Structural biology of CRISPR–Cas immunity and genome editing enzymes

Joy Y. Wang[1,2], *Patrick Pausch*[3]✉ *and Jennifer A. Doudna*[1,2,4,5,6,7,8,9]✉

Abstract | CRISPR–Cas systems provide resistance against foreign mobile genetic elements and have a wide range of genome editing and biotechnological applications. In this Review, we examine recent advances in understanding the molecular structures and mechanisms of enzymes comprising bacterial RNA-guided CRISPR–Cas immune systems and deployed for wide-ranging genome editing applications. We explore the adaptive and interference aspects of CRISPR–Cas function as well as open questions about the molecular mechanisms responsible for genome targeting. These structural insights reflect close evolutionary links between CRISPR–Cas systems and mobile genetic elements, including the origins and evolution of CRISPR–Cas systems from DNA transposons, retrotransposons and toxin–antitoxin modules. We discuss how the evolution and structural diversity of CRISPR–Cas systems explain their functional complexity and utility as genome editing tools.

CRISPRs and Cas proteins provide microorganisms with RNA-guided adaptive immunity and offer transformative technological opportunities for programmable genome manipulation[1,2]. Cas9 and related enzymes are now widely used to edit or regulate the genomes of cultured and primary cells, animals and plants, vastly accelerating the pace of fundamental research and enabling breakthroughs in agriculture and synthetic biology. In addition, genome editing offers the potential to both understand human genetics and cure genetic disease as never before. The biology and technological capabilities of CRISPR–Cas systems have driven efforts to understand the molecules responsible for CRISPR–Cas functions, including targeted DNA binding, cutting, editing and integration.

CRISPR–Cas systems are structurally and mechanistically diverse. These systems typically consist of the CRISPR array, an adaptation module and a CRISPR RNA (crRNA) biogenesis and DNA/RNA-interference module (reviewed in REFS[3,4]) (FIGS 1,2). To provide adaptive and heritable immunity, the CRISPR array stores the genetic information of mobile genetic elements (MGEs) as 'spacer' sequences (typically ~25–50 bp in size, although the size can range from ~17 to ~72 bp)[5,6] inserted between short palindromic repeats (reviewed in REF.[7]). The Cas1–Cas2 adaptation machinery captures a segment of viral or plasmid DNA, the protospacer, in a bacterial cell and integrates it into the CRISPR array (FIG. 1). In DNA-targeting CRISPR–Cas systems, protospacer selection depends on the presence of a 3–5-bp-long protospacer adjacent motif (PAM) that is not integrated into the CRISPR array and serves to distinguish self from non-self target sequences (reviewed in REF.[8]).

Once the CRISPR array is transcribed, dedicated pre-crRNA processing Cas ribonucleases, or *trans*-activating crRNA (tracrRNA) and 'host factor' RNases, associate with and cleave the pre-crRNA repeat sequence to release individual crRNAs (reviewed in REF.[9]) (FIG. 1). Depending on the CRISPR class, several Cas proteins (class 1), or a single Cas protein (class 2), recruit the mature crRNA guide for DNA or RNA interference (FIG. 1). To prevent self-targeting of the complementary spacer within the CRISPR array, DNA-targeting systems recognize a PAM before interrogation of double-stranded DNA (dsDNA). Hybridization of the crRNA spacer sequence to the unwound target nucleic acid results in a conformational shift in the interference complex, activating the Cas nuclease for DNA or RNA cutting and invader destruction.

During the ongoing arms race between prokaryotes and viruses, structurally and mechanistically diverse CRISPR–Cas systems and anti-CRISPR proteins (Acr proteins) evolved to provide their hosts with a competitive advantage (reviewed in REFS[10,11]). Acr proteins inhibit or inactivate CRISPR-effector complexes, preventing recognition and degradation of target sequences (reviewed in REF.[12]). Currently, the two CRISPR classes are subdivided into six types and more than 30 subtypes (reviewed in REF.[11]), differentiated by their interference module composition and nucleic acid target specificity (FIG. 2). For class I (types I, III and IV) and RNA-targeting class II (type VI) interference complexes, we refer the reader to recent reviews[13–16].

✉*e-mail: patrick.pausch@gmc.vu.lt; doudna@berkeley.edu*

https://doi.org/10.1038/s41579-022-00739-4

Advances in both biochemical and imaging methods have accelerated structural studies of Cas proteins and CRISPR–Cas complexes. In this Review, we focus on the mechanisms of CRISPR–Cas adaptive immunity, the diversity of DNA-targeting enzymes and emerging systems including CRISPR transposases that are advancing in part due to structural investigations. We discuss the process by which CRISPR–Cas systems provide adaptive immunity in microorganisms by capturing and integrating foreign DNA into CRISPR sequence arrays. We then turn to the mechanisms of genome surveillance and RNA-guided DNA targeting that lie at the heart of both bacterial adaptive immunity and genome editing applications. Finally, we explore the structural biology and functional implications of CRISPR transposases, whose biological activities and technological potential remain to be fully explored. Overall, the research discussed here represents the extraordinary pace of discovery and development that underscores the progression of genome editing as a transformative toolbox for targeted genetic manipulation.

Origins and mechanisms of CRISPR spacer capture

Origins and evolution of the CRISPR integrase. The emergence of the CRISPR adaptation module was a key event in the origins of prokaryotic adaptive immunity, enabling bacteria and archaea to record previous infections in the form of short foreign sequences inserted as spacers in host cell CRISPR arrays (FIG. 1). Phylogenetic studies suggest that the CRISPR adaptation module evolved from ancestral casposons, a novel class of DNA transposons that encode a Cas1 homologue, the caposase, as the transposase[17]. The remarkable parallels in structure and function between the casposase and CRISPR integrase provide insight into the elements necessary for the emergence of CRISPR adaptation (FIG. 3a). Cas1 and *Methanosarcina mazei* casposase share a canonical two-domain architecture with an eight-stranded amino-terminal domain and a carboxy-terminal (C-terminal) domain containing the conserved active site residues E-H-E[18]. The structure of the casposase dimer bound to the integration target and the single-stranded DNA (ssDNA) casposon ends resembles that of the Cas1 dimer bound to the CRISPR repeat and 3′ overhangs of the protospacer[19–22]. Interestingly, the casposase contains an additional C-terminal helix–turn–helix domain, although it is unclear what role it plays in integration. The structural resemblance between the casposase and Cas1 translates to many functional similarities. The *M. mazei* casposase's unique ability among transposases to integrate a variety of non-sequence-specific substrates provides the groundwork for the emergence of an adaptive immune system that needs to acquire diverse sequences to maintain a genetic record of prior infections[18]. Upon integration, the casposase generates 14–15-bp target site duplications[23,24], similarly to how the CRISPR integrase generates 25–50-bp CRISPR repeat duplications[25–28]. Strikingly, the casposase's mode of target recognition also parallels that of the CRISPR integrase. Unlike other DNA transposons that insert DNA into random locations[23], the casposase inserts substrates into a preferred target site[18,24], which is a key characteristic of the CRISPR integrase[29,30]. For the *M. mazei* casposase, a conserved target motif within the target site duplications and a sequence motif 12–17 bp upstream of the target site duplication are crucial for directing integration[18]. This is analogous to CRISPR spacer acquisition, which depends on both a specific repeat sequence and sequence motifs within the upstream leader[25,29,30]. These similarities give insight into the origins of many key elements of CRISPR adaptation.

The casposase structure also provides clues as to how the Cas1 integrase evolved to accommodate Cas2, a key subunit in most CRISPR integrases, as well as other binding partners involved in CRISPR adaptation[19,21,31] (FIG. 3a; Supplementary Table 1). The casposase forms a tetramer for integration, composed of two dimers holding the target site between them[18]. Although the casposase is structurally impeded from simultaneous tetramerization and binding to Cas2, similarities between the relevant interface of the casposase dimer and the Cas2-interacting interface of Cas1 suggest that surface mutations could allow the casposase dimer to bind to a Cas2 protomer. Introduction of a bridging Cas2 would maintain the orientation of the catalytic monomers towards each other and provide an extended binding surface for short DNA integration substrates. In addition to Cas2, the casposase also has interfaces that could accommodate binding to the CRISPR adaptation proteins Csn2 and Cas4, which assist Cas1–Cas2 in spacer acquisition in certain systems and are discussed in greater detail in later sections of this Review[32]. The relevant binding interface with the auxiliary protein Csn2 is the eight-stranded amino-terminal domain, which is remarkably similar in the Cas1 and casposase structures[18,33]. At the dimer level, the casposase also appears able to accommodate interactions with a Cas4 nuclease and even has a conserved asparagine residue that is important for Cas4 binding[18,34]. However, interestingly, in both the casposase and CRISPR Cas4–Cas1–Cas2 integrase structures, the relevant interface for Cas4 binding overlaps with the interface involved in integration target binding. Further work is required to elucidate when Cas1 acquired these binding partners in its evolution from the casposase into a functional CRISPR integrase.

Phylogenetic analyses imply the existence of an ancestral CRISPR integrase composed of only Cas1 before the adoption of Cas2 (REFS[5,35]). This Cas1 integrase

Author addresses

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA.
[2]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA.
[3]VU LSC-EMBL Partnership for Genome Editing Technologies, Life Sciences Center, Vilnius University, Vilnius, Lithuania.
[4]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA.
[5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA.
[6]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA, USA.
[7]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[8]Gladstone Institutes, University of California, San Francisco, San Francisco, CA, USA.
[9]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA.

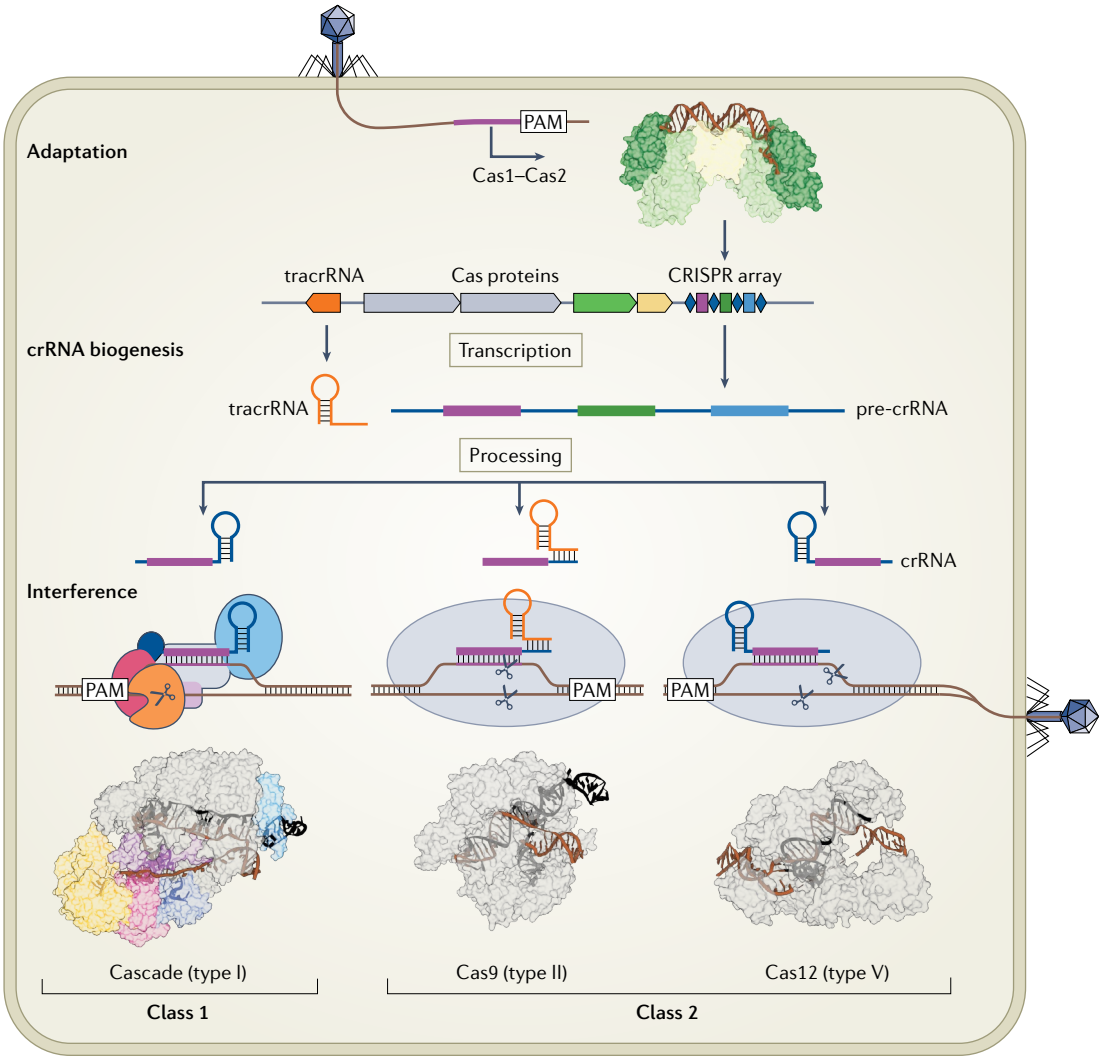


Fig. 1 | **CRISPR–Cas systems provide bacteria and archaea with adaptive immunity.** The three stages of CRISPR immunity: adaptation, CRISPR RNA (crRNA) biogenesis and interference. In adaptation, Cas1–Cas2 (Protein Data Bank (PDB) ID 5DS4)[20] inserts protospacers, derived from foreign genetic elements, into the CRISPR array as new spacers (represented as differently coloured rectangles) that are separated by CRISPR repeats (represented as blue diamonds). During crRNA biogenesis, the CRISPR array is transcribed into pre-crRNA, which is processed into mature crRNAs that each have a single spacer. The crRNA (or in some cases, the dual crRNA–*trans*-activating crRNA (tracrRNA)) assembles with the effector protein or complex to form a surveillance complex that recognizes and degrades foreign genetic elements complementary to the crRNA spacer during interference. Class 1 systems have multisubunit effector complexes, whereas class 2 systems have single-subunit effector proteins. Target cleavage by the class 1 type I Cascade–Cas3 effector complex (left), the class 2 type II Cas9 effector (centre) and the class 2 type V Cas12a effector (right) is depicted schematically, and representative structures of the effector complexes are shown beneath: Cascade–Cas3 bound to crRNA and target DNA (PDB ID 6C66)[181]; Cas9 bound to guide RNA (composed of crRNA and tracrRNA) and target DNA (PDB ID 4UN3)[91]; and Cas12a bound to crRNA and target DNA (PDB ID 5NFV)[124]. Nucleic acids in the structures are colour-coded: DNA, brown; RNA, black. PAM, protospacer adjacent motif.

appears to mark the emergence of a precise ruler mechanism characteristic of all CRISPR integrases that defines the length of integration substrates (FIG. 3). In a Cas1 phylogenetic tree from a recent review, the subtype V-C and V-D Cas1 genes along with solo Cas1 genes form a branch rooted near the ancestral casposon branch[36]. These systems lack Cas2 entirely and have unusually short spacers (17–20 bp) and repeats (25 bp) compared with other CRISPR–Cas systems. Like the casposase, the subtype V-C Cas1 also forms a tetramer for integration[37]; however, whereas the casposase has a wide tolerance for different substrate lengths[18], the subtype V-C integrase has a precise ruler mechanism that favours short integration substrates. More structural characterization is required to determine the evolutionary path from the casposase to a functional CRISPR integrase and the emergence of the ruler mechanism. Interestingly, compared with the casposase, the subtype V-C Cas1 integrase shows greater promiscuity for the integration site. Across different CRISPR–Cas systems, CRISPR integrases display differing levels of intrinsic specificity for the integration site, and some rely on host factors

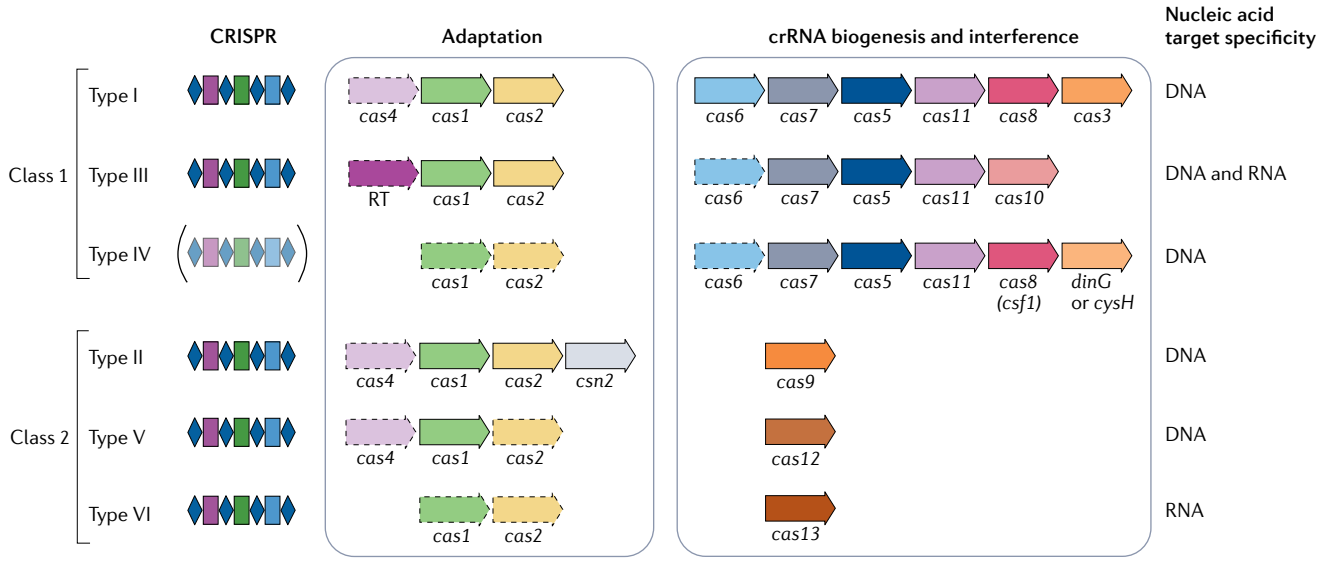


Fig. 2 | **Modular organization of CRISPR–Cas systems.** CRISPR–Cas systems are classified into two classes and six types. The functional modules involved in adaptation and CRISPR RNA (crRNA) biogenesis and interference for each type are illustrated, and the nucleic acid target specificity is indicated. Genes present in only some subtypes are marked by dashed outlines. RT, reverse transcriptase.

such as the integration host factor to guide them to the correct site (reviewed in REF.[32]). For systems that rely on integration host factor, the phase of the integration host factor binding motifs appears to be crucial to integration efficiency, as uncovered by a recent study examining the diversity and evolution of different CRISPR leader sequences[38]. It is unclear what host factors the subtype V-C Cas1 integrase may require for specific integration. Further investigation could shine light on the evolution of the mechanisms by which CRISPR–Cas systems identify the preferred integration site and coevolution of the adaptation module with their supporting host factors.

A key event in the evolution of the CRISPR integrase is the acquisition of Cas2, transforming the tetrameric structure of the casposase and subtype V-C Cas1 into the canonical heterohexameric Cas1–Cas2 complex observed in most CRISPR–Cas systems (Supplementary Table 1). On the basis of comparative sequence studies, the CRISPR adaptation module most likely adopted Cas2 from toxin–antitoxin (TA) modules[3,39] (FIG. 3a). Structural comparisons with the functional TA module of *Haemophilus influenzae* VapXD provide further insight into the evolutionary connection between TA systems and CRISPR adaptation[40]. The protein VapD is structurally related to Cas2: both form homodimers with a core ferredoxin fold and share conserved structural elements, including a similar nuclease active site, although Cas2 nuclease activity does not appear to be required for CRISPR acquisition[31,41,42]. Further work is required to understand VapD's mode of action in contributing to pathogenicity in the TA module and whether and how it is relevant to Cas2's role in CRISPR–Cas systems. The main structural differences between VapD and Cas2 occur in regions involved in interactions with the VapX antitoxin and Cas1 (REF.[40]). VapD has two additional helices involved in VapX antitoxin binding, which, on the basis of alignments with the Cas1–Cas2 structure bound to protospacer DNA, would sterically occlude protospacer binding[19–22,40]. Cas2, on the other hand, has an additional β-hairpin at the C terminus responsible for Cas1 interaction. These architectural differences hint at the loss of VapX as a binding partner and the adoption of Cas1 as a new binding partner in the transformation of VapD from a toxin in a TA system to a crucial component of the CRISPR adaptation module. There is a potential evolutionary advantage for the adoption of Cas2 as a structural unit in the integrase structure, which could provide increased stability and would lengthen the internal ruler to enable integration of longer sequences, offering greater targeting specificity of the crRNA for downstream CRISPR immunity processes[37]. However, the loss of the VapX antitoxin as a binding partner could have toxic effects, which the bacteria would need to account for to recruit VapD into the CRISPR integrase.

Recruitment of reverse transcriptases for RNA spacer acquisition. In the evolution of CRISPR spacer capture, some acquisition modules apparently recruited reverse transcriptases (RTs) from group II introns, enabling these systems to acquire spacers from RNA in addition to DNA[43–45] (FIG. 3). In several of these systems, the RT is fused with either Cas1 or the Cas6 maturase, enabling coordinated integration and crRNA generation[46,47] (FIG. 3b). The *Marinomonas mediterranea* Cas6–RT–Cas1–Cas2 integrase has been shown to carry out precise ligation of short ssRNA substrates into the CRISPR array in vitro and to conduct target primed reverse transcription using the free 3′ DNA end generated after RNA ligation[43]. Interestingly, for these fusions, the Cas6 domain is required for RNA spacer acquisition, and analysis of RT-containing systems suggests coevolution of the Cas6 and RT domains[46]. The recently characterized *Thiomicrospira* Cas6–RT–Cas1–Cas2 complex shows structural interactions between the Cas1–Cas2

integrase, the RT and the Cas6 maturase, enabling functional crosstalk between the three active sites[47] (FIG. 3a). A connecting α-helix links the RT and Cas1 active sites, and serves as a potential regulator of the two activities. Although much of the RT domain aligns closely to the group II intron RT structure[48], its palm region containing the catalytic centre appears to be in an inactive conformation[47]. The distinct RT conformation and the occluded Cas1 active site hint at larger conformational changes that may coordinate the different enzymatic activities of the complex. Further structural characterization is required to understand how the RT and Cas1 achieve CRISPR sequence integration from RNA substrates as well as Cas6's role in this process. How the RNA substrate is reverse transcribed, whether ssRNA ligation represents an actual intermediate step in vivo and how this intermediate is resolved into a full-site integration product are questions that remain to be addressed. It is also not known whether RNA spacer acquisition confers immunity against RNA bacteriophages and other foreign RNA elements.

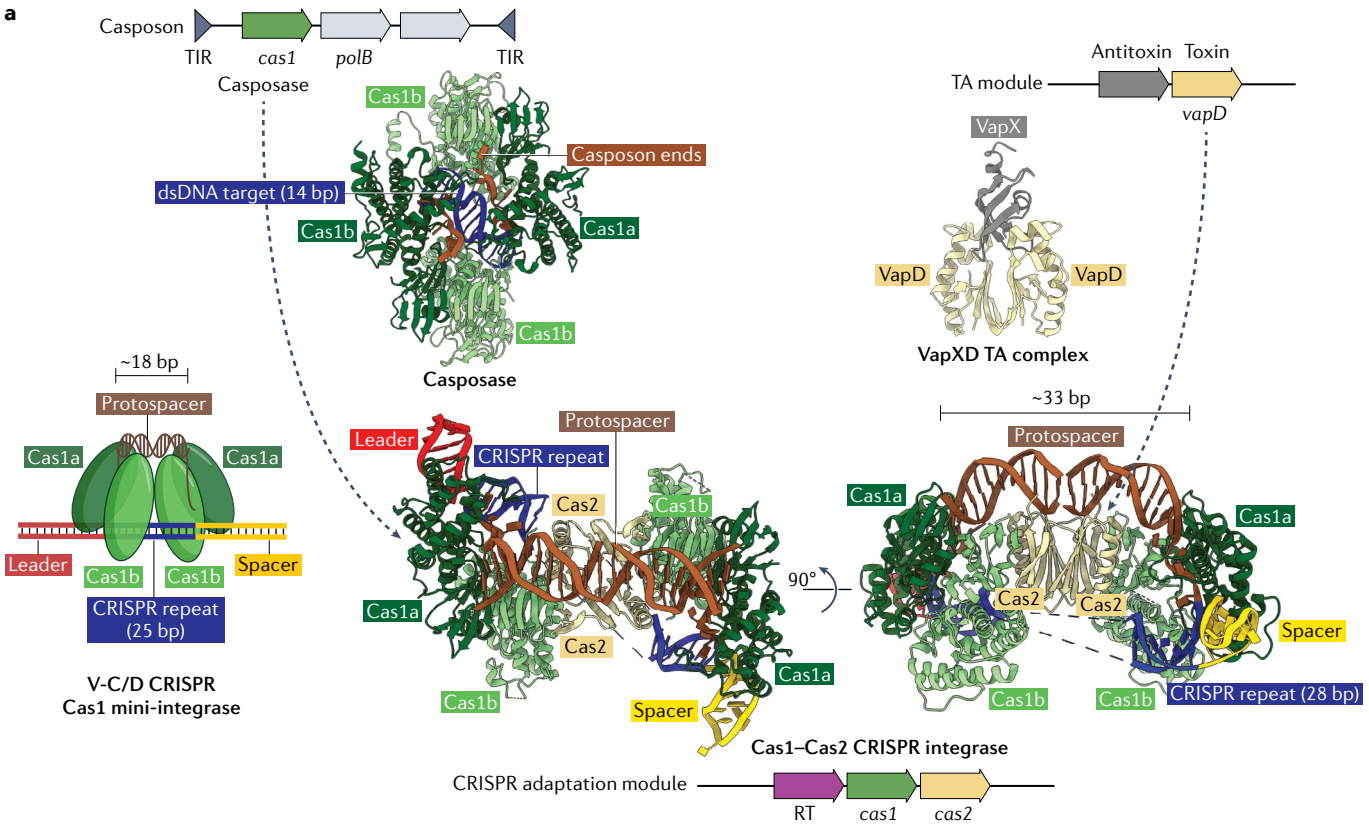
CRISPR integration substrate biogenesis, selection, processing and orientation. Whereas the CRISPR protospacer integration mechanism has been well characterized biochemically and structurally[32,49] (Supplementary Table 1), the processes upstream of integration, including substrate biogenesis and orientation, are less well understood. In the current model for spacer acquisition, RecBCD or AddAB generates DNA degradation products from foreign genetic material that are the precursors for CRISPR integration[50,51]. The molecular details of CRISPR substrate biogenesis remain unknown, although recent studies have elucidated how substrate precursors are selected, processed to the correct length and integrated in the correct orientation with respect to the PAM position to yield functional crRNA generation for downstream CRISPR–Cas immunity processes[19,34,52–56].

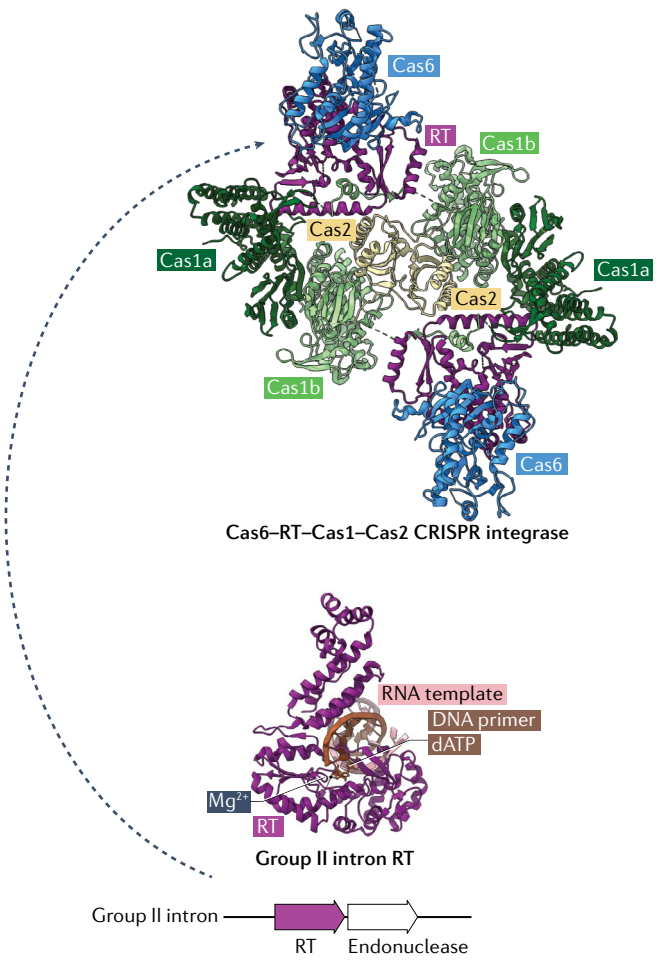
Cas1–Cas2 binds to suitable precursor DNA substrates, which are then processed into mature integration-competent molecules. The *Escherichia coli* Cas1–Cas2, which has a PAM-binding pocket formed by the groove of the C-terminal domain of the catalytic Cas1a and the C-terminal tail of the non-catalytic Cas1b, preferentially selects partially duplexed DNA substrates with long 3′ single-stranded overhangs that contain a PAM sequence in the correct location flanking the eventually integrated spacer[19,52] (FIG. 4a,b). Cas1–Cas2 has exquisite control over the length of the duplex region: the *E. coli* integrase caps the ends of a 23-bp duplex with tyrosine residues[19,20]. This capping is similarly observed in other Cas1–Cas2 complexes, such as the *Enterococcus faecalis* and *Synechocystis* sp. PCC6803 Cas1–Cas2, which use histidine and aspartate residues, respectively, to cap a 22-bp duplex[21,57]. The 3′ single-stranded overhangs are threaded into the Cas1 active site channels[19–21]. Cas1–Cas2 likely predetermines the length of the eventual spacer by protecting the correctly sized region of the substrate from being processed. In certain systems, Cas2 is naturally fused to a DnaQ-like exonuclease domain, which processes long 3′ overhangs of precursor prespacers to the correct length for integration[58]. For the *E. coli* system, which lacks such a natural fusion, DnaQ exonucleases, including DNA polymerase III and ExoT, have been shown to catalyse DNA substrate processing when provided in concert with Cas1–Cas2 (REFS[52,53]). Further structural characterization is required to understand the molecular details of the interaction between these processing enzymes and the Cas1–Cas2 integrase.

In certain type I, type II and type V CRISPR–Cas systems, the Cas4 nuclease works together with Cas1–Cas2 for spacer acquisition and assists in PAM selection and substrate maturation[34,54,55,59]. In these systems, Cas4, rather than Cas1, contains the PAM-binding pocket that sequesters the PAM-containing 3′ overhang[34] (FIG. 4a,b). For the *Geobacter sulfurreducens* system, only the presence of a short DNA molecule containing a PAM triggers assembly of a functional integrase complex consisting of the natural Cas4–Cas1 fusion and its associated Cas2, suggesting that Cas4 acts as a gatekeeper in selecting suitable substrate precursors. Cas4 further functions as an endonuclease that catalyses precise cleavage of PAM sequences before the substrate becomes fully integrated[55,59]. This step is crucial for preventing self-targeting of newly acquired CRISPR spacer sequences to the host cell CRISPR array when the immune system is active.

For both the *E. coli* Cas1–Cas2 and various Cas4–Cas1–Cas2 integrases, PAM binding establishes the directionality of integration. Sequestration of the PAM prevents its trimming by processing nucleases and temporarily blocks that side of the substrate from integration (FIG. 4a). Meanwhile, the non-PAM 3′ side of the DNA is trimmed efficiently and freed for integration into the leader-proximal end of the repeat, the preferred first integration site for both systems[34,52]. In the *E. coli* system, half-site integration enables further trimming of the PAM[52], although it is unclear whether this is done by processing enzymes or by Cas1. In the recently characterized *G. sulfurreducens* system, half-site integration activates Cas4 to cleave the PAM[34]. The structure of *G. sulfurreducens* Cas4–Cas1–Cas2 suggests that a suboptimally placed catalytic lysine residue in the conserved D-E-K motif in the Cas4 active site initially inhibits PAM cleavage, and a conformational change resulting from half-site integration can reposition the lysine to activate endonucleolytic PAM cleavage (FIG. 4c). In both *E. coli* and *G. sulfurreducens*, PAM processing frees that DNA strand for full integration of the prespacer. These studies show that conformational dynamics of PAM sequestration and delayed trimming is a recurring theme in the tightly regulated stepwise processing and directional integration of prespacers. Whether systems that lack PAMs coordinate spacer orientation remains to be determined.

In subtype II-A CRISPR–Cas systems, including those found in *Streptococcus pyogenes* and *Streptococcus thermophilus*, which do not encode Cas4, Cas9 is responsible for PAM recognition for spacer acquisition[60,61]. In these systems, all the genes encoded in the CRISPR locus (tracrRNA, *cas9*, *cas1*, *cas2* and *csn2*) are required for spacer acquisition[60]. Although Cas9's PAM-binding motif is not necessary to acquire new spacers, it is required to select functional PAM-adjacent spacers that would provide immunity[60]. Cas9's HNH domain appears





b

Adaptation modules	Type or subtype
Cas1–Cas2 integrase: *cas1* *cas2*	I, II, III, V, VI
Cas1 mini-integrase: *cas1*	V-C/D
Association with RTs: RT *cas1* *cas2*; RT *cas1* *cas2*; *cas6* RT *cas1* *cas2*	III, VI
Association with prespacer processing enzymes: *cas1* *cas2* DEDDh	I-E
	I, II, V
cas4 *cas1* *cas2*; *cas4* *cas1* *cas2*	
Association with CRISPR interference proteins: *cas1* *cas2* *cas3*	I-F
cas9 *cas1* *cas2* *csn2*	II

◀ Fig. 3 | **Evolution and diversity of the CRISPR integrase architecture. a** | Contributions of casposons, toxin–antitoxin (TA) modules and group II introns to the origins and evolution of the CRISPR adaptation module, and depiction of homologous structures. Casposase bound to integration product mimic (Protein Data Bank (PDB) ID 6OPM)[18]. DNA substrate is colour-coded: double-stranded DNA (dsDNA) target, dark blue; casposon ends, brown (top left). VapXD TA complex (PDB ID 6ZN8)[40] (top right). Subtype V-C or V-D CRISPR Cas1 mini-integrase cartoon (middle left), Cas1–Cas2 CRISPR integrase bound to full-site integration product mimic (PDB ID 5VVK)[22] (middle right) and Cas6–reverse transcriptase (RT)–Cas1–Cas2 CRISPR integrase (PDB ID 7KFU)[47] (bottom). Two of the four Cas6–RT domains are missing. DNA substrates are colour-coded: leader, red; CRISPR repeat, dark blue; spacer, yellow; protospacer, brown. Structure of group II intron RT bound to RNA template–DNA primer and deoxyadenosine triphosphate (dATP) (PDB ID 6AR3)[48] (bottom left). Nucleic acids are colour-coded: RNA, lavender; DNA brown. Protein structures are coloured by domain: catalytic Cas1a monomer, dark green; non-catalytic Cas1b monomer, light green; Cas2/VapD, light yellow; VapX, grey; Cas6, light blue; RT, magenta. Protospacer lengths are indicated for the subtype V-C/D Cas1 integrase and Cas1–Cas2 integrase. **b** | Diversity of domain organizations of CRISPR adaptation module. TIR, terminal inverted repeat sequence.

to have exonuclease activity independent of a guide RNA sequence that can trim DNA precursors to the correct size for integration[62]. Other host nucleases can also carry out this function since Cas9's nuclease activity is not required for spacer acquisition in vivo[60]. Cas9 has been shown to form a complex with Cas1–Cas2 and the auxiliary protein Csn2 (REFS[33,60,62]); however, further structural characterization is necessary to establish how these domains coordinate DNA substrate selection and integration.

In addition to naive adaptation, resulting from infection of previously unencountered bacteriophages or viruses, some CRISPR–Cas systems have evolved another pathway for spacer acquisition, known as primed adaptation, to efficiently adapt to previously encountered invading DNA that has acquired escape mutations. The current understanding of primed adaptation was reviewed recently[49]. In type I systems, this involves the formation of a primed acquisition complex, consisting of Cas1–Cas2, Cascade and Cas3, that translocates along DNA and hands over Cas3 cleavage products to Cas1–Cas2 for integration[26,56,63,64]. Although fluorescence and chromatin immunoprecipitation experiments have provided insight into the molecular mechanisms of type I priming[56,64,65], the nature of Cascade's conformational changes during priming, primed acquisition complex assembly and substrate handover from Cas3 to the Cas1–Cas2 integrase is not yet known. In type II systems, priming is directly correlated with target cleavage by Cas9 (REFS[66,67]). Structural and single-molecule data will also be required to understand substrate handover from Cas9 to the integrase in type II priming and whether it involves the Cas9–Csn2–Cas1–Cas2 supercomplex[33].

The distinctions between these CRISPR sequence integration systems demonstrate that CRISPR adaptation has evolved diverse mechanisms for PAM recognition and DNA substrate generation. This knowledge may help to advance the development of CRISPR adaptation-based molecular recording tools, which have been used for in vivo data storage in bacteria[68–71]. A greater understanding of the elements required for integration substrate selection and processing may be the key to increasing the efficiency of these tools and expanding their use in other microorganisms and mammalian cells to study horizontal gene transfer and lineage tracing.

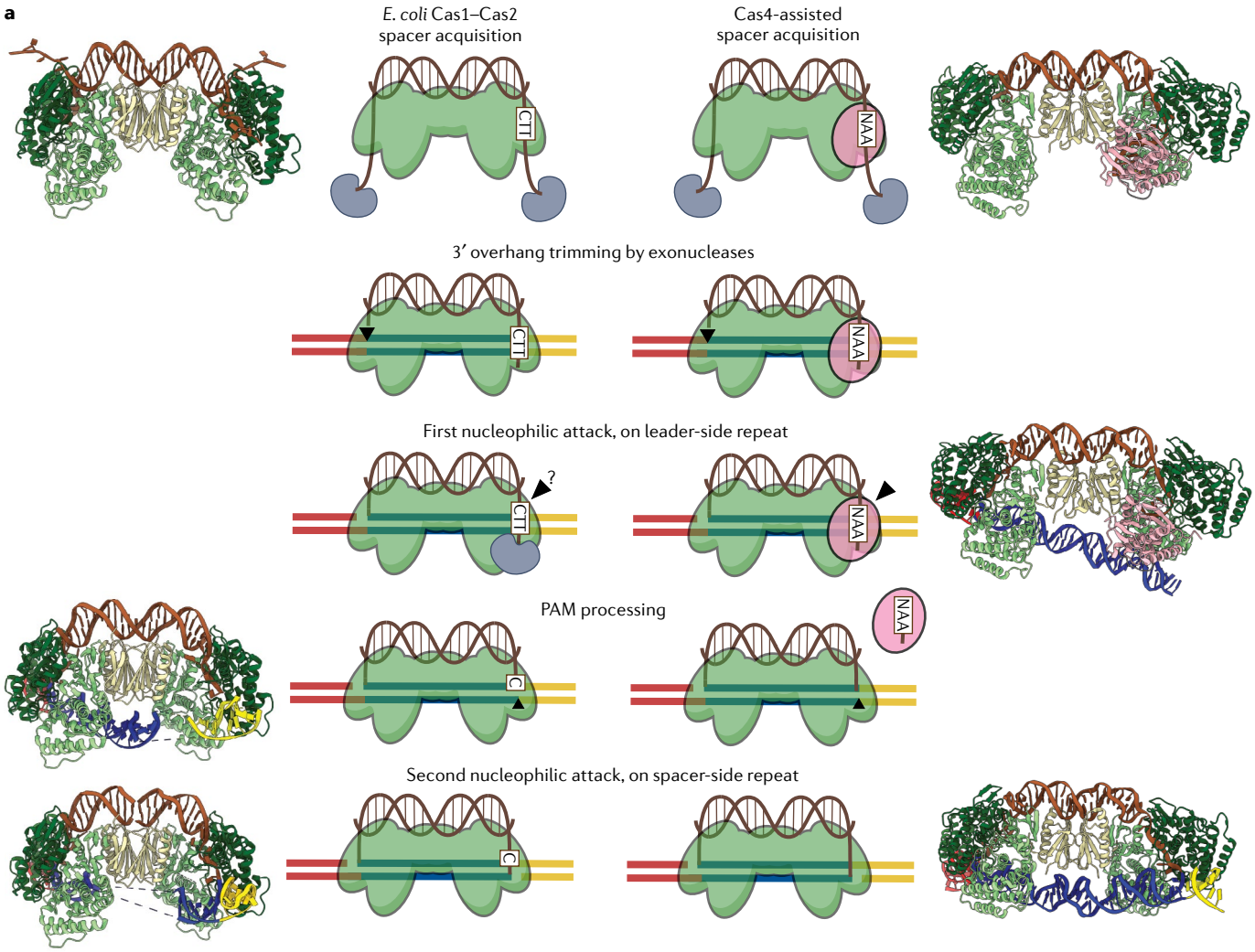
DNA-targeting class 2 systems

The DNA-targeting class 2 CRISPR–Cas9 system provided the first Cas effector that was harnessed for precision genome editing. Class 2 enzymes, including Cas9 (REFS[72–78]) and Cas12 (REFS[79–83]), are now used extensively for genome engineering across a wide range of species and have been co-opted as programmable DNA-targeting modules for biotechnological applications[2,84].

Cas9 and Cas12 rely on similar principles to recognize and cut DNA at an RNA-guide complementary DNA sequence (reviewed in REFS[85,86]) (FIG. 5a). These ribonucleoprotein (RNP) effectors become nuclease active only upon recognition of a cognate target DNA sequence. To identify the target, the RNP recognizes the PAM to initiate an ATP-independent DNA unwinding process that allows pairing of the target DNA strand (TS) to the RNA guide (reviewed in REF.[8]). During the RNA–DNA hybridization process, the 'non-target' DNA strand (NTS) is unpaired from the TS, and the Mg^{2+}-dependent endonuclease cuts both DNA strands individually, using either two separate active sites (HNH and RuvC in Cas9) or a single active site (RuvC in Cas12).

DNA-targeting mechanism of Cas9. Type II CRISPR Cas9 proteins are multidomain enzymes that range in size from 700 amino acids (subtype II-D) to more than 1,700 amino acids (subtype II-C)[11,87]. The functional RNP comprises Cas9 complexed with a dual-guide RNA formed by the crRNA hybridized to a tracrRNA scaffold or bound to an engineered single-guide RNA fusion of crRNA and tracrRNA[72] (FIG. 5b). Cas9's architecture resembles a bilobed structure composed of a crRNA–TS duplex recognition lobe (Rec lobe) and nuclease lobe (Nuc lobe)[88,89], which rearrange themselves relative to each other for target binding and cutting[85]. The lobes are further divided into several subdomains: Rec1, Rec2 and Rec3 of the Rec lobe; and RuvC, HNH and wedge–PAM-interacting domains of the Nuc lobe (Supplementary Table 2). In the absence of a guide RNA, apo-Cas9 resides in an open conformation that transitions into a closed conformation upon guide recruitment, which stabilizes the flexible PAM-interacting domain for PAM identification[88–91] (FIG. 5b). Upon target DNA binding, the RuvC and HNH nuclease domains shift their relative conformations in concert with the Rec lobe to enable nuclease activation[92–98], producing a blunt-ended DNA cut[72] (FIG. 5b).

Whereas nuclease activation and cutting have been studied in great detail[93–95,99–104], the process by which Cas9 interrogates each candidate sequence to identify a cognate target sequence has only recently been addressed. To locate candidate targets for sequence interrogation, Cas9 identifies PAM sequences during facilitated 1D diffusion along DNA or by random 3D collision[105,106]. Disulfide crosslinking of an N^4-cystamine-modified DNA to a cysteine-functionalized *S. pyogenes* Cas9 (T1337C) allowed trapping of the transient interrogation state and revealed the mechanism by which Cas9 'reads'



DNA[107] (FIG. 5c). Surprisingly, the cryogenic electron microscopy-based study revealed that the RNP associates with the PAM in an open bilobed conformation[107], previously only observed in the apo state[88] (FIG. 5b,c). To flip the PAM-adjacent bases for DNA interrogation, Cas9 bends and twists the DNA to locally melt the dsDNA, accompanied by a conformational shift of the bilobed structure to a closed conformation[107] (FIG. 5c). Following initial DNA melting, the TS is gradually paired to the guide[108], producing a 'linear' crRNA–DNA hybrid intermediate that transits to a 'kinked' duplex conformation upon complete target binding[96,97]. The kinked

◀ Fig. 4 | **Mechanisms of CRISPR prespacer biogenesis and orientation. a** | Prespacer processing and directional integration by *Escherichia coli* Cas1–Cas2 and Cas4-assisted CRISPR adaptation systems. In the schematic cartoons, the DNA prespacer (brown) is bound to Cas1–Cas2 (green) and the 3′ ends are trimmed by exonucleases (grey). The protospacer adjacent motif (PAM; shown in brown boxes) for each system is initially protected from trimming, and the non-PAM 3′ end is integrated into the leader side of the CRISPR array (leader, red; repeat, dark blue; spacer, yellow). PAM processing is performed by exonucleases or Cas1 in the *E. coli* system and by Cas4 (pink) in Cas4-assisted acquisition, and the processing sites are indicated by arrowheads. After PAM processing, the 3′ end is freed for spacer-side integration. The associated structures of *E. coli* Cas1–Cas2 (left) and *Geobacter sulfurreducens* Cas4–Cas1–Cas2 (right) at different stages of spacer acquisition are shown next to the cartoons. *E. coli* Cas1–Cas2 structures from top to bottom: Cas1–Cas2 bound to dual-PAM prespacer DNA (Protein Data Bank (PDB) ID 5DQZ)[19]; Cas1–Cas2 bound to half-site integration product mimic (PDB ID 5VVJ)[22]; Cas1–Cas2 bound to full-site integration product mimic (PDB ID 5VVK)[22]. As a point of clarification, the first *E. coli* structure uses a dual-PAM prespacer DNA as a substrate. It is expected that the natural substrate in cells only has a single PAM, shown in the cartoon model. If the cell does encounter a dual-PAM prespacer DNA, it is unclear how PAM protection on both sides affects PAM processing and integration. *G. sulfurreducens* Cas4–Cas1–Cas2 structures from top to bottom: Cas4–Cas1–Cas2 bound to single-PAM prespacer DNA (PDB ID 7MI5)[34]; Cas4–Cas1–Cas2 bound to half-site integration product mimic with Cas4 still engaged in PAM recognition (PDB ID 7MIB)[34]; Cas4–Cas1–Cas2 bound to full-site integration product mimic (PDB ID 7MI9)[34]. In the top two *G. sulfurreducens* Cas4–Cas1–Cas2 structures, the three non-PAM-interacting Cas4 proteins are missing. In the Cas4–Cas1–Cas2 bound to the full-site integration product mimic structure, all four Cas4 proteins are missing. **b** | Close-up of Cas1 PAM interaction (left) in *E. coli* Cas1–Cas2 (PDB ID 5DQZ)[19] and Cas4 PAM interaction (right) in *G. sulfurreducens* Cas4–Cas1–Cas2 (PDB ID 7MI4)[34]. PAM-interacting residues and sequence-specific hydrogen-bonding interactions are depicted. The arrow depicts the PAM-processing cleavage site. **c** | Close-up of Cas4 active site, with the arrow depicting the PAM-processing cleavage site. In the integrase structures, Cas4 is shown in pink, and the other domains follow the colour coding from FIG. 3. The DNA substrates are colour-coded: leader, red; repeat, dark blue; spacer, yellow; prespacer, brown.

conformation in turn facilitates docking of the HNH domain to the crRNA–TS hybrid as observed in recent structures, further permitting NTS binding within the RuvC for DNA cutting[94,95,97,98]. In the presence of extensive crRNA–TS mismatches, the kinked conformation cannot be assumed, thus resulting in inhibition of off-target DNA cutting[97]. Although the structure-guided design of high-fidelity Cas9 variants has yielded more accurate genome editing enzymes[97,99,109,110], alteration of the electrostatics governing the speed of DNA–protein and interdomain interactions of Cas9 during the early interrogation stage might enable the engineering of faster genome editing tools.

DNA-targeting mechanism of Cas12. Type V CRISPR Cas12 proteins are multidomain enzymes that range in size from 400–800 amino acids (subtype V-F; Cas14) up to 1,100–1,500 amino acids (subtype V-B; Cas12b). In contrast to Cas9, Cas12 proteins are highly diverse and are classified into more than ten subtypes (subtypes V-A to V-K, and several of subtype V-U)[11], which differ in their RNP biogenesis pathways, RNP composition and in some instances in their nucleic acid target preference (Supplementary Table 3). Functional RNPs comprise a Cas12 protein either bound to one crRNA guide (for example, subtypes V-A[79,111,112] and V-J[83,113,114]) or bound to a crRNA and tracrRNA hybrid (for example, subtypes V-B[5,115] and V-E[35,81,116]) or a crRNA and short-complementarity untranslated RNA (scoutRNA) hybrid in subtype V-C or V-D[5,117–119]. In some cases, these enzymes form homodimeric complexes on a crRNA and tracrRNA scaffold (subtype V-F)[120,121]. The structural and functional diversity of Cas12 enzymes has defied a simple naming system (Supplementary Table 3).

In general, Cas12 protein architectures possess a Cas9-analogous bilobed architecture[111] (FIG. 6a), composed of Rec and Nuc lobes that are characteristic of even the most compact Cas12 effectors, such as Cas14 (also known as Cas12f; subtype V-F)[121,122], Cas12g (subtype V-G)[123] and CasΦ (also known as Cas12j; subtype V-J)[113] (FIG. 6a). Structural studies of diverse Cas12 proteins revealed similarities beyond the bilobed architecture (Supplementary Table 3). In particular, a crRNA oligonucleotide-binding domain and the RuvC domain form a flexible platform from which the Rec lobe domains (Rec1 and Rec2) and DNA-loading 'nuclease' (Nuc) or zinc-ribbon domains emanate (FIG. 6b). In addition to this general architecture, other small domains are sometimes fused or inserted to aid in PAM identification (PAM-interacting domain) and NTS binding (NTS-binding domain) or to guide recruitment via a zinc-finger (zinc-finger domain) (Supplementary Table 3).

Following from their shared architecture, Cas12 proteins use similar mechanisms to bind and cut DNA as exemplified by Cas12a. In the absence of a guide RNA, Cas12a assumes a flexible 'open' conformation[112], which upon crRNA binding transitions to a 'closed' conformation[111,112,124] that is poised for PAM recognition[124,125]. In the closed conformation, the Rec domains structurally occlude the RuvC active site for nuclease repression[112,124]. Upon PAM-dependent unwinding of a dsDNA target and hybridization of the TS to the crRNA, the Rec domains rearrange themselves to accommodate the heteroduplex[124,126] (FIG. 6b). This rearrangement coincides with opening and activation of the RuvC active site to sequentially cleave first the single-stranded NTS and then the TS[124,127,128], producing a 5′-overhang staggered cut[79]. However, whether the sequential cutting mechanism applies to other type V enzymes remains to be demonstrated. Recent structural data revealed the ssDNA substrate bound to two magnesium cofactors within the RuvC active site of Cas12i and CasΦ/Cas12j, detailing the two metal ion catalysis mechanism of Cas12 enzymes[113,129]. After *cis* DNA cutting, the RuvC domain remains activated for nonspecific ssDNA 'shredding' in *trans*[130,131].

Structural information on the DNA interrogation process is not yet available, but opening of the duplex downstream of the PAM might involve DNA bending[132]. This mechanism is supported by the observation that distorted DNA may expose ssDNA segments, which could be recognized by Cas12 (REF.[133]). Analogously to Cas9, 1D diffusion along DNA may aid in PAM localization beyond 3D collision[134]. Notably, the PAM-interacting domain is often found associated with the unwound DNA, and might therefore be involved in DNA unwinding. However, whether these domains and other adjacent elements actively assist in DNA unwinding, or only associate with the unwound DNA for stabilization, is not clear. Producing a solid understanding of the DNA interrogation mechanism will be crucial to further develop Cas12-based genome editing tools.

A potential role of Cas9/Cas12-ancestral proteins in transposon homing. Cas9 and Cas12 are believed to have evolved from insertion sequence IS*200*/IS*605* transposon-associated nucleases of unknown functionality[135,136]. Recent bioinformatic analysis and experimental data show that the Cas9-ancestral IscB and the distantly related Cas12-ancestral TnpB nucleases are mechanistically akin to Cas9 and Cas12, respectively[87,137]. Similarly to the CRISPR-associated Cas proteins, IscB and TnpB both use a guide RNA to bind and cleave guide-complementary DNA in a target adjacent motif-dependent manner[87,137]. Rfam (RNA families database) searches to identify potential homologues of the IscB-associated guide suggested that the RNA-guide structure partially matches that of HEARO RNA[87], a non-coding RNA that was identified bioinformatically to be an HNH endonuclease-associated RNA and open reading frame[138]. Notably, HNH endonucleases are sometimes used by transposons as homing endonucleases[139,140]. On the basis of the RNA-guided DNA cleavage activity of TnpB, which might guide TnpB to the site of previously excised IS*200*/IS*605* elements, a process analogous to group I intron homing was hypothesized[137]. Whether TnpB and potentially IscB indeed perform RNA-guided transposon homing remains to be demonstrated.

Studies of IscB and TnpB further demonstrated that these RNA-guided nucleases can induce genome

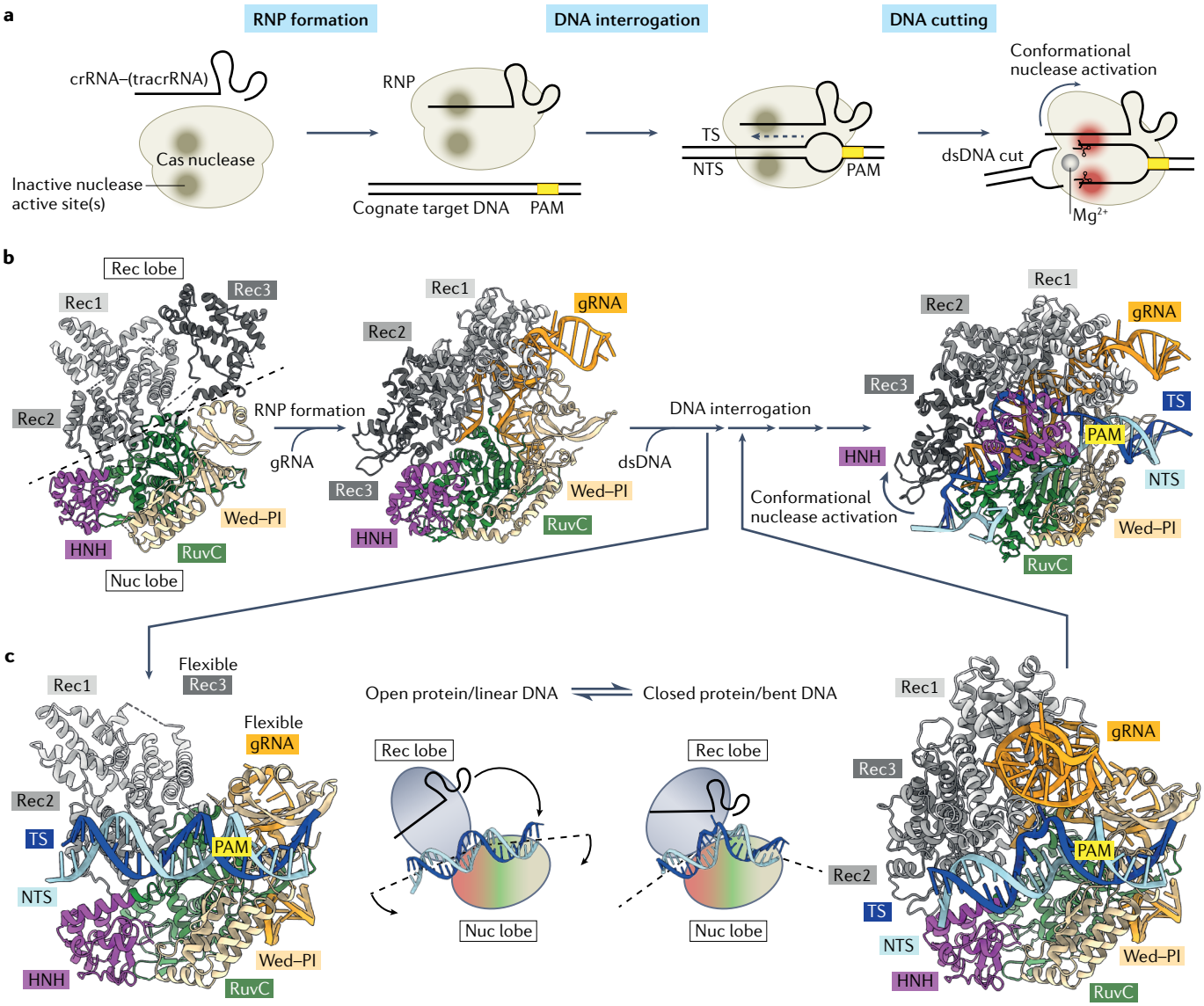


Fig. 5 | **Mechanism of Cas9. a** | General mechanism of Cas9-mediated and Cas12-mediated DNA interference. **b** | Architecture and DNA interference pathway of *Streptococcus pyogenes* Cas9. Structure of apo-Cas9 (Protein Data Bank (PDB) ID 4CMP)[88] (left). Structure of Cas9 bound to guide RNA (gRNA) (PDB ID 4ZT0)[90] (middle). Structure of Cas9 bound to gRNA and cut double-stranded DNA (dsDNA) (PDB ID 7S4X)[91] (right). **c** | DNA interrogation mechanism of Cas9. Structure of *S. pyogenes* Cas9 in the open conformation bound to a linear DNA[107] (left). *S. pyogenes* Cas9 in the closed conformation bound to a bent DNA, exposing the flipped target strand (TS) bases for DNA interrogation[107] (right). crRNA, CRISPR RNA; NTS, non-target stand; Nuc lobe, nuclease lobe; PAM, protospacer adjacent motif; Rec lobe, recognition lobe; RNP, ribonucleoprotein; tracrRNA, *trans*-activating CRISPR RNA; Wed–PI, wedge–protospacer adjacent motif-interacting domain. Coordinates were kindly provided by J. C. Cofsky prior to release of PDB IDs 7S3H and 7S36.

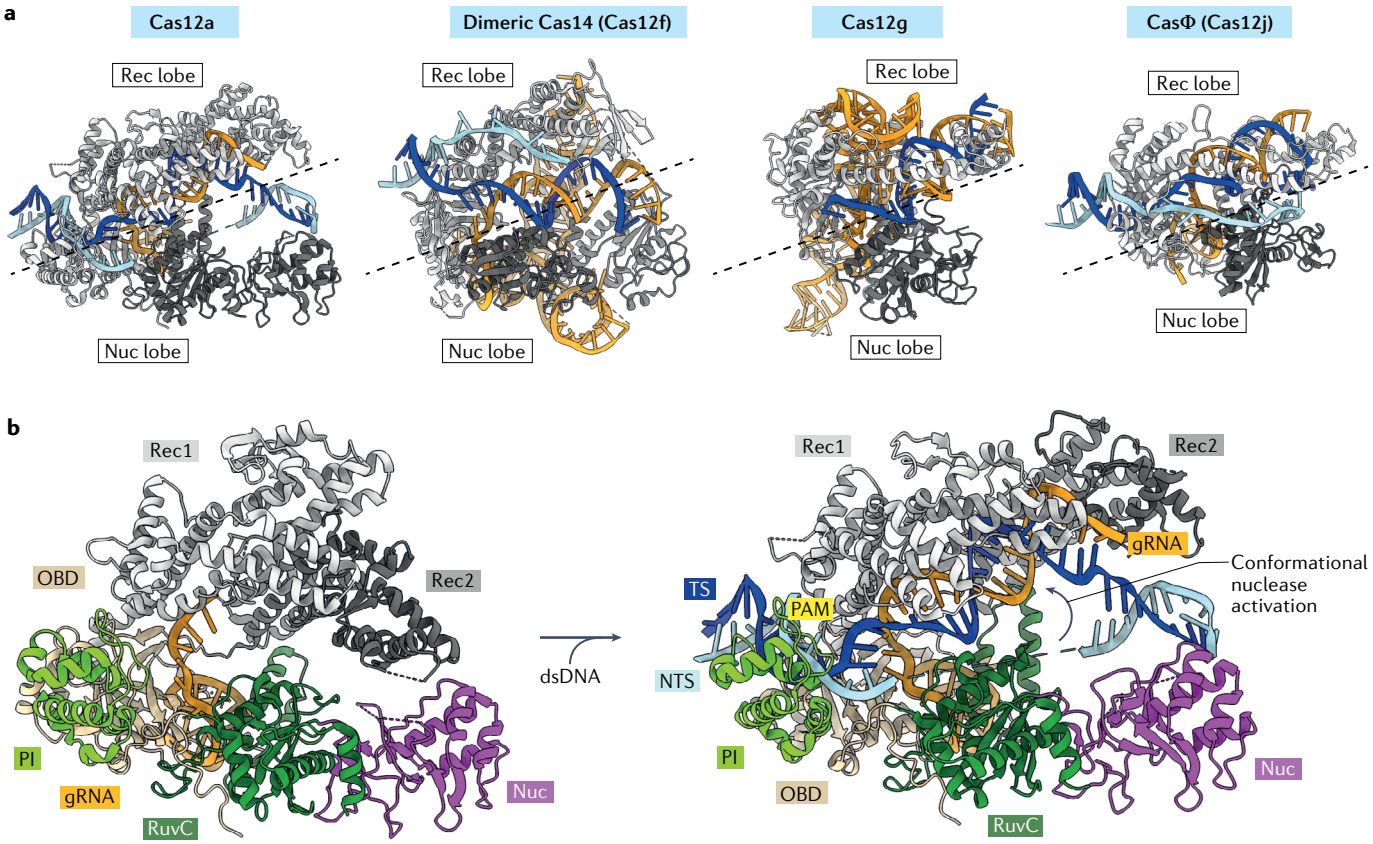


Fig. 6 | **Mechanism of Cas12. a** | Overview of Cas12 enzymes (light grey and dark grey) bound to diverse target nucleic acid (light blue and dark blue) and guide RNA (gRNA, orange): Cas12a (Protein Data Bank (PDB) ID 5NFV)[124], Cas14 (PDB ID 7C7L)[121], Cas12g (PDB ID 6XMG)[123] and CasΦ (PDB-ID 7LYT)[113] (from left to right). The dashed line indicates the boundary between the recognition lobe (Rec lobe) and the nuclease lobe (Nuc lobe). Of note, Cas12g targets RNA, in contrast to all other Cas12 enzymes. **b** | Architecture and DNA-interference pathway of *Leptospiraceae bacterium* Cas12a. Structure of Cas12a bound to gRNA (PDB ID 5NG6)[124] (left). Structure of Cas12 bound to gRNA and unwound double-stranded DNA (dsDNA) (PDB ID 5NFV)[124] (right). NTS, non-target stand; OBD, oligonucleotide binding domain; PAM, protospacer adjacent motif; PI, protospacer adjacent motif-interacting domain; TS, target strand.

editing in eukaryotic cells, although with markedly low efficacy[87,137]. Structural and biochemical studies will be instrumental in the development of IscB- or TnpB-based genome editing tools, and might reveal further similarities between IscB and Cas9, as well as TnpB and Cas12.

CRISPR-associated transposons

The CRISPR adaptation and interference modules discussed in this Review are examples of how bacteria and archaea have recruited various MGEs to evolve machinery for adaptive immunity. However, an opposing evolutionary process appears to have enabled Tn*7*-like transposons, on multiple independent occasions, to co-opt CRISPR–Cas machinery for transposon mobilization in type I, IV and V CRISPR–Cas systems[141–145]. These CRISPR-associated transposons (CASTs) include the core transposition machinery, TnsB and TnsC (and sometimes TnsA), as well as TniQ (a homologue of *E. coli* TnsD), which is involved in recruiting the transposase to the target site. Structural data have yielded the first insights into the interactions between these transposase complexes and CRISPR effectors, and how the crRNA guides the transposon to its complementary target sites for directional insertion at a precise distance from the PAM[146–152].

Type I CASTs. Type I CAST systems utilize the CRISPR effector–crRNA complex to recruit the accompanying transposase to the target site to direct RNA-guided DNA transposition[141] (FIG. 7a). Recent characterizations of the *Vibrio cholerae* TniQ–Cascade–crRNA complex have revealed key structural elements involved in this functional coupling[146–149]. The global architecture of the *V. cholerae* subtype I-F3 Cascade is similar to that of subtype I-F1 Cascade structures, which have been extensively described[153,154]. However, there are notable structural differences that facilitate specific interactions between Cascade and TniQ in the *V. cholerae* system. One major difference is the interaction between Cas6 and Cas7.1, which forms the interface at which TniQ binds (FIG. 7b). Whereas the ferredoxin-like domain of Cas6 mediates the Cas6–Cas7.1 interaction in the canonical *Pseudomonas aeruginosa* system, the Cas6 thumb domain provides this contact in *V. cholerae*, freeing Cas6's ferredoxin-like domain to interact with one monomer of the TniQ dimer[146–149]. Interestingly, a similar Cas6 rearrangement was previously observed in the minimal subtype I-F2 system[155], which is not TniQ associated. On the opposing side of the TniQ dimer, the second TniQ monomer is recruited to the adjacent

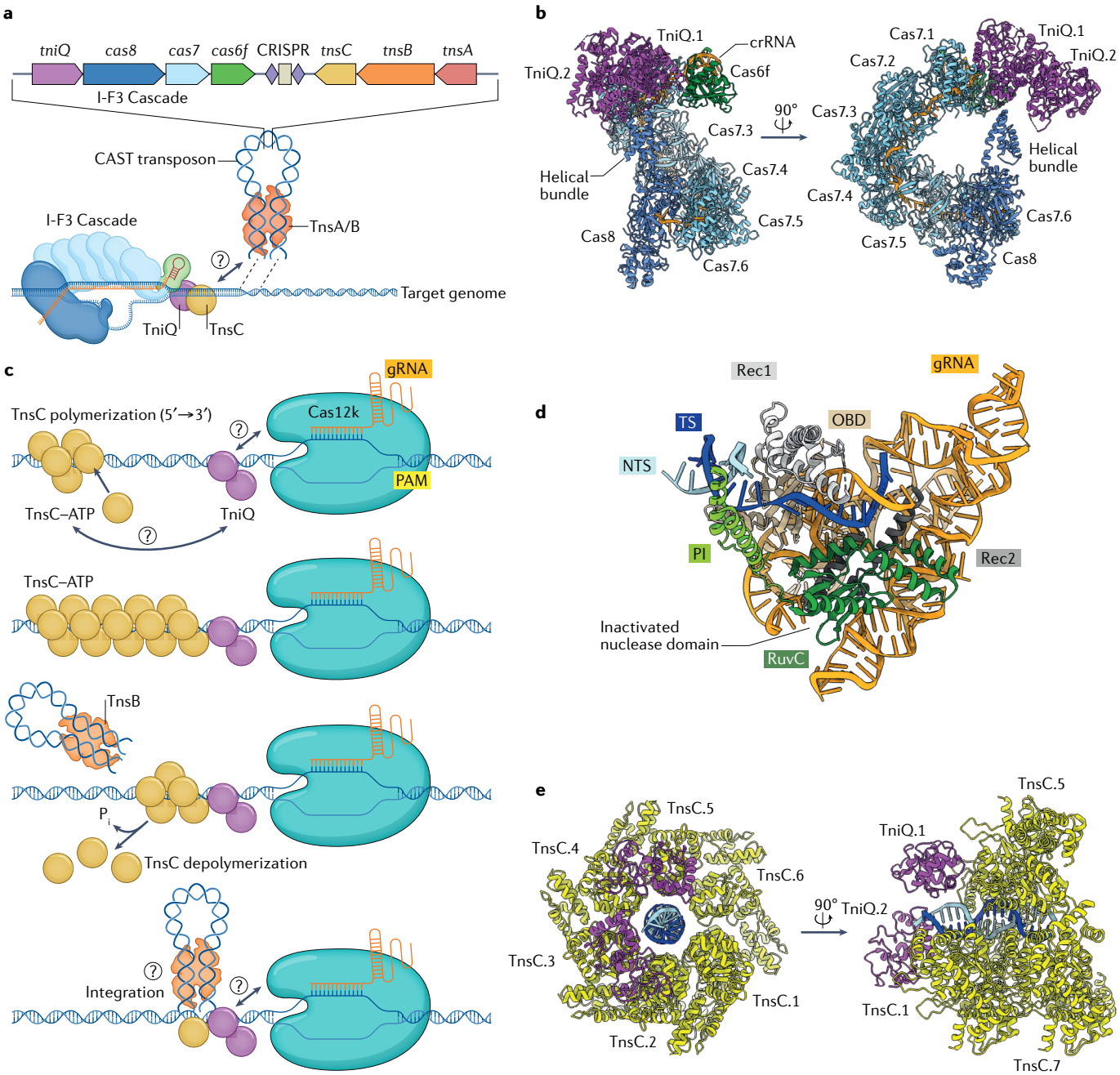


Fig. 7 | **Cas-mediated transposition. a** | Subtype I-F3 CRISPR-associated transposon (CAST) locus and transposition scheme. No structural information is currently available for the TniQ–Cascade and TnsABC transposon interaction (question mark). **b** | Subtype I-F3 CAST Cascade–TniQ structure (Protein Data Bank (PDB) ID 6UVN)[147]. **c** | Subtype V-K CAST transposition scheme. From top to bottom are shown the proposed steps for Cas12k-guided transposon integration via TniQ, the ATPase TnsC and TnsB[150]. It is unclear whether Cas12k interacts with TniQ and whether TniQ caps TnsC polymerization Cas12k proximally as shown[150], or distally, as recently proposed (question marks)[151]. Similarly to the subtype I-F3 CAST, structural data for the integration event are lacking (question mark). **d** | Structure of Cas12k–single-guide RNA–double-stranded DNA complex Cas12k (PDB ID 7N3P). **e** | TniQ–TnsC structures (PDB ID 7N6I)[150]. crRNA, CRISPR RNA; gRNA, guide RNA; NTS, non-target stand; OBD, oligonucleotide binding domain; PAM, protospacer adjacent motif; TS, target strand.

Cas7 (REFS[146–149]). Another important contact occurs between TniQ and the helical bundle domain of Cas8/5 (Cas8/5^{HB} domain) (FIG. 7b). The Cas8/5^{HB} domain typically recruits Cas2–Cas3 for target cleavage in canonical subtype I-F1 systems[153,156]; however, notably, the *cas3* genes are absent along with *cas1* and *cas2* in the type I CAST systems[141,142,144], suggesting that the Cas8/5^{HB} domain has adapted its function for TniQ recruitment. The remainder of the Cas8 subunit is responsible for PAM recognition and triggering R-loop formation, as in canonical Cascade–crRNA complexes[21,157]. These insights indicate key architectural changes necessary to co-opt the CRISPR effector from adaptive immunity to transposon mobilization. TniQ binding to Cascade–crRNA sets the stage for the eventual recruitment of the rest of the transposase TnsABC for transposon

insertion, although further characterization is required to understand that interaction and mechanism.

Type V CASTs. Subtype V-K CAST systems represent another example where transposons have co-opted the CRISPR effector, in this case using a Cas12k with a naturally inactivated nuclease domain for RNA-guided transposon insertion[142,143] (FIG. 7c). Cas12k requires a crRNA guide and a tracrRNA for target recognition, and is responsible for PAM identification and R-loop formation. Cas12k's architecture is similar to that of other type V Cas12 effectors, although it is missing the canonical RuvC active site D-E-D motif, rendering it incapable of target cleavage[151,152] (FIG. 7d). Additionally, in contrast to other Cas12 effectors, Cas12k is missing a Nuc domain and has a longer lid motif in a closed conformation that does not change upon target recognition[152]. The lid motif is well conserved and appears to be essential for RNA-guided DNA transposition activity, although it is unclear what role it plays.

TniQ is also crucial for RNA-guided DNA transposition in type V CAST systems[152]. Structural studies on the *Scytonema hofmanni* TniQ and AAA+ regulator protein TnsC[150] suggest the TniQ–TnsC interaction is important in target search and establishing the orientation of transposon insertion with respect to the PAM site (FIG. 7c,e). Similarly to MuB from bacteriophage Mu, the adenosine triphosphatase TnsC forms polymeric filaments on DNA in the presence of ATP[150,158]. This unidirectional filamentation stops once TnsC encounters TniQ, which selectively interacts with the polymerizing face, as shown in the TniQ–TnsC complex structure (FIG. 7e). This is hypothesized to specify the orientation of transposon insertion with respect to the PAM site.

How TniQ–TnsC and Cas12k communicate in determining the target site is still unclear. Data from immunoprecipitation assays with the *S. hofmanni* CAST system suggest that TniQ does not interact with Cas12k directly[151], hinting at a distinct role for TniQ in type V CAST systems compared with that in type I CAST systems. It is thus unclear whether TniQ caps TnsC polymerization proximal to or distal to Cas12k[150,151] (FIG. 7c). Notably, the C-terminal subtype I-F3 Cascade-interacting domain of TniQ is absent in type V-associated TniQ homologues, suggesting a divergent transposition mechanism that may not rely on direct interactions with the CRISPR effector[151]. Data in microbial systems show that at least one of the type V CAST systems has promiscuous DNA integration activity[159,160], implying some independence from the guide RNA or a requirement for additional factors to ensure target site specificity. It is hypothesized that Cas12k's involvement in DNA unwinding and R-loop formation may facilitate TnsC filament nucleation, which likely helps to recruit the transposase TnsB to catalyse DNA transposition[151]. Structural and functional data show that TnsC polymerization occurs on DNA that is structurally distorted to accommodate the TnsC helical filament and that the TnsC–DNA interactions are crucial for DNA transposition[151]. Immunoprecipitation assays suggest a direct interaction between TnsC and TnsB, but the molecular details remain unknown[151]. Biochemical data show that TnsB triggers disassembly of TnsC filaments by stimulating its ATPase activity. According to the hypothesized mechanistic model, TnsC filament disassembly exposes the target site to produce an integration competent state for TnsB, which is consistent with how systems that have both Mu and prototypic Tn*7* carry out transposition[158,161].

CASTs multitask for transposon mobilization and homing. CAST systems use two propagation mechanisms analogous to the two modes of transposition by CRISPR-less Tn*7* transposons[162,163]. For Tn*7* transposons, transposon propagation is facilitated either by TnsD, which directs transposition to the homing site, or by TnsE, which directs the transposon to MGEs[164]. Whereas the pathways described in the previous sections might be used to facilitate targeting of MGEs for horizontal transfer, structural variation of the transposon machinery by alternative RNA guides or TnsD homologues reprogrammes the machineries for homing[162,163]. In subtype V-K and subtype I-F3 CRISPR–Cas systems, a 'delocalized' and structurally distinct guide RNA, derived from a locus outside the CRISPR array, guides the CAST system to the homing site[162,163]. Distinct from this propagation mode, subtype I-B1 CASTs utilize two different TnsD homologues, one with similarity to subtype I-F3 TniQ for CRISPR–Cas-dependent mobilization and one, more closly related to Tn*7* TnsD, for CRISPR–Cas-independent homing[162]. A similar mechanism was recently described for a variant subtype I-F3 CAST system, which uses two divergent TniQ-family proteins for either mobilization or homing[165]. Structural and biochemical studies elucidating divergent CAST systems and the CAST homing pathway will be essential to understand the alternative mechanisms and might aid in the establishment of CASTs for genome engineering[141,142,159,160,165].

Conclusion and outlook

Fifteen years after the demonstration of the adaptive immunity function of CRISPR–Cas in bacteria[1], structural and analytical studies have uncovered a plethora of specific nucleic acid recognition, insertion and destruction mechanisms. CRISPR–Cas systems use sophisticated machinery to carry out their functions with exquisite control. From a biological perspective, the architectures of CRISPR–Cas enzymatic machineries reflect different drivers of genome evolution and reveal an intricate two-sided evolutionary relationship between CRISPR–Cas systems and MGEs. CRISPR–Cas systems have recruited MGEs to aid in their defence against invaders on multiple occasions in the evolutionary history of CRISPR–Cas, as observed in the origins of the Cas1–Cas2 adaptation module from casposons and TA modules, the recruitment of RTs from retrotransposons and the evolution of class 2 effectors from transposon-encoded IscB and TnpB nucleases. The discovery of CRISPR-associated transposons shows evidence of the reverse evolutionary trend, where transposons have hijacked CRISPR effectors for their own evolutionary success. This relationship between CRISPR–Cas and MGEs fits the 'guns for hire'

paradigm[44], where homologous proteins can be shuttled between different systems and exploited for their structural/functional properties, creating a toolbox for natural genetic engineering.

The toolbox that nature has developed for genetic manipulation along with the ease of programmability has catalysed the translation of CRISPR–Cas immune systems into versatile tools for genome editing and biotechnology[166,167]. The structural and biochemical characterization of naturally occurring CRISPR–Cas systems, and their engineered variants[116,168–175], will continue to be a cornerstone in the development of next-generation gene editing tools and other CRISPR-based technologies, including transcriptional regulation; nucleic acid imaging, detection and diagnostics; and molecular recording.

Recent developments in the computational prediction of protein and RNA structures, using deep learning strategies[176,177], have already begun to transform research in biology and are likely to further accelerate CRISPR–Cas research; however, structure prediction of protein–RNA/DNA complexes is a non-trivial task and will require the development of novel computational approaches that will likely integrate deep learning strategies and the prediction of protein–RNA/DNA binding interfaces[178]. Mapping of interfaces by experimental approaches, such as crosslinking and hydrogen–deuterium exchange mass spectrometry, might further inform these computational methods.

Although the recent computational methods are becoming increasingly capable, they are not suitable to predict the structural dynamics and intermediary stages of CRISPR-mediated nucleic acid binding and cutting. Cryogenic electron microscopy structure determination and molecular dynamics simulations are likely to continue contributing insights into the intricate nucleic acid identification and cutting mechanisms[107,179,180]. Structural insights into the fundamental mechanisms at the heart of CRISPR–Cas systems will enable continued development of these versatile enzymes for precise genome manipulation. In addition, this line of investigation will guide protein engineering strategies to unlock more efficient and safer tools for applications in research, medicine and agriculture.

Published online 13 May 2022

1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
3. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Phil. Trans. R. Soc. B* **374**, 20180087 (2019).
4. Hille, F. et al. The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
5. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60**, 385–397 (2015).
6. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172 (2007).
7. McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
8. Gleditzsch, D. et al. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).
9. Behler, J. & Hess, W. R. Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods* **172**, 12–26 (2020).
10. Marino, N. D., Pinilla-Redondo, R., Csörgő, B. & Bondy-Denomy, J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nat. Methods* **17**, 471–479 (2020).
11. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2019).
12. Jia, N. & Patel, D. J. Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nat. Rev. Mol. Cell Biol.* **22**, 563–579 (2021).
13. Liu, T. Y. & Doudna, J. A. Chemistry of class 1 CRISPR-Cas effectors: binding, editing, and regulation. *J. Biol. Chem.* **295**, 14473–14487 (2020).
14. Molina, R., Sofos, N. & Montoya, G. Structural basis of CRISPR-Cas type III prokaryotic defence systems. *Curr. Opin. Struct. Biol.* **65**, 119–129 (2020).
15. Taylor, H. N. et al. Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. *Front. Microbiol.* **12**, 674522 (2021).
16. O'Connell, M. R. Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR–Cas systems. *J. Mol. Biol.* **431**, 66–87 (2019).
17. Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D. & Koonin, E. V. Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol.* **12**, 36 (2014).
18. Hickman, A. B., Kailasan, S., Genzor, P., Haase, A. D. & Dyda, F. Casposase structure and the mechanistic link between DNA transposition and spacer acquisition by CRISPR-Cas. *eLife* **9**, e50004 (2020).
This work describes the structure of the casposase and its site specificity, providing insight into the evolutionary origins of the Cas1 protein.
19. Wang, J. et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* **163**, 840–853 (2015).
20. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* **527**, 535–538 (2015).
21. Xiao, Y., Ng, S., Nam, K. H. & Ke, A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **550**, 137–141 (2017).
22. Wright, A. V. et al. Structures of the CRISPR genome integration complex. *Science* **357**, 1113–1118 (2017).
23. Hickman, A. B. & Dyda, F. DNA transposition at work. *Chem. Rev.* **116**, 12758–12784 (2016).
24. Béguin, P., Charpin, N., Koonin, E. V., Forterre, P. & Krupovic, M. Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res.* **44**, 10367–10376 (2016).
25. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
26. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
27. Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. & Pul, Ü. Detection and characterization of spacer integration intermediates in type IE CRISPR–Cas system. *Nucleic Acids Res.* **42**, 7884–7893 (2014).
28. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
29. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* **4**, e08716 (2015).
30. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* **23**, 876–883 (2016).
31. Nuñez, J. K. et al. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
32. Sasnauskas, G. & Siksnys, V. CRISPR adaptation from a structural perspective. *Curr. Opin. Struct. Biol.* **65**, 17–25 (2020).
33. Wilkinson, M. et al. Structure of the DNA-bound spacer Capture complex of a type II CRISPR-Cas system. *Mol. Cell* **75**, 90–101.e5 (2019).
34. Hu, C. et al. Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas. *Nature* **598**, 515–520 (2021).
This work provides the mechanism and structural basis for Cas4-assisted PAM processing and describes a model in which PAM sequestration and delayed processing influences the orientation of spacer integration.
35. Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2016).
36. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and nomenclature of CRISPR-Cas systems: where from here? *CRISPR J.* **1**, 325–336 (2018).
37. Wright, A. V. et al. A functional mini-integrase in a two-protein type V-C CRISPR system. *Mol. Cell* **73**, 727–737 (2019).
This work describes a tetrameric CRISPR integrase that may represent the ancestral CRISPR integrase.
38. Santiago-Frangos, A., Buyukyoruk, M., Wiegand, T., Krishna, P. & Wiedenheft, B. Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. *Curr. Biol.* **31**, 3515–3524 (2021).
39. Koonin, E. V. & Makarova, K. S. CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.* **10**, 679–686 (2013).
40. Bertelsen, M. B. et al. Structural basis for toxin inhibition in the VapXD toxin-antitoxin system. *Structure* **29**, 139–150 (2021).
41. Kwon, A.-R. et al. Structural and biochemical characterization of HP0315 from Helicobacter pylori as a VapD protein with an endoribonuclease activity. *Nucleic Acids Res.* **40**, 4216–4228 (2012).
42. Ka, D., Kim, D., Baek, G. & Bae, E. Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. *Biochem. Biophys. Res. Commun.* **451**, 152–157 (2014).
43. Silas, S. et al. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234 (2016).
44. Koonin, E. V. & Makarova, K. S. Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* **9**, 2812–2825 (2017).
45. Silas, S. et al. On the origin of reverse transcriptase-using CRISPR-Cas systems and their hyperdiverse, enigmatic spacer repertoires. *mBio* **8**, e00897-17 (2017).

46. Mohr, G. et al. A reverse transcriptase-Cas1 fusion protein contains a Cas6 domain required for both CRISPR RNA biogenesis and RNA spacer acquisition. *Mol. Cell* **72**, 700–714 (2018).
47. Wang, J. Y. et al. Structural coordination between active sites of a CRISPR reverse transcriptase-integrase complex. *Nat. Commun.* **12**, 2571 (2021). **This work describes the structure of a Cas6–RT–Cas1–Cas2 complex, highlighting interactions between the three domains and the potential functional implications for CRISPR adaptation**.
48. Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. *Mol. Cell* **68**, 926–939 (2017).
49. Nussenzweig, P. M. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas immunity in bacteria. *Annu. Rev. Genet.* **54**, 93–120 (2020).
50. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).
51. Wigley, D. B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat. Rev. Microbiol.* **11**, 9–13 (2013).
52. Kim, S. et al. Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* **579**, 141–145 (2020). **This work describes how the kinetic coordination of prespacer processing and PAM trimming affects the orientation of spacer integration and presents a model for prespacer selection and processing**.
53. Ramachandran, A., Summerville, L., Learn, B. A., DeBell, L. & Bailey, S. Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. *J. Biol. Chem.* **295**, 3403–3414 (2020).
54. Kieper, S. N. et al. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* **22**, 3377–3384 (2018).
55. Lee, H., Dhingra, Y. & Sashital, D. G. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *eLife* **8**, e44248 (2019).
56. Musharova, O. et al. Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase. *Proc. Natl Acad. Sci. USA* **118**, e2021291118 (2021).
57. Wu, C. et al. Mechanisms of spacer acquisition by sequential assembly of the adaptation module in Synechocystis. *Nucleic Acids Res.* **49**, 2973–2984 (2021).
58. Drabavicius, G. et al. DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. *EMBO Rep.* **19**, e45543 (2018).
59. Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell* **70**, 48–59. e5 (2018).
60. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202 (2015).
61. Heler, R. et al. Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. *Mol. Cell* **65**, 168–175 (2017).
62. Jakhanwal, S. et al. A CRISPR-Cas9–integrase complex generates precise DNA fragments for genome integration. *Nucleic Acids Res.* **49**, 3546–3556 (2021).
63. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7**, e35888 (2012).
64. Dillard, K. E. et al. Assembly and translocation of a CRISPR-Cas primed acquisition complex. *Cell* **175**, 934–946.e15 (2018).
65. Xue, C., Whitis, N. R. & Sashital, D. G. Conformational control of Cascade interference and priming activities in CRISPR immunity. *Mol. Cell* **64**, 826–834 (2016).
66. Nicholson, T. J. et al. Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol.* **16**, 566–576 (2019).
67. Nussenzweig, P. M., McGinn, J. & Marraffini, L. A. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell Host Microbe* **26**, 515–526.e6 (2019).
68. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
69. Sheth, R. U., Yim, S. S., Wu, F. L. & Wang, H. H. Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).
70. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
71. Munck, C., Sheth, R. U., Freedberg, D. E. & Wang, H. H. Recording mobile DNA in the gut microbiota using an Escherichia coli CRISPR-Cas spacer acquisition platform. *Nat. Commun.* **11**, 95 (2020).
72. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
73. Gasiunas, G. & Barrangou, R. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
74. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
75. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
76. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
77. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
78. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
79. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
80. Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
81. Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
82. Strecker, J. et al. Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).
83. Pausch, P. et al. CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
84. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
85. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* **9**, e1481 (2018).
86. Stella, S., Alcón, P. & Montoya, G. Class 2 CRISPR–Cas RNA-guided endonucleases: Swiss Army knives of genome editing. *Nat. Struct. Mol. Biol.* **24**, 882–892 (2017).
87. Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
88. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
89. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014). **This study reveals for the first time how Cas9 recognizes DNA**.
90. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
91. 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).
92. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
93. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).
94. Zhang, Y. et al. Catalytic-state structure and engineering of Streptococcus thermophilus Cas9. *Nat. Catal.* **3**, 813–823 (2020).
95. Sun, W. et al. Structures of Neisseria meningitidis Cas9 complexes in catalytically poised and anti-CRISPR-inhibited states. *Mol. Cell* **76**, 938–952.e5 (2019).
96. Pacesa, M. & Jinek, M. Mechanism of R-loop formation and conformational activation of Cas9. Preprint at *bioRxiv* https://doi.org/10.1101/2021.09.16.460614 (2021).
97. Bravo, J. P. K. et al. Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature* **603**, 343–347 (2022). **This study demonstrates how excessive target mismatches inhibit DNA cutting by Cas9 and reveals a most comprehensive structure of Cas9 bound to the DNA cleavage product**.
98. Zhu, X. et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **26**, 679–685 (2019).
99. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
100. Palermo, G. et al. Protospacer adjacent motif-induced allostery activates CRISPR-Cas9. *J. Am. Chem. Soc.* **139**, 16028–16031 (2017).
101. 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).
102. Nierzwicki, L. et al. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *eLife* **10**, e73601 (2021).
103. 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).
104. Belato, H. B. et al. Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *J. Struct. Biol.* **214**, 107814 (2021).
105. Globyte, V., Lee, S. H., Bae, T., Kim, J. & Joo, C. CRISPR / Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* **38**, e99466 (2019).
106. 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).
107. Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E. & Doudna, J. A. CRISPR-Cas9 bends and twists DNA to read its sequence. *Nat. Struct. Mol. Biol.* **29**, 395–402 (2022). **This article for the first time reveals structural insights into how Cas9 opens dsDNA to interrogate target sequences**.
108. Ivanov, I. E. et al. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc. Natl Acad. Sci. USA* **117**, 5853–5860 (2020).
109. Liu, M.-S. et al. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
110. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
111. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016). **This study reveals for the first time how Cas12a recognizes DNA**.
112. Dong, D. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
113. Pausch, P. et al. DNA interference states of the hypercompact CRISPR-CasΦ effector. *Nat. Struct. Mol. Biol.* **28**, 652–661 (2021). **This article describes how a minimal Cas12 enzyme binds dsDNA and catalyses DNA cleavage.**
114. Carabias, A. et al. Structure of the mini-RNA-guided endonuclease CRISPR-Cas12j3. *Nat. Commun.* **12**, 4476 (2021).
115. Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell* **167**, 1814–1828 (2016).
116. Tsuchida, C. A. et al. Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol. Cell* **82**, 1199–1209.e6 (2022).
117. Harrington, L. B. et al. A scoutRNA is required for some type V CRISPR-Cas systems. *Mol. Cell* **79**, 416–424.e5 (2020).
118. Huang, C. J., Adler, B. A. & Doudna, J. A. A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. Preprint at *bioRxiv* https://doi.org/10.1101/2021.12.06.471469 (2021).
119. Kurihara, N. et al. Structure of the type V-C CRISPR-Cas effector enzyme. *Mol. Cell* **82**, 1–13 (2022).
120. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).

121. Takeda, S. N. et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell* **81**, 558–570 (2021).
122. Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).
123. Li, Z., Zhang, H., Xiao, R., Han, R. & Chang, L. Cryo-EM structure of the RNA-guided ribonuclease Cas12g. *Nat. Chem. Biol.* **17**, 387–393 (2021).
124. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221–233 (2017).
125. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* **26**, 901–913 (2016).
126. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).
127. Cofsky, J. C. et al. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *eLife* **9**, e55143 (2020).
128. Stella, S. et al. Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity. *Cell* **175**, 1856–1871 (2018). **This study provides extensive structural and mechanistic insights into the conformational activation of Cas12a**.
129. Huang, X. et al. Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.* **11**, 5241 (2020).
130. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
131. Swarts, D. C. & Jinek, M. Mechanistic insights into the cis- and trans-acting DNase activities of Cas12a. *Mol. Cell* **73**, 589–600 (2019).
132. Jiang, W. et al. CRISPR-Cas12a nucleases bind flexible DNA duplexes without RNA/DNA complementarity. *ACS Omega* **4**, 17140–17147 (2019).
133. Paul, B., Chaubet, L., Verver, D. E. & Montoya, G. Mechanics of CRISPR-Cas12a and engineered variants on λ-DNA. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkab1272 (2021).
134. Losito, M., Smith, Q. M., Newton, M. D., Cuomo, M. E. & Rueda, D. S. Cas12a target search and cleavage on force-stretched DNA. *Phys. Chem. Chem. Phys.* **23**, 26640–26644 (2021).
135. Kapitonov, V. V., Makarova, K. S. & Koonin, E. V. ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs. *J. Bacteriol.* **198**, 797–807 (2015).
136. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
137. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
138. Weinberg, Z., Perreault, J., Meyer, M. M. & Breaker, R. R. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462**, 656–659 (2009).
139. Stoddard, B. L. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* **5**, 7 (2014).
140. Stoddard, B. L. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19**, 7–15 (2011).
141. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
142. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
143. Faure, G. et al. CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
144. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).
145. Rybarski, J. R., Hu, K., Hill, A. M., Wilke, C. O. & Finkelstein, I. J. Metagenomic discovery of CRISPR-associated transposons. *Proc. Natl Acad. Sci. USA* **118**, e2112279118 (2021).
146. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernández, I. S. Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature* **577**, 271–274 (2020). **This work provides mechanistic insights into subtype I-F3 CRISPR transposases by describing the structures of a TniQ–Cascade complex and reveals interactions between TniQ and Cas6 and Cas7.1 within the Cascade complex**.
147. Li, Z., Zhang, H., Xiao, R. & Chang, L. Cryo-EM structure of a type I-F CRISPR RNA guided surveillance complex bound to transposition protein TniQ. *Cell Res.* **30**, 179–181 (2020).
148. Jia, N., Xie, W., de la Cruz, M. J., Eng, E. T. & Patel, D. J. Structure-function insights into the initial step of DNA integration by a CRISPR-Cas-Transposon complex. *Cell Res.* **30**, 182–184 (2020).
149. Wang, B., Xu, W. & Yang, H. Structural basis of a Tn7-like transposase recruitment and DNA loading to CRISPR-Cas surveillance complex. *Cell Res.* **30**, 185–187 (2020).
150. Park, J.-U. et al. Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021). **This work provides mechanistic insights into subtype V-K CRISPR transposases by describing a transposition regulator, TnsC, from a subtype V-K CAST system and its interaction with TniQ and proposing a model for subtype V-K CAST transposition**.
151. Querques, I., Schmitz, M., Oberli, S., Chanez, C. & Jinek, M. Target site selection and remodelling by type V CRISPR-transposon systems. *Nature* **599**, 497–502 (2021). **This work provides mechanistic insights into subtype V-K CRISPR transposases by describing target recognition by Cas12k and the role of the transposition regulator TnsC and proposing an alternative model for subtype V-K CAST transposition**.
152. Xiao, R. et al. Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition. *Mol. Cell* **81**, 4457–4466.e5 (2021).
153. Chowdhury, S. et al. Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47–57.e11 (2017).
154. Guo, T. W. et al. Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex. *Cell* **171**, 414–426.e12 (2017).
155. Pausch, P. et al. Structural variation of type I-F CRISPR RNA guided DNA surveillance. *Mol. Cell* **67**, 622–632.e4 (2017).
156. Rollins, M. F. et al. Structure reveals a mechanism of CRISPR-RNA-guided nuclease recruitment and anti-CRISPR viral mimicry. *Mol. Cell* **74**, 132–142.e5 (2019).
157. Hayes, R. P. et al. Structural basis for promiscuous PAM recognition in type I–E Cascade from E. coli. *Nature* **530**, 499–503 (2016).
158. Greene, E. C. & Mizuuchi, K. Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. *EMBO J.* **21**, 1477–1486 (2002).
159. Vo, P. L. H. et al. CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* **39**, 480–489 (2021).
160. Rubin, B. E. et al. Species- and site-specific genome editing in complex bacterial communities. *Nat. Microbiol.* **7**, 34–47 (2021).
161. Stellwagen, A. E. & Craig, N. L. Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.* **16**, 6823–6834 (1997).
162. Saito, M. et al. Dual modes of CRISPR-associated transposon homing. *Cell* **184**, 2441–2453 (2021).
163. Petassi, M. T., Hsieh, S.-C. & Peters, J. E. Guide RNA categorization enables target site choice in Tn7-CRISPR-Cas transposons. *Cell* **183**, 1757–1771.e18 (2020).
164. Waddell, C. S. & Craig, N. L. Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev.* **2**, 137–149 (1988).
165. Klompe, S. E. et al. Evolutionary and mechanistic diversity of type I-F CRISPR-associated transposons. *Mol. Cell* **82**, 616–628 (2022).
166. Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR-Cas toolbox and gene editing technologies. *Mol. Cell* **82**, 333–347 (2022).
167. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccia, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell* **82**, 348–388 (2022).
168. Lapinaite, A. et al. DNA capture by a CRISPR-Cas9 guided adenine base editor. *Science* **369**, 566–571 (2022).
169. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Mol. Cell* **61**, 886–894 (2016).
170. Chen, W. et al. Molecular basis for the PAM expansion and fidelity enhancement of an evolved Cas9 nuclease. *PLoS Biol.* **17**, e3000496 (2019).
171. Anders, C., Bargsten, K. & Jinek, M. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell* **61**, 895–902 (2016).
172. Guo, M. et al. Structural insights into a high fidelity variant of SpCas9. *Cell Res.* **29**, 183–192 (2019).
173. Nishimasu, H. et al. Structural basis for the altered PAM recognition by engineered CRISPR-Cpf1. *Mol. Cell* **67**, 139–147.e2 (2017).
174. Shams, A. et al. Comprehensive deletion landscape of CRISPR-Cas9 identifies minimal RNA-guided DNA-binding modules. *Nat. Commun.* **12**, 5664 (2021).
175. Donohoue, P. D. et al. Conformational control of Cas9 by CRISPR hybrid RNA-DNA guides mitigates off-target activity in T cells. *Mol. Cell* **81**, 3637–3649.e5 (2021).
176. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
177. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
178. Wei, J., Chen, S., Zong, L., Gao, X. & Li, Y. Protein-RNA interaction prediction with deep learning: structure matters. Preprint at *arXiv* https://arxiv.org/abs/2107.12243 (2021).
179. Nierzwicki, Ł. & Palermo, G. Molecular dynamics to predict cryo-EM: capturing transitions and short-lived conformational states of biomolecules. *Front. Mol. Biosci.* **8**, 641208 (2021).
180. Wang, J. et al. Gaussian accelerated molecular dynamics: principles and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1521 (2021).
181. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* **361**, eaat0839 (2018).

Acknowledgements
The authors thank J. C. Cofsky, K. M. Soczek and G. J. Knott for sharing the coordinates of Cas9 bound to linear and bent double-stranded DNA before publication. J.Y.W. is supported by the US National Science Foundation Graduate Fellowship and was previously supported by the Berkeley Graduate Fellowship. P.P. receives funding from the European Regional Development Fund under grant agreement number 01.2.2-CPVA-V-716-01-0001 with the Central Project Management Agency (CPVA), Lithuania, and from the Research Council of Lithuania (LMTLT) under grant agreement number S-MIP-22-10. This material is based upon work supported by the US National Science Foundation under award number 1817593 and by the Somatic Cell Genome Editing Program of the Common Fund of the US National Institutes of Health under award number U01AI142817-02. J.A.D. is a Howard Hughes Medical Institute investigator. The authors thank G. J. Knott and A. Lapinaite for helpful discussions.

Author contributions
All authors contributed to writing, reviewing and editing the manuscript.

Competing interests
J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, Algen Biotechnologies, Felix Biosciences, The Column Group and Inari. J.A.D. is the Chief Science Advisor of Sixth Street, is on the Board of Directors at Altos, Johnson & Johnson and Tempus, and has research projects sponsored by Biogen, Pfizer, AppleTree Partners and Roche. The Regents of the University of California have patents issued and pending for CRISPR technologies on which P.P. and J.A.D. are named as inventors. J.Y.W. declares no competing interests.

Peer review information
Nature Reviews Microbiology thanks Hong Li, Guillermo Montoya and Dinshaw Patel for their contribution to the peer review of this work.

Supplementary information
The online version contains supplementary material available at https://doi.org/10.1038/s41579-022-00739-4.