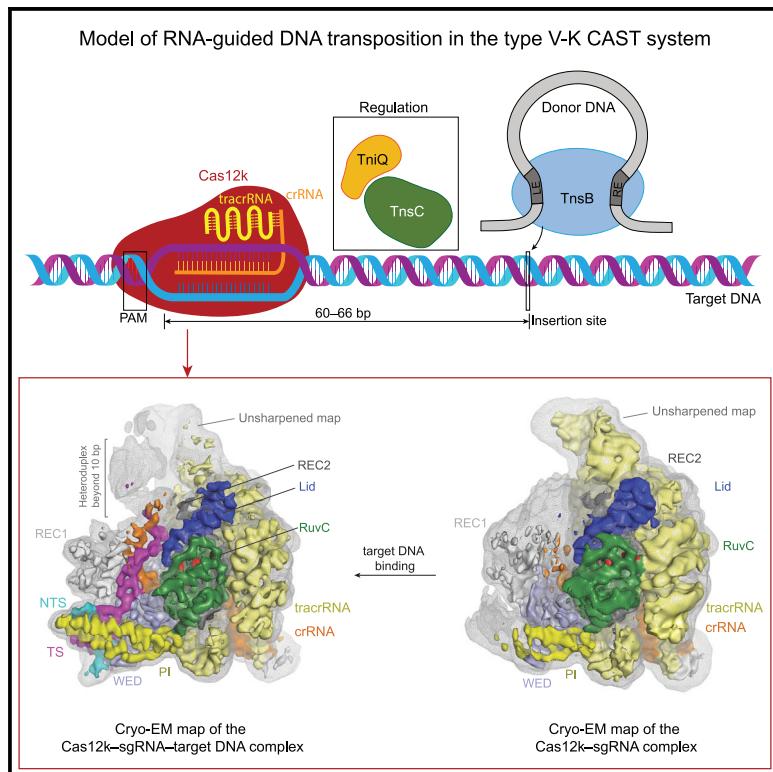


## Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition

### Graphical abstract



### Authors

Renjian Xiao, Shukun Wang, Ruijie Han, Zhuang Li, Clinton Gabel, Indranil Arun Mukherjee, Leifu Chang

### Correspondence

lchang18@purdue.edu

### In brief

CRISPR-associated transposase (CAST) systems allow for RNA-guided DNA insertion at precise locations in genomes and are promising tools for genome editing. Xiao et al. report the structure of Cas12k, the CRISPR effector protein of the type V-K CAST system, and present mechanistic insights into target DNA recognition and RNA-guided transposition.

### Highlights

- Cryo-EM structure of the Cas12k-sgRNA-target DNA complex is presented at 3.6 Å
- Cas12k uses an unusually large sgRNA for RNA-guided DNA transposition
- Identification of key residues for recognition of the PAM sequence
- Cryo-EM structure of the Cas12k-sgRNA complex is solved at 3.8 Å



## Article

# Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition

Renjian Xiao,<sup>1,3</sup> Shukun Wang,<sup>1,3</sup> Ruijie Han,<sup>1</sup> Zhuang Li,<sup>1</sup> Clinton Gabel,<sup>1</sup> Indranil Arun Mukherjee,<sup>1</sup> and Leifu Chang<sup>1,2,4,\*</sup>

<sup>1</sup>Department of Biological Sciences, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA

<sup>2</sup>Purdue University Center for Cancer Research, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: lchang18@purdue.edu

<https://doi.org/10.1016/j.molcel.2021.07.043>

## SUMMARY

The type V-K CRISPR-Cas system, featured by Cas12k effector with a naturally inactivated RuvC domain and associated with Tn7-like transposon for RNA-guided DNA transposition, is a promising tool for precise DNA insertion. To reveal the mechanism underlying target DNA recognition, we determined a cryoelectron microscopy (cryo-EM) structure of Cas12k from cyanobacteria *Scytonema hofmanni* in complex with a single guide RNA (sgRNA) and a double-stranded target DNA. Coupled with mutagenesis and *in vitro* DNA transposition assay, our results revealed mechanisms for the recognition of the GGTT protospacer adjacent motif (PAM) sequence and the structural elements of Cas12k critical for RNA-guided DNA transposition. These structural and mechanistic insights should aid in the development of type V-K CRISPR-transposon systems as tools for genome editing.

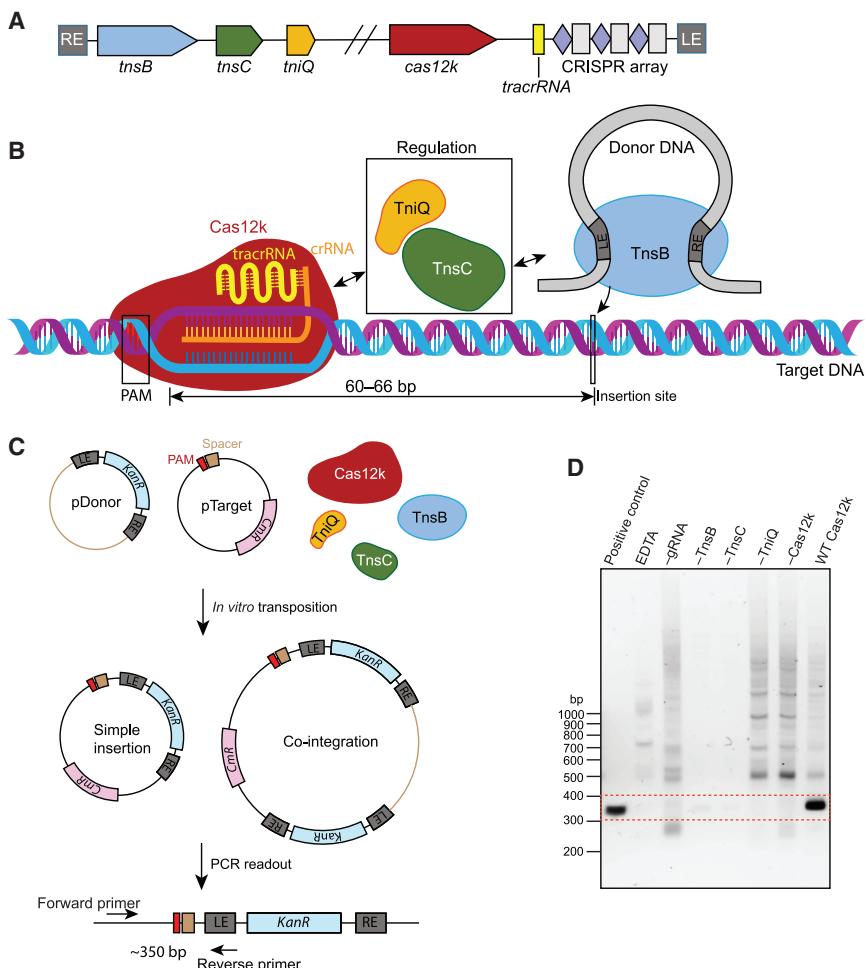
## INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein (Cas) systems are adaptive immunity systems in bacteria and archaea against mobile genetic elements (MGEs) and have been developed as tools for genome editing (Mohanraju et al., 2016; Sorek et al., 2013). These systems use guide RNAs and effector proteins to specifically target MGEs for degradation. The arms race between prokaryotes and foreign MGEs has resulted in diverse CRISPR-Cas systems that are divided into two classes (1 and 2) and six different types (I-VI) (Makarova et al., 2015; Shmakov et al., 2015). The class 2 type V system, featured by a conserved C-terminal RuvC nuclease domain in its effector Cas12 proteins, is abundant and further classified into 11 subtypes (V-A–V-K) (Makarova et al., 2020; Yan et al., 2019), which are promising for the development of genome editing tools. Among them, Cas12a, Cas12b, and Cas12e have been successfully applied in genome editing (Liu et al., 2019; Strecker et al., 2019a; Teng et al., 2018; Zetsche et al., 2015), such as gene knockout utilizing non-homologous end joining (NHEJ) to gene knockin using homology-directed repair (HDR) (Moreno-Mateos et al., 2017). However, the CRISPR-based gene knockin in mammalian cells relies heavily on endogenous HDR during the S/G2 phase of the cell cycle (Gratz et al., 2014; Moreno-Mateos et al., 2017). Recent discoveries in transposon-associated CRISPR-Cas systems have shed light on the development of an efficient knockin method independent of host DNA repair pathways, including

the type V-K, type I-F, and type I-B systems (Faure et al., 2019; Klompe et al., 2019; Saito et al., 2021; Strecker et al., 2019b). Among those systems, the type V-K system has the advantage of a small effector protein that is more amenable for delivery into mammalian cells.

Cas12k, the effector protein of the type V-K CRISPR-Cas system, was first identified with a featured inactive RuvC nuclease domain, which led to the subsequent discovery of its association with Tn7-like transposons (Faure et al., 2019; Strecker et al., 2019b). Strecker et al. (2019b) showed that a CRISPR-associated transposase from cyanobacteria *Scytonema hofmanni* (ShCAST) can be directed and inserted to target sites 60–66 bp downstream of the protospacer adjacent motif (PAM). The ShCAST system contains a CRISPR module composed of Cas12k with a CRISPR array and a transposon module (Figures 1A and 1B). The transposon module contains a single component transposase TnsB (similar to MuA transposase in the transposable phage Mu [Montaño et al., 2012]), an AAA+ (ATPases associated with diverse cellular activities) regulator TnsC, a target selector TniQ (homolog of TnsD in Tn7 transposon [Faure et al., 2019]), and the left and right ends of the transposon. This system was successfully repurposed for efficient RNA-guided DNA insertion in *Escherichia coli* (Strecker et al., 2019b). However, further development of this tool for genome editing, biomedical applications, and the eventual treatment of human diseases requires a deeper understanding of the molecular mechanism underlying RNA-guided DNA transposition.





**Figure 1. RNA-guided DNA transposition in the ShCAST system**

(A) Schematic of the genomic organization of the ShCAST system.

(B) Schematic of RNA-guided DNA transposition by components in the ShCAST system.

(C) Schematic for the *in vitro* DNA transposition assay.

(D) PCR results using reaction mixture of *in vitro* DNA transposition assay as template. Positions of expected PCR readout at ~350 bp are indicated by a red dashed box. A plasmid with expected insertion confirmed by sequencing is used as positive control. The result shown is representative of more than three experiments. See also Figure S1.

Both simple insertion and co-integration products were observed in RNA-guided insertions (Rice et al., 2020; Strecker et al., 2020; Vo et al., 2021a), with co-integration being the major product in our experiments as revealed by restriction enzyme digestion and DNA sequencing (Figures S1E and S1F).

#### Overall structure of Cas12k-sgRNA-target DNA

To understand the mechanism of RNA-guided target DNA recognition, we assembled a Cas12k-sgRNA-target DNA ternary complex by incubating Cas12k, sgRNA, and a target DNA containing a GGTT PAM sequence (Figure S1G; Table S1). Using cryo-EM, we reconstructed a

map of this ternary complex at 3.6 Å resolution (Figures 2 and S2; Table 1) that allowed us to build the atomic model (Figure S3A), except for residues 103–270 of Cas12k, the CRISPR RNA (crRNA)-target DNA heteroduplex beyond 10 bp from the PAM duplex, and small regions (e.g., 1–8 nt of sgRNA), which are not resolved in the map most likely due to flexibility.

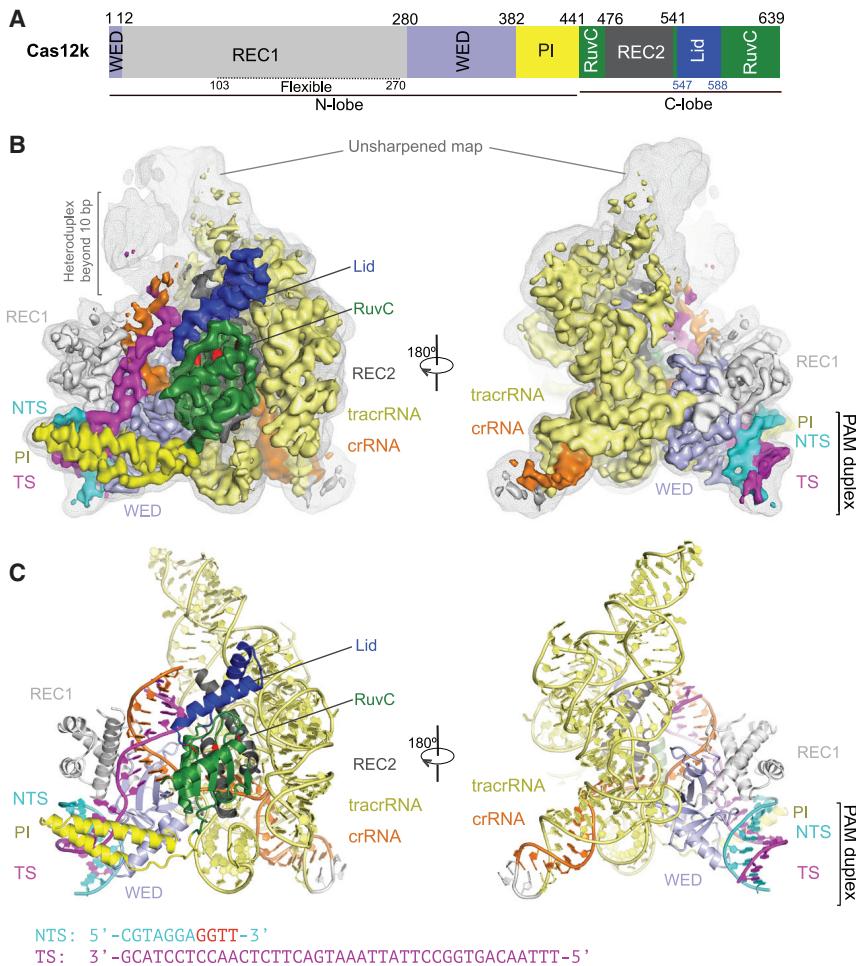
The overall structure of Cas12k resembles other Cas12 proteins, with Cas12f as the closest match by a search in the distance matrix alignment (DALI) server (Z score, 14.5) (Holm et al., 2008; Takeda et al., 2021; Xiao et al., 2021; Figures 3A and 3B). The 637-residue protein adopts a bi-lobed structure connected by a loop. The N-terminal lobe of Cas12k is composed of the WED, REC1, and PI domains. The WED domain, which plays a major role in recognizing sgRNA, contains seven strands ( $\beta$ 1– $\beta$ 7) with a helix  $\alpha$ 5 inserted between  $\beta$ 5 and  $\beta$ 6 (Figure 3C). The REC1 domain is inserted between  $\beta$ 1 and  $\beta$ 2 of the WED domain and composed of an N-terminal helical bundle  $\alpha$ 1– $\alpha$ 4 (REC1<sup>13–102</sup>) and a C-terminal flexible region (REC1<sup>103–270</sup>) that is predicted to form 6–7 helices (Figures 3D and S3B). Although sharing low sequence similarity, REC1<sup>13–102</sup> is structurally similar to the REC1<sup>C</sup> domain in Cas12f (Figures 3D and S3B and S3C), which forms the dimerization interface in Cas12f (Takeda et al., 2021; Xiao et al., 2021). However, the key hydrophobic residues

Here, we report the cryoelectron microscopy (cryo-EM) structures of a Cas12k-single guide RNA (sgRNA)-target DNA ternary complex and a Cas12k-sgRNA binary complex. The structures, combined with *in vitro* transposition assay, provide mechanistic insights into target recognition and RNA-guided DNA transposition by the ShCAST system.

## RESULTS

### In vitro DNA transposition

We first purified Cas12k, sgRNA, and transposition proteins (TnsB, TnsC, and TniQ) in the ShCAST system and tested their function using a previously established *in vitro* DNA transposition assay (Strecker et al., 2019b; Figures 1C and S1A). Our results suggest that TnsB, TnsC, and magnesium (required for transposon end cleavage and target joining [Skelding et al., 2002]) are strictly required for DNA transposition, whereas additional components including Cas12k, sgRNA, and TniQ are necessary for RNA-guided DNA transposition (Figures 1D and S1B–S1D). Omitting any of the latter three components leads to DNA transposition in a non-RNA-guided manner. Out of ten randomly selected colonies from the assay using all components, eight RNA-guided and two non-RNA-guided insertions were observed (Figure S1D).



(I118, Y121, Y122, and I126) in Cas12f are not conserved in REC1<sup>13–102</sup> (Figures 3D and S3B), which may be a reason why a Cas12k dimer is not observed. Following the WED domain is a PI domain composed of two helices,  $\alpha$ 6 and  $\alpha$ 7, which is absent in Cas12f but observed in some other Cas12 proteins such as Cas12i (Zhang et al., 2020).

The C-terminal lobe of Cas12k is composed of the RuvC and REC2 domains. Both the sequence and structure of the RuvC domain of Cas12k are conserved relative to Cas12f; however, the Cas12f's triplet of acidic residues required for nuclease activity is replaced by either serine or proline in Cas12k (Figures 3E and S3B and S3C). A Cas12k mutant restoring the catalytic residues (S452D, P546E, and P619D) did not reinstate target DNA cleavage (Figure S1H). In addition to the altered catalytic residues, two additional features are observed in the RuvC domain of Cas12k compared with that of Cas12f. First, there is no Nuc domain in Cas12k (Figures 3A, 3B, 3E, and S3B). The Nuc domains or equivalent domains are inevitably present in all Cas12 proteins with structures determined to date, including Cas12a (Dong et al., 2016; Gao et al., 2016; Nishimasu et al., 2017; Stella et al., 2017, 2018; Swarts and Jinek, 2019; Swarts et al., 2017; Yamano et al., 2016, 2017; Zhang et al., 2019), Cas12b (Liu et al., 2017; Wu et al., 2017; Yang et al., 2016), Cas12e

**Figure 2. Overall structure of the Cas12k-sgRNA-target DNA complex**

(A) Schematic of domain organization of Cas12k based on structure.

(B) Cryo-EM map of the Cas12k-sgRNA-target DNA complex at 3.6 Å in two views, with each domain color coded as in (A). The unsharpened map is shown in gray mesh.

(C) Atomic model of the Cas12k-sgRNA-target DNA complex shown in cartoon in the same views as in (B). Schematic of target DNA used for cryo-EM analysis is shown at the bottom. See also Figure S2.

(Liu et al., 2019), Cas12f (Takeda et al., 2021; Xiao et al., 2021), Cas12g (Li et al., 2021), and Cas12i (Huang et al., 2020; Zhang et al., 2020), and play an essential role in the nuclease activity. Second, the lid motif of Cas12k is longer than that of Cas12f and is in a closed conformation that covers the pseudonuclease site (Figures 3A and 3E). Both features are consistent with the lack of nuclease activity in the RuvC domain of Cas12k. Taken together, Cas12k is closely related to Cas12f in structure, but may have evolved these different features to meet the requirements for DNA transposition.

## Structure of sgRNA

The 265-nt sgRNA is composed of a 44-nt crRNA (30-nt spacer and 14-nt repeat)

and a 218-nt *trans*-activating CRISPR RNA (tracrRNA) connected by a short 3-nt linker (**Figures 4A** and **4B**; **Table S1**). The sgRNA contains two anti-repeat:repeat (AR:R) duplexes, AR:R1 and AR:R2, formed by the repeat region of crRNA (from -1 to -5 and from -6 to -14) and the AR sequences of tracrRNA (87–91 and 211–218). The AR:R1 duplex, clamped by the WED and REC2 domains, is the major region on the sgRNA that bridges the N- and C-lobes of Cas12k (**Figures 4C** and **S4A**). The AR:R2 duplex makes no contact with Cas12k and could function mainly for the recognition between crRNA and tracrRNA (**Figure 2C**).

The tracrRNA part (1–218) contains five stem loops formed by local regions including P1 (17–44), P3 (66–83), P6 (122–135), P7 (163–175), and P8 (195–210) and three duplexes formed by long-distance regions such as P2 (56–59 and 94–97), P4 (100–108 and 179–187), and P5 (112–121 and 150–160) (Figures 4A and 4B). Between P1 and P2 is a pseudoknot structure formed by three fragments (9–16, 45–55, and 140–149). P1 and the pseudoknot contact the C-lobe of Cas12k (Figures 4C and S4B), whereas P3 and P8 interact with the N-lobe of Cas12k primarily through electrostatic interactions (Figures 4C and S4C). A previous study showed that deletion of 1–47 nt in the sgRNA abolished RNA-guided DNA insertion, indicating that S1 is essential (Strecker et al., 2019b). P4–P7 show no contact with Cas12k

**Table 1. Cryo-EM data collection, refinement, and validation statistics**

	Cas12k-sgRNA-t arget DNA	Cas12k-sgRNA
<b>Data collection and processing</b>		
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	54	54
Defocus range (μm)	0.8–2.0	0.8–2.0
Pixel size (Å)	1.05	1.05
Symmetry imposed	C1	C1
Initial particle image (no.)	3,021,295	2,664,830
Final particle image (no.)	183,870	114,383
Map resolution (Å)	3.65	3.80
FSC threshold	0.143	0.143
Map resolution range (Å)	3.2~20	3.3~20
<b>Refinement</b>		
Initial model used	none	PDB: 7N3P
Model resolution (Å)	3.9	3.9
FSC threshold	0.5	0.5
Model resolution range (Å)	3.2–4.5	3.3–4.5
Map sharpening	–176.9	–196.4
B factor (Å <sup>2</sup> )		
Model composition		
Non-hydrogen atom	9,075	8,315
Protein residue	454	454
Nucleotide	260	223
Ligand	N/A	N/A
B factor (Å <sup>2</sup> )		
Protein	82.84	86.95
Nucleotide	149.88	134.96
Ligand	N/A	N/A
<b>RMSDs</b>		
Bond length (Å)	0.003	0.003
Bond angle (°)	0.579	0.593
<b>Validation</b>		
MolProbity score	1.49	1.79
Clashscore	8.04	8.28
Poor rotamer (%)	0.00	0.00
<b>Ramachandran plot</b>		
Favored (%)	97.76	96.41
Allowed (%)	2.24	3.59
Disallowed (%)	0.00	0.00

and display considerable flexibility as revealed by 3D variance analysis (Figures S2D–S2F).

### PAM recognition

The PAM duplex is enclosed in a positively charged groove formed by the REC1, WED, and PI domains (Figure 5A). All three domains

contribute residues that directly interact with the bases of the GGTT PAM sequence for sequence-specific recognition (Figures 5B and 5C and S4D and S4E). Specifically, R78 from the REC1 domain establishes bidentate hydrogen bond with the guanine base of G(–3) of the non-target strand (NTS). The hydroxyl group of T287 from the WED domain forms hydrogen bonds with the adenine base of A(–2) of the target strand (TS). R421 from the PI domain interacts with both T(–1) and T(–2) from the TS and NTS, respectively. Those three residues, particularly R78, are conserved in Cas12k (Figure 5A). In addition, a number of polar or positively charged residues recognize the PAM duplex through the phosphate backbones (Figures 5C and S4D and S4E). To test the structural observations, we mutated the three residues that recognize the bases and two positively charged residues that bind to phosphate backbones, R350 and R428 from the WED and PI domains, respectively. Alanine substitution of any of the residues reduced *in vitro* RNA-guided transposition activity by PCR readout (Figures 5D and S5A–S5C).

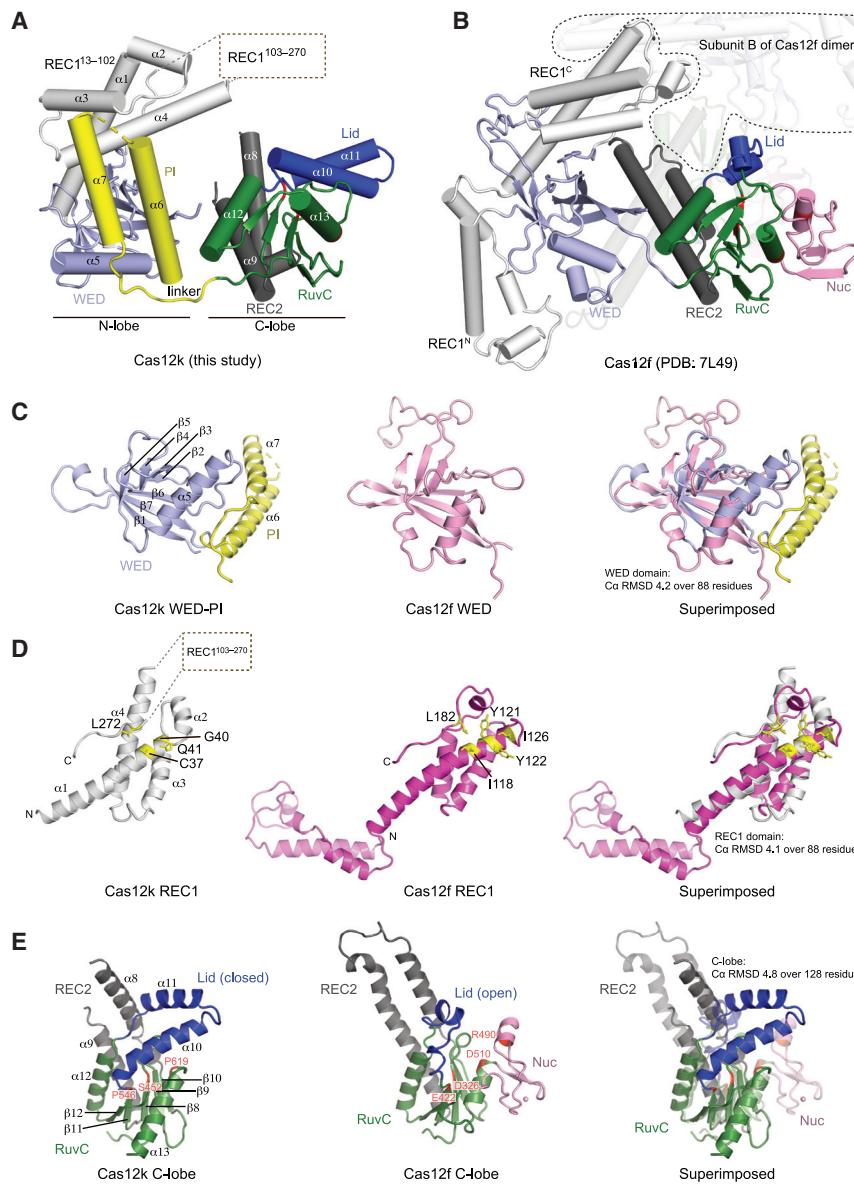
### crRNA-DNA heteroduplex recognition

A 10-bp crRNA-target DNA heteroduplex is observed in the ternary complex, which primarily contacts the REC1 and REC2 domains, as well as the lid motif (Figures 4C and S4F and S4G). The heteroduplex beyond 10 bp is likely formed, but in a flexible state, as it is visible in the unsharpened map at a lower contour level (Figures 2B and S4H). This shorter stabilized heteroduplex is in contrast to the usual 20-bp heteroduplex observed in other Cas12 structures where the PAM distal end of the heteroduplex is stabilized by either relatively large REC1 and REC2 domains (e.g., Cas12a [Dong et al., 2016; Gao et al., 2016; Nishimasu et al., 2017; Stella et al., 2017, 2018; Swarts and Jinek, 2019; Swarts et al., 2017; Yamano et al., 2016, 2017; Zhang et al., 2019], Cas12b [Liu et al., 2017; Wu et al., 2017; Yang et al., 2016], Cas12e [Liu et al., 2019], and Cas12i [Huang et al., 2020; Zhang et al., 2020]) or a second protein subunit (e.g., Cas12f [Takeda et al., 2021; Xiao et al., 2021]). The flexible REC1<sup>103–270</sup> might provide additional interactions with the heteroduplex. Cas12k with REC1<sup>103–270</sup> deleted ( $\Delta$ 103–270GSGS) is inactive for RNA-guided DNA transposition (Figures 5D and S5A–S5C).

Guided by our observation of a shorter stabilized heteroduplex, we set out to determine the minimum spacer length in sgRNA required for RNA-guided DNA transposition. We designed sgRNA with various spacer length, including 6, 8, 10, 12, 14, 16, 18, and 20 nt (Table S1). Our *in vitro* DNA transposition assay suggests that at least a 14-nt spacer length is required for detectable DNA transposition, and at least 16-nt is required for optimized activity (Figures 5E and S5D). This is consistent with a recent study showing that 16 nt is both sufficient and near the minimum length required for insertion in the type V-K system (Saito et al., 2021). This result suggests that one checkpoint for transposition in the type V-K system is the formation of the crRNA-target DNA heteroduplex at 14–16 bp.

### Conformational changes induced by target DNA recognition

To understand the conformational changes in Cas12k upon target DNA recognition, we reconstructed a cryo-EM structure



**Figure 3. Structure of Cas12k and comparison with Cas12f**

(A and B) Atomic models of Cas12k (A) and Cas12f (PDB: 7L49) (B). The subunit A of Cas12f is shown in the same view as Cas12k, whereas the subunit B is semi-transparent. The catalytic residues in Cas12f (B) and their equivalents in Cas12k (A) are highlighted in red.

(C–E) Comparison of the domains between Cas12k and Cas12f. See also Figure S3.

deletion of the lid motif ( $\Delta$ 549–588GSGS) abolishes RNA-guided DNA transposition (Figures 5D and S5A–S5C), suggesting it plays an essential function. Consistently, the lid motif in Cas12k is highly conserved (Figure 6F).

## DISCUSSION

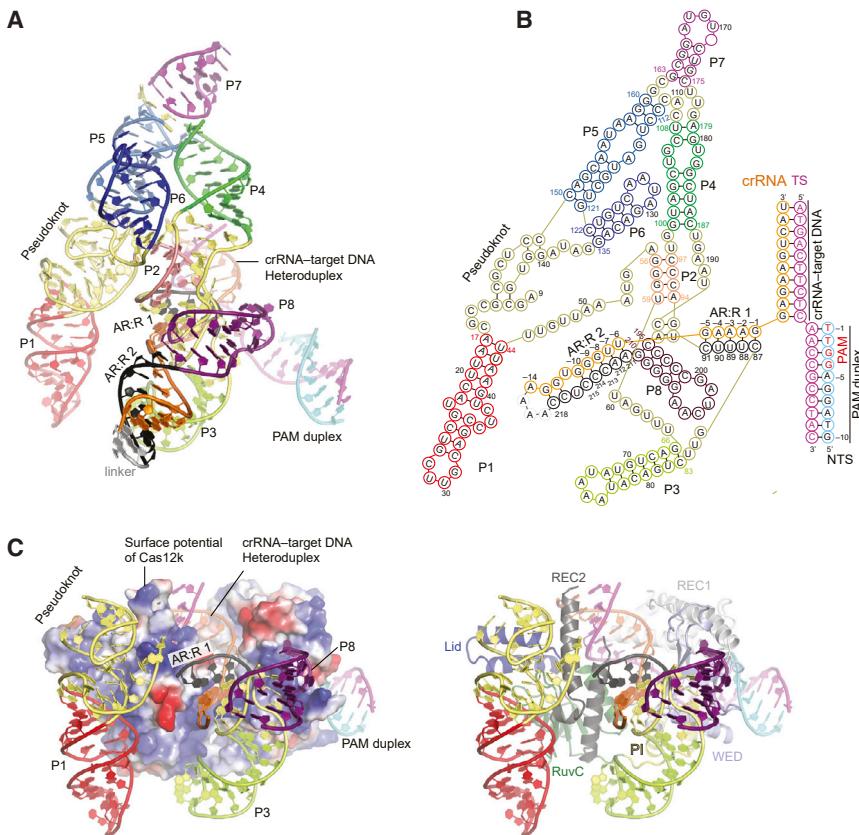
RNA-guided DNA transposition by the type V-K CAST system requires both the CRISPR module and the transposition module (Strecker et al., 2019b). The CRISPR module recognizes target DNA using guide RNAs and recruits the transposition machinery for DNA insertion. In this study, we showed the cryo-EM structure of Cas12k and the mechanism of target DNA recognition by the CRISPR module of the ShCAST system. The structure of Cas12k reported here is consistent with a preprint posted on bioRxiv recently (Querques et al., 2021).

Despite sharing similar architecture with other Cas12 proteins, Cas12k displays considerable differences that may be related to its association with the Tn7-like transposon, including an inactive RuvC domain, the absence of the Nuc domain, and a longer and closed lid motif.

Although undergoing no closed-to-open conformational change upon recognition of target DNA, the lid motif is kept in Cas12k and essential for RNA-guided DNA transposition. Given the essential function of the lid motif and the lack of nuclease activity of Cas12k, we speculate that the lid motif might play a role in either stabilization of the structure or the recruitment of transposition proteins.

Four stem loops (P4–P7) within the sgRNA show no interactions with Cas12k, raising a question about their function. Interestingly, when sgRNA is removed from the *in vitro* transposition assay, the number of colonies is significantly larger than that of other conditions (Figure S1B); however, none of the tested colonies show RNA-guided DNA insertion. This result may suggest that sgRNA might play an inhibitory role in the transposition machinery for non-RNA-guided DNA insertion. This may not be surprising because to direct the transposon machinery for RNA-

of the Cas12k-sgRNA binary complex at 3.8 Å (Figures 6A, 6B, and S6; Table 1). Although still largely flexible, REC1<sup>103–270</sup> is more visible in the binary complex and contacts the lid motif in the RuvC domain (Figure 6C). Structural superimposition with the ternary structure reveals minimal conformational change in Cas12k, with the exception of the REC1 domain that undergoes an ~6–8 Å shift to contact target DNA (Figures 6D and 6E; Video S1). This movement positions key residues (e.g., R78) for PAM recognition and also makes room for the crRNA-target DNA heteroduplex, which otherwise would clash with REC1 (Video S1). To be noted, the lid motif adopts a similar closed conformation in both the ternary and binary complex. The closed-to-open transition of the lid motif upon target DNA recognition is shown as a conserved mechanism for activation of the RuvC nuclease activity in Cas12 proteins (Stella et al., 2018; Xiao et al., 2021; Zhang et al., 2020). However,



guided DNA transposition, the CRISPR-Cas system may have evolved a mechanism to inhibit the transposon's original activity.

Recent studies showed the role of the AAA+ protein TnsC in transposition target site selection (Park et al., 2021; Querques et al., 2021; Shen et al., 2021). In the ShCAST system, TnsC forms filament structure on DNA that is capped by TniQ (Park et al., 2021). The TniQ-TnsC assembly is likely directly associated to Cas12k, establishing connection between the CRISPR module and the transposition module. The interactions between TnsB transposase and TnsC could direct DNA insertion in a fixed position relative to the target DNA recognition site of the CRISPR module, which is 60–66 bp downstream of the PAM in the ShCAST system. Although the core mechanisms of RNA-guided DNA transposition are likely conserved, there could be considerable differences between different systems. For example, it was previously reported that TniQ is bound to the CRISPR effector complex, the Cascade complex, in the type I-F CAST system (Halpin-Healy et al., 2020; Jia et al., 2020; Li et al., 2020; Wang et al., 2020). However, Cas12k-sgRNA or Cas12k-sgRNA-target DNA does not readily bind to TniQ according to our experiments and a recent study (Querques et al., 2021). The crystal structure of TniQ in the ShCAST system (ShTniQ) was determined recently (Querques et al., 2021) and shows considerable differences from TniQ in the type I-F system from *Vibrio cholerae* (VcTniQ). ShTniQ lacks the C-terminal HTH domain of VcTniQ, which is involved in dimerization of VcTniQ. ShTniQ also lacks the C-terminal helical domain prior the HTH domain, which contributes

**Figure 4. Overall structure of sgRNA**

(A) Structure of the sgRNA and target DNA in the Cas12k-sgRNA-target DNA complex in cartoon presentation with stem-loops (S1–S8), pseudoknot, and AR:R 1–2 duplexes color coded.

(B) Schematic of the sgRNA and target DNA, color coded as in (A).

(C) Contacts between sgRNA and Cas12k. Stem loops not in contact with Cas12k are not shown. Cas12k in surface potential and cartoon representations are shown in left and right panels, respectively. See also Figure S4.

the major binding site for Cas6 of the Cascade complex through hydrophobic interactions. The Cas12k structure reported here and these recent studies are beginning to unravel the underlying mechanism for RNA-guided DNA transposition for the type V-K systems.

Transposon-associated CRISPR-Cas systems are promising tools for gene insertion application; however, possible off-target insertion raises concerns because it can cause genome instability. In the case of the ShCAST system, non-RNA-guided DNA insertion is observed in *in vitro* transposition assay (Figure S1D) and in *E. coli* (Strecker et al., 2019b). The reduced specificity and co-integration

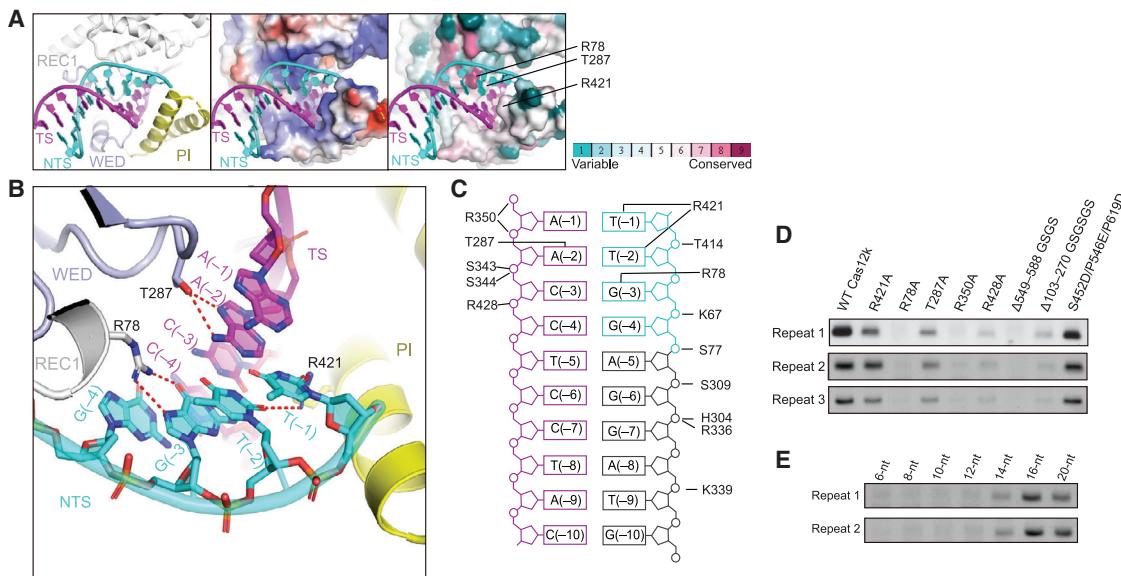
mechanism may limit the application of the type V-K system versus others such as the type I-F system (Vo et al., 2021b). Using linear or 5'-nicked DNA donors may prevent co-integration (Strecker et al., 2020); however, to reduce or eliminate unwanted DNA insertions, further studies are required to understand detailed mechanisms in the ShCAST system, including the interactions between the Cas12k-sgRNA-target DNA module and the whole transposition machinery; this will be especially vital for taking advantage of the small size of the type V-K systems for genome editing applications.

#### Limitations of the study

We determined the structures of Cas12k-sgRNA before and after target DNA binding and tested structural observations by *in vitro* transposition assay. However, more work is required to understand the structure of transposition proteins and more importantly the interactions between Cas12k and transposition proteins in the ShCAST system. The structures of TnsC and TniQ have been recently determined (Park et al., 2021; Querques et al., 2021). Further work is required to reveal the structure of TnsB bound to donor DNA and how this complex is recruited to the target site by a concerted action of Cas12k, TnsC, and TniQ.

#### STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:



**Figure 5. Target DNA recognition by Cas12k**

- (A) The PAM duplex is bound to a positively charged groove by REC1, WED, and PI domains. Left, middle, and right images show the cartoon model, surface potential, and conservation map, respectively.
- (B) Detailed interactions between the PAM duplex and Cas12k. Interactions are indicated by red dashed lines.
- (C) Schematic of the interactions between the PAM duplex and Cas12k.
- (D) PCR results of *in vitro* DNA transposition assay using reaction mixture as template using wild-type Cas12k and various Cas12k mutants. The results shown are from three replicates.
- (E) PCR results of *in vitro* DNA transposition assay using sgRNA with different spacer length. The results shown are from two replicates. See also Figures S4 and S5.

#### • KEY RESOURCES TABLE

#### • RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

#### • EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### • METHOD DETAILS

- Protein expression and purification
- sgRNA preparation
- Mutagenesis
- *In vitro* transposition assay
- Polymerase Chain Reaction (PCR)
- Electron Microscopy
- Image Processing
- Model building, refinement, and validation
- Structure-based sequence alignment
- Structural conservation
- Structural visualization

#### • QUANTIFICATION AND STATISTICAL ANALYSIS

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2021.07.043>.

#### ACKNOWLEDGMENTS

We thank Thomas Klose for help with cryo-EM, Steven Wilson for computation, and Heng Zhang for cloning at the initiation stage of this study. This

work made use of the Purdue Cryo-EM Facility. C.G. is supported by a grant from the NIH (T32GM132024). This work was supported by the NIH grant R01GM138675 to L.C.

#### AUTHOR CONTRIBUTIONS

L.C. supervised the study. R.X., S.W., and R.H. prepared samples. Z.L., R.X., and L.C. collected and processed cryo-EM data. S.W., R.X., and R.H. performed biochemical analysis with help from C.G. and I.A.M. All authors analyzed the data. R.X., S.W., R.H., and L.C. prepared the manuscript with input from other authors.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 8, 2021

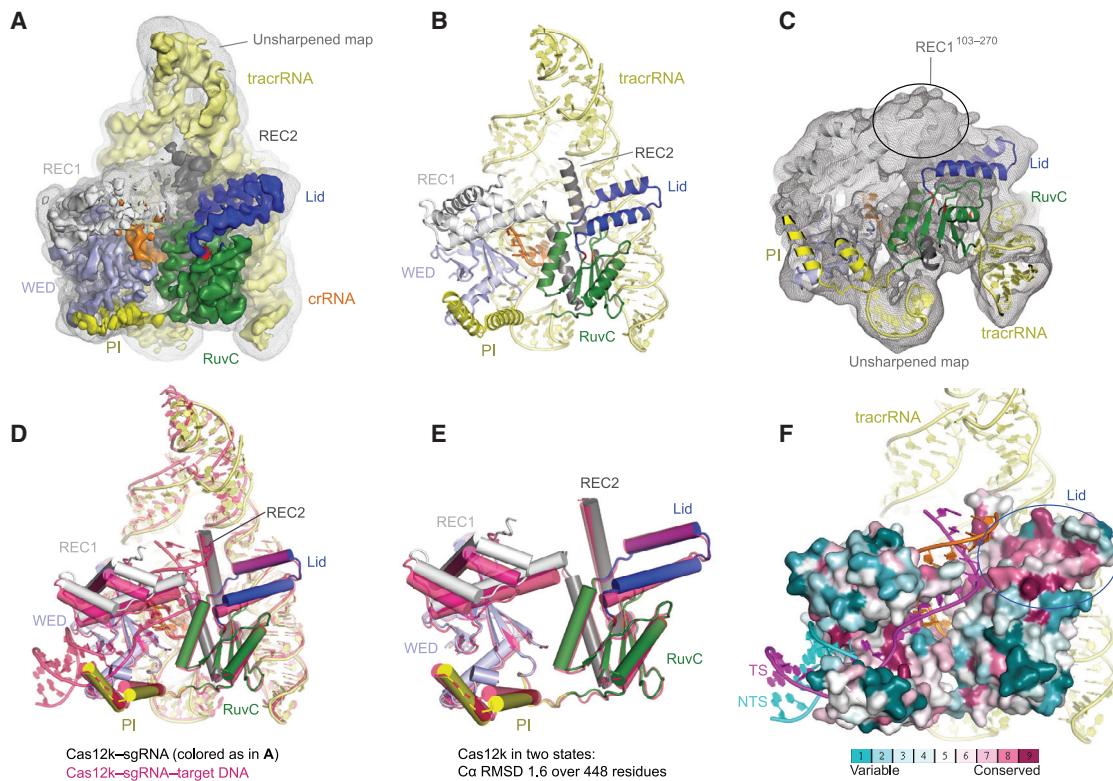
Revised: July 15, 2021

Accepted: July 29, 2021

Published: August 26, 2021

#### REFERENCES

- Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44** (W1), W344–W350.
- Biesiada, M., Purzycka, K.J., Szachniuk, M., Blazewicz, J., and Adamiak, R.W. (2016). Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol. Biol.* **1490**, 199–215.



**Figure 6. Structure of the Cas12k-sgRNA complex**

- (A) Cryo-EM map of the Cas12k-sgRNA complex at 3.8 Å, with each subunit color coded as in Figure 1A. The unsharpened map is shown in gray mesh.  
 (B) Atomic model of the Cas12k-sgRNA complex shown in cartoon in the same views as in (A).  
 (C) Cryo-EM density of REC1<sup>103-270</sup> (circled) in the binary complex.  
 (D) Structural superimposition of Cas12k-sgRNA (color-coded as in A) and Cas12k-sgRNA-target DNA (magenta) complexes.  
 (E) Structural superimposition of Cas12k protein in the two states shown in (D).  
 (F) Structural conservation map of Cas12k. Shown is the Cas12k-sgRNA-target DNA complex. The conserved lid motif is indicated by a circle. See also Figure S6.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 532, 522–526.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.

Faure, G., Shmakov, S.A., Yan, W.X., Cheng, D.R., Scott, D.A., Peters, J.E., Makarova, K.S., and Koonin, E.V. (2019). CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* 17, 513–525.

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* 26, 901–913.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38, W695–W699.

Gratz, S.J., Ukkun, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 196, 961–971.

Halpin-Healy, T.S., Klompe, S.E., Sternberg, S.H., and Fernández, I.S. (2020). Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature* 577, 271–274.

Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008). Searching protein structure databases with DALI Lite v.3. *Bioinformatics* 24, 2780–2781. <https://doi.org/10.1093/bioinformatics/btn507>.

Huang, X., Sun, W., Cheng, Z., Chen, M., Li, X., Wang, J., Sheng, G., Gong, W., and Wang, Y. (2020). Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.* 11, 5241.

Jia, N., Xie, W., de la Cruz, M.J., Eng, E.T., and Patel, D.J. (2020). Structure-function insights into the initial step of DNA integration by a CRISPR-Cas Transposon complex. *Cell Res.* 30, 182–184.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 571, 219–225.

Li, Z., Zhang, H., Xiao, R., and Chang, L. (2020). Cryo-EM structure of a type I-F CRISPR RNA guided surveillance complex bound to transposition protein TrIQ. *Cell Res.* 30, 179–181.

Li, Z., Zhang, H., Xiao, R., Han, R., and Chang, L. (2021). Cryo-EM structure of the RNA-guided ribonuclease Cas12g. *Nat. Chem. Biol.* 17, 387–393.

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., and Wang, Y. (2017). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Mol. Cell* 65, 310–322.

Liu, J.J., Orlova, N., Oakes, B.L., Ma, E., Spinner, H.B., Baney, K.L.M., Chuck, J., Tan, D., Knott, G.J., Harrington, L.B., et al. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 566, 218–223.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An

- updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83.
- Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353, aad5147.
- Montaño, S.P., Pigli, Y.Z., and Rice, P.A. (2012). The  $\mu$  transpososome structure sheds light on DDE recombinase evolution. *Nature* 491, 413–417.
- Moreno-Mateos, M.A., Fernandez, J.P., Rouet, R., Vejnar, C.E., Lane, M.A., Mis, E., Khokha, M.K., Doudna, J.A., and Giraldez, A.J. (2017). CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat. Commun.* 8, 2024.
- Nishimasu, H., Yamano, T., Gao, L., Zhang, F., Ishitani, R., and Nureki, O. (2017). Structural basis for the altered PAM recognition by engineered CRISPR-Cpf1. *Mol. Cell* 67, 139–147.e132.
- Park, J.-U., Tsai, A., Mehrotra, E., Petassi, M.T., Shan-Chi, H., Ke, A., Peters, J.E., and Kellogg, E.H. (2021). Structural basis for target-site selection in RNA-guided DNA transposition systems. *Science*. <https://doi.org/10.1126/science.abi8976>.
- Pei, J., Kim, B.H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36, 2295–2300.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Punjani, A., and Fleet, D.J. (2021). 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* 213, 107702.
- Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296.
- Querques, I., Schmitz, M., Oberli, S., Chanez, C., and Jinek, M. (2021). Molecular mechanism of target site selection and remodeling by type V CRISPR-associated transposons. *bioRxiv*. <https://doi.org/10.1101/2021.07.06.451292>.
- Rice, P.A., Craig, N.L., and Dyda, F. (2020). Comment on “RNA-guided DNA insertion with CRISPR-associated transposases”. *Science* 368, eabb2022.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324.
- Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745.
- Saito, M., Ladha, A., Strecker, J., Faure, G., Neumann, E., Altae-Tran, H., Macrae, R.K., and Zhang, F. (2021). Dual modes of CRISPR-associated transposon homing. *Cell* 184, 2441–2453.e18.
- Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W., et al. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 49 (D1), D10–D17.
- Shen, Y., Gomez-Blanco, J., Petassi, M.T., Peters, J.E., Ortega, J., and Guarné, A. (2021). Structural basis for DNA targeting by the Tn7 transposon. *bioRxiv*. <https://doi.org/10.1101/2021.05.24.445525>.
- Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* 60, 385–397.
- Skelding, Z., Sarnovsky, R., and Craig, N.L. (2002). Formation of a nucleoprotein complex containing Tn7 and its target DNA regulates transposition initiation. *EMBO J.* 21, 3494–3504.
- Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* 82, 237–266.
- Stella, S., Alcón, P., and Montoya, G. (2017). Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* 546, 559–563.
- Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcon, P., Jensen, S.B., Saligram, B., Moses, M.E., Hatzakis, N.S., and Montoya, G. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* 175, 1856–1871.e1821.
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104, 1104.
- Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019a). Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* 10, 212.
- Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019b). RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 365, 48–53.
- Strecker, J., Ladha, A., Makarova, K.S., Koonin, E.V., and Zhang, F. (2020). Response to Comment on “RNA-guided DNA insertion with CRISPR-associated transposases”. *Science* 368, eabb2920.
- Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* 151, 41–60.
- Swarts, D.C., and Jinek, M. (2019). Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Mol. Cell* 73, 589–600.e4.
- Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* 66, 221–233.e4.
- Takeda, S.N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., and Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell* 81, 558–570.e3.
- Teng, F., Cui, T., Feng, G., Guo, L., Xu, K., Gao, Q., Li, T., Li, J., Zhou, Q., and Li, W. (2018). Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov.* 4, 63.
- Vo, P.L.H., Acree, C., Smith, M.L., and Sternberg, S.H. (2021a). Unbiased profiling of CRISPR RNA-guided transposition products by long-read sequencing. *Mob. DNA* 12, 13.
- Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H., and Sternberg, S.H. (2021b). CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* 39, 480–489.
- Wang, B., Xu, W., and Yang, H. (2020). Structural basis of a Tn7-like transposase recruitment and DNA loading to CRISPR-Cas surveillance complex. *Cell Res.* 30, 185–187.
- Wu, D., Guan, X., Zhu, Y., Ren, K., and Huang, Z. (2017). Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Res.* 27, 705–708.
- Xiao, R., Li, Z., Wang, S., Han, R., and Chang, L. (2021). Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res.* 49, 4120–4128.
- Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* 165, 949–962.
- Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Mol. Cell* 67, 633–645.e633.

- Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V., et al. (2019). Functionally diverse type V CRISPR-Cas systems. *Science* 363, 88–91.
- Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* 167, 1814–1828.e1812.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771.
- Zhang, H., Li, Z., Daczkowski, C.M., Gabel, C., Mesecar, A.D., and Chang, L. (2019). Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins. *Cell Host Microbe* 25, 815–826.e4.
- Zhang, H., Li, Z., Xiao, R., and Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.* 27, 1069–1076.
- Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
BL21-CodonPlus (DE3)-RIL Competent cell	Agilent Technologies	Cat# 230245
Rosetta (DE3)pLysS Competent cell	Novagen	Cat# 70956
One Shot PIR1 Chemically Competent cell	Thermo fisher scientific	Cat# C101010
Stellar Competent Cell	Takara Bio	Cat# 636763
<b>Chemicals, peptides, and recombinant proteins</b>		
Terrific Broth (TB)	Thermo fisher scientific	Cat# BP97285
isopropyl β-D-thiogalactopyranoside (IPTG)	Thermo fisher scientific	Cat# R0393
Tris-Base	Thermo fisher scientific	Cat# BP152-1
HEPES	Thermo fisher scientific	Cat# BP310-1
Sodium chloride	Thermo fisher scientific	Cat# S271-10
Glycerol	Thermo fisher scientific	Cat# G33-500
2-Mercaptoethanol	Thermo fisher scientific	Cat# O3446I-100
Magnesium chloride	Thermo fisher scientific	Cat# M33-500
PMSF	Sigma-Aldrich	Cat# P7626
Imidazole	Sigma-Aldrich	Cat# I202-2KG
DTT	Thermo fisher scientific	Cat# AAJ1539722
EDTA	Thermo fisher scientific	Cat# AM9261
Potassium Chloride	Thermo fisher scientific	Cat# P217-500
Magnesium acetate tetrahydrate	Millipore Sigma	Cat# M5661-250G
Agarose	Thermo fisher scientific	Cat# BP1356-500
SYBR safe DNA Gel Stain	Thermo fisher scientific	Cat# S33102
Tobacco Etch Virus (TEV) protease	Homemade	N/A
FspI	New England Biolabs	Cat# R0135L
HindIII-HF	New England Biolabs	Cat# R3104S
BsrRI	New England Biolabs	Cat# R0581L
Bsal	New England Biolabs	Cat# R3733S
Proteinase K	Thermo fisher scientific	Cat# EO0491
Benzonase Nuclease	Millipore Sigma	Cat# E1014-25KU
ATP	Thermo fisher scientific	Cat# BP413-25
Phusion Hot Start II High-Fidelity PCR Master Mix	Thermo fisher scientific	Cat# F565L
Phusion High-Fidelity PCR Master Mix with HF Buffer	Thermo fisher scientific	Cat# F531L
PrimeSTAR Max Premix	Takara Bio	Cat# R045A
In-Fusion HD Cloning Kit	Takara Bio	Cat# 639650
Q5 Site-Directed Mutagenesis Kit	New England Biolabs	Cat# E0554
HiScribe T7 Quick High Yield RNA Synthesis Kit	New England Biolabs	Cat# E2050S
QIAprep Spin Miniprep Kit	QIAGEN	Cat# 27106
MinElute PCR Purification Kit	QIAGEN	Cat# 28006
ZymoPURE II Plasmid Gigaprep Kit	Zymo Research	Cat# D4204-B
Mini-PROTEAN TGX Precast Gels	Bio-rad	Cat# 4561086
Amicon Ultra-15 Centrifugal Filter Units 3kDa NMWL	Millipore sigma	Cat# UFC900324

(Continued on next page)

***Continued***

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Amicon Ultra-15 Centrifugal Filter Units 10kDa NMWL	Millipore sigma	Cat# UFC901024
Amicon Ultra-15 Centrifugal Filter Units 30kDa NMWL	Millipore sigma	Cat# UFC903024
Amicon Ultra-15 Centrifugal Filter Units 50kDa NMWL	Millipore sigma	Cat# UFC905024
NuPAGE LDS Sample Buffer (4X)	Thermo fisher scientific	Cat# NP0007
Gel Loading Dye, Purple (6X)	New England Biolabs	Cat# B7024S
Bovine Serum Albumin	Sigma-Aldrich	Cat# A2153-10G
Ampicillin, sodium salt	Thermo fisher scientific	Cat# BP1760-25
Chloramphenicol, 98%, ACROS Organics	Thermo fisher scientific	Cat# AC227920250
Kanamycin Monosulfate, USP Grade	Gold Biotechnology	Cat# K-120-5
SnakeSkin Dialysis Tubing, 7K MWCO, 22 mm	Thermo fisher scientific	Cat# 68700
Diethyl pyrocarbonate	Millipore sigma	Cat# D5758-25ML
<b>Critical commercial assays</b>		
Ni-NTA agarose resin	QIAGEN	Cat# 30430
HiTrap Heparin HP affinity column	Cytiva	Cat# 17040701
RESOURCE Q anion exchange column	Cytiva	Cat# 17117701
Superdex 200 10/300 column	Cytiva	Cat# 28990944
InstantBlue Coomassie Protein Stain	Abcam	Cat# ISB1L
<b>Deposited data</b>		
Cas12k-sgRNA-target DNA coordinate	This study	PDB: 7N3P
Cas12k-sgRNA-target DNA map	This study	EMDB: EMD-24143
Cas12k-sgRNA coordinate	This study	PDB: 7N3O
Cas12k-sgRNA map	This study	EMDB: EMD-24142
<b>Oligonucleotides</b>		
DNA primers	IDT	<a href="#">Table S1</a>
DNA oligos (for structure determination)	IDT	<a href="#">Table S1</a>
Cas12k sgRNA	IDT	<a href="#">Table S1</a>
Cas12k sgRNA variants	IDT	<a href="#">Table S1</a>
<b>Recombinant DNA</b>		
pDonor_ShCAST_kanR	Addgene	#127924
pTarget_CAST	Addgene	#127926
pET28-MKH8SUMO-ShTnsB	This paper	N/A
pET28-MKH8SUMO-ShTnsC	This paper	N/A
pET28-MKH8SUMO-ShTnIQ	This paper	N/A
pET-His6-StrepII-TEV LIC-Cas12k	This paper	N/A
pET-His6-StrepII-TEV LIC-Cas12k variant mutants	This paper	N/A
<b>Software and algorithms</b>		
PyMOL	Schrodinger LLC	<a href="https://pymol.org/2/">https://pymol.org/2/</a>
UCSF Chimera	Pettersen et al., 2004	<a href="https://www.cgl.ucsf.edu/chimera/">https://www.cgl.ucsf.edu/chimera/</a>
Phenix	Afonine et al., 2018	<a href="https://phenix-online.org/">https://phenix-online.org/</a>
Coot	Emsley et al., 2010	<a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/</a>
CryoSPARC	Punjani et al., 2017	<a href="https://cryosparc.com/">https://cryosparc.com/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Leginon	Suloway et al., 2005	<a href="https://emg.nysbc.org/redmine/projects/leginon/wiki/Leginon_Homepage">https://emg.nysbc.org/redmine/projects/leginon/wiki/Leginon_Homepage</a>
CTFFIND4	Rohou and Grigorieff, 2015	<a href="https://grigoriefflab.umassmed.edu/ctffind4">https://grigoriefflab.umassmed.edu/ctffind4</a>
MotionCor2	Zheng et al., 2017	<a href="https://github.com/singleparticle/MotionCor2">https://github.com/singleparticle/MotionCor2</a>
PROMALS3D	Pei et al., 2008	<a href="http://prodata.swmed.edu/promals3d/promals3d.php">http://prodata.swmed.edu/promals3d/promals3d.php</a>
ESPrpt	Robert and Gouet, 2014	<a href="https://esprpt.ibcp.fr/ESPrpt/ESPrpt/">https://esprpt.ibcp.fr/ESPrpt/ESPrpt/</a>
Sequence Manipulation Suite	Stothard, 2000	<a href="https://www.bioinformatics.org/sms2/ident_sim.html">https://www.bioinformatics.org/sms2/ident_sim.html</a>
GraphPad Prism 8	GraphPad Software	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
National Center for Biotechnology Information (NCBI)	Sayers et al., 2021	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Leifu Chang ([lchang18@purdue.edu](mailto:lchang18@purdue.edu)).

### Materials availability

Plasmids generated in this study will be made available upon request made to the lead contact.

### Data and code availability

- Cryo-EM reconstructions of Cas12k–sgRNA–target DNA and Cas12k–sgRNA complexes have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-24143 and EMD-24142, respectively. Coordinates for atomic models of Cas12k–sgRNA–target DNA and Cas12k–sgRNA complexes have been deposited in the Protein Data Bank under the accession numbers 7N3P and 7N3O, respectively.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

The plasmid DNAs were amplified in *E. coli* DH5 $\alpha$ , and One Shot PIR1 cells. Cas12k and transposition proteins were expressed in *E. coli* BL21-CodonPlus (DE3)-RIL and Rosetta (DE3)pLysS cells. Products of *in vitro* transposition were transformed into Stellar competent cells for amplification.

## METHOD DETAILS

### Protein expression and purification

Gene fragments of TnsB, TnsC, and TniQ were ordered from Integrated DNA Technologies (IDT) and cloned into bacterial expression plasmid pET28-MKH8SUMO (Addgene: #79526). The gene fragment for Cas12k was cloned into the bacterial expression plasmid pET-His6-StrepII-TEV LIC (Addgene: #29718). Cas12k, TnsB, and TniQ were expressed in *Escherichia coli* BL21-CodonPlus (DE3)-RIL (Agilent Technologies: # 230245) while TnsC was expressed in Rosetta(DE3)pLysS (Novagen: #70956) containing a pLysS-tRNA plasmid. Cells were grown to OD<sub>600</sub> = 0.6 in Terrific Broth (TB) and protein expression was induced by adding 0.3 mM of IPTG followed by overnight incubation at 16°C. The cells were collected and resuspended in lysis buffer (50 mM Tris-HCl, pH 7.6, 500 mM NaCl, 5% glycerol) supplemented with 1 mM PMSF and 5 mM β-mercaptoethanol, and then disrupted by sonication. Cell lysate was clarified by centrifugation. The supernatant was loaded onto Ni-NTA resin. After extensive washing with lysis buffer supplemented with 30 mM imidazole, target proteins were eluted with lysis buffer supplemented with 250 mM imidazole. The His-SUMO tag of TnsB, TnsC, and TniQ and His-StrepII tag of Cas12k were removed by overnight digestion with TEV protease at

4°C. The protein was diluted with buffer containing 50 mM Tris-HCl pH 7.6, 200 mM NaCl, and 5% glycerol and loaded onto a Heparin column (GE Healthcare), eluted with a linear NaCl gradient (0.1 to 1M). After concentration, the proteins were further purified by size exclusion chromatography (SEC) over a Superdex 200 increase 10/300 GL column (Cytiva) in buffer containing 25 mM Tris-HCl (pH 7.6), 500 mM NaCl, 10% glycerol, and 1 mM DTT (0.5 mM EDTA was added to the buffer for TnsC). Fractions were concentrated and stored at -80°C.

To assemble the Cas12k-sgRNA binary complex, Cas12k proteins were incubated with sgRNA (Table S1) at a ratio of 1:1.15 at 37°C for 30 min in buffer A (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 2 mM DTT and 1 mM MgCl<sub>2</sub>). To reconstitute the Cas12k-sgRNA-target DNA ternary complex, Cas12k protein was incubated with sgRNA at 37°C for 30 min followed by the addition of target DNA synthesized from IDT (Table S1) at a ratio of 1:1.1.5:1.3. After 30 min, the mixture was subjected to SEC over a Superdex 200 column (Cytiva) equilibrated with buffer A for further purification.

### sgRNA preparation

sgRNAs were produced by *in vitro* transcription using the HiScribe T7 High Yield RNA synthesis kit (NEB) with PCR amplified gBlocks (IDT) as templates. sgRNAs were purified over a Resource-Q column (Cytiva) and eluted with a linear NaCl gradient (50 mM–1000 mM) in 25 mM Tris-HCl, pH 8.0. The eluted sgRNAs were concentrated and stored at -80°C.

### Mutagenesis

Single amino acid mutations were introduced by the QuikChange site-directed mutagenesis method. Mutations with multiple amino acids were introduced by ligating inverse PCR-amplified backbone with mutations bearing DNA oligonucleotides via the In-Fusion Cloning Kit (ClonTech). All mutants were confirmed by Sanger sequencing.

### *In vitro* transposition assay

Donor plasmid (pDonor, kanamycin resistance) and target plasmid (pTarget, chloramphenicol resistance) were gifts from Feng Zhang (Addgene #127924 and #127926, respectively). *In vitro* transposition reaction was conducted as previously described unless otherwise stated. All proteins were diluted to 2 μM with 25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, and 25% glycerol. 50 nM of each proteins, 600 nM sgRNA, 20 ng pTarget, and 100 ng pDonor were added sequentially to the reaction buffer containing 26 mM HEPES pH 7.5, 4.2 mM Tris-HCl pH 8.0, 2.1 mM DTT, 0.05 mM EDTA, 0.2 mM MgCl<sub>2</sub>, 28 mM NaCl, 21 mM KCl, 1.35% glycerol, 50 μg/mL BSA, and 2 mM ATP (final pH 7.5) to a total volume of 20 μL. Reactions were incubated at 30°C for 40 min before being supplemented with 25 mM MgOAc<sub>2</sub> and incubated at 37°C for another 2 hours. 1 μL of the final products was taken out for direct PCR readout. The remaining sample was digested with 1 μL of Proteinase K (Thermo Fisher Scientific) at 37°C for 15 min before transformation into Stellar competent cells. Colonies were grown on kanamycin and chloramphenicol plates. Single colonies were randomly picked for plasmid preparation. After extraction, plasmids were analyzed by PCR, restriction enzyme (BamHI) digestion and sanger sequencing.

pDonor contains R6K-γ origin, which requires protein pi (encoded by the gene pir) to initiate replication. The Stellar competent cell used for *in vitro* transposition does not have protein pi, so pDonor cannot replicate in this host. Therefore, transformation of both pTarget and pDonor will not result in colonies. To obtain pDonor, One Shot PIR1 competent cell (Thermo Fisher Scientific) was used.

### Polymerase Chain Reaction (PCR)

Forward primer pTarget\_F, reverse primer pDonor\_R (Table S1), and *in vitro* transposition reaction product were mixed to a final volume of 25 μL for PCR reactions. Cycling conditions were as follows: 1 cycle, 94°C, 3 min; 35 cycles, 98°C, 10 s, 66.9°C, 15 s, 72°C, 8 s; 1 cycle, 72°C, 10 min. Plasmids extracted from single colonies were analyzed by PCR under cycling conditions as follows: 1 cycle, 98°C, 3 min; 35 cycles, 98°C, 10 s, 69.9°C, 15 s, 72°C, 12 s; 1 cycle, 72°C, 10 min.

### Electron Microscopy

Aliquots of 4 μL Cas12k-sgRNA binary complex (1 mg/mL) and Cas12k-sgRNA-dsDNA ternary complex (1 mg/mL) were applied to glow-discharged UltrAuFoil holey gold grids (R1.2/1.3, 300 mesh). The grids were blotted for 2 s and plunged into liquid ethane using a Vitrobot Mark IV. Cryo-EM data were collected with a Titan Krios microscope operated at 300 kV and images were collected using Leginon (Suloway et al., 2005) at a nominal magnification of 81,000x (resulting in a calibrated physical pixel size of 1.05 Å/pixel) with a defocus range of 0.8–2.0 μm. The images were recorded on a K3 electron direct detector in super-resolution mode at the end of a GIF-Quantum energy filter operated with 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<sup>2</sup>. Statistics for cryo-EM data are listed in Table 1.

### Image Processing

Movie frames were aligned using MotionCor2 (Zheng et al., 2017) with a binning factor of 2. The motion-corrected micrographs were imported into cryoSPARC (Punjani et al., 2017). Contrast transfer function (CTF) parameters were estimated using CTFFIND4 (Rohou and Grigorieff, 2015). A few thousand particles were auto-picked without template to generate 2D averages for subsequent template-based auto-picking. The auto-picked and extracted particles were processed for 2D classifications, which were used to exclude false

and bad particles that fell into 2D averages with poor features. An initial reconstruction was done in cryoSPARC using 100,000 particles (Punjani et al., 2017). Heterogenous refinement was further performed to sort out different conformational heterogeneity. To further screen homogeneous particles, 3D variance analysis (Punjani and Fleet, 2021) was performed and the resulting maps with different conformations (frame\_000.mrc and frame\_019.mrc) are used for supervised heterogenous refinement. The homogeneous dataset was used for final 3D refinement with C1 symmetry, resulting in 3.65 Å resolution from 183,870 particles based on the FSC = 0.143 criterion (Rosenthal and Henderson, 2003).

The Cas12k-sgRNA binary complex dataset were processed in a similar way as the ternary complex. 114,383 particles were selected for a final reconstruction at 3.80 Å resolution. Local resolution estimations were performed in cryoSPARC. Cryo-EM image processing is summarized in Table 1.

#### Model building, refinement, and validation

*De novo* model building of the Cas12k-sgRNA-target DNA structure was performed manually in COOT (Emsley et al., 2010) guided by secondary structure predictions from PSIPRED (Jones, 1999) of Cas12k protein and structure prediction of sgRNA by RNAComposer (Biesiada et al., 2016). Refinement of the structure models against corresponding maps were performed using the *phenix.real\_space\_refine* tool in Phenix (version 1.19.2) (Afonine et al., 2018). For the Cas12k-sgRNA complex, the structure model of the Cas12k-sgRNA-target-DNA complex was fitted into the cryo-EM map with models for target DNA deleted. The model is adjusted by all-atom refinement in COOT with self-restraints. The resultant model was refined against the corresponding cryo-EM map using the *phenix.real\_space\_refine* tool in Phenix.

#### Structure-based sequence alignment

PROMALS3D program (Pei et al., 2008) was used to align the sequences of Cas12k and Cas12f based on structure. The alignment diagram was plotted using ESPript (Robert and Gouet, 2014). Sequence identities and similarities were calculated using Sequence Manipulation Suite (Stothard, 2000). Root-mean-square deviation (RMSD) of the C $\alpha$  atomic was calculated using the *cealign* command in PyMOL.

#### Structural conservation

Structural conservation was calculated using ConSurf (Ashkenazy et al., 2016) based on multiple sequence alignment using Clustal Omega (Goujon et al., 2010). Sequences of Cas12k are from the ShCAST and AcCAST systems (Strecker et al., 2019b) and 86 Cas12k orthologs in the National Center for Biotechnology Information (NCBI) (Sayers et al., 2021).

#### Structural visualization

Figures were generated using PyMOL and UCSF Chimera (Pettersen et al., 2004).

#### QUANTIFICATION AND STATISTICAL ANALYSIS

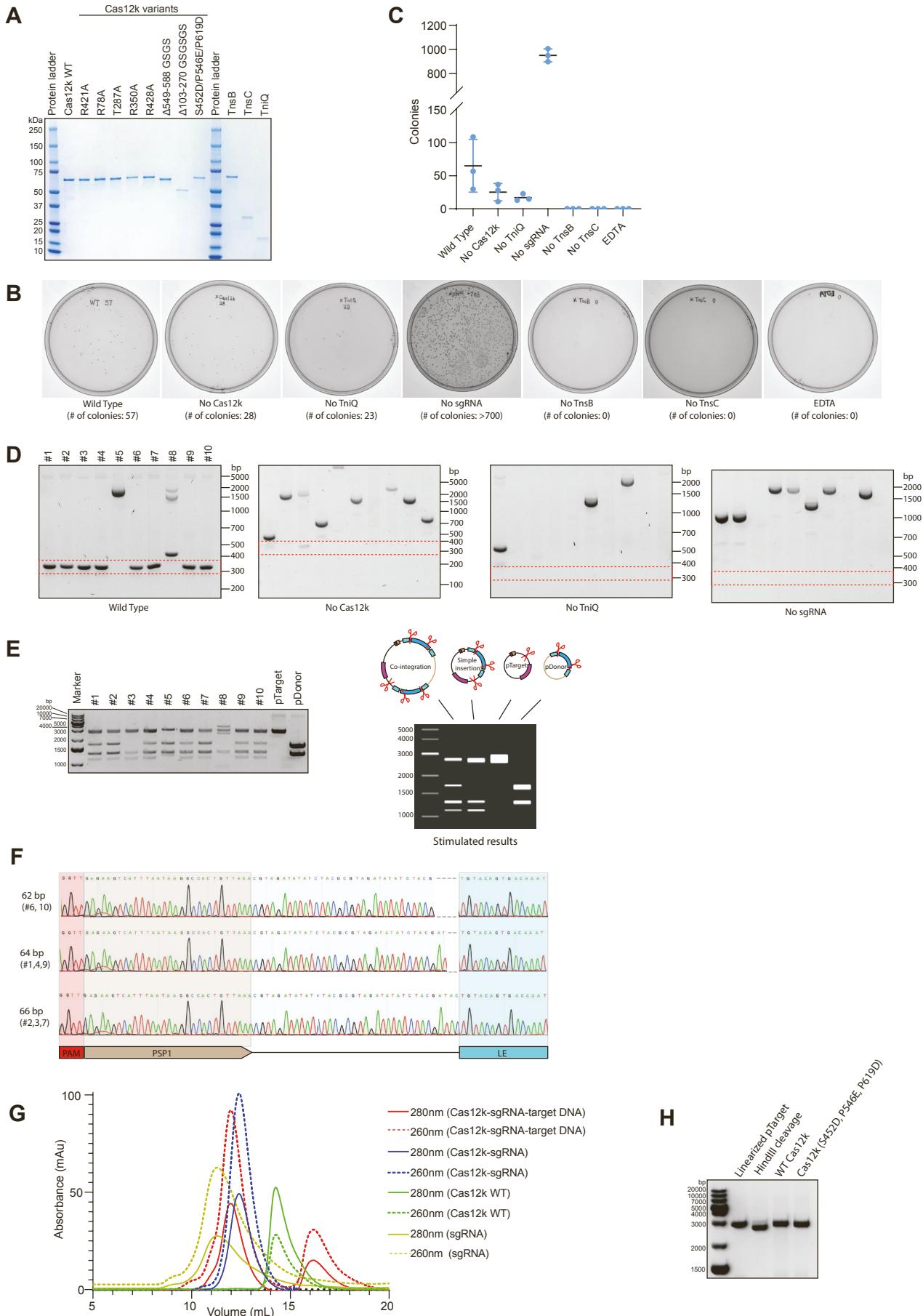
The number of replicates for each experiment is indicated in the corresponding figure legend. In Figures S1C and S5A, colony numbers for *in vitro* transposition were represented as mean  $\pm$  standard deviation (SD). Data were processed and plotted using GraphPad Prism 8. Statistical validation for the final structural models shown in Table 1 was performed using Phenix (Afonine et al., 2018).

**Supplemental information**

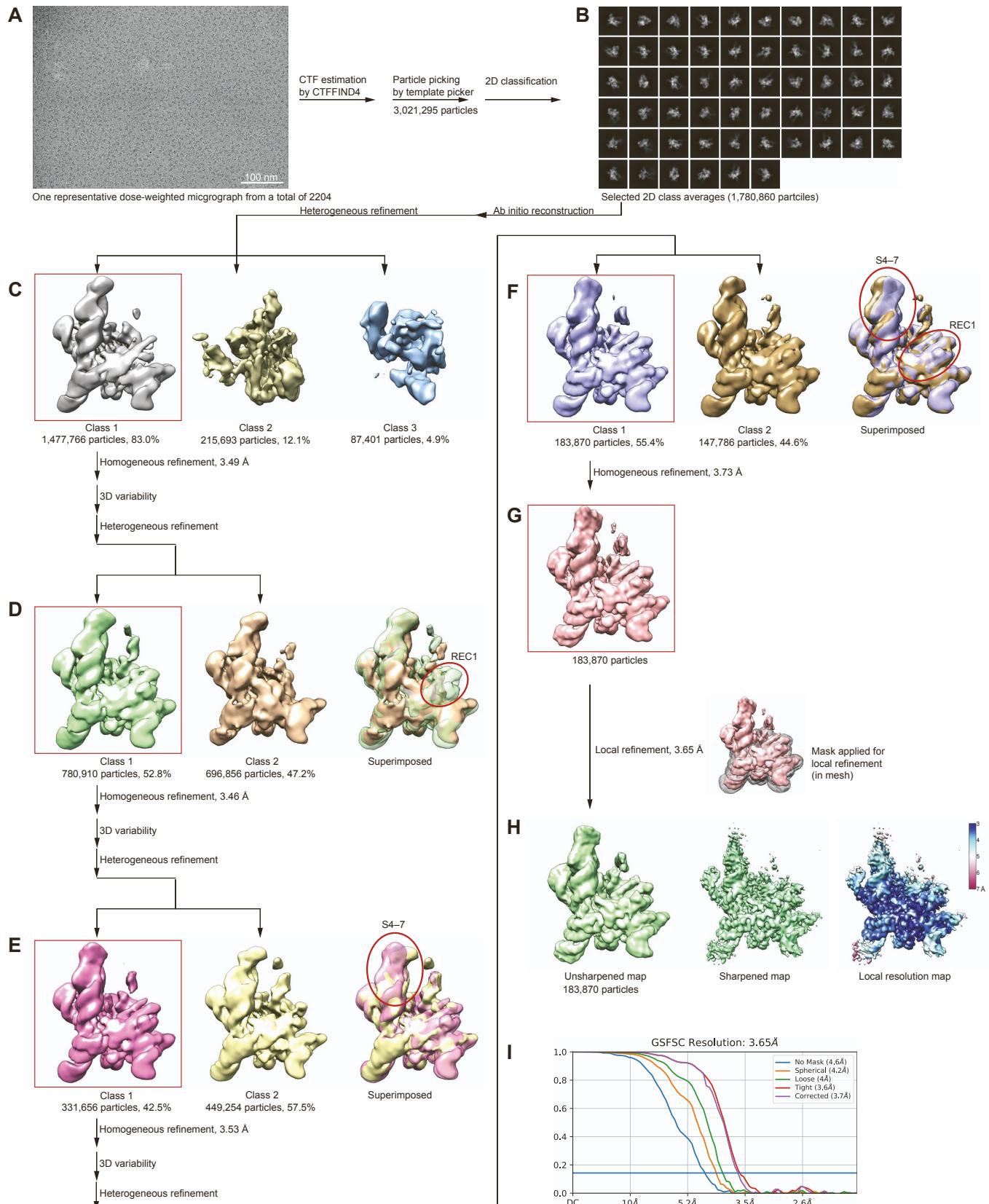
**Structural basis of target DNA recognition**

**by CRISPR-Cas12k for RNA-guided DNA transposition**

**Renjian Xiao, Shukun Wang, Ruijie Han, Zhuang Li, Clinton Gabel, Indranil Arun Mukherjee, and Leifu Chang**



**Figure S1. *In vitro* DNA transposition assay.** **(A)** SDS-PAGE of purified proteins used for the *in vitro* DNA transposition assay. **(B,C)** LB agar plate showing colonies after transformation of each transposition reaction. Graph in **F** shows mean $\pm$ SD ( $n=3$ ). **(D)** PCR results using purified plasmid as template. Ten colonies are randomly selected for plasmid extraction from each plate in **E**. Positions of expected PCR readout at ~350 bp are indicated by red dashed boxes. **(E)** Restriction enzyme digestion assay of ten plasmids extracted from the wild type condition in **D**. A simulated result based on the plasmid sequences is shown on the right. **(F)** Sanger sequencing of PCR products in the wild type condition in **D**. **(G)** Superimposed size exclusion chromatography profiles of Cas12k samples, including the Cas12k–sgRNA–target DNA ternary complex (red), the Cas12k–sgRNA binary complex (blue), Cas12k (green) and sgRNA (yellow), with UV absorbance curves at 280 nm and 260 nm shown in solid and dashed lines, respectively. **(H)** Target DNA (linearized pTarget plasmid) cleavage assay using wild type Cas12k and Cas12k with the catalytic acidic residues in RuvC restored by mutations. The results shown in **B–E** are representative of more than three experiments. Related to Figure 1.

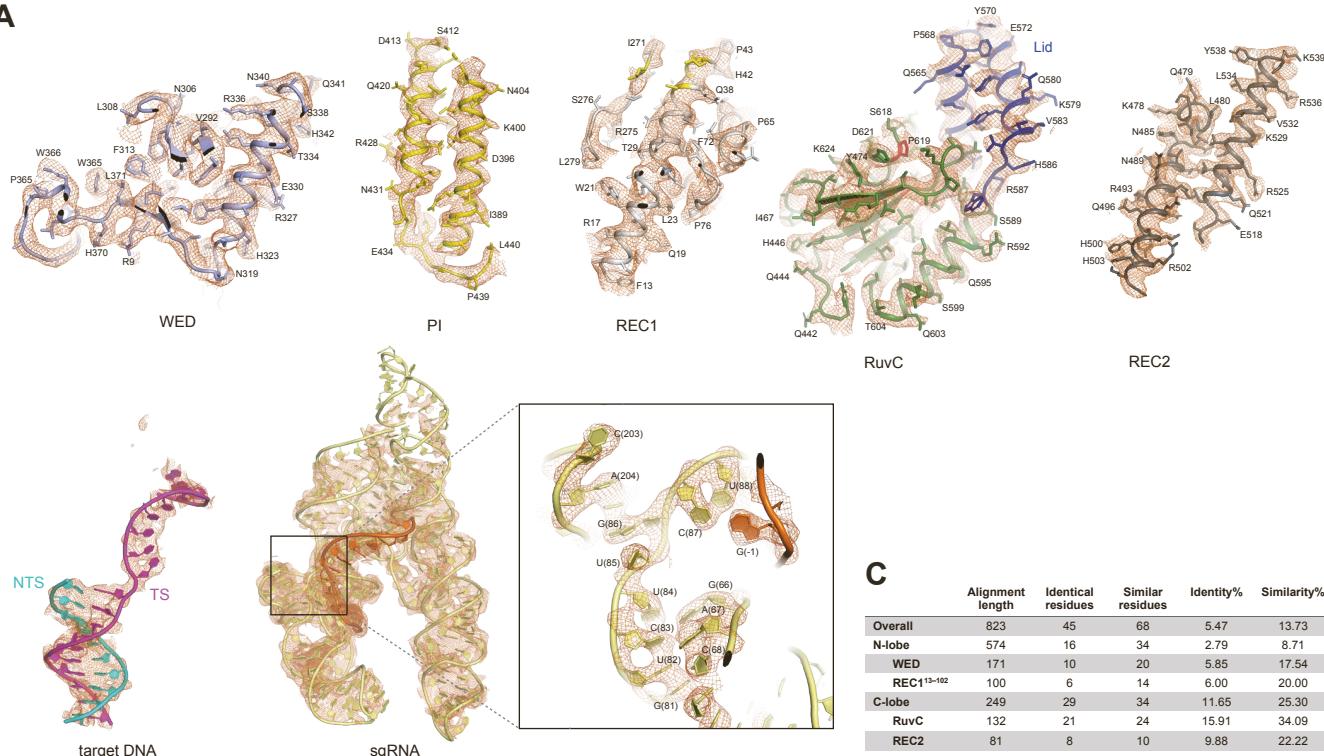


**Figure S2. Cryo-EM data processing for the Cas12k-sgRNA-target DNA ternary complex.**

(A) A representative raw cryo-EM micrograph of the Cas12k-sgRNA-target DNA complex from a total of 2204 micrographs. (B) Representative, good 2D class averages from a total of 100 images. (C) Three 3D reconstructions from heterogeneous refinement. (D–F) Three rounds of supervised heterogeneous

refinement using two maps from 3D variability analysis as templates. Variable regions are indicated by red circles. **(G)** Homogeneous refinement of final particle set. **(H)** Local refinement of final particle set using a mask as indicated. Shown from left to right are the unsharpened map, sharpened map, and local resolution map. **(I)** Plots of the half-map FSC. Related to Figure 2.

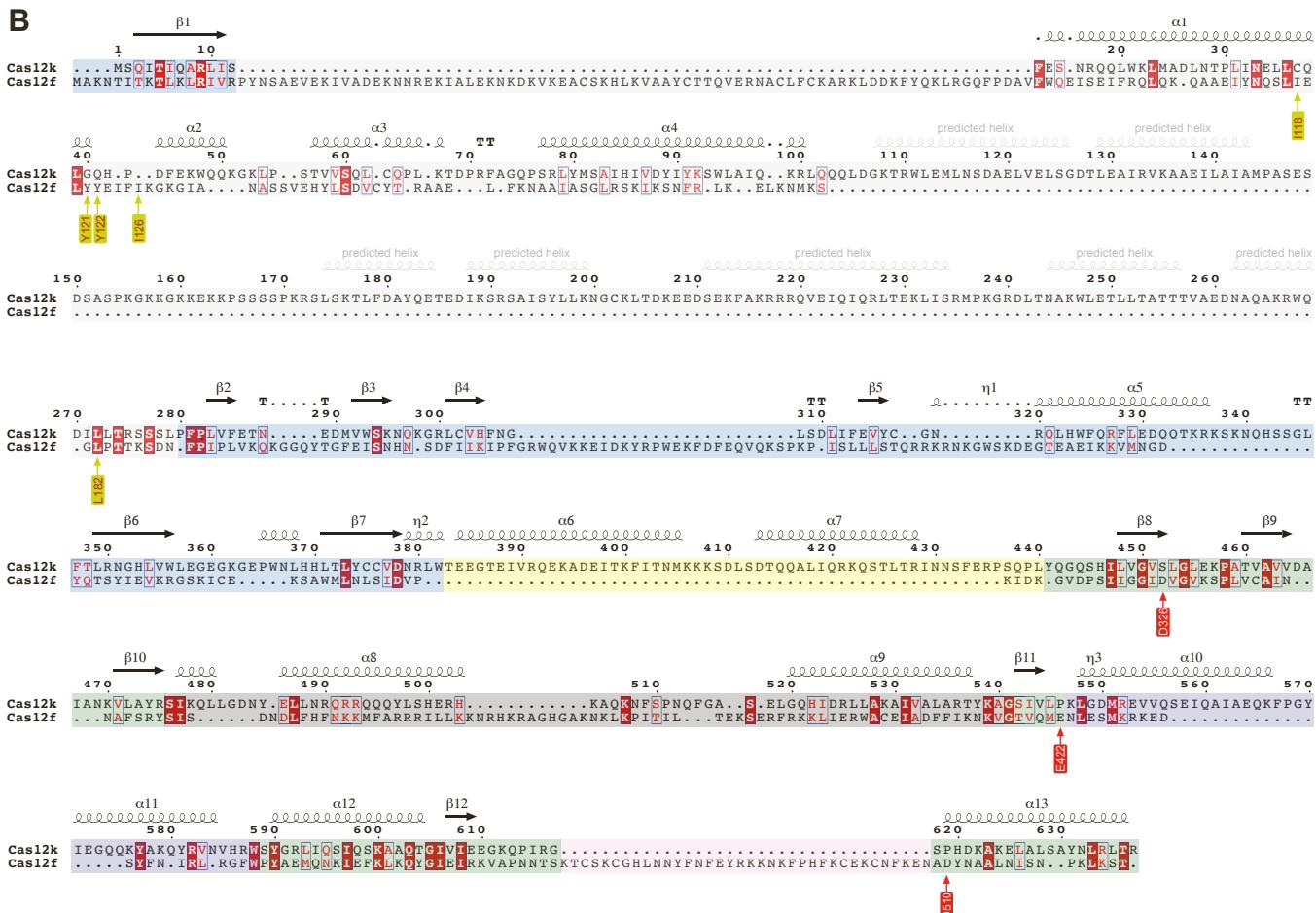
A



C

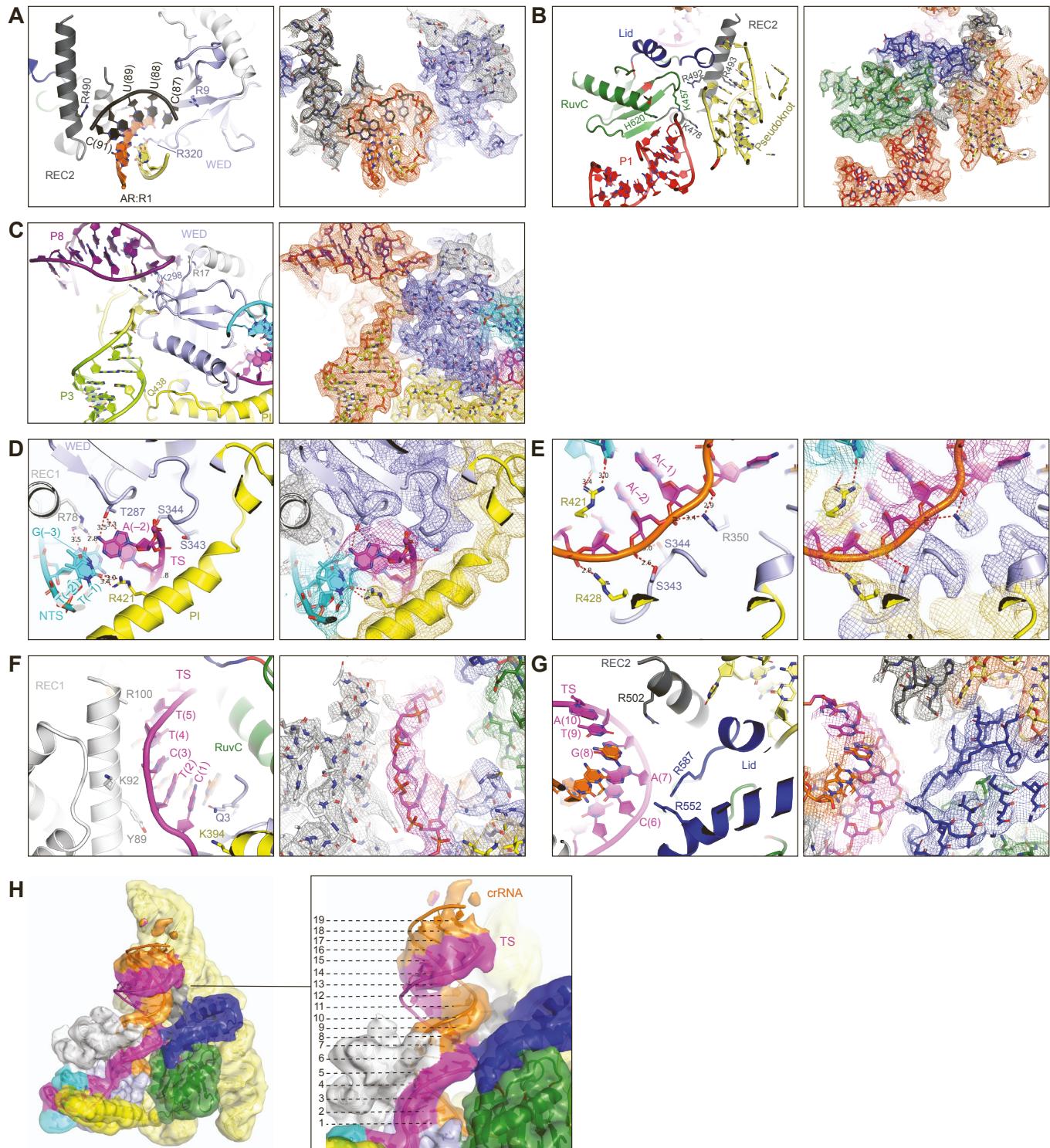
	Alignment length	Identical residues	Similar residues	Identity%	Similarity%
Overall	823	45	68	5.47	13.73
N-lobe	574	16	34	2.79	8.71
WED	171	10	20	5.85	17.54
REC <sup>13-102</sup>	100	6	14	6.00	20.00
C-lobe	249	29	34	11.65	25.30
RuvC	132	21	24	15.91	34.09
REC2	81	8	10	9.88	22.22

B



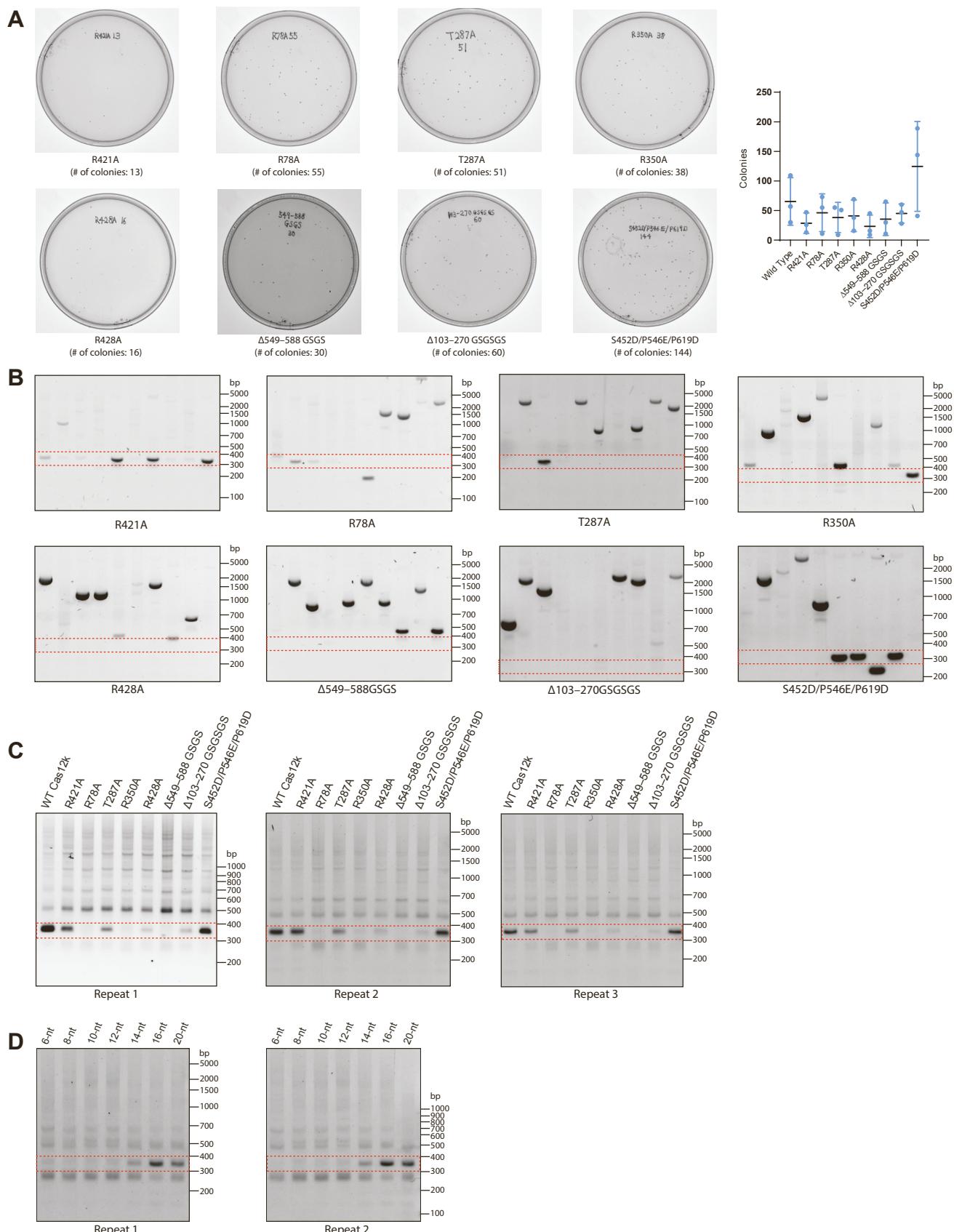
**Figure S3. Detailed cryo-EM density map and structure-based sequence alignment.** (A) Fitting between the cryo-EM map of the Cas12k–sgRNA–target DNA complex and the atomic model. (B) Structure-based sequence alignment of Cas12k and Cas12f. The top part shows the alignment of the N-lobe (residues 1–100) with secondary structure elements  $\alpha_1$  (20–30),  $\beta_1$  (1–10),  $\alpha_2$  (40–50), and  $\alpha_3$  (60–70). The bottom part shows the alignment of the C-lobe (residues 150–260) with  $\alpha$  helices  $\alpha_1$  through  $\alpha_5$  and  $\beta$  strands  $\beta_1$  through  $\beta_{10}$ . Cas12k is shown in blue and Cas12f in red. Yellow arrows indicate specific mutations or structural features.

Structure based sequence alignment of Cas12k and Cas12f. Residue numbers and secondary structures are labeled according to Cas12k. Arrowed residues are key residues for Cas12f dimerization (in yellow) and catalytic residues in the RuvC domain of Cas12f (in red). Each domain is indicated by background colors as in **Fig. 1A**. **(C)** Sequence identity and similarity based on alignment in **B**. Related to Figure 3.



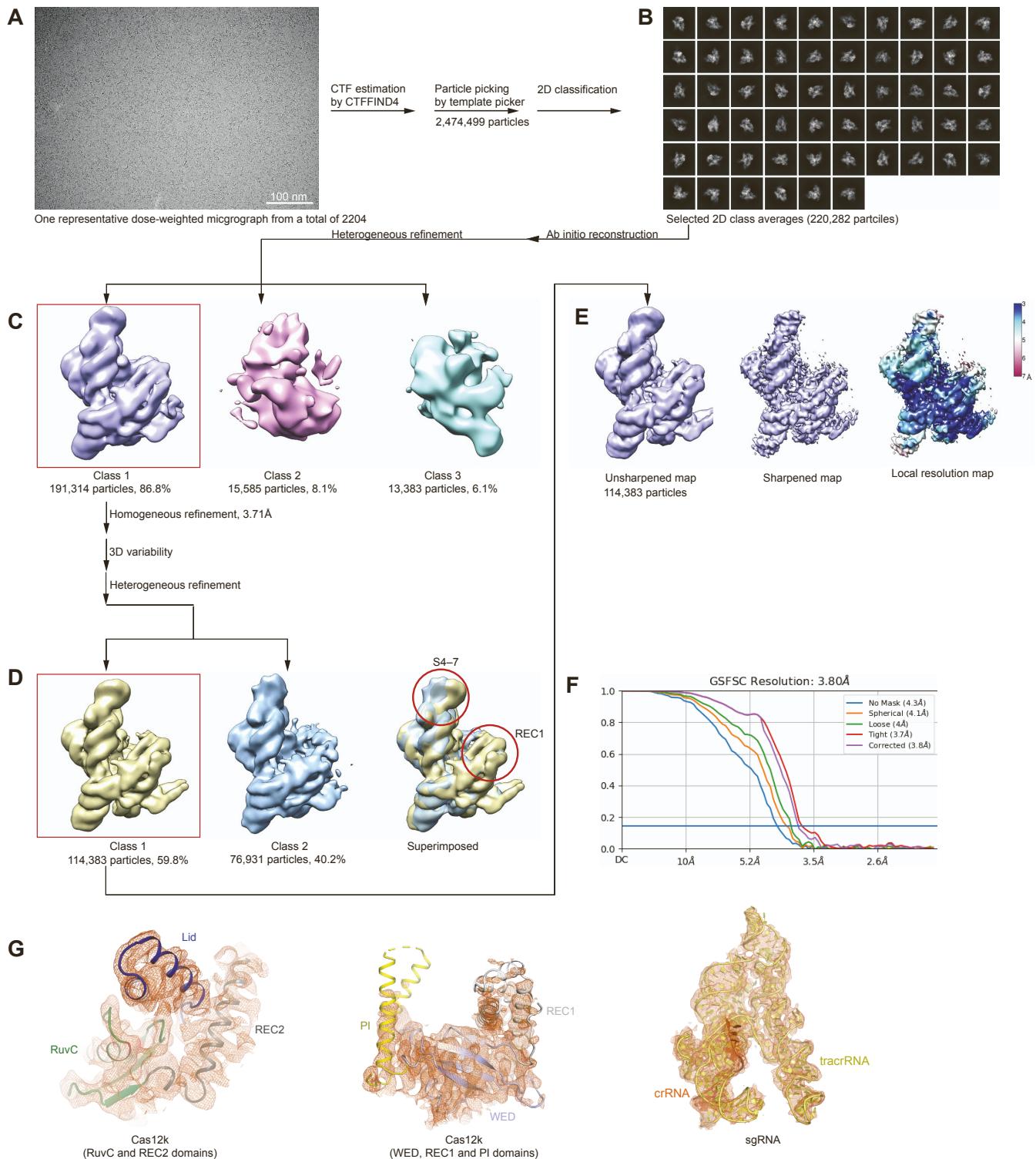
**Figure S4. Interactions between Cas12k and bound nucleic acids.** (A–C) Contacts between Cas12k and sgRNA. The AR:R1 duplex is clamped between the N-lobe and C-lobe of Cas12k (A). S1 and pseudoknot contact the C-lobe (B), whereas S3 and S8 contact the N-lobe of Cas12k (C). (D,E) Interactions between Cas12k and the PAM duplex of target DNA. Left panels are in cartoon representation, while right panels are in stick presentation with cryo-EM density shown in mesh. Key residues involved in interactions are labeled. Interactions are indicated by red dashed lines. (F) Contacts between the N-lobe of Cas12k and the TS of the crRNA-target DNA heteroduplex. (G) Contacts between the C-lobe of Cas12k and the TS of the crRNA-target DNA heteroduplex. (H) The unsharpened cryo-EM map of the Cas12k–sgRNA–target DNA complex color coded as in Fig.2B. Model of the heteroduplex

beyond 10 bp (11–19 bp) is built based on the corresponding low-resolution map of this region. Related to Figures 4 and 5.



**Figure S5. In vitro DNA transposition assay for Cas12k mutants.** (A) LB agar plate showing colonies after transformation of each transposition reaction. Graph on the right shows mean±SD (n=3). (B) PCR results using purified plasmid as template. Ten colonies are randomly selected for plasmid extraction.

from each plate in **A**. Positions of expected PCR readout at ~350 bp are indicated by red dashed boxes. The plates and PCR results were from the same batch as those from **Fig. S1E,F.** **(C,D)** PCR results of *in vitro* DNA transposition assay. Those are the full gels of results shown in **Fig. 5D,E**, respectively. Related to Figure 5.



**Figure S6. Cryo-EM data processing of the Cas12k-sgRNA binary complex.** (A) A representative raw cryo-EM micrograph of the Cas12k-sgRNA complex from a total of 1849 micrographs. (B) Representative, good 2D class averages from a total of 100 images. (C) Three 3D reconstructions from heterogeneous refinement. (D) Supervised heterogeneous refinement using two maps in 3D variability analysis as templates. Variable regions are indicated by red circles. (E) Homogeneous refinement of final particle set. Shown from left to right are the unsharpened map, sharpened map, and local resolution map. (F) Plots of the half-map FSC. (G) Fitting between the cryo-EM map of the Cas12k-sgRNA complex and the atomic model. Related to Figure 6.

**Table S1. Sequence of RNAs and DNA oligonucleotides utilized in this study.** Related to STAR Methods.

RNA	Sequences (5'-3')
sgRNA	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAUUUAU</u> <b>AAGGCCACUGUUAAA</b>
sgRNA_6nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAG</u>
sgRNA_8nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUC</u>
sgRNA_10nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAU</u>
sgRNA_12nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAU</u>
sgRNA_14nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAUUAA</u>
sgRNA_16nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAUUAAU</u>
sgRNA_20nt_spacer	AUAUAAAUGCGCCGCAAUUCAUGCUGCUUGCAGCCUCUGAAUUUUGUUA AAUGAGGGUUAGUUUGACUGUAUUUACAGCUUGCUUCUGACCCUGG UAGCUGCUCACCCUGAUGCUGCUAACUAGACAGGAUAGGUGCGCUCCC AGCAAUAAGGGCGCGGAUGUACUGCUGUAGUGGUACUGAAUCACCCCG AUCAAGGGGAACCCUCCAAA <u>AGGUGGUUGAAAGGAGAAGUCAUUAAU</u> <b>A</b> <b>AAGC</b>
DNA oligonucleotides	Sequences (5'-3')
Oligonucleotides used for structure determination	
Target strand	TTAACAGTGGCCTTATTAAATGACTTCTAACCTCCTACG
Non-target strand	CGTAGGAGGTT
Oligonucleotides used for insertion detection	
Target_F_1	TTCAGAGCAAGAGATTACGCGCAG
Donor_R_1	GTTGAAAGCAAGTCCTTTATCCGCT
Oligonucleotides used to introduce the Cas12k mutations	
R421A_F	ATTCAA <del>cg</del> AAGCAGAGTACGTTGACT
R421A_R	CTGCT <del>Tcg</del> TTGAATCAGTGCTTGTG
R78A_F	CCATCC <del>cg</del> TTGTATATGTCGGCCATC

---

R78A_R	ATACAAcgcGGATGGTTGCCCTGCAAA
T287A-F	TTTGAGcgAAATGAAGACATGGTATGG
T287A-R	TTCATTcgcCTCAAATACTAACGGAAA
R350_F	ACGTTGcgAACGCCATTGGTGTGG
R350_R	GCCGTTcgcAACGTGAATAACCCACT
R428_F	TTGACTgcgATCAATAACTCCTCGAG
R428_R	ATTGATcgcAGTCAACGTACTCTGCTT
S452D_F	GGAGTGgatCTGGGTCTGGAAAAACCA
S452D_R	ACCCAGAtcCACTCCTACTAAGATGTG
P546E_F	GTGCTTgaaAAACTGGGTGACATGCAG
P546E_R	CAGTTTtcAAGCACAATGCTCCCAGC
P619D_F	GGCTCTgatCATGACAAAGCCAAGGAA
P619D_R	GTCATGatcAGAGGCCACGGATCGGTTG
549-588 GSGS_F	TGGGTAGTGGTAGTTCGTATGGCCGCCTTATCC
549-588 GSGS_R	AACTACCACCTACCCAGTTAGGAAGCACAATGCTCCC
103-270 GSGSGS_F	agtggtagtATCTTGCTTACCCGTTCAAG
103-270 GSGSGS_R	accactaccTTGAAGGCCTTTGAATAG

---

Note: For sgRNA, tracrRNA, linker, and crRNA regions are indicated in black, grey, and orange (repeat region underlined), respectively.