Adaptation in CRISPR-Cas systems

Samuel H. Sternberg[1†], Hagen Richter[2†], Emmanuelle Charpentier[2,3,4,5*] and Udi Qimron[1*]

[1] *Department of Clinical Microbiology and Immunology, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*

[2] *Helmholtz Centre for Infection Research, Department of Regulation in Infection Biology, Braunschweig 38124, Germany*

[3] *Max Planck Institute for Infection Biology, Department of Regulation in Infection Biology, Berlin 10117, Germany*

[4] *The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Department of Molecular Biology, Umeå University, Umeå 90187, Sweden*

[5] *Hannover Medical School, Hannover 30625, Germany*

[†] Equal contribution

*To whom correspondence should be addressed. Emails: charpentier@mpiib-berlin.mpg.de and ehudq@post.tau.ac.il

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) constitute an adaptive immune system in prokaryotes. The system preserves memories of prior infections by integrating short segments of foreign DNA, termed spacers, into the CRISPR array in a process termed adaptation. During the last three years, significant progress has been made on the genetic requirements and molecular mechanisms of adaptation. Here we review these recent advances, with a focus on the experimental approaches that have been developed, the insights they generated, and a proposed mechanism for self versus non-self discrimination during the process of spacer selection. We further describe the regulation of adaptation and the protein players involved in this fascinating process that allows bacteria and archaea to harbor adaptive immunity.

Introduction

Most archaea (~90%) and many bacteria (~50%) encode CRISPR-Cas (clustered regularly interspaced short palindromic repeats – CRISPR associated) systems that are adaptive immune systems against mobile genetic elements (MGEs) (Makarova et al., 2015). The mechanisms involved in immunity rely on small CRISPR RNAs (crRNAs) that guide Cas protein(s) to cleave complementary foreign nucleic acids in a sequence-specific manner (Barrangou et al., 2007; Brouns et al., 2008; Garneau et al., 2010; Hale et al., 2009; Marraffini and Sontheimer, 2008). The hallmark of CRISPR-Cas systems is the CRISPR array that consists of short repeated sequences (repeats) interspersed by unique sequence elements (spacers), which frequently derive from mobile genetic elements such as plasmids and viruses (bacteriophages/phages) (Pourcel et al., 2005; Bolotin et al., 2005; Mojica et al., 2005). The array is preceded by an AT-rich leader containing a promoter. Transcription of the CRISPR array generates precursor RNA molecules that are further processed to generate the mature crRNAs. In this setting, spacer sequences provide the sequence specificity for interference with invading nucleic acids.

The *cas* genes, located proximal to the CRISPR array, encode the Cas proteins that play roles in the different stages of immunity. According to a recently updated classification, CRISPR-Cas systems can be grouped into class I and class II systems, where interference is carried out by multiple proteins or a single effector protein, respectively. Based on the presence of signature proteins, these classes are further subdivided into Types I, III, IV, and II, V, VI systems, respectively (Makarova et al., 2015). Following invasion of an MGE, the CRISPR-Cas system acts in three steps: i) adaptation (or acquisition), in which a new spacer derived from an invading sequence is inserted into the CRISPR array, ii) crRNA biogenesis, in which the CRISPR array is transcribed and the resulting precursor crRNA is processed into mature crRNAs, and iii) interference, in which the foreign nucleic acid is targeted and degraded by a Cas-crRNA ribonucleoprotein complex (Marraffini, 2015).

Barrangou and colleagues (Barrangou et al., 2007) were the first to demonstrate the adaptive feature of CRISPR-Cas immunity, confirming the earlier hypotheses that the CRISPR-Cas system conveys immunity against mobile genetic elements (Makarova et al.,

2006; Mojica et al., 2005; Pourcel et al., 2005). Challenging the bacterium *Streptococcus thermophilus* with phage resulted in the acquisition of phage-originating spacers into the CRISPR array, which provided resistance to matching phages upon further infection. The identification of spacer uptake upon phage challenge proved to be an efficient measure to analyze population dynamics, as several studies of the Banfield lab could demonstrate. Long-term co-culture experiments combined with metagenomic approaches showed that spacer uptake is a key factor that drives phage evolution (Paez-Espino et al., 2015; Sun et al., 2015).

Two modes of adaptation have been reported for Type I systems: i) naïve and ii) primed (Datsenko et al., 2012; Swarts et al., 2012; Yosef et al., 2012). During naïve adaptation, the organism obtains a spacer from a foreign DNA source. In contrast, primed acquisition relies on a pre-existing (priming) spacer that enables a biased and enhanced uptake of new spacers. Both modes are based on the action of two key proteins, Cas1 and Cas2. Naïve adaptation requires only Cas1 and Cas2 (Yosef et al., 2012), whereas primed adaptation additionally requires the Type I interference complex Cascade (CRISPR associated complex for antiviral defense) and the Cas3 nuclease (Datsenko et al., 2012; Fineran et al., 2014; Swarts et al., 2012). Other CRISPR-Cas types encode additional proteins that appear to be involved in spacer acquisition. In the Type I-A system of *Thermoproteus tenax*, a larger complex formed by Cas1, Cas2, Cas4 and Csa1 was reported *in vitro* (Plagens et al., 2012). *In vivo* acquisition in Type I-B system of *Haloarcula hispanica* demonstrated direct involvement of Cas4 in adaptation (Li et al., 2014b). In addition to Cas1 and Cas2, the two proteins Csn2 and Cas9 play essential roles for Type II-A acquisition of *S. thermophilus* and *Streptococcus pyogenes* (Heler et al., 2015; Wei et al., 2015b).

For identifying a suitable protospacer for acquisition, Type I and Type II systems employ short 3–7 base-pair (bp) protospacer adjacent motifs (PAM) (Deveau et al., 2008; Garneau et al., 2010; Horvath et al., 2008; Jinek et al., 2012; Mojica et al., 2009; Sapranauskas et al., 2011). During naïve adaptation in Type II systems Cas9 recognizes the PAM (Heler et al., 2015; Wei et al., 2015b), whereas in Type I-E systems, Cas1 and Cas2 are sufficient for PAM recognition (Wang et al., 2015; Yosef et al., 2012). Following protospacer selection and processing, the acquisition machinery performs site-specific integration of

the new spacer into the CRISPR array at the leader end, concurrent with duplication of the first repeat. Both the leader sequence and the first repeat were shown to be essential, and studies of the *Escherichia coli* Type I-E and the *S. thermophilus* Type II-A systems suggest that the leader-repeat boundary serves as an anchor for spacer integration (Wei et al., 2015a; Yosef et al., 2012).

In this review, we summarize recent advances in our understanding of adaptation in CRISPR-Cas immune systems. We describe the experimental approaches that have been developed to monitor spacer acquisition, and discuss features of naïve and primed adaptation, self- vs. non-self discrimination and regulatory components for the adaptation process. Finally, we highlight some future directions and remaining key questions. Table 1 summarizes the model organisms discussed in the review.

Methods for studying adaptation in CRISPR-Cas systems

Five years after the first demonstration of natural spacer acquisition in *S. thermophilus*, five assays were independently developed for the Type I-E CRISPR-Cas system, establishing a basic framework for later studies, including a subsequent *in vitro* adaptation assay as described below.

To determine the DNA elements and proteins required for adaptation, an assay was developed in which plasmid-encoded Cas1 and Cas2 were co-expressed for 1 to 3 days in an *E. coli* containing a CRISPR array but lacking *cas* genes (Yosef et al., 2012). PCR amplification of a short segment between the leader sequence and the existing spacers on the CRISPR array was conducted using genomic DNA. Gel electrophoresis analysis revealed two major bands: one band of a CRISPR array amplified from DNA of the parental bacteria (no acquisition) and a second band expanded by 61 bp in size, amplified from DNA of bacteria that did acquire a new spacer (Fig. 1A). Sequences of individual acquired spacers were subsequently analyzed. Reliable detection of adaptation could be achieved if 1% or more of the bacteria acquired new spacers. In following studies, gel extraction of the expanded band and high-throughput DNA sequencing allowed the detection of less than 0.01% bacteria that acquired spacers. This enabled the detection of spacer acquisition at more physiological conditions, such as low Cas1 and Cas2 expression

100 (Levy et al., 2015; Savitskaya et al., 2013; Yosef et al., 2013). A modified PCR protocol
101 using donor protospacer-specific primers results in DNA amplification only when the
102 spacer of the corresponding donor protospacer is inserted into the CRISPR array, enabling
103 greater sensitivity and direct comparison of adaptation efficiencies between different
104 potential protospacer sequences (Fig. 1B)(Yosef et al., 2013).

105 Another assay for monitoring spacer acquisition relies on positive selection of bacteria with
106 expanded CRISPR arrays (Diez-Villasenor et al., 2013). By placing the CRISPR array
107 upstream of an out-of-frame antibiotic-resistance gene, spacer acquisition and repeat
108 duplication with subsequent restoration of the open reading frame can be monitored via
109 recovery of an antibiotic resistance phenotype (Fig. 1C). Since rare acquisition events were
110 positively selected, spacer acquisition was detected even under conditions of low Cas1 and
111 Cas2 (Diez-Villasenor et al., 2013).

112 A significant advancement for studying adaptation has been the establishment of an *in vitro*
113 system that monitors spacer acquisition intermediates in the *E. coli* Type I-E system
114 (Nunez et al., 2015b). In reactions with purified *E. coli* Cas1 and Cas2, supercoiled plasmid
115 DNA containing a CRISPR array and dsDNA serving as a spacer donor, acquisition
116 products were observed by agarose gel electrophoresis and subjected to high-throughput
117 DNA sequencing (Fig. 1D). Importantly, the system enabled testing different spacer
118 donors, which had not been possible with the *in vivo* studies (Nunez et al., 2015b).
119 Nevertheless, the assay showed only intermediates of adaptation, i.e. half-site integration
120 rather than fully integrated spacers.

121 The above assays were used to detect naïve adaptation. An assay monitoring *E. coli* Type
122 I-E CRISPR-dependent plasmid curing over time revealed the involvement of interference
123 proteins in primed adaptation (Swarts et al., 2012). Bacteria expressing the entire set of
124 Cas proteins were propagated for 7-14 days. Plating the bacteria with or without antibiotic
125 was then used to monitor plasmid curing. Bacteria cured of the plasmid were analyzed for
126 spacer acquisition by sequencing of the amplified CRISPR array (Fig. 1E). A modified
127 assay monitored spacer acquisition under similar conditions but with bacterial strains that
128 already contained a plasmid-targeting spacer, thereby facilitating primed adaptation

(Savitskaya et al., 2013; Shmakov et al., 2014). This assay enabled high-throughput monitoring of spacer acquisition after just a single overnight growth. Primed spacer acquisition was also observed upon infection by M13 phage when a phage-targeting spacer already existed in the CRISPR array (Datsenko et al., 2012) (Fig. 1F). Taken together, the above-described assays are the major methods to monitor and characterize adaptation.

Adaptation in Type I CRISPR-Cas systems

Naïve adaptation

Using the PCR-based assay described above (Yosef et al., 2012), it was shown that in the *E. coli* Type I-E system, Cas1 and Cas2 are both necessary for spacer acquisition. The DNase activity of Cas1 is required since a $Cas1^{D221A}$ nuclease deficient mutant (Babu et al., 2011), did not support spacer acquisition *in vivo*. The study determined that a single repeat is both necessary and sufficient. By testing two variants of functional repeats followed by sequencing of the newly duplicated repeat, it was shown that the inserted repeat is identical to the leader-proximal repeat, indicating that this repeat is copied during spacer adaptation (Yosef et al., 2012). These experiments further demonstrated that the minimal required length of the leader is 40 to 60 bp upstream of the first repeat of the array, which was later refined to be 40 to 43 bp (Diez-Villasenor et al., 2013).

Protospacers with a flanking 5'-AAG-3' PAM sequence were selected as donors with ~35% frequency indicating that these spacers are overrepresented compared to the PAM frequency (~1.6%). The absence of selection for functional spacers during interference indicates that the acquisition machinery has an intrinsic ability to recognize the correct PAM. Lower expression of Cas1 and Cas2 resulted in more spacers having AAG PAMs (Qimron and colleagues, unpublished). Cascade and Cas3 further increased this stringency due to an unknown mechanism (see below). The GC content of different protospacers did not affect acquisition efficiency (Swarts et al., 2012). However, an AA motif at the end of the spacer sequence did increase adaptation efficiency, as determined by both low and high-throughput analyses (Fineran et al., 2014; Yosef et al., 2013).

Mechanistically, the above mentioned *in vitro* assay using purified Cas1 and Cas2 proteins revealed that dsDNA rather than ssDNA is the preferred substrate for adaptation (Nunez et

al., 2015b). The 3'-OH ends are essential features to make a nucleophilic attack on one strand of the repeat. Cas1 and Cas2 integrate a spacer with the correct PAM orientation by preferentially using the 3'-OH C nucleotide that is complementary to the G of the AAG PAM. These results, along with intermediate adaptation products identified *in vivo* (Arslan et al., 2014), suggest a model for naïve adaptation in the *E. coli* Type I-E system (Fig. 2). Because only reaction intermediates were monitored, some *in vivo* adaptation features were not observed using the *in vitro* assay (Nunez et al., 2015b). Whereas the integration of new spacers occurs adjacent to the leader-proximal repeat *in vivo*, spacer insertion was also observed at other repeats and even outside the CRISPR array *in vitro*. Furthermore, the length of new spacers varied substantially *in vitro*, while a strictly defined length of spacers is observed in natural arrays. The *in vitro* system, however, reflects only the final stages of spacer integration, i.e. the donor DNA is supplied in a ready-to-integrate form, and prior processing steps may specify spacer length and PAM preference (Yosef and Qimron, 2015). Selection and processing of the spacer donor from foreign DNA has yet to be demonstrated and it is not yet known how other proteins might facilitate hand-off of spacer donors to Cas1 and Cas2.

Interestingly, naïve adaptation was not observed in two studies on Type I-B and I-F systems. Here, adaptation strictly required the priming process (Li et al., 2014b; Richter et al., 2014). Given that self-adaptation may pose a serious threat in constitutively active systems, a prerequisite for priming may be reasonable. In addition, it is possible that naïve adaptation occurs in these systems under conditions that have not yet been determined.

Primed adaptation

Primed adaptation is characterized by an increased efficiency of spacer acquisition in the presence of Cas1, Cas2, Cascade, Cas3, and a 'priming' spacer targeting an existing protospacer. In the absence of any of these components, primed adaptation does not occur (Datsenko et al., 2012). In Type I-E system priming enhances acquisition 10 to 20 fold over naïve adaptation (Datsenko et al., 2012; Savitskaya et al., 2013). The overall efficiency of priming is significantly increased when the priming spacer has mutations in the seed sequence or if the protospacer has a non-cognate PAM (Datsenko et al., 2012).

This suggests that priming likely evolved as a mechanism to minimize infection by phage escape mutants that would otherwise evade the interference machinery (Datsenko et al., 2012).

Primed adaptation in Type I-E is biased to the strand orientation matching that of the protospacer targeted by the priming spacer. In early experiments, it was shown that multiple rounds of adaptation exclusively resulted in spacers acquired in the same orientation of the first spacer (Swarts et al., 2012). This observation was later validated in a controlled experiment in which primed adaptation was monitored from two plasmids harboring protospacers in either forward or reverse orientation. Increased spacer acquisition in one strand of the plasmid corroborated the orientation of the protospacer, indicating that acquisition is facilitated from a primed strand (Datsenko et al., 2012). This feature is not conserved in all Type I systems, as both Type I-B and Type I-F systems in contrast to the Type I-E system show primed adaptation in both strands (Li et al., 2014b; Richter et al., 2014). In Type I-F systems, the distribution of acquired spacers exhibits a gradient centered at the targeted protospacer (Fig. 3). A clear gradient is not observed for Type I-B, yet less acquisition in distant regions compared to the primed protospacer was observed.

Two major hypotheses have been proposed for the observed strand selection. The first hypothesis suggests that degradation products, generated by the interference machinery Cascade and Cas3, are preferentially used as spacer donors. The degradation fragments are produced in a defined orientation dictated by the helicase directionality of Cas3 primarily on one strand (Sinkunas et al., 2013; Swarts et al., 2012). This model, however, is difficult to reconcile with the fact that a protospacer with seed or PAM mutations, is less efficiently targeted by the interference machinery than a fully cognate protospacer, yet primes acquisition more efficiently. Another hypothesis suggests that DNA "sliding" by the interference proteins takes place following initial binding to a protospacer (Datsenko et al., 2012). This hypothesis assumes that sliding continues until an appropriate spacer adjacent to a cognate PAM is reached. Spacer acquisition in the region next to the targeted protospacer should be highest and gradually decrease as a function of distance from that spacer. Acquisition centered on the priming protospacer was indeed observed in the Type

I-F system, supporting the sliding hypothesis (Richter et al., 2014). However, both predictions of sliding were not fulfilled in a high-throughput analysis of spacers acquired from a plasmid in a Type I-E system (Savitskaya et al., 2013).

Recent biophysical studies of Type I-E started to shed light on the mechanism of primed adaptation, and in particular, how the interference machinery recruits the acquisition proteins. Early *in vitro* studies showed that Cascade weakly binds to protospacers containing mutated PAM or seed sequence (Semenova et al., 2011; Westra et al., 2013b), suggesting that cells would select against such sequences and/or Cas3 recruitment would fail. Recent single-molecule experiments demonstrated that a non-canonical Cascade binding mode persists at mutated protospacers and may be involved during primed adaptation (Blosser et al., 2015). Even protospacers with a mutated PAM can elicit highly stable Cascade binding (Szczelkun et al., 2014). These studies, together with the identification of a specific intermolecular Cascade-Cas3 interaction (Hochstrasser et al., 2014) indicate that recognition of a consensus PAM by Cascade is required for functional recruitment of Cas3 to promote an interference response (Hochstrasser et al., 2014), leaving open the question how mutated PAMs and/or protospacers elicit priming in a Cascade/Cas3-dependent process.

Real-time single-molecule fluorescence imaging was used to directly visualize Cascade and Cas3 binding to protospacers with either a consensus or mutated PAM that elicits interference or priming *in vivo*, respectively (Redding et al., 2015). The results confirmed that only consensus PAM binding promotes Cas3 recruitment. Strikingly, the addition of Cas1-Cas2 enabled Cascade to specifically recruit Cas3 to protospacers with a mutated PAM. Moreover, with the mutated PAM Cas3 could translocate in both directions, unlike the unidirectional translocation from the protospacer with consensus PAM. Based on these data, it was proposed that Cas3 may travel in complex with Cas1-Cas2, forming a larger spacer acquisition complex. The 3'→5' translocation behavior along both strands could explain the strand bias of newly acquired spacers relative to the priming protospacer (Datsenko et al., 2012; Li et al., 2014b; Richter et al., 2014). Supporting experiments with a Type I-F system revealed a direct interaction between Cas1 and Cas3 (Richter et al., 2012). Early proteomics experiments also revealed an interaction between Cas1 and

multiple subunits of Cascade (Babu et al., 2011), providing a molecular basis by which Cas1-Cas2 might assist in Cas3 recruitment. Recent experiments have similarly provided evidence for direct interactions between Cas9 and Cas1 in the Type II-A system (Heler et al., 2015), highlighting how interference and acquisition machineries are linked in diverse CRISPR-Cas systems.

Regulation of adaptation

Spacer adaptation can be a lethal process if self-spacers are acquired. Regulators of CRISPR-Cas interference that act mostly through transcription were identified in several systems. For example, the *E. coli* Type I-E system is regulated by the cAMP receptor protein (CRP) (Yang et al., 2014), histone-like nucleoid-structuring protein (H-NS) (Pougach et al., 2010; Pul et al., 2010), high-temperature protein G (Yosef et al., 2011) and LeuO (Westra et al., 2010). However, few studies have demonstrated regulation at the level of adaptation activity of the systems.

In *E. coli*, it was shown that both naïve and primed adaptation occur in the absence of the negative regulator H-NS (Swarts et al., 2012). H-NS represses the promoter of the operon that comprises the genes encoding Cascade, *cas1* and *cas2*, as well as transcription of the CRISPR array. H-NS decreases adaptation by its repressive effect on the promoter of the *cas* gene operon. A robust positive regulator of the adaptation system was recently identified in the *Sulfolobus islandicus* Type I-A system (Liu et al., 2015). Csa3a is encoded adjacent to the *cas* operon and was shown to bind two promoters regulating the adaptation genes. Its overexpression resulted in robust naïve adaptation with low specificity for spacer donors flanking cognate PAMs (67-74%). This concurs with low levels observed for naïve adaptation in the *E. coli* Type I-E system (Savitskaya et al., 2013; Yosef et al., 2012).

In the *Pectobacterium atrosepticum* Type I-F system, the CRP protein was shown to positively regulate the expression of the entire *cas* operon by binding to a consensus motif located upstream of the *cas1* gene (Patterson et al., 2015). Activation was cAMP-dependent and required the AMP cyclase gene, *cyaA*. Glucose and gene products such as GalM, which elevates the concentration of the AMP cyclase, reduced the CRP-CyaA-dependent *cas* operon transcription. Deletion of the CRP and CyaA activators decreased primed

adaptation, whereas deletion of the *galM* gene increased primed adaptation. Interestingly, the CRP regulator increases both adaptation and interference in *P. atrosepticum*. It further increases *cas* gene transcription in the *Thermus thermophilus* Type I-E and Type III-A systems, yet represses interference in the *E. coli* Type I-E system. These opposing regulatory roles may reflect the unique niches that different bacteria occupy. The reverse roles played by this regulator, reveal that regulation of CRISPR-Cas systems, even within a same subtype, is complex and not universal to all systems. Despite the above studies, it is still elusive whether adaptation is driven by an invasive genetic element or by a general stress-response-like phenomenon.

Adaptation in Type II CRISPR-Cas systems

The first documented adaptation was shown in *S. thermophilus* (Barrangou et al., 2007). New spacer sequences derived from infecting lytic phage integrated into the CRISPR array. Resulting bacteriophage insensitive mutants (BIMs) were immune to repeated infections. Using strains with a defective *csn2* gene demonstrated that this gene is required for efficient acquisition of new spacer sequences.

Long-term co-culture experiments of *S. thermophilus* with phage 2972 indicated that the uptake of chromosomal spacer sequences resulting in autoimmunity is lethal for the cell as selection against these sequences occurs (Paez-Espino et al., 2013; Wei et al., 2015a). Spacer contents of Type II-A CRISPR arrays are shown to be highly diverse (Horvath et al., 2008; Lopez-Sanchez et al., 2012), yet recent experiments show that a bias towards certain spacer sequences of the phage genome may reflect the effectiveness of an explicit spacer (Paez-Espino et al., 2013), albeit that in long-term experiments a selection against ineffective or un-functional spacer sequences occurs.

Using similar experimental setups as described above for Type I, Terns and co-workers set out to identify *cis*-acting elements involved in spacer acquisition in Type II systems. Similar to the findings of the *E. coli* Type I-E system, the leader and a single repeat are sufficient for efficient spacer uptake (Wei et al., 2015a). Furthermore, 10 bp of the leader sequence at the leader-repeat junction are essential for adaptation and an identified ATTGA motif directly at the leader-repeat junction is highly conserved among the CRISPR-Cas

systems of different streptococci (Wei et al., 2015a). Similar sequence dependencies were observed for the leader proximal nucleotides of the repeat; introduction of single or double nucleotide exchanges affected adaptation, whereas repeat alterations at the leader distal part had no effect. The leader proximal repeat nucleotides are essential for spacer acquisition but have no impact on crRNA biogenesis or interference (Wei et al., 2015a).

Remarkably, Type II also requires Cas9 and tracrRNA as another essential elements during adaptation (Heler et al., 2015; Wei et al., 2015b). Bacteria lacking Cas9 were unable to acquire new spacers, whereas Cas9 availability restored spacer uptake during phage infection (Heler et al., 2015; Wei et al., 2015b). Experiments using either Cas9 of *S. pyogenes* or *S. thermophilus* demonstrated a functional interchangeability of these proteins. The observed PAMs for newly acquired spacers perfectly matched the PAM specificity of the respective Cas9, leading to the hypothesis that Cas9 defines the PAM during adaptation. To confirm this, catalytically inactive Cas9 and variants with mutated PAM recognition residues were tested. Remarkably, catalytically inactive Cas9 still enabled robust spacer acquisition with the correct PAM, whereas new spacers without defined PAM were acquired when the PAM-interacting residues were mutated (Heler et al., 2015; Wei et al., 2015b). Earlier studies in Type I demonstrated that proteins involved in adaptation form large multi-subunit complexes (Nunez et al., 2014; Plagens et al., 2012). Similarly, possible interactions between all four proteins required for adaptation (Cas9, Cas1, Cas2 and Csn2) were indicated for Type II-A (Heler et al., 2015).

In most cases, during a brief period of phage attack, the CRISPR-Cas system must acquire a spacer from an invading phage, generate mature crRNAs, assemble the interference complex and target nucleic acids from the phage in order to prevent lysis. Using defective phages to infect *S. thermophilus* Hynes and colleagues demonstrated that this treatment is similar to classical vaccination (Hynes et al., 2014). Phages damaged by UV-light result in defective phages that are less potent in killing the bacteria. The frequency of spacer uptake was higher upon infections with a defective phage compared to infections using wild-type phage. The defective phage thus appears to “buy” time and allow the CRISPR-Cas system to adapt and mount an immune response.

Adaptation in Type III CRISPR-Cas systems and other pathways

Given the high diversity of CRIPSR-Cas subtypes, it is not surprising that pathways other than the described naïve and primed adaptation exist. Studies of the CRISPR-Cas systems of *Sulfolobales* identified a unique spacer acquisition pattern. The different species of *Sulfolobus* that are under investigation in the Garrett laboratory contain multiple systems of Type I and Type III. Garrett and colleagues were able to observe spacer acquisition after challenging *S. solfataricus* with a mixture of phages isolated from Yellowstone National Park. Interestingly, the sequenced spacers matched to open reading frames of a conjugative plasmid indicative that this spacer uptake was only observed during co-infection with the conjugative plasmid (Erdmann and Garrett, 2012). Analysis of the integration site revealed that most of the spacers were integrated proximal to the leader as was reported previously for Type I (Yosef et al., 2012). A few exceptions were found for one CRISPR locus in which new spacers were inserted at different repeats of the array with the majority being integrated after the fourth repeat. Similar results were obtained when *S. islandicus* was infected with a phage mixture and spacers were only acquired from one of the two phages (Erdmann et al., 2014). In the latter case, adaptation was directly associated to a Type I-A system. Many thermophilic organisms tend to have more than one CRISPR array and it is often difficult to directly assign the arrays to a specific CRISPR-Cas system. It is possible that acquisition events observed in *S. solfataricus* were actually adaptation events associated to the activity of the Type I-A system (Erdmann and Garrett, 2012; Erdmann et al., 2014). In agreement, it was shown that the Type III-B system of *S. solfataricus* exploited the Cas6a enzyme of the Type I-A system to process the Type III-B crRNAs (Deng et al., 2013). The identification of PAMs for the newly acquired spacers in *S. solfataricus* (Erdmann and Garrett, 2012) can be seen as further evidence, since Type III systems do not utilize PAMs during target interference. Furthermore, Type III-B systems often lack an adaptation module (Erdmann et al., 2013; Shah et al., 2013). Therefore, it is tempting to speculate about a cross-talk among these types with respect to adaptation. Interestingly, spacers were not acquired from the wild-type virus that stimulated spacer acquisition from other foreign DNA sources. These observations highlight yet another possible adaptation mechanism that should be further elucidated.

Self versus non-self discrimination during adaptation

Spacers represent potential targets for the interference machinery, as they bear perfect complementarity to the crRNAs they encode. In Type III, the extended base-pair complementarity of crRNAs with the repeats in the array prevents autoimmunity (Marraffini and Sontheimer, 2010). In Type I and Type II, targeting during interference occurs through specific recognition of the PAM adjacent the protospacer (Sashital et al., 2011; Semenova et al., 2011; Westra et al., 2013a). The absence of PAMs in spacer-flanking sequences in the CRISPR array prevents self-recognition. Consistent with early studies proposing that PAM specificity occurs upon spacer selection (Mojica et al., 2009), naïve adaptation assays have revealed that the acquisition machinery also has an intrinsic PAM specificity (Yosef et al., 2012; Yosef et al., 2013). Independent recognition of the PAM during both adaptation and interference may increase the ability to prevent a lethal autoimmune response. This redundancy may further explain the enhanced specificity of spacer acquisition with consensus PAM during primed over naïve adaptation (Savitskaya et al., 2013). PAM specificity during adaptation and interference of Type I are overlapping but non-identical (Fineran et al., 2014; Yosef et al., 2012).

Mutations in the PAM or seed sequence of a protospacer (Deveau et al., 2008; Fineran et al., 2014; Semenova et al., 2011), substantially reduce the binding affinity of Cascade (Semenova et al., 2011; Westra et al., 2013a) and perturb the recruitment and/or cleavage activity of the Cas3 nuclease (Hochstrasser et al., 2014; Rutkauskas et al., 2015) in Type I-E systems leading to evasion of interference. A high-throughput study examining sequence determinants of interference and priming demonstrated that up to 13 mutations within the protospacer and at least 22 mutated PAMs still elicit a priming response in Type I-E (Fineran et al., 2014). Similarly, up to 19 mutated PAMs elicited priming for Type I-B, as well as the 4 consensus PAMs that function for interference (Li et al., 2014a), highlighting the plasticity with which the interference machinery can adapt to escape phage. In the latter study, priming was abrogated when the mutated PAM matched the 3' end of the spacer flanking repeat sequence, explaining how self-priming is specifically avoided.

Naïve adaptation experiments demonstrated that Cas1 and Cas2 of Type I-E prefer spacers from plasmids rather than the chromosome, despite the large excess of chromosomal DNA (Yosef et al., 2012). The chromosomal DNA content was ~25-50 times greater than the plasmid DNA content. Yet, chromosome-derived spacers represented only ~2-22% of the total spacers acquired, depending on Cas1 and Cas2 expression levels. This represents a 100-1000 fold preference in spacer acquisition from plasmid over the chromosome. The observed ratio seems rather due to inherent preference for foreign DNA in the adaptation process, than selective pressure against chromosomal spacer acquisition. In contrast, such preference was not observed for naïve adaptation in the Type II-A system of *S. thermophilus,* indicating that protection from self-immunity may operate via a different pathway (Wei et al., 2015b).

A recent model proposes that two main mechanisms account for this preference (Levy et al., 2015). One mechanism was deciphered after discovering that the recombination/repair complex RecBCD facilitates adaptation. RecBCD is known to interact with an octamer sequence called Crossover Hotspot Instigator (Chi) (Smith, 2012). It was shown that acquisition is significantly higher from protospacers located immediately upstream of a Chi site, suggesting that Cas1 and Cas2 may depend on RecBCD for spacer acquisition, and that Chi sites may attenuate their activity (Levy et al., 2015). The observation that the *E. coli* chromosome contains a ~15-fold overrepresentation of Chi sites compared to plasmids suggests a mechanism for specific avoidance of self-DNA. In fact, spacer acquisition was specifically reduced from a plasmid containing additional Chi sites (Levy et al., 2015) (Fig. 4). The second mechanism explaining foreign DNA preference was revealed when spacer acquisition hotspots were found upstream of replication stalling sites called Ter sites (Levy et al., 2015). During bidirectional DNA replication, Ter sites stall the faster-moving replication fork until the slower replication fork completes replication, allowing for chromosome decatenation (Neylon et al., 2005). DNA nicks and double-strand breaks, which also stall the replication fork, exhibited additional hotspots. These findings suggested that spacer acquisition is highest from DNA regions undergoing frequent replication stalls. The high copy number of plasmids consequently means a greater occurrence of termination replication events, compared to chromosomal DNA replication,

which terminates only once in each growth cycle. Thus, the acquisition machinery should exhibit a natural preference for any high-copy DNA (Fig. 4).

Structural insights into adaptation

The nuclease/integrase function of Cas1 was predicted in 2006 (Makarova et al., 2006), and the adjacency of the *cas1* and *cas2* genes in most *cas* operons led to the hypothesis of functional cooperation between both proteins in spacer acquisition (Makarova et al., 2006; Makarova et al., 2011). Numerous studies have attempted to make inroads into the mechanism of adaptation through structural studies of Cas1 (~30 kDa) and Cas2 (~10 kDa).

Cas1 proteins from six different bacteria and one phage have been crystallized (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009); 4 unpublished). Cas1 adopts a novel fold that can be divided into an N-terminal domain comprising primarily β-strands and a C-terminal α-helical domain (Fig. 5A). Interactions between the N-terminal domains from adjacent protomers promote stable Cas1 dimerization, and the C-terminal domain contains three highly conserved residues (E141, H208, D221; *E. coli* numbering) coordinating a divalent metal ion and forming the putative active site (Wiedenheft et al., 2009). Studies of Cas1 from *P. aeruginosa*, *E. coli* and *A. fulgidus* demonstrated that mutating any of these residues largely eliminated the observed nuclease activity. However, the structures failed to provide significant insights into substrate specificity, and nuclease activity was observed against a range of different substrates, including ssRNA, ssDNA, dsDNA, branched DNA, and plasmid DNA structures (Babu et al., 2011; Kim et al., 2013; Wiedenheft et al., 2009).

Studies of Cas2 similarly show various enzyme activities. An early study detected nuclease activity on ssRNA substrates but not DNA for six different Cas2 homologs (Beloglazova et al., 2008), whereas subsequent studies observed either no nuclease activity (Samai et al., 2010), or degradation of dsDNA (Ka et al., 2014; Nam et al., 2012). Regardless of whether these differences are artifacts or result from the evolutionary distance separating the tested homologs, Cas2 crystal structures presented in these studies, reveal a highly conserved three-dimensional architecture. Cas2 is a small, single-domain protein that exhibits the ferredoxin-like fold also found in Cas5 and Cas6 proteins (Li, 2015), and like Cas1, Cas2 forms a stable homodimer (Fig. 5A).

A breakthrough in the structural understanding of spacer acquisition was the finding that Cas1 and Cas2 assemble into a larger complex, and that complex formation facilitates *in vitro* spacer integration, CRISPR DNA binding, and is critical for adaptation *in vivo* (Nunez et al., 2014; Nunez et al., 2015b). The crystal structure of Cas1-Cas2 from *E. coli* (Nunez et al., 2014) reveals a heterohexameric, crab-like architecture with pseudo-two-fold symmetry, in which a central Cas2 dimer interacts on two opposite faces with separate Cas1 dimers (Fig. 5A). The Cas1-Cas2 interface is stabilized by a combination of electrostatic and hydrophobic interactions. Impairment of this interface perturbs complex formation *in vitro*, as well as spacer acquisition *in vivo*. While purified Cas1-Cas2 showed non-specific DNA binding, perhaps reflecting the flexibility of spacer donor sequences, overexpressed Cas1-Cas2 in cell lysate specifically bound DNA containing a CRISPR array and leader sequence. This suggests that yet unidentified accessory factors may facilitate loading of Cas1-Cas2 onto the CRISPR array for efficient integration. Furthermore, specific CRISPR DNA binding by Cas1 was dependent on the presence of Cas2, suggesting that the primary role of Cas2 may be conformational and involves restructuring of Cas1. In agreement with this hypothesis, spacer acquisition *in vivo* was completely unaffected by Cas2 active site mutations, however abolished by Cas1 active site mutations.

Two recently reported crystal structures of the *E. coli* Cas1-Cas2 complex bound to spacer donor DNA substrates, shedding light on the mechanisms of foreign DNA capture, PAM recognition and CRISPR integration (Nunez et al., 2015a; Wang et al., 2015). The Cas1-Cas2 complex binds a splayed, dual-fork DNA substrate in which 23-bp of dsDNA are flanked by 3' single-stranded overhangs that are threaded into two, symmetry-related Cas1 active sites (Fig. 5B). Tyrosine residues from the same Cas1 monomers (Y22) bracket the double-stranded region *via* stacking interactions, explaining how Cas1-Cas2 acts as a caliper to accurately measure the length of new spacers. These residues may act like a wedge to generate double-strand/single-strand junctions. Wang *et al.* succeeded in crystallizing Cas1-Cas2 with a spacer donor containing the PAM, revealing the molecular details of PAM selection during naïve adaptation (Fig. 5D) (Wang et al., 2015). Cas1 specifically recognizes the PAM-complementary 5'-CTT-3' sequence within the 3' single-stranded overhang region. This helps to position the phosphodiester bond following the C

nucleotide within the Cas1 active site, that spacer donor precursors can be trimmed down to the correct length (Fig. 5B). Finally, using data from alternative crystal forms, Nunez *et al.* proposed an intriguing structural model that explains how Cas1-Cas2 may position spacer donor and acceptor CRISPR DNA to promote the integration reaction (Nunez et al., 2015a).

Conclusions and future directions

Numerous studies have been published on the interference activity of CRISPR-Cas immune systems and only a few dozen studies have been published on the adaptation step. It remains the least understood pathway in CRISPR-Cas immunity. Below, we list several key issues that in our opinion must be resolved for a deeper understanding of the adaptation process.

Spacer biogenesis

How foreign DNA is processed into the spacer donor that is further integrated into the CRISPR array is a question that requires thorough examination. For example, it is yet unclear how spacer donors are excised and loaded onto the Cas1-Cas2 complex. Other questions involve the precise roles of RecBCD and Cas3 during naïve and primed adaptation in Type I-E systems.

Leader recognition

New spacers are mostly integrated at the leader-repeat boundary of CRISPR arrays (Yosef et al., 2012). However, it was shown *in vitro* that spacers can be inserted at sites even outside of the CRISPR array (Nunez et al., 2015b). It is possible that the recognition of the leader-proximal repeat sequence requires additional proteins other than Cas1 and Cas2.

Mechanism of primed adaptation

The priming mechanism is still elusive. A gradient of spacers centered near the primed protospacer was observed in one experimental set-up (Richter et al., 2014). Thus, sliding from the primed site is a reasonable mechanism for primed adaptation. The components of this sliding complex could be Cascade, Cas3, Cas1 and Cas2, or a complex of Cas1, Cas2

and Cas3. In the latter case, the main role of Cascade may be to mediate interactions between the Cas1-Cas2 complex and Cas3. The interplay between these proteins is a major question to be elucidated.

More studies required for Type III adaptation

Knowledge about spacer acquisition in Type III systems is sparse. The typical lack of Cas1 and Cas2 in Type III-B and the co-occurrence of Type III systems together with another Type I or Type II system highlight the potential of cross-talks among these systems. It is noteworthy that the vast amount of these systems exists in bacteria and archaea that are not easy to manipulate or for which the access to phages is limited. Heterologous expression systems are more experimentally tractable and could help to overcome these drawbacks.

Acknowledgements

We thank Moran Goren and Konstantin Severinov for constructive comments and Michael Namestnikov for graphic help. U.Q. is supported in part by the European Research Council Starting Grant 336079, the Israel Science Foundation grant 268/14, and the Israeli Ministry of Health grant 9988-3. E.C. is supported in part by the Alexander von Humboldt Foundation (Professorship), the German Federal Ministry for Education and Research, the Helmholtz Association, the Max Planck Society, the Swedish Research Council and Umeå University. H.R. is a recipient of the Helmholtz Post-doctoral programme.

Author Contribution

S.H.S, H.R, E.C. and U.Q wrote the manuscript.

References

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. Nucleic Acids Res.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A*., et al.* (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol Microbiol *79*, 484-502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W*., et al.* (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem *283*, 20361-20371.

Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Kunne, T., Sobota, M., Dekker, C., Brouns, S.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol Cell *58*, 60-70.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun *3*, 945.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Mol Microbiol *87*, 1088-1099.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol *190*, 1390-1400.

Diez-Villasenor, C., Guzman, N.M., Almendros, C., Garcia-Martinez, J., and Mojica, F.J. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biol *10*, 792-802.

Erdmann, S., and Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol *85*, 1044-1056.

Erdmann, S., Le Moine Bauer, S., and Garrett, R.A. (2014). Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus. Mol Microbiol *91*, 900-917.

Erdmann, S., Shah, S.A., and Garrett, R.A. (2013). SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in Sulfolobus solfataricus P2. Biochem Soc Trans *41*, 1449-1458.

Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A *111*, E1629-1638.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.
Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *519*, 199-202.
Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America *111*, 6618-6623.
Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol *190*, 1401-1412.
Hynes, A.P., Villion, M., and Moineau, S. (2014). Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat Commun *5*, 4399.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science.
Ka, D., Kim, D., Baek, G., and Bae, E. (2014). Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. Biochem Biophys Res Commun *451*, 152-157.
Kim, Y.K., Kim, Y.G., and Oh, B.H. (2013). Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of Pyrococcus furiosus. Proteins *81*, 261-270.
Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature *520*, 505-510.
Li, H. (2015). Structural Principles of CRISPR RNA Processing. Structure *23*, 13-20.
Li, M., Wang, R., and Xiang, H. (2014a). Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. Nucleic Acids Res *42*, 7226-7235.
Li, M., Wang, R., Zhao, D., and Xiang, H. (2014b). Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res *42*, 2483-2492.
Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., She, Q., and Peng, N. (2015). Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. Nucleic acids research *43*, 1044-1055.
Lopez-Sanchez, M.J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., Poyart, C., Rosinski-Chupin, I., and Glaser, P. (2012). The highly dynamic CRISPR1 system of Streptococcus agalactiae controls the diversity of its mobilome. Mol Microbiol *85*, 1057-1071.
Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). 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.
Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nature reviews. Microbiology.
Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55-61.
Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

622 Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR
623 RNA-directed immunity. Nature *463*, 568-571.
624 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif
625 sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*,
626 733-740.
627 Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences
628 of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*,
629 174-182.
630 Nam, K.H., Huang, Q., and Ke, A. (2012). Nucleic acid binding surface and dimer interface
631 revealed by CRISPR-associated CasB protein structures. FEBS Lett *586*, 3956-3961.
632 Neylon, C., Kralicek, A.V., Hill, T.M., and Dixon, N.E. (2005). Replication termination in
633 Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. Microbiol Mol Biol
634 Rev *69*, 501-526.
635 Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign
636 DNA capture during CRISPR-Cas adaptive immunity. Nature.
637 Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-
638 Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.
639 Nature structural & molecular biology *21*, 528-534.
640 Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer
641 acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193-198.
642 Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., Barrangou, R., and
643 Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to
644 phage. Nat Commun *4*, 1430.
645 Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., and Banfield, J.F.
646 (2015). CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus.
647 MBio *6*.
648 Patterson, A.G., Chang, J.T., Taylor, C., and Fineran, P.C. (2015). Regulation of the Type I-F
649 CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. Nucleic
650 Acids Res.
651 Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the
652 CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J
653 Bacteriol *194*, 2491-2500.
654 Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., and
655 Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia
656 coli. Mol Microbiol *77*, 1367-1379.
657 Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire
658 new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for
659 evolutionary studies. Microbiology *151*, 653-663.
660 Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and
661 characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*,
662 1495-1512.
663 Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B.,
664 Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the
665 Escherichia coli CRISPR-Cas System. Cell *163*, 854-865.
666 Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals,
667 R.H., and Fineran, P.C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-
668 independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res
669 *42*, 8516-8526.

Richter, C., Gristwood, T., Clulow, J.S., and Fineran, P.C. (2012). In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System. PLoS One *7*, e49549.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell reports.
Samai, P., Smith, P., and Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta Crystallogr Sect F Struct Biol Cryst Commun *66*, 1552-1556.
Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res.
Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nature structural & molecular biology *18*, 680-687.
Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. RNA Biol *10*.
Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). 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.
Shah, S.A., Erdmann, S., Mojica, F.J., and Garrett, R.A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. RNA biology *10*, 891-899.
Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M.D., Datsenko, K.A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. Nucleic Acids Res *42*, 5907-5916.
Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J *32*, 385-394.
Smith, G.R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol Mol Biol Rev *76*, 217-228.
Sun, C.L., Thomas, B.C., Barrangou, R., and Banfield, J.F. (2015). Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. Isme j.
Swarts, D.C., Mosterd, C., van Passel, M.W.J., and Brouns, S.J.J. (2012). CRISPR Interference Directs Strand Specific Spacer Acquisition. PLoS One *7*, e35888.
Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci U S A *111*, 9798-9803.
Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell *163*, 840-853.
Wei, Y., Chesne, M.T., Terns, R.M., and Terns, M.P. (2015a). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Res *43*, 1749-1758.
Wei, Y., Terns, R.M., and Terns, M.P. (2015b). Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. Genes Dev *29*, 356-361.
Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L*., et al.* (2010). H-NS-mediated repression of CRISPR-based

717 immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol
718 Microbiol *77*, 1380-1393.
719 Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and
720 Brouns, S.J. (2013a). Type I-E CRISPR-cas systems discriminate target from non-target DNA
721 through base pairing-independent PAM recognition. PLoS genetics *9*, e1003742.
722 Westra, E.R., Staals, R.H., Gort, G., Hogh, S., Neumann, S., de la Cruz, F., Fineran, P.C., and
723 Brouns, S.J. (2013b). CRISPR-Cas systems preferentially target the leading regions of MOBF
724 conjugative plasmids. RNA Biol *10*.
725 Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009). Structural
726 basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense.
727 Structure *17*, 904-912.
728 Yang, C.D., Chen, Y.H., Huang, H.Y., Huang, H.D., and Tseng, C.P. (2014). CRP represses the
729 CRISPR/Cas system in Escherichia coli: evidence that endogenous CRISPR spacers impede phage
730 P1 replication. Molecular microbiology *92*, 1072-1091.
731 Yosef, I., Goren, M.G., Kiro, R., Edgar, R., and Qimron, U. (2011). High-temperature protein G is
732 essential for activity of the Escherichia coli clustered regularly interspaced short palindromic
733 repeats (CRISPR)/Cas system. Proc Natl Acad Sci U S A *108*, 20136-20141.
734 Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the
735 CRISPR adaptation process in Escherichia coli. Nucleic acids research *40*, 5569-5576.
736 Yosef, I., and Qimron, U. (2015). Microbiology: How bacteria get spacers from invaders. Nature
737 *519*, 166-167.
738 Yosef, I., Shitrit, D., Goren, M., Burstein, D., Pupko, T., and Qimron, U. (2013). DNA motifs
739 determining the efficiency of adaptation into the Escherichia coli CRISPR array. Proc Natl Acad
740 Sci U S A *110*, 14396-14401.

Table 1. Model organisms discussed in this review.

CRISPR-Cas class	CRISPR-Cas type	organism	naïve / primed	reference
1	I-A	*Sulfolobus solfataricus, Sulfolobus islandicus*	naïve	(Erdmann and Garrett, 2012; Erdmann et al., 2014)
1	I-B	*Haloarcula hispanica*	primed	(Li et al., 2014a; Li et al., 2014b)
1	I-E	*Escherichia coli*	naïve + primed	(Datsenko et al., 2012; Diez-Villasenor et al., 2013; Fineran et al., 2014; Levy et al., 2015; Savitskaya et al., 2013; Shmakov et al., 2014; Swarts et al., 2012; Yosef et al., 2012; Yosef et al., 2013)
1	I-F	*Pectobacterium atrosepticum*	primed	(Richter et al., 2014)
2	II-A	*Streptococcus thermophilus*	naïve	(Barrangou et al., 2007; Deveau et al., 2008; Garneau et al., 2010; Wei et al., 2015a; Wei et al., 2015b)
2	II-A	*Streptococcus agalactiae*	naïve	(Lopez-Sanchez et al., 2012)
2	II-A	*Streptococcus pyogenes*		(Heler et al., 2015)
1	III-B / I-A ?	*S. solfataricus, S. islandicus*	naïve	(Erdmann and Garrett, 2012; Erdmann et al., 2014)

Figure legends

Figure 1. Assays for detecting adaptation. A. Bacteria with plasmid (green) driven expression of Cas1 and Cas2 are grown for 1-3 days. PCR amplification of the CRISPR array using genomic DNA as template followed by gel electrophoresis analysis reveals both parental and expanded arrays. **B.** Same as in A, but using a spacer-specific primer to allow only amplification of expanded arrays. A band on the gel is only observed when a specific spacer was integrated. **C.** A plasmid-based CRISPR array leads to an antibiotic resistance only upon spacer insertion and repeat duplication. **D.** Donor spacer, acceptor plasmid, and Cas1 and Cas2 are mixed in a test tube. Intermediates of adaptation are observed by gel electrophoresis analysis and further investigated by high-throughput DNA sequencing. **E.** Bacteria harboring a plasmid are grown for 7-14 days. Adaptation is recorded by plasmid curing. Antibiotic sensitive bacteria are then selected for spacer acquisition analysis. **F.** Infection of bacteria by a phage results in low naïve acquisition, as detected by PCR analysis (top). Acquisition primed by an existing spacer that targets the invading DNA results in higher adaptation frequency (bottom).

Figure 2. Model for spacer integration into the CRISPR array. Donor DNA (green, PAM in red) is inserted into a CRISPR array in a multi-step process. The processing events that generate the mature donor DNA are unknown. Cas1 and Cas2 catalyze a nucleophilic attack of the 3'-OH in an orientation dictated by the last nucleotide of the PAM. The opposite 3'-OH end likely initiates another nucleophilic attack on the opposite strand of the array followed by gap filling.

Figure 3. Primed adaptation in Type I-E and I-F. Adaptation is significantly enhanced by the presence of Cascade, Cas3 and a 'priming' spacer matching the target DNA. Arrows represent newly acquired spacers. In Type I-E, a gradient of spacers was not detected, yet significantly preferred acquisition from one strand was observed. In Type I-F, a gradient of spacers peaking near the priming protospacer was observed on both strands.

Figure 4. A model for self and non-self discrimination in Type I-E. A. In a RecBCD-dependent mechanism Chi sites reduce the amount of sequences available for spacer donor selection. **B.** Major hotspots for adaptation are replication-stall sites, such as replication termination (lightning symbol). High-copy elements have more replication-stall sites than the chromosome, leading to more frequent acquisition of spacers from these elements.

Figure 5. Structure of the Type I-E acquisition complex. A. Crystal structure of *E. coli* Cas1-Cas2 complex (PDB 4P6I) (Nunez et al., 2014). A Cas2 dimer (light and dark green) is sandwiched by two Cas1 dimers (light blue and grey), forming a ~150 kDa heterohexameric complex. **B.** Crystal structure of the *E. coli* Cas1-Cas2 complex bound to PAM-containing DNA (PDB 5DQZ) (Wang et al., 2015), colored as in A. The DNA contains 23 base-pairs (bp) of dsDNA flanked by ssDNA overhangs. The inset (right) shows how the PAM-complementary 5'-CTT-3' sequence is specifically recognized, positioning the scissile phosphate in the active site.