

Cas12e orthologs evolve variable structural elements to facilitate dsDNA cleavage

Received: 22 March 2024

Accepted: 11 November 2024

Published online: 30 December 2024

Danyuan Li^{1,2,6}, Shouyue Zhang^{3,4,6}, Shuo Lin^{1,2,6}, Wenjing Xing^{1,6}, Yun Yang^{1,2}, Fengxia Zhu⁵, Dingding Su⁵✉, Chunlai Chen¹✉ & Jun-Jie Gogo Liu^{1,2}✉

Exceptionally diverse type V CRISPR-Cas systems provide numerous RNA-guided nucleases as powerful tools for DNA manipulation. Two known Cas12e nucleases, DpbCas12e and PlmCas12e, are both effective in genome editing. However, many differences exist in their *in vitro* dsDNA cleavage activities, reflecting the diversity in Cas12e's enzymatic properties. To comprehensively understand the Cas12e family, we identify and characterize six unreported Cas12e members that vary in their CRISPR-locus architectures, PAM preferences, and cleavage efficacies. Interestingly, among all variants, PlmCas12e exhibits the most robust *trans*-cleavage activity and the lowest salt sensitivity in *cis*-cleavage. Further structural comparisons reveal that the unique NTSB domain in PlmCas12e is beneficial to DNA unwinding at high salt concentrations, while some NTSB-lacking Cas12e proteins rely on positively charged loops for dsDNA unwinding. These findings demonstrate how divergent evolution of structural elements shapes the nuclease diversity within the Cas12e family, potentially contributing to their adaptations to varying environmental conditions.

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated proteins) functions as a defense system in bacteria, archaea, and viruses against invading genetic material^{1–3}. The CRISPR interference effector typically consists of two components: non-coding CRISPR RNA (crRNA) and Cas proteins. crRNA contains a sequence complementary to the target DNA sequence known as the protospacer, which is adjacent to the protospacer adjacent motif (PAM). The protospacer acts as a memory bank of infection events^{4,5}. Via complexing with Cas proteins, crRNA serves as the guide RNA (gRNA) and directs the effector to the complementary target, allowing the effector to invade and cleave the target^{1,6–8}. In some systems, a longer non-coding RNA, *trans*-activating CRISPR RNA (tracrRNA) derived from the CRISPR locus, is essential for target interference via forming the hybrid guide with crRNA^{9,10}. Over the last decade, *in-silico*

analyses have boosted the diversity of CRISPR systems, especially type V systems, which employ a single conserved RuvC nuclease domain for target cleavage. To date, double-stranded DNA (dsDNA) cleavage activity has been experimentally validated in 11 subtypes of the type V family, including Cas12a-b, Cas12d-f, Cas12h-j, Cas12l, Cas12n, and Cas12 λ ^{3,11–18}. Cas12 nucleases predominantly recognize 5' T-rich PAMs and cleave the non-target strand (NTS) and target strand (TS) at the PAM distal end, forming a staggered end. Notably, some Cas12 proteins, including Cas12l, Cas12f, and Cas12m recognize 5'-CCN PAMs^{15,17,19,20}, while Cas12n recognizes a 5'-AAN PAM¹⁸. Additionally, activated by a specific target sequence, most active Cas12 nucleases are able to degrade the non-specific ssDNA *in trans*. This characteristic has led to the development of Cas12 as a nucleic acid detection platform^{21–23}.

¹Beijing Frontier Research Center for Biological Structure, State Key Laboratory of Membrane Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China. ³CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. ⁴College of Life Science, University of Chinese Academy of Sciences, Beijing, China. ⁵Peking University Institute of Advanced Agricultural Sciences, Weifang 261325, China. ⁶These authors contributed equally: Danyuan Li, Shouyue Zhang, Shuo Lin, Wenjing Xing. ✉e-mail: dingding.su@pku-iaas.edu.cn; chunlai@mail.tsinghua.edu.cn; junjiegogoliu@tsinghua.edu.cn

CRISPR-Cas systems have become powerful tools for genome editing due to their simplicity and high efficiency²⁴. A subtype of the type V family, the CRISPR-Cas12e (also known as CasX) system has been reported with high editing efficacy in human cells¹⁴. So far, only two Cas12e nucleases, *Deltaproteobacteria* Cas12e (DpbCas12e) and *Planctomycetes* Cas12e (PlmCas12e) have been biochemically and structurally characterized²⁵. They recognize a 5'-(T)TCN PAM and cleave the NTS and TS sequentially. While these two Cas12e nucleases share a 68.5% protein sequence identity, they diverge significantly in biochemical properties, indicating a correlation between structures and catalytic activities²⁶. Based on the overall structural similarity, domain switches between two Cas12e proteins, along with gRNA modifications, notably improve the genome editing efficacy in human cells²⁶. This underscores the importance of comprehending the interrelation between structures and activities, providing valuable insights for the fine-tuning and enhancement of CRISPR-Cas systems as more effective manipulation tools. However, a comprehensive understanding of the Cas12e family remains elusive.

Here, we screen and identify six unreported Cas12e orthologs, which recognize 5' T- or C- rich PAMs. They all could interfere with plasmids in *Escherichia coli* (*E. coli*) guided by a single guide RNA (sgRNA) combining crRNA and tracrRNA. Notably, Cas12e enzymes exhibit higher activity under low salt concentration compared to high salt concentration. Among the characterized Cas12e nucleases, PlmCas12e and DpbCas12e, which possess unique non-target strand binding (NTSB) domains beneficial for dsDNA unwinding, demonstrate the most robust *cis*- and *trans*-cleavage activities and remain effective across a wide range of salt concentrations. In contrast, Cas12e orthologs lacking NTSB domains, represented by *Verrucomicrobia* Cas12e (VemCas12e) and *Lentisphaerae* bacterium Cas12e (LesCas12e), exhibit lower activities and are more vulnerable to high salt concentration, which was shown to hinder R-loop formation. Structural analyses and biochemical assays on NTSB-containing and -lacking Cas12e enzymes

further reveal the importance of the NTSB domain in securing the dsDNA cleavage under high salt conditions. These results provide valuable insights into the cleavage mechanisms of Cas12e family members, emphasizing how the nucleases attain enhanced cleavage activity through the evolution of gaining distinct structural elements.

Results

Cas12e orthologs recognize varied PAMs and are active for DNA interference

CRISPR-Cas12e subtype was initially reported with two orthologous systems, PlmCas12e and DpbCas12e, both containing *Cas1*, *Cas2*, and *Cas4* in the CRISPR locus. Based on these two Cas12e nucleases, we constructed Hidden Markov Models (HMM) and identified six unreported candidates from metagenomic databases, named *Planctomycetota* bacterium 2 Cas12e (Plm2Cas12e), *Candidatus Kerfeldbacteria* bacterium Cas12e (CkbCas12e), *Opitutae* bacterium Cas12e (OpbCas12e), *Lentisphaerae* bacterium Cas12e (LesCas12e), *Verrucomicrobia* Cas12e (VemCas12e), and bacterium HR35 Cas12e (HrbCas12e). Phylogenetic analysis of these Cas12e orthologs with representative reported type V effectors revealed that all orthologs clustered with the two known PlmCas12e and DpbCas12e proteins (Fig. 1a). Comparative analyses of predicted structures and sequence alignments inform that these Cas12e orthologs contain similar domain arrangements as PlmCas12e and DpbCas12e, except for the counterpart of the NTSB domain, which varies in size among orthologs (Fig. 1b). In contrast to the homology in Cas effector, the genomic architectures of these CRISPR-Cas12e systems exhibit various patterns (Fig. 1c). In the genomic loci of VemCas12e, LesCas12e, CkbCas12e, and Plm2Cas12e, tracrRNA-encoding region is located between the *Cas2* gene and the CRISPR array, while for HrbCas12e locus, tracrRNA-encoding region is located between the *Cas12e* and adaptation modules (*cas1*, *cas2*, and *cas4* genes). The CRISPR-OpbCas12e locus lacks an adaptation module and its tracrRNA-encoding region locates between *Cas12e* and the

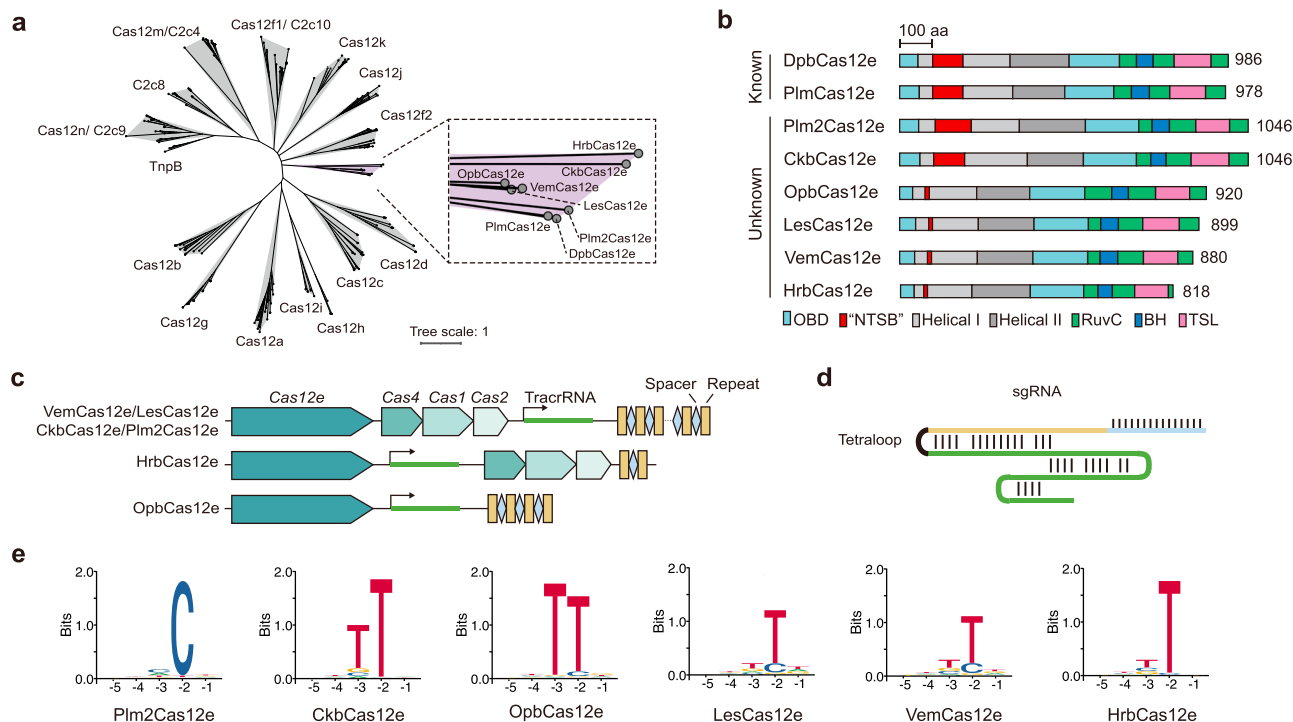


Fig. 1 | Identification of phylogenetic branches of the CRISPR-Cas12e system. **a** Phylogenetic analysis of Cas12e orthologs in this study and representative reported type V effectors. The enlarged dashed box illustrates Cas12e proteins positioned in different branches. **b** Size and domain comparisons of Cas12e effector proteins based on AlphaFold-predicted structures. **c** In-silico prediction of

tracrRNA and crRNA based on base pairing between tracrRNA and crRNA across the CRISPR-Cas locus. Arrows indicate the predicted transcription start sites for tracrRNAs. **d** A single guide RNA (sgRNA) is obtained by fusing tracrRNA and crRNA with a tetraloop. **e** PAM screening results of Cas12e orthologs. Source data are provided as a Source Data file.

CRISPR array. A recent study also reported several Cas12e proteins classified as Cas12e1 and Cas12e2 clades²⁷, displaying different genomic architectures from the systems identified in this study (Supplementary Fig. 1a). Potentially, the genomic loci of the Cas12e family are highly active, thereby resulting in significant evolutionary diversity. Nevertheless, these systems all feature the tracrRNA components and are highly conserved at the 3' termini of their direct repeats (Supplementary Fig. 1b).

To assess the homology of Cas12e proteins, we performed a phylogenetic analysis and comparative structural analysis. The results demonstrated that Cas12e orthologs could be clustered into four major branches, each containing at least one variant characterized in our study (Supplementary Fig. 1c). In terms of predicted atomic structures, Cas12e proteins can generally be divided into two types: one type is characterized by a significant “NTSB” counterpart, while the other type lacks it (Supplementary Fig. 1c). To further explore the nuclease activities of Cas12e orthologs in different types, we first adapted an *in vitro* cleavage-based method to identify the PAMs²⁸. For each Cas12e system, the sgRNA generated via fusing crRNA and tracrRNA with a 5′-“UUCG” tetra loop was used for the following assays (Fig. 1d). The results showed that five out of six Cas12e nucleases recognized canonical T-rich PAMs (Fig. 1e). Notably, PlmCas12e preferred a 5′-CN PAM, which was also observed in several other Cas12 nucleases like Cas12I, Cas12f and Cas12m^{15,17,19}. To further validate the PAM preference, we performed additional PAM depletion assays for VemCas12e and LesCas12e in *E. coli*. Both nucleases showed the 5′-TTG-PAM preference (Supplementary Fig. 2a–c), and this TTG PAM exhibited the highest fold-change increase in the previous *in vitro* screening assays (Supplementary Fig. 2d, e).

We next tested if all these CRISPR-Cas12e systems could interfere with plasmids in bacteria like PlmCas12e and DpbCas12e²⁵. We co-expressed the Cas12e orthologs and sgRNAs using a single plasmid in *E. coli*, which contained a pre-transformed and inducible ccdB-encoding plasmid designed to kill the host in the presence of 10 mM arabinose in the medium (Supplementary Fig. 3a)¹⁷. We observed that all eight Cas12e systems could be guided to target ccdB-coding plasmid in bacteria, as indicated by the more survival clones in the targeting group than the non-targeting group (Supplementary Fig. 3b). Cleavage on the plasmid with a single PAM revealed that cutting on TS occurred mainly after the 20th to 23rd nucleotide (nt) (Supplementary Fig. 3c, d). The cutting on NTS predominantly occurred within the spacer complementary region (20 bp), with a portion of the cleavage events occurring after 21-bp position to 29-bp position, possibly resulting from the nonspecific ssDNA cleavage activity of the RuvC domain on the locally unwound plasmid DNA^{29,30}.

Cas12e orthologs show diverse DNA cleavage activities at varying salt concentrations

To further evaluate the biochemical diversity of Cas12e orthologs, we tested their *cis*-cleavage and *trans*-cleavage activities *in vitro*. The *cis*-cleavage of 1-kilobase (kb) dsDNA demonstrated the specific DNA cleavage activity of the Cas12e orthologs, albeit with varying effectiveness (Fig. 2a). Additionally, these orthologs varied when cutting ssDNA in *trans* (Fig. 2b and Supplementary Fig. 4a). Among them, PlmCas12e showed previously unobserved *trans*-cleavage activity triggered by dsDNA activators²⁶. As indicated by different research on the biochemical activities of Cas9 or Cas12, the cleavage activity of a nuclease can be altered according to the reaction conditions^{17,31}. We noted that a significant difference between this study and previous studies is the salt concentrations, with 400 mM KCl used in the previous studies and 150 mM NaCl in this study. Therefore, we further assessed the *trans*-cleavage activity at different salt concentrations and found that decreasing the salt concentration could enhance the *trans*-cleavage activity of PlmCas12e (Fig. 2c and Supplementary Fig. 4b). Notably, changes in salt concentration had similar effects on *trans*-

cleavage activity for DpbCas12e (Supplementary Fig. 4c). Moreover, LbCas12a, known for its remarkably robust *trans*-cleavage activity²¹, also exhibited enhanced *trans*-cleavage activity with decreased salt concentrations (Supplementary Fig. 4d). Similar phenomenon for LbCas12a was also observed in other studies^{32,33}.

Leveraging a fluorophore-quencher reporter system, we confirmed that low salt concentration can improve *trans*-cleavage rates of PlmCas12e in a multi-turnover reaction assay (reporter substrates: activated RNP = 50: 1, Supplementary Fig. 5a). To evaluate the catalytic efficacy of *trans*-cleavage by PlmCas12e, we measured the enzyme kinetics of PlmCas12e, using LbCas12a as a reference (Supplementary Fig. 5b, c). Specifically, PlmCas12e demonstrated an enzyme turnover rate (k_{cat}) and Michaelis-Menten constant (K_m) of 0.47 s⁻¹ and 1.57 × 10⁻⁶ M, respectively, with a catalytic efficiency (k_{cat}/K_m) of 2.98 × 10⁵ M⁻¹ s⁻¹ activated by target dsDNA (Fig. 2d and Supplementary Fig. 5d). For LbCas12a, we measured a k_{cat} and K_m of 0.79 s⁻¹ and 1.41 × 10⁻⁷ M, respectively, and a k_{cat}/K_m of 5.65 × 10⁶ M⁻¹ s⁻¹ (Supplementary Fig. 5e). The difference in K_m indicated that PlmCas12e exhibits diminished binding affinity to *trans*-ssDNA, which is a major factor contributing to its lower *trans*-cleavage activity compared to LbCas12a. Therefore, Cas12e holds the potential to be manifested as a nucleic acid detection platform at low salt concentrations.

Similar to *trans*-cleavage, low salt concentration also promoted the *cis*-cleavage by PlmCas12e (Supplementary Fig. 6a). To investigate the effect of salt concentration on the *cis*-cleavage of Cas12e orthologs, cleavage fractions were measured across a gradient of salt concentrations over 30 min. As salt concentration increased, the *cis*-cleavage efficiency of all Cas12 orthologs exhibited a decreasing trend (Fig. 2e and Supplementary Fig. 6b). Overall, we observed the salt-dependent *cis*- and *trans*-activities for Cas12e nucleases.

Salt concentration affects the PAM recognition and R-loop formation

To further understand the salt-dependent behavior, we explored the target interference process, including dsDNA recognition and invading by Cas12e orthologs under different salt concentrations. Considering the structural diversity of the Cas12e family (Fig. 1b), NTSB-containing PlmCas12e and DpbCas12e and NTSB-lacking VemCas12e and LesCas12e were selected for the analysis. These Cas12e systems exhibit the same *Cas* gene arrangement in the CRISPR locus (Supplementary Fig. 1a) and contain identical sequences at the 3' end of direct repeats, which informed the formation of the conserved triplex region in sgRNA structures (Supplementary Fig. 1b)^{14,26}. These similarities among these four orthologs provide a foundation for conducting comprehensive and parallel analyses.

PAM recognition is a critical step for the target interference and contributes to the initial melting of the dsDNA duplex^{34–37}. Hence, we first evaluated the PAM preferences at different salt concentrations. PlmCas12e showed a more stringent PAM requirement when salt concentrations were increased from 50 mM to 300 mM NaCl (Fig. 2f and Supplementary Fig. 7a). For instance, as the salt concentration increased, PAM position -2 for PlmCas12e exhibited an increased requirement for -2dC. The -2dC-preference was also highlighted using PAM depletion assays conducted in *E. coli*²⁵. Similarly, VemCas12e also showed more strict PAM requirements, especially at -3dT position, as the salt concentration increased (Fig. 2g and Supplementary Fig. 7b). Further cleavage assays revealed that dsDNA containing a preferred PAM exhibited relatively lower cleavage reduction than that with a less-preferred PAM during cleavage by Cas12e proteins as salt concentrations increased (Supplementary Fig. 8a–d). This indicated that the optimal PAM with stringent sequence determined under high salt conditions might secure more robust gene-editing efficacy by Cas editors.

PAM recognition is followed by TS: gRNA heteroduplex (R-loop) propagation. In Cas effectors, R-loop formation is an essential and rate-

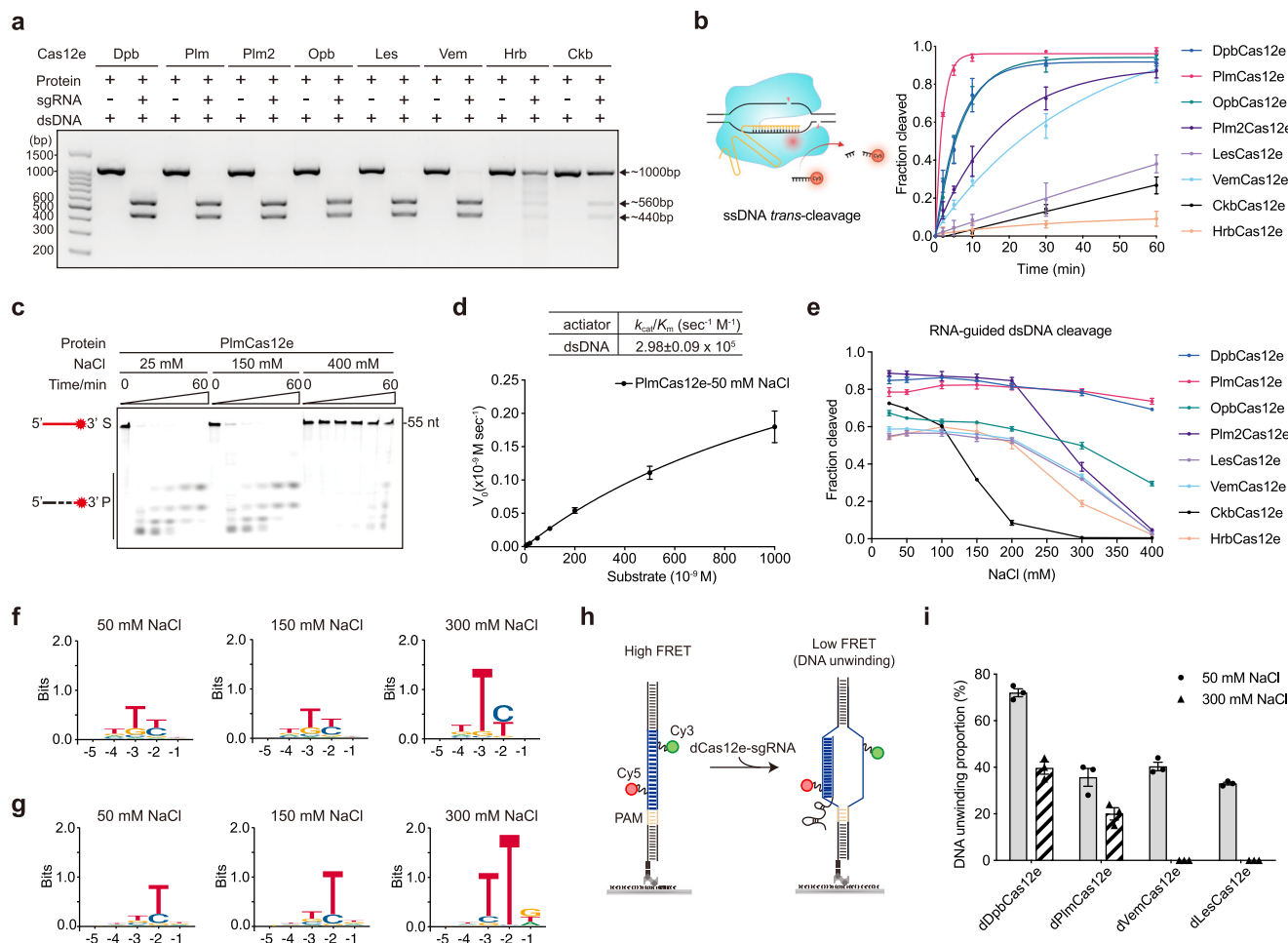


Fig. 2 | High salt concentration reduces the *trans*- and *cis*-cleavage activities of Cas12e orthologs. **a** sgRNA-guided 1 kb-dsDNA cleavage assays for Cas12e orthologs. Similar results were obtained in three independent experiments. **b** Left, schematic of Cas12e (blue)-sgRNA (yellow) RNP complex cleaving target dsDNA (black), resulting in activation of the RuvC nuclease (denoted by a red dot) domain. This schematic is created in BioRender. Liu, J. (2023) BioRender.com/x40b358. Right, comparison of ssDNA *trans*-cleavage kinetics among Cas12e orthologs at a physiological salt concentration of 150 mM NaCl. Data are shown as mean \pm SD from $n = 3$ independent experiments. **c** *trans*-cleavage of 3'-Cy5 labeled 55-nt ssDNA by PlmCas12e under gradients of different salt concentrations. Similar results were obtained in three independent experiments. **d** Michaelis-Menten kinetic studies of

reporter *trans*-cleavage by PlmCas12e under 50 mM NaCl condition. Data are shown as mean \pm SD from $n = 3$ independent experiments. **e** Comparison of dsDNA *cis*-cleavage efficacy among Cas12e orthologs under a gradient of salt concentrations in 0.5 h. Data are shown as mean \pm SD from $n = 3$ independent experiments. Direct lines connecting different points are used to illustrate variations in efficiency. **f–g** PAM screening results of PlmCas12e (**f**) and VemCas12e (**g**) at three salt concentrations. Source data are provided as a Source Data file. **h** Schematic of smFRET assay to quantify dsDNA unwinding with Cy3-labeling on NTS and Cy5-labeling on TS. **i** DNA unwinding proportion for different Cas12e orthologs after a 5-min incubation of DNA substrate and Cas12e-sgRNA complex. Data are shown as mean \pm SEM from $n = 3$ technical replicates.

limiting step for target interference³⁸, we then postulated that the high salt concentration might affect the R-loop formation by Cas12e effectors. To test this hypothesis, we conducted single-molecule fluorescence resonance energy transfer (smFRET) assays and deactivated Cas12e (dCas12e) orthologs were used for capturing the stable R-loop conformation³⁹. In the assays, Cy3 was attached to the non-target strand at the 16th position from PAM, and Cy5 was attached to the target strand at the 7th position from PAM. The labeled dsDNA alone displayed a single high-FRET state as Cy3 and Cy5 were close to each other, representing an unperturbed DNA duplex. The labeled dsDNA invaded by Cas effectors would generate a stable low FRET signal, the proportion of which corresponds to the extent of DNA unwinding and R-loop formation (Fig. 2h)⁴⁰. At a lower salt concentration (50 mM NaCl), all tested dCas12e orthologs could lead to a shift to a stable low FRET, indicating the R-loop formation (Supplementary Fig. 9a–d). In agreement with our hypothesis, as the salt concentration increased to 300 mM, less R-loop populations were detected for all orthologs (Fig. 2i). Notably, the proportion of R-loop is lower than the cleavage

fraction for each Cas12e protein (Fig. 2e, i), and remains largely unchanged when the incubating time was extended from 5 min (used in Fig. 2i) to 30 min (used in Fig. 2e) (Supplementary Fig. 10a–e). We speculate that the discrepancy between results of smFRET assays and cleavage assays is likely attributable to the differing properties that these assays actually measure and present. The smFRET assay captures R-loop formation with deactivated Cas12e under near-equilibrium conditions, where the R-loop proportion remains largely constant. In contrast, the cleavage assay was conducted under non-equilibrium conditions with active Cas12e protein, allowing for the observation of the proportion of cleavage products accumulated over time. Taken together, our results indicate that high salt concentration can hinder the R-loop formation by Cas12e effectors, likely resulting in decreased target interference activity.

Notably, at 300 mM NaCl, dVemCas12e and dLesCas12e rarely generated R-loop, whereas dDpbCas12e and dPlmCas12e were still capable of forming stable R-loops, albeit with a two-fold activity reduction compared to 50 mM NaCl (Fig. 2i). This observation

suggests that VemCas12e and LesCas12e are more sensitive to elevated salt concentrations in R-loop formation compared to PlmCas12e and DpbCas12e.

Structural analysis of VemCas12e and LesCas12e

To gain mechanistic insights into the diverse salt sensitivities of the above-analyzed Cas12e orthologs, we reconstituted the VemCas12e and LesCas12e ternary complexes *in vitro* for structural analysis, aligning with published PlmCas12e and DpbCas12e structures^{14,26}. Considering that the tracrRNA sequence (TracrRNA-T0) used in *E. coli* assays was determined by bioinformatically predicted TSS (Transcription Start Site), we experimentally explored the optimal tracrRNA length for VemCas12e and LesCas12e via screening assay (Supplementary Fig. 11a–f). Among the tracrRNAs of different lengths, TracrRNA-T2 for VemCas12e and TracrRNA-T2 for LesCas12e enabled the highest cleavage activity. These optimal tracrRNAs for both Cas12es were then used for complex reconstitution and related analysis.

Using single-particle cryo-electron microscopy (cryo-EM), we reconstituted and obtained the EM maps of dVemCas12e (D576A/E662A/D837A)-sgRNA-dsDNA ternary complex and dLesCas12e (D578A/E664A/D841A)-sgRNA-dsDNA ternary complex with resolutions of 2.51 Å and 2.92 Å, respectively (Supplementary Fig. 12a–h and 13a–c). Overall, VemCas12e and LesCas12e proteins are highly conserved at the domain level and share a bilobed architecture comprising a REC lobe (composed of Helical-I and Helical-II domain) and a NUC lobe (composed of RuvC, TSL, BH domain, and OBD) (Fig. 3a–d, Supplementary Fig. 13d, left). Structural alignment using Dali server⁴¹ against the Protein Data Bank (PDB) showed that both VemCas12e and LesCas12e share the highest structural similarity with DpbCas12e (PDB: 6NY3) with Z-scores of 21.8 and 20.0, respectively (Supplementary Fig. 13d, middle). Further detailed structural comparison between VemCas12e and DpbCas12e revealed high similarity across Helical-I, Helical-II, OBD, and RuvC domains (Supplementary Fig. 13d, right). However, both VemCas12e and LesCas12e lack the NTSB domain.

Furthermore, conserved PAM recognition details were observed between VemCas12e and LesCas12e proteins. In VemCas12e, Y422 forms two hydrogen bonds with dA(–2), and K479 forms a hydrogen bond with dT(–2*) (Fig. 3e, f). Additionally, K479 also recognizes the O4' and N3 of dG(–1*) via hydrogen bonds, confirming the G-preference at dN(–1*) position in PAM screening results (Supplementary Fig. 2b). The side chains of S99 and Y101 from the Helical-I domain engage in van der Waals interactions with the base of dG(–1*) and the methyl group of dT(–2*), respectively (Supplementary Fig. 13e, left). LesCas12e displays a similar PAM recognition pattern, with Y424, K481, and Y104 exhibiting analogous interactions to the PAM duplex as Y422, K479, and Y101 in VemCas12e (Fig. 3g and Supplementary Fig. 13e, right). However, unlike K479 of VemCas12e, K481 of LesCas12e does not form hydrogen bonds with the base of dG(–1*), aligning with the weaker preference at dN(–1*) position suggested by PAM screening assays (Fig. 1e). The dT(–3*):dA(–3) base pair in the PAM duplex does not form base-specific interaction with VemCas12e or LesCas12e, aligning with the weak cleavage preferences for dT(–3*) suggested by PAM screening assays (Fig. 1e). Of note, NTSB-containing Cas12e indicated by reported structures, like PlmCas12e, showed different PAM recognition details from VemCas12e and LesCas12e²⁶. In PlmCas12e, the Q512 and Y513 from the OBD form hydrogen bonds with dA(–3) and dG(–2), respectively (Supplementary Fig. 13f). Meanwhile, K227 from the Helical-I domain recognizes the bases of both dT(–3*):dA(–3) pair and dG(–2) through hydrogen bonds. These interactions collectively contribute to the 5'-TCN PAM preference by PlmCas12e.

The sgRNAs of VemCas12e, LesCas12e, and PlmCas12e shared the similar overall architecture, including a triplex region composed of triple base pairs, a scaffold stem, and an extended stem formed by pairing of repeat and anti-repeat (Fig. 4a, b). Furthermore, the

interactions between the REC lobe and scaffold stem and between the NUC lobe and triplex region are well-observed across three complexes. Beyond, these sgRNAs have many varied details. The 5' end of sgRNA base pairs with the scaffold stem in PlmCas12e complex but not in VemCas12e and LesCas12e complexes. Additionally, the sgRNA of VemCas12e but not PlmCas12e or LesCas12e features an extra stem (termed outward stem) free from the VemCas12e protein. Truncation of this extra stem showed minimal effect on the dsDNA cleavage by VemCas12e (Supplementary Fig. 14a, b). These observations highlight the conservation and diversity of sgRNAs in different Cas12e systems.

Our structural analysis has shown that, with the exception of the NTSB regions, Cas12e orthologs exhibit a high level of conservation at the domain level and utilize sgRNAs with similar architecture, indicating that NTSB is the primary factor in differentiating their DNA interference behaviors.

NTSB-DNA interactions facilitate R-loop formation under high salt concentration

Unlike NTSB-lacking Cas12e, NTSB-containing Cas12e maintained considerable R-loop populations at high salt concentrations (Fig. 2i). To validate the beneficial role of NTSB, we introduced alanine mutations (K106A/K148A/K150A) at some of the residues responsible for DNA R-loop interaction and stabilization in PlmCas12e NTSB domain (Fig. 5a, b), and measured the dsDNA cleavage activities across varying salt concentrations (Fig. 5c and Supplementary Fig. 15a, b)³⁹. Compared with the wild-type (WT) PlmCas12e, the NTSB-mutant showed similar cleavage activities at lower salt concentrations (50, 100, and 150 mM NaCl) but substantially decreased activities at high salt concentrations (200 and 300 mM NaCl). Notably, Plm2Cas12e naturally lacks positively charged residues at analogous positions to K106 and R192 in PlmCas12e (Fig. 5d and Supplementary Fig. 15c), which are involved in NTS and TS interaction, respectively³⁹. Indeed, Plm2Cas12e behaved similarly to PlmCas12e NTSB-mutant across the salt concentrations (Fig. 5e and Supplementary Fig. 15d). Reintroducing a single arginine residue at Q217 (corresponding to R192 in PlmCas12e) in Plm2Cas12e enhanced the cleavage activity at a high salt concentration (300 mM NaCl) but showed minimal effect to the DNA cleavage at low salt concentrations (25–200 mM) (Fig. 5e), highlighting the role of the arginine for the dsDNA cleavage by Plm2Cas12e at high salt concentration.

Our findings collectively indicate that the stable interactions between the DNA target and the NTSB domain, primarily facilitated by positively charged residues, play a crucial role in promoting R-loop formation and dsDNA cleavage under high salt concentrations.

VemCas12e and LesCas12e adopt positively charged loops for DNA duplex melting

Though much less effective compared to NTSB-containing Cas12e, NTSB-lacking orthologs, like VemCas12e and LesCas12e, were still able to invade dsDNA and display detectable dsDNA cleavage signal under high salt concentrations (Fig. 2e). This suggests that NTSB-lacking orthologs might contain some less-dominant elements for dsDNA invading. Indeed, in the analogous position to the NTSB domain, both VemCas12e and LesCas12e possess a positively charged loop (termed Loop 1) composed of about 12 residues (Fig. 5f and Supplementary Fig. 16). The alanine substitutions at the positively charged residues (K91A/K92A/K96A/K97A) within Loop 1 of VemCas12e nearly abolished the cleavage of dsDNA but remained effective for bubbled dsDNA cleavage (Fig. 5g and Supplementary Fig. 17a). This indicates that Loop 1 plays critical roles in dsDNA unwinding prior to complete R-loop formation. The density of NTS that may interact with Loop1 cannot be observed in the EM map, likely due to structural flexibility in this local region. This suggests a dynamic or transient interaction between Loop 1 and NTS within the R-loop window, unlike the strong and stable NTSB-NTS interaction observed in NTSB-containing Cas12e^{14,26}.

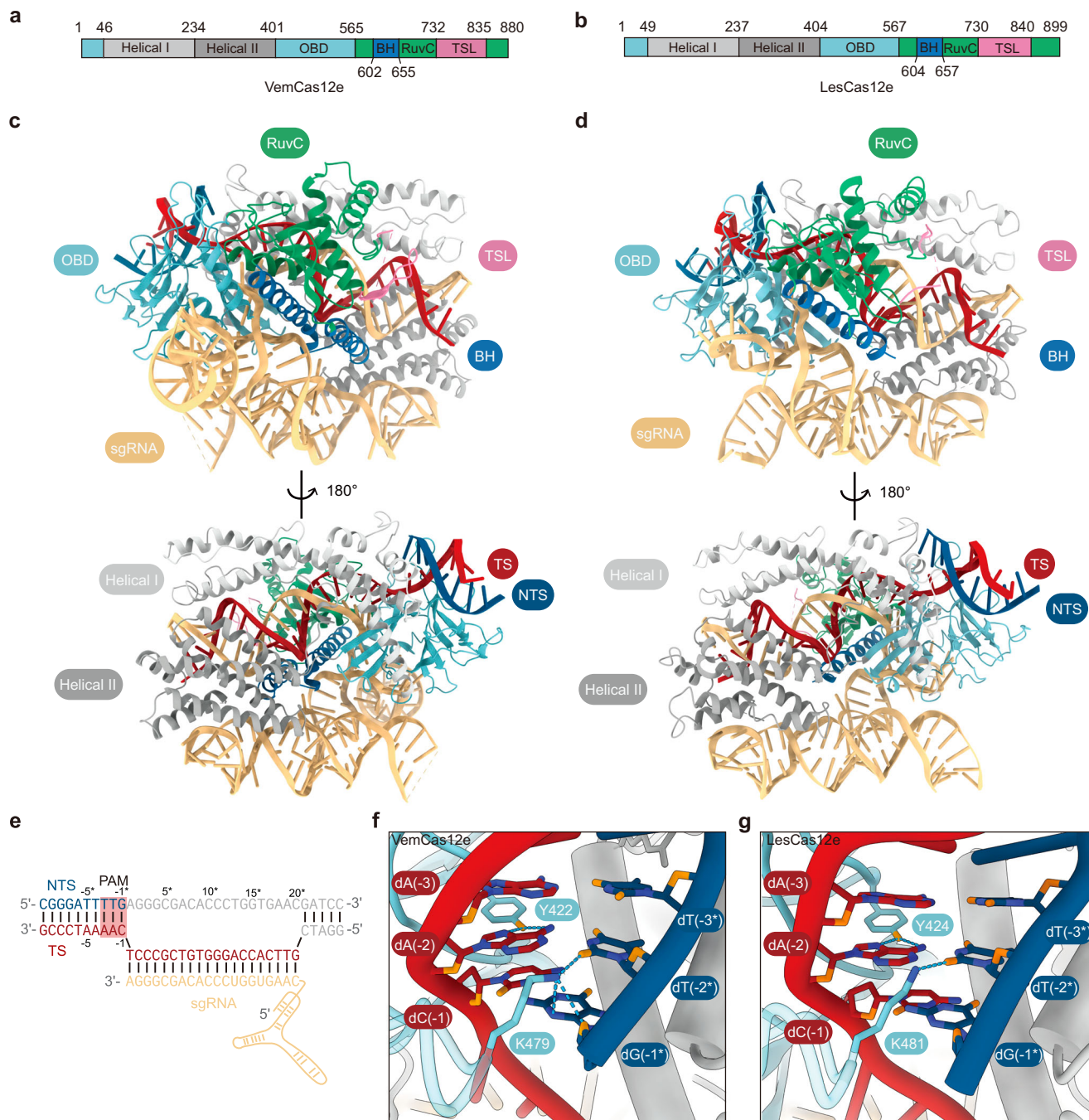


Fig. 3 | Cryo-EM structures of VemCas12e and LesCas12e ternary complexes. a, b Domain architectures of VemCas12e (a) and LesCas12e (b). **c, d** Overall structures of dVemCas12e (c) and dLesCas12e (d) in complex with sgRNA and target dsDNA. Domains are colored identical to those in a and b. **e** Partially paired dsDNA target

used in both VemCas12e (c) and LesCas12e (d) complexes. Structurally disordered nucleotides are colored gray. **f, g** PAM recognition patterns of VemCas12e (f) and LesCas12e (g). Base-specific hydrogen-bonding interactions are indicated with blue dashed lines.

Notably, within the OBD of both VemCas12e (475–486 aa) and LesCas12e (477–488 aa), another positively charged loop, denoted Loop 2, was observed to insert into the minor groove of the PAM duplex (Fig. 5f). Although PlmCas12e and DpbCas12e also contain a Loop 2, their Loop 2s are six to nine residues shorter and do not insert into the DNA duplex. Additionally, Loop 2 in VemCas12e and LesCas12e is engaged in the recognition of dT(−2*) in the PAM duplex via a lysine residue (K479 in VemCas12e and K481 in LesCas12e) (Fig. 3f, g). Additionally, the side chain of an arginine residue (R478 in VemCas12e, R480 in LesCas12e) within Loop 2 is inserted directly downstream of

the PAM, likely to act as a physical barrier to preserve DNA strand separation (Fig. 5h, i). Both VemCas12e-K479A and -R478A mutants showed significantly reduced DNA cleavage activity (Supplementary Fig. 17b, c). The activity was restored by introducing one or two mismatched base pairs proximal to the PAM (Supplementary Fig. 17b, c), suggesting that these residues contribute to R-loop formation by facilitating the melting of dsDNA proximal to the PAM. Together, these observations indicate that some short structural elements, such as Loop 1 and Loop 2, in NTSB-lacking Cas12e may play a similar but weaker role as NTSB in unwinding dsDNA (Supplementary Fig. 17d).

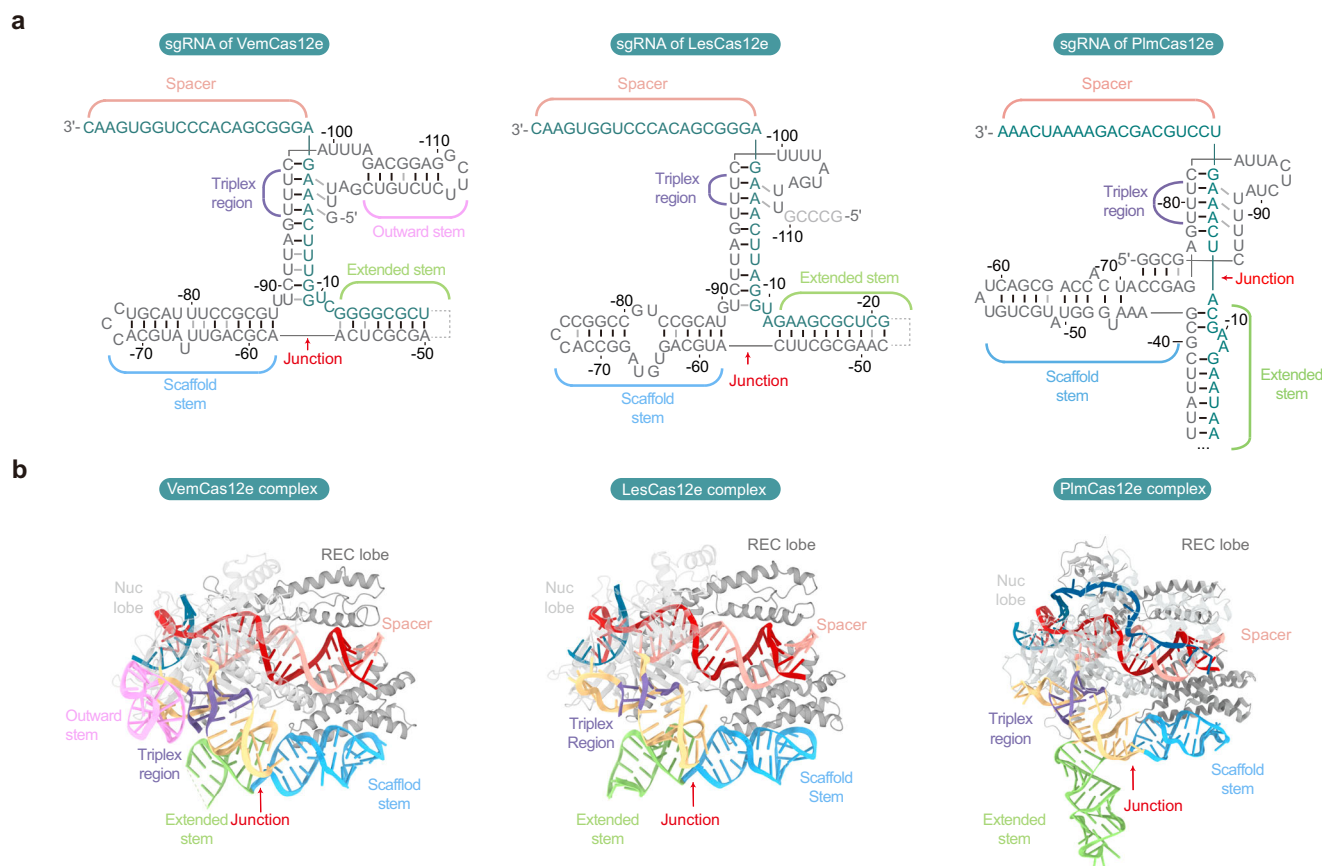


Fig. 4 | sgRNA architectures of VemCas12e and LesCas12e. a Schematic of guide RNA structures of VemCas12e (left), LesCas12e (middle), and PlmCas12e (right, PDB: 7WB1). The junction connecting the triplex region and two stems is highlighted with a red arrow. The black line indicates the Watson-Crick base pair. The grey line

indicates a Non-Watson-Crick base pair. The grey dashed lines indicate disordered regions. **b** Guide RNA structures of VemCas12e (left), LesCas12e (middle), and PlmCas12e (right, PDB: 7WB1) in ternary complexes. The structure of sgRNA is colored similarly as in **a**. NTS is colored in blue and TS is colored in red.

NTSB-lacking OpbCas12e adopts a unique helix-loop structure for robust R-loop formation and target cleavage

Although OpbCas12e, like other NTSB-lacking members, contains Loop 1 and Loop 2 (Supplementary Fig. 18a, b), it exhibits unique biochemical behavior compared to VemCas12e and LesCas12e. OpbCas12e showed higher activity than other NTSB-lacking members across all tested salt concentrations (Fig. 2e) and significantly maintained comparable R-loop signals in both low and high salt concentrations (50 mM and 300 mM NaCl) (Supplementary Fig. 18c).

To reveal the molecular details for OpbCas12e's robust dsDNA unwinding ability without NTSB, we carefully compared the structures of OpbCas12e (predicted) and VemCas12e and identified a unique helix-loop structure in OpbCas12e RuvC domain (Supplementary Fig. 18d). This helix-loop shows minimal sequence similarity to known proteins. FRET assays showed that helix-loop-deleted dOpbCas12e mutant could still generate a considerable R-loop signal at low salt concentration (50 mM NaCl) but not at high salt concentration (300 mM NaCl) (Supplementary Fig. 18e). Meanwhile, the dsDNA cleavage activity of OpbCas12e was abrogated after deletion of the helix-loop regardless of the salt concentrations (Supplementary Fig. 18f). These results suggested that the unique helix-loop in OpbCas12e is not only beneficial for dsDNA unwinding but also essential for subsequent DNA cleavage.

Based on our results, we would like to propose a model to illustrate the impact of salt concentration on R-loop formation. In the reaction system with low salt concentrations (Fig. 6a), the free DNA target was inadequately decorated by a small number of cations, like Na^+ , and displayed a vulnerable conformation for Cas12e invading.

Likely, the overall electrostatic interaction between Cas12e lobes and dsDNA was efficient for allowing the R-loop formation. Under such conditions, the extra DNA-unwinding capability provided by NTSB and the loops was likely beneficial but not essential. While under the high salt concentrations (Fig. 6b), the overall electrostatic interaction between Cas12e lobes and dsDNA was largely dampened by the competition of concentrated cations. Therefore, NTSB and the loop elements turned essential for DNA-unwinding and R-loop formation.

Discussion

Divergent evolution of CRISPR-Cas12e subtype with unique traits for adaptations

Our structural comparisons revealed the NTSB domain as an enhancer for R-loop formation and cleavage efficiency. Based on the bilobed configuration, Cas12e proteins employ various modules, such as the NTSB domain and certain positively charged loops close to the PAM duplex, and the helix-loop within the RuvC domain, to facilitate dsDNA unwinding and cleavage (Fig. 6c). Other Cas12 systems adopt similar strategies with different structural components. For instance, Cas12a⁴² and Cas12b⁴³ evolved a "phosphate lock" interaction model to facilitate the unwinding of PAM-adjacent base pairs. Some Cas12 proteins insert small modules between the two strands of target DNA to unzip PAM-proximal DNA or maintain the unwound conformation. For example, Fncas12a⁴⁴ contains a loop-lysine helix-loop (LKL) within the PAM interacting domain, Cas12f⁴⁵ features a helix within REC1 domain, and Cas12i⁴⁶ comprises a zipper motif within the Helical-I domain. Overall, for robust dsDNA unwinding, the Cas12 proteins took various

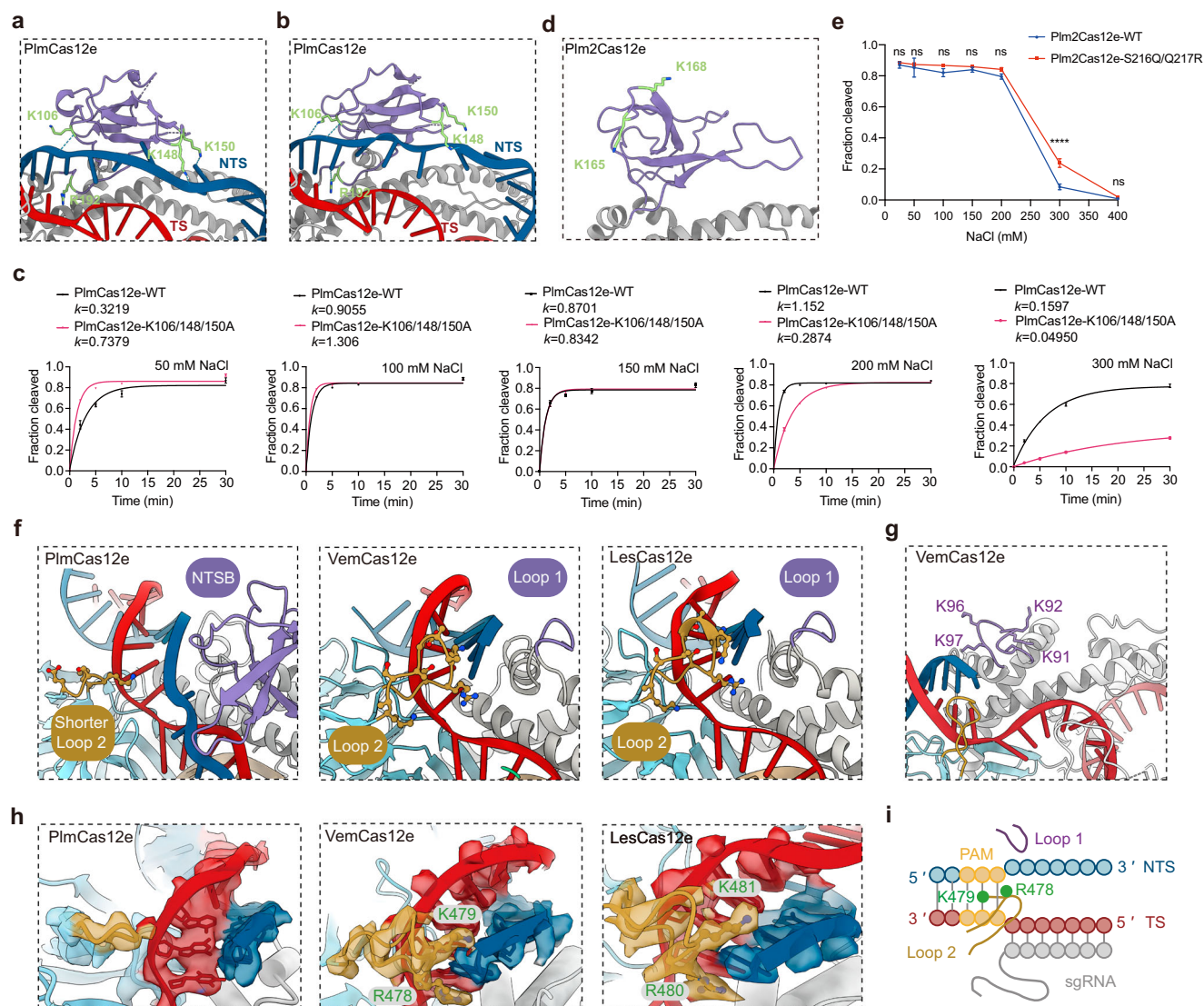


Fig. 5 | dsDNA melting mechanisms by NTSB-containing PlmCas12e and NTSB-lacking VemCas12e and LesCas12e. **a, b** Positively charged residues involved in DNA duplex interactions in the PlmCas12e NTSB (**a** PDB: 7WB1) and PlmCas12e NTSB (**b**, PDB: 7WAZ). The blue dashed lines indicate hydrogen-bonding interactions. **c** Comparisons of *cis*-cleavage on dsDNA by wild-type (WT) PlmCas12e and lysine residues-mutated PlmCas12e at varying salt concentrations. The cleaved fraction for each reaction was fit into the One Phase Decay model. k value is the rate constant. Data are shown as mean \pm SD from $n = 3$ independent experiments. **d** Positively charged residues at corresponding positions to K148 and K150 within the Plm2Cas12e NTSB (predicted). **e** dsDNA cleavage efficiencies by wild-type Plm2Cas12e and arginine-restored Plm2Cas12e mutant at different salt concentrations. Data are shown as mean \pm SD from $n = 3$ independent experiments. Statistical

analysis was performed using a two-way ANOVA test with Sidak's multiple comparison test. (**** p value < 0.0001 ; "ns" means not significant with the p value > 0.05). The exact p values for the comparison at 25, 50, 100, 150, 200, 300, and 400 mM NaCl are 0.9704, 0.8939, 0.1051, 0.9125, 0.1220, < 0.0001 , and 0.9953, respectively. Source data are provided as a Source Data file. **f** Comparison of structural elements of NTSB-containing PlmCas12e (PDB: 7WB1) and NTSB-lacking VemCas12e and LesCas12e. NTSB and Loop 1 are colored in purple, and Loop 2 is colored in brown. **g** Positively charged Loop 1 of VemCas12e. Four lysine residues are denoted. **h** The electron density map of Loop 2 among three Cas12e orthologs. Residues interacting with dsDNA duplex are denoted. **i** Schematic diagram showing different structural elements responsible for dsDNA invading in VemCas12e.

evolutionary directions, including domain expansion, acquisition of new structural elements, and residue variability, leading to significant diversity in both size and nuclease activity.

This study reveals that the different DNA unwinding capacities are responsible for the varying salt sensitivities among Cas12e systems (Fig. 2i). Indeed, these Cas12e systems are identified from diverse environments with various salt concentrations. Furthermore, bacterial cells could adjust the intracellular osmotic pressure to provide the turgor pressure against the living environments^{47,48}. Therefore, we conceive that the high salt tolerance of some Cas12e systems is crucial for host bacteria to cope with fluctuations in the osmolarity of the external environment.

Less effective binding of TS by VemCas12e and LesCas12e

In the VemCas12e and LesCas12e ternary complexes, we did not observe the density of the NTS downstream of the PAM duplex, whereas PlmCas12e and DpbCas12e complexes formed a complete R-Loop with displaced NTS bound to the surface of the NTSB, RuvC, and TSL domains^{14,26}. Additionally, the TSL domains in VemCas12e and LesCas12e are disordered, likely resulting in the missing of state II, at which the RNA-DNA heteroduplex is bent back to access the RuvC catalytic pocket for the cleavage of the target strand in PlmCas12e and DpbCas12e. These structural differences indicate that VemCas12e and LesCas12e may bind the TS less effectively than PlmCas12e and DpbCas12e.

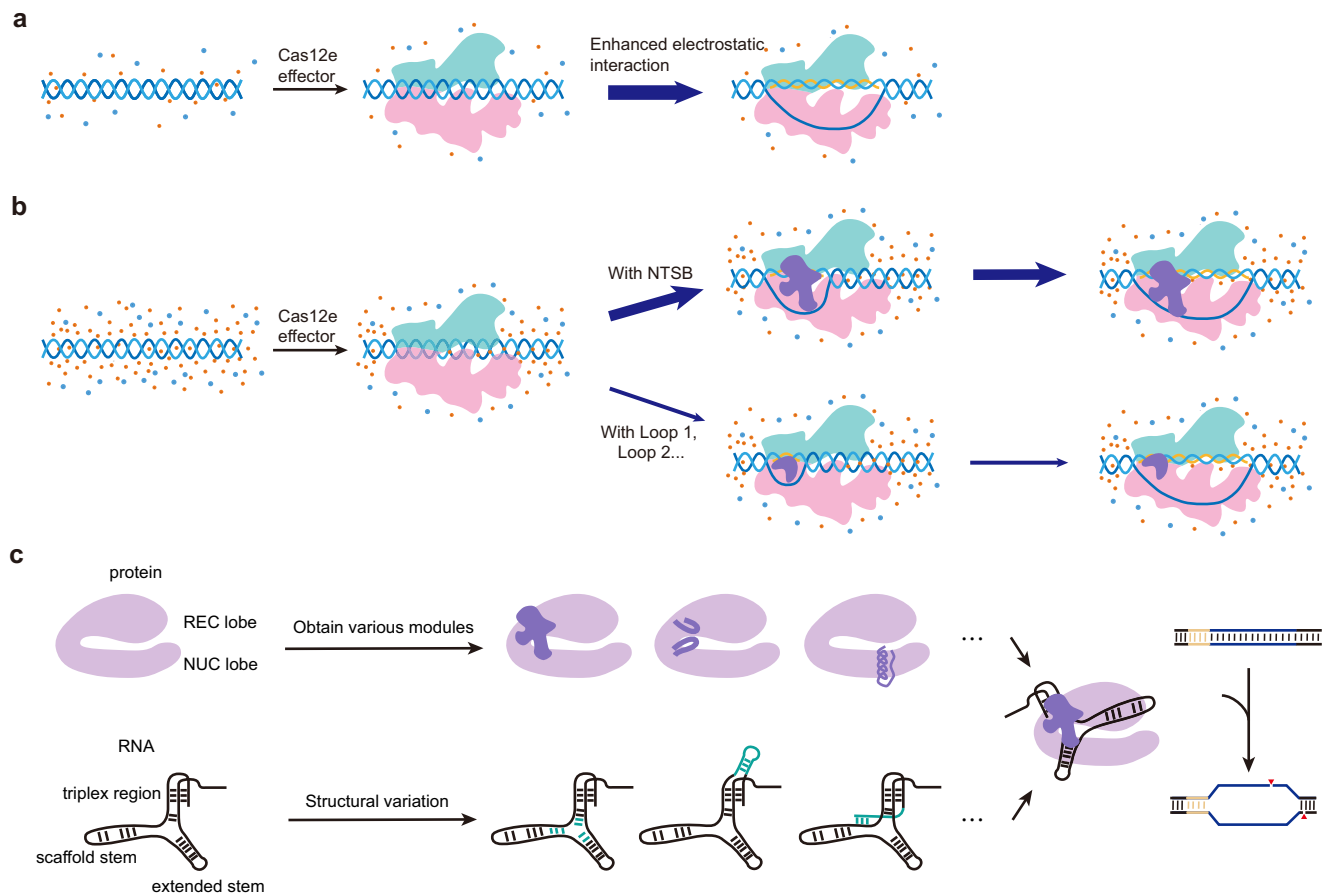


Fig. 6 | Modulation of the dsDNA unwinding by Cas12e orthologs at low and high NaCl concentrations. **a** Cas12e effector mediated R-loop formation of the dsDNA duplex at low salt concentration. Orange dot, positively charged ion; blue dot, negatively charged ion. Green module, REC lobe; pink module, NUC lobe. **b** dsDNA unwinding and complete R-loop formation performed by NTSB-containing Cas12e and NTSB-lacking Cas12e at high salt concentration. Purple

module, NTSB domain or loops. The thickness of the blue arrowed lines represents the likelihood of reaction occurrence. **c** The diversities of Cas12e and its cognate sgRNA. Different elements colored dark purple were attained based on the bilobed configuration. The sgRNA scaffolds (in black) with varying details in base pairing (colored green) indicate various sgRNAs. The PAM region on the dsDNA target is colored in yellow. Red triangles indicate the cleavage.

Mechanisms of salt sensitivities in dsDNA *cis*-cleavage and ssDNA *trans*-cleavage

Our study observed a substantial increase in the dsDNA cleavage activity of Cas12e proteins at lower salt concentrations, which is inherently comprehensible. Cations (e.g., Na⁺) in the salt solution may neutralize the negatively charged DNA strands, reducing the repulsion between the two strands and, therefore, stabilizing the double-stranded structure⁴⁹. Consequently, reduced cations facilitate the interaction between Cas12e proteins and DNA backbone, thereby promoting the R-loop formation. This aligns with previous studies showing mutations that increase the interaction between the protein and the DNA backbone may improve genomic editing capacities^{50–53}. On the other hand, in vitro systems can exploit salt concentration as a selective pressure to identify more efficient DNA manipulation tools. Furthermore, under the background of increasing salinization of cultivated land around the world, this study has significant application potential for the genome editing in plants that thrive in saline-alkali land.

We also observed noticeable reduced *trans*-cleavage activity as the salt concentration increased (Fig. 2c). The *trans*-cleavage of ssDNA by the RuvC domain is primarily driven by non-specific digestion²¹, where ssDNA recognition depends on the electrostatic interaction between the positively charged pocket of the RuvC domain and the negatively charged phosphate backbone of the ssDNA. The increase of the positively charged ions likely screens the negative charges on the

ssDNA, thereby reducing the binding between the ssDNA and the RuvC domain.

In summary, our work elucidates the structural mechanisms underlying the R-loop formation across Cas12e orthologs. Additionally, we emphasize the crucial role of the NTSB domain for salt-tolerant R-loop formation and target cleavage. These findings not only expand the toolkit for in vitro implications but also pave the way for the rational design of Cas12e variants with enhanced activity.

Methods

Statistics & reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Cas12e candidates mining

Previous reported Cas12e nucleases were aligned using ClustalO⁵⁴ with parameter “iterations” changed to 10. Using HMMER, alignment was subsequently used to build the Hidden Markov Model (HMM), which was further searched against the NCBI database for homologous proteins⁵⁵. Cas protein and CRISPR array were predicted from candidates’ up/downstream sequences using local version of CRISPRCasFinder⁵⁶. Functions of these proteins were further

analyzed with eggNOG mapper (version 2.1.9, based on eggNOG Database version 5.0.2)⁵⁷.

Phylogenetic analysis

The representatives of Cas12 family members, PlmCas12e and DpbCas12e, were collected. The multiple sequence alignment of all Cas12 effectors was generated using ClustalO with parameter “iterations” changed to 10. The phylogenetic tree was further constructed using IQ-TREE2 (version 2.0.7)⁵⁸ with the following parameters: -alrt 1500 -bb 1500 nm 1500 m VT + F + R6. Tree was visualized in iTOL (version 6.8)⁵⁹. The phylogenetic tree of Cas12e family members was built similarly.

Next-generation sequencing data processing

All sequencing data was processed by Trim Galore (version 0.6.5) to remove adapters and low-quality reads. Q20, Q30, and GC content of the clean data were calculated by FasQC (version 0.12.0). Next, paired reads were merged by FLASH (version 2.2.00)⁶⁰ for subsequent analysis. For the PAM screening assay, the N5 region was extracted between known up/downstream sequences using fasta-grep and visualized using WebLogo (version 2.8.2a)⁶¹. For cleavage site determination, cleavage sites were exposed after adapter removal, and reads representing different cleavage sites were merged and calculated for cleavage ratio.

TracrRNA and crRNA determination

For CRISPR-Cas12e systems, the 3' region of the tracrRNA was determined based on the complementarity of repeat and anti-repeat. Specifically, repeat sequence was used to search for anti-repeat region against the window spanning from 1 kb upstream of the Cas proteins to the end of the CRISPR arrays in the genomes using Pairwise Sequence Alignment in EMBOSS Needle with default parameters. The 5'-boundary of tracrRNA was determined by promoter prediction using BPROM⁶². All tracrRNAs were determined in the same way as shown in Fig. 1c and Supplementary Fig. 1a, and the sequences were shown in Supplementary Data 1. The alignment of all the repeats in Supplementary Fig. 1b was performed using Multiple Sequence Alignment in Clustal Omega (1.2.4) with default parameters⁵⁴.

Plasmids construction

Bacteria- and human-codon optimized *Cas12e* genes were cloned to the protein expression pET28b-based vector with an N-terminal hexahistidine tag and a SUMO tag by homologous recombination (One Step Seamless Cloning Mix, CWBIO). For the *Cas12e* mutant, mutations were introduced via PCR using mutagenetic primers. The purification of DpbCas12e and PlmCas12e proteins utilized the 10xHis-MBP-tagged plasmid reported in previous studies^{14,26}. For the purification of gRNA in vitro, a T7 promoter and DNA template were inserted into the pUC19. Then, a 0.7 kb-long dsDNA template was amplified from that plasmid, serving as a transcription template in the system. For PAM screen assays, the pUC19-derived plasmid library containing five randomized nucleotides upstream of the protospacer was constructed as previously described²⁸. Briefly, a single-stranded oligonucleotide containing five randomized nucleotides (N5-oligo) was synthesized. Then, the N5-oligo was annealed to and amplified with another oligonucleotide containing a complementary region to the 3' end of the N5-oligo. The PCR products were purified using DNA Clean & Concentrator-25 (Zymo Research) and cloned into the linearized pUC19 plasmid by homologous recombination (One Step Seamless Cloning Mix, CWBIO). Using a similar method, the plasmid with each single PAM of Cas12e orthologs was obtained for cleavage site determination. To express Cas12e proteins and gRNAs for plasmid interference assays in bacteria, a pTcr-promoted Cas12e gene, and J23119-promoted sgRNA were inserted into the pCDFDuet vector to obtain pCDFDuet-Cas12e-gRNA. Two BbsI- or SapI- recognizing sites reside at

the 3' end of the gRNA, enabling the convenient spacer insertion through the Golden Gate method. The plasmids used in this study are listed in Supplementary Data 2.

Protein expression and purification

The plasmid pET28b-10xHis-SUMO-Cas12e was transformed into *E. coli* Rosetta (DE3) competent cells by heating at 42 °C for 1 min before the 5-min resting on ice. Then, a 1 mL Luria broth (LB) was added to the mixture and incubated at 37 °C for 1 h to recover. The cells were transferred to a solid LB dish containing 50 µg/mL kanamycin and incubated at 37 °C overnight. The next day, around ten clones were picked and transferred into 100 mL LB medium for further cultivation. After six hours, 20 mL of the culture was added to 500 mL terrific broth (TB) medium supplemented with 100 µg/mL ampicillin and 4 mL 50% glycerol. Once the optical density (OD₆₀₀) reached 1.0–1.2, expression was induced with 0.4 mM IPTG before the cells were grown overnight at 16 °C. The cells were gathered by centrifugation and resuspended in lysis buffer [20 mM HEPES pH 7.5 at 25 °C, 600 mM NaCl, 10% (v/v) glycerol, 50 mM imidazole, 1 mM TCEP] supplemented with 4 mM β-mercaptoethanol and 1 mM PMSF. After the sonication, lysed cells were centrifuged at 24,100 g for 60 min at 4 °C. The supernatant was loaded onto the pre-equilibrated Ni-NTA (QIAGEN) agarose beads three times. Then the resin was washed with a tenfold volume of lysis buffer. The His-tagged protein was eluted by lysis buffer supplemented with 300 mM imidazole. To remove the His-SUMO tag, the elute was incubated with Ubl-specific protease 1 (Ulp1) at 4 °C for 30 min. Next, the mixture was loaded into a 5 mL HiTrap Heparin HP column (GE Healthcare) and eluted with a linear gradient of heparin elution buffers [buffer A (20 mM HEPES pH 7.5, 200 mM NaCl, 10% Glycerol, 1 mM TCEP), and buffer B (20 mM HEPES pH 7.5, 1 M NaCl, 10% Glycerol, 1 mM TCEP)]. Peak fractions containing target proteins were pooled and further purified by size-exclusion chromatography (Superdex increase 200 10/300 GL; GE Healthcare) gel filtration column equilibrated with SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM TCEP). The protein fractions validated by polyacrylamide gel electrophoresis (PAGE) were aliquoted, flash-frozen in liquid nitrogen, and stored at –80 °C. DpbCas12e and PlmCas12e were purified using a similar protocol with some modifications. Both proteins were tagged with maltose binding protein (MBP) in 2CT-10 expression vector. Following incubation with Ni-NTA resin, the enriched Cas12e proteins were incubated with tobacco etch virus (TEV) protease overnight in a buffer containing 20 mM HEPES (pH 7.5 at 25 °C), 600 mM NaCl, 10% (v/v) glycerol, and 1 mM TCEP. The flow-through was then collected and applied to a Heparin column for further purification. For gel filtration, DpbCas12e and PlmCas12e proteins were eluted using a buffer containing 20 mM HEPES (pH 7.5 at 25 °C), 400 mM NaCl, 10% glycerol, and 1 mM TCEP. Protein sequences are listed in Supplementary Data 1.

RNA preparation

All the gRNAs were produced using in vitro transcription (IVT). First, DNA sequences containing T7 RNA polymerase promoter upstream of the gRNA were assembled by overlap PCR before it was inserted into a pUC19 empty vector. The sequence was further validated by Sanger sequencing. Then, a 0.7 kb linearized DNA templates were acquired by PCR using two primers, one of which could anneal to the 3' end of spacer region, enabling the precise termination. The remaining steps were conducted as described previously¹⁷. Briefly, DNA template was incubated with 0.5 mg/mL T7 polymerase and 4 mM NTP mix in IVT buffer (30 mM Tris pH 8.1, 25 mM MgCl₂, 0.01% Triton, 2 mM spermidine) for 3 hrs at 37 °C. The mixture was then loaded into 10% Urea-PAGE with 2x formamide loading (95% formamide, 0.02% SDS, 0.02% BPB, 0.01% xylene cyanole FF, 1 mM EDTA) for electrophoresis. Correct band was extracted, smashed, and soaked in soaking buffer (0.38 M NaAc pH 5.2, 0.8 mM EDTA, 0.8% SDS) overnight at 4 °C. The gRNA was

further washed twice using nuclease-free water in a 3 kDa MWCO concentrator (Merck Millipore), and finally concentrated into 400 μ L. The RNA sample was then transferred to a 1.5 mL tube. To precipitate the RNA, 0.3 volumes of 3 M NaOAc and 3 volumes of chilled 100% EtOH were added to the sample, followed by incubation at -80°C for 2 h. Next, the sample was centrifuged at 12000 g for 30 min. The resulting pellet underwent two additional washes with 70% EtOH before being dissolved in 200 μ L nuclease-free water. To achieve proper folding, the concentrated gRNA was transferred into refolding buffer (50 mM KCl, 5 mM MgCl_2) and incubated at 72°C for 5 min, then slowly cooled down to room temperature (RT) before being frozen and stored at -80°C . The gRNA sequences are listed in Supplementary Data 1.

PAM screening and cleavage site determination

The PAM screening using *in vitro* systems is adapted from a previous study with modifications²⁸. The N5 plasmid library was incubated with Cas12e protein and sgRNA at a ratio of 10:15:1 (Cas12e:sgRNA:plasmid) at 37°C for 30 min in a cleavage buffer (20 mM Tris-HCl pH 7.5 at 37°C , 50/150/300 mM NaCl, 10 mM MgCl_2 , 1 mM DTT). The reaction was stopped by adding 15 mM EDTA with 1 mg/mL proteinase K and was incubated at 50°C for 30 min. Linearized plasmid was extracted and end-repaired by T4 DNA polymerase (Thermo Fisher Scientific) with 1 mM dNTP (Sangon Biotech). A dA was added to the 3' end of the products by DreamTaq polymerase (Thermo Fisher Scientific) with 1 mM dATP (Sangon Biotech). Next, the adapters with a 3' dT overhang were ligated to the purified products with the same mass concentration by fast T4 DNA ligase (Beyotime). A fragment was amplified from the region containing the N5 PAMs using two primers, one pairing to the upstream of the PAM, and the other pairing to the adapter. Next, the purified PCR products were applied to the TIANSeq Fast DNA Library Prep Kit (TIANGEN) for Illumina Novaseq PE150 next-generation sequencing (NGS). As a control, N5 plasmid library was incubated in the blank cleavage buffer at 37°C for 30 min. Then a fragment was PCR amplified by two primers paired to the upstream and downstream of the PAM before being applied to NGS library preparation using kits. The depletion fold-change for each PAM was analyzed using the number of matched reads in Cas12e and control groups normalized with total reads.

To determine the cleavage sites of Cas12e on plasmids, a plasmid with a single PAM was applied to the cleavage system described above. The linearized products were treated using the same methods as those in PAM screening assays, with the exception that the fragment covering the PAM-distant end after the cleavage was also enriched by two primers paired to the adapter and downstream of the PAM, enabling the cleavage sites determination on the non-target strand. The primers used in PAM screening assays are listed in Supplementary Data 3.

Plasmid interference assay in bacteria

E. coli BW25141 competent cells carrying the ccdB toxin plasmid (p11-LacY-wtx1) were prepared following the protocol previously described⁶³. Briefly, the p11-LacY-wtx1 plasmid was transformed into chemically competent *E. coli* BW25141. The transformed cells were cultured at 37°C with vigorous shaking in 1 L LB medium until the OD₆₀₀ reached 0.4–0.5. The cells were then chilled and harvested by centrifugation. After harvesting, the cells were washed sequentially with 800 mL water followed by 400 mL 10% glycerol. Finally, the cells were resuspended in 2.5 mL 10% glycerol, frozen in aliquots, and stored at -80°C . 50 μ L competent cells were electroporated with 200 ng plasmid (pCDFDuet-Cas12e-sgRNA) with 0.2 cm cuvette (BIO-RAD) under 1.8 kV using Eppendorf eporator (Eppendorf). After 1.5-hr of recovering in 5 mL SOC medium at 37°C , the bacterial cells were enriched by centrifugation and resuspended with 5 mL LB containing 50 $\mu\text{g}/\text{mL}$ streptomycin, and cultured for an extra 8–12 h. To induce the expression of the ccdB toxin, 5 μ L of culture with gradient dilutions

from 10^0 to 10^{-7} was spotted onto the LB-agar plate containing 50 $\mu\text{g}/\text{mL}$ streptomycin and 10 mM arabinose, incubating overnight at 37°C .

PAM depletion assay

PAM depletion assays were conducted as previously described²⁵ with modification. Specifically, ampicillin-resistant plasmid libraries containing randomized N5-PAM sequences were transformed into *E. coli* DH5 α , which had been pre-transformed with a streptomycin-resistant plasmid encoding Cas12e proteins and the cognate sgRNA targeting the N5 PAM plasmid. In the control group, the pre-transformed plasmids encoded the Cas12e protein and a non-targeting cognate sgRNA. After transformation, 1 mL LB medium was added into the cells for 1-hour culturing at 37°C . Then, the cells were plated on selective medium containing streptomycin (50 mg/mL) and Ampicillin (100 mg/mL) for 12 h at 37°C . Plasmid DNA was extracted, and the PAM-including region was amplified with adapters for Illumina sequencing. The 5-nucleotide PAM sequences were extracted, and their frequencies were calculated. PAM sequences depleted by more than 5-fold were identified and used to generate sequence logos via WebLogo3⁶¹.

In vitro cleavage assays

The 65 bp Cy5-labeled dsDNA substrates were obtained using a PCR method described in a previous study⁶⁴. Briefly, a 5'-Cy5 labeled 16 nt primer was mixed with a 65-nt ssDNA template with certain PAMs in PCR system (Q5 polymerase, NEB). The PCR products after 35-cycle amplification were purified using DNA Clean&Concentrator-25 (Zymo Research). The concentration was further determined by comparing the strength of the band to the standard Cy5-labeled ssDNA sample using urea-PAGE gel. Generally, 500 nM Cas protein was incubated with 750 nM gRNA at 25°C to form the ribonucleoprotein (RNP) complex in a cleavage buffer (20 mM Tris-HCl pH 7.5 at 37°C , 150 mM NaCl, 10 mM MgCl_2 , 1 mM DTT). The salt concentration for the *trans*-cleavage was adjusted as required. Then, 10 nM dsDNA substrates were added and mixed with RNP on ice. The assays were performed at 37°C , and aliquots were collected at the indicated time points described in figure legends. The aliquots were mixed with 2x urea loading buffer (8 M urea, 25 mM EDTA, 100 $\mu\text{g}/\text{mL}$ Heparin) and were incubated at 95°C for 3 min before being loaded into 15% Urea-PAGE. The cleavage signals were visualized using a Gel scanner (Amersham Typhoon 5, GE Healthcare). For the cleavage of 1-kb dsDNA, 30 nM substrates were added into the cleavage system for better imaging using nucleic acid stain. The reaction was stopped by adding 25 mM EDTA and 1 mg/mL proteinase K (Solarbio). After a 1 h incubation at 50°C , the samples were run on 2% agarose gel.

For the cleavage of bubbled dsDNA by VemCas12e and its mutant, a 5'-Cy5 labeled TS (ordered from Sangon Biotech) was annealed with another non-labeled NTS at a molar ratio of 1:1 in an annealing buffer (25 mM Tris-HCl pH 8.0, 25 mM KCl), followed by being heated for 5 min at 95°C and slowly cooling down to room temperature. Cleavage reactions were initiated by mixing 500 nM RNP with 10 nM dsDNA substrate in a cleavage buffer (20 mM Tris-HCl pH 7.5 at 37°C , 150 mM NaCl, 10 mM MgCl_2 , 1 mM DTT) at 37°C and product aliquots were collected at the following time points: 0, 2, 5, 10 and 30 min.

For the *trans*-cleavage of a random 55-nt Cy5 labeled ssDNA, 300 nM Cas protein and 450 nM gRNA were mixed and incubated at 25°C for 30 min in a cleavage buffer (20 mM Tris-HCl pH 7.5 at 37°C , 150 mM NaCl, 10 mM MgCl_2 , 1 mM DTT). To activate the *trans*-cleavage activity of the Cas protein, 450 nM cis-dsDNA activator was incubated with the RNP at 37°C for 60 min. Then, the *trans*-cleavage was initiated by adding 10 nM 55-nt ssDNA at 37°C . Sample aliquots were taken at the following time points: 0, 2, 5, 10, 30, and 60 min. The salt concentration for the *trans*-cleavage was adjusted as required.

For the *trans*-cleavage of fluorophore-quencher reporter by PlmCas12e, Cas12e-gRNA complex was pre-assembled by incubating 50 nM PlmCas12e with 62.5 nM gRNA and 1 nM dsDNA activator at

37 °C for 60 min in a reporter cleavage buffer (20 mM Tris-HCl pH 7.5 at 37 °C, 50–300 mM NaCl, 10 mM MgCl₂, 100 ng/μL BSA, 1 mM DTT). F-Q reporter was added into the system on ice to a final concentration of 50 nM. Then, 20 μL mixture was transferred to one well of a 96-well plate. Two wells for each reaction were prepared to reduce random errors. The reactions were incubated in a real-time PCR detection system (Bio-Rad CFX96) for up to 120 min at 37 °C with fluorescence measurements taken per 30 s.

For Michaelis-Menten analysis of PlmCas12e, 100 nM PlmCas12e, 125 nM gRNA, and 2 nM activator were prepared as described above. Reaction was initiated by diluting the complexes to a final concentration of 50 nM PlmCas12e: 62.5 nM gRNA: 1 nM activator (effective complex = 1 nM) by an equal volume of reporter cleavage buffer containing 5, 10, 20, 50, 100, 200, 500, 1000 nM F-Q reporter. Then, 20 μL mixture was transferred to one well of a 96-well plate. Three wells for each reaction were prepared to reduce random errors. The fluorescence was measured as described above. For Michaelis-Menten analysis of LbCas12a, 10 nM LbCas12a, 12.5 nM crRNA, and 0.2 nM activator were prepared as described above. The reaction was initiated by diluting the complexes to a final concentration of 5 nM PlmCas12e: 6.25 nM gRNA: 0.1 nM activator (effective complex = 0.1 nM) by an equal volume of reporter cleavage buffer containing 5, 10, 20, 50, 100, 200, 500, 1000 nM F-Q reporter. The initial velocity of fluorescence was calculated by fitting a linear regression to the data points within 600 s and converted to the initial reaction velocity (V_0). V_0 values from the different reactions were plotted against substrate concentration to determine the Michaelis-Menten constants using (GraphPad Prism), according to the following equation: $Y = V_{\max} * X / (K_m + X)$, where X is the substrate concentration and Y is the V_0 . The turnover number (k_{cat}) was determined by the following equation: $k_{\text{cat}} = V_{\max} / [E]$, where $[E] = 1$ nM for PlmCas12e and 0.1 nM for LbCas12a.

The sequences of the substrates used in cleavage assays are listed in Supplementary Data 3.

Single-molecule FRET assay

Single-molecule FRET experiments were performed at 37 °C using a home-built objective-type total internal reflection fluorescence (TIRF) microscope in single-molecule FRET buffer (50 mM Tris-HCl pH 7.5 at 37 °C, 50/300 mM NaCl, 10 mM MgCl₂) with an oxygen scavenging system containing glucose (3 mg/mL), glucose oxidase (100 μg/mL) (Sigma-Aldrich), catalase (40 μg/mL) (Roche), 1 mM cyclooctatetraene (COT, Sigma-Aldrich), 1 mM 4-nitrobenzylalcohol (NBA, Sigma-Aldrich), 1.5 mM 6-hydroxy-2,5,7,8-tetramethyl-chromane-2-carboxylic acid (Trolox, Sigma-Aldrich). The dCas12e-gRNA complex was assembled at the ratio of 1:1.5 at 25 °C for 40 min before the FRET assays. Cy5 is attached to the target strand in the 7th position from PAM, and Cy3 is attached to the non-target strand in the 16th position from PAM. Target DNA was immobilized on the cover slide via biotin-streptavidin interaction and incubated with 1 μM dCas12e/sgRNA effectors. Fluorescent signals were recorded at 100 ms/frame at 10 mW laser power.

Collected movies were analyzed by a custom-made software program developed as an ImageJ plugin (<http://rsb.info.nih.gov/ij>). Fluorescence spots were fitted by a 2-D Gaussian function within a 9-pixel by 9-pixel area, matching the donor and acceptor spots using a variant of the Hough transform. The background-subtracted total volume of the 2-D Gaussian peak was used as raw fluorescence intensity I . FRET traces displayed anti-correlation behaviors between donor and acceptor fluorescent signals were picked. FRET efficiency (E) of the selected single-molecule traces was calculated as:

$$E = I_{\text{acceptor}} / (I_{\text{donor}} + I_{\text{acceptor}})$$

The nucleotides and related information are provided in Supplementary Data 3.

Reconstitution of the ternary complex

The dCas12e-gRNA complex was assembled by incubating 20 μM dCas12e and 28 μM gRNA at 25 °C for 30 min, followed by additional incubation on ice in a complex buffer (20 mM HEPES pH7.5, 150 mM KCl, 2 mM MgCl₂, 1 mM DTT, 0.25% glycerol) for 30 min on ice. An equal volume of complex buffer containing 30 μM dsDNA was then added to the dCas12e-gRNA complex, followed by incubation for 30 min at 25 °C and 60 min on ice. The sample was loaded into a Superdex 200 10/300 column using a 0.5 mL sample loop buffered in a SEC complex buffer (20 mM HEPES pH7.5, 150 mM KCl, 1 mM TCEP, 0.25% glycerol). Correct fractions were collected after analyzing the protein component by SDS-PAGE and RNA, DNA components by Urea-PAGE. The ternary complex was diluted to 1.1 μM using the SEC complex buffer, and then cross-linked by adding 1/9 volume of freshly prepared 10 mM BS3 (Sigma-Aldrich) cross-linker at 25 °C for 30 min, followed by 60 min on ice.

Cryo-EM sample preparation

To prepare cryo-EM samples, 4 μL of above-mentioned purified complexes were applied to graphene oxide grids (Quantifoil Au 1.2/1.3, 300 mesh), which were glow-discharged (in a HARRICK PLASMA) for 15 s at low level after 2 min evacuation. The grid was then blotted with a pair of 55 mm filter papers (Ted Pella) for 3 s at 8 °C and 100% humidity, and flash-frozen in liquid ethane using an FEI Vitrobot Mark IV. Cryo-EM data were collected on a Titan Krios electron microscope operated at 300 kV and equipped with a Gatan K3 direct electron detector with a Gatan Quantum energy filter. Micrographs were recorded in counting mode at a nominal magnification of 81,000×, resulting in physical pixel sizes of 1.0825 Å. The defocus was set between −1.3 μm and −1.5 μm. The total exposure time of each movie stack led to a total accumulated dose of 50 electrons per Å², which was fractionated into 32 frames.

Image processing and 3D reconstruction

Raw dose-fractionated image stacks were aligned, dose-weighted, and summed using MotionCorr⁶⁵. CTF-estimation, particle picking, 2D reference-free classification, initial model generation, heterogeneous refinement, non-uniform refinement and local resolution estimation were performed in cryoSPARC (version 4.3.0).

Model building and refinement

Protein, RNA, and DNA models were built manually in Coot (version 0.9.8.1) with assistance of AlphaFold2, based on the refined cryo-EM maps. The complete models were then refined against the EM maps by PHENIX (version 1.20.1) in real space with secondary structure and geometry restraints. The final models were validated with the PHENIX and visualized by ChimeraX.

Reporting summary

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

Data availability

The electron density maps have been deposited to the Electron Microscopy Data Bank (EMDB) under the accession numbers [EMD-38856](#) and [EMD-38768](#), which are publicly available as of the date of publication. The atomic coordinates and structure factors have been deposited to the Protein Data Bank (PDB) under the accession numbers [8Y2I](#) and [8XYC](#), which are publicly available as of the date of publication. The raw cryo-EM micrographs used in this study will be shared by corresponding author upon request. Any additional information required to re-analyze the data reported in this paper is available from the corresponding author upon request. Previously published structural models discussed in this work refer to entries [6NY3](#), [7WB1](#), and [7WAZ](#). Sequence data that support the findings of this study are available via the NIH Sequence Read Archive via

BioProject: PRJNA1172643 [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1172643>]. Source data are provided with this paper.

References

- Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
- Pausch, P. et al. CRISPR–CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
- Al-Shayeb, B. et al. Diverse virus-encoded CRISPR–Cas systems include streamlined genome editors. *Cell* **185**, 4574–4586. e4516 (2022).
- Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Sapranaukas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
- Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
- Huang, C. J., Adler, B. A. & Doudna, J. A. A naturally DNase-free CRISPR–Cas12c enzyme silences gene expression. *Mol. Cell* **82**, 2148–2160. e2144 (2022).
- Wu, W. Y. et al. The miniature CRISPR–Cas12m effector binds DNA to block transcription. *Mol. Cell* **82**, 4487–4502. e4487 (2022).
- Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
- Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
- Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell* **60**, 385–397 (2015).
- Harrington, L. B. et al. A scoutRNA is required for some type V CRISPR–Cas systems. *Mol. Cell* **79**, 416–424. e415 (2020).
- Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
- Karvelis, T. et al. PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res.* **48**, 5016–5023 (2020).
- Yan, W. X. et al. Functionally diverse type V CRISPR–Cas systems. *Science* **363**, 88–91 (2019).
- Sun, A. et al. The compact Cas11 (Cas12l) ‘bracelet’ provides a unique structural platform for DNA manipulation. *Cell Res.* **33**, 229–244 (2023).
- Chen, W. et al. Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell* **83**, 2768–2780. e2766 (2023).
- Kong, X. et al. Engineered CRISPR–OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat. Commun.* **14**, 2046 (2023).
- Bigelyte, G. et al. Innate programmable DNA binding by CRISPR–Cas12m effectors enable efficient base editing. *Nucleic Acids Res.* **52**, 3234–3248 (2024).
- Chen, J. S. et al. CRISPR–Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
- Broughton, J. P. et al. CRISPR–Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* **38**, 870–874 (2020).
- Chen, J. CRISPR–Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **371**, 791 (2021). (371, eabh0317, 2021).
- Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR–Cas toolbox and gene editing technologies. *Mol. Cell* **82**, 333–347 (2022).
- Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
- Tsuchida, C. A. et al. Chimeric CRISPR–CasX enzymes and guide RNAs for improved genome editing activity. *Mol. Cell* **82**, 1199–1209. e1196 (2022).
- Altae-Tran, H. et al. Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. *Proc. Natl. Acad. Sci.* **120**, e2308224120 (2023).
- Karvelis, T. et al. Rapid characterization of CRISPR–Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* **16**, 1–13 (2015).
- Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. & Russell, R. Kinetic Basis for DNA Target Specificity of CRISPR–Cas12a. *Mol. Cell* **71**, 816–824. e813 (2018).
- Stephenson, A. A., Raper, A. T. & Suo, Z. Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *J. Am. Chem. Soc.* **140**, 3743–3750 (2018).
- Gasiunas, G. et al. A catalogue of biochemically diverse CRISPR–Cas9 orthologs. *Nat. Commun.* **11**, 5512 (2020).
- Fuchs, R. T., Curcuru, J., Mabuchi, M., Yourik, P. & Robb, G. B. Cas12a trans-cleavage can be modulated in vitro and is active on ssDNA, dsDNA, and RNA. Preprint at <https://biorxiv.org/content/10.1101/600890v1> (2019).
- Smith, C. W. et al. Probing CRISPR–Cas12a Nuclease Activity Using Double-Stranded DNA-Templated Fluorescent Substrates. *Biochemistry* **59**, 1474–1481 (2020).
- Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
- Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
- 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).
- 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. e224 (2017).
- Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA unwinding is the primary determinant of CRISPR–Cas9 activity. *Cell Rep.* **22**, 359–371 (2018).
- Xing, W., Li, D., Wang, W., Liu, J. G. & Chen, C. Conformational dynamics of CasX (Cas12e) in mediating DNA cleavage revealed by single-molecule FRET. *Nucleic Acids Res.* **52**, 9014–9027 (2024).
- Okafor, I. C. et al. Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res.* **47**, 11880–11888 (2019).
- Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
- Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
- 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).
- Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).
- Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).
- Zhang, B. et al. Mechanistic insights into the R-loop formation and cleavage in CRISPR–Cas12i. *Nat. Commun.* **12**, 3476 (2021).
- Abdulkarim, S., Fatimah, A. & Anderson, J. Effect of salt concentrations on the growth of heat-stressed and unstressed *Escherichia coli*. *J. Food Agriculture Environ.* **7**, 51–54 (2009).

48. Shabala, L. et al. Ion transport and osmotic adjustment in *Escherichia coli* in response to ionic and non-ionic osmotic stress. *Environ. Microbiol.* **11**, 137–148 (2009).
49. Singh, A. & Singh, N. Effect of salt concentration on the stability of heterogeneous DNA. *Phys. A: Stat. Mech. its Appl.* **419**, 328–334 (2015).
50. Zhang, L. et al. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nat. Commun.* **12**, 3908 (2021).
51. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
52. Hino, T. et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* **186**, 4920–4935.e4923 (2023).
53. Xu, X. et al. Engineered miniature CRISPR–Cas system for mammalian genome regulation and editing. *Mol. cell* **81**, 4333–4345.e4334 (2021).
54. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
55. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 1–11 (2010).
56. Couvin, D. et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
57. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evolution* **38**, 5825–5829 (2021).
58. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evolution* **37**, 1530–1534 (2020).
59. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
60. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
61. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
62. Salamov, V. S. A. & Solovyev, A. Automatic annotation of microbial genomes and metagenomic sequences. *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, 61–78 (2011).
63. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
64. Deng, P. et al. Structural RNA components supervise the sequential DNA cleavage in R2 retrotransposon. *Cell* **186**, 2865–2879.e2820 (2023).
65. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

Acknowledgements

The work was supported by Science and Technology Innovation - Major Project Youth Scientist Project 2022ZD0213900, National Key Research and Development Program 2022YFF1002801, Ministry of Agriculture and Rural Affairs of China and National Natural Science Foundation of China No.32150018 to J.J.G.L.; National Natural Science Foundation of China No.21877069, 22277063 and 22061160466 to C.C.; Taishan Innovation Industry Leading Talent Project to D.S., National Natural Science Foundation of China No.32101195 and China Postdoctoral Science Foundation,

General Funding No.2022M711848 to S.Z. EM data were collected at the Tsinghua Cryo-EM facility and analyzed using the Bio-Computation platform at the Tsinghua University Branch of the Chinese National Center for Protein Sciences (Beijing). We thank J.L. Lei, X.M. Li, and F. Yang for expert electron microscopy assistance and We thank T. Yang, Y.K. Wang, A.B. Jia, and J.H. Chen for computational support.

Author contributions

J.J.G.L., C.C., and D.S. supervised the project. J.J.G.L., C.C., D.L., S.Z., and W.X. designed the experiments. D.L. purified the proteins, RNAs and reconstituted the ternary complexes. S.Z. prepared the cryo-EM samples and collected the cryo-EM data. S.Z. and S.L. did the structural analysis and built the atomic model. S.L. provided bioinformatics support. D.L. and S.L. analyzed the structural variations among different orthologs. D.L. and Y.Y. performed the cleavage assays and *E. coli* assays and analyzed the results. W.X. performed the FRET assays and analyzed the results. D.L., S.L., and J.J.G.L. wrote the manuscript with the help from all authors including F.Z.

Competing interests

J.J.G.L., S.Z., and D.L. have submitted patent applications to the China patent office pertaining to the following aspects of this work: (1) DNA manipulating tools using CRISPR–LesCas12e, –VemCas12e, –HrbCas12e, and –CkbCas12e systems. Applicant: Tsinghua University. Status: Granted. Application number: 2022106204929 [P]. (2) DNA manipulating tools using CRISPR–OpbCas12e and –Plm2Cas12e systems. Applicant: Tsinghua University. Status: Filed. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54491-9>.

Correspondence and requests for materials should be addressed to Dingding Su, Chunlai Chen or Jun-Jie Gogo Liu.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

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

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

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

© The Author(s) 2024