Viruses, including those that infect bacteria (known as bacteriophages) and archaea, are the most abundant biological agents on our planet. In response to viral predation, bacteria and archaea have evolved a range of defence mechanisms, and many of these protective systems, such as restriction–modification systems (R–M systems), abortive infection and the modification of virus receptors, provide innate immunity. However, the genomes of almost all archaea and of about one-half of the bacteria contain CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) loci, which are responsible for adaptive immunity. The sequences and lengths of CRISPR arrays vary, but they all have a characteristic pattern of alternating repeat and spacer sequences. In addition, CRISPR arrays are usually located adjacent to the cas genes (FIG. 1).

In 2005, three groups recognized that the sequences of some CRISPR spacers were identical to sequences from mobile genetic elements (MGEs), including viruses and conjugative plasmids. In addition, a positive correlation was found between the possession of virus-derived spacers and resistance to the corresponding virus, which suggested that CRISPR loci might participate in a nucleic acid-based immune system. This hypothesis was tested by phage-challenge experiments, which revealed that CRISPR loci acquire fragments of invading DNA and that these new spacers result in sequence-specific resistance to the corresponding phage. Moreover, it was found that the cas genes are required for this process.

Subsequent research has shown that CRISPR-mediated adaptive immunity occurs in three stages: the recruitment of new spacers (known as the acquisition stage), transcription of the CRISPR array and subsequent processing of the precursor transcript into smaller CRISPR RNAs (crRNAs) (known as the expression stage), and crRNA-directed cleavage of invading DNA by the Cas nucleases or other nucleases (known as the interference stage) (FIG. 1). In this Review, we discuss the recent mechanistic insights that have been gained from structural and functional analyses of Cas proteins and CRISPR ribonucleoprotein (crRNP) complexes, which emphasize both conserved and unique features of adaptive immunity in bacteria and archaea.

CRISPR–Cas diversity
CRISPR–Cas systems are highly diverse, which is probably due to the rapid evolution of immune systems as a result of the dynamic selective pressures that are imposed by invading MGEs. Initial comparative analyses of CRISPR loci revealed that there are major differences in CRISPR repeat sequences, in cas gene sequences and in the architecture of the cas operons. On the basis of these differences, CRISPR–Cas systems have been classified into three main types and several subtypes (FIG. 2; Supplementary information S1 (table)). Each type has a specific ‘signature’ Cas protein: type I systems all...
Restriction–modification systems
(R–M systems). Innate defence systems in bacteria and archaea that enable the discrimination of ‘non-self’ DNA from ‘self’ DNA. These systems typically consist of an endonuclease that specifically recognizes and cleaves a short palindromic sequence motif in invading DNA and a methyltransferase that methylates a nucleotide within the same motif in the genomic DNA of the host cell, thereby protecting self DNA from degradation.

Cascade (CRISPR-associated complex for antiviral defence). A multisubunit Cas (CRISPR-associated protein) complex that associates with a CRISPR RNA (crRNA) in type I CRISPR–Cas systems. Recent insights have revealed that the Cascade core is conserved in type III CRISPR ribonucleoprotein (crRNP) complexes.

contain the Cas3 nuclease–helicase, type II systems are defined by the Cas9 nuclease, and type III systems all have Cas10, which is a large protein of unknown function \( ^{12} \) (FIG. 2; Supplementary information S1 (table)). Type I and type III systems seem to be distantly related, whereas type II systems are phylogenetically and structurally distinct \( ^{13} \). In order to target and cleave invading nucleic acid, crRNAs and Cas proteins form crRNP complexes, the nomenclature of which is defined by their composition \( ^{12} \). Type I-A to type I-F crRNP complexes are known as Cascade (CRISPR-associated complex for antiviral defence), whereas all crRNPs in type II systems (that is, type II-A, type II-B and type II-C systems) are known as Cas9 complexes. In addition, type III-A crRNP complexes are known as Csm complexes, whereas those that belong to type III-B systems are known as Cmr complexes.

Type I and type III CRISPR–Cas systems are found in various combinations among phylogenetically diverse bacteria and archaea, whereas the distinct type II systems (sometimes in combination with other CRISPR–Cas types) are restricted to bacteria \( ^{16,14} \). Interestingly, CRISPR–Cas systems have also been found in viral genomes and plasmids \( ^{15–19} \), which is consistent with phylogenetic studies that suggest that these systems are frequently exchanged via horizontal gene transfer \( ^{11} \).
their diversity, all Cas proteins can be grouped into four functional categories (FIGS 1, 2): nucleases and/or recombinases, which are involved in spacer acquisition; ribonucleases, which catalyse the processing of crRNA guides; proteins that assemble with the RNA guides to form the crRNP complexes for target surveillance; and nucleases, which are responsible for degradation of the DNA or RNA targets.

**Acquisition of spacers**

The acquisition of new invader-derived spacers generally proceeds in a polarized manner at the leader-end of the CRISPR locus6,20 (FIGS 1, 3a), which results in a chronological record of previously encountered foreign nucleic acid. The most recent experimental data support the following model for the step-wise acquisition of novel spacers (FIG. 3a). The recognition and fragmentation of invading DNA is likely to be the first step in the process. A recent study reported functional synergy between Cas1 and Cas2, which is consistent with the finding that Cas1 and Cas2 are required for the integration of new spacers28. The strict conservation of Cas1 and Cas2 in all CRISPR–Cas systems suggests that the basic mechanism of CRISPR adaptation is conserved (FIG. 3). Although the simultaneous expression of both Cas1 and Cas2 enables spacer acquisition28, their precise functions in the adaptation process remain elusive. Cas1 is a metal-dependent endonuclease that catalyses the cleavage of double-stranded DNA (dsDNA), single-stranded DNA (ssDNA) and branched DNA in a sequence-independent manner31,32. Crystal structures of the homodimeric Cas1 protein have shown that it consists of an amino-terminal β-strand domain and a carboxy-terminal α-helical domain31,32 (FIG. 3b). The C-terminal domain contains a conserved binding site for a divalent metal ion31, which is crucial for DNA degradation in vitro and spacer acquisition in vivo24. The metal-binding site is surrounded by a cluster of basic residues that form a positively charged strip across the surface of the C-terminal domain. This
Cas2 is a metal-dependent nuclease that contains a RAMP-like fold with a typical β₁,α₁,β₂,α₂,β₃,α₃ arrangement, in which the two α-helices are positioned together on one face of a four-stranded antiparallel β-sheet [31–34, fig. 3b]. The β-sheets from two Cas2 protomers form a β-sandwich, and conserved amino acids are positioned along the dimer interface. The substitution of a conserved aspartic acid residue in each protomer, located at the dimer interface, does not affect their assembly (fig. 3b), but it perturbs the binding of a metal ion and disrupts nuclease activity [35]. Although several studies have reported that Cas2 proteins are endoribonucleases [36], other Cas2 proteins mainly catalyse the cleavage of dsDNA, which indicates that they are deoxyribonucleases [37]. Differences in the loop regions explain differences in substrate preference: for example, Cas2 proteins that have a long loop connecting α₁ to β₁ have a relatively narrow substrate-binding cleft and correspond to ribonucleases. By contrast, Cas2 proteins that have long β₁–α₁ loops contain wider substrate-binding clefts and show deoxyribonuclease activity [38]. A recent study [39] has revealed that Cas1 and Cas2 from E. coli form a stable complex that interacts with the CRISPR locus. The data show that an intact Cas1–Cas2 complex is essential for spacer acquisition in vivo. Importantly, although Cas1 activity is required for protospacer processing and/or spacer integration, Cas2 activity is not needed for spacer acquisition.

Other factors involved in spacer acquisition. In addition to the participation of Cas1 and Cas2, there are indications that a variable set of accessory factors might be involved in spacer acquisition. Pulldown assays have shown that Cas1 of Escherichia coli interacts with RecBCD and RuvB, which are housekeeping proteins that are involved in general DNA repair and recombination [40]. Moreover, several cases of gene fusion and conserved gene clustering suggest that CRISPR acquisition might require additional Cas proteins, such as Csn2, Cas4, Csa1 and Cas3 (Supplementary information S1 (table)). Attempts have been made to verify the putative roles of some of these proteins in CRISPR adaptation, as discussed below.

Csn2 is encoded by all type II-A systems and has been shown to be involved in CRISPR adaptation in Streptococcus thermophilus. Several structural studies have revealed that Csn2 forms a tetrameric ring-shaped complex with a positively charged central cavity that binds to, and slides along, DNA fragments. The apparent lack of Csn2 catalytic activity suggests that it might have an accessory role during spacer acquisition (such as stabilizing the double-strand break during spacer integration) or that it might be involved in the recruitment of additional factors [41].

Cas4 and Csa1 share amino acid sequence similarity with RecB- and AddB-type nuclease–helicases [42, 43]. The Cas4 protein of Sulfolobus solfataricus is a ring-shaped decamer that has DNA-targeting 5’ to 3’ exonuclease activity [44, 45]. In addition, some Cas4 homologues have
Box 1 | ‘Self’ versus ‘non-self’ discrimination by CRISPR–Cas systems

All immune systems must efficiently distinguish ‘self’ from ‘non-self’ to avoid autoimmunity. In DNA-targeting CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) systems, the mechanism of discrimination occurs during CRISPR surveillance. The protospacer itself cannot be used for discrimination, as the crRNA spacer is also complementary to its template in the CRISPR locus on the host chromosome. Instead, in silico analyses of sequences that flank the protospacers recognized by CRISPR–Cas type I and type II systems have revealed that type-specific short sequences (of 2–3 nucleotides), which are collectively known as protospacer adjacent motifs (PAMs)76–78, are necessary for discrimination. The most important feature of the PAM is that it differs from the corresponding sequence of the CRISPR repeat41, which enables discrimination between a non-target and a self non-target. Indeed, experimental analyses of CRISPR interference by type I (REFS 54, 99, 101, 103) and type II (REFS 76, 77) systems have confirmed an important role for the PAM motif. Moreover, studies of CRISPR adaptation in these systems72,25,27,40 indicate that the PAM is also important for spacer acquisition. This makes sense, as only functional protospacers (that is, those that provide immunity) are selected for integration into the CRISPR array.

Type I and type II systems use a ‘non-self activation’ strategy that involves protein-mediated detection of a PAM that is located adjacent to the targeted protospacers in the invading DNA. This eventually results in the ‘switching on’ of interference, most probably by a conformational change that triggers either the recruitment of a nuclease to the crRNA complex (for example, Cas3 in type I systems) or the induction of intrinsic crRNA nuclease activity (for example, Cas9 in type II systems)94,100,101. In type I systems, PAMs are located downstream (at the 3′ end) of the protospacer on the target strand, whereas PAMs of type II systems are located upstream (at the 5′ end) of the protospacer27,22,102. Recognition of PAMs may occur in a single-stranded conformation, which either exclusively involves the strand that base pairs with the crRNA (in type I systems)103 or the displaced strand (in type II systems)91.

Type III systems seem to lack the PAM-based system; instead, the type III-A system uses a ‘self inactivation’ strategy that involves base pairing between the 5′ and 3′ handles of the crRNA (as part of the Csm complex) and the repeat sequence in the CRISPR locus on the host chromosome. Base pairing in this region of the crRNA signals binding to the chromosomal CRISPR array (self DNA), which seems to trigger the ‘switching off’ of the interference process, possibly by preventing the recruitment of the nuclease111.

been reported to have endonuclease activity as well as helicase activity44,45. Fusions of Cas4 and Cas1 occur in several bacterial and archaeal type I and type III systems, which indicates that the two proteins are functionally related12,28. Cas4 from Thermotoga maritima has been shown to form a complex in vitro with a Cas1–Cas2 fusion protein and Cas1 (REF 46). However, such complexes have not yet been isolated from a natural system, which may indicate that the proteins interact only transiently in vitro. Furthermore, it is likely that fusion proteins (such as Cas4–Cas1 and Cas1–Cas2) might contribute to stabilizing these complexes11,46.

Cas3 is a multidomain nuclease–helicase that is fused to Cas2 in type I-F systems49 (Supplementary information S1 (table)). In the type I-F system of Pectobacterium atrosepticum, a direct interaction between Cas1 and the Cas2–Cas3 fusion protein has been observed, which suggests that Cas3 has a dual role, functioning during CRISPR interference as well as during spacer acquisition41. The proposed role for Cas3 during both acquisition and interference might be related to a phenomenon that is known as ‘primed spacer acquisition’ (REFS 25, 49). Priming refers to the positive-feedback loop that accelerates the formation of new spacers from previously encountered genetic elements85. In the type I-E system, this process requires Cas1, Cas2, Cas3 and an RNP complex that is composed of crRNA and multiple Cas proteins (that is, Cascade), which suggests that many proteins participate in this process. However, the mechanism of primed spacer acquisition is currently unknown.

Processing of crRNA guides

Transcription of the CRISPR array generates a long precursor transcript (known as a pre-crRNA) (FIG. 4a), Primary processing of the pre-crRNA involves endoribonuclease cleavage within the repeat sequences, either by Cas6 homologues (FIG. 4b) or by RNase III (FIG. 4c).

Type I and type III systems. In type I and type III systems, Cas6-like nucleases are responsible for the primary processing of the pre-crRNA (FIG. 4b), which has been extensively reviewed elsewhere44. Although Cas6 homologues generally consist of two RAMP domains, a considerable level of structural variation has been described111.

Despite the structural variability, Cas6 enzymes specifically cleave the pre-crRNA by hydrolysing a single phosphodiester bond in the repeat sequences of the transcript. This typically results in crRNAs that have a repeat-derived 5′ handle of 8 nucleotides, followed by the complete spacer sequence and a repeat-derived 3′ handle of variable size that forms a hairpin structure in some systems (FIG. 4b). Cas6 variants are metal-independent endoribonucleases that generate crRNAs that have a 5′ hydroxyl group and either a 3′ phosphate or a cyclic 2′–3′ phosphate52–55. This suggests that Cas6 has a general acid–base catalytic mechanism, in which the deprotonated hydroxyl at the 2′ position of the ribose functions as a nucleophile. The catalytic sites of all characterized Cas6-like enzymes are composed of an invariant histidine residue, a tyrosine or serine residue and, in some cases, a lysine residue. However, the relative positions of these residues are poorly conserved (FIG. 4d), which might explain the observed functional variations in Cas6 activity (see below)56,57.

In the type I-F and type I-F systems, the Cas6 proteins (which are known as Cas6e and Cas6f, respectively) have a high affinity for the cleaved crRNA product, which results in single-turnover reaction kinetics96. This is consistent with the observation that Cas6e and Cas6f are core components of Cascade complexes, in which they remain firmly associated with the hairpin at the 3′ handle of the crRNA11,56,58–65 (FIG. 4d). By contrast, the Cas6 variants of other type I systems and all type III systems function as stand-alone nucleases that deliver primary crRNAs to the respective crRNAP complexes15,56. After transfer of the crRNAs to these complexes, the 3′ handles are accessible for nucleolytic trimming65. This secondary processing seems to result in short 3′ handles in type I-A, type I-B, type I-C and type I-D systems. In type III systems, differential trimming typically results in the production of two mature crRNA species that differ by 6 nucleotides66–70 (FIG. 4b).

In type I-C systems, Cas6 is substituted by a Cas5 variant (known as Cas5d; FIG. 4c) that cleaves the pre-crRNA to form the mature crRNA71–73. The resulting
crRNA contains an 11 nucleotide 5′ handle (rather than the 8 nucleotide handle that is generated by Cas6) and a 21–26 nucleotide 3′ handle. Similarly to Cas6, Cas5d is a RAMP protein with an active site that is composed of a catalytic triad (containing tyrosine, lysine and histidine), and like Cas6e and Cas6f, Cas5d remains associated with its crRNA product and assembles with other Cas proteins to form the multisubunit Cascade complex. In all other type I systems, catalytically inactive Cas5 homologues are a subunit of Cascade and have been proposed to interact with the 5′ handle of the crRNA (see below).

**Type II systems.** In type II systems, processing of pre-cRNA relies on a completely different mechanism (FIG. 4c). In addition to a cas operon and a CRISPR array, the CRISPR locus of these systems includes a gene that encodes a transactivating crRNA (tracrRNA). The crRNA contains a 25 nucleotide sequence that is complementary to the repeat region of the pre-cRNA transcript. Base pairing between these two RNAs results in a double-stranded RNA that is cleaved by RNase III (FIG. 4c). Cas9 is required for primary crRNA processing, most probably for binding and positioning the RNA molecules for cleavage by RNase III (REF. 14). After initial processing by RNase III, the crRNA–tracrRNA hybrid remains firmly associated with Cas9 (REFS 76, 77). The 5′ end of the crRNA spacer (which is 24–27 nucleotides in length) is trimmed by an unknown nuclease, which typically results in a spacer that is 20 nucleotides long. The mature crRNA–tracrRNA hybrid is required for target interference — probably for proper anchoring and positioning of the crRNA in Cas9 — in a way that might be analogous to the binding of the crRNA hairpin in...
Cascade. In addition, recently obtained crystal structures of Cas9 reveal that major domain rearrangements occur following the binding of target nucleic acids (see below).

Assembly of crRNAP complexes

Mature crRNAs associate with Cas proteins to form stable crRNAP complexes. Type I systems form multisinuotide surveillance complexes that are called Cascade (FIG. 2). Although the proteins associated with the type III crRNAPs are phylogenetically distinct from those in the type I system, recent structural studies have shown that there are striking architectural similarities between the type III and type I crRNAPs. By contrast, the type II crRNAP complex (in which Cas9 is the only protein component) is fundamentally different (REF. S1; Supplementary information S1 (table)).

Type I crRNAP complexes. The first crRNAP complex to be identified was the Cascade complex of the E. coli type I-E system. Owing to striking structural similarities, the crRNAP complexes of other type I variants are generally referred to as Cascade (REFS 34,35). The type I-E crRNAP (Cascade/I-E) from E. coli is composed of a core complex (Cas5, Cas6, Cas7 and a single 61 nucleotide crRNA) and two less tightly associated subunits (Cse1 and Cse2) (REFS 36,37). The complete complex has an uneven subunit stoichiometry: (Cse1)2-(Cse2)2-(Cas5)2-(Cas7)2-(Cas6)2 (REFS S2, S5) (FIG. 5a), which is a typical feature of all type I and type III crRNAP complexes and is controlled by differential translation of the encoding polycistronic mRNA (REF. 38). The overall architectures of Cascade complexes of type I-C and type I-F systems share a helical backbone structure that is composed of Cas7 (known as Cys7 in type I-F systems), Cas8 (known as Cys1 in type I-F systems), Cas5 (known as Cys2 in type I-F systems) and a crRNA (FIG. 5a). Interestingly, in vitro assembly of Cascade/I-A has shown that, in addition to Cas7, Cas8, Cas5 and Cas5 (known as the small subunit; Supplementary information S1 (table)), the truncated domains of Cas3 (the helicase domain, which is known as Cas3′) and the nuclease domain, which is known as Cas3′′) form part of the complex (FIG. 5a). Thus, despite many similarities, the occurrence of structural differences suggests that there are minor functional variations.

A major advance in our understanding of crRNA-guided surveillance came from two cryo-electron microscopy (cryo-EM) structures (of 8–9 Å resolution) of E. coli Cascade/I-E (FIG. 5b). These structures revealed an overall seahorse-shaped architecture in which the 3′ and 5′ handles of the crRNA are anchored at opposite ends of the complex and the 32 nucleotide spacer sequence is displayed along the helical backbone, which is composed of six Cas7 proteins. This is a typical feature of Cascade complexes, in which a string of Cas7 subunits provides a backbone that has an elongated binding cleft for the crRNA guide (FIG. 5a). Cas7 of Cascade/I-A has a crescent-shaped structure that contains a central RAMP domain. Although the Cas7 RAMP domain resembles a typical RNA-recognition motif (RRM), it seems to lack some of the conserved aromatic residues that are responsible for RNA binding by canonical RRMs. By mapping the highly conserved residues onto the three-dimensional structure, two conserved clusters were identified on the concave surface of the Cas7 structure that is involved in binding to the crRNA (REFS 39,40).

Type II crRNAP complexes. The recently established high-resolution structures of Cas9 have been a major breakthrough in the field (REFS 80,81). Crystal structures have been obtained of two Cas9 proteins (from the type II-A system of Streptococcus pyogenes and the II-C system of Actinomyces naeslundii) in the absence of nucleic acids (REF. 40). In a second study, the structure of Cas9 from a type II-A system (from S. pyogenes) was solved, with a single-guide RNA (sgRNA; which is an engineered functional fusion of crRNA and tracrRNA) hybridized to a 20 nucleotide DNA target (Supplementary information S2 (figure)). These structures show that Cas9 has a conserved architecture that consists of two distinct lobes: the α-helical recognition lobe, which is primarily involved in coordinating the guide RNA, and the nuclease lobe, which is responsible for PAM recognition and subsequent cleavage of the target DNA (see below) (Supplementary information S2 (figure)). In the apo state, the lobes are oriented in a position that would preclude the binding and cleavage of target DNA. Complementary single-particle electron microscopy reconstructions show that the two structural lobes undergo a reorientation following binding of the sgRNA, which results in the formation of a central channel that allows for the binding of DNA substrates (FIG. 5b). This is in agreement with the structure of Cas9–sgRNA with a single-stranded DNA target, in which the RNA–DNA heteroduplex is located in the positively charged groove at the interface of the two lobes (FIG. 5b; Supplementary information S2 (figure)). The observation that the catalytic sites of the two nuclease domains — HNH and RuvC — are not positioned properly for cleavage may suggest that an R-loop configuration is required to reach the ultimate cleavage-competent state of Cas9.

Type III crRNAP complexes, and similarities with type I systems. Structures of the type III-A crRNAP complex from S. solfatarius (known as the Csm complex (REF. 76)) and two type III-B crRNAP complexes from Pyrococcus furiosus and Thermus thermophilus (known as Cmr complexes; FIG. 5b) were recently determined by electron microscopy (REFS 74,75). Type III complexes have a multicopy backbone (which is composed of Csm3 in type III-A systems and Cmr4 in type III-B systems; FIG. 5a, b) that is morphologically similar to the Cas7 backbone of type I Cascade complexes (FIG. 5a, b). Indeed, the crystal structure of Csm3 has revealed that it is a structural homologue of Cas7 (REF. 90) (FIG. 5a, b). Moreover, cryo-EM structures of the Cmr complex from P. furiosus have shown that the crRNA is positioned along the backbone, similarly to the crRNA in Cascade (REF. 79).

To denote potential structural and/or functional similarities between components of type I and type III crRNAP complexes, the terms ‘large subunits’ and ‘small subunits’ have been introduced (REF. 84). In most type I systems,
the large subunits are Cas8 homologues (such as Cas8a, Cas8b, Cas8c, Cse1 and Csy1), whereas in type III systems, the large subunits are Cas10 homologues (such as Csm1 and Cmr2) (Supplementary information S1 (table)). The large subunit in type I Cascade complexes is positioned adjacent to Cas5, close to the 5’ handle of the crRNA (FIG. 5a, b). Similarly, native mass spectrometry of a type III-A Csm complex revealed that the large subunit Csm1 (which is a Cas10 homologue) interacts with Csm4 (which is a Cas5 homologue) (FIG. 5a, b). In addition, evidence for a similar interaction was obtained from a crystal structure of a partial Cmr complex that consisted of Cmr2 (a Cas10 homologue) and Cmr3 (a Cas5 homologue) (FIG. 5a, b). Comparison of the crystal structures of the type I-E subunit Cse1 (REFS 84, 94) (which is a Cas8 homologue) and the type III-B subunit Cmr2 (REFS 95, 96) (which is a Cas10 homologue) showed that these two proteins do not share obvious connections.
structural similarity\textsuperscript{24}. However, the conserved position of the large subunits in type I and type III crRNA complexes, as well as the apparent substitution of Cas8 by Cas10 in type I-D crRNA complexes (Supplementary information S1 (table)), suggests that they have analogous roles.

The small subunits are Cas5 and Cse2 in type I systems and Csm2 and Cmr5 in type III systems (FIG. 5a; Supplementary information S1 (table)). The absence of genes that encode small subunits in certain cas operons (for example, types I-B, I-C, I-D and I-F) has been proposed to be compensated for by extensions of the large subunits\textsuperscript{91,92}. A structural comparison of small subunits has shown that there is structural conservation between Cmr5 (from the type III-B Cmr complex) and the N-terminal domain of Cse2 (from Cascade/I-E), and between Cas5 (from Cascade/I-A) and the C-terminal domain of Cse2 (from Cascade/I-E)\textsuperscript{99}. In Cascade/I-E, the Cse2 dimer constitutes a protein bridge that connects Cas6 in the head of the complex to Cse1 in the tail of the complex (FIG. 5a, b). Moreover, electron microscopy structures of Csm and Cmr complexes have shown that the small subunits constitute a second helical string of subunits, which run parallel to the Cas7 backbone and form a solid bridge that connects the ‘bottom’ (large subunit and the Cas5-like proteins Csm4 or Cmr3) to the head (Csm5 or Cas1 and Cmr6) in all type I and type III complexes. The set of small subunits in type III complexes consists of three copies of Csm2 in the type III-A complex\textsuperscript{94} and three copies of Cmr5 in the type III-B complexes\textsuperscript{89,97} (FIG. 5a).

Target surveillance and interference

The targeting of invading MGEs by the different crRNA complexes seems to proceed in a stepwise manner\textsuperscript{13,24,34,38,89} (FIG. 6). Finding a protospacer sequence that is complementary to the crRNA involves scanning of the invading DNA, discriminating self from non-self (BOX 1) and base pairing between the 7–8 nucleotide seed region of the spacer and the complementary protospacer, followed by extended base pairing between the spacer and protospacer, which eventually results in complete strand displacement\textsuperscript{82,100}. Hybridization of crRNA to the target strand generates an R-loop structure, which — at least in some cases — has been shown to trigger a conformational change in the crRNA backbone\textsuperscript{72,74,79,80}. This structural transition may function as a signal that recruits a trans-acting nuclease (for type I and type III-A systems) or lead to the activation of intrinsic nuclease activity (for type II and type III-B systems) (FIG. 6).

Type I systems. Scanning of invader DNA seems to be strongly dependent on nonspecific interactions between the crRNA complex and the invading DNA. In type I-E systems, the Cse1 subunit of Cascade is important for nonspecific association with DNA and also has a key role in preventing autoimmunity (the targeting of chromosomal sequences)\textsuperscript{22,84,101}. Structural and biochemical studies suggest that a flexible loop in Cse1 interacts with the 3 nucleotide PAM motif\textsuperscript{84,94}. The PAM is an antigenic signature, and PAM recognition by Cse1 might destabilize the DNA duplex, thereby enabling the crRNA to access the target DNA for hybridization\textsuperscript{84}. For successful interference, base pairing between the seed region of the crRNA spacer and the complementary target protospacer is essential at positions closest to the PAM. In Cascade/I-E, the seed region of the crRNA includes nucleotides 1–5 and 7–8 at the 5’ boundary of the spacer\textsuperscript{89}. Mutated targets, in which base pairing within this seed region is imperfect, generally escape detection by Cascade\textsuperscript{99,104}. In the case of PAM recognition and successful seed base pairing, crRNA-guided strand invasion of the dsDNA proceeds in an ATP-independent manner, which generates an R-loop that might be stabilized by the positively charged surface of the two small subunits (Cse2)\textsuperscript{94,98}.

The formation of a complete R-loop coincides with a major conformational change of the Cascade/I-E complex\textsuperscript{24,74} and local bending of the target DNA\textsuperscript{86,105}. In type I systems, these structural changes seem to trigger the recruitment of the Cas3 nuclease–helicase\textsuperscript{REF. 103}. Single-particle EM reconstructions of dsDNA-bound Cascade have recently shown that Cascade positions the kinked DNA duplex in such a way that the PAM motif contacts the Cse1 subunit\textsuperscript{89}. Interestingly, this study showed that docking of Cas3 on Cascade occurs at a site that is provided by Cse1, in close proximity to Cas5 (FIG. 5b). Although the observed density for Cas3 in this reconstruction only corresponds to part of Cas3 (REF. 89), its binding site on Cascade is consistent with the occurrence of some natural Cas3 fusions, such as Cas3–Cse1 in type I-E systems\textsuperscript{103} and Cas5–Cas3 in type I-B systems (J.v.d.O, unpublished observations).

Cas3 consists of a HD-nuclease domain\textsuperscript{104,105} (Supplementary information S3 (figure)) fused to an SF2-helicase domain (superfamily 2 helicase domain)\textsuperscript{87}. In Cascade/I-E, it has been shown that the aforementioned loop in Cas1 is not only involved in PAM recognition but is also required for Cascade-associated Cas3 nuclease activity\textsuperscript{89}. The ATP-dependent helicase activity of Cas3, combined with its metal-dependent nuclease activity, mediates complete degradation of the target DNA\textsuperscript{54,105}. After initial endonucleolytic cleavage of the displaced strand of the R-loop, exonucleolytic degradation proceeds in the 3’ to 5’ direction\textsuperscript{14,103,104} (FIG. 6a). The other DNA strand undergoes endonucleolytic and exonucleolytic degradation\textsuperscript{14,103,104}.

Type II systems. The interference mechanism of type II systems is completely different from that of type I and type III systems (FIG. 6b). In type II systems, interference is mediated by the Cas9–RNP complex that consists of Cas9 and two RNAs (that is, crRNA and tracrRNA)\textsuperscript{75,77,107} (FIG. 5a, b). Targeting of a complementary DNA fragment by Cas9 has recently been shown to proceed in a step-wise manner\textsuperscript{106}. As mentioned above, loading of the sgRNA triggers a structural rearrangement that leads to the formation of a central channel that binds to the target DNA\textsuperscript{98}. Next, the Cas9–RNA complex scans the DNA for a PAM motif (BOX 1; FIG. 6b). Scanning and identification of the target DNA by type II systems seems to be a mirror image of the initial steps.
Figure 6 | Surveillance and interference by crRNP complexes. Proposed mechanisms of targeting for the three different types of CRISPR–Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) systems. a | In type I systems, the Cascade complex searches for a complementary protospacer in the invader DNA via target scanning. The large subunit (Cse1 or Cas8) of the complex recognizes the protospacer adjacent motif (PAM) sequence by a ‘non-self activation’ strategy (BOX 1), which is followed by hybridization between the seed sequence and the target DNA. If these initial criteria are met, complete base pairing results in R-loop formation and a simultaneous conformational change in the Cascade complex, which probably triggers Cas3 recruitment and subsequent degradation of the displaced target DNA strand (red triangles indicate endonucleolytic cleavage). The dashed arrow indicates processivity by the concerted helicase (green triangle) and exonuclease activities in the 3’ to 5’ direction.

b | In type II systems, the Cas9 complex, bound to the CRISPR RNA (crRNA)–transactivating crRNA (tracrRNA) duplex, follows a similar mechanism of PAM-dependent recognition of invading DNA. However, unlike type I systems, the PAM is located upstream (at the 5’ end) of the protospacer and both target DNA strands are cleaved by Cas9-mediated nuclease activity.

c | In type III-A systems, the Csm complex targets DNA in a PAM-independent process, using a ‘self inactivation’ strategy (BOX 1). The stand-alone nuclease that is responsible for DNA degradation has been proposed to be Csm6 (also known as Csx1) (FIG. 2; Supplementary information S1 (table)).

d | In type III-B systems, the crRNA-guided Cmr complex targets invading RNA in a PAM-independent process. After recognition and hybridization of crRNA and a complementary target RNA sequence, cleavage of this target occurs at multiple sites (red triangles). The nuclease that is responsible for RNA degradation has been proposed to be a subunit of the Cmr complex (Cmr4; Supplementary information S1 (table)). With the exception of type I systems, in which Cas3 mediates target degradation (part a), all other systems (parts b–d), are thought to involve non-Cas nucleases for complete target degradation.
of type I interference; the PAM motif resides on the displaced strand of the target DNA and is located close to the 3′ handle of the crRNA guide. Two tryptophan-containing flexible loops in the C-terminal domain of the Cas9 nuclease lobe are involved in PAM recognition, and mutation of these residues affects both the binding and cleavage of target DNA. Interaction with the PAM motif is required for DNA binding, and subsequent DNA strand displacement and R-loop formation initiate at the PAM. Base pairing progresses over a 12 nucleotide seed sequence towards the distal end of the target sequence, with little room for mismatch (Fig. 6b). During the final stage of interference, Cas9—which has adopted a cleavage-competent state (as described above)—uses its intrinsic nucleolytic activity to cleave the DNA. Endonuclease activity is catalysed by two active sites at separate locations in the nuclease lobe. The HNH-like nuclease domain cleaves the DNA strand that base pairs with the crRNA, and the RuvC-like nuclease domain cleaves the displaced DNA strand. Cas9-mediated nuclease activity results in a blunt double-stranded end at a specific site, which is typically 3 nucleotides from the 3′ end of the protospacer (Fig. 6b). Cas9 has been reported to be a single-turnover enzyme.

Type III systems. Csm complexes of type III-A systems typically consist of at least five distinct proteins (Csm1–Csm5) (Fig. 2; Supplementary information S1 (table)) and crRNAs with conserved 8 nucleotide 5′ handles and variable 3′ ends. The mechanistic details underlying the targeting of complementary protospacers by type III crRNPs remain elusive, but some information is emerging. Type III-A systems discriminate self DNA from non-self DNA in a PAM-independent manner (BOX 1), with the possible involvement of the Cas10-like protein Csm1 (that is, the large subunit). Similarly to some type I systems, the 5′ end of the spacer region of the crRNA has been proposed to contain a seed sequence (Fig. 6c). The type III-A system has been shown to target DNA in vivo; however, no in vitro DNA-degrading activity has so far been reported for a purified Csm complex. Genetic analysis suggests that an additional protein, Csm6 (also known as Casx1), is required for interference. It is tempting to speculate that Csm6 is a helicase and/or nuclease that is recruited for DNA interference, analogous to Cas3 in type I systems (Fig. 6a, c), but this remains to be determined.

Cmr complexes of type III-B systems consist of at least six distinct proteins (Cmr1–Cmr6) (Fig. 2; Supplementary information S1). They are unique among CRISPR–Cas systems in that they target RNA rather than DNA. It has recently been shown that the RNA target is cleaved by the T. thermophilus Cmr complex at 4–5 distinct sites (Fig. 5d). Cleavage of the target RNA occurs processively, in the 3′ to 5′ direction, in 6 nucleotide intervals. This sequential slicing of the target RNA has been confirmed for two additional Cmr complexes (M. Terns and S. Bailey, personal communication). Whether or not there is a seed-like sequence at the 5′ end of the spacer remains to be determined. The regularly spaced cleavage pattern of Cmr complexes indicates that there are multiple catalytic sites along the backbone of the complex, which suggests that Cmr4 might be the catalytic subunit, as discussed below.

Two Cmr complexes (Cmr-A and Cmr-β) are encoded by different gene clusters in Sulfolobus islandicus. Unexpectedly, the Cmr-A complex (which is composed of Cmr1–Cmr6) has been reported to target plasmid DNA in vivo, using a Cas3-dependent mechanism that requires the transcription of its target; this seems to be a functional analogue of the aforementioned Csm complex of III-A systems. By contrast, the S. solfataricus Cmr-β complex (which is composed of Cmr1–Cmr7) targets RNA in vitro; however, the reported catalytic mechanism differs substantially from that which has been described for the III-B systems of P. furiosus and T. thermophilus. These studies suggest that there is further mechanistic and/or functional diversity among type III systems.

Heteroduplex formation. Among the different CRISPR–Cas types, the molecular details of crRNA binding by crRNPs differ substantially. As mentioned above, Cas5 of Cascade I-E is most probably involved in binding the 5′ handle of the crRNA. At the other end of the crRNA, the 3′ hairpin (which consists of a 6 base pair stem and a 4 nucleotide loop) is firmly bound by the Csaε nucleosome subunit (Fig. S5). These stable interactions at both ends of the mature crRNA constrain base pairing with target nucleic acids to 5–6 helical segments that are each separated by short non-helical sequences (Supplementary information S4 (figure)). This type of crRNA–target interaction is supported by biochemical analysis and genetic data, which indicate that some nucleotides in the spacer do not base pair and, as such, are not essential for target recognition. Interestingly, this base-pairing pattern is reminiscent of the DNA–DNA interaction that is mediated by the RecA protein, which forms a nucleoprotein filament during homologous recombination. Strands of RecA molecules, which are complexed with a ssDNA template, invade a dsDNA helix. After the displacement of the non-complementary strand, the newly formed hybrid of complementary DNA strands is globally underwound and stretched but is locally allowed to adopt a classic B-form conformation, which resembles the crRNA–target hybrid configuration of the Cascade R-loop (Supplementary information S5 (figure)).

In the Cascade complexes of type I-A and type I-C systems, as well as in the crRNPs of type II and type III systems, crRNAs are bound by only a single handle. This probably results in increased flexibility of the crRNA and possibly leads to a different structure of the crRNA–target heteroduplex. Indeed, the structure of Cas9 reveals that there is complete base pairing between the sgRNA and the 20 nucleotide target DNA, which resembles the crRNA–target hybrid configuration of the Cascade R-loop (Supplementary information S2 (figure)).

In RNA targeting by type III-B systems, the situation might resemble eukaryotic RNA interference (RNAi), in which an Argonaute nuclease typically uses a 21 nucleotide guide RNA with
a firmly anchored 5’ end to recognize a complementary mRNA target. Interactions between the Argonaute complex and a complementary target nucleic acid trigger the release of the 3’ end of the guide RNA, which results in the formation of at least 15 contiguous base pairs in an A-form conformation.118

Conclusions and outlook
Since the publication of the landmark paper by Barrangou et al.,1 which describes the discovery of the CRISPR–Cas system as a prokaryotic adaptive immune system, impressive progress has been made with respect to understanding many of the unique mechanistic features that are associated with these remarkable systems. Although in silico studies initially showed that there was overwhelming variation in CRISPR–Cas systems, subsequent comparative sequence analyses resulted in the identification of three major types12 and in potential scenarios for the evolution of CRISPR–Cas variants.13 Moreover, recent structural analyses of Cas proteins and crRNPs have revealed that there are unanticipated similarities between the type I and type III crRNPs. In contrast to these multisubunit complexes, the single-protein Cas9–crRNA complex from type II systems is structurally unrelated to other crRNA complexes, and there are major mechanistic differences at the levels of CRISPR expression and interference.

Outstanding gaps in our mechanistic understanding of CRISPR–Cas functionality include the molecular details of the spacer acquisition process, particularly the role of Cas1 and Cas2, the involvement of Cas3 and/or Cas4 and assistance by general repair and/or recombination enzymes. At the level of CRISPR interference, the general picture has become clear, but several relevant details regarding self versus non-self discrimination, transcriptional activation and repression of synthetic regulators, transcriptional activation and repression of synthetic regulators, and genome editing in eukaryotic cells, ranging from yeast to plant and from zebrafish to human.14 In addition, RNA targeting by type III-B systems (as well as by Cas9) may function as an alternative system for the directed silencing of gene expression.113 In terms of applications of CRISPR-associated nucleases in general, and Cas9 in particular, the sky seems to be the limit. However, even for Cas9, there is still room for improvement — for example, by lowering the stringency of its PAM dependence and reducing its off-target cleavage. We anticipate that fundamental details of CRISPR–Cas structure and function will not only further improve our understanding of these unique defence systems but will also be crucial for optimizing and further expanding the applicability of CRISPR–Cas systems.

23. This paper reports the discovery of the PAM motif.
Characterization of the CRISPR/Cas subtype with an iron–sulfur cluster.

NATURE REVIEWS Microbiology
VOLUME 12 | JULY 2014 | 491

© 2014 Macmillan Publishers Limited. All rights reserved
Mediated adaptive immune systems in bacteria and USA
governed by a seed sequence. Interspaced short palindromic repeat (CRISPR) RNA is
bending of negatively supercoiled DNA.
Makarova, K. S., Wolf, Y. I. & Koonin, E. V. The basic
Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. Cell 159, 945–956 (2014). This study provides evidence that RNA (rather than DNA) is targeted by type III-B systems.

Acknowledgements
The authors thank members of the van der Oost laboratory (R. Staals and S. Brons) and the Wiedenheft laboratory (specifically M.-C. Rollin and S. Golden) for discussions and critical reading of this manuscript. This work was supported by a Netherlands Organisation for Scientific Research (NWO) Earth and Life Sciences (ALW-TOP) grant (grant number 854.10.005) and NWO ALW Open competition to J.O. (grant number 820.02.003). E.R. was received from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 273606. Postdoctoral training for R.N.J. is supported by the US National Institutes of Health (grant number R01GM108888) N.I.M. is supported by the US National Institutes of Health (grant numbers R01GM103500 and R01GM108888). US National Science Foundation EPS- CoR (grant number EPS-1011354), the M.J. Murdock Charitable Trust and the Montana State University Agricultural Experimental Station.

Competing interests statement
The authors declare no competing interests.

DATABASES
Electron Microscopy Data Bank (EMDB)
http://www.ebi.ac.uk/pdbe/emdb
Protein Data Bank (PDB)
http://www.rcsb.org/pdb/home/home.do

SUPPLEMENTARY INFORMATION
See online article: 1ALN (table), 4AL7 (figure), 4F3M (figure), 5860 (figure), 5740 (figure), 5314 (figure), 5760 (figure), 4921 (figure), 4923 (figure)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF