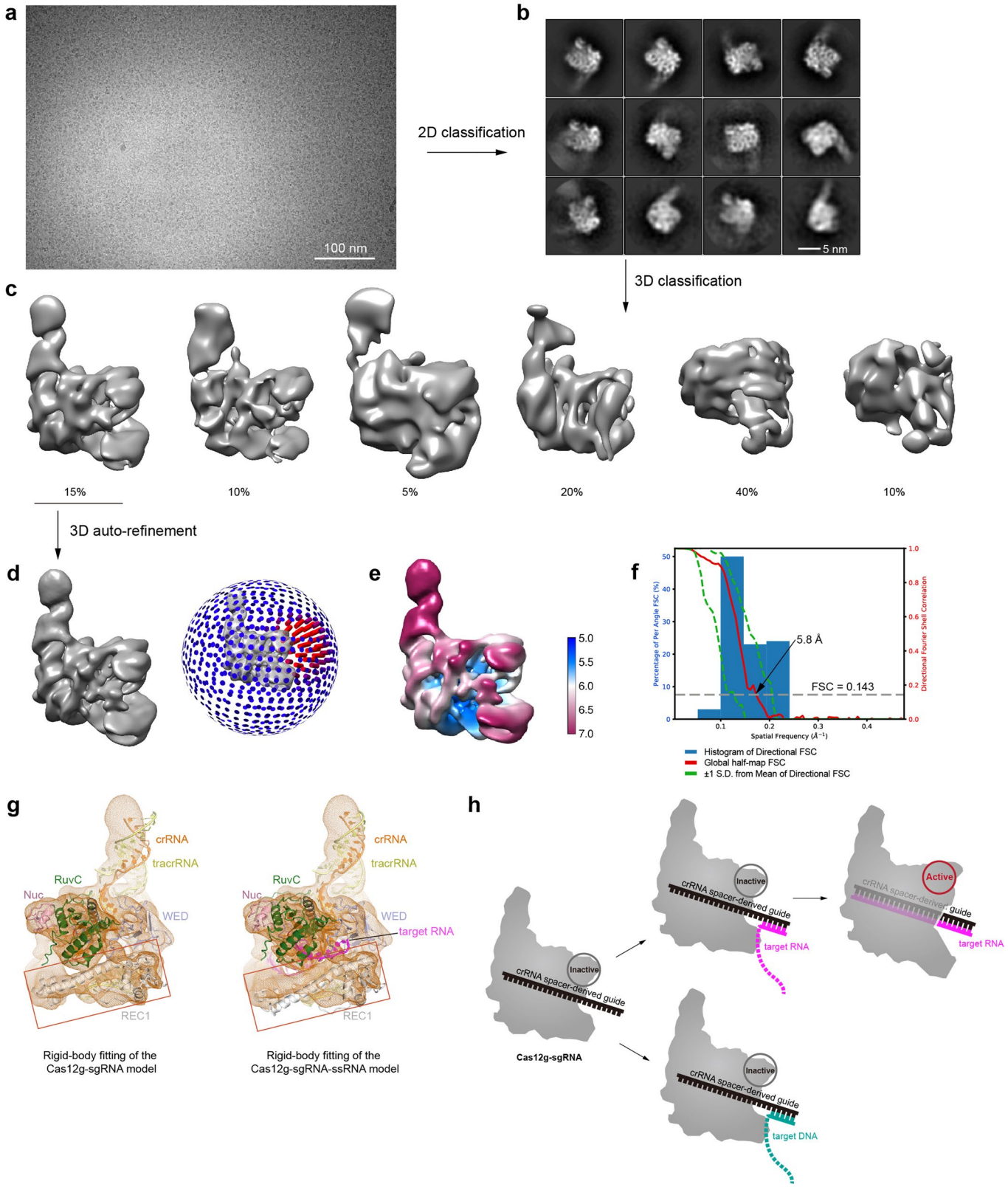
Extended Data Fig. 6 | Cryo-EM data processing for the Cas12g-sgRNA-target RNA complex. a, A representative raw cryo-EM micrograph of the Cas12g-sgRNA-target RNA complex from a total of 987 micrographs. **b**, Representative 2D class averages from a total of 100 images. **c**, 3D classification. Three major classes were observed. Classes II and III did not result in good reconstructions and likely represent damaged particles. **d**, 3D refinement for Class I. Angular distribution is shown on the right. **e**, Plot of the global half-map FSC (solid red line), map-to-model FSC (solid orange line), and spread of directional resolution values (±1σ from mean, green dotted lines; the blue bars indicate a histogram of 100 such values evenly sampled over the 3D FSC). FSC plot for the reconstruction suggests an average resolution of 4.8 Å. **f**, Local resolution map in two views.



Extended Data Fig. 7 | Structural and biochemical analysis of the Cas12g-sgRNA-target RNA complex. a-**b**, Structural elements involved in crRNA-target RNA duplex recognition. Three such structural elements are indicated by red lines, including 1) an N-terminal extension enriched in arginines within the WED domain, 2) a loop within the REC2 domain (608-AKKATLQP-615), and 3) $REC1^{220-354}$. **c**, Substrate RNA cleavage assay using wild-type Cas12g and Cas12g with mutations in two loops involved in the recognition of the crRNA-target RNA duplex. The results shown are representative of three experiments. **d**-**e**, Comparison of the REC1 and REC2 domains for substrate recognition between Cas12g and Cas12b.



Extended Data Fig. 8 | Binding assay between the Cas12g-sgRNA complex and target nucleic acids. Upper: SEC profiles for the Cas12g-sgRNA complex incubated with different targets including target RNA, target RNA with mismatches at positions 16 and 17, target RNA with mismatches at positions 18 and 19, and target ssDNA. UV absorbance curves at 280 nm are shown. Lower: Urea-PAGE (15%) analysis of the elution fractions from SEC as indicated. The results shown are representative of three experiments.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Cryo-EM analysis for the Cas12g-sgRNA-target DNA complex. a, A representative raw cryo-EM micrograph of the Cas12g-sgRNA-target DNA complex from a total of 1083 micrographs. **b**, Representative 2D class averages from a total of 100 images. **c**, 3D classification. Classes 2-6 did not result in good reconstructions and likely represent damaged particles. **d**, 3D refinement for Class 1 with angular distribution. **e**, Plot of the global half-map FSC (solid red line), and spread of directional resolution values ($\pm 1\sigma$ from mean, green dotted lines; the blue bars indicate a histogram of 100 such values evenly sampled over the 3D FSC). FSC plot for the reconstruction suggests an average resolution of 5.8 Å. **f**, Local resolution map. **g**, Comparison of the rigid-body fitting between the Cas12g-sgRNA complex and the Cas12g-sgRNA-target RNA complex. Red squares indicate the fitting of the REC1 domain, which shows that the model of the Cas12g-sgRNA complex fits better to the map (left panel). **h**, Schematic model for Cas12g activation. Target DNA fails to form fully assembled duplex required for Cas12g activation.

nature research

Corresponding author(s): Leifu Chang

Last updated by author(s): Nov 7, 2020

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 our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Leginon 3.4, Image Quant 1.2
Data analysis	Appion 3.3, Relion 3.0, Gctf 1.06, Motioncor2, cryoSPARC v2.15.0, PyMOL 2.2.3, UCSF Chimera 1.13, Coot 0.8.9.1, Phenix 1.12, ImageJ 1.48, PROMALS3D.

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

Cryo-EM reconstruction of Cas12g-sgRNA binary complex and Cas12g-sgRNA-target RNA ternary complex have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-22257, EMD-22258 respectively. Coordinates for atomic models of Cas12g-sgRNA binary complex and Cas12g-sgRNA-target RNA ternary complex have been deposited in the Protein Data Bank under the accession numbers 6XMF and 6XMG respectively. Source data are provided with this paper.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No sample-size calculation was performed. The actual sample size was increased so that desired structure features were obtained in cryo-EM maps.
Data exclusions	No data were excluded.
Replication	All biochemical experiments were carried out as triplicates, and were readily and reliably reproduced. For stoichiometric titration assay, the resulting mean +/- deviation are shown as error bars.
Randomization	Since no human or animal subject were used in this study, no randomization was performed.
Blinding	Blinding was not relevant to this study, since no manual counting or scoring was performed to obtain data.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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