Molecular mechanisms of CRISPR–Cas spacer acquisition

Jon McGinn and Luciano A. Marraffini

Abstract | Many bacteria and archaea have the unique ability to heritably alter their genomes by incorporating small fragments of foreign DNA, called spacers, into CRISPR loci. Once transcribed and processed into individual CRISPR RNAs, spacer sequences guide Cas effector nucleases to destroy complementary, invading nucleic acids. Collectively, these two processes are known as the CRISPR–Cas immune response. In this Progress article, we review recent studies that have advanced our understanding of the molecular mechanisms underlying spacer acquisition and that have revealed a fundamental link between the two phases of CRISPR immunity that ensures optimal immunity from newly acquired spacers. Finally, we highlight important open questions and discuss the potential basic and applied impact of spacer acquisition research.

Bacteria and archaea are frequently exposed to both beneficial and parasitic foreign nucleic acids[1]. On the one hand, this enables organisms to access and incorporate diverse genetic material, such as plasmids that encode antibiotic resistance genes or virulence factors encoded by temperate phages. On the other hand, this renders cells vulnerable to parasitic elements that compromise the fitness of the population, such as plasmid-encoded transposons or virulent phages. To balance these costs and benefits, bacteria and archaea have evolved a number of pathways to curate the nucleic acids that enter the cell. These pathways have profound implications for the evolution of bacterial and archaeal populations[1].

CRISPR and their associated *cas* genes encode one such mechanism by which cells restrict incoming nucleic acids[2,3]. CRISPR–Cas systems have the unique ability to heritably alter the host genome by incorporating small fragments of foreign nucleic acids, known as spacers, in between the repeats of the CRISPR locus[2] (FIG. 1a). This process is known as spacer acquisition[4–6]. Spacers are transcribed and parsed into individual CRISPR RNAs (crRNAs), which guide effector Cas nucleases to cleave cognate nucleic acids. Thus, spacer sequences define the specificity of the CRISPR–Cas immune response, bestowing immunity to both the host and its progeny[2,3].

Found in approximately 45% of bacteria and 85% of archaea, CRISPR systems have been categorized by *cas* gene content into two classes, six types and more than 20 subtypes[7]. Each of the six types uses functionally distinct effector complexes that mediate the destruction of foreign nucleic acids. Whereas type I, type II and type V systems target DNA, the type VI system targets RNA, and the type III system targets both RNA and DNA (type IV systems have not yet been experimentally characterized)[7]. By contrast, the core machinery that mediates spacer acquisition, encoded primarily by *cas1* and *cas2*, is well conserved across the different types. Spacer acquisition can be conceptually divided into two phases: capture of spacer sequences in the invading genome (known as protospacers) and spacer integration. During the first phase, protospacers are selected and extracted from foreign genomes. In the second phase, spacers are processed and incorporated into the CRISPR locus. Recent studies have revealed several aspects of the molecular mechanism of spacer acquisition and how these correlate with the specific targeting mechanism of each different CRISPR type. In this Progress article, we review the current models of spacer acquisition and discuss the future of the field, both in terms of basic scientific research and technological applications.

Integration of new spacers

The Cas1–Cas2 integrase. Cas1 is the most highly conserved Cas protein and can be found in all six CRISPR types[7]. It is believed that Cas1 evolved as the core enzyme of a class of transposons called casposons and was eventually co-opted to form the basis of the CRISPR–Cas immune system[8]. In the context of CRISPR immunity, Cas1 interacts with Cas2 to form a complex that acts as the spacer integrase[9–13]. This heterohexameric complex [$(Cas1_2–Cas2)_2$] contains two separate DNA-binding regions, one that binds the incoming protospacer and one that binds the CRISPR array. Once loaded with the incoming spacer, the Cas1–Cas2 complex catalyses two cleavage–ligation reactions, first at the leader end of the first repeat of the CRISPR array and subsequently at the spacer end of the repeat[12–14] (FIG. 1a). In this reaction, the terminal 3′-OH of each strand of the protospacer DNA carries out a nucleophilic attack on each end of the repeat DNA. The product of this reaction is an intermediate in which the 3′ ends of a double-stranded DNA (dsDNA) protospacer are ligated to single-stranded DNA (ssDNA) repeat sequences. These ssDNA ‘gaps’ are presumably filled by a DNA polymerase and then ligated, resulting in a simultaneous spacer insertion and repeat duplication[12–18].

Site-specific integration of new spacers. The integration of new spacers by the Cas1–Cas2 complex is polarized; that is, new spacers are predominantly added to the leader end of the CRISPR array. For this reason, CRISPR loci have been likened to molecular fossil records of past infections, with the newest memories located at the leader end and the most ancestral spacer sequences positioned at the trailer end[18] (BOX 1). By ordering spacers chronologically, CRISPR systems optimize their immune response against the most recent invaders, as leader end spacers provide more robust immunity relative to more downstream positions[19]. This has been proposed to be a bet-hedging strategy in which resources are prioritized to defend the host against

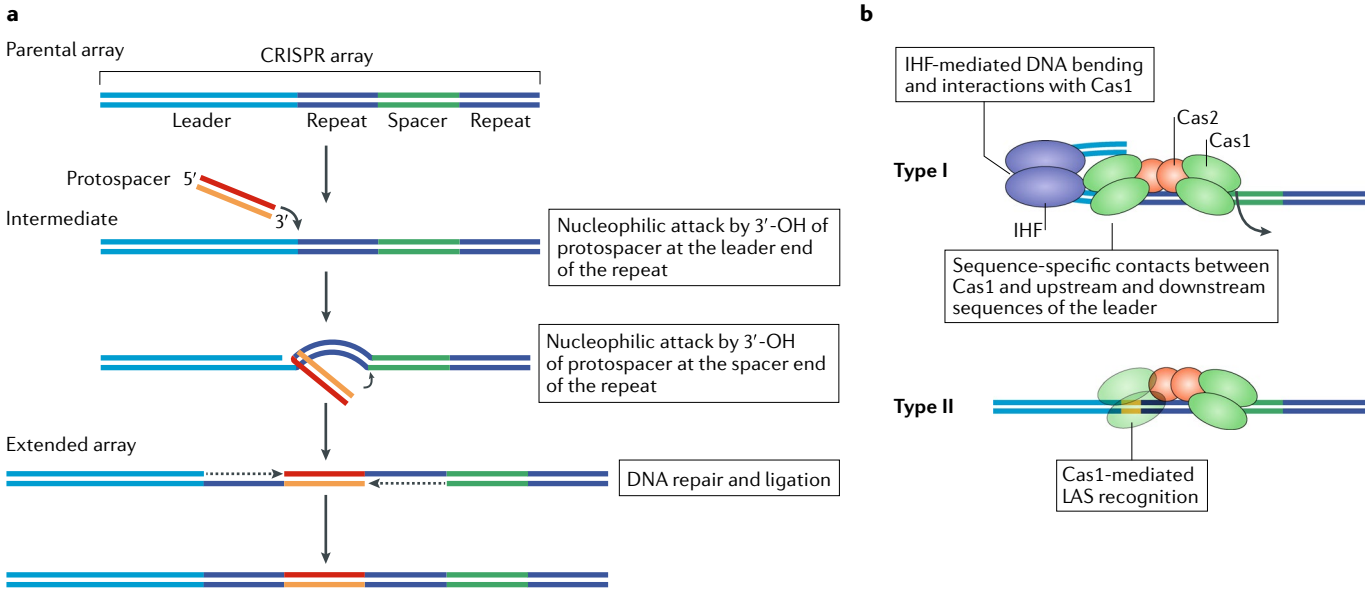


Fig. 1 | **Integration of new spacers into the CRISPR locus. a** | General schematic of the spacer integration reaction. The CRISPR array comprises a series of direct, semi-palindromic repeats intercalated with variable spacer sequences of foreign origin. The array is preceded by an AT-rich leader sequence. Integration of new spacers begins with a concerted cleavage–ligation reaction that occurs preferentially at the leader end of the first repeat, whereby the terminal 3′-OH of the protospacer carries out a nucleophilic attack. Next, the repeat DNA is bent, and a second cleavage–ligation reaction takes place at the spacer side of the repeat. The product of this reaction is an intermediate in which the 3′ ends of a double-stranded (dsDNA) protospacer are ligated to single-stranded DNA (ssDNA) repeat sequences. The ssDNA repeats are presumably filled by DNA polymerase and ligated to complete the spacer integration process. **b** | Two mechanisms for preferential spacer acquisition at the leader end of the CRISPR array. In the type I system, integration host factor (IHF) binds to a conserved binding site in the leader and induces DNA bending, which enables the Cas1–Cas2 complex to perform the first cleavage–ligation reaction. Cas1–Cas2 makes specific contacts with upstream and downstream sequences in the leader, as well as with IHF. In the type II system, Cas1-mediated recognition of the leader anchoring sequence (LAS) in the leader is sufficient for polarized spacer integration.

the most recent invader, which is also most likely to be the greatest threat to the cell[19,20]. One potential mechanism underlying this phenomenon is the differential expression of crRNAs across the CRISPR array, which has been observed in many CRISPR types[21–25]. In one study of the *Streptococcus pyogenes* type II-A system, a twofold difference was observed when comparing the abundance of crRNAs originating from the same spacer sequence in the first or fifth position in the array[19]. Given that a single Cas9 ribonucleoprotein can take hours to find its target (according to single-molecule experiments)[26] and that it is reported to be a single-turnover enzyme[27], it is possible that modest differences in crRNA abundance can give rise to larger differences in the level of immunity afforded to the host[19].

Several mechanisms have the potential to 'reactivate' ancestral memories that are positioned further downstream in the CRISPR array. Spacer deletions have been frequently observed in laboratory and natural contexts[28,29]. Indeed, in one study of the type II-A system from *S. pyogenes*, deletion of four spacers from the 5′ end of the array was strongly selected for during phage infection, which enabled a spacer in position 5 to be shifted to position 1 to enable maximum levels of immunity[19]. Alternatively, another study revealed that internal promoters contained within spacer sequences can enable high expression of downstream crRNAs[30]. In addition, we speculate that it is possible that lower expression of downstream crRNAs that are unable to provide full immunity could enable a primed immune response (see below).

How polarized addition of new spacers is achieved differs by CRISPR type (FIG. 1b). In type I CRISPR systems, an α-helix of Cas1 makes sequence-specific contacts with a minor groove in the 3′ end of the leader[12,31], but this is not sufficient to enforce spacer addition at the leader end[15]. Rather, factors encoded by the host genome are required for site-specific integration. In type I-E and type I-F systems, a protein called integration host factor (IHF) is required for polarized spacer integration in vitro and is required for spacer acquisition in vivo. Type I leaders contain a conserved IHF-binding site, and binding of IHF induces a distortion of the CRISPR array DNA. This creates the ideal target substrate for the Cas1–Cas2 integrase specifically at the first repeat[12,32]. Additionally, the Cas1–Cas2 integrase makes contacts with IHF and upstream sequences in the leader as a result of DNA bending induced by IHF[12] (FIG. 1b). This mechanism leads to a conundrum for type I CRISPR systems found in Gram-positive bacteria, which lack IHF homologues[32]. It is possible that related DNA-bending proteins (such as HU or H-NS) could fulfil this role. Indeed, an archaeal type I-A system, whose host lacks IHF, exhibits leader specificity for spacer integration in a manner dependent on an as of yet unidentified host factor or host factors[33].

Type II CRISPR systems, which are commonly found in Gram-positive hosts[34], also exhibit strictly polarized spacer integration[2,14,19,35]. In contrast to the type I machinery, the type II Cas1–Cas2 complex can induce the necessary DNA topology for the spacer integration reaction without any additional host factors[13,14]. Similar to the type I system, an α-helix of the type II Cas1 makes sequence-specific contacts with the minor groove of the leader DNA (for type II, this is termed the leader anchoring sequence (LAS))[13,14,19]. These contacts are sufficient to position spacer integration at the leader end of the array without the need for any additional host factors[13,14,19] (FIG. 1b). This is achieved owing to the additional stabilizing contacts between the LAS and Cas1 that improve the kinetics of the cleavage–ligation reaction at

the leader-repeat junction[13,14]. Because the second cleavage–ligation reaction occurs at the spacer–repeat junction, the target substrate varies, and this requires some flexibility in the LAS-interacting domain of Cas1 for catalysing the reaction[13]. Most likely as a result of this flexibility, in the absence of a proper LAS, the type II CRISPR systems can undergo 'ectopic spacer integration', that is, integration of new spacers in the middle of the array[19].

Protospacer capture

Identification of foreign nucleic acids. CRISPR systems can acquire self-targeting spacers from the host chromosome[18,36,37], which results in autoimmunity and cell death[38–40]. To avoid this, CRISPR systems use various mechanisms to bias spacer acquisition to foreign genetic elements.

For the generation of spacer substrates, CRISPR systems use the DNA repair machinery of the host: RecBCD in Gram-negative organisms[17,36] and its homologue, AddAB, in Gram-positive organisms[41]. RecBCD, which binds free ends of dsDNA to perform end resection during homologous recombination, stimulates spacer acquisition from double-strand breaks. This activity is limited by *chi* sites, which are eight nucleotide sequence motifs that slow RecBCD activity. Because *chi* sites are enriched in the host chromosome relative to genomes of phages or plasmids (for example, in the *Escherichia coli* genome, *chi* sites are found at rates 14 times higher than expected[36]), this is a mechanism to constrain spacer acquisition from the host genome and differentiate self from non-self nucleic acids[36,41]. Furthermore, the free dsDNA end that is presented to the cell during infection by dsDNA phages is exploited by the CRISPR system to preferentially acquire spacers from the phage DNA (FIG. 2a), as the bacterial chromosome is circular and lacks free DNA ends (with the exception of accidental dsDNA breaks, most common at the terminus). This also biases the pool of acquired spacers to the end of the phage genome that is being injected. This results in acquisition of spacers that facilitate the immediate recognition and cleavage of invading DNA at the very beginning of the infection and results in more effective immunity[41]. Although RecBCD is important for efficient spacer acquisition, its degradation products are reported to be ssDNA fragments[42,43]. Given that the in vitro spacer integration studies showed that dsDNA protospacer substrates are markedly favoured over ssDNA ones[10–15,44], it remains unresolved if and how RecBCD degradation products could be used for spacer integration[45,46]. Alternatively, it is possible that the Cas1–Cas2 machinery physically associates with RecBCD[47] to either directly uptake degradation products from RecBCD[36] or to sample intact dsDNA upstream of RecBCD. Moreover, given that spacer acquisition can occur in the absence of RecBCD[36] and AddAB[41], there may be alternative pathways for spacer generation, which will be an interesting area of future study.

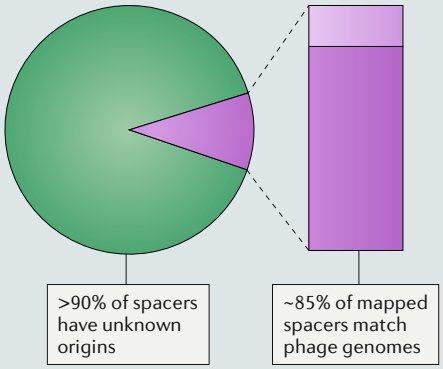
There is also evidence that CRISPR systems have evolved to avoid deleterious levels of autoimmunity by limiting the rate of spacer acquisition. In laboratory settings, successful acquisition of new spacers against phages is an extremely rare event, estimated to occur in only 1 in 10^7 cells[19,48,49]. Spacer acquisition from the host genome is equally rare and does not pose substantial fitness costs to the host. However, increased rates of spacer acquisition in mutants have been shown to lead to higher levels of toxicity[50], which suggests that the rate of spacer acquisition has been evolutionarily tuned to balance the benefits of protection with autoimmunity. To mitigate growth rate costs associated with autoimmunity, it is also possible for spacer acquisition to be temporally regulated. Indeed, quorum sensing has been implicated as a regulator of CRISPR activity in at least two bacterial species[51,52].

Selection of functional targets. Only a subset of sequences in foreign genomes can become functional spacers because of type-specific targeting requirements. In type I and type II, a protospacer-adjacent motif (PAM) located at one of the flanks of the target is required to license target cleavage and to prevent the cleavage of the spacer sequence in the CRISPR array (repeats lack properly positioned PAMs)[53,54]. Although type III flanking sequence requirements are more flexible[55–57], transcription across the target is needed for targeting of the DNA[58–61]. To ensure functional immune responses, CRISPR systems must select protospacers that are flanked by the correct PAM or are actively transcribed.

In type I and type II, the spacer acquisition machinery preferentially samples protospacers with functional PAMs. To accomplish this, various mechanisms have evolved (FIG. 2b). In type I-E, the Cas1–Cas2 complex has direct, sequence-specific interactions with the PAM that bias acquisition to PAM-adjacent protospacers[11]. By contrast, the Cas1–Cas2 complex from the type II CRISPR system does not exhibit any PAM selectivity[13,48]. Instead, the PAM-interacting domain of Cas9 enforces PAM-specific spacer acquisition through direct interactions with the Cas1–Cas2 complex (as well as the type II accessory protein Csn2)[48]. Recently, another accessory protein, Cas4, was found to also be involved in biasing spacer acquisition to protospacers

Box 1 | Dark matter of CRISPR space(rs)

CRISPR arrays provide unique insight into the genetic material encountered and selected against during the course of prokaryotic evolution. As a molecular fossil record of past invasions, we can infer not only the genomes that a given organism encountered but also in what order. Thus, there has been considerable interest in analysing the origin of spacer sequences. Of the spacers that can be mapped to sequenced genomes, 80–90% of spacers map to phage genomes (see the figure). The remainder of mapped spacers match genes associated with mobile genetic elements. Surprisingly, though, a vast majority of spacer sequences (>90%) originate from unknown sources, comprising the CRISPR 'dark matter'[82].



For these mysterious spacer sequences, perhaps the most parsimonious explanation is that sequence databases are missing a vast diversity of phage and mobile genetic element (MGE) sequences. Indeed, sequence analysis has determined that many of these unmapped spacer sequences share similar properties with the mapped spacer sequences, which suggests that they map to uncharacterized classes of phages and MGEs[82]. These dark matter spacers have also been found for a large majority of reverse transcriptase-Cas1-derived spacers, which suggests a diversity of RNA phages and RNA MGEs that remain to be discovered as well[83]. It is also a possibility that these spacer sequences foreshadow the discovery of unexpected sources of nucleic acids.

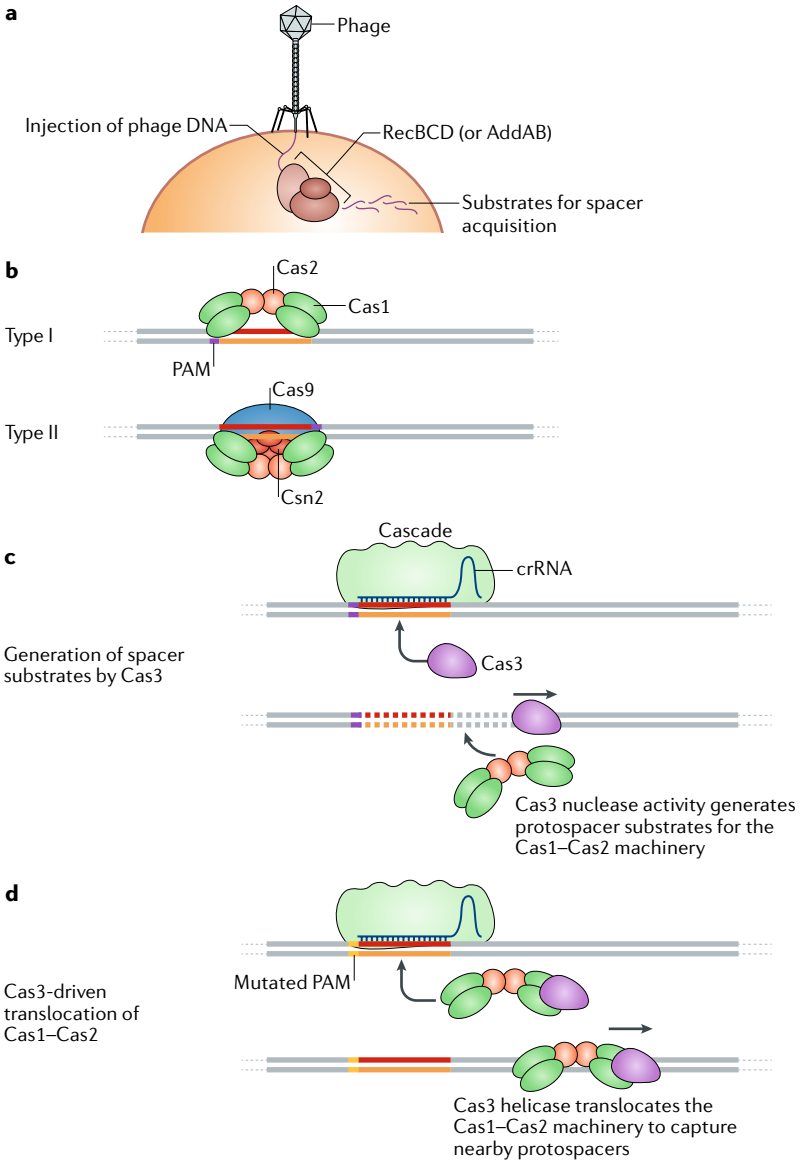


Fig. 2 | **Protospacer selection and capture. a** | RecBCD in Gram-negative organisms (or AddAB in Gram-positive organisms) generates substrates for spacer acquisition following the injection of viral DNA, possibly by producing more invader DNA molecules that contain free ends. **b** | Two mechanisms for selection of functional targets. In the type I-E system, Cas1–Cas2 has inherent substrate preference for protospacers with a canonical protospacer-adjacent motif (PAM). In type II, the PAM-interacting domain of Cas9 (loaded with trans-activating CRISPR RNA (tracrRNA), not shown) guides the Cas1–Cas2 complex (as well as the accessory protein Csn2) in selecting protospacers. **c** | The CRISPR RNA (crRNA)-guided CRISPR-associated complex for antiviral defence (Cascade) binds to a foreign target in a PAM-dependent manner, and it subsequently recruits the nuclease Cas3, which results in the generation of suitable substrates for spacer acquisition. **d** | Imperfect target recognition by Cascade results in an altered conformation of the Cse1 subunit. This leads to the recruitment of a nuclease-inactive Cas3 in a Cas1–Cas2-dependent manner, which mediates primed spacer acquisition.

with functional PAMs in other type I systems[62–64]. Spacer acquisition has not been observed experimentally for many of the transcription-dependent type III CRISPR systems. A small fraction of type III systems has been shown to harbour reverse transcriptase (RT)–Cas1 fusion proteins. Indeed, one such system has been observed to preferentially acquire spacers from transcribed regions of genomes[65]. Although the exact mechanism remains unclear, it has been demonstrated that RT–Cas1 fusion proteins can acquire new spacers directly from RNA transcripts. These RNA-derived spacers provide a mechanism to ensure that their targets are transcribed and can be recognized by the type III RNA-guided nucleases.

Primed spacer acquisition

Pre-existing spacers can enhance the rate of spacer acquisition in a sequence-dependent manner through a process known as primed spacer acquisition (also known as priming); this enhanced rate is relative to 'naive acquisition', in which no spacers in the CRISPR array have homology to the target genome. This boosted rate of spacer acquisition can result from either perfectly matching spacers or partially matching spacers, such as those against mutated phages or related phages with point mutations in the spacer region or PAM. Thus, once a host acquires a single spacer against a phage, it becomes more likely to subsequently acquire additional spacers from the vicinity of the priming target region in the phage genome[30,66].

As of yet, priming has been observed only for type I CRISPR–Cas systems, with the type I-E and type I-F systems being the best studied examples. Close associations and interactions between the spacer acquisition machinery and the interference machinery drive this feedforward cycle. The importance of these associations is underscored by the existence of fusion proteins in which interference genes have been fused to spacer acquisition genes, such as the Cas2–Cas3 fusion protein present in type I-F systems[44,67]. During type I-E targeting, the crRNA-guided CRISPR-associated complex for antiviral defence (Cascade) binds to a foreign target in a PAM-dependent manner, and it subsequently recruits the nuclease Cas3 for target destruction[68]. In addition to eliminating the foreign genome, the nuclease and helicase activities of Cas3 also drive the generation of spacer substrates[45] (FIG. 2c). This mode of interference-driven spacer acquisition seems to result in markedly higher rates of spacer acquisition than in primed acquisition resulting from partially matching targets[67,69,70]. In the absence of a proper PAM, Cascade can still bind the target and recruit Cas3 in a manner dependent on Cas1–Cas2 (REF.[71]). In this context, the nuclease domain of Cas3 is inactive, and it is believed that its helicase activity is used to translocate the Cas1–Cas2 complex along the nearby DNA and drive primed spacer acquisition by the integrase complex[67,71] (FIG. 2d). These distinct modes of Cas3 recruitment may result from distinct conformations adopted by Cascade, in particular the Cse1 subunit, upon binding of either a perfect target or a mutated target[71–74]. Alternatively, given the clear preference for dsDNA end substrates during naive spacer acquisition[36], it has also been suggested that the Cas3 cleavage activity produces the free

ends required for the generation of spacer substrates during primed acquisition and that different cleavage rates that result from perfect or partial homology to the target could sufficiently explain the different rates of acquisition[75].

Primed spacer acquisition enables organisms to defend themselves against rapidly evolving phage populations, such as phages that evade CRISPR–Cas immunity through the introduction of target mutations or related phages with conserved but not identical targets[66,76,77]. Priming can also be a mechanism to bias spacer acquisition activity to foreign DNA substrates that harbour the priming target; that is, using the molecular memories stored in the CRISPR array to differentiate between self and foreign DNA[76]. However, the increased rate of spacer acquisition resulting from priming poses a dilemma for the host cell: the more spacers a CRISPR system acquires, the more likely it is that a spacer with a partial match to the host chromosome is incorporated, which would elicit primed spacer acquisition from the host genome and lead to autoimunity[70]. How CRISPR systems balance these costs and benefits remains to be addressed. Whether priming occurs in other CRISPR types also remains to be determined. Similarly to the interaction between Cascade and Cas3 during priming, Cas9 has been shown to interact with the Cas1–Cas2 integrase[48], an observation that opens up the possibility of priming in the type II CRISPR–Cas immune response. By contrast, for type III systems, the presence of mismatches between the crRNA and the target sequence does not abrogate immunity[57], and it therefore seems unlikely that primed spacer acquisition similar to that seen for type I systems can occur.

Conclusions and outlook

More than 10 years after the discovery of the function of CRISPR systems as a form of adaptive immunity in bacteria and archaea, we are fast approaching a comprehensive understanding of the molecular mechanisms underlying spacer acquisition. Through complementary studies that use genetic, biochemical and structural approaches, we now have key insights into several major steps of this process, from the selection of spacers from foreign genomes to the cleavage–ligation reactions of new spacers into the first repeat of the CRISPR array. However, gaps in our knowledge exist, particularly at the boundaries of protospacer capture and spacer integration. For instance, although it has been demonstrated that RecBCD and AddAB are involved in the generation of substrates for spacer acquisition, the mechanism of this process is largely unsolved. Furthermore, what kind of processing protospacers undergo and how spacer size is regulated remain understudied areas. It is possible that accessory proteins that are not well characterized (such as Cas4 and Csn2) and that are associated with spacer acquisition modules could have a role in these processes[78]. The development of new in vivo and in vitro experimental systems will be key to gaining insight into these processes. Indeed, a few recent studies have already shed light on the role of Cas4 in PAM-dependent protospacer processing[33,62–64].

More broadly, it will be interesting to learn how spacer acquisition affects other aspects of prokaryotic life. From a cell biological perspective, it will be valuable to learn more about the interplay between CRISPR spacer acquisition and other host-encoded genes or pathways, such as other phage defence mechanisms. From an ecological standpoint, the extent to which spacer acquisition occurs in different environmental contexts remains to be explored. Finally, metagenomic studies combined with the establishment of new in vivo systems will be instrumental in understanding how spacer acquisition and CRISPR immunity influence the evolution of natural prokaryotic populations in physiologically and ecologically relevant contexts.

The spacer acquisition machinery has begun to be repurposed for technological applications. Rather than being used for directed genome editing like Cas9, the Cas1–Cas2 machinery has been repurposed as a synthetic molecular recorder[79]. This could be used for storing digital information in genomes of bacterial populations[80]. Additionally, an alternative technique has been developed for recording environmental signals encountered by bacterial populations[81]. However, there are currently several factors that limit the utility and function of these techniques. Most notably, the low frequency of spacer acquisition events makes it impossible to have reliable recording within single cells; therefore, the current technologies rely on deep sequencing of large populations to detect spacer acquisition. It is possible that tools such as hyper-Cas9, a hyperactive spacer acquisition mutant[50], could accelerate both basic and technological research in the spacer acquisition field.

The storage of information is a fundamental aspect of all biological systems. CRISPR loci constitute a unique form of biological memory, providing heritable and adaptive immunity to bacteria and archaea. Future research into the biology of CRISPR and spacer acquisition will enhance our understanding of biological memory, prokaryotic evolution and host–pathogen interactions.

Jon McGinn[1] *and Luciano A. Marraffini*[1]*

Laboratory of Bacteriology, The Rockefeller University, New York, NY, USA.

**e-mail: marraffini@rockefeller.edu*

https://doi.org/10.1038/s41579-018-0071-7

Published online 31 August 2018

1. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
2. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
3. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
4. Amitai, G. & Sorek, R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67–76 (2016).
5. Jackson, S. A. et al. CRISPR-Cas: adapting to change. *Science* **356**, eaal5056 (2017).
6. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-Cas systems. *Mol. Cell* **61**, 797–808 (2016).
7. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
8. Krupovic, M., Beguin, P. & Koonin, E. V. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36–43 (2017).
9. Nunez, J. K. et al. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).
10. Nunez, 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).
11. Wang, J. et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* **163**, 840–853 (2015).
12. Wright, A. V. et al. Structures of the CRISPR genome integration complex. *Science* **357**, 1113–1118 (2017).
13. 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).
14. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.* **23**, 876–883 (2016).
15. Nunez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
16. Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. & Pul, U. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* **42**, 7884–7893 (2014).
17. Ivancic-Bace, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.* **43**, 10821–10830 (2015).
18. 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).
19. McGinn, J. & Marraffini, L. A. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. *Mol. Cell* **64**, 616–623 (2016).
20. Weinberger, A. D. et al. Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* **8**, e1002475 (2012).
21. Deltcheva, E. et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

22. Elmore, J. R. et al. Programmable plasmid interference by the CRISPR-Cas system in *Thermococcus kodakarensis*. *RNA Biol.* **10**, 828–840 (2013).
23. Nickel, L. et al. Two CRISPR-Cas systems in *Methanosarcina mazei* strain Go1 display common processing features despite belonging to different types I and III. *RNA Biol.* **10**, 779–791 (2013).
24. Randau, L. RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol.* **13**, R63 (2012).
25. Richter, H. et al. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. *Nucleic Acids Res.* **40**, 9887–9896 (2012).
26. Jones, D. L. et al. Kinetics of dCas9 target search in *Escherichia coli*. *Science* **357**, 1420–1424 (2017).
27. 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).
28. Kupczok, A., Landan, G. & Dagan, T. The contribution of genetic recombination to CRISPR array evolution. *Genome Biol. Evol.* **7**, 1925–1939 (2015).
29. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–2007 (2008).
30. Richter, C. 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).
31. 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* https://doi.org/10.7554/eLife.08716 (2015).
32. Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* **62**, 824–833 (2016).
33. Rollie, C., Graham, S., Rouillon, C. & White, M. F. Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Res.* **46**, 1007–1020 (2018).
34. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105 (2014).
35. Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res.* **43**, 1749–1758 (2015).
36. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).
37. Paez-Espino, D. et al. Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat. Commun.* **4**, 1430 (2013).
38. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L. A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12**, 177–186 (2012).
39. Edgar, R. & Qimron, U. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J. Bacteriol.* **192**, 6291–6294 (2010).
40. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
41. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
42. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642–671 (2008).
43. Yeeles, J. T., van Aelst, K., Dillingham, M. S. & Moreno-Herrero, F. Recombination hotspots and single-stranded DNA binding proteins couple DNA translocation to DNA unwinding by the AddAB helicase-nuclease. *Mol. Cell* **42**, 806–816 (2011).
44. Fagerlund, R. D. et al. Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proc. Natl Acad. Sci. USA* **114**, E5122–E5128 (2017).
45. Künne, T. et al. Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* **63**, 852–864 (2016).
46. Musharova, O. et al. Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* **45**, 3297–3307 (2017).
47. Babu, M. et al. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502 (2011).
48. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202 (2015).
49. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* **5**, 4399 (2014).
50. 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).
51. Hoyland-Kroghsbo, N. M. et al. Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl Acad. Sci. USA* **114**, 131–135 (2017).
52. Patterson, A. G. et al. Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell* **64**, 1102–1108 (2016).
53. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
54. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
55. Elmore, J. R. et al. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev.* **30**, 447–459 (2016).
56. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
57. Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad targeting specificity during bacterial Type III CRISPR-Cas immunity constrains viral escape. *Cell Host Microbe* **22**, 343–353 (2017).
58. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633–637 (2014).
59. Samai, P. et al. Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell* **161**, 1164–1174 (2015).
60. Deng, L., Garrett, R. A., Shah, S. A., Peng, X. & She, Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Mol. Microbiol.* **87**, 1088–1099 (2013).
61. Peng, W., Feng, M., Feng, X., Liang, Y. X. & She, Q. An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Res.* **43**, 406–417 (2015).
62. Kieper, S. N. et al. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* **22**, 3377–3384 (2018).
63. 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 (2018).
64. Shiimori, M., Garrett, S. C., Graveley, B. R. & Terns, M. P. Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell* **70**, 814–824 (2018).
65. Silas, S. et al. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234 (2016).
66. Fineran, P. C. et al. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl Acad. Sci. USA* **111**, E1629–E1638 (2014).
67. Rollins, M. F. et al. Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl Acad. Sci. USA* **114**, E5113–E5121 (2017).
68. Westra, E. R. 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).
69. Semenova, E. 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. USA* **113**, 7626–7631 (2016).
70. Staals, R. H. et al. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
71. Redding, S. et al. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* **163**, 854–865 (2015).
72. Blosser, T. R. et al. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* **58**, 60–70 (2015).
73. Xue, C. et al. CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43**, 10831–10847 (2015).
74. 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).
75. Severinov, K., Ispolatov, I. & Semenova, E. The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Front. Mol. Biosci.* **3**, 45 (2016).
76. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
77. Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7**, e35888 (2012).
78. Mohanraju, P. et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).
79. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
80. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345–349 (2017).
81. 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).
82. Shmakov, S. A. et al. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* **8**, e01397–17 (2017).
83. 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).

Acknowledgements
The authors thank the members of the laboratory for helpful discussions. J.M. is supported by a US National Science Foundation Graduate Research Fellowship. L.A.M. is supported by a Burroughs Wellcome PATH Award, a National Institutes of Health (NIH) Director's Pioneer Award (DP1GM128184-01) and an HHMI-Simons Faculty Scholar Award.

Author contributions
J.M. researched data for the article. J.M. and L.A.M. made substantial contributions to discussions of the content, wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

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

Reviewer information
Nature Reviews Microbiology thanks K. Severinov and the other anonymous reviewer(s) for their contribution to the peer review of this work.