nature catalysis

ARTICLES

https://doi.org/10.1038/s41929-020-00506-9

Check for updates

Catalytic-state structure and engineering of *Streptococcus thermophilus* Cas9

Yifei Zhang[1,2], Hongyuan Zhang[1,2], Xuexia Xu[3,4], Yujue Wang[1,2], Weizhong Chen[1], Yannan Wang[1,2], Zhaowei Wu[1], Na Tang[1,2], Yu Wang[5], Suwen Zhao[3,4], Jianhua Gan[6]✉ and Quanjiang Ji[1]✉

Cas9 nucleases recognize and cleave their target DNA through base pairing of a guide RNA with a spacer adjacent to a protospacer adjacent motif (PAM). *Streptococcus thermophilus* Cas9 (St1Cas9), a smaller Cas9 orthologue than *Streptococcus pyogenes* Cas9, enables robust genome editing in diverse organisms. Here we report high-resolution structures of St1Cas9 in complex with a single-guide RNA and different PAM-containing DNAs. All of the structures represent an HNH catalytic state that is rarely observed in other Cas9 structures, clearly depicting the active conformation. A unique wing region in the REC domain forms intensive interactions with the HNH domain, playing a key role in regulating St1Cas9 DNA cleavage activity and probably stabilizing the active conformation. Furthermore, St1Cas9 applies a strategy distinct from those of other Cas9 orthologues for PAM recognition. Structure-guided engineering of St1Cas9 substantially expanded its targeting scope. These molecular-level characterizations will facilitate the rational engineering of St1Cas9.

Advances in the understanding and development of clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated (Cas) systems have revolutionized gene editing methods in biology and medicine[1–9]. Through targeting a protospacer adjacent motif (PAM) and a complementary sequence with a single-guide RNA (sgRNA), the RNA-guided Cas9 nuclease efficiently and programmably cuts the target DNA and creates a double-stranded break, leading to robust in vivo gene editing in diverse organisms[6–12]. Despite its ease of use and high efficiency, several major issues are still present in current CRISPR–Cas9 technology, such as a high off-target rate, a strict PAM requirement, high cell toxicity and large gene size[13–16].

One potential way to overcome these current CRISPR–Cas9 technology limitations is to discover, characterize and utilize Cas9 orthologues from other bacteria that may possess rich and distinct biochemical properties for gene editing. Recent progress in the mining and engineering of CRISPR–Cas9 systems from different bacterial species has greatly expanded genetic manipulation platforms for in vivo editing applications[7–9,17–19]. Several Cas9 orthologues from distinct bacterial species have been harnessed for genome editing in living cells, such as *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9) and *Streptococcus thermophilus* Cas9 (St1Cas9) from a Class 2 type II-A CRISPR system, *Francisella novicida* Cas9 (FnCas9) from a Class 2 type II-B CRISPR system and *Neisseria meningitidis* Cas9 (Nm1Cas9) from a Class 2 type II-C CRISPR system[7–9,15,19–26]. SpCas9, SaCas9 and St1Cas9 are three of the earliest identified and harnessed Cas9 nucleases for gene editing[7–9,15,20–24]. SpCas9 has a high DNA-cleavage activity, but its large size, imperfect specificity and strict requirement for an NGG PAM restricts its applications in therapeutics. SaCas9 (1,053 amino acids) and St1Cas9 (1,121 amino acids) are smaller Cas9 orthologues than SpCas9 (1,368 amino acids), rendering them suitable for being packed into a single AAV vector for in vivo editing[15,27]. Furthermore, SaCas9 and St1Cas9 have relatively longer PAM requirements compared with SpCas9 (NNGRRT and NNRGAA for SaCas9 and St1Cas9, respectively; underlined text represents the differences of the PAM specificities among Cas9 orthologues), lower cell toxicities and higher DNA-targeting specificities[15–17], suggesting distinct DNA targeting and cleavage mechanisms present in these enzymes.

In addition to the exploration of Cas9 orthologues from different bacterial species, recent progress in directed evolution or structure-guided protein engineering has developed a few Cas9 variants to recognize expanded PAM sequences, such as SpCas9 VQR, SpCas9 EQR, SpCas9 VRER, xCas9, SpCas9-NG, KKH SaCas9, SpCas9-NRRH, SpCas9-NRTH, SpCas9-NRCH and SpRY[20,28–32]. Moreover, different high-fidelity Cas9 variants with improved targeting specificity in living cells have been engineered, such as SpCas9-HF1, eSpCas9, HypaCas9, evoCas9, SpCas9_R63A/Q768A and SaCas9-HF[33–38].

Structural studies of Cas9 nucleases from different bacterial species provide insights into RNA-guided DNA targeting and cleavage mechanisms. The structures of SpCas9 in the unbound, sgRNA-bound, and sgRNA- and DNA-bound forms unveil the detailed molecular mechanism of DNA recognition and cleavage by Cas9, with notable conformational rearrangement on binding to sgRNA and DNA[39–42]. The crystal structure of SaCas9 in complex with sgRNA and DNA reveals a conserved sgRNA-induced conformational activation mechanism, but distinct chemistry for PAM DNA and sgRNA recognition compared with that of SpCas9[43]. Recent progress in cryogenic electron microscopy (cryo-EM) structural studies and computational simulations of SpCas9[44,45] and the crystal structures of Nm1Cas9[46] provide certain insights into the HNH catalytic mechanism that was not observed in previous Cas9 crystal structures. Despite the recent determination of cryo-EM structures of St1Cas9[47], the low resolution and lack of an HNH domain hamper a detailed understanding of the DNA targeting and catalytic mechanism of St1Cas9.

[1]School of Physical Science and Technology, ShanghaiTech University, Shanghai, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]iHuman Institute, ShanghaiTech University, Shanghai, China. [4]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [5]College of Life Science and Engineering, Jiangxi Agricultural University, Nanchang, China. [6]Shanghai Public Health Clinical Center, State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Physiology and Biophysics, School of Life Sciences, Fudan University, Shanghai, China. ✉e-mail: ganjhh@fudan.edu.cn; quanjiangji@shanghaitech.edu.cn

In this study we report the high-resolution crystal structures of St1Cas9 in complex with sgRNA and double-stranded DNA that contains one of three different PAMs (AAAGGA, AAAGAA and AAGGAA). Interestingly, all of the structures represent an HNH catalytic state that has rarely been observed in the previous Cas9 crystal structures, clearly depicting the active enzyme conformation. A wing region composed of two β strands in the REC domain forms intensive interactions with the HNH domain. Furthermore, an aromatic residue, F693, at a linker region that is extended from the C-terminus of the HNH domain, forms π–π stacking interactions with the first base of the target DNA. These interactions are unique in St1Cas9 and not present in SaCas9, SpCas9 or Nm1Cas9, playing important roles in regulating St1Cas9 activity and probably contributing to stabilizing the active conformation of the HNH domain. Moreover, St1Cas9 uses a chemistry for PAM recognition that is distinct from that of other Cas9 orthologues. As St1Cas9 has a strict requirement for a long PAM for targeting, we broadened the PAM compatibility of St1Cas9 to NNRNAA via structure-guided protein engineering. These structural and biochemical characterizations of the active St1Cas9 conformation provide insights into the DNA targeting and catalytic mechanisms of Cas9 nucleases and facilitate rational engineering of St1Cas9.

Results

Overall structure of the St1Cas9–sgRNA–dsDNA ternary complex. The unique biochemical properties of St1Cas9 compared with those of its Cas9 orthologues suggest different DNA targeting and catalysis mechanisms for St1Cas9. We sought to determine the crystal structure of the St1Cas9–sgRNA–dsDNA complex to understand the detailed molecular mechanism. Inspired by the strategy used for the crystallization of SaCas9-sgRNA-dsDNA[43], we deleted stem loop 2 from the sgRNA and mutated the catalytic residue H599 to A599 to facilitate crystallization (Fig. 1a). In vitro DNA cleavage assays confirmed that the truncation of stem loop 2 did not reduce St1Cas9 DNA cleavage activity (Fig. 1b), whereas a bacterial survival assay revealed that the truncation retained the high genome-targeting activity of St1Cas9 in living *Klebsiella pneumoniae* cells (*K. pneumoniae*) (Extended Data Fig. 1a,b). With the aforementioned method, we successfully solved the crystal structures of St1Cas9 (H599A) in complex with the truncated 70-nucleotide (nt) sgRNA, a 28-nt target DNA strand (TS) and an 8-nt non-target DNA strand that contains one of three different PAMs (AAAGAA, AAGGAA and AAAGAA) at 2.76, 3.0 and 2.56 Å resolution, respectively (Supplementary Table 1).

Similar to other Cas9 structures[25,39–41,43,44,46,48,49], the St1Cas9 structure adopts a bi-lobed architecture composed of a recognition (REC) lobe and a nuclease (NUC) lobe (Fig. 1c,d and Supplementary Fig. 1a–c). The NUC lobe comprises the RuvC, HNH, Wedge (WED) and PAM-interaction (PI) domains, as well as linker regions Linker1 (L1) and Linker2 (L2). The RuvC domain is further divided into three RuvC motifs (RuvC-I, RuvC-II and RuvC-III). The REC lobe contains three domains: REC1, REC2 and a wing region. Notably, the wing region composed of two β strands is unique to St1Cas9 and is absent in all of the previously reported Cas9 structures (Extended Data Fig. 2a–g). In the large Cas9 nucleases such as SpCas9 and *Francisella novicida* Cas9 (FnCas9), an extra REC domain comprising several α helices is present between REC1 and REC3 (Extended Data Fig. 2b,c). In the small Cas9 orthologues such as SaCas9, Nm1Cas9, *Corynebacterium diphtheria* Cas9 (CdCas9) and *Campylobacter jejuni* Cas9 (CjCas9), neither the wing region nor the additional REC domain is present (Extended Data Fig. 2d–g). The REC lobe and NUC lobe are connected by an arginine-rich bridge helix.

The sgRNA is composed of a guide segment (spacer), a repeat:antirepeat region, a tetraloop, stem loop 1 and a single-stranded linker (Fig. 1a,d). The guide segment occupies the groove between the REC and NUC lobes. The repeat:antirepeat region, stem loop 1 and the single-stranded linker are recognized and stabilized by the arginine-rich bridge helix, REC1, RuvC, WED and PI domains (Extended Data Fig. 3). In the presence of the target DNA, the St1Cas9–sgRNA complex recognizes the PAM via the WED and PI domains and then the guide sequence of the sgRNA forms Watson–Crick pairs with the target DNA strand (TS) and the HNH domain chelates Mg^{2+} and rotates toward the active state to initiate DNA cleavage (Supplementary Fig. 2a and Fig. 2a).

A structural comparison revealed that the sgRNAs of SpCas9, SaCas9 and St1Cas9 share a similar A-form conformation (Supplementary Fig. 2b). Although the St1Cas9 sgRNA adopts the same RNA structure with the SaCas9 sgRNA, there are notable sgRNA recognition differences between the two Cas9 orthologues (Extended Data Fig. 3)[43]. Moreover, recent cryo-EM structures of St1Cas9 revealed that full-length St1Cas9 sgRNA is composed of two stem loops (stem loop1 and stem loop 2)[47], whereas full-length SpCas9 sgRNA contains an additional stem loop 3 (Supplementary Fig. 2b).

The catalytic conformation of the HNH domain. Despite recent advances in determining the structures of Cas9 nucleases from different bacterial species, the catalytic conformation of the HNH domain is still not fully known. In the SpCas9 structures, the HNH domain from the cryo-EM structure[44] is pointed towards the catalytic site but lacks density for the water molecules and the central Mg^{2+} ion (Extended Data Fig. 4a). In the SaCas9 and FnCas9 crystal structures[25,43], the HNH domain adopts an inactive conformation and the active site is positioned more than 40 Å away from the TS cleavage site (Extended Data Fig. 4b,c). The pre- and post-catalytic conformations of the Nm1Cas9-sgRNA-dsDNA ternary complex were recently revealed[46] (Extended Data Fig. 4d,e). However, the water molecules coordinating the Mg^{2+} in the HNH domain catalytic centre were not visible in the structures. In the determined structures of the St1Cas9–sgRNA–dsDNA complex, all of the structures represent an HNH catalytic conformation, clearly depicting the active conformation (Fig. 2a,b).

Close inspection of the HNH catalytic conformation revealed that the HNH nuclease domain adopts a ββα-metal fold and chelates Mg^{2+} for TS cleavage between the third and fourth nucleotides upstream of the PAM (Fig. 2b). Similar to H588A in the DNA-bound Nm1Cas9 catalytic structure, the side chain of the catalytic residue H599A points toward the scissile phosphate and the Cα atom of H599A is approximately 6 Å away from the scissile phosphate (Extended Data Fig. 4d,f). The Mg^{2+} ion adopts a pseudo-octahedral coordination geometry with D598, N622, two water molecules and two oxygen atoms of the scissile phosphate of TS nucleotides 3 and 4 as the coordination molecules (Fig. 2b). Y578, E596 and K625—located in the active centre—form strong interactions with the metal-chelating residues through hydrogen bonding (Fig. 2b). Furthermore, K625 forms hydrogen bonds with I600 and R628 in the HNH domain (Extended Data Fig. 5a), similar to that in Nm1Cas9[46] and *Actinomyces naeslundii* Cas9 (AnCas9)[40] (Extended Data Fig. 5b,c), which probably contributes to structural stabilization of the ββα-metal fold of the HNH domain.

To examine the role of these active site residues in St1Cas9 catalytic capacity, we separately mutated these residues to a non-functional residue (alanine) and tested the DNA cleavage capacity of the mutant proteins. As shown in Fig. 2c, the mutation of the active residue H599 or the residues D598 and N622 that directly chelate Mg^{2+} almost abolished St1Cas9 catalytic activity. The mutation of K625—which forms a hydrogen bond with D598—drastically reduced the DNA cleavage activity, whereas the single mutations or even double mutation of Y578 and E596—which interact with D598 and N622, respectively—did not affect St1Cas9 DNA cleavage activity (Fig. 2c and Extended Data Fig. 5d). Bioinformatics analysis of the Cas9 protein family revealed that H599, N622 and K625 are

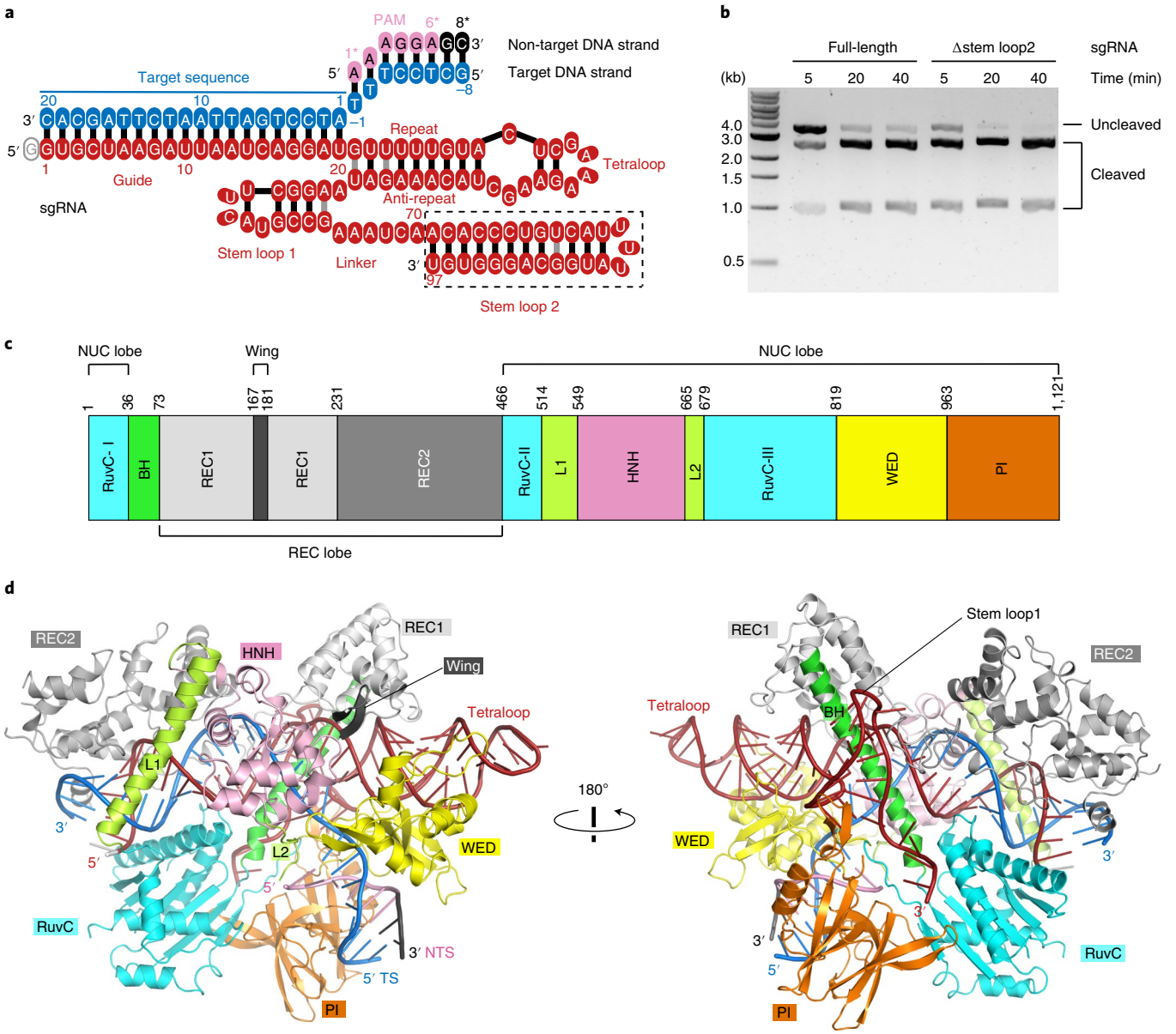


Fig. 1 | Overall structure of St1Cas9–sgRNA–dsDNA ternary complex. a, A schematic representation of sgRNA and the target DNA. Stem loop 2 (dashed box) was truncated for crystallization. **b**, An in vitro DNA cleavage assay of wild-type St1Cas9 using the full-length sgRNA (97-nt) and the truncated sgRNA (70-nt). The truncated nucleotides of the 70-nt sgRNA are shown in the dashed-line box of Fig. 1a. The linearized plasmid containing the 5′-AAAGAA-3′ PAM was used as the cleavage substrate. The experiment was performed in triplicate. **c**, The domain organization of St1Cas9. **d**, The crystal structure of the St1Cas9–sgRNA–dsDNA (AAAGGA PAM) ternary complex. Individual St1Cas9 domains were coloured according to the scheme in Fig. 1c. NTS, non-target DNA strand.

highly conserved, whereas Y578 can be substituted with F, E596 can be substituted with D or Q, and D598 can be substituted only with E in other Cas9 orthologues, further supporting the different roles of these residues in St1Cas9 activity (Fig. 2d,e).

The interactions between the HNH domain and the REC lobe. In all of the structures of the St1Cas9–sgRNA–dsDNA complex, the HNH domain is in an active conformation that is not observed in other Cas9 crystal structures except that of Nm1Cas9[46], probably due to the flexibility of the HNH domain in Cas9. We sought to probe the possible mechanism that results in the active conformation of the St1Cas9 HNH domain. Previous computational simulation and experimental validation pinpoint the key role of the REC lobe in sensing, regulating and locking the catalytic HNH domain[50]. We therefore carefully examined the interactions between the HNH domain and the REC lobe in the structures of the St1Cas9–sgRNA–dsDNA complex, revealing two major interaction mechanisms. First, S417 in the REC2 domain forms a hydrogen bond with H553 of the HNH domain (Fig. 3a, upper right). Second, the wing region forms intensive interactions with the HNH domain, including a hydrogen bond between T170 and D638, a salt bridge between R179 and E624, and several van der Waals interactions (Fig. 3a, middle right). In addition to the interactions between the REC lobe and the HNH domain, we observed π–π stacking interactions between F673 located at L2 extended from the C-terminus of the HNH domain and the first base dT-1 of the TS, a unique mechanism not observed in other Cas9 structures (lower right in Fig. 3a and Supplementary Fig. 3a,b).

To further investigate the role of these interactions in regulating St1Cas9 activity, we separately mutated or truncated S417, F673

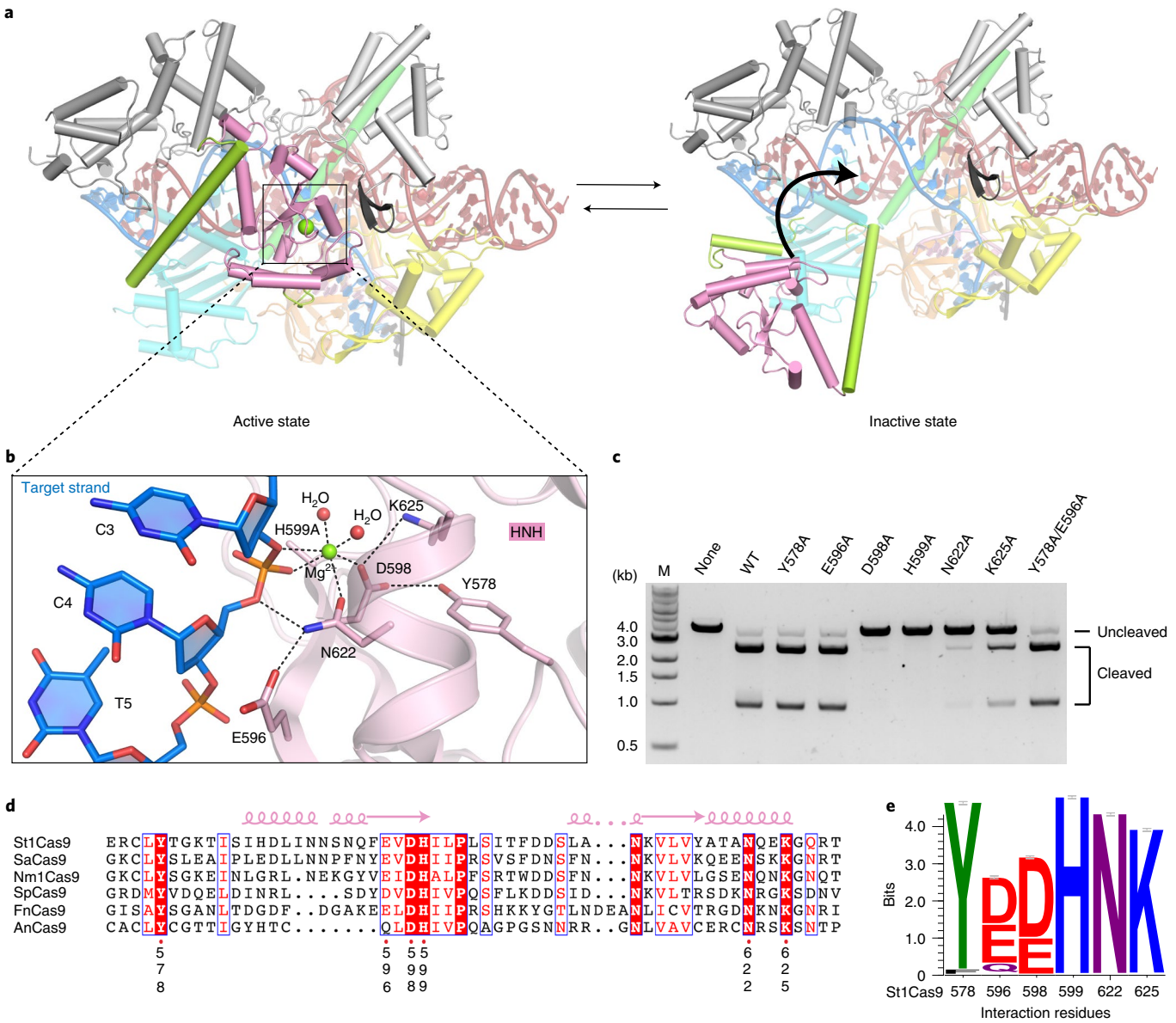


Fig. 2 | The catalytic conformation of the HNH domain. a, The St1Cas9–sgRNA–dsDNA structure represents a HNH catalytic state. The HNH inactive-state structure was obtained using structural modelling with the structure of SaCas9 (PDB ID: 5AXW) as the template. **b**, A magnified view of the active site of the HNH domain. **c**, An in vitro DNA cleavage assay of the wild-type and various active-site mutants of St1Cas9. The linearized plasmid containing a target sequence was used as the substrate and the full-length sgRNA was used as the guide RNA in the assay. The experiment was performed in triplicate. WT, wild-type St1Cas9. **d**, The sequence alignment of HNH domains. The secondary structure of St1Cas9 was shown above the sequences. The active-site residues were highlighted. **e**, Conservation of the active-site residues of the HNH domains from the Cas9 protein family (IPR028629).

and the wing region and then tested the capacities of the resulting mutant proteins in DNA cleavage. As shown in Fig. 3b, the S417A mutant protein retained a DNA cleavage activity similar to that of wild-type St1Cas9; however, the truncation of the wing region or the mutation of F673 to A673 substantially reduced St1Cas9 DNA cleavage activity, suggesting the important role of these interactions in regulating St1Cas9 activity and stabilizing the HNH active conformation (Fig. 3b,c). A subsequent electrophoretic mobility shift assay confirmed that the truncation of the wing region did not notably affect St1Cas9 DNA binding ability despite the relatively long sequence deletion (Extended Data Fig. 6). Although the wing region has been observed only in the St1Cas9 structures, bioinformatics analysis revealed that this wing region is widely distributed in numerous Cas9 nucleases from different bacterial species, including both the type-II-A and type-II-C class Cas9 nucleases (Extended Data Fig. 7).

As the stabilization of the HNH active conformation by the wing region may make the HNH domain of St1Cas9 spend more time in the docked conformation than other orthologues, we sought to probe the role of the wing region in the fidelity of St1Cas9 by using the in vitro DNA cleavage assay. First, we determined the targeting specificity of wild-type St1Cas9 by using DNAs containing different types of mismatches as the substrates (Extended Data Fig. 8a). As shown in Extended Data Fig. 8b, no notable cleavage activity was observed when two mismatch sites were present in the target DNA. When only one mismatch site was present in the target DNA, wild-type St1Cas9 had substantial cleavage activity on substrates with mismatches on positions 20 or 19, but much lower activity on

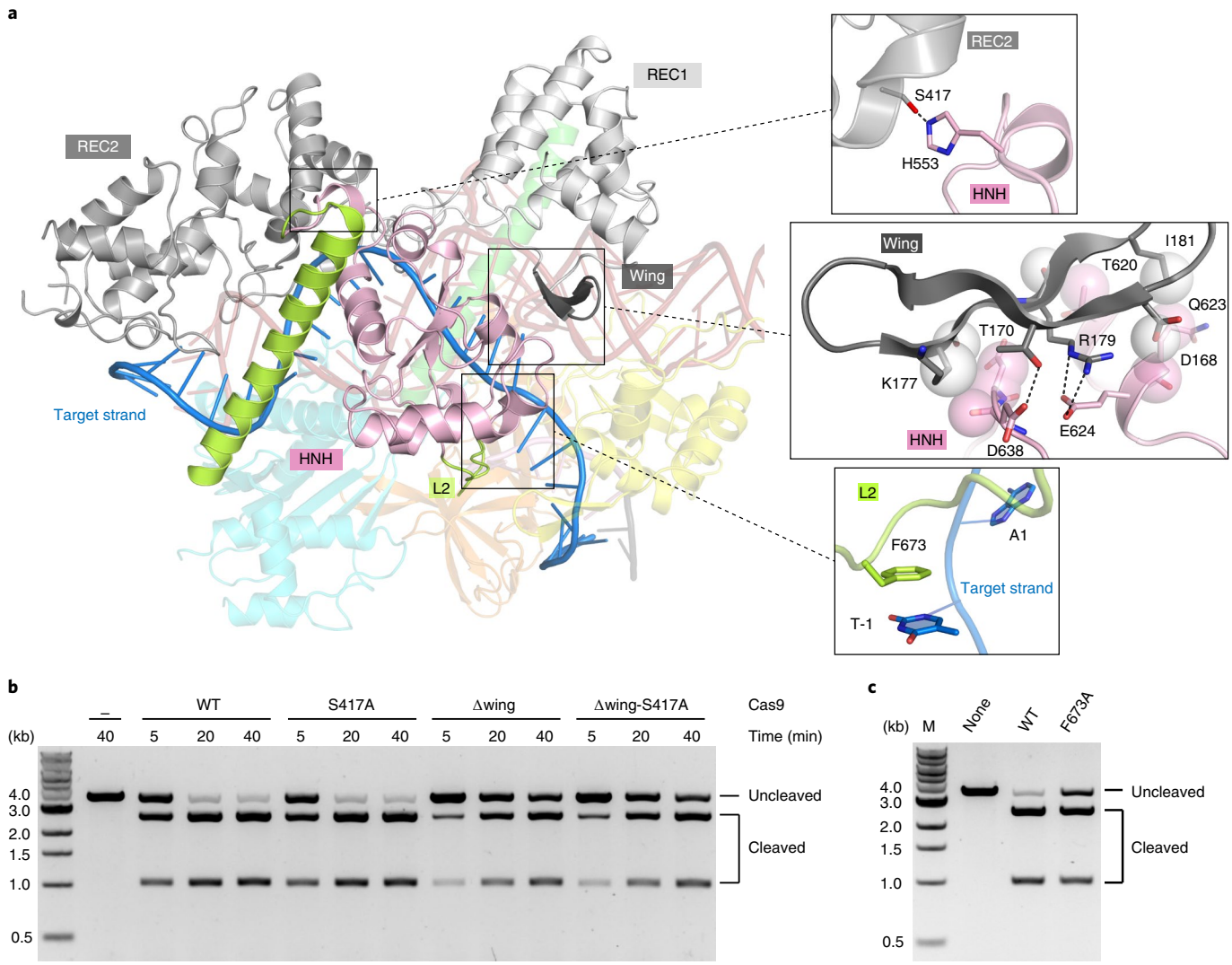


Fig. 3 | The interactions between the HNH domain and the REC lobe. a, A detailed view of the interactions between the HNH domain with REC2 domain (upper right) and the wing region (middle right). The π-π stacking interaction between the residue F673 of L2 and dT-1 of target strand is indicated (lower right). **b**, In vitro DNA cleavage assay of the wild-type and various mutants of St1Cas9. The linearized plasmid containing a target sequence was used as the substrate and the full-length sgRNA was used as the guide RNA in the assay. The experiment was performed in triplicate. **–**, reaction without the addition of St1Cas9 was used in the lane of none as the control; Δwing, the truncation of the wing region. **c**, In vitro DNA cleavage assay of the wild type and F673A mutant of St1Cas9. The linearized plasmid containing a target sequence was used as the substrate and the full-length sgRNA was used as the guide RNA in the assay. The experiment was performed in triplicate.

substrates with mismatches at positions 18–16 (Extended Data Fig. 8c). As wild-type SpCas9 has a strong DNA cleavage activity on substrates with two mismatches at positions 19–18, 18–17 or 16–15[51], this assay confirmed that St1Cas9 has higher DNA targeting specificity than SpCas9. Next, we used the target DNA with mismatches at positions 20 or 19 to probe the fidelity of the wing-region-deleted St1Cas9 (St1Cas9$_{\Delta wing}$). As St1Cas9$_{\Delta wing}$ has a lower DNA cleavage activity than wild-type St1Cas9 (St1Cas9$_{WT}$) (Fig. 3b), we adjusted the concentrations of St1Cas9$_{\Delta wing}$ and St1Cas9$_{WT}$ to achieve a similar activity on the fully matched DNA substrate for comparison. As shown in Extended Data Fig. 8d, St1Cas9$_{\Delta wing}$ and St1Cas9$_{WT}$ had a similar cleavage activity on the fully matched DNA substrate at the time points tested. Notably, St1Cas9$_{\Delta wing}$ showed no cleavage activity on substrates with single mismatches on positions 20 or 19 at all of the time points tested, whereas St1Cas9$_{WT}$ showed obvious cleavage activity on the same mismatched DNAs at all of the time points tested, suggesting that the presence of the wing region impairs the fidelity of St1Cas9.

PAM recognition mechanism. The PAM duplex is located in the groove between the WED and PI domains and the PAM is recognized by the PI domain (Fig. 4a,b). St1Cas9 applies distinct chemistries for PAM recognition compared with those of other Cas9 orthologues[52]. The first base of the PAM, dT-1, forms non-base-specific π–π stacking interactions with the residue F673 of L2 (Fig. 4c). The second base pair, dA2*:dT-2, does not directly interact with the protein in the determined structures. N6 and N7 of dA3* form bidentate hydrogen bonds with the side chain of Q1084, which is further anchored by E1057 via a hydrogen bonding interaction (Fig. 4c). When dA3* was substituted with dG3*, the side chain of Q1084 rotated away from E1057 and formed only a single hydrogen bond with N7 of dG3* (Fig. 4c,d). O6 and N7 of the fourth base, dG4*, form two hydrogen bonds with K1086 (Fig. 4c). In contrast to the hydrogen bonding interactions between dA3*-dG4* and the protein, the dA5*:dT-5 and dA6*:dT-6 base pairs interact with M1049 via van der Waals interactions. In the structures that contain the AAAGAA or AAGGAA PAM, N6 of dA5* and O4 of dT-6

form van der Waals interactions with the sulfur atom of M1049, and O4 and C5 of dT-6 form van der Waals interactions with the methyl group of the side chain of M1049 (Fig. 4c), whereas in the structure containing the AAAGGA PAM, the side chain of M1049 rotates by approximately 150° compared with that in the structure that contains the AAAGAA PAM, resulting in a weakened interaction between dG5* and M1049 (Fig. 4c,d). In agreement with the interaction mechanisms, St1Cas9 had a high cleavage activity towards the DNA that contains the AAAGAA PAM, whereas the activities were reduced substantially when the third or fifth A base was substituted with G (Fig. 4e). Based on the in vitro DNA cleavage activities and previous in vivo editing studies[16], we concluded that the PAM requirement of St1Cas9 is NNRGAA (where R is A or G). Alanine substitution mutation of any of the PAM recognition residues (E1057, Q1084, K1086 and M1049) substantially reduced the St1Cas9 DNA cleavage activity, further confirming the key role of these residues in PAM targeting (Fig. 4f).

The PAM recognition mechanism of St1Cas9 is distinct from that of SpCas9 and SaCas9 (Fig. 4g,h). SpCas9 recognizes the NGG PAM using R1333/E1219/R1335, and SaCas9 recognizes the NNGRRT PAM using E993/R1015/N985/N986/R991, whereas St1Cas9 uses E1057/Q1084/K1086/M1049 for NNRGAA PAM recognition. Both in SpCas9 and SaCas9, the side chains of R (R1335 of SpCas9 and R1015 of SaCas9) form bidentate hydrogen bonds with G bases in their PAMs and are further anchored by E (E1219 of SpCas9 and E993 of SaCas9) via salt bridges (Fig. 4g,h). By contrast, in St1Cas9, the corresponding residue R is substituted with Q, which showed higher preference for the third A base than G (Fig. 4e). R1333 of SpCas9 and K1086 of St1Cas9 form bidentate hydrogen bonds with G bases, whereas N985 and N986 of SaCas9 form a single hydrogen bond with the fourth and fifth R bases (where R is A or G) of the NNGRRT PAM, respectively (Fig. 4c,g,h). Moreover, St1Cas9 uses M1049 to recognize the rest PAM bases via van der Waals interactions, a mechanism not observed in SpCas9 or SaCas9, which recognize the PAM sequences only by hydrogen-bond interactions (Fig. 4c,g,h).

Expansion of St1Cas9 PAM recognition via structure-guided engineering. St1Cas9 has a strict requirement for a long PAM for DNA targeting, substantially limiting its editable sites. Inspired by the successes of structure-guided engineering for PAM expansion of Cas9 nucleases[20,25,30], we sought to relieve the PAM restriction of St1Cas9 through structure-guided protein engineering. First, K1086A, K1086L and K1086H were separately introduced into St1Cas9 to relax the specificity for the fourth base G by disrupting the interaction between it and St1Cas9. Notably, the single amino-acid mutation of K1086 to L1086 drastically enhances St1Cas9 DNA cleavage activity towards the DNA that contains the NNAHAA PAM (where H is A, C or T), but subtantially impaired its activity toward the DNA containing the NNAGAA PAM, which can be targeted by wild-type St1Cas9 (Extended Data Fig. 9a). Although the K1086L variant has a high activity with the NNATAA PAM, the substitution of the third base A with any of the other three bases (C, T or G) abolished the DNA cleavage activity (Extended Data Fig. 9b). Next we introduced the E1057L mutation into the K1086L variant to relieve the requirement for the third base A of the PAM for DNA recognition because E1057 plays an important role in fixing the third base recognition residue Q1084 (Fig. 4c). As shown in Extended Data Fig. 9c, the introduction of the E1057L mutation into the K1086L variant slightly increased the DNA cleavage activity with the NNGTAA PAM, but reduced the activity with the NNATAA PAM. Structural analysis revealed that D939 and N1081—located in the WED and PI domains, respectively—are adjacent to the phosphate backbone of the target DNA (Fig. 5a,b). We constructed a D939K/E1057L/N1081K/K1086L variant (referred as KLKL) to investigate whether substitutions of D939 and N1081 with a positive residue K could form a non-base-specific interaction with the target DNA and enhance the cleavage activity of St1Cas9. As shown in Fig. 5c and Extended Data Fig. 9c, compared with wild-type St1Cas9, the evolved KLKL variant had a substantially increased DNA cleavage activity toward DNA containing the NNRNAA PAM, thereby relaxing the PAM requirement of St1Cas9 from NNRGAA to NNRNAA.

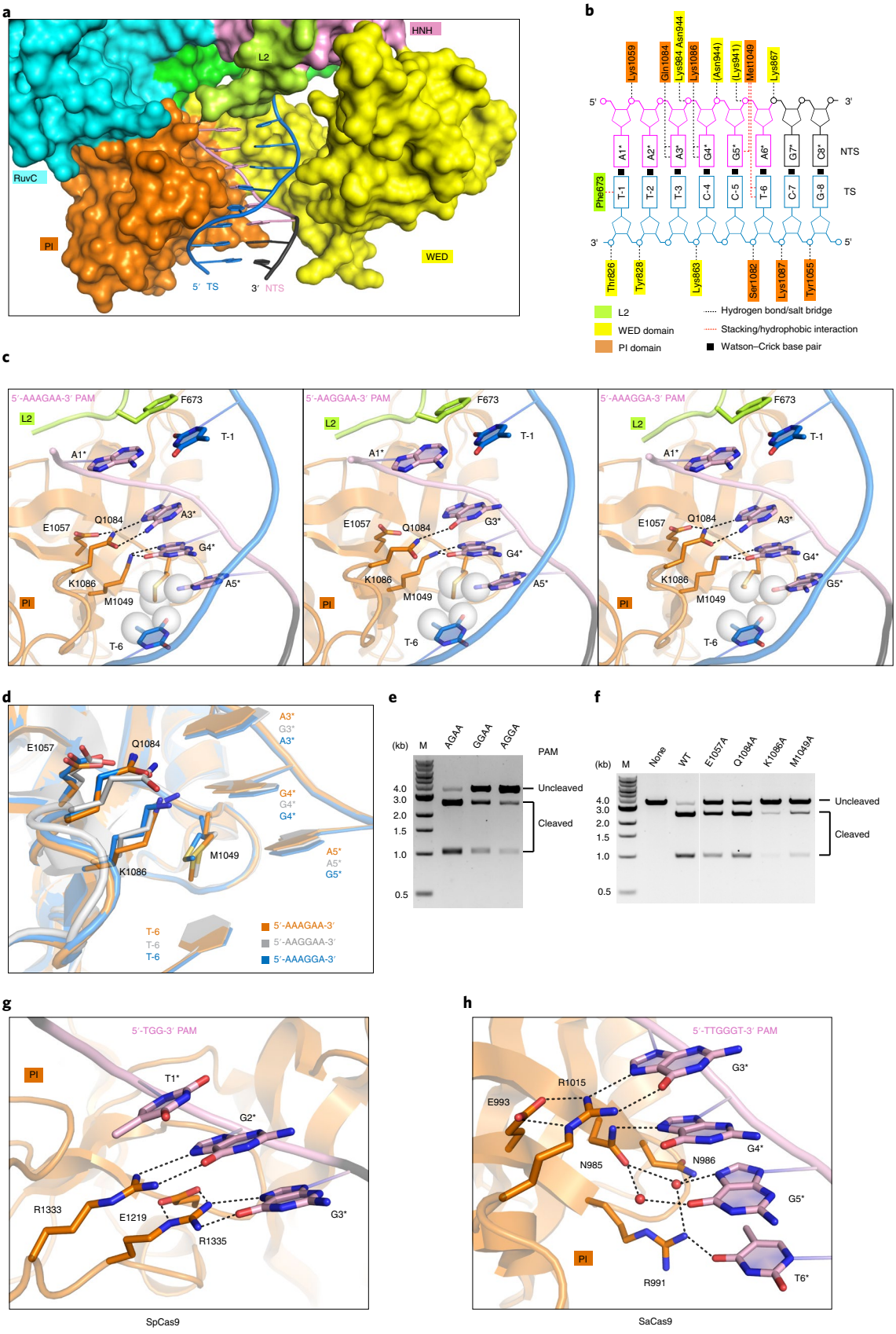
Activity evaluation of St1Cas9 in living cells. We sought to develop a St1Cas9-mediated base-editing system in *K. pneumoniae* to probe the capacity of the engineered KLKL variant for gene editing in living cells. Directly fusing a catalytically impaired Cas nuclease with a nucleotide deaminase is an effective strategy to construct Cas nuclease-mediated base-editing systems[53–58]. We thereby constructed a base-editing system in *K. pneumoniae* by fusing the St1Cas9 D9A nickase with the rat APOBEC1 cytidine deaminase through a 32-residue linker (Extended Data Fig. 10a). First we determined the editing window of this system by transforming different plasmids containing distinct spacers into *K. pneumoniae* cells, revealing an editing window from positions 4 to 11 (Extended Data Fig. 10b), which is narrower than that in mammalian cells[27]. Next we compared the capacities of wild-type St1Cas9 and the KLKL variant to facilitate base editing using the established base-editing system. Notably, the KLKL variant exhibited much higher base-editing activity than wild-type St1Cas9 toward the NNAAAA, NNATAA, NNACAA, NNGAAA, NNGTAA and NNGCAA PAMs, but reduced activity towards the NNAGAA and NNGGAA PAMs compared with wild-type St1Cas9 (Fig. 5d), which is consistent with the in vitro DNA cleavage activities (Fig. 4c). Together, these results demonstrate that the KLKL variant has an expanded in vivo DNA-targeting scope, with NNRNAA PAM specificity.

Discussion

Exploration of alternative Cas9 nucleases from different bacterial species is a potential way to overcome the limitations of current CRISPR–Cas9 technology. St1Cas9—a relatively smaller Cas9 orthologue of SpCas9—enables robust in vivo gene editing in diverse organisms, with the properties of lower cell toxicity, higher specificity and longer PAM requirement than SpCas9[6,16,17]. We systematically characterized St1Cas9 through determining multiple crystal structures of the St1Cas9–sgRNA–dsDNA complex. All of the structures represent an HNH catalytic state, which is not common in other determined Cas9 crystal structures. We also probed the possible mechanisms that result in the active conformation of St1Cas9, revealing that a wing region in the REC lobe and an aromatic residue in a linker region extended from the HNH domain play important roles in regulating St1Cas9 activity and probably contribute to stabilizing the active HNH conformation.

Fig. 4 | PAM recognition mechanism. a, A magnified view of the PAM duplex binding to the groove between the WED and PI domain. **b**, Detailed interactions of St1Cas9 with the PAM duplex. **c**, A detailed view of the interactions among the PI domain, L2 and the PAM duplex. **d**, A superimposition of structures that contain the 5′-AAAGAA-3′ PAM (orange), 5′-AAGGAA-3′ PAM (grey) and 5′-AAAGGA-3′ PAM (blue). **e**, The in vitro DNA cleavage activity of wild-type St1Cas9 towards linearized plasmid containing different PAMs. The experiment was performed in triplicate. **f**, In vitro DNA cleavage assay of the wild-type and various PAM-interacting-residue mutants of St1Cas9. The linearized plasmid containing the 5′-AAAGAA-3′ PAM was used as the cleavage substrate. The experiment was performed in triplicate. The results of the none and WT lanes here and in Fig. 3c were from the same experiment. **g**,**h**, Detailed interactions of SpCas9 (**g**, PDB ID: 4UN3) and SaCas9 (**h**, PDB ID: 5AXW) with their PAM duplexes.

The stabilization of the HNH active conformation by the wing region may impair the fidelity of St1Cas9. These mechanisms are unique to St1Cas9 and are absent in other currently determined Cas9 structures. Furthermore, St1Cas9 uses different strategies for PAM recognition compared to the Cas9 orthologues. As St1Cas9 has a relatively long PAM requirement for recognition, we engineered a St1Cas9 variant KLKL via structure-guided protein engineering, substantially expanding the targeting scope of this nuclease.



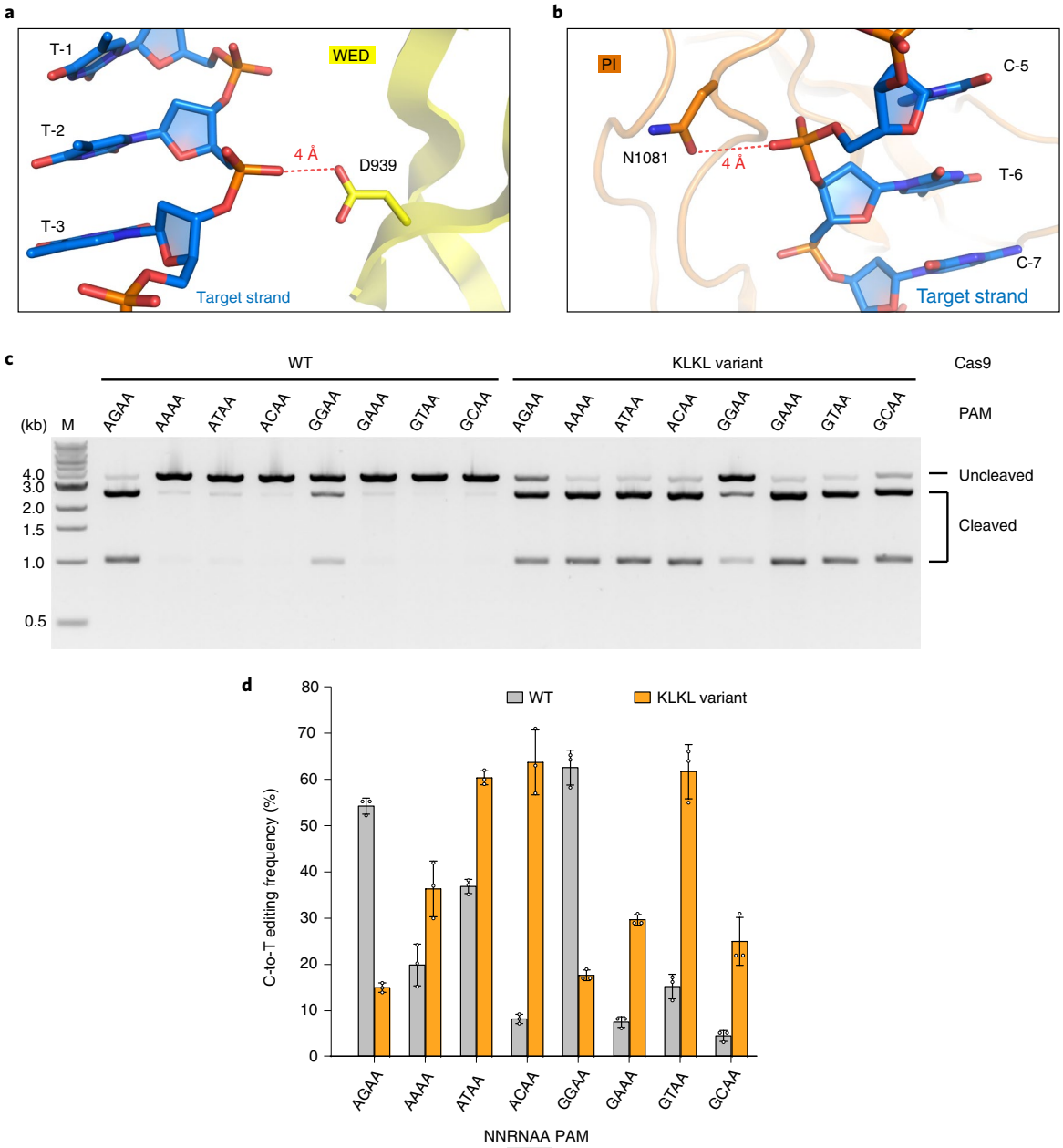


Fig. 5 | PAM expansion of St1Cas9 via structure-guided engineering. a,**b**, Residues D939 and N1081 were located adjacent to the phosphate backbone of the target DNA strand. **c**, In vitro DNA cleavage assay of the wild-type and the D939K/E1057L/N1081K/K1086L (KLKL) variant of St1Cas9. The linearized plasmids containing different PAM sequences were used as the substrates. The engineered St1Cas9 KLKL variant could recognize and cleave the DNA containing the 5′-AARNAA-3′ PAMs (R is A and G, N is any nucleotide). The experiment was performed in triplicate. **d**, C-to-T conversions using the fusions of St1Cas9 (WT)-cytidine deaminase and St1Cas9 (KLKL)-cytidine deaminase. The assay was performed in *K. pneumoniae* cells. The experiment was performed in triplicate. Error bars reflect mean ± s.d.

The recent cryo-EM structures of St1Cas9 uncovered an allosteric inhibition mechanism by an anti-CRISPR protein[47], but the lack of the HNH domain and the low resolution of the structures hamper the detailed understanding of the HNH catalytic and DNA recognition mechanisms (Supplementary Fig. 4a–f). Our high-resolution St1Cas9 crystal structures clearly unveil the HNH catalytic mode and the PAM-recognition mechanism. The subsequent structural and biochemical analyses pinpointed two major mechanisms that may result in the active conformation of the HNH domain. Given the important roles of those mechanisms in St1Cas9 catalytic activity, engineering the two mechanisms into other Cas9 nucleases could be an effective strategy to boost the catalytic activity of Cas9 nucleases. The detailed PAM–protein interaction mechanism also obtained in this study greatly aided the engineering of a St1Cas9 variant, KLKL, with a substantially expanded targeting scope. In addition to PAM expansion, structure-guided engineering has also been successfully applied for the evolution of high-fidelity Cas9 nucleases[33–35]. The high-resolution St1Cas9 structures may thus facilitate the rational engineering of high-fidelity St1Cas9 variants for therapeutic applications.

Methods

St1Cas9 expression and purification. The wild-type St1Cas9 gene was cloned into a modified pET28a vector encoding an N-terminal His6-tag and a human

rhinovirus 3C (HRV3C) protease cleavage site. Point mutations were introduced into the plasmid encoding the wild-type St1Cas9 by site-directed mutagenesis or Gibson assembly[59] (NEB) and further confirmed by DNA sequencing. All of the St1Cas9 variant proteins were expressed in the BL21 (DE3) *Escherichia coli* strain with the addition of 0.5 mM isopropyl b-D-1-thiogalactopyranoside at 18 °C for 14 h after the OD_{600} of the cells reached ~0.5–0.6. The cells were harvested by centrifugation at 6,000 g for 10 min and then resuspended in the lysate buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5% glycerol, 1 mM DTT) and disrupted by French Press. Clarified lysate was loaded onto a Ni-NTA column (GE Healthcare). The proteins bound to the column were eluted by a stepwise increase of imidazole (50 mM wash and 250 mM elution). To remove the N-terminal His6-tag, the eluted proteins were incubated with the HRV3C protease at 4 °C overnight and then clarified with a Ni-NTA column. Finally, the cleaved proteins were further purified by using a HiLoad Superdex 200 16/60 column (GE Healthcare) in the elution buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM DTT). All of the St1Cas9 variants were concentrated to 20 mg ml^{-1} using an Amicon Ultra 10 K filter (Millipore) and the proteins were supplemented with 20% glycerol, snap frozen in liquid nitrogen and stored at −80 °C.

Preparation of sgRNA and DNA. The transcription templates of sgRNAs were synthesized by GENEWIZ (Suzhou) and amplified by a PCR. sgRNAs were prepared by in vitro transcription using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) and further purified using phenol/chloroform extraction followed by ethanol precipitation. The oligonucleotides of the target and non-target DNA strands for crystallization were synthesized by Sangon. Target and complementary non-target DNA strands were mixed with a 1:1.1 molar ratio in the hybridization buffer (10 mM Tris-HCl (pH 7.5), 100 mM NaCl). The mixture was heated at 95 °C for 5 min and then slowly cooled to room temperature. The sequences of sgRNAs and DNAs are listed in Supplementary Table 2.

Protein crystallization and structure determination. The purified H599A protein, the truncated 70-nt sgRNA and the hybridized dsDNA (containing the AAAGAA PAM, the AAGGAA PAM and AAAGGA PAM, respectively) were mixed with a 1:1.5:2 molar ratio in the buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT) and incubated on ice for 20 min. Next, the St1Cas9–sgRNA–dsDNA complex was purified by using a HiLoad Superdex 200 16/60 column in the same buffer and then the elution product was concentrated to 13 mg ml^{-1} for crystallization. Crystals were obtained by mixing 1 µl of the St1Cas9–sgRNA–dsDNA complex with 1 µl reservoir solution and the mixture was incubated at 16 °C. Crystals containing the AAAGAA PAM were crystallized from the condition containing 0.012 M NaCl, 0.08 M KCl, 0.04 M sodium cacodylate trihydrate (pH 7.0), 40% v/v (±)-2-methyl-2,4-pentanediol and 0.012 M spermine tetrahydrochloride. Crystals containing the AAGGAA PAM or the AAAGGA PAM were crystallized from the condition containing 0.08 M KCl, 0.02 M $BaCl_2$, 0.04 sodium cacodylate trihydrate (pH 7.0), 40% v/v (±)-2-methyl-2,4-pentanediol and 0.012 M spermine tetrahydrochloride. All of the crystals were snap frozen in liquid nitrogen for X-ray diffraction.

X-ray diffraction data were collected at beamline BL19U1 of the Shanghai Synchrotron Radiation Facility. The data were processed by HKL3000[60]. The complex structure containing the AAAGGA PAM was determined using molecular replacement with Phaser from CCP4i[61]. The SaCas9-sgRNA-dsDNA complex structure (PDB ID: 5AXW) was used as the search model. Other structures were determined using the complex structure containing the AAAGGA PAM as the search model. All of the structure models were improved by iterative model building using COOT[62] and refinement using phenix.refine from PHENIX[63]. Data collection statistics are summarized in Supplementary Table 1. Structural figures were prepared using Pymol (http://www.pymol.org).

In vitro cleavage assay. The sequence containing the AAAGAA PAM was amplified by PCR using 2× Phanta Max Master Mix (Vazyme Biotech Co.) and then cloned into vector pUC19 via Gibson assembly (referred as pUC-AG). The plasmids with distinct PAM sites were constructed by site-directed mutagenesis, using pUC-AG as the template (Supplementary Table 2). The *Kpn*I-linearized plasmids were purified by SanPrep column from the PCR product purification kit (Sangon). In vitro cleavage reactions were performed in a 10 µl reaction buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT) containing 5 nM linearized plasmids, 250 nM St1Cas9 and 500 nM sgRNA (unless otherwise noted). Unless otherwise specified, reactions were incubated at 37 °C for 40 min and then quenched by adding 25 mM EDTA and 10 µg Proteinase K at 58 °C for 10 min. Reaction products were separated by a 1% agarose gel, stained with the 4S red plus dye (Sangon), visualized by the ChemiDoc MP System (Bio-Rad) and then verified by comparison with DNA standards (1 kb DNA ladder, NEB).

Electrophoretic mobility shift assays. Non-target strands and single 5′-FAM-labelled target strands were synthesized by Sangon. Target and non-target strands were hybridized with a 1:1.5 molar ratio in the hybridization buffer (10 mM Tris-HCl (pH 7.5), 100 mM NaCl). Catalytically inactive wile-type St1Cas9 (D9A/H599A, referred as $dSt1Cas9_{WT}$) and Cas9 with truncation of the wing region (truncated region: F169 to L180; D9A/H587A, referred as $dSt1Cas9_{\Delta wing}$) were used for the electrophoretic mobility shift assays. Cas9 and sgRNA were mixed with a 1:2 molar ratio in the reaction buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM $MgCl_2$, 50 µg ml^{-1} heparin, 0.3% Nonidet P-40, 1 mM DTT) and pre-incubated at 37 °C for 15 min before being incubated with 20 nM FAM-labelled dsDNA at 37 °C for 30 min. Reaction products were separated by a native 6% polyacrylamide gel at room temperature using 0.5× TBE supplemented with 5 mM $MgCl_2$ as the running buffer. The results were visualized by the ChemiDoc MP System (Bio-Rad).

Phylogenetic tree of wing region containing Cas9s. MAFFT version 7.450 (ref. [64]) was used to generate a multiple sequence alignment for the Cas9 superfamily (InterPro accession: IPR028629). Based on the multiple sequence alignment of the Cas9 superfamily, first 630 Cas9 sequences with the length containing the wing region no less than 23 residues (that is, search region from G163 to P185 containing the wing region in St1Cas9) were selected, then the selected sequences were clustered at 70% sequence identity. The representative sequence in each cluster was used for generating a phylogenetic tree. Neighbour joining method with 2,000 steps bootstrapping in Mega7[65] was used to generate the tree.

Bacterial survival assay and cytosine base editing assay in *K. pneumoniae*. The St1Cas9-mediated bacterial survival assay plasmid (referred as pCasKP-St1) was constructed by cloning the sgRNA, L-arabinose-inducible *ParaB* promoter and wild-type St1Cas9 gene into the backbone of pBECKP[66] via Gibson assembly[59]. A 20-nt spacer that targets the *dhaB3* gene (the spacer sequence: CCATTGACCGACATCACCCT) was inserted into the *Bsm*BI sites of pCasKP-St1. After electroporation with the constructed plasmids, the *K. pneumoniae* KP_1.6366 cells were recovered by adding 1 ml LB medium containing 0.2% L-arabinose. The cultures were incubated at 40 °C for 2 h and then plated onto an LB agar plate supplemented with 50 µg ml^{-1} kanamycin and 0.2% L-arabinose. The plates were incubated at 40 °C overnight.

The St1Cas9-mediated cytosine base editing plasmid (referred as pBECKP-St1) was constructed by cloning the rAPOBEC1-St1Cas9(D9A) gene into the backbone of pCasKP-St1 via Gibson assembly[59]. The St1Cas9 KLKL variant gene was subsequently introduced into the plasmid by Gibson assembly using pBECKP-St1 as the template. The 20-nt spacers were inserted into pBECKP-St1 by Golden Gate assembly. After electroporation with the constructed plasmids, the *K. pneumoniae* KP_1.6366 cells were recovered by adding 1 ml LB medium containing 0.2% L-arabinose. The cultures were incubated at 37 °C for 1.5 h before being plated onto an LB agar plate supplemented with 50 µg ml^{-1} kanamycin and 0.2% L-arabinose. After growth at 37 °C overnight, all of the colonies were collected and lysed by heating at 95 °C for 10 min. The editing regions were amplified by PCR using the lysed colonies as the templates (2× Phanta Max Master Mix, Vazyme Biotech Co.). Editing efficiencies were determined by Sanger sequencing of the PCR products and quantified using the EditR[67] software. The sequences of the spacers were listed in Supplementary Table 3. The sgRNA and protein sequences were listed in Supplementary Table 4.

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

Data availability

Conservation analysis of the active-site residues of the HNH domains was performed from the Cas9 protein family (InterPro accession: IPR028629). The structures of St1Cas9 in complex with different PAMs have been deposited in the Protein Data Bank under the accession codes: St1Cas9/GGAA (6M0V), St1Cas9/AGAA (6M0W) and St1Cas9/AGGA (6M0X). The data that support the plots within this paper and other findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

Received: 26 February 2020; Accepted: 30 July 2020;
Published online: 07 September 2020

References

1. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
2. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, 7 (2006).
3. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
4. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).

5. Sapranauskas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
6. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–2586 (2012).
7. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
8. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
9. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
10. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
11. Chen, W., Zhang, Y., Yeo, W. S., Bae, T. & Ji, Q. Rapid and efficient genome editing in *Staphylococcus aureus* by using an engineered CRISPR/Cas9 system. *J. Am. Chem. Soc.* **139**, 3790–3795 (2017).
12. Bao, Z. et al. Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **4**, 585–594 (2015).
13. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
14. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
15. Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
16. Rock, J. M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.* **2**, 16274 (2017).
17. Muller, M. et al. *Streptococcus thermophilus* CRISPR–Cas9 systems enable specific editing of the human genome. *Mol. Ther.* **24**, 636–644 (2016).
18. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
19. Ibraheim, R. et al. All-in-one adeno-associated virus delivery and genome editing by *Neisseria meningitidis* Cas9 in vivo. *Genome Biol.* **19**, 137 (2018).
20. Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
21. Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
22. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
23. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2**, 17092 (2017).
24. Hille, F. et al. The biology of CRISPR–Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
25. Hirano, H. et al. Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
26. Acharya, S. et al. *Francisella novicida* Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proc. Natl Acad. Sci. USA* **116**, 20959–20968 (2019).
27. Agudelo, D. et al. Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9. *Genome Res.* **30**, 107–117 (2020).
28. Kleinstiver, B. P. et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR–Cas9 by modifying PAM recognition. *Nat Biotechnol.* **33**, 1293–1298 (2015).
29. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
30. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
31. Miller, S. M. et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471–481 (2020).
32. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* **368**, 290–296 (2020).
33. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
34. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
35. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
36. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
37. Tan, Y. et al. Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high genome-wide specificity. *Proc. Natl Acad. Sci. USA* **116**, 20969–20976 (2019).
38. Bratovic, M. et al. Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat. Chem. Biol.* **16**, 587–595 (2020).
39. 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).
40. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
41. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
42. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
43. Nishimasu, H. et al. Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).
44. Zhu, X. et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **26**, 679 (2019).
45. Zuo, Z. et al. Structural and functional insights into the bona fide catalytic state of *Streptococcus pyogenes* Cas9 HNH nuclease domain. *eLife* **8**, e46500 (2019).
46. Sun, W. et al. Structures of *Neisseria meningitidis* Cas9 complexes in catalytically poised and anti-CRISPR-inhibited states. *Mol. Cell* **76**, 938 (2019).
47. Fuchsbauer, O. et al. Cas9 allosteric inhibition by the anti-CRISPR protein AcrIIA6. *Mol. Cell* **76**, 922–937 (2019).
48. Yamada, M. et al. Crystal structure of the minimal Cas9 from *Campylobacter jejuni* reveals the molecular diversity in the CRISPR–Cas9 systems. *Mol. Cell* **65**, 1109–1121 (2017).
49. Hirano, S. et al. Structural basis for the promiscuous PAM recognition by *Corynebacterium diphtheriae* Cas9. *Nat. Commun.* **10**, 1968 (2019).
50. Palermo, G. et al. Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* **51**, e91 (2018).
51. Guo, M. H. et al. Structural insights into a high fidelity variant of SpCas9. *Cell Res.* **29**, 183–192 (2019).
52. Jiang, F. & Doudna, J. A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
53. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
54. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
55. Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
56. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
57. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
58. Li, X. et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
59. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).
60. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
61. Bailey, S. The Ccp4 Suite—programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
65. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
66. Wang, Y. et al. CRISPR–Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Appl. Environ. Microbiol.* **84**, e01834-18 (2018).
67. Kluesner, M. G. et al. EditR: a method to quantify base editing from Sanger sequencing. *CRISPR J.* **1**, 239–250 (2018).

Acknowledgements

We thank staff from BL19U1 beamlines of the National Facility for Protein Science Shanghai at Shanghai Synchrotron Radiation Facility for help with data collection. This work was financially supported by the National Natural Science Foundation of China (grant nos. 21922705, 91753127 and 31700123), National Key R&D Program of China Grant (no. 2017YFA0506800) and the Shanghai Science and Technology Committee Rising-Star Program (grant no. 19QA1406000) to Q.J.

Author contributions

Q.J. conceived and directed the project. Y.Z. and Q.J. designed the research. Y.Z. performed most of the experiments. H.Z., Z.W. and N.T. helped express and purify St1Cas9 proteins. H.Z., Yujue Wang, W.C., Yannan Wang and Yu Wang helped perform the cytosine base editing assay in *K. pneumoniae*. H.Z., Yujue Wang, W.C., Yannan Wang, Z.W., N.T. and Yu Wang helped with other experiments. X.X. and S.Z. performed the bioinformatics analysis. Y.Z. and Q.J. analysed the data. Y.Z., J.G. and Q.J. carried out crystallographic studies. Y.Z. and Q.J. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41929-020-00506-9.

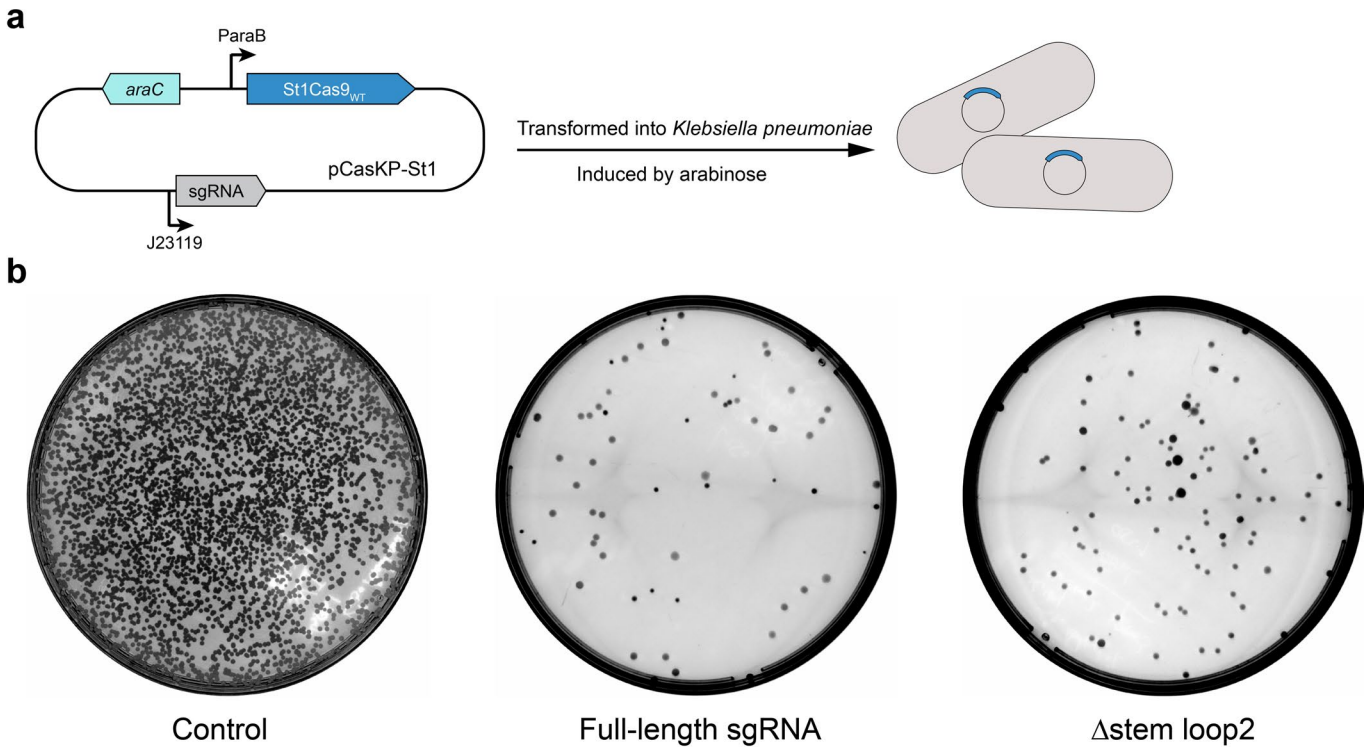
Supplementary information is available for this paper at https://doi.org/10.1038/s41929-020-00506-9.

Correspondence and requests for materials should be addressed to J.G. or Q.J.

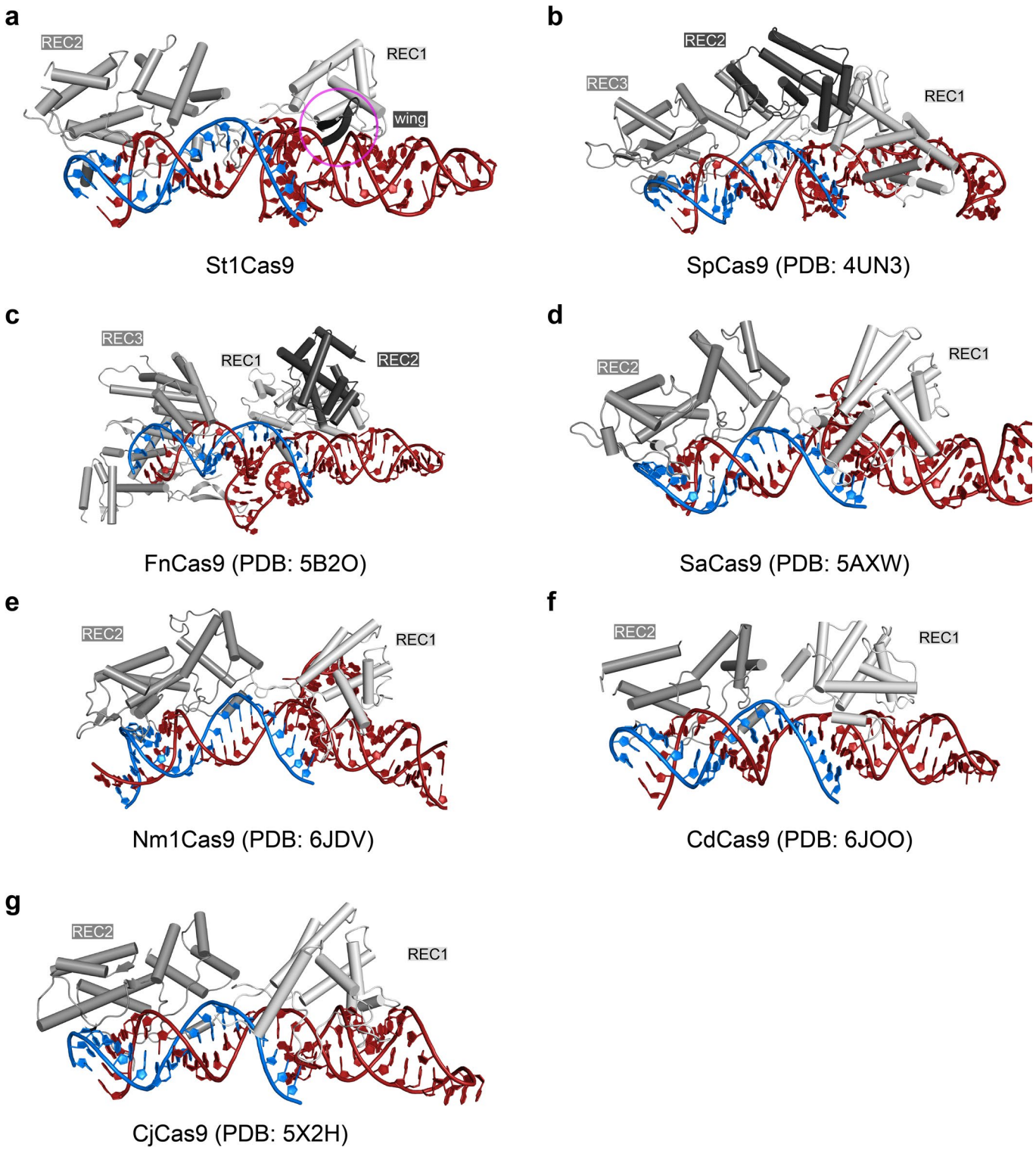
Reprints and permissions information is available at www.nature.com/reprints.

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

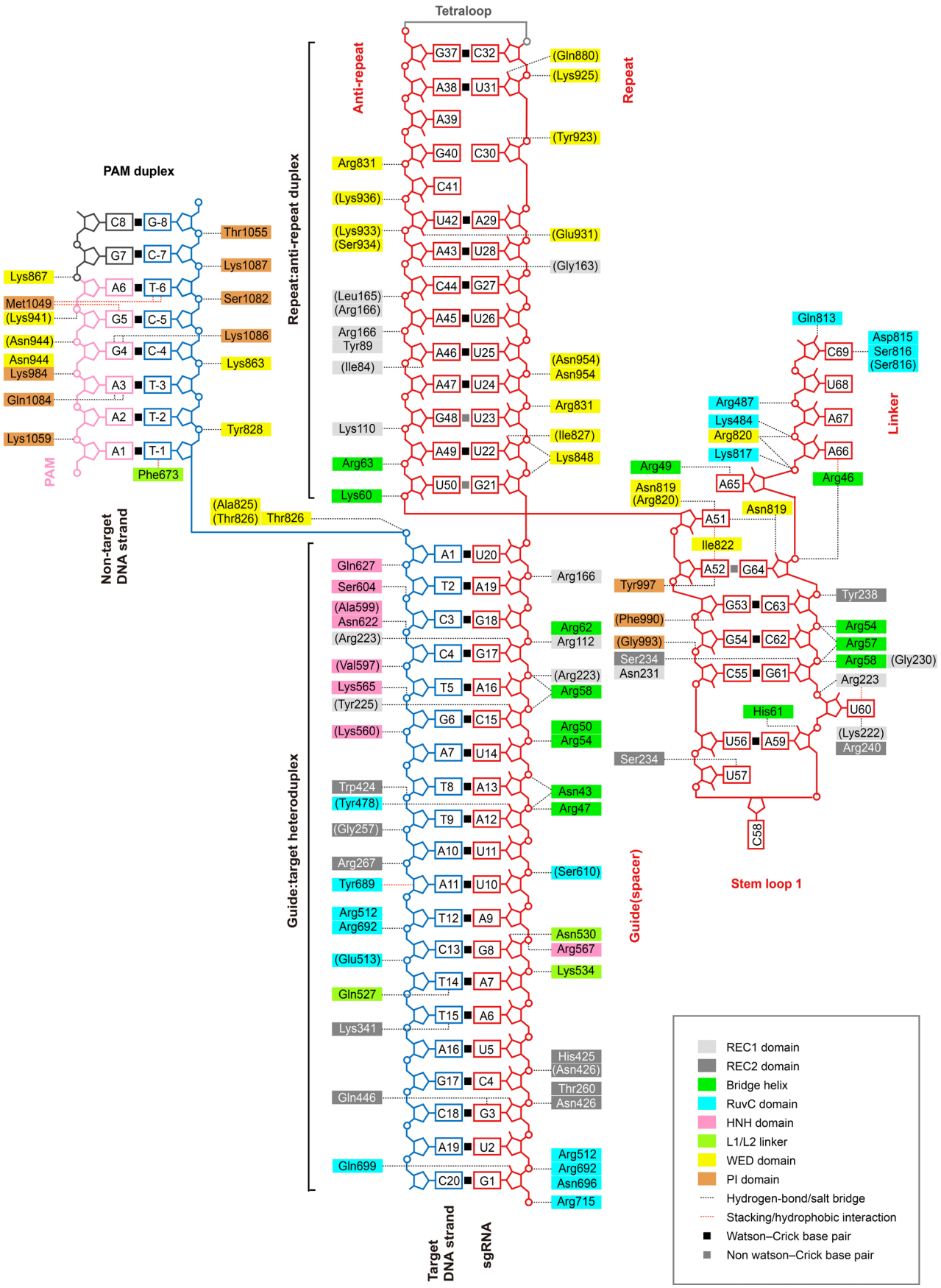
© The Author(s), under exclusive licence to Springer Nature Limited 2020



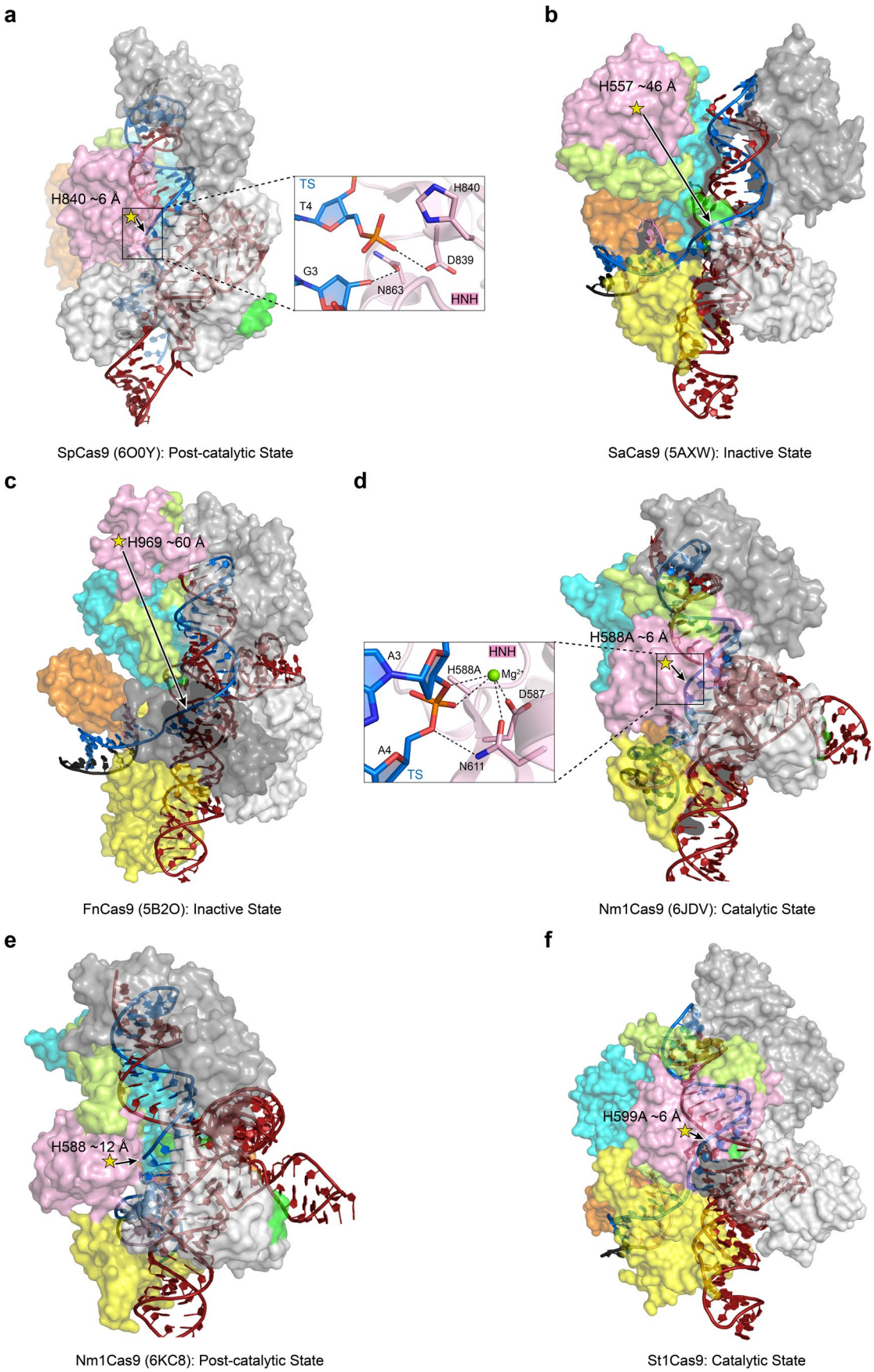
Extended Data Fig. 1 | Bacterial survival assay of St1Cas9 with full-length or truncated sgRNA. a, Scheme of St1Cas9-mediated bacterial survival assay in *Klebsiella pneumoniae* cells. **b**, The dhaB3 spacer-introduced pCasKP-St1Cas9 plasmids containing the full-length sgRNA (middle) or the sgRNA with the deletion of stem loop2 (right) efficiently killed *K. pneumoniae* cells. The empty pCasKP-St1Cas9 was transformed into *K. pneumoniae* cells as a control (left).



Extended Data Fig. 2 | The unique wing region is only present in St1Cas9 among the currently determined Cas9 structures. a, The REC lobe of St1Cas9 in this study. The wing region is highlighted by the pink cycle. **b**–**g**, The REC lobes of SpCas9 (**b**), FnCas9 (**c**), SaCas9 (**d**), Nm1Cas9 (**e**), CdCas9 (**f**) and CjCas9 (**g**), respectively.

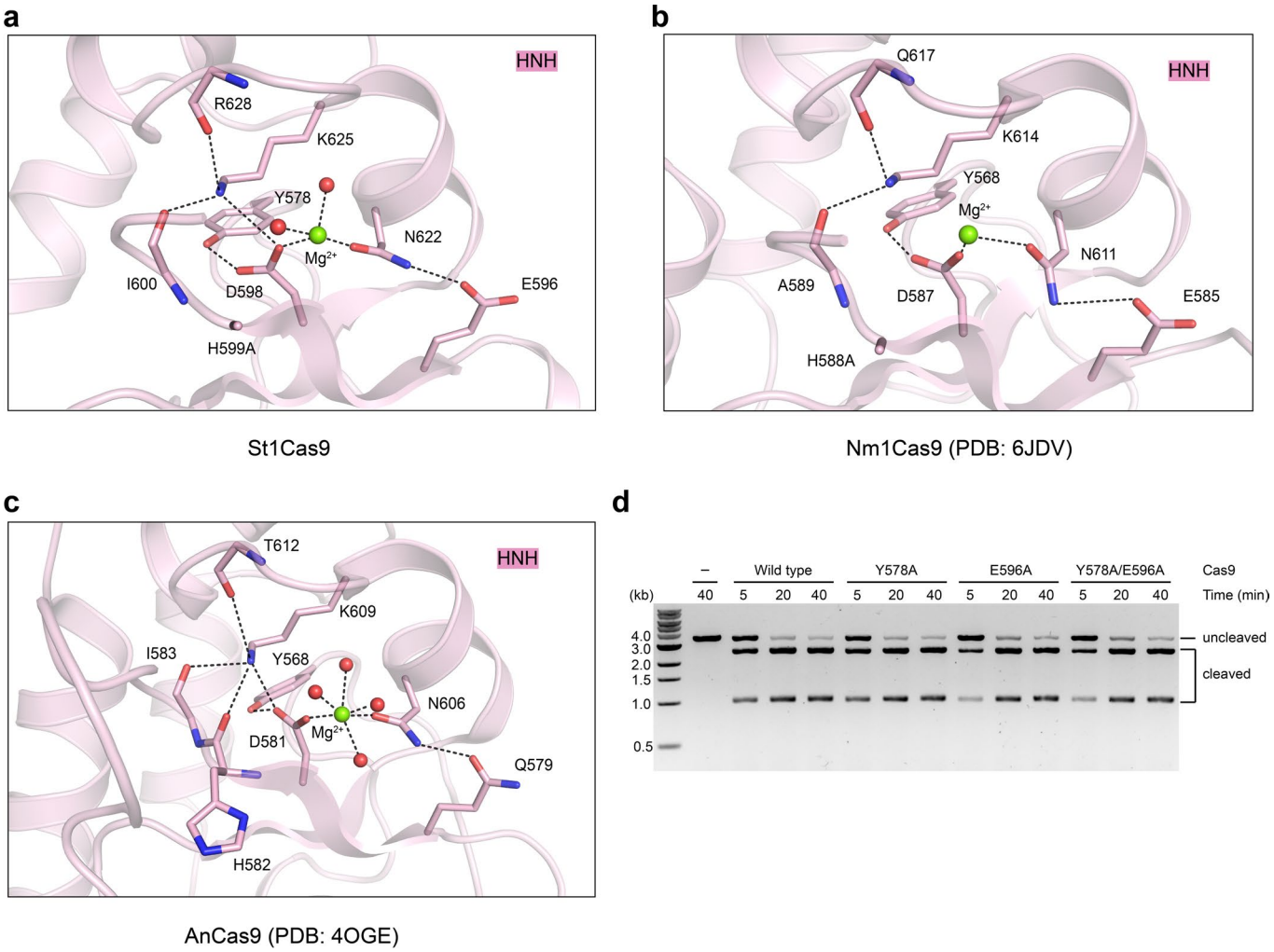


Extended Data Fig. 3 | Interactions between St1Cas9 and the nucleic acids. St1Cas9 residues that interact with the sgRNA-dsDNA duplex via their main chains are shown in parentheses. Water-mediated hydrogen-bonding interactions are not shown.

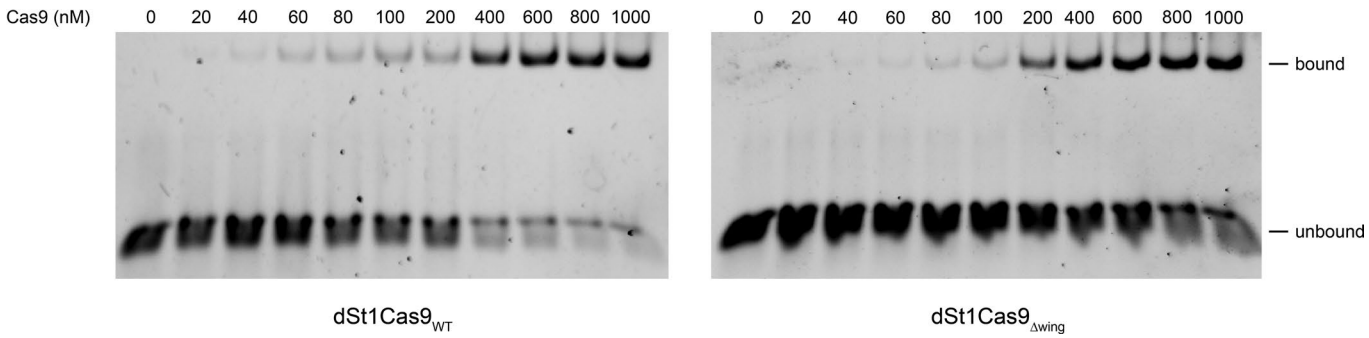


Extended Data Fig. 4 | See next page for caption.

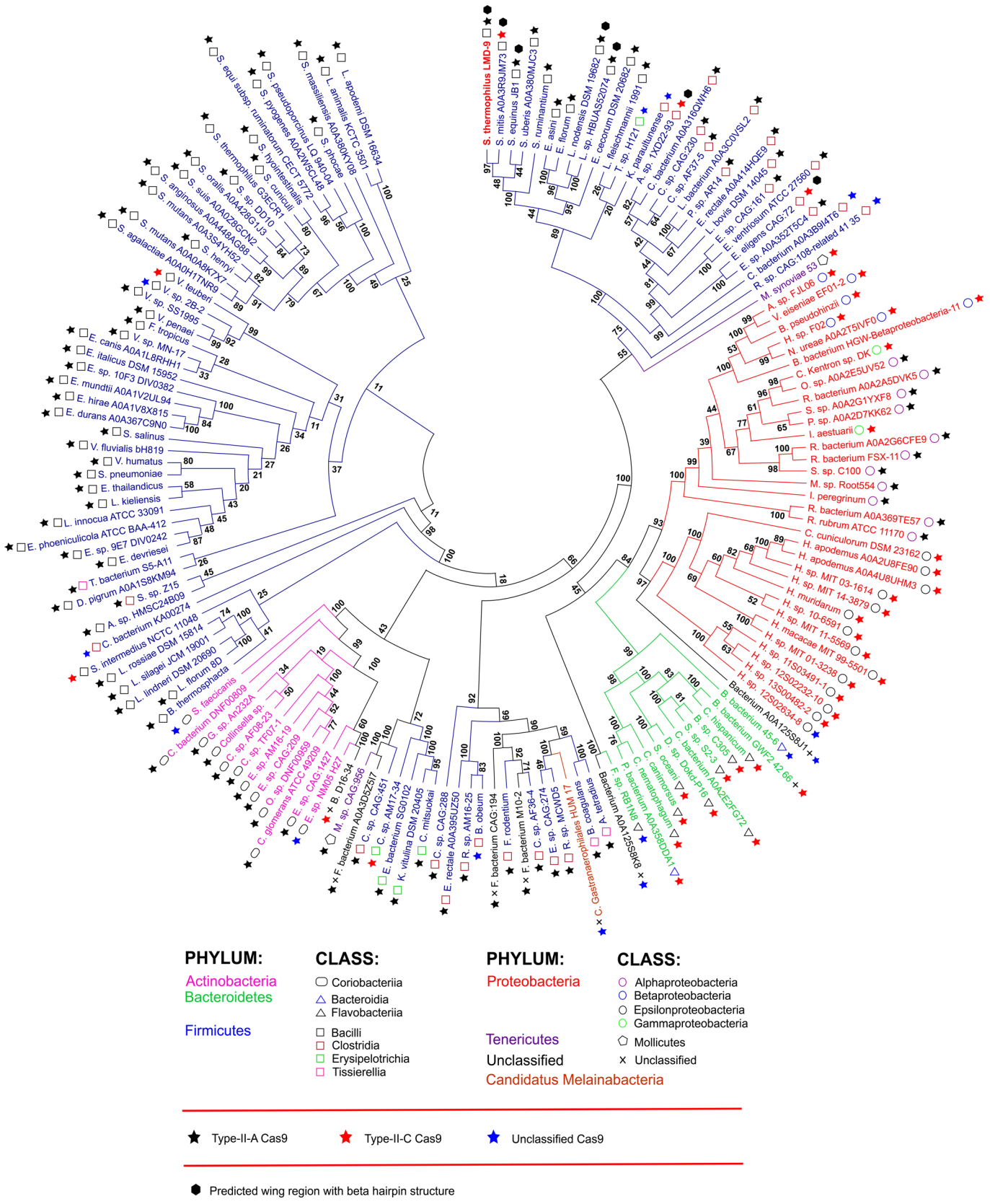
Extended Data Fig. 4 | Distinct conformations of the Cas9 orthologs. a, Post-catalytic state of DNA-bound SpCas9. Zoom-in view of the active sites of the SpCas9 HNH domain is shown in the right figure. **b** and **c**, Inactive states of DNA-bound SaCas9 (**b**) and DNA-bound FnCas9 (**c**). **d**, Catalytic state of DNA-bound Nm1Cas9. Zoom-in view of the active sites of the Nm1Cas9 HNH domain is shown in the left figure. **e**, Post-catalytic state of DNA-bound Nm1Cas9. **f**, Catalytic state of DNA-bound St1Cas9. All distances were measured between the Cα atom of the catalytic residue H and the scissile phosphate of TS nucleotides 3 and 4 in the catalytic or inactive state Cas9 structures (**b**–**d**, **f**), or between the Cα atom of the catalytic residue H and the 5′-PO_3 of TS nucleotide 4 in the post-catalytic state Cas9 structures (**a**, **e**). The positions of the catalytic residue H were shown as yellow stars.



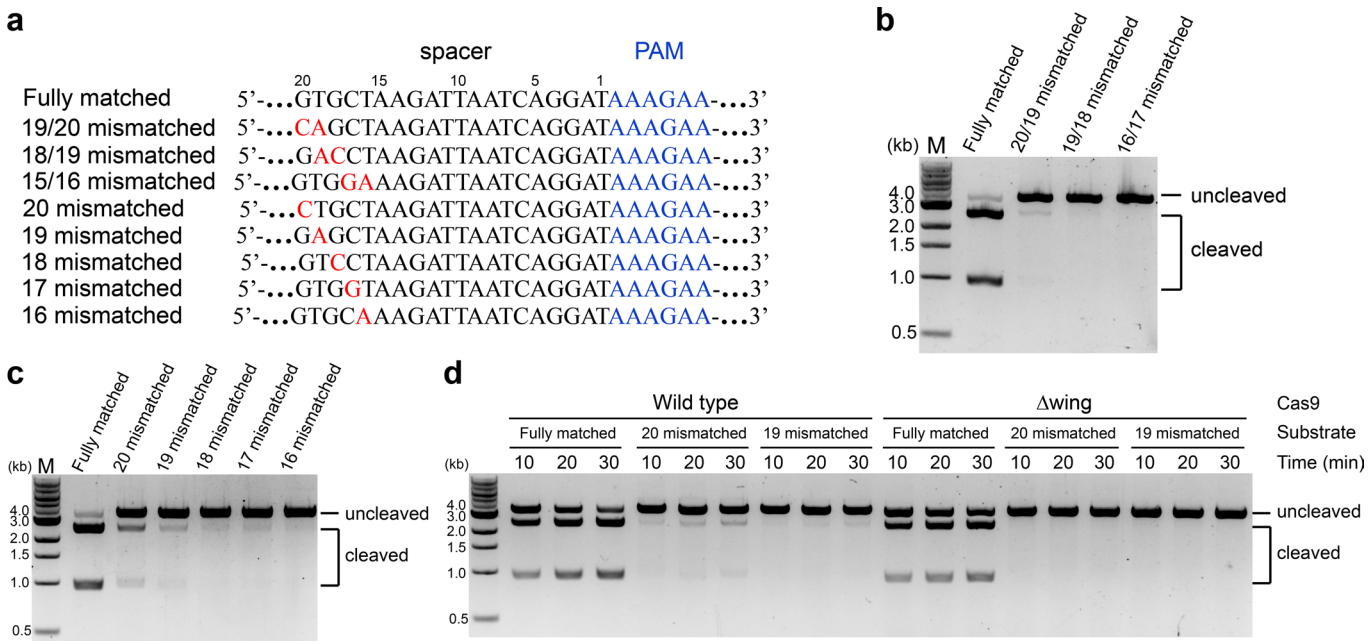
Extended Data Fig. 5 | Catalytic pockets of the HNH domains. a–c, Close-up view of Mg^{2+}-binding catalytic pockets of the HNH domains of St1Cas9 (**a**), Nm1Cas9 (**b**) and AnCas9 (**c**). Mg^{2+} and water molecules were coloured green and red, respectively. **d**, In vitro DNA cleavage assay of the wild type and various mutants of St1Cas9. The linearized plasmid containing a target sequence was used as the substrate and the full-length sgRNA was used as the guide RNA in the assay.



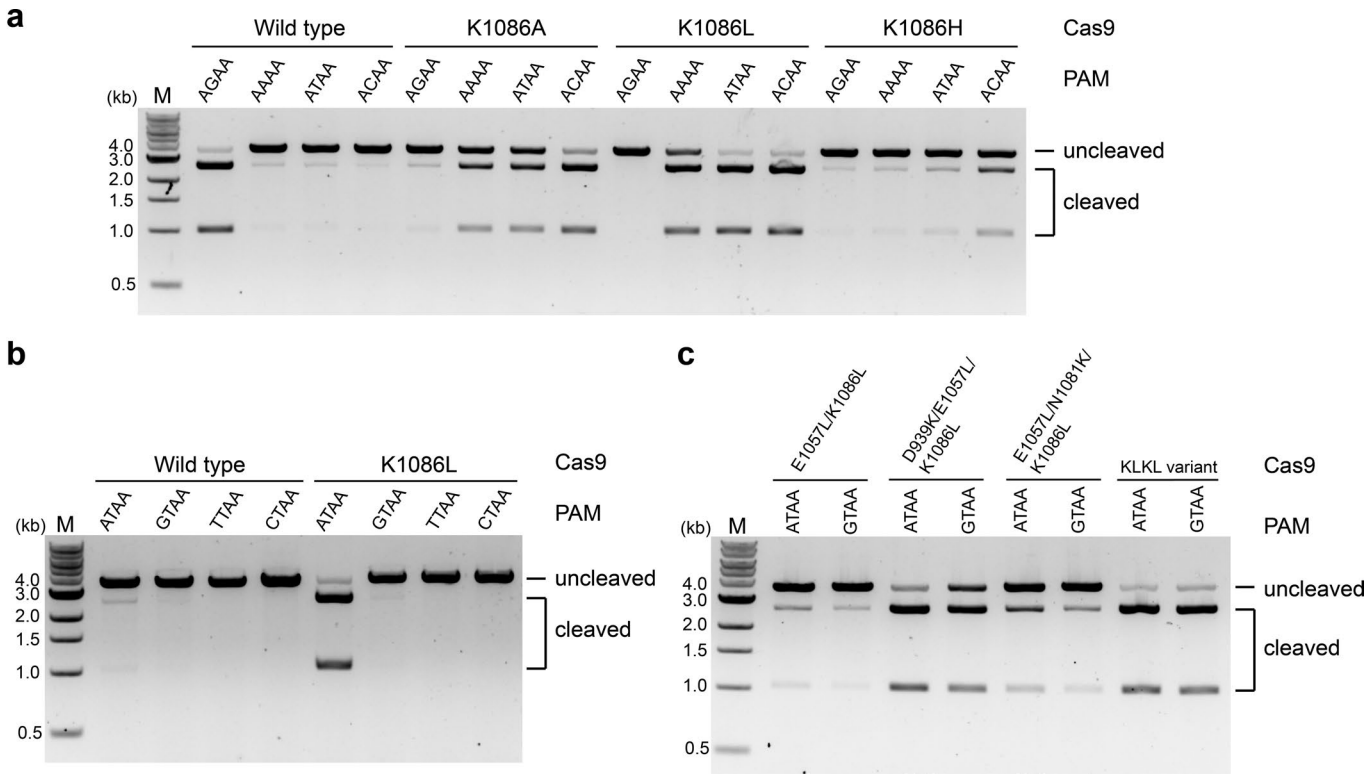
Extended Data Fig. 6 | DNA binding activities of the wild type St1Cas9 and the variant with the truncation of the wing region. Electrophoretic mobility shift assay was performed using catalytic inactive dCas9-sgRNA complexes and the FAM-labelled target dsDNA.



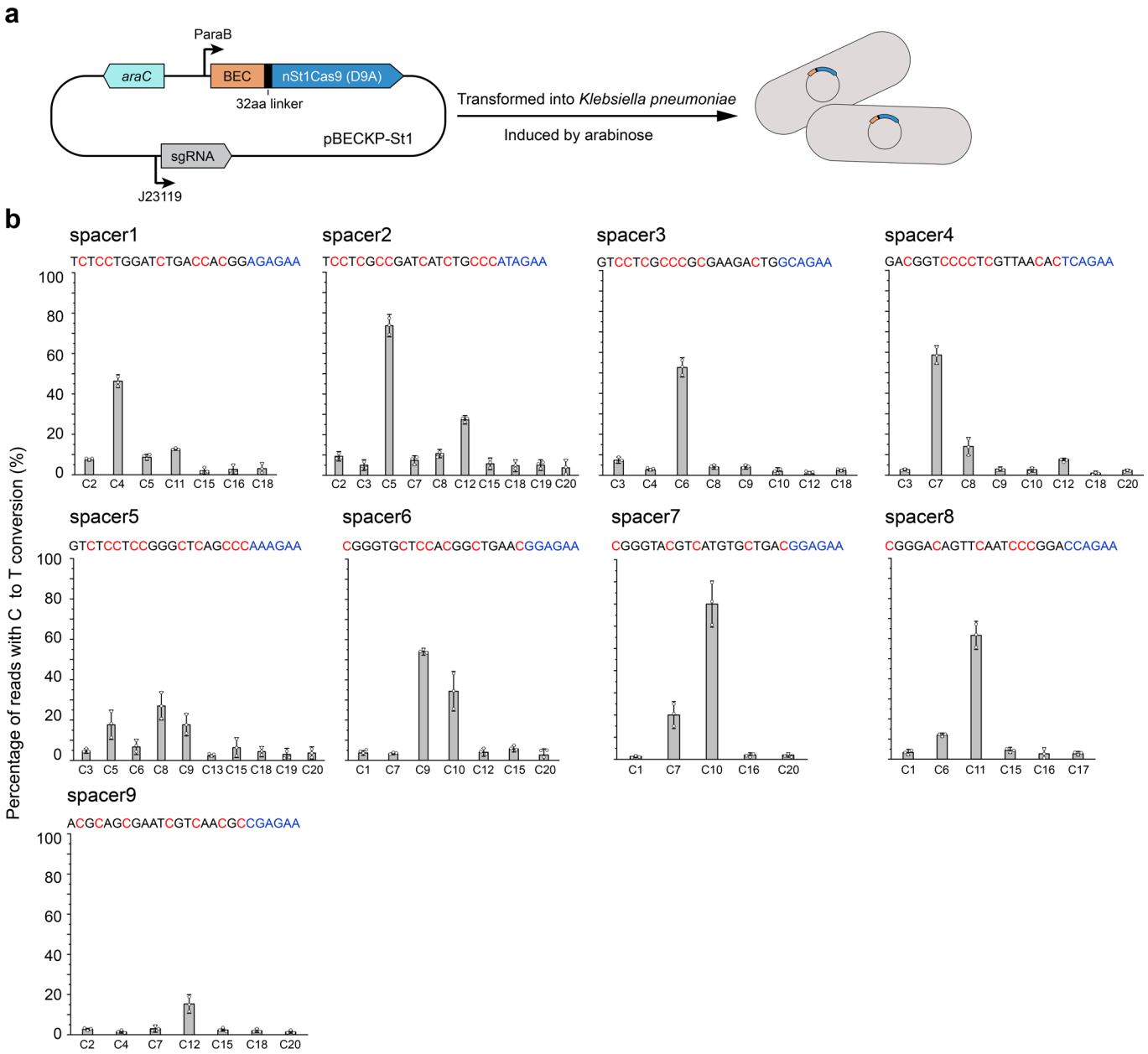
Extended Data Fig. 7 | Phylogenetic tree of putative wing region containing Cas9s. Each sequence represents sequences with >70% sequence identity.



Extended Data Fig. 8 | In vitro cleavage activities of wild type St1Cas9 and the variant with the truncation of the wing region. a, The spacer sequences of the linearized plasmids with single or double mismatches at positions 20-16. **b**, In vitro DNA cleavage assay of wild type St1Cas9. The linearized plasmids containing double mismatches at positions 20-16 were used as the substrates. **c**, In vitro DNA cleavage assay of wild type St1Cas9. The linearized plasmids containing single mismatches at positions 20-16 were used as the substrates. **d**, In vitro DNA cleavage assay of wild type St1Cas9 and the variant with the truncation of the wing region. The linearized plasmids containing single mismatches at positions 20 or 19 were used as the substrates. The reaction condition of wild-type St1Cas9 was 5 nM linearized plasmid, 120 nM Cas9, and 240 nM sgRNA. The reaction condition of the St1Cas9 variant with the truncation of the wing region was 5 nM linearized plasmid, 500 nM Cas9, and 1 μM sgRNA.



Extended Data Fig. 9 | In vitro cleavage activities of the wild type and various mutants of St1Cas9. a, In vitro DNA cleavage assay of the wild type and various K1086 mutants of St1Cas9. The linearized plasmids containing the 5′-AAANAA-3′ PAM sequences were used as the substrates. **b**, In vitro DNA cleavage assay of the wild type and K1086L mutant of St1Cas9. The linearized plasmids containing the 5′-AANTAA-3′ PAM sequences were used as the substrates. **c**, In vitro DNA cleavage assay of the wild type and various engineered mutants of St1Cas9. The linearized plasmids containing the 5′-AARTAA-3′ PAM sequences were used as the substrates.



Extended Data Fig. 10 | Determination of the editing window of St1Cas9-mediated cytosine base editing system in *K. pneumoniae*. a, Scheme of St1Cas9-mediated cytosine base editing. **b**, The editing results of various C-rich spacers in the KP_1.6366 strain. The Cs were coloured red and the target PAM sites were coloured blue. The experiment was performed in triplicate. Error bars report standard deviations of the mean (s.d.).

nature research

Corresponding author(s): Quanjiang Ji; Jianhua Gan

Last updated by author(s): Jul 12, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	HKL3000
Data analysis	CCP4i version 7.0.073; Phenix 1.13; COOT version 0.8.9; PyMol (version 1.8.2.3); MAFFT (webserver, January-June 2019); MEGA version 7.0

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Conservation analysis of the active-site residues of the HNH domains was performed from the Cas9 protein family (InterPro assession: IPR028629). The structures of St1Cas9 in complex with different PAMs have been deposited in the Protein Data Bank under the accession codes: St1Cas9/GGAA (6M0V), St1Cas9/AGAA (6M0W), and St1Cas9/AGGA (6M0X).

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No sample size calculations were made because this kind of analysis is not relevant for this work.
Data exclusions	No data were excluded from analysis
Replication	All biochemical experiments were independently reproduced at least three times. All attempts at replication were successful.
Randomization	Randomization was not performed for this study.
Blinding	Blinding was not performed for this study because this is not required in this field.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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