

Review Article

The biology and type I/III hybrid nature of type I-D CRISPR–Cas systems

 Tess M. McBride^{1,2},  Shaharn C. Cameron^{1,2,3},  Peter C. Fineran^{1,2,3} and  Robert D. Fagerlund^{1,2,3}

¹Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand; ²Genetics Otago, University of Otago, Dunedin, New Zealand; ³Bioprotection Aotearoa, University of Otago, PO Box 56, Dunedin 9054, New Zealand

Correspondence: Peter C. Fineran (peter.fineran@otago.ac.nz) or Robert D. Fagerlund (robert.fagerlund@otago.ac.nz)



Prokaryotes have adaptive defence mechanisms that protect them from mobile genetic elements and viral infection. One defence mechanism is called CRISPR–Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins). There are six different types of CRISPR–Cas systems and multiple subtypes that vary in composition and mode of action. Type I and III CRISPR–Cas systems utilise multi-protein complexes, which differ in structure, nucleic acid binding and cleaving preference. The type I-D system is a chimera of type I and III systems. Recently, there has been a burst of research on the type I-D CRISPR–Cas system. Here, we review the mechanism, evolution and biotechnological applications of the type I-D CRISPR–Cas system.

Introduction

Prokaryotes are under constant predation by their viruses and mobile genetic elements (MGEs), such as plasmids [1]. To counteract infection, prokaryotes have evolved an arsenal of diverse defence mechanisms [2]. CRISPR–Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) systems are heritable, adaptive immune systems that are present in ~80% and ~40% of all sequenced archaeal and bacterial genomes, respectively [2–5]. The loci of CRISPR–Cas systems consist of CRISPR arrays, which store foreign nucleic acid sequences from previous infections, and *cas* genes that encode the structural and catalytic proteins of the system. When re-exposed to an MGE, the invading nucleic acids are sequence-specifically targeted and degraded by CRISPR–Cas complexes to eliminate the threat.

There are three stages of CRISPR–Cas immunity: adaptation, expression and processing, and interference (Figure 1) [6]. The core adaptation machinery, Cas1 and Cas2, are conserved throughout almost all CRISPR–Cas systems [7]. During adaptation, Cas1 and Cas2 associate to form an adaptation complex that captures a foreign nucleic acid fragment, a prespacer, and incorporates it into the CRISPR array [8–10]. During expression and processing, the *cas* genes are transcribed and translated, while the CRISPR array is transcribed into one long RNA known as precursor-CRISPR RNA (pre-crRNA). The pre-crRNA is processed by an endoribonuclease, which generates a shorter crRNA [11–13]. During interference, the Cas protein(s) and the crRNA form an interference complex [14]. This interference complex uses the crRNA as a guide to bind the complementary invading nucleic acid target known as a ‘protospacer’. Once the complex has bound the target, Cas protein(s) degrade the invading nucleic acid(s), protecting the host cell [15].

CRISPR–Cas systems have a wide diversity of structural and mechanistic attributes and are classified into two classes, six types, and over 30 subtypes [16]. Class 1 CRISPR–Cas systems are characterised by their multi-protein interference complexes and account for ~90% of all sequenced and classified CRISPR–Cas systems, while class 2 systems are defined by having a single interference protein [4]. Class 1 systems can be further divided into types I, III, and IV based on their ‘signature’ genes.

Received: 24 September 2022
Revised: 16 January 2023
Accepted: 17 January 2023

Version of Record published:
13 April 2023

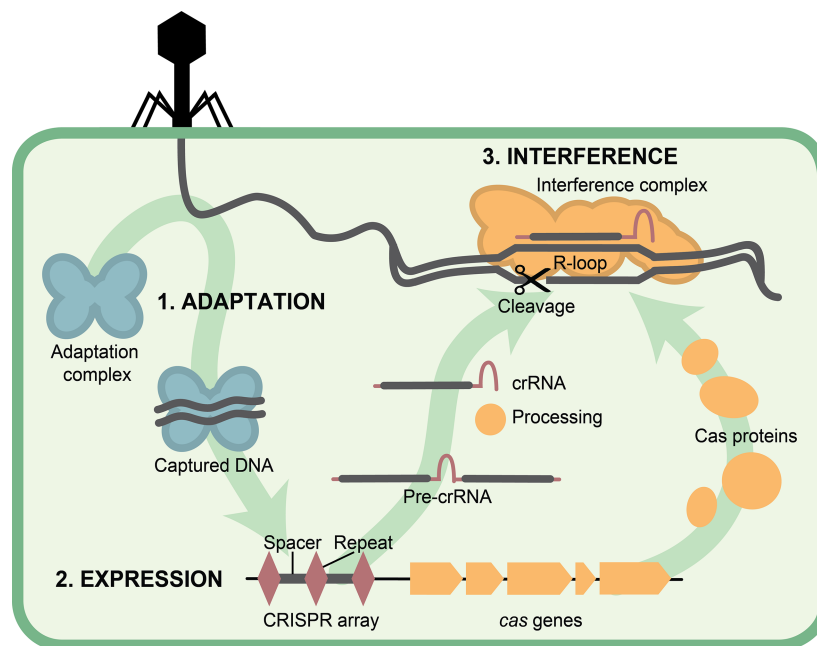


Figure 1. CRISPR–Cas stages of defence.

Adaptation occurs when the adaptation complex (blue) captures DNA, which it then incorporates into the CRISPR array as a spacer (grey line). The expression stage occurs when the cas genes (yellow) are expressed, and the CRISPR array is transcribed into pre-crRNA, which is subsequently processed by a Cas protein into crRNA. Interference occurs when the Cas proteins and crRNA assemble into the interference complex and then binds to complementary DNA and degrades it. Stages of defence based on type I CRISPR–Cas systems.

This review focuses on the type I-D CRISPR–Cas system, which is a hybrid of type I and III systems (Figure 2A) [7]. Early bioinformatic evidence indicated the type I-D system was likely an evolutionary ancestor of class 1 systems, linking a type III-like ancestor to the typical type I system [7]. Recently, the mechanism of type I-D immunity has begun to be discovered, revealing structural and mechanistic features of both type I and type III systems [7,17–19]. After the burst of research into class 2 complexes as genetic editing tools, the spotlight has started to fall on class 1 systems to expand the CRISPR toolbox [20]. Indeed, type I-D Cascade shows potential for novel biotechnological applications [21–23].

Class 1 systems

The class 1 type I and III systems have a common evolutionary lineage and share a similar interference complex architecture [4,24]. The interference complexes typically have a major filament comprised of multiple Cas7 (Csm3/Cmr4 in type III) subunits that bind along the length of the crRNA in a helical fashion (Figure 2B). The 3' end of the crRNA can be capped with Cas6 in type I systems or with a non-catalytic Cas7 (Csm5/Cmr6) in type III systems. The 5' end of the crRNA is capped with Cas5 and the large subunit, which is Cas8 or Cas10 in type I and III systems, respectively [25–29]. Typically, the C-terminus of the large subunit and multiple copies of Cas11, also known as the small subunit, runs along the belly of the complex as the minor filament [25,30].

Type I system classification and mechanism

Type I systems, subtypes I-A through I-G, are characterised by the presence of the signature gene *cas3* (Figure 2A) [7]. Type I systems use an interference complex, termed Cascade (CRISPR-associated complex for antiviral defence), that recognises a protospacer adjacent motif (PAM) and targets dsDNA [31–33]. The PAM is a two to five base pair motif present on the invading dsDNA, but not in the CRISPR array, that allows Cascade to discriminate between 'self' and 'non-self' sequences [34,35]. The PAM is identified by the PAM recognition pocket in the large subunit, Cas8, which makes a combination of sequence- and structure-based

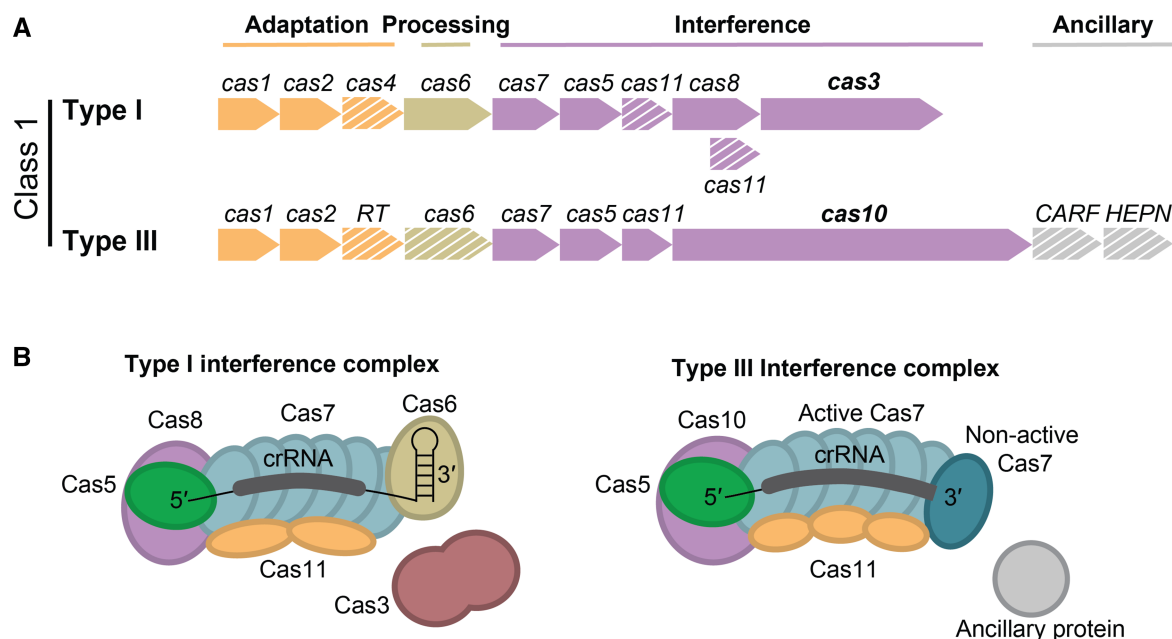


Figure 2. Classification Type I and type III CRISPR–Cas systems.

(A) Comparison of type I and III CRISPR–Cas systems, with their ‘signature’ genes in bold. Genes are organised by role; adaptation (yellow), processing (olive green), interference (purple), and ancillary (grey). Genes absent from some subtypes are indicated with white diagonal lines. Diagram adapted from Makarova et al. [16]. (B) Schematic of type I and type III interference complexes with proteins annotated.

interactions. Following PAM detection, the crRNA binds to the complementary target sequence and forms an R-loop in the dsDNA target [36,37]. For Cascade to fully bind the target, the first 8–10 nt (except position 6), which define the seed sequence of the crRNA, must have complete complementarity to the protospacer [36,38]. The R-loop displaces the non-target strand of foreign dsDNA and permits the crRNA to fully hybridise with the target strand [39,40]. The binding of the protospacer to the entire length of the crRNA induces conformational changes in the complex that lead to the recruitment of Cas3 [39,41]. Cas3 typically contains a metal-dependent histidine-aspartate (HD) nuclease domain (Cas3^{''}) and a helicase domain (Cas3[']) [42]. Cas3 nicks the non-target strand and generates a free end of ssDNA, which is then threaded through the helicase domain and is processively unwound via 3'–5' helicase activity and degraded [37,39,42,43]. Targeting and degradation of invading DNA protects the host cell from viruses that may result in cell lysis and plasmids that contain potentially detrimental genes [44].

Type III system classification and mechanism

Type III systems share a common interference complex architecture to type I systems; however, the targeted nucleic acids and mechanism of interference differ [14,45–48]. All type III CRISPR–Cas systems have multi-protein interference complexes that are evolutionarily related to type I Cascade and feature Cas protein homologues (Csm in types III-A/D/E/F and Cmr in types III-B/C), except the recently discovered type III-E system (also known as gRAMP and Cas7–11) that involves a single protein with multiple domains [16]. Type III CRISPR–Cas systems are defined by the signature gene *cas10* (Figure 2A). The Cas10 large subunit features an N-terminal HD nuclease domain and two polymerase-like cyclic (palm) domains, although only one site appears active [4]. Type III interference complexes bind to target RNAs that are complementary to their crRNA and have a unique mechanism to distinguish self from non-self targets. These type III interference complexes detect non-self-transcripts by recognising mismatches between the repeat-derived 5' handle of the crRNA and the 3' protospacer flanking sequence (PFS) of the target RNA. If complementarity between these sequences is detected — such as when antisense transcripts derived from CRISPR arrays are produced — Cas10 HD nuclease and cyclase activity is inactivated to prevent autoimmunity [47,49,50]. The detection of

non-complementarity between the crRNA 5' handle and the 3' PFS of the target RNA license a conformational change and activation of Cas10 [49]. The Cas10 HD domain non-specifically degrades ssDNA, and the active palm domain catalyses the conversion of ATP to a range of cyclic oligonucleotides (cOAs) [47,48,50,51]. Cyclic oligoadenylates act as secondary messengers that allosterically activate ancillary nucleases that non-specifically degrade foreign and host-derived DNA/RNA and can induce host cell dormancy or death [47,48,52–55]. In addition, catalytically active RNA Recognition Motif (RRM) domains within the Cas7 backbone cleave the mRNA target at six nucleotide intervals [28,56,57]. The cleavage and release of the RNA fragments return the Cas10 HD and cyclase domains to their inactive state [58]. Targeting RNA by type III systems allows specific targeting of transcriptionally active plasmids and viral genomes [46,59].

The type I-D CRISPR-Cas system

Classification of the type I-D system

The type I-D CRISPR–Cas system is abundant in archaea and cyanobacteria [4,24,60,61]. Type I-D is a hybrid system that contains signature genes of both type I and type III systems, *cas3* and *cas10*, respectively (Figure 3) [7]. However, *cas3* is split into its helicase (*cas3'*) and nuclease (*cas3''*) domains, with the nuclease domain fused to *cas10* containing an inactivated palm domain, forming the large subunit *cas10d* (also known as *csc3*) [24]. Although the domain organisation of Cas10d resembles Cas10 with an HD nuclease domain at the N-terminus, this domain in Cas10d lacks the circular permutation of the histidine residues that is the hallmark of the type III Cas10 HD domain, and instead is reminiscent of the HD domain from other type I Cas3 proteins [4,17,24]. While the type I-D locus can vary between strains, it typically also contains type I-D orthologs of *cas1*, *cas2*, *cas4*, *cas5* (*cas5d*; *csc1*), *cas6*, (*cas6d*) *cas7* (*cas7d*; *csc2*), and frequently a transcriptional regulator and CRISPR array(s) [7,19]. Additionally, within *cas10d* is an internal sequence (designated *cas11d*; Figure 3) that encodes the small subunit Cas11d [7,19]. Current studies have primarily used type I-D systems derived from several archaea and bacteria, notably the cyanobacterial strain *Synechocystis* sp. PCC6803 (Hereafter *Synechocystis*) (Figure 3). Some investigated systems appear to lack genes, such as *cas2* in *Sulfolobus islandicus*, and *cas4* and *cas6* in *Thermophilum pendens* (Figure 3). The lack of some genes may indicate these systems are inactive or may expropriate proteins or crRNA from other CRISPR–Cas systems in the host [4,60]. One example of cross-talk between CRISPR–Cas systems is the co-option of type I-F crRNAs by the type III-B system in *Marinomonas mediterranea* [62].

Type I-D naïve adaptation

There are two types of adaptation in CRISPR–Cas systems, naïve and priming [8,63–65]. Naïve adaptation occurs when the adaptation complex establishes immunity against nucleic acids which have not been previously encountered [63]. Primed adaptation can occur when the interference complex associates with escape mutants or even during the normal processes of interference [65–67]. During primed adaptation, the interference complex interacts with the adaptation complex and allows the rapid acquisition of new spacers to bolster or restore immunity. Since there is currently no data on priming by type I-D systems, we will limit our discussion to naïve adaptation.

The mechanism of type I-D adaptation has been determined predominantly through experiments using the *Synechocystis* type I-D adaptation genes in the heterologous host *Escherichia coli* [68–70]. The nucleic acid source of the type I-D adaptation complex are fragments of dsDNA produced by DNA repair enzymes, and there is no orientation bias to a particular strand [68,71]. Adaptation by type I-D involves three proteins; Cas1, Cas2, and Cas4 (Figure 4A). Cas1 has nuclease and integrase domains [72], while Cas2 generally plays a structural role [72], and Cas4 has a PD-(D/E)XK RecB endonuclease domain [73,74]. During type I-D adaptation, two asymmetrical complexes exist within the cell to process the prespacer, Cas₁₂:Cas₄₁ and Cas₁₂:Cas₂₂ [70]. These asymmetrical complexes differ from other type I systems which typically form a single heterohexameric Cas₁₄:Cas₂₂ complex and can contain one to two Cas4 subunits [9,75]. The type I-D Cas₁₂:Cas₄₁ complex identifies and binds the PAM-containing 3' overhang regions of DNA fragments, while the Cas₁₂:Cas₂₂ complex binds the opposite non-PAM containing 3' overhang end [70,76]. While Cas1 and Cas2 are sufficient to capture and integrate spacers, Cas4 allows for the enrichment of spacers containing PAMs that support interference [68]. When the complexes have captured the dsDNA, Cas4 endonuclease activity is required to remove the PAM, leaving an overhang that is typically 6 nt in length [70]. Single nucleotide slipping can occur during PAM cleavage, which may play a role in primed adaptation [70,77]. It is unknown if the PAM

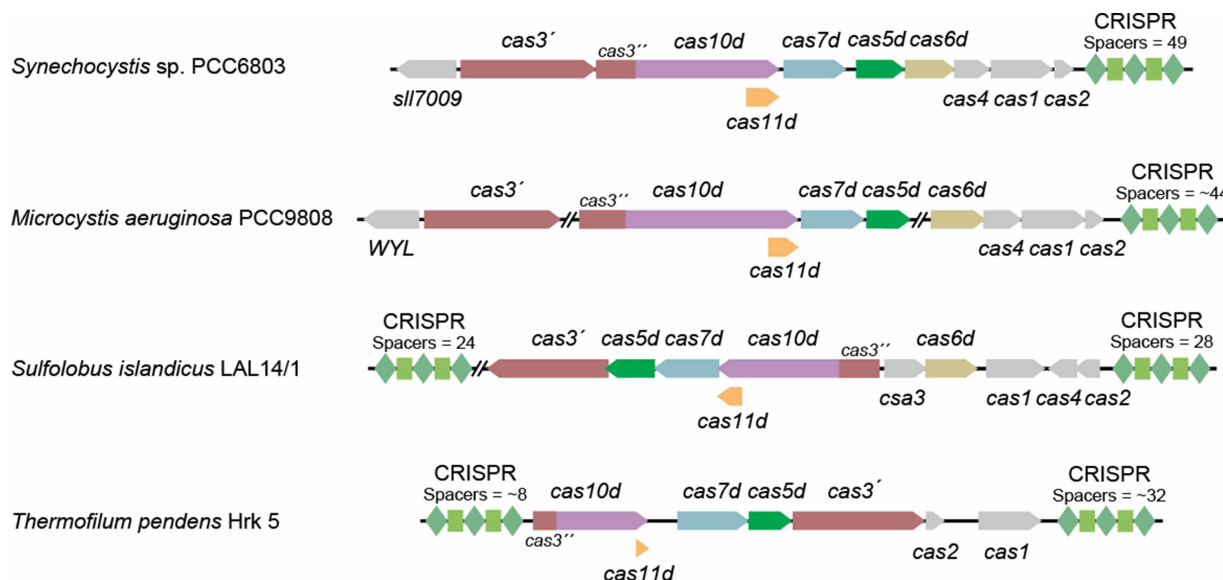


Figure 3. Schematic of representative type I-D cas genes and CRISPR arrays.

CRISPR repeats are indicated by green diamonds, and spacers are indicated by green boxes. CRISPR arrays represent the position in locus, and the length is not correlated to the actual size of the CRISPR array.

recognition domain is present in Cas1 or Cas4, but the PAM sensing C-terminal tail in type I-E Cas1 is not conserved in type I-D Cas1 [70,78]. The non-PAM containing 3' overhang is processed by an unknown endogenous nuclease [68,70].

The leader region of the CRISPR array is an A-T-rich region important for adaptation and contains the promoter for array expression [79]. The type I-D leader sequence in cyanobacteria is 202–220 bp, which is longer than other systems, such as type I-E, which is typically less than 100 bp (Figure 4B) [69]. Within the type I-D leader sequence, three conserved motifs enhance spacer uptake: motifs I, II, and III [69]. While the exact function of each motif is yet to be identified, Kieper et al. [69] noted that motif III resembled the integrase anchoring site in type I-E system, a sequence that interacts with the adaptation machinery upon DNA bending by Integration Host Factor (IHF). However, as there is no homologue of IHF in *Synechocystis*, the authors speculated that another DNA-binding host factor may bind at motif II and facilitate DNA bending to position motif III in proximity of the adaptation complex [69,80]. Kieper et al. [70] proposed a model (Figure 4A) in which the type I-D adaptation complex sequence-specifically identifies the leader region and positions the Cas₁₂:Cas₂₂ complex to catalyse half-site integration of the non-PAM end into the leader proximal end of the repeat. The Cas₁₂:Cas₄₁ complex is predicted to pause to allow the correct orientation of the spacer within the CRISPR array. The Cas₁₂:Cas₄₁ complex next catalyses the integration of the end that previously contained the PAM into the downstream repeat-spacer junction. Endogenous DNA repair systems then repair the repeat regions, resulting in the new spacer being fully integrated between two repeats into the leader-proximal region of the CRISPR array [70,81]. Spacers integrated into the type I-D system CRISPR array were most frequently 35 nt in length; however, the size varied by 5 to 6 nt, which is common in other systems containing Cas4 [68,82]. When Cas4 was removed from the system, the spacer size increased by 1 nt to 36 nt, indicating that Cas4 plays a role in correct spacer processing [68].

Type I-D expression and processing

Before type I-D Cascade can form, cas genes and crRNA are expressed. The type I-D interference genes *cas3'*, *cas5d*, *cas6d*, *cas7d*, and *cas10d* are transcribed and then translated (Figure 5) [83]. The small subunit of the type I-D system, Cas11d, is expressed from an in-frame alternative translational start site within the 3' end of *cas10d* [19]. This phenomenon is conserved across type I-D, I-B, and I-C systems [17,19,23].

The expression and processing of type I-D crRNAs have primarily been studied in *Synechocystis* [19,83]. Like other type I systems, the CRISPR array is expressed as a pre-crRNA, and the palindromic repeats present in

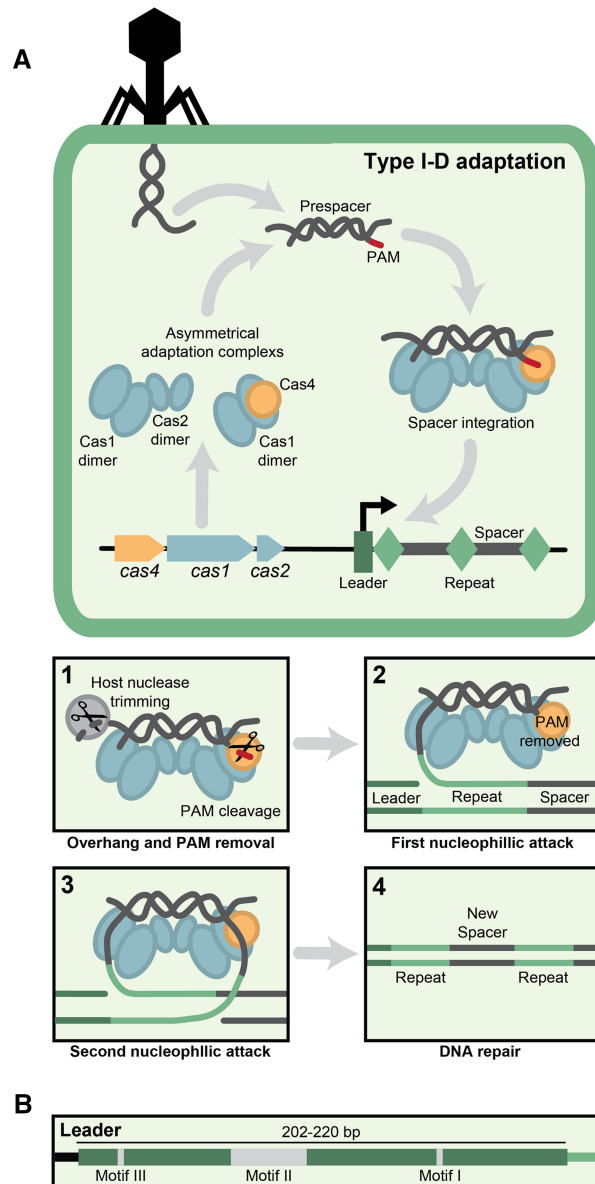


Figure 4. Predicted model of naïve type I-D adaptation.

(A) The adaptation Cas proteins form asymmetrical complexes Cas₁₂:Cas₂₂ and Cas₁₂:Cas₄₁, which capture a prespacer. The PAM end is removed by Cas4, and opposite 3' overhang is trimmed by a host nuclease (1). Half-site integration occurs via nucleophilic attack of the prespacer into the leader-repeat junction (2). Full-site integration occurs via nucleophilic attack of the second strand of the prespacer into the repeat-spacer junction (3). DNA repair completes integration to form a repeat-spacer-repeat at the leader proximal end of the array (4). Adapted from the model proposed by Kieper et al. [70]. (B) Schematic of the *Synechocystis* leader sequence and associated motifs.

each repeat form stem-loop secondary structures (Figure 5) [83–85]. Cas6d binds the pre-crRNA in a sequence- and secondary structure-dependent manner and cleaves the repeat sequence after the stem-loop to create an ~72 nt crRNA with an 8 nt 5' and a 29 nt 3' repeat-derived handles [83,86]. This 72 nt crRNA was the predominant crRNA species present within type I-D Cascade when recombinantly expressed and purified from *E. coli* [19]. However, in the natural host *Synechocystis*, there appears further maturation to 45 and 39 nt [83]. Similar maturation of the *S. islandicus* type I-D crRNA was observed when extracted from type I-D Cascade expressed in *S. islandicus*, revealing crRNAs of 49 ± 1 nt [18]. In *Synechocystis*, the 6 nt difference in

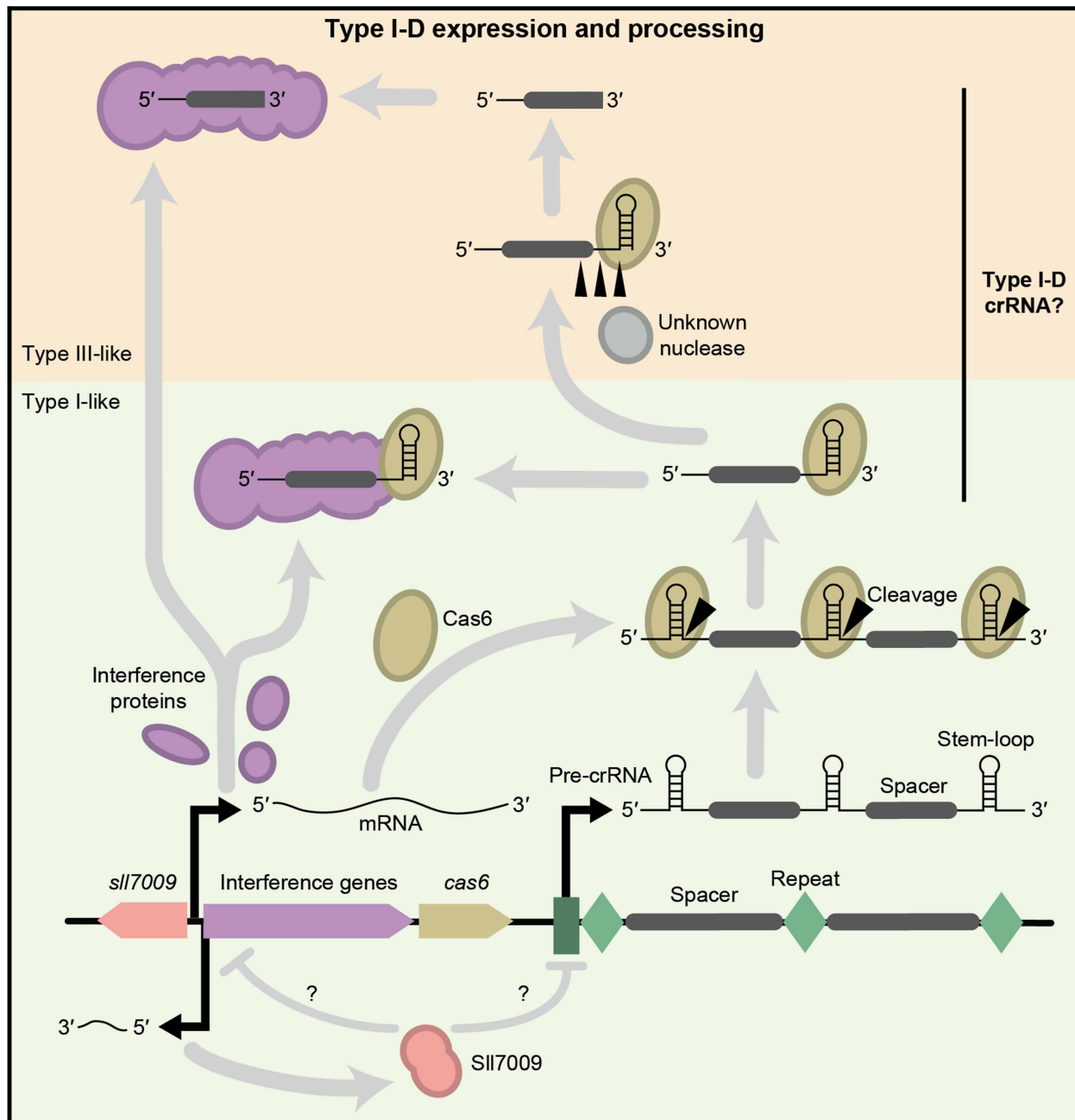


Figure 5. Predicted model of type I-D expression and processing.

The *cas* (purple and brown) and regulator (*sII7009*) genes are transcribed and translated while the CRISPR array is transcribed, forming the pre-crRNA. *SII7009* may regulate type I-D interference via any of the promoters. Cas6 associates with the stem-loops of the pre-crRNA and cleaves at the base of the stem, forming type I-like crRNA (green background), as observed by McBride et al. [19]. Shorter type I-D crRNA also occurs (Scholz et al. [83]), resembling type III-like mature crRNA (yellow background).

mature crRNA lengths indicates step-wise processing by a host nuclease(s), similar to type III systems and suggests complexes lacking Cas6 and a Cas7 subunit(s) [83,87]. The type I-D crRNA may have undergone maturation with unidentified endoribonucleases present within its native host, *Synechocystis*, but not in an *E. coli* heterologous host.

Type I-D systems frequently co-occur with a DNA binding regulator that likely responds to signal(s) to elicit changes in CRISPR–Cas activity (Figure 3) [4,60]. In *Synechocystis*, this regulator is called *SII7009* and

contains a WYL (tryptophan, tyrosine, leucine) motif that is predicted to bind small negatively charged ligands and a winged helix-turn-helix domain that binds DNA [88–90]. Regulators are associated with other type I-D systems, including the WYL-containing protein from *M. aeruginosa* and Csa3 from *S. islandicus*, which contains a CRISPR-associated Rossman fold (CARF) domain. The *Synechocystis* type I-D system includes three promoters, one that controls the *cas* genes, a second that controls the CRISPR array, and a third that controls the Sll7009 regulator (Figure 5) [91]. Hein and colleagues identified Sll7009 as a negative regulator of the type I-D system by knocking out *sll7009* and observing an increase in mature crRNA [90]. Sll7009 likely regulates the type I-D system by coupling cell signals with the repression of at least one of the type I-D promoters.

Type I-D interference

The type I-D system has been demonstrated to be an active and functional system within *Synechocystis* [68]. The *Synechocystis* type I-D system uses a 5'-GTN-3' interference PAM *in vivo* and *in vitro*, where 'N' stands for any nucleotide [18,19,68]. The *M. aeruginosa* interference PAM differs slightly, with a 5'-GTH-3' PAM, where 'H' can be A, C, or T [21]. Insight into the interference mechanism has been gained by high-resolution structures of the *Synechocystis* type I-D Cascade and the Cas10d subunit bound to an anti-CRISPR protein from *S. islandicus* [17,92]. Type I-D Cascade has a helical major filament composed of multiple Cas7d subunits (Figure 6A). Cas6d likely binds the stem region at the 3' end of the crRNA; however, despite Cas6d being co-purified with the complex, it was not resolved in the structure [17]. The 5' end of the crRNA is capped with Cas5d and Cas10d. The C-terminus of Cas10d and the alternatively translated Cas11d subunits run along the belly of the complex, forming the minor filament [17,19]. The overall architecture of type I-D Cascade better resembles type III complexes rather than type I Cascades, particularly as the angular trajectory of the crRNA is consistent with type III crRNA [17,19].

Type I-D Cascade is capable of binding dsDNA [19], ssDNA [17,18], and ssRNA [17] with high affinity. During dsDNA interference, Cascade randomly samples dsDNA to identify a PAM (Figure 6B–D) [18,32,68,93]. The most well-studied type I system is type I-E, which uses three elements within the PAM binding pocket in the large subunit to detect the PAM: a glycine-rich loop, a lysine finger, and a glutamine wedge [94]. The glycine-rich loop is inserted into the minor groove of the DNA and results in DNA bending; both the glycine-rich loop and the lysine finger directly interact with the nucleotides to determine their specificity. The glutamine wedge is inserted after the PAM and causes bifurcation of the dsDNA [94,95]. PAM detection by type I-D Cascade was elucidated with the structure of the *Synechocystis* type I-D Cascade bound to a 5'-GTT-3' PAM-protospacer [17]. The PAM is identified by the PAM recognition domain within Cas10d. A glycine-rich loop from the PAM recognition domain is inserted into the minor groove of the dsDNA, where a glycine (G433) interacts with the guanine nucleotide on the non-target strand in the –3 PAM position (G_{NTS-3}) (Figure 6B,C). The lysine finger (K326) in the PAM recognition domain interacts with nucleotides C_{TS-3}, T_{NTS-2}, and G_{NTS-3}. These specific interactions of Cas10d with the –2 and –3 positions of the PAM are consistent with the third nucleotide being flexible as 'N' in GTN. The dsDNA is split directly after the PAM by a cleft composed of the bottom of Cas7d and Cas10d. Type I-D Cascade does not possess a conventional glutamine wedge from the large subunit; however, a Cas5 glutamine residue (Q110) inserts between the dsDNA where the strands separate, which may constitute the type I-D glutamine wedge (Figure 6B).

The bifurcation of the dsDNA allows the target strand to hybridise with the complementary crRNA in Cascade and displaces the non-target strand to form an asymmetrical R-loop [17]. When the crRNA binds the protospacer, every sixth base, starting at position one, is flipped out by the thumb domain of the Cas7 subunits, allowing for mismatches at these sites [17,22,96]. The recognition of the PAM and the protospacer causes a conformational shift of the Cas3' and Cas11d domains of Cas10d and the Cas11d subunits [17]. This conformational change may support the stabilisation of the R-loop through interactions of the non-target strand with positively charged residues in the C-terminus of Cas10d and Cas11d subunits [17,19] and direct the non-target strand towards the Cas10d active site. The Cas3' domain of Cas10d has a type I-like HD active site responsible for target cleavage [17,18,22,97]. Cas3' has been shown to interact with Cas10d [22], and this HD domain is likely where the Cas3' helicase docks [17]. Upon the recruitment of Cas3', Cas3' nicks the non-target strand and, in an ATP-dependent manner, nicks further sequences upstream of the protospacer on the non-target strand and both sides of the protospacer on the target strand [18]. The observed bidirectional cleavage upstream of the protospacer is consistent with the helicase-nuclease activity of Cas3'–Cas10d on each strand after the initial cleavage at the protospacer region. While most type I systems display unidirectional cleavage

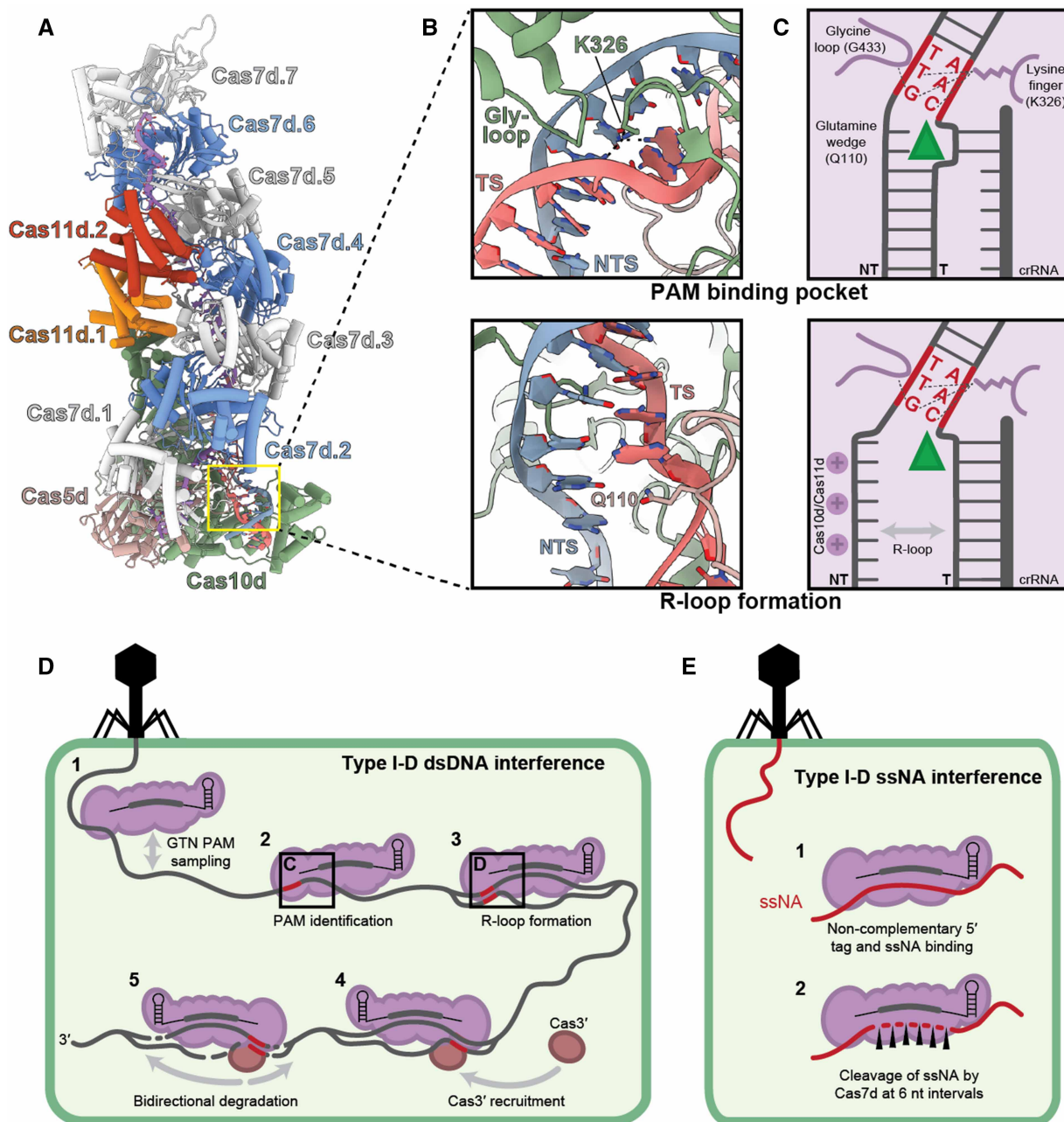


Figure 6. Model of type I-D interference for dsDNA and single-stranded nucleic acids (ssNA).

(A) Structure of *Synechocystis* type I-D Cascade (Protein Data Bank ID, 7SBA). Subunits displayed: Cas10d (green); Cas5d (pink); Cas11d (orange); Cas7d (grey and blue); crRNA (purple); target strand (TS; red); non-target strand (NTS; cyan). Structure (B) and schematic (C) of the PAM binding pocket, including the glycine loop and lysine finger from Cas10d (top) and the glutamine wedge from Cas5d positioned where dsDNA bifurcation occurs (bottom). R-loop formation is likely stabilised by a positive patch within Cas10d and Cas11d. (C) adapted from Hayes et al. [94]. (D) The predicted mechanism of type I-D interference of dsDNA. Type I-D Cascade samples dsDNA looking for a GTN PAM (1), the PAM is identified with the PAM binding pocket in Cas10d (2), and an R-loop is formed (3), Cas3' is recruited to Cascade (4), bidirectional degradation of both strands occurs (5). (E) The predicted mechanism of type I-D interference of single-stranded nucleic acids (ssNA). Type I-D Cascade binds the ssNA (1), followed by the cleavage of the ssNA by Cas7d (2).

[23,98,99], the type I-A Cascade from *Pyrococcus furiosus* was also recently shown to cleave in a bidirectional manner [100]. Interestingly, types I-A and I-D both have split *cas3* helicase and nuclease genes [24], suggesting a separately encoded helicase might be important for bidirectional cleavage. A similar model of two helicases

operating on different DNA strands was recently proposed for type IV-A interference, where helicase CasDinG facilitated bidirectional DNA depletion [101]. The precise molecular details of how bidirectional cleavage occurs are still emerging and further characterisation is required.

Type I-D Cascade can also bind single-stranded DNA and RNA *in vitro* [17,18]. Type I-D Cascade bound ssDNA and RNA with a 5'-AAC-3' PFS, complementary to a 5'-GTT-3' PAM, and a scrambled PFS (5'-ACG-3') [17], indicating single-stranded binding by type I-D Cascade appears to have a broader range of PFS requirements and different from those required for dsDNA binding [17,19,68]. Furthermore, when type I-D Cascade is bound to ssRNA, the PAM recognition domain becomes flexible, in contrast with the rigid form bound to dsDNA [17]. When type III systems bind target RNA, the palm domain of Cas10 converts ATP to cOA signalling molecules, and the Cas7 subunits cleave the target [28,47,48]. Type I-D is not expected to produce cOA molecules as the palm domain is predicted to be inactive in Cas10d [7]. Unexpectedly, the Cas10d subunit from *Microcystis aeruginosa* PCC9808 was recently shown to have ATPase activity; however, the role of this function is unknown [22]. The Cas7 subunits in Cascade have both sequence and structural homology to their active type III counterpart [24,96]. When type I-D Cascade is bound to ssDNA, the active Cas7 subunit cleaves the ssDNA at 6 nt intervals (Figure 6E) [18], similar to the cleavage of RNA in type III systems [28,56]. However, the exact mechanism of ssDNA cleavage and which residues are involved is unknown [18].

Type I-D Cascade can bind dsDNA [18,19], ssDNA [17,18], and ssRNA [17] in a sequence-specific manner, which may indicate that the type I-D system is a broad defence mechanism. However, *in vivo* interference has only been tested and demonstrated against dsDNA targets [21–23,68]. Interestingly, type I-D Cascade appears to have a binding preference for single-stranded nucleic acids over double-stranded ones, based on the binding affinity *in vitro* [17,19]. This preference may reflect the type of bacteriophage (phage) or stage of phage replication that type I-D Cascade targets. For example, if type I-D Cascade does target ssRNA *in vivo*, it may target single-stranded RNA phages or phage mRNAs produced during transcription. Alternatively, if type I-D does target ssDNA *in vivo*, it may be an ssDNA phage or ssDNA exposed during replication or transcription of the phage genome.

Viral evasion mechanisms of the type I-D system

Viruses have evolved mechanisms to counteract prokaryotic immune mechanisms, allowing for phage propagation [2,5]. To evade CRISPR–Cas immunity, phages can acquire mutations that escape detection, use anti-CRISPR proteins or protect their DNA within a nucleus-like structure [102–104]. Anti-CRISPR proteins inactivate CRISPR–Cas immunity via diverse mechanisms [44]. The only known anti-CRISPR for the type I-D system is AcrID1 (Anti-CRISPR against type I-D) from the virus SIRV3 that infects *S. islandicus* [92,97]. AcrID1 is a 13 kDa protein that forms a homodimer with a net negative charge [97]. This dimer interacts directly with Cas10d, contacting all domains within Cas10d [92,97]. Addition of AcrID1 inhibited type I-D Cascade from degrading dsDNA [97]. Manav and colleagues suggested that two AcrID1 dimers sequester two Cas10d subunits forming an AcrID1₄:Cas10d₂ complex, which would prevent Cas10d from complexing with the other Cas proteins to form a functional Cascade. Another potential inhibition method by AcrID1 was proposed by Schwartz and colleagues, who modelled AcrID1 bound to the positively charged belly of Cas10d in Cascade [17]. In this model, AcrID1 was hypothesised to obstruct the path of the non-target strand toward the Cas10d HD domain; therefore, preventing R-loop formation, subsequent Cas3' recruitment and target degradation.

Type I-D system as an evolutionary intermediate

Type I-D CRISPR–Cas systems are frequently found in archaea and cyanobacteria [24,60,105]. Phylogenetic analysis of the Cas10d subunit indicated four lineages represented by two cyanobacterial groups, one with predominantly other bacteria, and one with mostly archaea [7,19]. Across all groups, the internal expression of Cas11d appeared widely conserved. The features of type I-D systems that are reminiscent of type I and III systems have led to it being a predicted evolutionary intermediate [7,17–19]. Combined with a CRISPR–Cas origin model proposed by Koonin and Makarova [106], these observations allow for the proposal of a model for the evolution of type I-D systems (Figure 7).

The type I-D system was likely formed from a type III-like ancestor through the inactivation of the cyclase palm domain within *cas10*, the loss of associated ancillary nucleases and a *cas7* gene (Figure 7) [7,24]. The ancestral system then acquired the helicase, *cas3'*. The circular permutation of histidines within the *cas10* HD domain was lost, forming a type I-like HD domain, *cas3''* [106]. The type I-D Cas7d subunit has nuclease activity, comparable to the type III Cas7 but in contrast with other type I Cas7 subunits [7,16,18,96]. From the

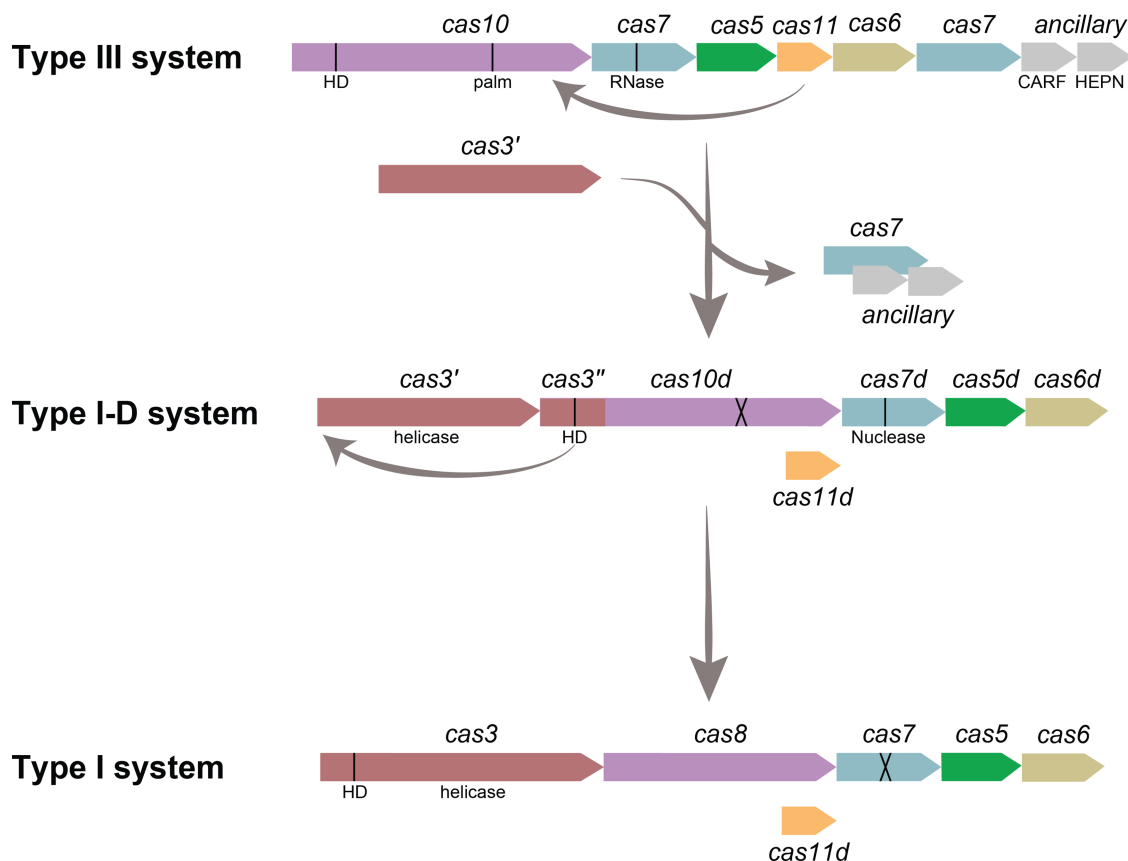


Figure 7. Predicted model of the type I-D system as an evolutionary intermediate between type III and type I systems.

Type I-D likely evolved from a type III-like ancestor system through the acquisition of a helicase domain, the loss of the circular permutation in the *cas10d* HD, *cas11* becoming alternatively expressed from the large subunit, and the loss of one *cas7* and the auxiliary genes. Type I systems may have evolved from type I-D via the transfer of the HD domain, *cas3''*, onto the helicase domain, *cas3'*, to form the typical *cas3* nuclease-helicase. In some type I systems, duplication of *cas11* occurs, forming a separate gene. A black line indicates key domains, and inactive domains are indicated by a cross.

type III-like ancestor to type I-D, the *cas11d* gene appears to fuse onto the end of *cas10*, retaining its translational start site [19]. Comparison of type I-D systems from different genomes revealed that the genes from the effector module were shuffled and the adaptation module genes, and/or *cas6*, remained conserved [106,107]. We compared diverse type I-D systems in genomes from cyanobacteria, other bacteria and archaea phyla groups [19] and found the gene order of the effector module differed between these groups, with notable differences with the associated regulator and the positions of *cas3'* and *cas6* (Supplementary Figure S1).

A model for the evolution of type I systems from type I-D begins with the HD domain, *cas3''*, being transferred onto the helicase *cas3'*, resulting in the evolution of a smaller, enzymatically inactive large subunit (*cas8*) [106]. The alternative translational start site for *cas11* was retained in some systems and may have been duplicated to become a separate gene in others. However, the evolutionary lineage of Cas11 in type I systems is unknown, and further analysis is required [19]. The Cas7d nuclease activity was lost, which resulted in inactive Cas7 subunits within other type I systems [17,106]. Further sequence and structural analysis of type I and III systems may provide further insight into the evolutionary steps that formed class 1 systems.

Biotechnological applications of type I-D systems

Over the past decade, many CRISPR–Cas systems have been co-opted as sequence-specific genome engineering technologies, which have sparked significant advances in fundamental science, precision medicine, and biotechnology [20]. Initial CRISPR–Cas technologies were developed using class 2 CRISPR–Cas systems, such as Cas9 and Cas12, which have single multi-domain effector proteins that can be programmed to generate double-

stranded breaks within/adjacent to the target site [108,109]. The class 2 systems are easily engineered, smaller than class 1 systems, and amenable for use across heterologous host species [109–112]. In contrast, the more abundant type I CRISPR–Cas systems are underutilised as biotechnological tools yet possess unique attributes that are attractive for exploitation [113]. Cas3-mediated processive degradation introduces long-range deletions into the target genome – a distinct function that has been harnessed to develop a new suite of CRISPR–Cas tools [113]. For example, the type I-E system was used to make large deletions at endogenous targeted loci in human cells [98]. Furthermore, given the natural abundance of type I systems in clinically and industrially relevant bacteria, these endogenous systems can be harnessed for applications such as precision antimicrobials and strain engineering [99,114–117]. Another advantage of harnessing type I systems is their diverse PAM requirements that differ from those required by Cas9 and Cas12-based tools, which expands the range of available target sites [32]. The newly characterised type I-D system is among the most recent additions to the type I CRISPR–Cas toolbox and, thus far, has been developed as a gene editing platform [21]. As a hybrid between a type I and III system, the type I-D system has distinct structural and functional attributes that display potential for further exploitation in developing programmable molecular tools [17–19].

The type I-D system is among several type I CRISPR–Cas effectors that have been converted into sequence-specific genome editing technologies [21–23,98,100,118–120]. The exploitation of type I CRISPR–Cas systems to generate deletions was originally performed by Vercoe *et al.*, who used an endogenous type I-F system to isolate mutants with 30–100 kb deletions in the core and accessory genome of *Pectobacterium atrosepticum* [99]. In addition, several type I-E systems have been shown to induce similar long-range chromosomal (30–100 kb) deletions in human cells [98,99,118,119]. In all type I-E editors studied, the Cas3-generated deletions were limited to regions upstream of the PAM and target site. Subsequent studies of other type I editors showed the type I-B and I-C systems also generated long-range unidirectional deletions upstream of the PAM when applied in human cells [23].

Recently, the type I-D system from *M. aeruginosa* was adapted for gene editing of human and plant genomes [21,22]. The type I-D system generated a spectrum of bidirectional long-range chromosomal deletions from 2.5 to 19 kb with $\leq 57\%$ editing efficiency in both eukaryotic cell types – a pattern consistent with type I-D *in vitro* cleavage pattern [18]. The type I-D complex also introduced double-stranded breaks and small deletions (1–12 bp) within the target site at up to 19% editing efficiency [22]. Interestingly, the type I-A system also produced bidirectional deletions in human cells that spanned up to 2.2 kb [100]. Notably, unidirectional Cas3-induced deletions in other systems (types I-B, -C and -E) appear longer (≤ 100 kb) than their bidirectional counterparts (≤ 19 kb) [21–23,98,100,118,119]. The disparity in length between the two types of deletion may be due to the unidirectional deletions preserving the PAM and protospacer and allowing multiple rounds of editing [118]. In contrast, the bidirectional deletions induced by types I-A and I-D destroyed the PAM and target site, preventing multiple rounds of editing [21,22,100]. Further characterisation of Cas3-driven interference mechanisms will provide insight into how CRISPR–Cas type I systems can be further exploited for gene editing technologies.

As previously discussed, diverse type I CRISPR–Cas complexes require the independent translation of the small subunit Cas11 from an alternative start site located within the large subunit, including Cas11d from within *cas10d* of the type I-D system [19]. The internal translation of Cas11 is a critical consideration when implementing these specific type I systems for biotechnological applications in eukaryotes due to translational machinery differing from their natural prokaryotic hosts. As predicted by McBride *et al.* [19] the reconstitution of specific type I-B, I-C, and I-D systems in eukaryotic cells may require *cas11* encoded from a separate gene. Tan and colleagues recently confirmed these predictions with gene editing studies using examples of type I-B, I-C, and I-D systems, including the *Synechocystis* type I-D system. In all cases, gene editing improved when Cas11 was separately expressed, including in *Synechocystis* type I-D, which achieved 5% editing efficiency with Cas11 but had 0% efficiency without [23]. Combined with the role of Cas11 in dsDNA binding, the decreased editing efficiency is predicted to be caused by a dramatic decrease in dsDNA binding capability [19,23]. Interestingly, Osakabe and colleagues reported that the type I-D system from *M. aeruginosa* could introduce gene edits in human cells without the need for separately encoded Cas11d. However, it was unclear from these studies whether Cas11 was expressed at sufficient levels to maximise editing efficiency [21].

Until recently, type I-D systems were largely uncharacterised; therefore, the full spectrum of biotechnological applications for these systems is still emerging. Like other type I systems, the type I-D system could be utilised for endogenous chromosomal targeting applications such as selective killing and strain engineering [99,120–122]. The type I-D complex could also be engineered for programmable gene regulation, as has previously been

achieved with catalytically inactivated type I-E and I-F systems [20,123–126]. Recent studies of type I-D systems from *S. islandicus* and *Synechocystis* have shown that the hybrid features of type I-D complexes allow the binding of dsDNA, ssDNA and ssRNA [17–19]. Further studies into how these nucleic-acid targeting activities are co-ordinated during type I-D interference may reveal a suite of complementary biotechnological applications that will further expand the CRISPR–Cas toolkit.

Concluding remarks

The type I-D CRISPR–Cas system displays genetic, structural, and mechanistic aspects of both type I and type III CRISPR–Cas systems and are often referred to as chimeric [7,17–19]. The type I-D system is likely an evolutionary intermediate of class 1 CRISPR–Cas systems, where it evolved from a type III-like ancestor and later evolved into the ‘typical’ type I system [7,106]. The chimeric nature of the type I-D system led to research into its mechanisms of adaptation, expression and processing, and interference, and has further confirmed that the type I-D system displays features of both type I and III systems [7,17–19,68–70,83]. For example, the structure of the type I-D Cascade interference complex better resembles a type III complex, and it binds and cleaves substrates in a type I- and type III-like manner [7,17–19]. The unusual features of the type I-D CRISPR–Cas system also lends itself to exploitation as novel biotechnological tools, such as for gene editing [21–23,98,100,118–120]. Despite these recent advances, there are still many key questions (see Key considerations box) into the mechanism of type I-D systems. As new details emerge into how type I-D functions, new opportunities for biotechnological exploitation may arise.

KEY CONSIDERATIONS

- Does the type I-D system elicit primed adaptation, and if so, does it occur in a bidirectional manner?
- Is Cas6d part of the final Cascade complex *in vivo*, or is it removed during crRNA processing?
- How does the regulator associated with the type I-D system control expression and what signals does it respond to?
- What is the mechanism for bidirectional deletions by the type I-D system?
- Does the single-stranded nucleic acid targeting of the type I-D system have a physiological role?
- Does Cas11d provide new insight into the evolutionary lineage of class 1 CRISPR–Cas systems?

Competing Interests

All authors are inventors on a patent application based on this research titled ‘Type I-D CRISPR–Cas Systems and Uses Thereof,’ which has been assigned Australian Patent Application No. PCT/NZ2023/050034, and the filing date of 17 March 2023.

CRedit Author Contribution

Peter C. Fineran: Conceptualization, Supervision, Funding acquisition, Project administration, Writing – review and editing. **Tess M. McBride:** Conceptualization, Visualization, Writing – original draft, Writing – review and editing. **Shaharn C. Cameron:** Conceptualization, Writing – original draft, Writing – review and editing. **Rob Fagerlund:** Conceptualization, Supervision, Visualization, Writing – original draft, Project administration, Writing – review and editing.

Acknowledgements

We thank David Taylor, Jack Bravo and Evan Schwartz for helpful discussions on the *Synechocystis* type I-D system. This work was supported by a Marsden Fund Fast-Start Grant (to R.D.F.) from the Royal Society Te Apārangi, New Zealand (RSNZ), the Marsden Fund (to P.C.F.), and Bioprotection Aotearoa (Tertiary Education Commission, NZ) (to R.D.F., S.C.C. and P.C.F.). P.C.F. was supported by an Experienced Researcher Fellowship

from the Alexander von Humboldt Foundation. T.M.M. and S.C.C were supported by the University of Otago Doctoral Scholarship, and T.M.M by a University of Otago Doctoral Publishing Bursary.

Abbreviations

HD, histidine-aspartate; IHF, Integration Host Factor; PAM, protospacer adjacent motif; PFS, protospacer flanking sequence; CRISPR-Cas, clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins; Cascade, CRISPR-associated complex for antiviral defence; MGE, mobile genetic element.

References

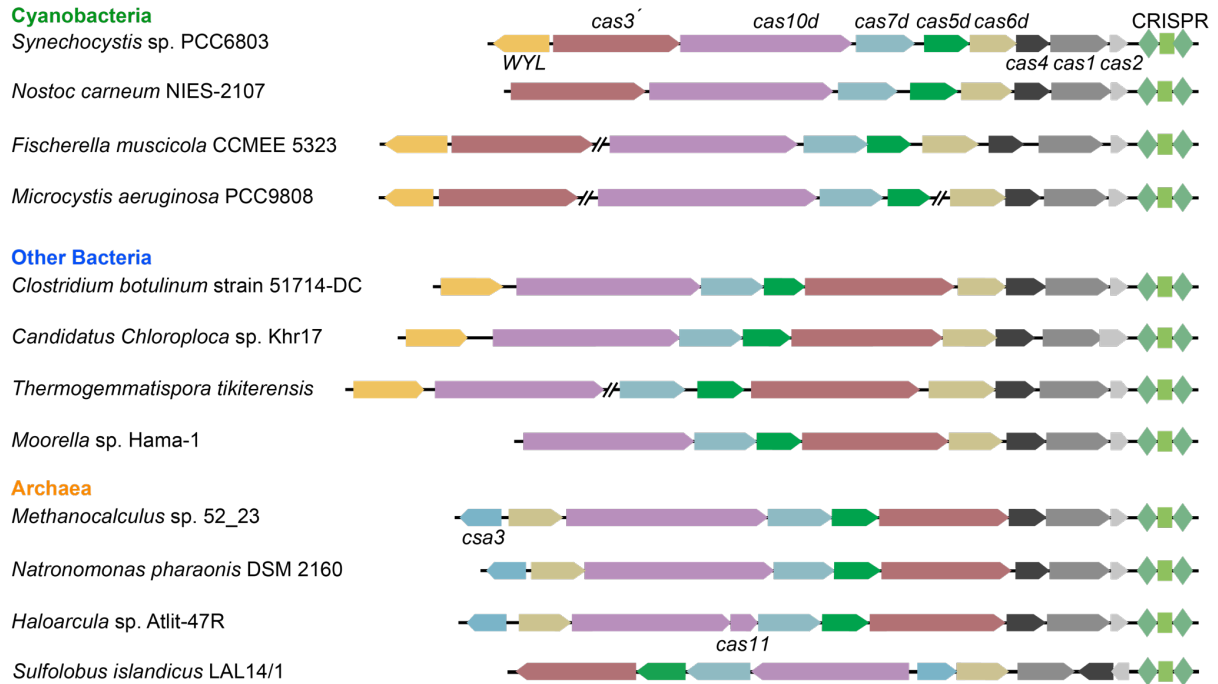
- Price, J.E. and Chapman, M.R. (2018) Phaged and confused by biofilm matrix. *Nat. Microbiol.* **3**, 2–3 <https://doi.org/10.1038/s41564-017-0078-2>
- Hampton, H.G., Watson, B.N. and Fineran, P.C. (2020) The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 <https://doi.org/10.1038/s41586-019-1894-8>
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S. et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 <https://doi.org/10.1126/science.1138140>
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J. et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 <https://doi.org/10.1038/nrmicro3569>
- Safari, F., Sharifi, M., Farajnia, S., Akbari, B., Karimi Baba Ahmadi, M., Negahdaripour, M. et al. (2020) The interaction of phages and bacteria: the co-evolutionary arms race. *Crit. Rev. Biotechnol.* **40**, 119–137 <https://doi.org/10.1080/07388551.2019.1674774>
- Barrangou, R. (2013) CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip. Rev. RNA* **4**, 267–278 <https://doi.org/10.1002/wrna.1159>
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P. et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 <https://doi.org/10.1038/nrmicro2577>
- Yosef, I., Goren, M.G. and Qimron, U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 <https://doi.org/10.1093/nar/gks216>
- Núñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 <https://doi.org/10.1038/nsmb.2820>
- 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 <https://doi.org/10.1371/journal.pone.0049549>
- Carte, J., Wang, R., Li, H., Terns, R.M. and Terns, M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 <https://doi.org/10.1101/gad.1742908>
- Hochstrasser, M.L. and Doudna, J.A. (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* **40**, 58–66 <https://doi.org/10.1016/j.tibs.2014.10.007>
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 <https://doi.org/10.1126/science.1192272>
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuys, R.J., Snijders, A.P. et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 <https://doi.org/10.1126/science.1159689>
- Garneau, J.E., Dupuis, M.É., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P. et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 <https://doi.org/10.1038/nature09523>
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J. et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 <https://doi.org/10.1038/s41579-019-0299-x>
- Schwartz, E.A., McBride, T.M., Bravo, J.P.K., Wrapp, D., Fineran, P.C., Fagerlund, R.D. et al. (2022) Structural rearrangements allow nucleic acid discrimination by type I-D cascade. *Nat. Commun.* **13**, 2829 <https://doi.org/10.1038/s41467-022-30402-8>
- Lin, J., Fuglsang, A., Kjeldsen, A.L., Sun, K., Bhoobalan-Chitty, Y. and Peng, X. (2020) DNA targeting by subtype ID CRISPR–Cas shows type I and type III features. *Nucleic Acids Res.* **48**, 10470–10478 <https://doi.org/10.1093/nar/gkaa749>
- McBride, T.M., Schwartz, E.A., Kumar, A., Taylor, D.W., Fineran, P.C. and Fagerlund, R.D. (2020) Diverse CRISPR-Cas complexes require independent translation of small and large subunits from a single gene. *Mol. Cell* **80**, 971–979.e977 <https://doi.org/10.1016/j.molcel.2020.11.003>
- Pickar-Oliver, A. and Gersbach, C.A. (2019) The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 <https://doi.org/10.1038/s41580-019-0131-5>
- Osakabe, K., Wada, N., Miyaji, T., Murakami, E., Marui, K., Ueta, R. et al. (2020) Genome editing in plants using CRISPR type I-D nuclease. *Commun. Biol.* **3**, 648 <https://doi.org/10.1038/s42003-020-01366-6>
- Osakabe, K., Wada, N., Murakami, E., Miyashita, N. and Osakabe, Y. (2021) Genome editing in mammalian cells using the CRISPR type I-D nuclease. *Nucleic Acids Res.* **49**, 6347–6363 <https://doi.org/10.1093/nar/gkab348>
- Tan, R., Krueger, R.K., Gramelspacher, M.J., Zhou, X., Xiao, Y., Ke, A. et al. (2022) Cas11 enables genome engineering in human cells with compact CRISPR-Cas3 systems. *Mol. Cell* **82**, 852–867.e855 <https://doi.org/10.1016/j.molcel.2021.12.032>
- Makarova, K.S., Aravind, L., Wolf, Y.I. and Koonin, E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **6**, 38 <https://doi.org/10.1186/1745-6150-6-38>
- Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P. et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 <https://doi.org/10.1038/nsmb.2019>
- van Duijn, E., Barbu, I.M., Barendregt, A., Jore, M.M., Wiedenheft, B., Lundgren, M. et al. (2012) Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced short-palindromic-repeats (CRISPR)-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Mol. Cell. Proteomics* **11**, 1430–1441 <https://doi.org/10.1074/mcp.M112.020263>
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G. et al. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell* **52**, 124–134 <https://doi.org/10.1016/j.molcel.2013.08.020>

- 28 Staals, R.H.J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E. et al. (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol. Cell* **52**, 135–145 <https://doi.org/10.1016/j.molcel.2013.09.013>
- 29 Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R. et al. (2013) Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol. Cell* **52**, 146–152 <https://doi.org/10.1016/j.molcel.2013.09.008>
- 30 Venclovas, Č. (2016) Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes. *FEBS Lett.* **590**, 1521–1529 <https://doi.org/10.1002/1873-3468.12179>
- 31 Mojica, F.J.M., Diez-Villaseñor, C., Garcia-Martínez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)* **155**, 733–740 <https://doi.org/10.1099/mic.0.023960-0>
- 32 Shah, S.A., Erdmann, S., Mojica, F.J. and Garrett, R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.* **10**, 891–899 <https://doi.org/10.4161/ma.23764>
- 33 Nuñez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N. and Doudna, J.A. (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535–538 <https://doi.org/10.1038/nature15760>
- 34 Sashital, D.G., Wiedenheft, B. and Doudna, J.A. (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* **46**, 606–615 <https://doi.org/10.1016/j.molcel.2012.03.020>
- 35 Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. et al. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.* **9**, e1003742 <https://doi.org/10.1371/journal.pgen.1003742>
- 36 Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B. et al. (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 <https://doi.org/10.1073/pnas.1104144108>
- 37 Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L. et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by cascade and Cas3. *Mol. Cell* **46**, 595–605 <https://doi.org/10.1016/j.molcel.2012.03.018>
- 38 Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J. et al. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 <https://doi.org/10.1038/nature10402>
- 39 Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E. et al. (2014) Cas3 mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl Acad. Sci. U.S.A.* **111**, 6618–6623 <https://doi.org/10.1073/pnas.1405079111>
- 40 Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P. et al. (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 <https://doi.org/10.1073/pnas.1402597111>
- 41 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 Rep.* **10**, 1534–1543 <https://doi.org/10.1016/j.celrep.2015.01.067>
- 42 Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342 <https://doi.org/10.1038/emboj.2011.41>
- 43 Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y. et al. (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* **21**, 771–777 <https://doi.org/10.1038/nsmb.2875>
- 44 Bondy-Denomy, J. and Davidson, A.R. (2014) To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* **22**, 218–225 <https://doi.org/10.1016/j.tim.2014.01.007>
- 45 Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L. et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 <https://doi.org/10.1016/j.cell.2009.07.040>
- 46 Goldberg, G.W., Jiang, W., Bikard, D. and Marraffini, L.A. (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633–637 <https://doi.org/10.1038/nature13637>
- 47 Kazlauskienė, M., Kostiuik, G., Venclovas, Č., Tamulaitis, G. and Siksnys, V. (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 <https://doi.org/10.1126/science.aao0100>
- 48 Niewoehner, O., Garcia-Doval, C., Rostøl, J.T., Berk, C., Schwede, F., Bigler, L. et al. (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 <https://doi.org/10.1038/nature23467>
- 49 Sofos, N., Feng, M., Stella, S., Pape, T., Fuglsang, A., Lin, J. et al. (2020) Structures of the Cmr-beta complex reveal the regulation of the immunity mechanism of Type III-B CRISPR-Cas. *Mol. Cell* **79**, 741–757.e747 <https://doi.org/10.1016/j.molcel.2020.07.008>
- 50 Estrella, M.A., Kuo, F.T. and Bailey, S. (2016) RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes Dev.* **30**, 460–470 <https://doi.org/10.1101/gad.273722.115>
- 51 Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A. and Marraffini, L.A. (2015) Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas immunity. *Cell* **161**, 1164–1174 <https://doi.org/10.1016/j.cell.2015.04.027>
- 52 Smalakyte, D., Kazlauskienė, M., Havelund, J.F., Rukšėnaitė, A., Rimaite, A., Tamulaitiene, G. et al. (2020) Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains. *Nucleic Acids Res.* **48**, 9204–9217 <https://doi.org/10.1093/nar/gkaa634>
- 53 Rostøl, J.T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Fromm, R. et al. (2021) The Card1 nuclease provides defence during type III CRISPR immunity. *Nature* **590**, 624–629 <https://doi.org/10.1038/s41586-021-03206-x>
- 54 Zhu, W., McQuarrie, S., Gruschow, S., McMahon, S.A., Graham, S., Gloster, T.M. et al. (2021) The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence. *Nucleic Acids Res.* **49**, 2777–2789 <https://doi.org/10.1093/nar/gkab073>
- 55 Mayo-Muñoz, D., Smith, L.M., Garcia-Doval, C., Malone, L.M., Harding, K.R., Jackson, S.A. et al. (2022) Type III CRISPR-Cas provides resistance against nucleus-forming jumbo phages via abortive infection. *Mol. Cell* **82**, 4471–4486.e4479 <https://doi.org/10.1016/j.molcel.2022.10.028>
- 56 Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A. et al. (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol. Cell* **56**, 518–530 <https://doi.org/10.1016/j.molcel.2014.10.005>
- 57 Tamulaitis, G., Kazlauskienė, M., Manakova, E., Venclovas, Č., Nwokeoji, A.O., Dickman, M.J. et al. (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol. Cell* **56**, 506–517 <https://doi.org/10.1016/j.molcel.2014.09.027>
- 58 Kazlauskienė, M., Tamulaitis, G., Kostiuik, G., Venclovas, Č. and Siksnys, V. (2016) Spatiotemporal control of Type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol. Cell* **62**, 295–306 <https://doi.org/10.1016/j.molcel.2016.03.024>

- 59 Gudbergstottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q. et al. (2011) Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49 <https://doi.org/10.1111/j.1365-2958.2010.07452.x>
- 60 Bernheim, A., Bikard, D., Touchon, M. and Rocha, E.P. (2020) Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic Acids Res.* **48**, 748–760 <https://doi.org/10.1093/nar/gkz1091>
- 61 Cai, F., Axen, S.D. and Kerfeld, C.A. (2013) Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biol.* **10**, 687–693 <https://doi.org/10.4161/ma.24571>
- 62 Silas, S., Lucas-Elio, P., Jackson, S.A., Aroca-Crevillen, A., Hansen, L.L., Fineran, P.C. et al. (2017) Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife* **6**, e27601 <https://doi.org/10.7554/eLife.27601>
- 63 Fineran, P.C. and Charpentier, E. (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202–209 <https://doi.org/10.1016/j.virol.2012.10.003>
- 64 Swarts, D.C., Mosterd, C., van Passel, M.W. and Brouns, S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7**, e35888 <https://doi.org/10.1371/journal.pone.0035888>
- 65 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 <https://doi.org/10.1038/ncomms1937>
- 66 Staals, R.H., Jackson, S.A., Biswas, A., Brouns, S.J., Brown, C.M. and Fineran, P.C. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. *Nat. Commun.* **7**, 1–13 <https://doi.org/10.1038/ncomms12853>
- 67 Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A. et al. (2016) Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl Acad. Sci. U.S.A.* **113**, 7626–7631 <https://doi.org/10.1073/pnas.1602639113>
- 68 Kieper, S.N., Almendros, C., Behler, J., McKenzie, R.E., Nobrega, F.L., Haagsma, A.C. et al. (2018) Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.* **22**, 3377–3384 <https://doi.org/10.1016/j.celrep.2018.02.103>
- 69 Kieper, S.N., Almendros, C. and Brouns, S.J.J. (2019) Conserved motifs in the CRISPR leader sequence control spacer acquisition levels in type I-D CRISPR-Cas systems. *FEMS Microbiol. Lett.* **366**, fnz129 <https://doi.org/10.1093/femsle/fnz129>
- 70 Kieper, S.N., Almendros, C., Haagsma, A.C., Barendregt, A., Heck, A.J. and Brouns, S.J. (2021) Cas4–Cas1 is a protospacer adjacent motif–processing factor mediating half-site spacer integration during CRISPR adaptation. *CRISPR J.* **4**, 536–548 <https://doi.org/10.1089/crispr.2021.0011>
- 71 Vink, J.N.A., Baijens, J.H.L. and Brouns, S.J.J. (2021) PAM-repeat associations and spacer selection preferences in single and co-occurring CRISPR-Cas systems. *Genome Biol.* **22**, 281 <https://doi.org/10.1186/s13059-021-02495-9>
- 72 Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W. and Doudna, J.A. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904–912 <https://doi.org/10.1016/j.str.2009.03.019>
- 73 Lee, H., Zhou, Y., Taylor, D.W. and Sashital, D.G. (2018) Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell* **70**, 48–59.e45 <https://doi.org/10.1016/j.molcel.2018.03.003>
- 74 Zhang, J., Kasciukovic, T. and White, M.F. (2012) The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS ONE* **7**, e47232 <https://doi.org/10.1371/journal.pone.0047232>
- 75 Lee, H., Dhingra, Y. and Sashital, D.G. (2019) The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. *Elife* **8**, e44248 <https://doi.org/10.7554/eLife.44248>
- 76 Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M. et al. (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* **163**, 840–853 <https://doi.org/10.1016/j.cell.2015.10.008>
- 77 Jackson, S.A., Birkholz, N., Malone, L.M. and Fineran, P.C. (2019) Imprecise spacer acquisition generates CRISPR-Cas immune diversity through primed adaptation. *Cell Host Microbe* **25**, 250–260.e254 <https://doi.org/10.1016/j.chom.2018.12.014>
- 78 Kim, S., Loeff, L., Colombo, S., Jergic, S., Brouns, S.J.J. and Joo, C. (2020) Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* **579**, 141–145 <https://doi.org/10.1038/s41586-020-2018-1>
- 79 Jansen, R., Embden, J.D., Gastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 <https://doi.org/10.1046/j.1365-2958.2002.02839.x>
- 80 Nuñez, J.K., Bai, L., Harrington, L.B., Hinder, T.L. and Doudna, J.A. (2016) CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* **62**, 824–833 <https://doi.org/10.1016/j.molcel.2016.04.027>
- 81 Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y. et al. (2017) Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in *sulfolobus islandicus*. *Nucleic Acids Res.* **45**, 8978–8992 <https://doi.org/10.1093/nar/gkx612>
- 82 Li, M., Gong, L., Zhao, D., Zhou, J. and Xiang, H. (2017) The spacer size of I-B CRISPR is modulated by the terminal sequence of the protospacer. *Nucleic Acids Res.* **45**, 4642–4654 <https://doi.org/10.1093/nar/gkx229>
- 83 Scholz, I., Lange, S.J., Hein, S., Hess, W.R. and Backofen, R. (2013) CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS ONE* **8**, e56470 <https://doi.org/10.1371/journal.pone.0056470>
- 84 Sashital, D.G., Jinek, M. and Doudna, J.A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **18**, 680–687 <https://doi.org/10.1038/nsmb.2043>
- 85 Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 <https://doi.org/10.1038/nature10886>
- 86 Jesser, R., Behler, J., Benda, C., Reimann, V. and Hess, W.R. (2019) Biochemical analysis of the Cas6-1 RNA endonuclease associated with the subtype I-D CRISPR-Cas system in *Synechocystis* sp. PCC 6803. *RNA Biol.* **16**, 481–491 <https://doi.org/10.1080/15476286.2018.1447742>
- 87 Hatoum-Aslan, A., Mariv, I. and Marraffini, L.A. (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl Acad. Sci. U.S.A.* **108**, 21218–21222 <https://doi.org/10.1073/pnas.1112832108>
- 88 Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 <https://doi.org/10.1016/j.femsre.2004.12.008>

- 89 Makarova, K.S., Anantharaman, V., Grishin, N.V., Koonin, E.V. and Aravind, L. (2014) CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front. Genet.* **5**, 102 <https://doi.org/10.3389/fgene.2014.00102>
- 90 Hein, S., Scholz, I., Voß, B. and Hess, W.R. (2013) Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol.* **10**, 852–864 <https://doi.org/10.4161/ma.24160>
- 91 Kopf, M., Klähn, S., Scholz, I., Matthiessen, J.K., Hess, W.R. and Voß, B. (2014) Comparative analysis of the primary transcriptome of *Synechocystis* sp. PCC 6803. *DNA Res.* **21**, 527–539 <https://doi.org/10.1093/dnares/dsu018>
- 92 He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S. et al. (2018) Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat. Microbiol.* **3**, 461–469 <https://doi.org/10.1038/s41564-018-0120-z>
- 93 Xue, C., Zhu, Y., Zhang, X., Shin, Y.K. and Sashital, D.G. (2017) Real-time observation of target search by the CRISPR surveillance complex cascade. *Cell Rep.* **21**, 3717–3727 <https://doi.org/10.1016/j.celrep.2017.11.110>
- 94 Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S. et al. (2016) Structural basis for promiscuous PAM recognition in type I-E cascade from *E. coli*. *Nature* **530**, 499–503 <https://doi.org/10.1038/nature16995>
- 95 Guo, T.W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A. et al. (2017) Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex. *Cell* **171**, 414–426.e412 <https://doi.org/10.1016/j.cell.2017.09.006>
- 96 Hrle, A., Maier, L.K., Sharma, K., Ebert, J., Basquin, C., Urlaub, H. et al. (2014) Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *RNA Biol.* **11**, 1072–1082 <https://doi.org/10.4161/ma.29893>
- 97 Manav, M.C., Van, L.B., Lin, J., Fuglsang, A., Peng, X. and Brodersen, D.E. (2020) Structural basis for inhibition of an archaeal CRISPR-Cas type I-D large subunit by an anti-CRISPR protein. *Nat. Commun.* **11**, 5993 <https://doi.org/10.1038/s41467-020-19847-x>
- 98 Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E. et al. (2019) CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* **10**, 5302 <https://doi.org/10.1038/s41467-019-13226-x>
- 99 Vercoe, R.B., Chang, J.T., Dy, R.L., Taylor, C., Gristwood, T., Clulow, J.S. et al. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* **9**, e1003454 <https://doi.org/10.1371/journal.pgen.1003454>
- 100 Hu, C., Ni, D., Nam, K.H., Majumdar, S., McLean, J., Stahlberg, H. et al. (2022) Allosteric control of type I-A CRISPR-Cas3 complexes and establishment as effective nucleic acid detection and human genome editing tools. *Mol. Cell* **82**, 2754–2768.e2755 <https://doi.org/10.1016/j.molcel.2022.06.007>
- 101 Domgaard, H., Cahoon, C., Armbrust, M.J., Redman, O., Thomas, A. and Jackson, R.N. (2022) Casding is an ATP-dependent 5'-3' DNA helicase with accessory domains essential for type IV CRISPR immunity. *bioRxiv*, 1–40 <https://doi.org/10.1101/2022.08.23.504870>
- 102 Hwang, S. and Maxwell, K.L. (2019) Meet the anti-CRISPRs: widespread protein inhibitors of CRISPR-Cas systems. *CRISPR J* **2**, 23–30 <https://doi.org/10.1089/crispr.2018.0052>
- 103 Malone, L.M., Birkholz, N. and Fineran, P.C. (2021) Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Curr. Opin. Biotechnol.* **68**, 30–36 <https://doi.org/10.1016/j.copbio.2020.09.008>
- 104 Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gummy, L.F. et al. (2020) A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat. Microbiol.* **5**, 48–55 <https://doi.org/10.1038/s41564-019-0612-5>
- 105 Wang, M., Jiang, Y.Y., Kim, K.M., Qu, G., Ji, H.F., Mittenenthal, J.E. et al. (2011) A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.* **28**, 567–582 <https://doi.org/10.1093/molbev/msq232>
- 106 Koonin, E.V. and Makarova, K.S. (2019) Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180087 <https://doi.org/10.1098/rstb.2018.0087>
- 107 Hudaiberdiev, S., Shmakov, S., Wolf, Y.I., Terns, M.P., Makarova, K.S. and Koonin, E.V. (2017) Phylogenomics of Cas4 family nucleases. *BMC Evol. Biol.* **17**, 232 <https://doi.org/10.1186/s12862-017-1081-1>
- 108 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* **337**, 816–821 <https://doi.org/10.1126/science.1225829>
- 109 Zetsche, B., Gootenberg, J.S., Abudayeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P. et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 <https://doi.org/10.1016/j.cell.2015.09.038>
- 110 Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 <https://doi.org/10.1038/nbt.2508>
- 111 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 <https://doi.org/10.1126/science.1231143>
- 112 Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S. et al. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 <https://doi.org/10.1038/nbt.2675>
- 113 Hidalgo-Cantabrana, C. and Barrangou, R. (2020) Characterization and applications of Type I CRISPR-Cas systems. *Biochem. Soc. Trans.* **48**, 15–23 <https://doi.org/10.1042/BST20190119>
- 114 Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N. et al. (2016) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res.* **44**, e34 <https://doi.org/10.1093/nar/gkv1044>
- 115 Pyne, M.E., Bruder, M.R., Moo-Young, M., Chung, D.A. and Chou, C.P. (2016) Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *clostridium*. *Sci. Rep.* **6**, 25666 <https://doi.org/10.1038/srep25666>
- 116 Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W. et al. (2019) Characterization and repurposing of the endogenous Type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res.* **47**, 11461–11475 <https://doi.org/10.1093/nar/gkz940>
- 117 Hidalgo-Cantabrana, C., Goh, Y.J. and Barrangou, R. (2019) Characterization and repurposing of Type I and Type II CRISPR-Cas systems in bacteria. *J. Mol. Biol.* **431**, 21–33 <https://doi.org/10.1016/j.jmb.2018.09.013>
- 118 Dolan, A.E., Hou, Z., Xiao, Y., Gramelspacher, M.J., Heo, J., Howden, S.E. et al. (2019) Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. *Mol. Cell* **74**, 936–950.e935 <https://doi.org/10.1016/j.molcel.2019.03.014>
- 119 Cameron, P., Coons, M.M., Klompe, S.E., Lied, A.M., Smith, S.C., Vidal, B. et al. (2019) Harnessing type I CRISPR-Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 <https://doi.org/10.1038/s41587-019-0310-0>

- 120 Csörgő, B., León, L.M., Chau-Ly, I.J., Vasquez-Rifo, A., Berry, J.D., Mahendra, C. et al. (2020) A compact Cascade-Cas3 system for targeted genome engineering. *Nat. Methods* **17**, 1183–1190 <https://doi.org/10.1038/s41592-020-00980-w>
- 121 Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R. and Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio* **5**, e00928-13 <https://doi.org/10.1128/mBio.00928-13>
- 122 Caliendo, B.J. and Voigt, C.A. (2015) Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat. Commun.* **6**, 6989 <https://doi.org/10.1038/ncomms7989>
- 123 Luo, M.L., Mullis, A.S., Leenay, R.T. and Beisel, C.L. (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43**, 674–681 <https://doi.org/10.1093/nar/gku971>
- 124 Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P.R. and Lundgren, M. (2015) Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.* **43**, 237–246 <https://doi.org/10.1093/nar/gku1257>
- 125 Chen, Y., Liu, J., Zhi, S., Zheng, Q., Ma, W., Huang, J. et al. (2020) Repurposing type I-F CRISPR-Cas system as a transcriptional activation tool in human cells. *Nat. Commun.* **11**, 3136 <https://doi.org/10.1038/s41467-020-16880-8>
- 126 Young, J.K., Gasior, S.L., Jones, S., Wang, L., Navarro, P., Vickroy, B. et al. (2019) The repurposing of type I-E CRISPR-Cascade for gene activation in plants. *Commun. Biol.* **2**, 383 <https://doi.org/10.1038/s42003-019-0637-6>



Supplementary Figure 1. Alternative gene arrangements occur in the type I-D CRISPR-Cas system according to phylogenetic group. The figure schematically shows selected CRISPR–*cas* loci of the type I-D CRISPR-Cas system grouped according to phylum. Homologous genes are colour-coded and identified by a family name. Internally expressed *cas11* annotations have been removed, although *Haloarcula* sp. Atlit-47R appears to have a separate *cas11* gene unlike other type I-D systems [7, 19]. Archaeon *Sulfolobus islandicus* LAL14/1 had distinct gene rearrangements compared to other systems, which is supported by previous phylogenetic analysis that indicated Cas10d from archaeon *Sulfolobus islandicus* LAL14/1 did not group with other archaea or bacteria [7, 19]. Note that only the CRISPR array directly adjacent to the adaptation genes is shown for each representative. Other arrays may be present nearby or elsewhere in these genomes.