CRISPR-Cas
Adapting to change

Jackson, Simon A.; McKenzie, Rebecca E.; Fagerlund, Robert D.; Kieper, Sebastian N.; Fineran, Peter C.; Brouns, Stan J.J.

DOI
10.1126/science.aal5056

Publication date
2017

Document Version
Accepted author manuscript

Published in
Science

Citation (APA)
Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333), 1-9. Article eaal5056. https://doi.org/10.1126/science.aal5056

Title:

CRISPR-Cas: adapting to change

Authors:

Simon A. Jackson[1]†, Rebecca E. McKenzie[2]†, Robert D. Fagerlund[1], Sebastian N. Kieper[2], Peter C. Fineran[1,3]* and Stan J.J. Brouns[2,4*]

Affiliations:

[1]Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[2]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.

[3]Bio-Protection Research Centre, University of Otago, PO Box 56, Dunedin 9054, New Zealand.

[4]Laboratory of Microbiology, Wageningen University, The Netherlands.

†These authors contributed equally

*Co-corresponding authors. E-mail. peter.fineran@otago.ac.nz; Tel. +64 (0)3 479 7735; Fax. +64 (0)3 479 8540 or Email. stanbrouns@gmail.com; Tel: +31-15 278 3920

Abstract:

Bacteria and archaea are engaged in a constant arms race to defend against the ever-present threats of viruses and invasion by mobile genetic elements. The most flexible weapons in the prokaryotic defense arsenal are the CRISPR-Cas adaptive immune systems, which are capable of selective identification and neutralization of foreign elements. CRISPR-Cas systems rely on stored genetic memories to facilitate target recognition. Thus, to keep pace with a changing pool of hostile invaders, the CRISPR memory banks must be regularly updated by the addition of new information, through a process termed adaptation. In this review, we outline the recent advances in our understanding of the molecular mechanisms governing adaptation and highlight the diversity between systems.

One Sentence Summary:

How prokaryotes adapt their CRISPR memory to constantly-evolving invaders

Main Text:

Adaptive immunity in prokaryotes

Bacteria and archaea are constantly threatened by phage infection and invasion by mobile genetic elements (MGEs) through conjugation and transformation. In response, a defense arsenal has evolved, including various 'innate' mechanisms and the CRISPR-Cas adaptive immune systems (*1-3*). CRISPR-Cas systems are widely distributed, present in 50% and 87% of complete bacterial and archaeal genomes, respectively, and are classified into two major classes consisting of 6 types according to their Cas proteins (*4, 5*). CRISPR-Cas systems function as RNA-guided nucleases that provide sequence-specific defense against invading MGEs (*6, 7*). Their repurposing, particularly Cas9, has stimulated a biotechnological revolution in genome editing that has resulted in breakthroughs across many biological fields (*8*). In native hosts, the advantage conferred by CRISPR-Cas systems over innate defenses lies in the ability to update their resistance repertoire in response to infection (termed CRISPR adaptation). Adaptation is achieved by incorporating short DNA fragments from MGEs into CRISPR arrays to form memory units termed spacers, which are subsequently transcribed and processed to CRISPR RNAs (crRNAs) (**Fig. 1)**. Cas proteins associate with crRNAs to form crRNA-effector complexes, which seek and destroy invading MGEs. Thus, adaptation of CRISPR arrays is a crucial process required to ensure persistent CRISPR-Cas defense (*9, 10*).

Adaptation in nature appears widespread, highlighting the dynamic interaction between hosts and invaders (*11-13*). When a prokaryotic community undergoes CRISPR adaptation, individual cells acquire different, and often multiple spacers. This population diversity increases defense by limiting the reproductive success of MGE variants that evade recognition through genetic mutations (escape mutants) (*14*). The CRISPR polymorphisms resulting from adaptation enable differentiation of species subtypes, including economically and clinically relevant isolates, and allow tracking of pathogen outbreaks (*15, 16*).

Typically, new spacers are inserted at one end of the array in a position closest to the promoter driving CRISPR transcription – termed the leader (**Fig. 1**) (*6, 17-19*). This polarization of the CRISPR records provides a chronological account of the battle between phages and bacteria, analyses of which can provide insights into phage-host co-occurrences, evolution and ecology (*20, 21*). Moreover, spacer integration at the leader end enhances defense against recently encountered MGEs, potentially due to elevated crRNA abundance (*22*). However, in some systems, the repeats themselves contain internal promoters, which might make leader-proximal spacer integration less important (*23*). CRISPR arrays typically contain 10-30 spacers, but some species contain arrays with over 500 spacers (*24*). Spacers that may no longer be under evolutionary selection can be lost via recombination between CRISPR repeats (*11, 25*).

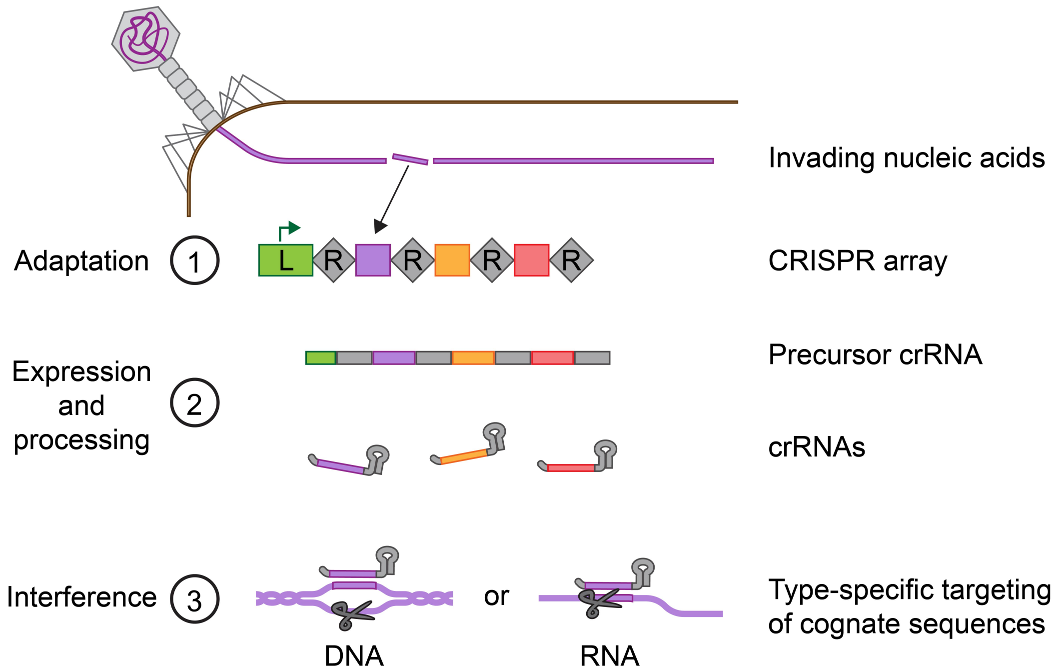

Fig. 1: CRISPR-Cas adaptation and defense. A simplified schematic of CRISPR-Cas defense, which consists of an array of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins encoded by *cas* genes (omitted for clarity). CRISPR-Cas defense consists of three defined stages 1) Adaptation, the creation of memory of prior infections formed via the insertion of small foreign DNA sequences into the leader (L) end of the CRISPR array, where they are stored as spacers (colored squares) between duplicated repeats (R). 2) Expression and CRISPR-RNA (crRNA) biogenesis, the transcription and processing of the array into small guide RNA sequences. 3) Interference, degradation of the target foreign invader by sequence-specific binding and cleavage.

Early bioinformatic studies showed many spacers were of foreign origin, hinting that CRISPR loci would form the memory of an immune system (*15, 26-28*). Subsequent confirmation of this link between spacers and resistance to phage and MGEs was gained experimentally (*6, 7, 29*). Despite the elegance of memory-directed defense, CRISPR adaptation is not without complications. Paradoxically, the spacers required for defense must be added to CRISPRs during exposure to MGEs (*30, 31*). In addition, the inadvertent acquisition of spacers from host DNA must be avoided because this will result in cytotoxic self-targeting – akin to autoimmunity (*32, 33*). Recently, significant progress has been made toward understanding the molecular mechanisms governing how, when and why CRISPR spacers are acquired. Here, we review these studies and highlight the insights they shed on both the function and evolution of CRISPR-Cas systems.

Molecular mechanism of adaptation

At the forefront of adaptation are Cas1 and Cas2 proteins, which form a $Cas1_4$-$Cas2_2$ complex (*34, 35*) (hereafter Cas1-Cas2) – the 'workhorse' of spacer integration (**Fig. 2**). Illustrative of their key roles in spacer integration, the *cas1* and *cas2* genes are associated with nearly all CRISPR-Cas systems (*4*). Cas1-Cas2-mediated spacer integration prefers dsDNA substrates and proceeds via a mechanism resembling retroviral integration (*36, 37*). In addition to Cas1-Cas2, a single repeat, at least part of the leader sequence (*17, 18, 22, 38*), and additional host factors for repair of the insertion sites (e.g. DNA polymerase) are required (*39*). Spacer integration requires three main processes: 1) substrate capture 2) recognition of the CRISPR locus and 3) integration within the array.

Substrate capture

During substrate capture, Cas1-Cas2 is loaded with an integration-compatible pre-spacer, which is thought to be partially duplexed DNA. In the Cas1-Cas2:pre-spacer complex, each single-stranded 3'OH end of the pre-spacer DNA extends into a single active subunit of each Cas1 dimer (*40*) located either side of a central Cas2 dimer (*41, 42*) (**Fig. 2**). The branch points of the splayed DNA are stabilized by a Cas1 wedge, which acts as a molecular ruler to control spacer length. Although it is likely that Cas1-Cas2 rulers exist and measure different spacer sizes in all systems, the mechanism has only been demonstrated in the *Escherichia coli* type I-E system, where two tyrosine residues bookend the core 23 nt dsDNA region (*41, 42*). Details of how pre-spacer substrates are produced from foreign DNA is discussed later.

Recognition of the CRISPR locus

Prior to integration, the substrate-bound Cas1-Cas2 complex must locate the CRISPR leader-repeat sequence. Adaptation complexes of several systems display intrinsic affinity for the leader-repeat region in vitro (*36, 43*), yet this is not always wholly sufficient to provide the specificity observed in vivo. For the type I-E system, leader-repeat recognition is assisted by the integration host factor (IHF) heterodimer, which binds in the leader (*44*). IHF binds DNA in a sequence-specific manner and induces ~120˚ DNA bending, providing a cue to accurately localize Cas1-Cas2 to the leader-repeat junction (*44, 45*). A conserved leader motif upstream of the IHF pivot is proposed to stabilize the Cas1-Cas2-leader-repeat interaction and increase adaptation efficiency, supporting bipartite binding of the adaptation complex to DNA sites either side of bound IHF (*45*).

IHF is absent in many prokaryotes, including archaea and gram-positive bacteria, suggesting other leader-proximal integration mechanisms exist. Indeed, type II-A Cas1-Cas2 from *Streptococcus pyogenes* catalyzed leader-proximal integration in vitro, at a level of precision comparable to the type I-E system with IHF (*43, 44*). Hence, type II-A systems may rely solely on intrinsic sequence specificity for the leader-repeat. A short leader-anchoring site (LAS) adjacent to the first repeat and ≤6 bp of this repeat were essential for adaptation (*22, 38, 43*) and are conserved in systems with similar repeats. Placement of an additional LAS in front of a non-leader repeat resulted in adaptation at both sites (*38*), whereas LAS deletion caused ectopic integration at a downstream repeat adjacent to a spacer containing a LAS-like sequence (*22*). Taken together, this shows specific sequences upstream of CRISPR arrays direct leader-polarized spacer integration, both via direct Cas1-Cas2 recognition and assisted by host proteins, such as IHF.

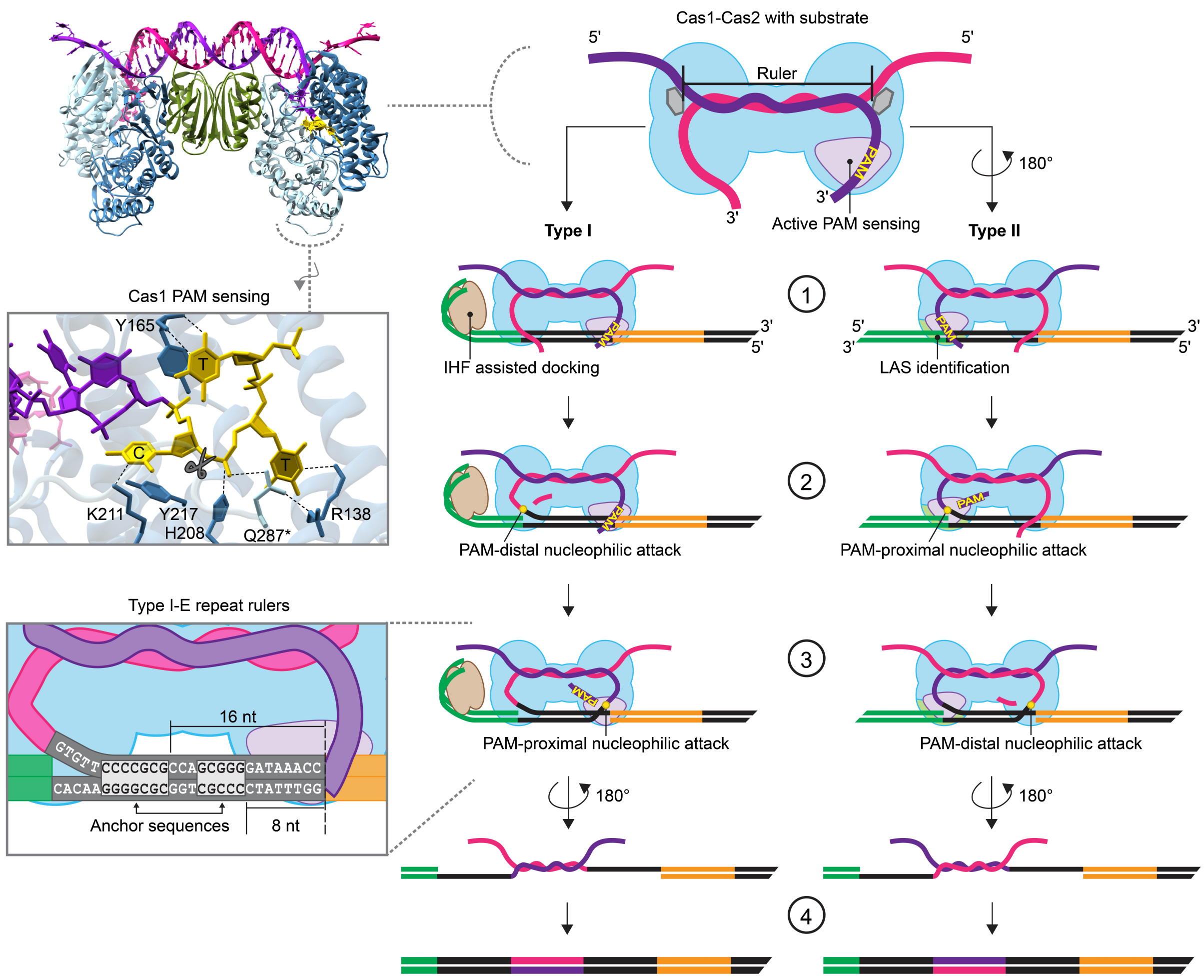

Fig. 2: Cas1-Cas2-mediated spacer acquisition. The substrate loaded Cas1-Cas2 protein complex (*E. coli* type I-E structure shown top left; PDB 5DQZ) with the active PAM sensing domain highlighted (light purple) and a partially duplexed DNA pre-spacer substrate (strands are purple and pink) (*41, 42*). The Cas1 PAM sensing insert shows the canonical type I-E PAM (CTT), residue-specific interactions (a residue from the non-catalyic Cas1 monomer is annotated with *), and site of PAM processing (scissors). The ruler mechanism determining spacer length for the type I-E systems uses two conserved tyrosine residues (grey hexagons). Spacer integration proceeds as follows: 1) the Cas1-Cas2:pre-spacer complex binds the leader (green) and first repeat (black). 2) The first nucleophillic attack occurs at the leader-repeat junction and gives rise to a half-site intermediate. 3) The second nucleophillic attack occurs at the repeat-spacer (orange) boundary resulting in full site integration. The type I-E repeat is magnified (lower left) to indicate the inverted repeats within its sequence and highlight the anchoring sites of the molecular rulers that determine the point of integration. 4) Host DNA repair enzymes fill the intergration site. For additional details, see the text.

Integration into the CRISPR array

In almost all types of CRISPR-Cas systems, the presence of a short sequence motif in the target nucleic acid adjacent to where the crRNA basepairs is essential for interference (the target-strand that the crRNA pairs to is known as the protospacer) (**Fig. 3**) (*46*). This sequence motif is termed a protospacer adjacent motif (PAM) and is a key feature for spacer selection during adaptation (*17, 27, 47, 48*). Acquisition of interference-proficient spacers requires processing of the pre-spacer substrate at a specific position relative to a PAM and also integration into the CRISPR array in the correct orientation. The active site of each Cas1 monomer contains a PAM sensing domain (*41, 42*) and the presence of a PAM within the pre-spacer substrate ensures integration in the appropriate orientation (*49-51*). Accordingly, PAM proximal processing, resulting in complete or partial (in the case of type I-E) removal of the PAM, is likely to occur after Cas1-Cas2 orients and docks at the leader-repeat. In contrast, if complete processing occurred before docking to the CRISPR locus, then the PAM directionality cue would be lost. Cas1-mediated processing of the pre-spacer creates two 3'OH ends required for nucleophilic attack on each strand of the leader-proximal repeat (*36, 37, 52*). The initial nucleophilic attack most likely occurs at the leader-repeat junction and forms a half-site intermediate, then a second attack at the existing repeat-spacer junction generates the full-site integration product (**Fig. 2**). The precise order of the pre-spacer processing and integration steps remains to be fully determined, yet considerable progress toward elucidating the reaction mechanisms has been made.

Following the first nucleophilic attack, Cas1-Cas2 employs molecular rulers that harness the intrinsic sequence-specificity of the complex to define the site of the second attack and ensure accurate repeat length duplication. CRISPR repeats are often semi-palindromic, containing two short inverted repeat (IR) elements, but the location of these can vary (*53*). In type I-B and I-E systems, the IRs occur close to the center of the repeat (**Fig. 2**) and are important for adaptation (*54, 55*). In the type I-E system, both IRs act as anchors for the Cas1-Cas2 complex, positioning the active site for the second attack at the repeat-spacer boundary (*54*). However, in the type I-B system from *Haloarcula hispanica*, only the first IR was essential for integration, and thus a single molecular ruler directed by an anchor between the IRs was proposed (*55*). In contrast, in the type II-A systems of *Streptococcus thermophilus* and *S. pyogenes* the IRs are located distally within the repeats, suggesting these short sequences may directly position the nucleophilic attacks without molecular rulers (*38, 43*). Although further work is required to determine how the spacer integration events are directed in different CRISPR-Cas systems, it seems likely the conserved leader-repeat regions at the beginning of CRISPR arrays maintain recognizable sequences to ensure Cas1-Cas2 localizes appropriately and spacer insertion and repeat duplication is of the correct length.

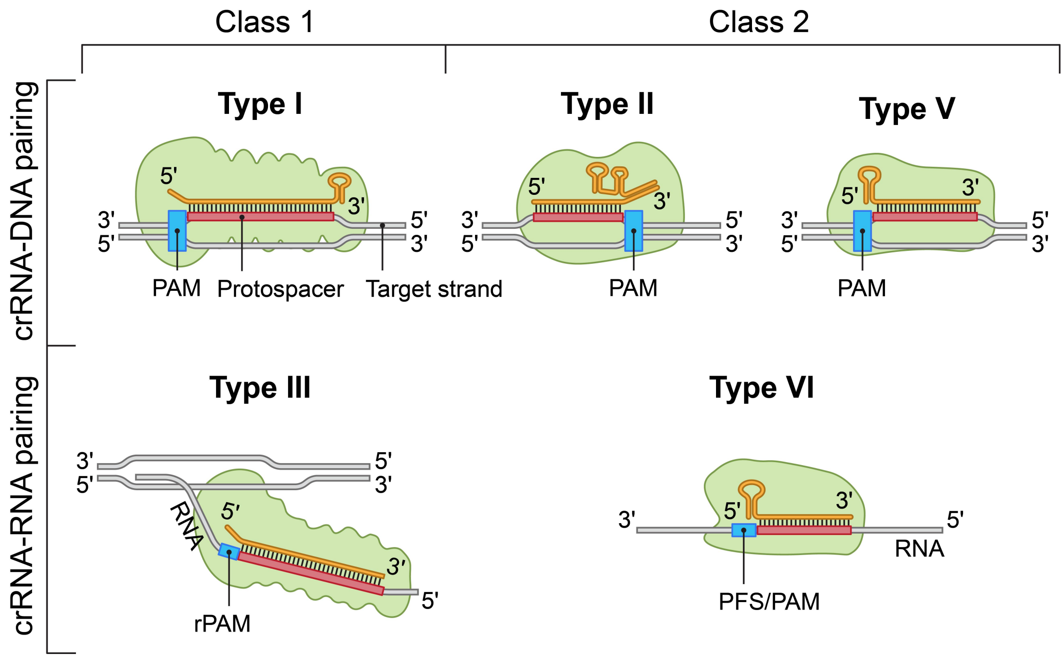

Fig. 3: Target interactions and the PAMs of different CRISPR-Cas types. DNA targets are recognized by the crRNA-effector complexes of types I, II and V, resulting in formation of an R-loop with the non-target strand displaced. The target strand contains the protospacer (red), which is complementary to the spacer (crRNA, orange) sequence. The protospacer adjacent motif (PAM, blue) is located at either the 3' end of the protospacer (type I and type V) or the 5' end (type II). The PAM assignment is consistent with target-centric nomenclature (*46*). Type III and VI recognize RNA targets, with type III exhibiting transcription-dependent DNA targeting. Some type III systems require an RNA-based PAM (rPAM). Type VI systems exhibit a protospacer flanking sequence (PFS) specificity, which is analogous to a PAM.

Production of spacers from foreign DNA

Naive adaptation

Acquisition of spacers from MGEs that are not already catalogued in host CRISPRs is termed naïve adaptation (*56*) (**Fig. 4**). To facilitate naïve adaptation, pre-spacer substrates are generated from foreign material and loaded onto Cas1-Cas2. Currently, the main known source of these precursors is the host RecBCD complex (*57*). Stalled replication forks that occur during DNA replication can result in double strand breaks (DSBs), which are repaired via RecBCD-mediated unwinding and degradation of the dsDNA ends back to the nearest Chi sites (*58*). During this process, RecBCD produces ssDNA fragments that are proposed to anneal, forming substrates suitable for use by Cas1-Cas2 (*57*). Loading of substrates into Cas1-Cas2 is likely enhanced by interaction between Cas1 and RecBCD (*59*), positioning the adaptation machinery adjacent to the site of substrate generation. The increased number of active origins of replication and the paucity of Chi sites on MGEs, versus the host chromosome, biases naïve adaptation toward foreign DNA. Furthermore, RecBCD recognizes unprotected dsDNA ends, which are commonly present in phage genomes upon injection or prior to packaging, thereby providing an additional phage-specific source of naïve adaptation substrates (*57, 60*).

Despite the clear role of RecBCD in substrate generation, naïve adaptation also occurs in its absence, albeit with reduced bias toward foreign DNA (*57*). Events other than DSBs might also stimulate naïve adaptation, such as R-loops that prime plasmid replication (*61*), lagging ends of incoming conjugative elements (*62*), and even CRISPR-Cas mediated spacer integration events themselves (*51, 57*). Furthermore, it is unknown whether all CRISPR-Cas systems display an intrinsic adaptation bias towards foreign DNA. Complicating results, spacer acquisition from the host genome in native systems could be underestimated because the resulting self-targeting means these genotypes are typically lethal (*32, 33, 51, 63*). For example, in the *S. thermophilus* type II-A system, adaptation appears biased toward MGEs, yet nuclease-deficient Cas9 (dCas9) failed to discriminate between acquisition from host versus foreign DNA (*63*) and it is unknown whether the adaptation was reliant on DNA break repair. Further studies in a range of host systems are required to clarify how diverse CRISPR-Cas systems balance the requirement for naïve adaptation from MGEs against the risk of self-acquisition events.

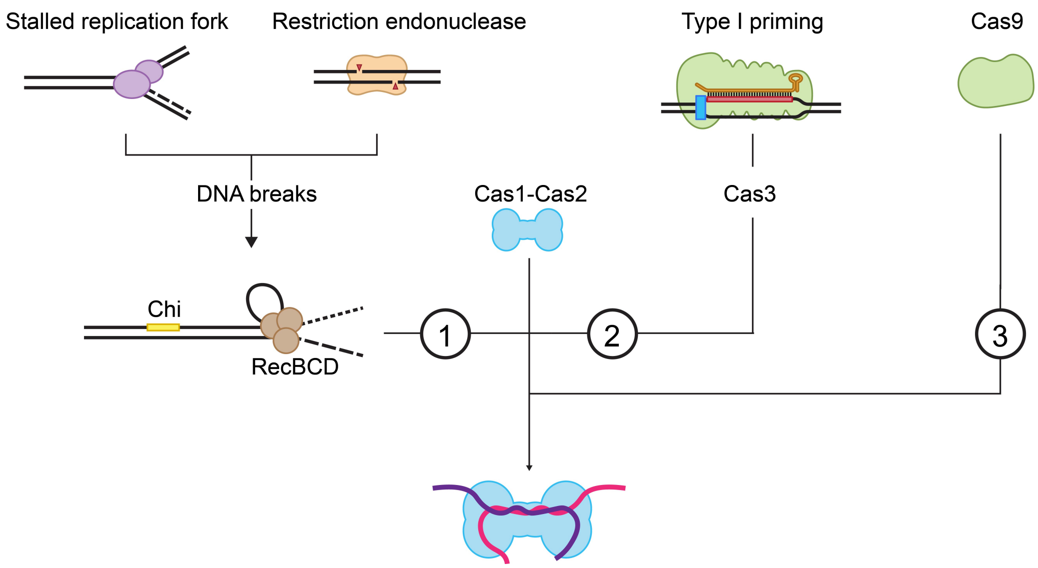

Fig. 4: Cas1-Cas2 substrate production pathways. 1) Naïve generation of substrates by RecBCD activity on DNA ends resulting from DSBs from stalled replication forks, innate defenses such as restriction endonuclease activity or from the ends of phage genomes (not shown). 2) Primed substrate production in type I systems. 3) Cas9-dependent spacer selection in type II systems. For details, see the text.

crRNA-directed adaptation (Priming)

Mutations in the target PAM or protospacer sequences can abrogate immunity, allowing MGEs to escape CRISPR-Cas defenses (*47, 64, 65*). Furthermore, the immunological effectiveness of individual spacers varies: often several target-specific spacers are required to both mount an effective defense (*66, 67*) and prevent proliferation of MGE escape mutants (*13, 14*). Thus, CRISPR-Cas systems need to adapt faster than the foreign element can evade targeting. Indeed, type I systems have evolved a mechanism known as primed adaptation (priming) to facilitate rapid CRISPR adaptation (*68, 69*), even against highly divergent invaders (*65*) (**Fig. 4**). In contrast to naïve adaptation, priming utilizes target recognition by crRNAs from pre-existing spacers to direct spacer acquisition toward invaders whose proliferation exceeds the existing defense capabilities. This often occurs with MGE escape mutants, but also when the CRISPR-Cas expression level is insufficient to provide immunity – even with spacers perfectly targeting the MGE (*65, 68-72*).

Priming begins with target recognition by crRNA-effector complexes. Therefore, factors that influence target recognition (i.e. the formation and stability of the R-loop – see **Fig. 3**), including PAM sensing and crRNA:target complementarity, affect the efficiency of primed adaptation (*64, 65, 67, 73-80*). Furthermore, these same factors influence conformational rearrangements in the target-bound crRNA-effector complex, coalescing to favor either interference or priming (*67, 74, 75, 78, 81*). In type I-E systems, the Cas8e (Cse1) subunit of Cascade can adopt one of two conformational modes (*78, 81*), which may promote either direct or Cas1-Cas2-stimulated recruitment of the effector Cas3 nuclease (*74, 75, 81*).

Cas3, found in all type I systems, exhibits 3' to 5' helicase and endonuclease activity that nicks, unwinds and degrades target DNA (*82-85*). In vitro activity of the type I-E Cas3 produces ssDNA fragments of ~30-100 nucleotides that are enriched for PAMs in their 3' ends, which anneal to provide partially duplexed pre-spacer substrates (*73*). The spatial positioning of Cas1-Cas2 during primed substrate generation has not been clearly established, although Cas1-Cas2-facilitated recruitment of Cas3 would imply the adaptation machinery is localized close to the site of substrate production (*74, 81*). In support of this, Cas3 in type I-F systems is fused to the C-terminus of Cas2 and forms a Cas1-Cas2-3 complex (*35*) that couples the adaptation machinery directly to the source of substrate generation during primed adaptation (*51, 86*).

Despite different crRNA-effector:target interactions favoring distinct Cas3 recruitment modes, primed adaptation can occur from both escape mutants and interference-proficient targets (*51, 68, 69, 87*). When target copy-number influences are excluded for type I-E and type I-F systems, interference-proficient targets promote stronger spacer acquisition than escape targets (*51, 87*). This provides a positive feedback loop, reinforcing immunity against recurrent threats even in the absence of escapees (*51, 69*). However, because target interference rapidly destroys the invader, more spacer acquisition is provoked by escape mutants where replication of the MGE outpaces its destruction. Over time, the prolonged presence of the invader, combined with the priming-centric target recognition mode, results in higher net production of pre-spacer substrates from escape mutants (*51, 72, 73, 87*).

Because priming initiates with site-specific target recognition (i.e. targeting a 'priming' protospacer), Cas1-Cas2 compatible substrates are subsequently produced from MGEs with locational biases (**Fig. 5**). Mapping the MGE sequence positions and strands targeted by newly acquired spacers (i.e. their corresponding protospacers) revealed subtype-specific patterns and has provided much of our insight into the priming mechanisms (*50, 51, 68, 69, 86, 88, 89*). In type I-E systems, new protospacers map to the same strand (*50, 69*) as the priming protospacer (**Fig. 5**). For type I-B priming, Cas3 is predicted to load onto either strand at the priming protospacer, resulting in a bidirectional distribution of new protospacers (*88*). For type I-F priming, the first new protospacer typically maps to the strand opposite the priming protospacer, in a direction consistent with Cas3 loading and helicase activity on the non-target strand. Furthermore, once the first spacer is acquired, two targets in the MGE will be recognized and substrate production can be driven from both locations (*51, 86*) (**Fig. 5**). However, in a head-to-head contest interference-proficient targets dominate, thus, subsequent spacers (i.e. the second and beyond) generally result from targeting by the first new spacer and are typically located back towards the original priming protospacer(*51*) (**Fig. 5**). The dominance of the first new spacer also holds true for type I-E (*69, 87*) and likely all other systems that display priming. However, these are generalized models and many questions remain unresolved, such as the mechanisms resulting in strand selection and why some spacer sequences are more highly acquired from MGEs than others. Further analyses of priming in different systems, particularly the order of new spacers acquired, will greatly inform our understanding of primed Cas1-Cas2 substrate production.

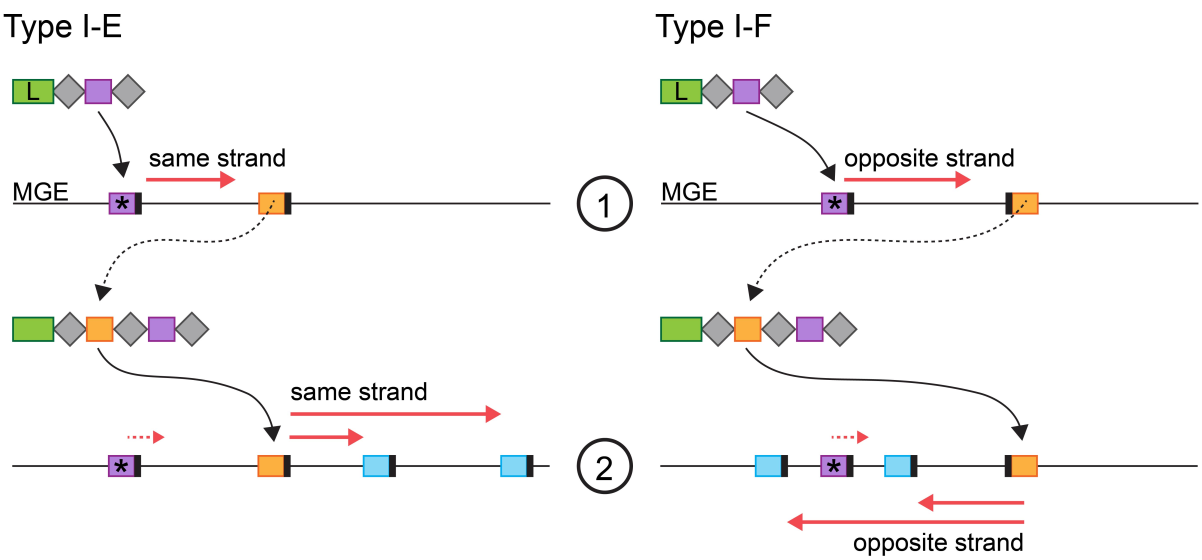

Fig. 5: Primed adaptation from a multi-copy MGE by type I-E and I-F CRISPR-Cas systems. 1) An existing spacer (purple) with homology to an MGE sequence that has escaped interference (the 'priming' protospacer denoted with an asterisk) directs target recognition – the PAM adjacent to the protospacer is shown in black (PAMs at the right or left of protospacers indicate the strand each protospacer is on). The crRNA-effector complex recruits Cas3 and the 3' to 5' helicase activity (illustrated by the red arrow) results in the acquisition of a new spacer that maps to a protospacer (orange) from a site distal to the initial priming location. 2) The new interference-proficient spacer directs targeting of the MGE and recruitment of Cas3. Hence, subsequent spacers (mapping to blue protospacers) typically originate from Cas3 activity (red arrows) beginning at this location. See text for details.

Cas protein-assisted production of spacers

Given the apparent advantages conferred by priming in type I systems, mechanisms to utilize existing spacers to direct adaptation are likely to exist in other CRISPR-Cas types. For example, DNA breaks induced by interference activity of class 2 CRISPR-Cas effector complexes could trigger host DNA repair mechanisms (e.g. RecBCD), thereby providing substrates for Cas1-Cas2. In agreement with a generalized DNA break-stimulated adaptation model, restriction enzyme activity stimulated RecBCD-facilitated adaptation (*57*). This may also partially account for the enhanced adaptation observed during phage infection of a host possessing an innate defense restriction-modification system (*31*), but whether this was RecBCD-dependent is unknown. For CRISPR-Cas-induced DNA breaks, spacer acquisition would be preceded by target recognition, hence the resulting adaptation could be considered related to 'priming' (*90*). Although direct evidence to support this concept is lacking, adaptation in type II-A systems requires Cas1-Cas2, Cas9, a tracrRNA and Csn2 (*63, 90*). In support of a role for Cas9 in substrate generation, the PAM-sensing domain of Cas9 enhances the acquisition of spacers with compatible PAMs (*90*). However, Cas9 nuclease activity is dispensable (*63*) and existing spacers are not strictly necessary (*90*), suggesting that PAM interactions of Cas9 could be sufficient to select appropriate new spacers. Some Cas9 variants can also function with non-CRISPR RNAs and tracrRNA (*91*), raising the possibility that host or MGE-derived RNAs might direct promiscuous Cas9 activity, resulting in DNA breaks, or replication fork stalling and trigger spacer integration.

Roles of accessory Cas proteins in adaptation

Although Cas1 and Cas2 play a central role in adaptation, type-specific variations in *cas* gene clusters occur. In many systems, Cas1-Cas2 is assisted by accessory Cas proteins, which are often mutually exclusive and type-specific (*4*). For example, in the *S. thermophilus* type II-A system, deletion of *csn2* impaired the acquisition of spacers from invading phages (*6*). Csn2 assembles into ring-shaped homo-tetramers with a calcium-stabilized central channel (*92, 93*) that binds cooperatively to the free ends of linear dsDNA and can translocate by rotation-coupled movement (*94, 95*). Given that substrate-loaded type II-A Cas1-Cas2 is capable of full-site spacer integration in vitro (*43*), Csn2 is likely to play an earlier role in either pre-spacer substrate production, selection or processing. Potentially, Csn2 binding to the free ends of dsDNA provides a cue to direct nucleases necessary for substrate generation (*94*).

Cas4, another ring-forming accessory protein, is found in type I, II-B and V systems (*4*). Confirming its role in adaptation, Cas4 is necessary for type I-B priming in *H. hispanica* (*88*) and interacts with a Cas1-2 fusion protein in the *Thermoproteus tenax* type I-A system (*96*). Fusions between Cas4 and Cas1 are found in several systems, supporting a functional association with adaptation. Cas4 contains a RecB-like domain and four conserved cysteine residues, which are presumably involved in the coordination of an iron-sulfur cluster (*97*). However, Cas4 proteins appear to be functionally diverse with some possessing uni- or bi-directional exonuclease activity (*97, 98*), while others exhibit ssDNA endonuclease activity and unwinding activity on dsDNA (*98*). Due to its nuclease activity, Cas4 is hypothesized to trim pre-spacer substrates and aid adaptation by generating 3' overhangs in the duplex pre-spacer substrate.

To provide immunity, type III systems require spacers complementary to RNA transcribed from MGEs (**Fig. 3**) (*99, 100*). Some bacterial type III systems contain fusions of Cas1 with reverse transcriptase domains (RTs), which provide a mechanism to integrate spacers from RNA substrates (*101*). The RT-Cas1 fusion from *M. mediterranea* can integrate RNA precursors into an array, which are subsequently reverse transcribed to generate DNA spacers (*101*). However, integration of DNA-derived spacers also occurs, indicating that the RNA derived-spacer route is not exclusive (*101*). Hence, the integrase activity of RT-Cas1-Cas2 is extended by the reverse transcriptase activity, enabling enhanced build-up of immunity against highly transcribed DNA MGEs and potentially from RNA-based invaders.

Despite evidence that accessory Cas proteins are involved in spacer acquisition, their roles mostly remain elusive. Furthermore, other host proteins may also be required for pre-spacer substrate production. For example, RecG is required for efficient primed adaptation in type I-E and I-F systems, but its precise role remains speculative (*39, 102*). Additionally, it remains enigmatic why some CRISPR-Cas systems appear to require accessory proteins, whilst closely related types do not. For example, type II-C systems lack *cas4* or *csn2* that assist in type II-A and II-B adaptation, respectively. These type-specific differences exemplify the diversity that has arisen during evolution of CRISPR-Cas systems.

Evolution of adaptation

The expanding knowledge of spacer integration has led to a promising theory for the evolutionary origin of CRISPR-Cas systems (*103*). Casposons are transposon-like elements typified by the presence of Cas1 homologs, casposases, which catalyze site-specific DNA integration and result in the duplication of repeat sites analogous to CRISPR adaptation (*104, 105*). It is proposed that ancestral innate defenses gained DNA integration functionality from casposases, seeding the genesis of prokaryotic adaptive immunity (*106*). The innate ancestor remains to be determined, but is likely to be a nuclease-based system. Co-occurrence of casposon-derived terminal inverted repeats and casposases in the absence of full casposons might represent an intermediate of the CRISPR signature repeat-spacer-repeat structures (*107*). However, the evolutionary journey from the innate immunity-casposase hybrid to full adaptive immunity remains unclear. Nevertheless, comparative genomics indicate that all known CRISPR-Cas systems evolved from a single ancestor (*4, 5*).

The more compact class 2 CRISPR-Cas systems likely evolved from class 1 ancestors, through acquisition of genes encoding new single-subunit effector proteins and loss of additional *cas* genes (*5*). Evolution of CRISPR-Cas types would have required stringent co-evolution of the adaptation machinery, leader-repeat sequences (*108*), crRNA processing mechanisms and effector complex function. However, despite the subsequent divergence of CRISPR-Cas systems into several types, Cas1-Cas2 remains the workhorse of spacer acquisition, central to the success of CRISPR-Cas systems (*4, 5*). As long as spacers can be acquired from MGEs, unique effector machineries capable of utilizing the information stored in CRISPRs will continue to evolve.

Mechanisms to generate Cas1-Cas2 compatible substrates, such as primed adaptation might have arisen because naïve acquisition is an inefficient and undirected process, potentially leading to high rates of lethal self-targeting spacers. However, despite the apparent advantages of primed adaptation, it was recently reported that promiscuous binding of crRNA-effector complexes to the host genome results in a basal level of self-priming, the extent of which is likely underrepresented due to the lethality of such events (*51*). Host *cas* gene regulation mechanisms have arisen to balance the likelihood of self-acquisition events against the requirement to adapt to new threats, for example, when the risk of phage infection or HGT is high (*109, 110*). Alternatively, it has been proposed that selective acquisition of self-targeting spacers could provide benefits such as invoking altruistic cell death (*111*), rapid genome evolution (*33*), regulation of host processes (*112, 113*), or even preventing the uptake of other CRISPR-Cas systems (*114*).

Outlook

The past four years has seen rapid progress to understand the adaptation phase of CRISPR-Cas immunity. Despite this progress, many facets of CRISPR adaptation require further attention.

Synergy between innate defense systems and adaptation is relatively unexplored, but two roles can be envisioned; DNA breaks (*57*) stimulating generation of substrates for spacer acquisition (**Fig. 4**) or stalling of infection to 'buy time' for adaptation (*31, 115, 116*). Analogously, it remains to be determined whether interference by CRISPR-Cas systems other than type I can also stimulate primed adaptation. If not, the benefits of priming might provide an explanation for why type I systems are more prevalent than other types.

It is also unclear why many CRISPR-Cas systems have multiple arrays used by a single set of Cas proteins, rather than a solo array. Given that Cas1-Cas2 is directed to leader-repeat junctions during integration, multiple arrays might provide additional integration sites, increasing adaptation efficiency. In addition, parallel CRISPR arrays should increase crRNA production from recently acquired spacers (i.e. due to polarization) (*22*). Whereas some strains have multiple CRISPR arrays belonging to the same type, other hosts have several types of CRISPR-Cas systems simultaneously (*117*). The benefits of harboring multiple CRISPR-Cas systems are not entirely clear, but can result in spacers used by different system to extend targeting to both RNA and DNA (*118*). From an adaptation perspective, multiple systems might enable a wider PAM repertoire to be sampled during spacer selection. Additional systems in a single host could also be a response to defy phage- and MGE-encoded anti-CRISPR proteins, which can inhibit both interference and primed adaptation (*119-121*), or may allow some systems to function in defense, while others perform non-canonical roles in gene regulation (*113*).

While Cas effector nucleases (e.g. Cas9) have been harnessed for many biotechnological applications, the use of repurposed CRISPR-Cas adaptation machinery has yet to be widely exploited. The sequence-specific integrase activity holds promise in synthetic biology, such as for the insertion of specific sequences (or barcodes) to mark and track cells in a population. In *E. coli* the feasibility of such an approach is evident (*49*), but transition to eukaryotic systems will provide the greatest utility where lineage tracking and cell fate could be followed, as has been performed with Cas9 (*122*). The elements required for leader-specific integration must be carefully considered for the introduction of CRISPR-Cas adaptation into eukaryotic cells, as unintended ectopic integrations could be problematic given the larger eukaryotic sequence space. Ultimately, our understanding of adaptation in prokaryotes may lead to applications where entire CRISPR systems are transplanted into eukaryotic cells to prevent viral invaders. As we begin to comprehend adaptation in more detail the opportunities to repurpose other parts of these remarkable prokaryotic immune systems is increasingly becoming reality.

References and Notes:

1. R. L. Dy, C. Richter, G. P. Salmond, P. C. Fineran, Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu Rev Virol* **1**, 307-331 (2014).
2. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**, 675-687 (2013).
3. L. A. Marraffini, CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55-61 (2015).
4. K. S. Makarova *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736 (2015).
5. P. Mohanraju *et al.*, Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).
6. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
7. S. J. Brouns *et al.*, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
8. A. V. Wright, J. K. Nuñez, J. A. Doudna, Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44 (2016).
9. G. Amitai, R. Sorek, CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* **14**, 67-76 (2016).
10. S. H. Sternberg, H. Richter, E. Charpentier, U. Qimron, Adaptation in CRISPR-Cas Systems. *Mol Cell* **61**, 797-808 (2016).
11. M. J. Lopez-Sanchez *et al.*, The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol* **85**, 1057-1071 (2012).
12. G. W. Tyson, J. F. Banfield, Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10**, 200-207 (2008).
13. A. F. Andersson, J. F. Banfield, Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047-1050 (2008).
14. S. van Houte *et al.*, The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385-388 (2016).
15. C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653-663 (2005).
16. F. Liu *et al.*, Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* **77**, 1946-1956 (2011).
17. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* **40**, 5569-5576 (2012).
18. C. Díez-Villaseñor, N. M. Guzmán, C. Almendros, J. García-Martínez, F. J. Mojica, CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol* **10**, 792-802 (2013).
19. S. Erdmann, R. A. Garrett, Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* **85**, 1044-1056 (2012).
20. C. L. Sun, B. C. Thomas, R. Barrangou, J. F. Banfield, Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J* **10**, 858-870 (2016).
21. D. Paez-Espino *et al.*, Uncovering Earth's virome. *Nature* **536**, 425-430 (2016).
22. J. McGinn, L. A. Marraffini, CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol Cell* **64**, 616-623 (2016).
23. Y. Zhang *et al.*, Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol Cell* **50**, 488-503 (2013).

24. A. Biswas, R. H. Staals, S. E. Morales, P. C. Fineran, C. M. Brown, CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016).
25. P. Horvath *et al.*, Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1401-1412 (2008).
26. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182 (2005).
27. A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).
28. K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, E. V. Koonin, 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).
29. L. A. Marraffini, E. J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845 (2008).
30. S. T. Abedon, Facilitation of CRISPR adaptation. *Bacteriophage* **1**, 179-181 (2011).
31. A. P. Hynes, M. Villion, S. Moineau, Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat Commun* **5**, 4399 (2014).
32. A. Stern, L. Keren, O. Wurtzel, G. Amitai, R. Sorek, Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* **26**, 335-340 (2010).
33. R. B. Vercoe *et al.*, Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* **9**, e1003454 (2013).
34. J. K. Nuñez *et al.*, Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528-534 (2014).
35. C. Richter, T. Gristwood, J. S. Clulow, P. C. Fineran, In vivo protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas System. *PLoS One* **7**, e49549 (2012).
36. J. K. Nuñez, A. S. Lee, A. Engelman, J. A. Doudna, Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193-198 (2015).
37. Z. Arslan, V. Hermanns, R. Wurm, R. Wagner, U. Pul, Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res* **42**, 7884-7893 (2014).
38. Y. Wei, M. T. Chesne, R. M. Terns, M. P. Terns, Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res* **43**, 1749-1758 (2015).
39. I. Ivančić-Baće, S. D. Cass, S. J. Wearne, E. L. Bolt, Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res* **43**, 10821-10830 (2015).
40. B. Wiedenheft *et al.*, Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904-912 (2009).
41. J. Wang *et al.*, Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840-853 (2015).
42. J. K. Nuñez, L. B. Harrington, P. J. Kranzusch, A. N. Engelman, J. A. Doudna, Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535-538 (2015).
43. A. V. Wright, J. A. Doudna, Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol* **23**, 876-883 (2016).
44. J. K. Nuñez, L. Bai, L. B. Harrington, T. L. Hinder, J. A. Doudna, CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* **62**, 824-833 (2016).
45. K. N. Yoganand, R. Sivathanu, S. Nimkar, B. Anand, Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res*, (2016).

46. R. T. Leenay, C. L. Beisel, Deciphering, communicating, and engineering the CRISPR PAM. *J Mol Biol*, (2016).
47. H. Deveau *et al.*, Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1390-1400 (2008).
48. F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).
49. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
50. S. Shmakov *et al.*, Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res* **42**, 5907-5916 (2014).
51. R. H. Staals *et al.*, Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun* **7**, 12853 (2016).
52. C. Rollie, S. Schneider, A. S. Brinkmann, E. L. Bolt, M. F. White, Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* **4**, (2015).
53. V. Kunin, R. Sorek, P. Hugenholtz, Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* **8**, R61 (2007).
54. M. G. Goren *et al.*, Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. *Cell Rep* **16**, 2811-2818 (2016).
55. R. Wang, M. Li, L. Gong, S. Hu, H. Xiang, DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res* **44**, 4266-4277 (2016).
56. P. C. Fineran, E. Charpentier, Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202-209 (2012).
57. A. Levy *et al.*, CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505-510 (2015).
58. D. B. Wigley, Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* **11**, 9-13 (2013).
59. M. Babu *et al.*, A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* **79**, 484-502 (2011).
60. L. W. Enquist, A. Skalka, Replication of bacteriophage lambda DNA dependent on the function of host and viral genes. I. Interaction of red, gam and rec. *J Mol Biol* **75**, 185-212 (1973).
61. J. Gowrishankar, J. K. Leela, K. Anupama, R-loops in bacterial transcription: their causes and consequences. *Transcription* **4**, 153-157 (2013).
62. E. R. Westra *et al.*, CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids. *RNA Biol* **10**, 749-761 (2013).
63. Y. Wei, R. M. Terns, M. P. Terns, Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev* **29**, 356-361 (2015).
64. E. Semenova *et al.*, Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-10103 (2011).
65. P. C. Fineran *et al.*, Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A* **111**, E1629-1638 (2014).
66. D. Paez-Espino *et al.*, Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4**, 1430 (2013).
67. C. Xue *et al.*, CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res* **43**, 10831-10847 (2015).
68. K. A. Datsenko *et al.*, Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945 (2012).
69. D. C. Swarts, C. Mosterd, M. W. van Passel, S. J. Brouns, CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888 (2012).

70. E. Savitskaya, E. Semenova, V. Dedkov, A. Metlitskaya, K. Severinov, High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol* **10**, 716-725 (2013).
71. A. G. Patterson, J. T. Chang, C. Taylor, P. C. Fineran, Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res* **43**, 6038-6048 (2015).
72. K. Severinov, I. Ispolatov, E. Semenova, The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. *Front Mol Biosci* **3**, 45 (2016).
73. T. Künne *et al.*, Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol Cell* **63**, 852-864 (2016).
74. S. Redding *et al.*, Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **163**, 854-865 (2015).
75. T. R. Blosser *et al.*, Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell* **58**, 60-70 (2015).
76. D. G. Sashital, B. Wiedenheft, J. A. Doudna, Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol Cell* **46**, 606-615 (2012).
77. M. F. Rollins, J. T. Schuman, K. Paulus, H. S. Bukhari, B. Wiedenheft, Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*. *Nucleic Acids Res* **43**, 2216-2222 (2015).
78. R. P. Hayes *et al.*, Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* **530**, 499-503 (2016).
79. P. B. van Erp *et al.*, Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res* **43**, 8381-8391 (2015).
80. M. Li, R. Wang, H. Xiang, *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res* **42**, 7226-7235 (2014).
81. C. Xue, N. R. Whitis, D. G. Sashital, Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol Cell* **64**, 826-834 (2016).
82. T. Sinkunas *et al.*, Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* **30**, 1335-1342 (2011).
83. S. Mulepati, S. Bailey, In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J Biol Chem* **288**, 22184-22192 (2013).
84. E. R. Westra *et al.*, CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**, 595-605 (2012).
85. Y. Huo *et al.*, Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol* **21**, 771-777 (2014).
86. C. Richter *et al.*, Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* **42**, 8516-8526 (2014).
87. E. Semenova *et al.*, Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc Natl Acad Sci U S A* **113**, 7626-7631 (2016).
88. M. Li, R. Wang, D. Zhao, H. Xiang, Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* **42**, 2483-2492 (2014).
89. C. Rao *et al.*, Active and adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen. *Cell Microbiol* **18**, 1319-1338 (2016).
90. R. Heler *et al.*, Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199-202 (2015).
91. T. R. Sampson, S. D. Saroj, A. C. Llewellyn, Y. L. Tzeng, D. S. Weiss, A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254-257 (2013).

92. K. H. Nam, I. Kurinov, A. Ke, Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *J Biol Chem* **286**, 30759-30768 (2011).
93. P. Ellinger *et al.*, The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J Struct Biol* **178**, 350-362 (2012).
94. Z. Arslan *et al.*, Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Res* **41**, 6347-6359 (2013).
95. K. H. Lee *et al.*, Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins* **80**, 2573-2582 (2012).
96. A. Plagens, B. Tjaden, A. Hagemann, L. Randau, R. Hensel, Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* **194**, 2491-2500 (2012).
97. J. Zhang, T. Kasciukovic, M. F. White, The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One* **7**, e47232 (2012).
98. S. Lemak *et al.*, Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. *J Am Chem Soc* **135**, 17476-17487 (2013).
99. C. R. Hale *et al.*, RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945-956 (2009).
100. G. W. Goldberg, W. Jiang, D. Bikard, L. A. Marraffini, Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633-637 (2014).
101. S. Silas *et al.*, Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234 (2016).
102. G. E. Heussler, J. L. Miller, C. E. Price, A. J. Collins, G. A. O'Toole, Requirements for *Pseudomonas aeruginosa* Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm Enrichment Assay. *J Bacteriol* **198**, 3080-3090 (2016).
103. M. Krupovic, K. S. Makarova, P. Forterre, D. Prangishvili, E. V. Koonin, Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC Biol* **12**, 36 (2014).
104. A. B. Hickman, F. Dyda, The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. *Nucleic Acids Res* **43**, 10576-10587 (2015).
105. P. Beguin, N. Charpin, E. V. Koonin, P. Forterre, M. Krupovic, Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res*, (2016).
106. E. V. Koonin, M. Krupovic, Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* **16**, 184-192 (2015).
107. M. Krupovic, S. Shmakov, K. S. Makarova, P. Forterre, E. V. Koonin, Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. *Genome Biol Evol* **8**, 375-386 (2016).
108. O. S. Alkhnbashi *et al.*, Characterizing leader sequences of CRISPR loci. *Bioinformatics* **32**, i576-i585 (2016).
109. A. G. Patterson *et al.*, Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell*, (2016).
110. N. M. Høyland-Kroghsbo *et al.*, Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A*, (2016).
111. E. V. Koonin, F. Zhang, Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *Bioessays*, (2016).
112. R. Li *et al.*, Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. *Cell Res* **26**, 1273-1287 (2016).
113. E. R. Westra, A. Buckling, P. C. Fineran, CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* **12**, 317-326 (2014).

114. C. Almendros, N. M. Guzman, J. Garcia-Martinez, F. J. Mojica, Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems. *Nat Microbiol* **1**, 16081 (2016).

115. K. S. Makarova, V. Anantharaman, L. Aravind, E. V. Koonin, Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct* **7**, 40 (2012).

116. M. E. Dupuis, M. Villion, A. H. Magadán, S. Moineau, CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun* **4**, 2087 (2013).

117. R. H. J. Staals, S. J. J. Brouns, in *CRISPR-Cas Systems: RNA-mediated Adaptive Immunity in Bacteria and Archaea,* R. Barrangou, J. van der Oost, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 145-169.

118. J. Elmore, T. Deighan, J. Westpheling, R. M. Terns, M. P. Terns, DNA targeting by the type I-G and type I-A CRISPR-Cas systems of *Pyrococcus furiosus*. *Nucleic Acids Res* **43**, 10353-10363 (2015).

119. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432 (2013).

120. A. Pawluk *et al.*, Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol* **1**, 16085 (2016).

121. D. Vorontsova *et al.*, Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res* **43**, 10848-10860 (2015).

122. S. D. Perli, C. H. Cui, T. K. Lu, Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, (2016).

Acknowledgments: Work in the Fineran laboratory on CRISPR-Cas is supported by a Rutherford Discovery Fellowship (PCF) from the Royal Society of New Zealand (RSNZ), the Marsden Fund (RSNZ), the University of Otago and the Bio-protection Research Centre (Tertiary Education Commission). The Brouns laboratory is funded by an ERC starting grant (639707), an NWO VIDI grant (864.11.005), and FOM projectruimte (15PR3188). We thank members of the Fineran and Brouns groups for useful feedback on the manuscript.