Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems

Mitchell R. O'Connell

Department of Biochemistry and Biophysics, *School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, USA*
Center for RNA Biology, *University of Rochester, Rochester, NY 14642, USA*

https://doi.org/10.1016/j.jmb.2018.06.029
Edited by Mazhar Adli

Abstract

Prokaryotic adaptive immune systems use Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and CRISPR-associated (Cas) proteins for RNA-guided cleavage of foreign genetic elements. The focus of this review, Type VI CRISPR–Cas systems, contain a single protein, Cas13 (formerly C2c2) that when assembled with a CRISPR RNA (crRNA) forms a crRNA-guided RNA-targeting effector complex. Type VI CRISPR–Cas systems can be divided into four subtypes (A–D) based on Cas13 phylogeny. All Cas13 proteins studied to date possess two enzymatically distinct ribonuclease activities that are required for optimal interference. One RNase is responsible for pre-crRNA processing to form mature Type VI interference complexes, while the other RNase activity provided by the two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains, is required for degradation of target-RNA during viral interference. In this review, I will compare and contrast what is known about the molecular architecture and behavior of Type VI (A–D) CRISPR–Cas13 interference complexes, how this allows them to carry out their RNA-targeting function, how Type VI accessory proteins are able to modulate Cas13 activity, and how together all of these features have led to the rapid development of a range of RNA-targeting applications. Throughout I will also discuss some of the outstanding questions regarding Cas13's molecular behavior, and its role in bacterial adaptive immunity and RNA-targeting applications.

Introduction

Prokaryotic Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNAs and CRISPR-associated (Cas) proteins act together to function as an adaptive immune system to protect bacteria and archaea against foreign genetic elements such as bacteriophages and plasmids [1–5]. CRISPR–Cas systems specifically act as RNA-guided programmable nucleases to degrade DNA and/or RNA of predominately exogenous nucleic acids by harnessing a genetic molecular memory of previous infections [6–14]. At the molecular level, mounting this adaptive response requires three effectively distinct processes: adaptation, CRISPR RNA (crRNA) biogenesis, and interference (for review, see Ref. [2]). Adaptation (i.e., updating the molecular memory bank to include information about the most recent infection) requires the acquisition of foreign DNA sequences into a CRISPR array. A CRISPR array is repeating genomic sequence containing multiple copies of a short segment of DNA comprising alternating semi-palindromic repeats and "spacer" sequences, which are inserted by integrase-like Cas proteins (in most cases, Cas1/Cas2; for review, see Ref. [15]).

The CRISPR array is then transcribed to generate a pre-CRISPR-RNA (pre-crRNA), which is then processed nucleolytically by either Cas proteins and/or host factors to generate individual mature crRNAs (for review, see Ref. [16]). These crRNAs then form a complex with a subset of the system's Cas proteins to form an "effector" or "interference" complex that enables crRNA-guided scanning of nucleic acid present in the cell (DNA and/or RNA depending on the CRISPR–Cas type) and once sufficient base-pairing complementarity between the crRNA-spacer and a target is found, the Cas effector complex is able to cleave and promote degradation of the identified nucleic acid molecule, helping to prevent further infection [2, 3, 6–14].

CRISPR–Cas systems can be classified into two general classes based on the protein constituents of their Cas effector complexes [17]. Class 1 CRISPR–Cas effector complexes are assembled from a crRNA and multiple protein subunits, typically encoded by three to six distinct *cas* genes [18] . In contrast, Class 2 CRISPR–Cas effector complexes are assembled from a crRNA, which in some cases also hybridized with another CRISPR–Cas encoded small RNA known as a trans-activating crRNA (tracrRNA) [19], and a single Cas protein [6, 8, 20–23]. Interestingly, Class 2 systems are significantly rarer within prokaryotes compared to Class I systems, with only 10% of CRISPR–Cas loci within sequenced genomes belonging to Class II systems, with almost all within the bacterial kingdom [24, 25]. Class 1 and Class 2 CRISPR–Cas systems can be further divided into different types based on the composition of their Cas protein cache: types I, III, and IV (Class 1), and types II, V, and VI (Class 2), which to date are currently subdivided into approximately 30 distinct subtypes (usually denoted using a proceeding letter, e.g., Type II-A) [26]. The abbreviated nature of Class II systems and their ability to be readily reprogrammed to bind, and/or cleave specific sequences of DNA or RNA have resulted a revolution in genome interrogation and engineering. CRISPR–Cas proteins have been used to edit genomes, track molecules in live cells, or perturb the regulation of specific genes, in a range of organisms [2–4].

The focus of this review, Type VI CRISPR–Cas systems, can be divided in four subtypes (A–D) based on the phylogeny of their effector complexes [26–28] (Fig. 1a). All Type VI systems include a single "effector" protein, known as Cas13 (formerly C2c2 in type VI-A systems), that when assembled with crRNA forms a crRNA-guided RNA-targeting effector complex [27–31]. All Cas13 proteins studied to date possess two enzymatically distinct ribonuclease activities that

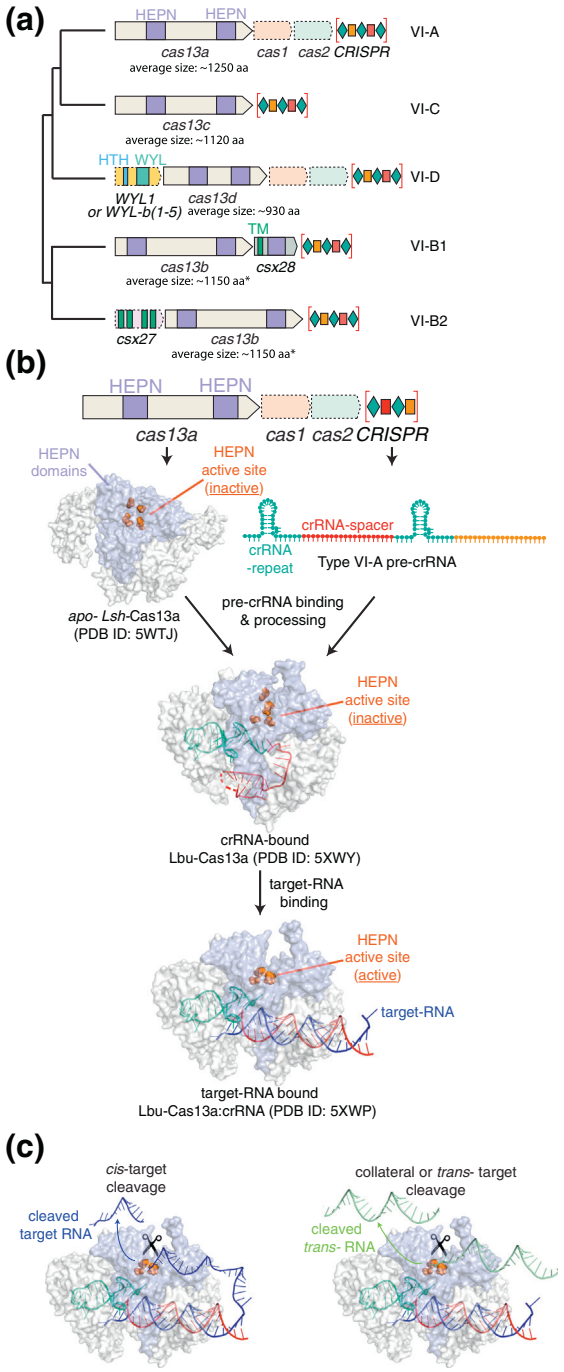

Fig. 1. Diversity and functional overview of Type VI CRISPR–Cas systems. (a) Schematic comparing the architectures of the genomic loci of known Type VI subtypes. The phylogenetic tree (adapted from Ref. [28]) is for illustrative purposes only to highlight general relationships between the subtypes; branch lengths are not to scale, and evolutionary distances are not indicated. It should be noted that the sequence identity is very low (~11–16%) when comparing Cas13 across type VI subtypes. Gene cartoons are labeled with gene names and are approximately to scale. The average size of the Cas13 protein within each subtype is indicated (* the size shown for Cas13b includes both VI-B1 and VI-B2 subtypes). Gene cartoons outlined with dashed lines indicate that these genes are not always found within the indicated Type VI subtype. Abbreviations: HEPN, Higher Eukaryotes and Prokaryotes Nucleotide-binding domain; HTH, helix-turn-helix domain; WYL, WYL domain; TM, predicted transmembrane-spanning region. Within each CRISPR array, green diamonds represent the CRISPR repeats, while orange and red rectangles represent the spacer sequences. (b) Schematic detailing the process of Cas13 HEPN nuclease activation through complementary target-RNA binding. The HEPN nuclease domains are colored in purple and the HEPN active site residues are shown as spheres in orange, crRNA-repeat is colored green, the crRNA-spacer is colored red, and the target-RNA is colored blue. Protein Data Bank coordinates are provided for each structure used to make this fig. (c) A schematic highlights that Cas13's HEPN domains once activated can cleave target-RNAs in both *cis*- (left, cleaved RNA is colored blue) and *trans*- (right, cleaved RNA is colored green). Cas13 coloring is the same as in (b).

are required for optimal interference [27, 28, 30–32]. One RNase is responsible for pre-crRNA processing to help form mature Type VI interference complexes [27, 28, 30–32], while the other RNase activity, provided by the two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains, is required for degradation of target-RNA during viral interference [27–31] (Fig. 1b, c). These properties of Cas13 have together help to herald the rapid development of a new generation of RNA-targeting tools for applications in research and therapeutics [30, 33–37]. In addition to Cas13, some Type VI CRISPR–Cas systems contain additional "accessory" proteins that, while not essential for Cas13 RNA-guided RNase activity, are able to modulate RNA interference activity, either positively or negatively [27, 31] (Fig. 1a).

In this review, I will compare and contrast what is known about the molecular architecture and behavior Type VI (A–D) CRISPR–Cas13 interference complexes, how this allows them to carry out their RNA-targeting function, and how these features have led to the rapid development of a range of RNA-targeting applications. Throughout I will also discuss some of the outstanding questions regarding Cas13's molecular behavior, and its role in bacterial adaptive immunity and RNA-targeting applications.

Type VI CRISPR–Cas systems: discovery, diversity, and abundance

Spurred on by the discovery of Type V CRISPR systems (containing the effector Cpf1/Cas12) [25, 38–40], which suggested that other Class II CRISPR–Cas effector proteins beyond Cas9 exist in nature, Shmakov *et al*. [20] developed a computational pipeline to search the whole NCBI WGS database based on the presence of *cas 1* to identify unclassified candidate Class 2 CRISPR loci. These candidate loci were then further filtered to identify loci that contained large uncategorized (>500 amino acid) proteins. This approach yielded approximately 50 new CRISPR–Cas loci that vastly expanded the number of Type V systems, and first introduced VI-A CRISPR–Cas systems and a new Class 2 effector, C2c2 (now known as Cas13), to the world. More recent extensions of this type of approach (but instead using CRISPR-repeat arrays as an anchor) have expanded the known Class II CRISPR–Cas systems to include to now include at least 13 subtypes (plus five additional tentative subtypes) [24, 26, 31, 41]. Importantly, these analyses revealed that at the time, four distinct Type VI CRISPR–Cas subtypes exist (A, B1, B2, and C, with D to be discovered more recently, see below) (Fig. 1a). The resident Cas13 within each of these subtypes shares extremely low-sequence identity with one another, which is practically limited to the active site motif of the HEPN domain [26, 31]. This Type VI subtype classification also highlighted that the HEPN domains can reside at different positions across the length of Cas13, and additional features exist in Type VI-B loci, which is further divided into B1 and B2 subtypes [26, 31]. Specifically, the VI-B1 systems possess an additional ORF, *Csx28,* that encodes a small protein with one putative transmembrane domain (or possibly signal peptide) and one divergent HEPN domain (Fig. 1a), while the VI-B2 systems in many cases (but not all) possess a different ORF, Csx27, that also encodes a small protein, which appears to contain three to four predicted transmembrane domains (Fig. 1a). It is also worth noting that in many cases, Type VI CRISPR–Cas loci lack adaptation genes (*cas1*/*cas2)*, indicating that they may utilize adaptation modules from other CRISPR–Cas loci within the same genome, or have lost their adaptation modules and are no longer actively acquiring new protospacers (Fig. 1a). More work is required within native Type VI CRISPR system hosts to understand the role of Type VI systems in nascent adaptive immunity.

Very recently, this Class II effector discovery approach was further expanded by two groups that utilized additional genomics and metagenomics data sets and an updated bioinformatics pipeline, which together enabled researchers to case a wider net to find undiscovered Class II effector proteins [27, 28]. These searches identified additional Type VI CRISPR–Cas loci, and excitingly, this led to the identification of a new subtype, Type VI-D, and an associated effector protein, Cas13d, which were found predominately in two bacterial genuses, *Eubacterium* and *Ruminococcus* [27, 28] (Fig. 1a). This subtype, although most similar to Cas13a [27] and/or Cas13c [28], only shares minimal sequence identity with Cas13a and is much smaller than all of previously discovered subtypes (e.g., median size is ~300 aa or 26% smaller than that of previously identified Cas13 proteins) [27, 28] In addition, much like Type VI-B systems, a substantial majority Type VI-D systems were shown to contain associated accessory proteins, and in this case, these proteins all contained WYL domains [27], a domain often associated with prokaryotic defense systems [42] (Fig. 1a). Experimental verification of Cas13d's activity showed that the effector protein behaved in most ways similar to previously studied Cas13s, with a few small differences, which I will discuss in more detail below.

Beyond phylogenetic characterization, functional differences within subtypes have been observed. For example, Seletsky *et al*. [32] explored the pre-crRNA processing behaviors of 11 Cas13a homologs and found that the Cas13a enzyme family comprises two distinct functional groups (named the U- and A-cleaving subfamilies) that recognize orthogonal sets of crRNAs and possess different single-stranded RNA (ssRNA) cleavage specificities (discussed in more detail below) (Fig. 1a).

Cas13 domain organization

Similar to other Class 2 effector proteins, Cas9 and Cas12, high-resolution structures of Cas13a from several orthologs have revealed that Cas13a adopts a "bi-lobed" globular protein architecture consisting of one crRNA "Recognition" lobe (REC) and one "Nuclease" lobe (NUC) [43–46] (Fig. 2a). In step with their very

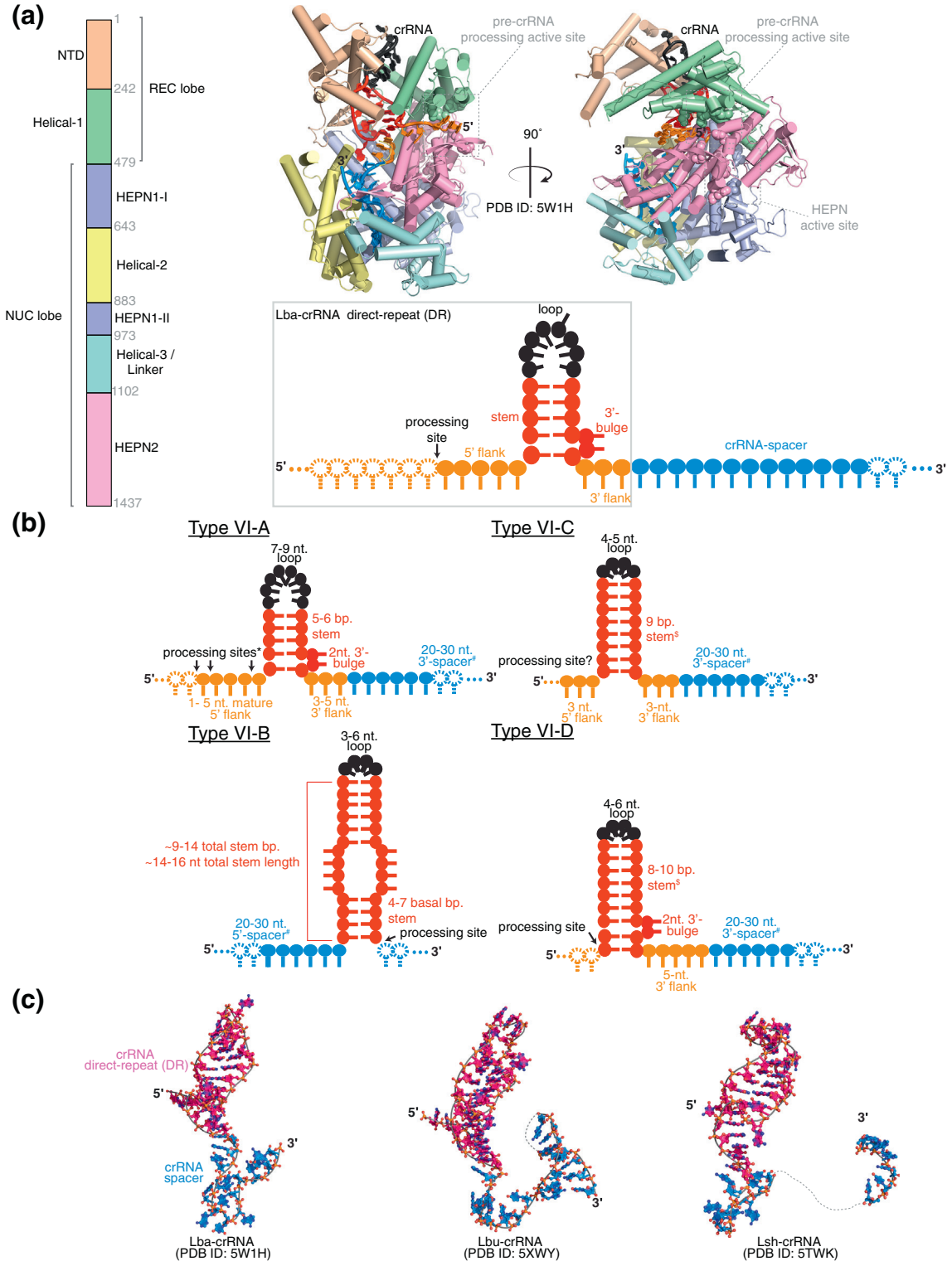

Fig. 2 (*legend on next page*)

different biochemical activities, this is where the similarity between Cas13 and Cas9/Cas12 ends, as no additional structural similarity can be detected between these proteins [43–45]. Indeed, while Cas9 and Cas12 can be defined by the presence of a common RuvC nuclease domain, whereas Cas13 invariably contains two distinct HEPN nuclease domains [20], which are commonly found in multiple systems involved in biological conflicts and processing of cellular RNAs, including (but not limited to) prokaryotic toxin–antitoxin and abortive infection defense systems, as well as in KEN (kinase-extension nuclease) domain-containing eukaryotic ribonucleases such as RNase L and Ire1 (for review, see Ref. [47]). This led to the initial hypothesis that Cas13 was an RNA-targeting enzyme [20]. High-resolution structural studies of Cas13 have allowed for the determination of domains within the REC and NUC lobes, namely, an N-terminal domain (NTD) and a Helical-1 domain that form the REC lobe, and a split HEPN domain that contains a Helical-2 domain insertion followed by a Linker/Helical-3 domain, and a second HEPN domain that together forms the NUC lobe (Fig. 2a) [43–45]. Interestingly, from primary sequence analysis, the equivalent region to the Helical-1 domain of Cas13a appears to be absent in the smaller Cas13d family [27, 28], suggesting that Cas13d may rely solely on the NTD domain for crRNA direct repeat (DR) recognition or utilize another part of the protein to carry out the same function. Indeed, a very recent cryo-EM structure of Cas13d in complex with its cognate crRNA confirmed that the NTD alone is primarily responsible for DR stem recognition, with additional contacts contributed by the HEPN1/2 domain to the lower DR stem [46] (see below for more details on crRNA recognition and processing).

Type VI crRNA architecture

Like all other CRISPR–Cas systems, crRNAs in type VI systems are first transcribed as pre-crRNAs that require processing to generate single mature crRNA species (for review, see Ref. [16]). In Class 2 systems, there are two ways in which this has been shown to occur. In Type II systems, the DR region of the crRNA hybridizes to a tracrRNA which is then bound by Cas9 while being recognized by the host factor nuclease RNase III, which acts to cleave the double-stranded RNA region of the repeat:tracrRNA duplex and generate mature crRNAs:tracrRNA hybrids [19]. The requirement for a tracrRNA and presumably RNaseIII activity has been also shown for Type V-B and Type V-E systems [20, 24]. In contrast, it has been shown for at least Type V-A systems and all Type VI systems tested to date that crRNA processing is carried out by a dedicated RNA nuclease within Cas12a [48] and Cas13 [30] (Fig. 2a). Type VI crRNAs adopt a comparatively simple structure of a DR forming single short hairpin loop "handle," which is flanked by a 5′ or 3′ ~20–30-nt spacer "guide" sequence (Fig. 2a–c) [20, 26, 29–35].

In high-resolution structures of Cas13a:crRNA binary complexes (Fig. 2b), the DR forms a stem of five to six Watson–Crick base pairs, disrupted by an invariant 2-nt bulge (either AC or AA) and depending on the Cas13a homolog, a stem–loop region of 7–9 nt [43–45]. From high-resolution structures of crRNA-bound Cas13a from *Leptotrichia shahii* (Lsh) [43], *Leptotrichia buccalis* (Lbu) [44], and *Lachnospiraceae bacterium* (Lba) [45], the bulge nucleotides are flipped out, adopting a single-stranded conformation and make specific contacts with the protein, suggesting that they are important for crRNA-recognition (Fig. 2b, c,). Upstream of the stem–loop in type VI-A pre-crRNAs is a 5′ 8- to 12-nt single-stranded region required for initial Cas13 recognition and downstream pre-crRNA processing [29, 30, 45]. Pre-crRNA processing results in cleavage of this single-stranded region to a final length of 1–5 nt depending on the Cas13a ortholog (Fig. 2b) [20, 29, 30, 32, 45]. Downstream of the stem–loop, a 3- to 5-nt 3′ single-stranded region is present and is required for Cas13a function [29, 30].

Although currently no high-resolution structures of Type VI-B crRNAs exist, RNA secondary structure prediction and conservation analysis show that in contrast to VI-A, VI-C, and VI-D crRNAs, VI-B crRNAs possess much longer DR stems (~9–14 bp in total) with often several unpaired regions and bulges, and consequently smaller stem–loops (Fig. 2b) [26, 31, 35]. In addition, some Type VI-B CRISPR loci contain additional crRNA units with much longer DRs (~88-nt length) can be processed by Cas13b and form active interference complexes, the reason for their existence is currently not understood [31]. Beyond stem–loop

Fig. 2. Structural features of Cas13a and Type VI crRNAs. (a) Lba Cas13a domain organization schematic (left), a crystal structure of Lba-Cas13a bound to its cognate crRNA (upper right), and a cartoon schematic of the Lba-crRNA (lower right). HEPN and crRNA-processing nuclease active sites are shown in the crystal structure as spheres and are labeled in the figure. (b) Schematic highlighting the important similarities and differences between crRNAs from each of the different Type VI system subtypes. [#]Spacer length can vary when expressed in *E. coli* due to trimming by host nucleases; ranges are shown. [$]Stems in Type VI-C and VI-D system sometimes contain non Watson–Crick base pairing, mismatches, or bulges. Known pre-crRNA processing sites are indicated by an arrow. *In Type VI-A systems, the processing site is different between groups of Cas13 orthologs. (c) Crystal structures of Cas13a-bound crRNAs (for clarity, only crRNAs are shown) from Type VI-A systems from Lba (left), Lbu (middle), and Lsh (right). The crRNA repeat is shown in pink, and the crRNA-spacer is shown in blue. Regions where the crRNA is disordered in the crystal structure are shown using a dotted line.

structure, the order of the direct repeat relative to the spacer is also different in Type VI-B systems, with the direct-repeat stem–loop residing at the 3′ of the spacer, in contrast to the 5′ DR—spacer structure observed in type VI-A, VI-C and VI-D systems (Fig. 2b) [31]. In contrast to Type VI-A and VI-B crRNAs, Type VI-C crRNAs DRs are shorter in length (~30-nt) [20], suggesting that they may have co-evolved in the presence of a different adaptation module than VI-A, B, and D systems. Finally, the most recent addition to the Type VI family, Type VI-D, contains crRNAs with a 36-nt DR forming a conserved 8- to 10-nt stem with a 4- to 6-nt loop, a 5- to 10-nt 5′ flanking single-stranded region (in the pre-crRNA), and a 5- to 7-nt 3′ flanking single-stranded region that contains a conserved terminal AAAAC motif (Fig. 2b) [27, 28]. In a very recent cryo-EM structure of a binary Cas13d:crRNA complex from *Eubacterium siraeum (*Es-Cas13d), the DR forms a stem of 9 bp, six of which are surprisingly formed through non-Watson–Crick interactions [46]. This stem is also disrupted by 2-nt bulge and contains a stem–loop region of 4 nt, which may suggest that this 2-nt bulge is a common feature of at least Type VI-A, VI-D, and perhaps VI-C crRNAs.

It is also worth noting that protospacer matches have been identified for a few crRNA repeats within Type VI loci. For example, Smargon *et al*. [31] found that 36 crRNA spacers from the Type VI-B systems in their study matched with 80% or more homology to unique sequences within phage genomes. Of these, the vast majority (75%) were complementary to the coding strand of phage protein genes, suggesting that in line with the RNA-substrate preference of Cas13b, they are likely targeting phage mRNAs. Most recently, when analyzing the spacers within Type VI-D systems studied, Yan *et al*. [27] discovered that seven protospacers matched with high confidence to either dsDNA phage sequences, prophage regions within other bacterial genomes, or open-reading frames not associated with prophages in other prokaryotic sequences. More work needs to be done to understand how mechanistically Type VI systems elicit antiviral defense within their host organisms.

crRNA-direct repeat recognition by Cas13

crRNA-processing Cas proteins and Cas interference complexes have evolved to bind their cognate crRNAs with high affinity and specificity (for review, see Ref. [16]). This is generally achieved through a combination of conserved RNA base and structure-specific contacts. High-resolution structures for Lsh-, Lbu-, and Lba-Cas13a, and Es-Cas13d have begun to reveal the requirements for this interaction [43–45]. In all Cas13a structures, the stem–loop formed by the crRNA DR is bound between a cleft formed mostly by an interface between the NTD and Helical-I domains of Cas13a, and an extensive number of conserved non-covalent contacts stabilize the interaction between the phosphate backbone of the crRNA stem–loop and Cas13a. Additional stabilization is provided by extensive interactions between the 2′-hydroxyl groups of the crRNA and amino acid side chains that extended NTD, Helical-1, and HEPN2 domains. crRNA recognition requires a large conformation change by apo-Cas13a, with the Helical-2 domain rotating toward the HEPN2 and Linker domains, generating the crRNA-binding channel [43]. The more compact Es-Cas13d on the other hand, as highlighted above, lacks a large portion of the Helical-1 domain, and in the binary Es-Cas13d:crRNA structure, the NTD domain is mostly responsible for contacting the crRNA DR with a smaller set of contacts emanating from the HEPN1/2 domains [46]. In addition, in contrast to all other crRNA structures solved to date, the Es-Cas13d:crRNA complex chelates an Mg^{2+} ion within this bulge region using a combination of contacts from the Cas13d and the crRNA, likely facilitating additional stabilization of the binary complex [46].

Beyond contacts with the stem of the crRNA DR, extensive contacts are made by Cas13a with the large (7–9 nt) crRNA loop. In each of the three currently available structures of Cas13a in complex with their cognate crRNAs, the loop regions of each crRNA adopt a different conformation (Fig. 2c) [43–45]. For Lba-Cas13a, the crRNA loop is condensed into a helical stack and only a single adenosine loop nucleotide is flipped into the solvent [45]. This is in contrast to the structure of Lsh-Cas13a's crRNA, in which ~4 nt in total are flipped out of the loop [43] (Fig. 2c). In both cases, several base-stacking and hydrogen binding interactions between nucleotides within loop help to stabilize its structure, which is further stabilized by extensive intermolecular interactions with mostly Helical-I and to a lesser extent the NTD domains of Cas13a [43, 45]. This is not the case with Es-Cas13d and its crRNA, where the upper portion of the crRNA DR stem and the 4-nt loop both extend out into solvent and no interactions between Cas13d and the crRNA are made [46].

In addition to the stem–loop, the mature VI-A crRNA repeat contains a single-stranded 5′ flank that sits in a non-A-form conformation inside a shallow surface-exposed groove at the interface between Helical-1 and HEPN2 domains [43–45] (Fig. 2a). Here it is stabilized by a combination of specific hydrogen bonds, stacking interactions with aromatic residues, and charge–charge interactions with the crRNA phosphate backbone [43–45]. In the A-cleaving Cas13a sub-family [32], after pre-crRNA processing, mature crRNAs maintain a 5-nt 5′ flank that wraps around the exterior of the protein, with A(−25) and G(−26) stacking with F1300 and F422, respectively [45]. In addition, the two conserved adenosine nucleotides are recognized in a base-specific manner by hydrogen bonds by side chains from the Helical-I and HEPN2 domains, suggesting that specific recognition of these adenosines is required to obtain crRNA specificity [45].

In contrast, most Cas13a homologs from the U-cleaving subfamily generate a 4-nt 5′ flank, with the only exception to date being Lsh-Cas13a, which generates a 1-nt 5′ flank (Fig. 2b). It is less clear about the specific 5′ flank nucleotide requirements for the U-cleaving subfamily, as only a single adenosine (A2) is read out in a base-specific manner by Cas13a, through a backbone carbonyl H-bond at F1102 [44]. Mutation of this single adenosine (to a guanosine) results in a ~5-fold decrease in binding affinity between Lbu-Cas13a and its cognate crRNA [30]. In Lsh-Cas13a, it appears that only the 5′ C(1) is read out by base-specific interactions with Cas13a side chains [43].

In contrast, nucleotides within type VI-A crRNA's DR 3′ flanking region maintain a nearly perfect A-form conformation along the ribose-phosphate backbone, with a slight kink between the last two nucleotides of the repeat (Fig. 2c) [43–45]. To maintain this helical conformation, extensive hydrogen bonds are formed between conserved side chains from Cas13's HEPN2 domain and phosphate and 2′-hydroxyl groups of the crRNA backbone [43–46]. Together, the organization of the DR 3′ flanking region probably serves to correctly position the 5′ end of the spacer for optimal target recognition. The length of the DR 3′ flanking region differs across the Cas13a family, with the U-cleaving subfamily possessing a 5–6 nt long flanking region, whereas this region in the A-cleaving sub family is generally short (3–4 nt) [32].

The base specificity of the loop and 5′ and 3′ DR flanks is underscored by a number of mutational studies that showed mutation of these nucleotides resulted in loss of Cas13 activity [29, 30]. In contrast, the identity of the nucleotides within the crRNA stem appears to be less important, as long as the base pairing, structure, and the length of the stem are altered significantly [29, 30] . Overall, this sensitivity to both flanking regions of the hairpin is reminiscent of the sequence and structural motifs required by many Cas6 and Cas5d enzymes (see Ref. [16] for review). Due to a lack of high-resolution structures or mutational data for Cas13b and Cas13c, little is known about how they specifically interact with their crRNA DRs.

Cas13a pre-crRNA processing mechanism

East-Seletsky *et al.* [30] were first to demonstrate that Cas13a possessed pre-crRNA processing activity, and that this activity did not require the HEPN nuclease domains. Although initial primary sequence analysis failed to determine the domain/active site responsible for this activity, extensive amino-acid mutagenesis across the length of the protein enabled the identification of residues involved in Cas13a's pre-crRNA processing activity. It was found that mutation of a single arginine residue to alanine (R1079A) within the HEPN2 domain of Lbu-Cas13a was able abrogate all pre-crRNA processing activity without effecting HEPN-domain-mediated RNA cleavage [30]. It was also shown that the pre-crRNA catalytic mechanism was (a) single turnover with the mature-crRNA remaining bound after cleavage, (b) did not require the presence of any divalent metals, (c) and resulted in the formation of 5′OH and 3′ cyclic phosphate reaction products [30]. Interestingly, pre-crRNA nucleotide substitutions on both sides of the scissile phosphate affected processing, suggesting that sequence-specific recognition is required across the processing site [30].

In contrast, the first high-resolution structure of Cas13a (from Lsh) bound to a mature-crRNA showed that while the DR 5′ flank lies in a cleft between Helical-1 and HEPN2, the pre-crRNA processing active site is located wholly within the Helical-1 domain, rather than within the HEPN2 domain [43]. Specifically, Helical-I residues including a conserved R438, which lies adjacent to the ribose of C(−27) and non-conserved K441, are both critical for pre-crRNA processing (and were confirmed through mutational analysis) [43]. Interestingly, although an extensive number of residues from the HEPN2 domain are in somewhat close proximity to the 5′ crRNA terminus, individual mutagenesis of all but one (N1315) of these residues failed to impair pre-crRNA processing [43].

In light of this apparent contradiction, in a follow-up study, East-Seletsky *et al.* [32] carried out further mutagenesis of Lbu-Cas13a focusing in on the HEPN2 and Helical-1 domains an in attempt to reconcile these differences in Cas13a pre-RNA processing mechanism. It was concluded that for at least Lbu-Cas13a, specific residues within mostly HEPN2 and to a lesser extent the Helical-1 domains play significant roles in crRNA biogenesis for Lbu-Cas13a [32]. In addition, the pre-crRNA processing abilities for nine other Cas13a homologs were determined and all but Lsh-Cas13a cleaved 4–5 nt upstream of the conserved crRNA-repeat hairpin (Fig. 2b) [32].

This disagreement is more than likely due to the divergent pre-crRNA processing activity of Lsh-Cas13a. The participation of residues from the both HEPN2 and Helical-I domains in pre-crRNA processing in Lbu-Cas13a has subsequently been supported by a high-resolution structure of Lbu-Cas13a bound to a mutated pre-crRNA [44]. All of the residues previously identified to be involved in Lbu-Cas13a pre-crRNA processing [30, 32] were shown to lie in close proximity to the scissile phosphate [44]. Specifically, the side chains of R1072, R1079, and K1082 contact the scissile phosphate group, suggesting that these amino acids play important roles for pre-crRNA processing.

Most recently, an Lba-Cas13a:crRNA crystal structure and a detailed analysis of the pre-crRNA processing activity finally help reveal details regarding the precise catalytic mechanism of crRNA cleavage [45]. Similar to the previous structures, Lba-Cas13a's

crRNA DR 5′ ssRNA flank is cradled in a groove formed by the Helical-1 and HEPN2 domains forming an acid–base catalytic center for processing [45]. Lba-Cas13a positions the 5′ flank within the processing groove where two residues from Helical-1 and HEPN2, respectively (W325 and N1232), stabilize the ribose 5′ of the scissile phosphate in a distorted C2′-endo conformation [45]. It has been hypothesized that this coordination sets up the ribose 2′ hydroxyl for nucleophilic attack on the scissile phosphate, at which point a number of residues from both Helical-1 and HEPN2, as well as an activated water, participate in proton transfer to facilitate the formation of a 2′–3′-cyclic phosphate of the nucleotide 5′ to the cleavage site leaving behind a mature crRNA with a ribose 5′ hydroxyl [45]. Interestingly, it should be noted that none of the biochemically implicated residues are within hydrogen bond distance of the ribose 2′ hydroxyl [45]. The pre-crRNA-bound structure revealed that the 2′ hydroxyl of G(−29) contacts N1232 and a single water molecule that is coordinated by HEPN2 residues K1305 and K1320. These structural observations led to the hypothesis that the G(−29) 2′ hydroxyl is deprotonated by an activated water [45]. Consistent with this hypothesis, K1305 is surrounded by two acidic moieties (D1268 and E1235) that may enable K1305 to act as a general base to activate the coordinated water [45].

Cas13b and Cas13d also exhibit pre-crRNA processing activity, and the cleavage sites have been mapped using RNA sequencing and *in vitro* pre-RNA processing assays (Fig. 2b) [27, 28, 31]. In the case of Cas13b, the pre-crRNA is processed into a 66-nt mature crRNA, with a 30-nt 5′ spacer followed by the 36-nt 3′ DR (Fig. 2b) [31]. Based on the predicted structures of the repeat, it is predicted that Cas13b pre-crRNA processing active site recognizes the ssRNA junction at the base of the stem and cleaves at the ds/ss junction [31]. On the other hand, Cas13d pre-RNA processing produces mature crRNA of a similar repeat-spacer structure to the Cas13a family, specifically cleaving at the DR hairpin junction (Fig. 2b) [27, 28]. However, unlike Cas13a, which does not require divalent cations for pre-crRNA processing, it appears that Rsp-Cas13d's pre-crRNA processing activity is substantially inhibited in the presence of EDTA [27], indicating that divalent cations are required for efficient pre-crRNA cleavage, while interestingly Es-Cas13d pre-cRNA processing activity is less affected by EDTA [27, 28]. Recently, this observation was reconciled when (as noted above) an Mg^{2+} ion was observed coordinated by the crRNA in cryo-EM structure of the Es-Cas13d complex, and additional biochemical experiments show that it is required for optimal pre-crRNA processing [46]. This interesting observation might be reminiscent of the divalent metal requirements seen within the Type V effector, Cas12a, where optimal crRNA binding to Cas13a requires a coordinated Mg^{2+} [48–50]; however, the catalytic mechanism is strictly metal ion independent [49]. Given the conservation of pre-crRNA processing in Type VI-A, B, and D systems, it is likely that Cas13c will also maintain the ability to cleave pre-crRNAs. Further biochemical and structural studies will be required to elucidate whether Cas13b (and c) shares similarities with Cas13a with respect to pre-crRNA processing active site location and chemistry.

In addition to initial Cas13-mediated pre-crRNA processing, the crRNA-spacer also undergoes additional processing *in vivo*. Most crRNA-spacers start off much longer due to the combined length of the genomically encoded protospacer (~30–40 nt depending on the locus), and for Type VI-A and -D (and possibly -C) systems, the additional 3′ constant region from the downstream DR 5′ flanking region. Heterologous expression of crRNA arrays in *Escherichia coli* leads to a distribution of crRNA-spacer lengths; for example, for Lsh-Cas13a, ~14- to 28-nt length spacers were detected [29], while for Es-Cas13d (*E. siraeum*), a median spacer length of 23 nt was observed with a range from 20 to 30 nt [27], and for Ur-Cas13d (*Uncultured Ruminoccocus* sp.), a spacer range from 14 to 26 nt was observed [28]. It is likely that host RNases are responsible for this spacer-trimming activity. Because of this variability, biochemical and cell-based experiments have been required to determine "optimal" minimal crRNA-spacer length and it appears to vary anywhere from 20 to 22 nt (for Lbu-Cas13a, Lba-Cas13a, Es-Cas13d, and Lwa-Cas13a) [27, 28, 30, 33, 45] to 28–30 nt (for Lsh-Cas13a and a number of Cas13b orthologs tested in human cells) [34, 35] depending on which assay (biochemical *versus* cell-based) is being used, and which Cas13 ortholog is being tested. Further work is required to understand the structural and mechanistic details regarding optimal crRNA-spacer length. For now, the crRNA-spacer length needs to be determined empirically for each new Cas13 ortholog in question, as well as within each specific application.

Finally, one surprising result worth noting is that pre-crRNA processing activity is not strictly required for efficient trans-ssRNA cleavage, despite being conserved in most Cas13a enzymes [32]. This is in contrast to other CRISPR–Cas systems, where in the absence of pre-crRNA processing, CRISPR–Cas interference complexes are not competent for phage defense [12, 16, 19, 51–53]. In light of this observation, it was shown that the role of pre-crRNA processing is more subtle—its function serves to liberate each individual crRNA from the confines of a long CRISPR array transcript, relieving RNA folding constraints and potential steric hindrance of neighboring Cas13a:crRNA-spacer species during crRNA loading and/or ssRNA targeting. In support of this idea, it was shown that in the context of an unprocessed pre-crRNA containing six distinct mature-crRNAs, the rate of Cas13a target-RNA HEPN-nuclease cleavage activity by a crRNA-processing inactive mutant was significantly reduced

for all crRNAs sequences within the pre-crRNA array relative to wild-type Cas13a, and was more pronounced with each successive crRNA within the array, with the last crRNA resulting an 85% reduction in HEPN nuclease activity relative to WT Lbu-Cas13a [32].

The mechanism Cas13 RNA-target search and the role of crRNA-spacer organization

A hallmark of a majority of nucleic-acid guided, nucleic-acid targeting complexes studied to date is the presence of a so-called "seed" region within their guide sequence that helps direct the target-search process and stabilize the initial hybridization of the guide-strand to the target nucleic-acid [54, 55]. In high-resolution structures of guide-RNAs bound to their effector complexes, this seed region is often observed as a partially pre-ordered guide, often in an A-form or pseudo A-form helical conformation, which is often thought to reduce the entropic penalty incurred by guide: target binding hybridization [54, 55]. For example, guide pre-ordering by the argonaute proteins has been shown to increase the on-rate of RNA-guide RNA-target hybridization relative to "naked" RNA, and also increase the sensitivity to mismatches (particularly in the seed region), thus improving the specificity of the molecular interaction [56]. Similarly, a number of CRISPR–Cas interference complexes employ a similar approach to facilitate target search and increase sensitivity to mismatches (for review, see Ref. [54]).

In early work, Abudayyeh *et al*. [29] explored the effect that mismatches between crRNA-spacer of Lsh-Cas13a and a target-RNA had on the HEPN nuclease activity of Lsh-Cas13a. They revealed that mismatches in the middle of the crRNA-spacer led to the largest reductions in HEPN nuclease activity; however, it was not clear whether this effect was due to an initial binding defect, or due to the effect that mismatches have on the ability of RNA binding to activate the HEPN nuclease. It has been previously shown that these two processes can often be decoupled in RNA-guided complexes (e.g., argonaute [56, 57] and Cas9 [55, 58]).

Following on from this early work, and as discussed above, several high-resolution structures of Cas13a and Cas13d proteins have been solved in the presence of a crRNA. Interestingly, all of these structures have posed an interesting conundrum regarding the formation of an expected "pre-ordered" segment within the spacer. In the high-resolution structure of Lsh-Cas13a, unambiguous electron density was observed for almost half of the nucleotides of the 28-nt crRNA-spacer segment, predominately at the very 5′ and 3′ ends of the crRNA-spacer (Fig. 2c) [43]. Specifically, 4 nt immediately 3′ of the 3′- flanking DR extend out in a single-stranded conformation, and bind in the cleft formed by the HEPN1 and Helical-2 domains [43]. This conformation is then interrupted by a U-turn conformation at nt 4–5, resulting in nt 6–9 binding inside a groove formed by the Linker and HEPN2 domains [43]. The following 8–10 nt, which constitute the central region of the crRNA-spacer, appear to be disordered. The presence of partial, discontinuous density, allowed Liu *et al.* [43] to hypothesize that the central region is bound in a shallow/wide-open, solvent-exposed groove formed between the NTD and Helical-2 domains. Unambiguous electron density of the 3′ portion of the crRNA-spacer revealed that this portion is housed in a groove within the larger subdomain of the NTD domain (Fig. 2c) [43]. Overall, the 28-nt crRNA-spacer in this structure adopts a non-helical conformation with its 5′ and 3′ portions buried within Cas13a and its central portion exposed to solvent. Based on this, a novel mode of target search was proposed by Liu *et al*. [43], whereby a solvent-exposed disordered "lasso" is responsible for initial target binding and hybridization in this region drives a conformation change that exposes the 5′ and 3′ segments of the crRNA-spacer for a complete guide–target interaction. This mode would indeed represent a novel way for a RNA-guided protein to search for its targets [54]. Alternatively, given that the effect of mismatches on target-RNA binding has not been tested for Lsh-Cas13a, the central region might instead be required for HEPN activation rather than the initial target search and this structure of Lsh-Cas13: crRNA may instead represent a "closed" conformation, or that "seed" ordering is a meta-stable phenomenon and was not captured in this structure.

More recent structural studies of Lbu- and Lba-Cas13a in complex with crRNA have revealed several similarities to Lsh-Cas13a and some differences with respect to guide organization within Cas13a (Fig. 2c). Both Lbu- and Lba-Cas13a also contort their crRNA-spacer into a distorted U-turn configuration, with some small differences observed in the overall trajectory of the crRNA-spacer [44, 45]. In the Lba-Cas13a: 24-nt spacer structure, 13 consecutive nucleotides from the 5′ end of the crRNA repeat could be unambiguously modeled into the electron density, whereas the 11 remaining spacer nucleotides are disordered in the structure. In the ordered crRNA-spacer region, Knott *et al*. [45] observed that HEPN1/2, Helical-2, and Helical-3 make a number of conserved contacts with the sugar-phosphate backbone of the crRNA-spacer. Interestingly, the 3′ end of the ordered region of crRNA-spacer, nt 11–13, which forms part of the central "seed" region previously identified, exits the NUC lobe forming an A-form conformation that appears more solvent accessible than the preceding 10 nt of the crRNA-spacer. Interestingly, the part of the crRNA-spacer emerging from the NUC lobe adopts a similar conformation in the second structure of Lba-Cas13a (containing a 20-nt spacer), as well as in the Lsh-Cas13a:crRNA structure [43, 45].

In the Lbu-Cas13a:crRNA structure, one of the most surprising differences observed compared to

the Lba- and Lsh-Cas13a structures is the even more complete ordering of the central region of the crRNA-spacer (Fig. 2c) [44]. In this structure, crRNA-spacer nt 9–15 are ordered close to an A-form helical conformation by an interaction between their sugar-phosphates and residues that line the surface of the Helical-2 domain [44]. This conformation orients the Watson–Crick edges of the central portion of the crRNA-spacer toward the solvent, potentially facilitating the target search process. One large caveat here is that this ordering might be an artifact due to nt 9–15 forming intramolecular mismatched base pairs with nt 22–26 of the same crRNA-spacer, which naturally could order these crRNA-spacer nucleotides close to an A-form conformation. One might hypothesize that this structure more closely represents a "pseudo-ternary" complex given that the "seed" region of the spacer-crRNA is engaged in base-pairing interactions not so dissimilar to the eventual interactions it will have with complementary target-RNA. Therefore, the question remains: is this "guide-folding/organization" a general feature of Cas13a (or Lbu-Cas13a for that matter), or is it a crRNA-spacer sequence and a length-specific phenomenon? That is, in terms of length, is Lbu-Cas13a also able to engage in crRNA-spacer intramolecular base-pairing when the length of the spacer is 20-nt as opposed 28-nt? Particularly, given we know that a 20-nt spacer is sufficient for maximal cleavage activity.

Surprisingly and in contrast to all the Cas13a:crRNA structures solved, the crRNA-spacer within the very recent Cas13d:crRNA cryo-EM structure adopts yet another different structure. Instead of forming a kinked "U-turn" structure, the entire spacer is organized in an extended pseudo-helical structure and held in a solvent exposed channel by extensive backbone and 2′ hydroxyl contacts by the Helical1/2 as well as the HEPN1/2 domains [46]. This arrangement is perhaps more reminiscent of the extended seed regions observed for some Hfq-binding sRNAs (for review, see Ref. [54]), rather than what is generally observed with argonaute or CRISPR–Cas systems. Taken together, all the structural studies to date on Cas13 in complex with a crRNA suggest that the crRNA-spacer may exist in multiple conformations, and potentially only some of which can be captured in high-resolution structures.

Cas13 and crRNA conformational change upon target-RNA:crRNA duplex formation

In the absence of target-RNA, Cas13a–crRNA complexes exhibit negligible HEPN-nuclease activity, suggesting that the Cas13a and its HEPN domains exist in an auto-inhibited conformation that prevents non-discriminate RNA cleavage (Fig. 1b) [29, 30]. Recently, a structure of Lbu-Cas13a:crRNA in complex with a completely complementary target-RNA revealed for the first time how Cas13a:crRNA recognizes a ssRNA target, and the large conformation change required to accommodate a 28-bp crRNA:target-RNA duplex [44]. The availability of both a binary crRNA-bound Cas13a complex and a ternary RNA target-bound complex enables us to track the movements of each Cas13a domain (Fig. 3a), as well as conformation change within the crRNA upon stable target-RNA binding (Fig. 3b). In Cas13, the Helical-2 and HEPN1 domains undergo significant movement, while the NTD, Helical-1, and HEPN2 domains experience less movement when comparing the binary and ternary structures (Fig. 3a). In particular, the Helical-2 domain rotates away from the HEPN2 domain, allowing the HEPN1 domain to rotate relative to the HEPN2 domain, bringing the two halves of the HEPN active site which reside in HEPN1 and HEPN2 close enough together to form a catalytically competent active site capable of ssRNA cleavage (Fig. 3a, c) [44]. The crRNA also experiences significant conformation change upon stable binding of a complementary target-RNA (Fig. 3b). Most noticeable is the large change observed in the position of both the 3′ DR ssRNA flank and the crRNA-spacer within the Cas13a complex. The 3′ DR ssRNA flank experiences a dramatic shift from a near A-form conformation in the binary complex to a distorted with alternately flipped consecutive bases making a number of stacking and base-specific contacts with the HEPN-1 and HEPN-2 domains [44]. Ultimately, this large conformational shift results in an overall 15-Å movement of the first 5′ phosphate of the crRNA-spacer toward the NTD and Helical-2 domains (Fig. 3b) [44]. Downstream of the 3′ DR ssRNA flank, the crRNA-spacer reorients from a highly distorted U-turn conformation in the binary complex to forming close to a regular A-form helix with the target-RNA in the ternary complex (Fig. 3b) [44]. Once this conformation change has occurred, the resultant crRNA-spacer:target-RNA duplex is surrounded almost wholly by the nuclease lobe with two alpha helices of the Linker domain and a beta-hairpin of the HEPN2 forming a "lock" over a central positively charged channel formed by the lobe [44]. Of the 28-nt crRNA-spacer:target-RNA duplex that is formed, only the first 24 bp is encompassed by Cas13a, and the last 4 bp does not make any contact with Lbu-Cas13a [44].

Multiple contacts are also made by Lbu-Cas13a to the crRNA-spacer once duplexed with a target-RNA. These are mostly clustered within two specific regions of the crRNA-spacer. The largest cluster, encompassing nt 7–15 and the central region of the crRNA, interacts via its sugar-phosphate backbone with at least 12 residues from the Helical-2, HEPN1, and Linker domains [44]. The second much smaller cluster encompasses nt 18–20 and 23–24 and again interacts via sugar-phosphate backbone with residues from the Helical-2 and Linker domains [44]. Collectively, these clusters are likely to be involved in a combination of

stabilizing the bound duplex, and also detecting the presence of a stable A-form crRNA-spacer and transmitting this information to permit a conformational change to form an active HEPN-nuclease. This can be emphasized by the fact that the single-amino-acid mutations in several residues in these clusters results in reduced target-RNA cleavage [44]. Conversely, nt 2–6 of the crRNA-spacer do not make any direct contacts with LbuCas13a in the ternary complex, and this is recapitulated in biochemical experiments mismatches where multiple mismatches between the crRNA-spacer and target-RNA at spacer nucleotide positions 2–5 resulted in no defect in HEPN-nuclease activity [59] (see below for more details).

Looking across the length of target-RNA, contacts are made with the Helical-2, Linker, and HEPN1 domains via sugar-phosphate sequence-independent contacts. It appears that the highest density of contacts made by Cas13a occurs in at the 3′ distal end of the crRNA-spacer–target-RNA interaction (i.e., 5′ of nt 16–21′ of the ssRNA; see Fig. 3d), and again, the importance of base pairing in this distal region is also exhibited in biochemical assays where mismatches in the 18′ to 20′ region lead to large defects in HEPN-nuclease activity [59] . Beyond this large cluster of contacts, several other important contacts are made to the target-RNA by Cas13a. From mutational and biochemical studies, the most important of these is (1) R1335, which makes contacts with 2′ hydroxyl of the target-RNA nucleotide 13′ (within the “seed” region) and when mutated to alanine leads to a large defect in HEPN-nuclease activity, and (2) S786, which is shown to make contacts with the “switch region” nucleotides responsible for HEPN activation [59]. Finally, another interesting observation is that only 21 nt of the target-RNA are specifically contacted/recognized by Lbu-Cas13a [44]. This may explain why Lbu-Cas13a can use short crRNA-spacers (on order of 21 nt) without significant loss of HEPN-nuclease activity.

The Protospacer Flanking Sequence

It is well understood that there is a requirement for Cas interference complexes to detect the presence of a protospacer adjacent motif (PAM) sequence flanking a dsDNA target site to prevent interference complexes from cleaving their own CRISPR genomic loci (as PAM sequences are not found in the flanking repeat sequences) [60]. With this in mind, Abudayyeh *et al*. [29] asked whether Cas13a exhibits any equivalent sequence preference for nucleotides flanking each crRNA target site. They discovered that for Lsh-Cas13a, the first nucleotide flanking 3′ end of the protospacer exhibited a strict preference for A, U, or C (H) and not G (Fig. 3d). This result was confirmed through *in vitro* RNA cleavage experiment that showed that the presence of a G nucleotide in this position significantly reduced HEPN-nuclease activity, while A, U, and C led to maximal activity [29]. Intelligently, the authors decided to call this sequence preference a “Protospacer Flanking Site” (PFS) rather than a “Protospacer Adjacent Motif” (PAM), in order to avoid confusion with the defined role of a PAM in preventing “self” CRISPR genomic loci damage, given that target- *versus* self- discrimination should be dispensable when targeting RNA [29].

Since the initial report of a PFS preference for Lsh-Cas13a, PFS preferences have been also confirmed to exist for several other Cas13 enzymes. Using *in vitro* cleavage assays, Lwa-Cas13a was initially shown to possess a slight ‘H' PFS preference, albeit robust HEPN-nuclease activation was still observed with G-PFS-containing ssRNA targets in Cas13a-based RNA detection assays [33]. Later studies by the same authors indicated that when using a bacterial PFS screen [34] or a human cell line plasmid library screen [35], Lwa-Cas13 does not exhibit any detectable PFS preference and that targeting G-PFS containing ssRNAs in human cells does not result in any defect in targeting efficiency [34, 35]. This was also shown to occur with Lbu-Cas13a, where no detectable PFS preferences were observed using *in vitro* cleavage assays [30] or high-throughput mismatch target binding screens [59]. Thus, it is likely that different Cas13a homologs possess different sensitivities to a 3′ PFS and it might also depend the target sequence context and the sensitivity of the assay being used.

The PFS requirements for Cas13b enzymes are also similarly as varied. Initially, it was shown using both an *E. coli* essential gene and a Kanamycin gene library screen that for two different Cas13b's from *Bergeyella zoohelcum* and *Prevotella buccae* that both a D (A, U or G) 5′ PFS and a NAN or NNA 3′ PFS are required for optimal RNA targeting (Fig. 3d) [31]. This result was confirmed using *in vitro* RNA cleavage assays, which highlighted that a cytosine 5′ PFS inhibits HEPN-nuclease activity, while the presence of an adenosine in the second or third 3′ PFS position led to optimal HEPN-nuclease activity [31]. This additional requirement of a 5′ PFS is not all that surprising given the inverted Type VI-B crRNA-spacer-repeat orientation relative to Type VI-A crRNAs (Fig. 3d). More recent efforts used a bacterial screen to identify PFS preferences for an additional 13 Cas13b homologs. This experiment revealed that a mixture PFS preferences with some Cas13bs exhibits only a 5′ PFS preference, while others seemed to require both a 5′ and a 3′ PFS [35]. Conversely, in the same study when mismatched target library experiments are carried out for Cas13b from *Prevotella* sp. *P5–125* in a human cell line, almost all PFS combinations resulted in robust target cleavage, indicating that a preferred PFS is not required for efficient target-RNA cleavage in this assay [35]. In contrast to Cas13a or Cas13b, no PFS sequences were detected using a

bacterial screening strategy for any Cas13d orthologs studied to date [27, 28]. Currently, nothing is known about whether PFS preferences exist for Cas13c enzyme family.

It is worth highlighting that when PFS screens are done at the level of target-RNA binding (as opposed to HEPN-nuclease cleavage), no evidence for sequence restrictions at either the 5′ and 3′ flanking sequences is observed for both Lbu-Cas13a [59] or dPsp-Cas13b [35], suggesting that a disfavored PFS may inhibit Cas13a activity at the level of HEPN activation (conformation change) rather than affecting the stability of the crRNA:target-RNA interaction.

From a structural standpoint (at least for Lbu-Cas13a), the mechanism for a 5′ PFS preference is likely to be a result of aberrant base pairing between the disfavored PFS nucleotide (i.e., a guanine in target-RNA position –1 for Cas13a) and the crRNA DR (i.e., a conserved cytosine in crRNA position –1) [44]. This base-pairing interaction is likely to disrupt the tight

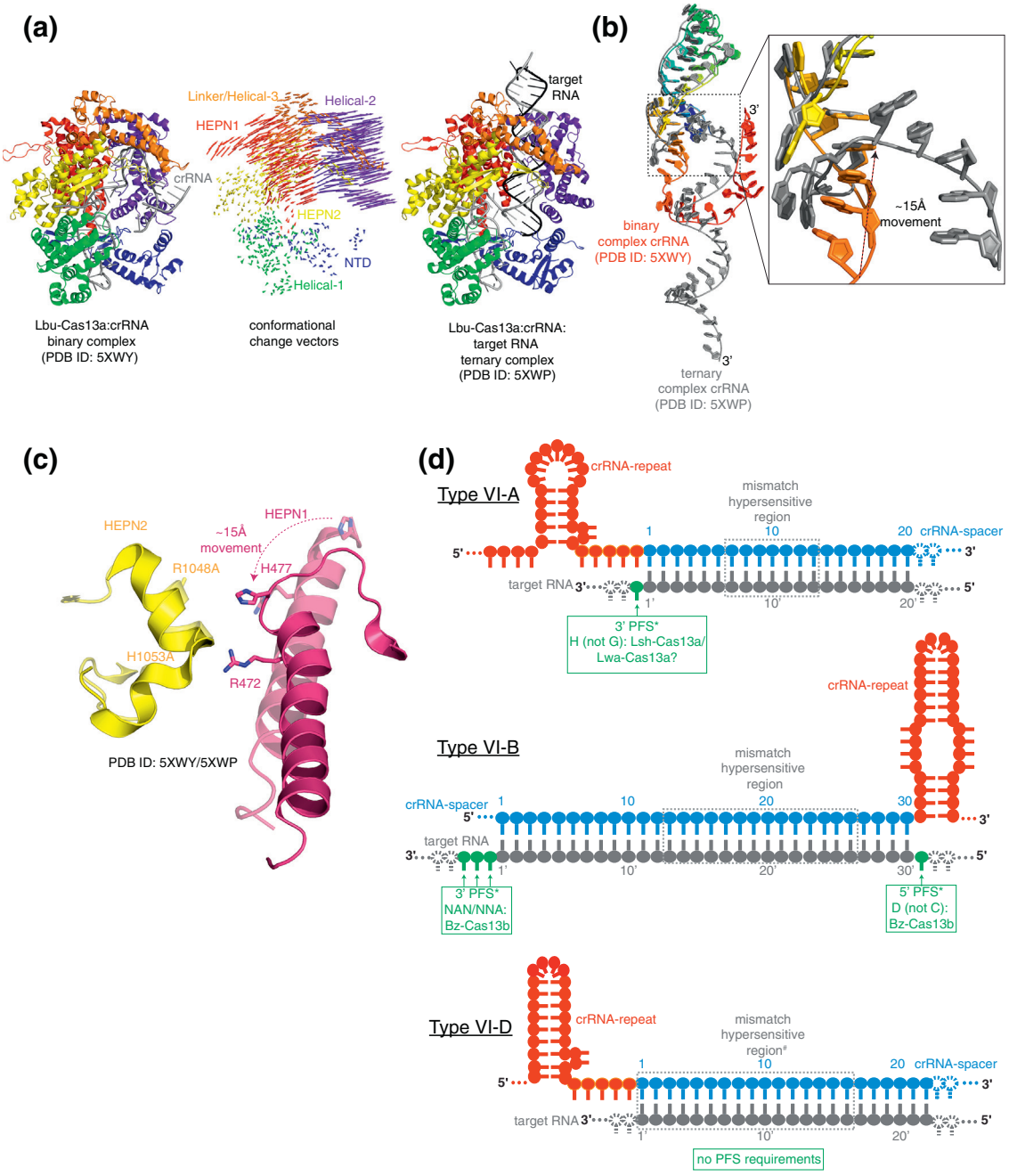

Fig. 3 (*legend on next page*)

network of H-bond interactions between crRNA C-1, amino acid residues from the Cas13a HEPN1 domain (especially S786), and multiple target-RNA phosphates [44], which together we have hypothesized to be likely required for efficient HEPN-nuclease activation [59]. Interestingly, the same conserved crRNA DR cytosine (C-1) is flipped and held away from the target-RNA in the Eu-Cas13d target-bound structure and thus unable to aberrantly participate in any target-RNA base pairing. This might explain why Cas13d does not appear to have a PFS requirement [46].

It is important to note that this is not the first instance where the "avoidance" of a particular nucleotide (rather than an affinity for a defined motif) has been observed for CRISPR–Cas systems. For example, the DNA targeting modality of Type III-A CRISPR–Cas systems prevents self-targeting through a mechanism that requires mismatches between the 5′ repeat-derived region of the crRNA and the target DNA strand for activation of the DNA nuclease, such that when the 5′ constant region from the crRNA is paired with target DNA the nuclease is inhibited [61–66]. And perhaps even more similar to Cas13a, RNA cleavage by the Cmr complex (Type III-B system) from *S. solfataricus* requires mismatches between the 3′ end of the target-RNA and the 5′ repeat-derived constant region [66]. However, much like Cas13, this discrimination may be more of a structural "quirk" of the enzyme rather than necessary for self- *versus* non-self-recognition.

Taken together, while it is clear that some sequence restrictions likely exist for sequences flanking the ssRNA protospacer targets for Cas13, the ability to detect these preferences varies widely depending on which assay is used to detect them, and at this stage, it is not clear whether the PFS has any physiological role. Further work is needed to understand if and how PFS preferences are able to modulate interference efficacy within Type-VI system hosts.

HEPN domain conservation, conformation change, and collateral cleavage

When the first amino acid sequences for Cas13 were published, the HEPN domains were only conserved domains in Cas13a that could be identified by primary sequence analysis, and were thus important in the initial hypothesis that Cas13 was an RNA-targeting enzyme [20]. Approximately 40 distinct families of HEPN domains exist based on sequence conservation, are typically 100–120 amino acids long, and adopt 4-helical "up-down" fold [47]. These domains are primarily characterized by the presence of a conserved amino acid motif often positioned at the end of helix-3 and in the loop connecting helix-3 and helix-4, with a consensus sequence of R-$X_{4–6}$-H where the amino acid after R is typically polar (often N, D, or H). Many HEPN domains also contain an additional conserved acidic residue, typically as part of an E-X_3-[KR] motif [47]. It has been shown for several HEPN domain-containing ribonucleases that the R-$X_{4–6}$-H motif within HEPN domain is required for a metal-independent RNase catalytic activity that generates a product with a 3′ terminal 2′–3′ cyclic phosphate [47, 67–73] . It has been hypothesized that the conserved histidine within this R-$X_{4–6}$-H motif acts as a base to activate a 2′OH to carry out a nucleophilic attack of the RNA phosphodiester backbone, while the conserved arginine could either act to stabilize the pentavalent transition state intermediate or position the backbone of the RNA substrate [47].

HEPN domains have also been found in several other CRISPR–Cas proteins including Csx1 [72, 73] and Csm6 [71] (which both reside in type Type-III systems), and more recently, a putative divergent HEPN was hypothesized to exist in Csx28 [26, 31]. Interestingly, both Csx1 and Csm6 do not interact directly (or form part of) with the interference complexes

Fig. 3. Conformation change in Cas13a and its bound crRNA upon target-RNA recognition. (a) Crystal structures of Lbu-Cas13a:crRNA complex (left) and Lbu-Cas13a:crRNA:target-RNA complex (right), with each domain colored as indicated. In the center is a vector representation of the Cas13a conformational change that occurs upon target-RNA binding. Vector length represents the magnitude of the conformational change between to the two states. (b) An overlay of the crystals structures of the Lbu-crRNA when bound by Cas13a alone ("binary complex crRNA," colored from blue to red going 5′ to 3′) or in complex with a target-RNA ("ternary complex crRNA," colored gray; target-RNA is omitted for clarity). Inset: a zoomed region of the 3′ DR flank region highlighting the large conformational change that occurs upon target-RNA binding. (c) A close-up overlay of the HEPN-nuclease active site from Lbu-Cas13a highlighting the conformation change that occurs in HEPN1 after target-RNA binding leading to the activation of HEPN-nuclease activity. HEPN1 is shown in maroon, HEPN2 is shown in yellow. Active site residues in HEPN2 in both of these crystal structures have been mutated to alanine. (d) A schematic of the crRNA:target-RNA interaction highlighting important elements required for stable target-RNA binding and/or HEPN-nuclease activation. PFS, Protospacer Flanking Sequence. The PFS is indicated for orthologs where a PFS has been shown to regulate HEPN-nuclease cleavage activity. The asterisk "*" denotes that the requirement for a PFS is not always observed, as appears to be dependent on the Cas13 ortholog studied and assay used to measure its effect. "Mismatch hypersensitive regions" are outlined with a gray dotted box and indicated the region where mismatches between the crRNA-spacer and the target-RNA result in the largest defect in target-RNA binding and/or HEPN-nuclease activation for a number of homologs within each subtype. [#]Cas13d mismatch-sensitivity experiments have only been carried out for Es-Cas13d (see Ref. [46]).

from their respective CRISPR–Cas systems but rather act as auxiliary endoribonucleases that act to amplify the nuclease activities of the interference complex [71–76]. A striking commonality exists across most if not all HEPN-domain proteins (i.e., Csm6, Csx1, and eukaryotic ribonucleases RNase L and Ire1) experimentally studied to date: the HEPN domains all appear to require dimerization to be active nucleases [47, 67–73]. Upon HEPN dimerization and in some cases additional ligand-driven conformational change, it has been proposed that the two R-$X_{4–6}$-H motifs become closely juxtaposed leading to the formation of a composite active site that is able to bind substrate RNA asymmetrically [68, 70, 71].

Thus, it was very exciting to see that when the first Cas13 sequences were released and analyzed [20], the only conserved domains within these proteins that could be identified were not one but two HEPN domains. Indeed, shortly after this initial discovery, several biochemical studies of Cas13a [29, 30] and Cas13b [31] were able to show that both HEPN domains were required for RNA cleavage, as single-point mutations in either HEPN domain completely abrogated RNA cleavage [29–31]. These biochemical analyses also revealed several other interesting mechanistic features regarding the HEPN-nuclease activity. Perhaps most fundamentally, it was shown that upon recognition of a complementary target-RNA through base-pairing interactions with the crRNA, the HEPN domain was able to cleave ssRNA but not dsRNA, ssDNA, or dsDNA [29–31], and when folded structures are present in an RNA target, Cas13 preferentially cuts at non-base-paired ssRNA regions [29]. Furthermore, it has been shown that Lsh-Cas13a [29], Lbu-Cas13a [30], and a number of closely-related homologs prefer to cleave mostly after uridines [32], while a distinct clade exhibits a preference for cleavage after adenosines [32]. Similarly, the first Cas13b protein tested (Bz) was also shown to have a substrate preference for uracil and to lesser extent cytosines [31].

More recently, library cleavage screens were used to further refine our understanding of cleavage preferences for several Cas13a and Cas13b homologs [36]. From di-nucleotide screens, it was found that several Cas13 homologs displayed orthogonal substrate preferences to the extent that Lwa-Cas13a, Cca-Cas13b, Lba-Cas13a, and Psm-Cas13b only exhibited activity on only one of the following dinucleotides AU, UC, AC, and GA, respectively [36]. This feature is useful for RNA detection applications (discussed in more detail below). These orthogonal preferences could be extended to RNA hexamer substrates for some Cas13 homologs [36], suggesting that the HEPN domains may exhibit specificity for nucleotide identity beyond the active site cleft. Recently, Es-Cas13d was shown to have a preference for cleaving after uracils, displaying low but detectable activity on other nucleotide substrates [28]. To date, nothing is known about Cas13c HEPN-nuclease preferences. What was also evident from these HEPN-nuclease studies is that a wide range of HEPN-nuclease activities are observed (several orders of magnitude) across the Cas13 family [29–32, 34, 35], but it is currently unclear precisely what differences at the protein or crRNA sequence level are responsible for this large variation in nuclease activity.

As already noted, HEPN-nucleases have been generally thought not to require the presence of metal ions for catalytic activity; thus, it was a surprise when it was first shown that Cas13's HEPN-nuclease activity strictly requires the presence of a divalent metal ion [29, 30], and in most cases the preference is Mg^{2+}, with some Cas13 homologs displaying an additional preference/flexibility for Ca^{2+}, Ni^{2+}, or Mn^{2+} [36]. This suggests that either the HEPN domains within Cas13 are divergent and have evolved a catalytic requirement for metal ions or that the presence of metal ions helps stabilize substrate binding. The latter hypothesis is significantly more attractive because the products of the Cas13 HEPN-nuclease cleavage (5′-hydroxyl and 2′–3′ cyclic phosphate) [36] are only known to be generated through a metal-independent mechanism (for review, see Ref. [77]). Interestingly, as discussed above, an Mg^{2+} ion was found coordinated by both Cas13d and its crRNA in a recent cryo-EM structure, and while this appears to be unique to Cas13d, at least in the structures solved so far, it likely contributes to the ability for Cas13d to target-RNA efficiently [46]. Nevertheless, the role of metal ions in promoting Cas13 HEPN-nuclease cleavage needs further study.

The most striking observation from the early biochemical studies was the pattern of the cleavage products generated by Cas13 in the *in vitro* cleavage experiments. Cas13 rapidly generates multiple RNA cleavage products, many of which were outside the crRNA-spacer:target-RNA duplex (in contrast to Cas9). In addition, an accumulation of increasingly smaller cleavage products over the course of the cleavage reaction and concomitant decrease in larger products was observed [27–31], suggesting that even when these larger products are released from the bound target-RNA, they continue to be cleaved by the enzyme. The combination of this observation, and the observation that targeting RFP mRNA in *E. coli* with Lsh-Cas13 programmed with an RFP-targeting crRNA leads to a sizeable growth defect [29], prompted a closer look at the HEPN-nuclease behavior of Cas13.

To study this phenomenon in more detail, *in vitro* RNA cleavage experiments were designed where in addition to Cas13:crRNA and a complementary target-RNA, non-complementary RNAs were added into the reaction to see whether they could also be cleaved [29, 30]. It was very clear that once Cas13:crRNA binds to a complementary target-RNA (a.k.a. an “activator-RNA”), the HEPN-nuclease becomes active and is able to cleave not only the target-RNA in *cis* but also any other RNA present in *trans*, including non-complementary RNAs, as well as unbound complementary target-RNAs or target-RNAs bound

to other Cas13:crRNA molecules (Fig. 1c) [29, 30]. Mutation to either HEPN domain active site residues abrogated this activity confirming that the HEPN-nuclease was responsible for both "cis" and "trans"-RNA cleavage [29, 30]. This activity has been collectively referred to as "collateral cleavage." Collateral cleavage is also a feature of Cas13b [31] and Cas13d [27, 28]. Although Cas13c HEPN-cleavage activity has not been biochemically interrogated yet, it is expected also possess this activity, given its similarity to Cas13a.

Much like the relatively indiscriminate cleavage of viral and host RNAs by toxin–antitoxin systems, Type III HEPN-containing CRISPR proteins Csm6 and Csx1 described above, as well as the eukaryotic RNase-L system, it has been hypothesized that this collateral activity exhibited by TypeVI CRISPR–Cas systems has evolved as a robust mechanism to induce a short dormancy-like state in cells to allow for sufficient time for adaptive immunity to be activated, and/or additionally induce suicide or permanent dormancy to prevent or reduce phage burst size in lieu of successful immunity [29, 30, 47, 78]. More work is required to understand the role/function of Cas13's "collateral cleavage" in the broader context of immunity.

Fascinatingly, so far this "collateral cleavage" effect has not been observed when Cas13 has been used to target-RNA in either human or plant cell lines [34, 35, 37], suggesting that there is something fundamentally different relative to bacteria regarding either the behavior of Cas13 within the context of a human or plant cell, the cellular environment (RNA concentration, availability and/or protection of RNA by secondary structure or bound proteins) or the regulatory response to initial '*cis*' target-RNA cleavage by the cell. Given the ability for RNase-L (which also contains a type of HEPN domain) to readily induce significant collateral RNA degradation within human cells [79–81], the most favorable hypothesis is the latter one, that is, the cell is able to rapidly inactivate Cas13 through its response to the generation of a cleaved "*cis*" target-RNA. This hypothesis also raises the possibility that collateral cleavage may not be a predominate mechanism in some bacterial hosts, given the diversity in host RNA degradation mechanisms across bacterial phyla (for review, see Ref. [82]). More work needs to be done to understand rates of *cis*- *versus trans*-cleavage in highly purified systems, as well as in the cellular milieu to further understand this interesting phenomenon.

In all Cas13a structures solved to date, the two HEPN domains are expectedly found close in three-dimensional space, with their active-site residues solvent exposed on the external surface of protein facing away the crRNA:target duplex (Figs. 1b, c and 2a) [43–45]. In these structures, the HEPN1 domain is interrupted by the insertion of the helical-2 domain, and thus, HEPN1 can be considered as HEPN1I-II (Fig. 2a). This split nuclease architecture is reminiscent of the split RuvC architecture in Type II interference proteins Cas9 [83] and Cpf1/Cas12 [84], and in both cases, this architecture possibly evolved as a means to conformationally control catalytic activity. HEPN1-II is then followed by the Linker [43, 44] or Helical-3 domains [45], which connect HEPN1-II to the second HEPN domain (HEPN2). The Linker/Helical-3 domain varies in length between the different homologs: 271 amino acids in Lsh-Cas13a, compared to 129 and 133 in Lba- and Lbu-Cas13a, respectively. It is currently unclear what the exact role of the expanded Lsh-Cas13a Linker/Helical-3 domain is, but it may be hypothesized that it helps accommodate the longer crRNA-spacer required for optimal Lsh-Cas13a activity.

Comparing the Lbu-Cas13a binary cryo-EM structure to the ternary crystal structure reveals that the HEPN1 domain rotates inward moving closer to the crRNA-spacer:target-RNA duplex, which results in approximately 15-Å movement of H477 of HEPN1 toward the corresponding catalytic residues R1048 and H1053 within the HEPN2 domain (Fig. 3c) [44]. In addition, a loop from HEPN2 (residues ~990–1007) that helps to shield H477 from the HEPN1 active site half from solvent moves to allow H477 to gain access to the catalytic site cleft [44]. Similarly, in the Lba-Cas13a:crRNA binary structure, the conformation of the second alpha helix of HEPN1 is distorted such that the histidine (H605) from of the $RX_{4-6}H$ motif is buried away from the putative active site cleft, obscured beneath a loop of HEPN2 [45]. Given that this residue is required for HEPN-nuclease catalytic activity [45], the second helix from HEPN1 must also reorient upon activator-RNA binding to reveal H605 to the active site. The same HEPN1 histidine (H602) in the Lsh-Cas13a crystal structure is also concealed from the active site cleft by a loop from the HEPN2 [43] and a similar mode of activation is seen in Cas13d [46], suggesting that this auto-inhibition by the HEPN2 loop might be a conserved feature of HEPN-nuclease activation in Cas13 proteins.

Understanding Cas13's target-RNA binding and cleavage specificity

Our first understanding of the specificity requirements of Cas13a was revealed exploring the effect that mismatches between the crRNA-spacer and the ssRNA-activator had on the ability of Lsh-Cas13a to cleave target-RNA [29]. Using MS2 phage infection assays, it was found that single mismatches between a targeting crRNA-spacer and the target MS2 RNA genome in most cases did not affect Lsh-Cas13a's ability to inhibit phage infection [29]. Interestingly, mismatches at positions 5 and 17 of the crRNA-spacer appeared to cause the largest effect, with close to a 10-fold decrease in phage resistance [29]. When

consecutive tandem mismatches were introduced into the crRNA-spacer, the interference defect was much more pronounced, with the largest effect occurring with double mismatches toward the center of the spacer (~1000-fold reduction in phage resistance) with smaller to no defects toward each end of the crRNA-spacer [29]. This observation was also confirmed using *in vitro* RNA cleavage assays with crRNA-spacers containing consecutive double and triple mismatches [29] (Fig. 3d).

More recently, the RNA-cleavage specificities of another Cas13a homolog (Lwa-Cas13a) and a Cas13b homolog (Psp-Cas13b) were tested in human cell lines using either mismatched crRNA-spacers to a number of transcripts [34] and/or a plasmid library that expresses RNA-targets containing single and double mismatches across the crRNA-spacer complementary region [34, 35]. It was found that both of these proteins also exhibited a mismatch-sensitive "seed" in the central portion of the crRNA-spacer. For Lwa-Cas13a, both studies generally agreed that the central region of the crRNA-spacer exhibits the greatest sensitivity to mismatches [34, 35]; however, the magnitude and precise location of maximal sensitivity differ between assays and RNA targets, suggesting that additional parameters such as spacer/target sequence composition probably play a role in modulating specificity. In addition, while a 21/24-nt crRNA-spacer (depending on the *in vitro* assay) was sufficient for maximal activity for *in vitro* RNA cleavage, in cells shortening the crRNA-spacer below 28-nt reduces the on-target-RNA cleavage efficiency by a factor of two, suggesting that additional parameters (e.g., crRNA-loading) may dictate the crRNA-spacer length requirement in human cells [34, 35]. It is worth noting that an understanding of Lwa-Cas13a mismatch sensitivities (at least in the crRNA-repeat proximal region) has been previously utilized for SNP detection [33]. For Psp-Cas13b, plasmid mismatch library experiments showed that "mismatch-sensitive" region roughly extends from spacer nt 12–26 (for a 30-nt spacer), suggesting also that while the very central portion of the crRNA-spacer (positions ~14–18) possesses the largest sensitivity to mismatches, the region proximal to the crRNA-repeat (positions 18–26) is more sensitive to mismatches than the crRNA-repeat distal region (positions 1–12) (Fig. 3d) [35]. These results may suggest a different mode of crRNA-spacer organization in Cas13b complexes and further work is required to fully understand Cas13b's RNA-targeting specificity. During revision of this review, Es-Cas13d's *in vitro* mismatch tolerance was also reported, and it was found that the mismatch-sensitive region extends all the way from crRNA-spacer positions 1–16, which fits well with the how the crRNA is organized within the complex (see above). It will be exciting to see whether this holds for multiple Cas13d homologs [46].

Cas13 HEPN-cleavage specificity within the context of human RNA knockdown experiments has also been analyzed. In an experiment targeting a *Gluc* reporter transcript in human cells, both Lwa-Cas13a and Psp-Cas13b exhibit high efficiency (an average of 40.1% and 92.3% knockdown, respectively) and remarkable specificity with negligible significant off-target-RNA knockdown events detected (zero and one off-target detected for Lwa-Cas13a in [34, 35], respectively). This level of specificity is striking, particularly when compared to a guide-matched short-hairpin RNA (shRNA) knockdown, which exhibited hundreds of significant off-targets [34, 35]. Lwa-Cas13a RNA knockdown specificity was also tested for endogenous *KRAS* and *PPIB* transcripts, and although knock down efficiencies were low (27.1% and 29.2% knockdown, respectively), in both cases no significant off-target was detected. Notably, a similarly impressive level of efficiency and specificity was also observed for Cas13d when used to knockdown *B4GALNT and ANXA4* RNA transcripts in HEK293T cells [28].

Given the interest in Cas13 as not only a tool for specific RNA-cleavage but also a tool specific RNA binding, there is a need to understand what the minimal base-pairing requirements (in terms of total number and position of base pairs) for stable RNA association. With this question in mind, Tambe *et al.* [59] used a combination of *in vitro* high-throughput sequencing-based and traditional biochemical assays to study the relationship between stable crRNA-mediated RNA-target binding and HEPN domain nuclease activation. It was found that Cas13a's crRNA-guided RNA binding affinity and HEPN-nuclease activity are differentially affected by the number and position of mismatches between the crRNA and activator-RNA. In particular, mismatches in some regions are able maximally active HEPN-nuclease activation despite lower binding affinities, while mismatches in other regions of the crRNA-spacer promote tight binding but are unable to activate the HEPN-nuclease [59]. This observation highlights not only the presence of an interesting plasticity in targeting RNA that may be integral to maintaining an adaptive response to rapidly mutating phage sequences, but also that care needs to be taken when thinking about optimal target site design and the range of likely off-targets when designing crRNAs for RNA-binding *versus* RNA-cleavage applications.

Type VI accessory proteins modulate Cas13 nuclease activity

A number of CRISPR–Cas systems contain ORFs within their genomic loci that do not appear to be required for spacer acquisition, pre-crRNA processing or nucleic-acid interference [17, 85–87]. This includes Type VI systems, which contain an array of different 'accessory' proteins within close vicinity of their CRISPR loci. This notably includes the recently designated Csx27 and Csx28 Cas proteins, which

coincide with two distinct branches of the phylogenetic Cas13b tree [20, 31]. These two variant Type VI systems are currently categorized as VI-B1 and VI-B2 systems, respectively [20, 31]. Almost all VI-B2 systems contain a Csx28 ORF, while Csx27 is only present in approximately 50% of VI-B1 systems. Both of this ORFs display no identifiable protein sequence similarity to other Cas proteins and are not found in any other CRISPR–Cas systems [20, 31]. The only features that can be predicted for both proteins are the presence of one (Csx28) or multiple (Csx27) transmembrane-spanning regions (TMs) [31], which could be indicative of an inner membrane location in Gram-negative bacteria (in which all type VI-B systems reside). Smargon *et al*. tested this prediction by heterologously expressing N- or C- terminus RFP-tagged Csx27 from *B. zoohelcum* and Csx28 from *P. buccae* in *E. coli*, but found that neither of these constructs exhibited distinct membrane localization when observed under a fluorescence microscope. Given, the demonstrated predictive power of bacterial subcellular localization algorithms (for review, see Ref. [88]), further experimental validation is required to completely rule out inner membrane localization, particularly for Csx27. In addition to these transmembrane segments, Csx28 also possesses a predicted divergent HEPN domain at its C-terminus [26, 31].

To understand the role of Csx27 and Csx28 in modulating Cas13b interference activity, Smargon *et al*. [31] co-expressed either Csx27 or Csx28 and Cas13b in the presence of either a MS2 targeting or non-targeting crRNA, and tested their effect on Cas13b's interference ability in a MS2 drop plaque assay. They found that while Csx27 or Csx28 did not influence interference alone, when co-expressed with Cas13b and a targeting crRNA, Csx27 was able to inhibit Cas13b's interference activity by at least 2 (and up to 5) orders of magnitude, while conversely, Csx28 enhanced interference by up to 4 orders of magnitude [31]. Interestingly, Bz-Csx27 was able to inhibit both its associated Cas13b and Pb-Cas13b (which contains an associated Csx28 but not a Csx27 protein) [31], suggesting that Csx27 has broad inhibitory properties that can function across multiple Type VI-B systems.

In terms of mechanism, a suitable hypothesis is that Csx28 dimerizes to bring its two HEPN domains together to form a composite RNase active site, much like what is observed with Csm6, Csx1, and all other HEPN domain proteins studied to date [47, 71–73, 76] . Once dimerized and/or activated, Csx28's active HEPN domains may also act as a non-specific RNA nuclease significantly boosting the efficacy of the Cas13b driven anti-viral response. For Csx27, it is less clear how it might be inhibiting Cas13b at the molecular level, but given the evidence that despite the presence of multiple TMs, Csx27 localizes to the cytoplasm, it is plausible that it harnesses these hydrophobic helices to tightly associate with Cas13b and inhibit its activity, similar to what is observed in some toxin-anti-toxin systems [89]. Alternatively, it is also equally possible that in its host organism, Csx27 indeed localizes to the inner membrane and acts to tether Cas13b to the cytoplasmic side of the inner membrane. Upon infection, Csx27 responds and releases its inhibitory hold on Cas13b in preparation for interference. Given the widespread presence of putative TMs within Cas accessory proteins, the use of membrane association by CRISPR–Cas systems to improve the fight against mobile genetic elements seems plausible [90].

Very recently, a new family of Cas accessory proteins was discovered to reside within Type VI-D system loci [27]. These proteins were all shown to all contain one or multiple WYL domains [27]. WYL domains are approximately 170 amino acids long, are very frequently found at the C-terminal of a helix-turn-helix (HTH) or ribbon–helix–helix (RHH) DNA-binding domains, and are often associated within prokaryotic defense systems [42]. WYL-domain-containing proteins have been shown to act as crRNA transcriptional repressors within a *Synechococcus* Type I-D CRISPR–Cas system [91], as well as a ssDNA binding domain that modulates ATPase/helicase activity of the non-CRISPR-related Pif1 helicase [92]. In contrast to these activities, RspWYL1 (WYL1 from *Ruminococcus* sp.) type VI-D CRISPR–Cas system was shown to stimulate Cas13d-driven collateral RNA cleavage in both bacterial negative selection screens and in *in vitro* RNA cleavage assays, but exhibited no RNA cleavage activity in the absence of Cas13d:crRNA and a complementary target-RNA [27].

Further studies are required to understand how Cas accessory proteins such as WYL1 and Csx28 are able to enhance total RNA cleavage rates when target-RNA-activated Cas13 is also present. Finally, it is interesting to note that WYL-domain containing proteins also appear to be associated with other Class II CRISPR systems, particularly Type II, V and VI-A and B systems, suggesting that this domain has been widely adopted by CRISPR–Cas systems to regulate their activity [26]. That said, Koonin and colleagues have wisely suggested to exercise caution when wondering whether to consider these genes as *bona fide* components of CRISPR–Cas loci given that many CRISPR–Cas loci belong exist in “islands” that contain a range of other highly-mobile genes such as toxins–antitoxins, transposases, and components of other defense systems [17].

Biotechnological and therapeutic applications of Cas13 RNA targeting

The ability to direct a protein to a specific RNA sequence in a programmable manner to degrade the RNA or help to localize tethered effector domains (e.g., GFP or splicing factor domains) would enable a wide variety of ways to specifically regulate RNAs to investigate or manipulate any aspect of their function

[93]. In 2014, it was first shown that that CRISPR–Cas9 in *Streptococcus pyogenes*, the revolutionary genome editing tool also binds and cleaves RNA in a programmable manner [94]. Building on this discovery, RNA-targeting Cas9 has already been used to image RNA transport [95], alter protein expression [96], and eliminate aberrant RNAs in myotonic dystrophy patient cells [97] and in a mouse model [98]. Recently, several studies have shown that other CRISPR–Cas9 orthologs or CRISPR-associate argonaute proteins may be even better suited to RNA targeting applications [99–102]. However, even with special sgRNA design rules that promote RNA *versus* DNA recognition [94], one concern with using Cas9 to target-RNAs is that they also retain the ability to interact with DNA, potentially leading to unintended off-target DNA binding and/or cleavage. With this in mind, the discovery of Cas13, which only targets RNA has led to a new era of excitement around the development of RNA-targeting tools for research and therapeutics.

The first set of applications developed has involved the use of the "collateral cleavage" activity of Cas13's HEPN-nuclease for ultrasensitive RNA detection [30, 33, 36]. This system relies on the use of a short quenched-fluorescent reporter RNA that when cleaved by an RNA-target-activated Cas13 molecule leads to a large increase in fluorescence, and due to the robust multiple turnover nature of Cas13 enables thousands of reporter cleavage events per single target-RNA binding event, acting to substantial amplify the initial "detection" event [30]. This initial demonstration of Cas13-based sensitive RNA detection has since been extended to incorporate pre-amplification of the input nucleic acid sample substantially increasing the sensitivity of the approach to the point where it outperforms current state-of-the-art RNA detection technology, achieving attomolar detection with single-nucleotide specificity, all while being extremely cost effective [33]. Furthermore, the authors were able to demonstrate that this methodology (named SHERLOCK; Specific High-Sensitivity Enzymatic Reporter UnLOCKing) can be carried out using freeze dried reagents deposited on paper strips potentially enabling rapid, low cost point-of-care diagnostics for the detection of a range of maladies from viral infection to cancer diagnosis [33]. More recently, further developments to the SHERLOCK platform have harnessed the orthogonal crRNA requirements and HEPN-nuclease cleavage preferences of Cas13a and Cas13b to develop multiplexed detection, the use of accessory proteins (e.g., Csm6) to boost the output signal, as well as alternative detection modalities compatible with colorimetric lateral flow detection devices such that RNA sequences can be detected with very high sensitivity without the requirement of a fluorimeter [36, 103].

Beyond RNA detection, there has been a number of newer applications developed focusing on the use of Cas13 as tool to observe and manipulate RNA biology in live cells. Cas13a, Cas13b, and most recently Cas13d have all been shown be proficient for targeted RNA knock down in human and plant cells [28, 34, 35, 37]. All Cas13s tested for human RNA knockdown to date have demonstrated remarkably high specificity, exhibiting negligible and significantly less off-targets compared to matched RNAi (shRNA) target sequences [28, 34, 35]. These studies have also effectively harnessed the ability of Cas13's crRNA-processing activity to enable the knockdown of several transcripts in parallel from a single pre-crRNA transcript [28, 34, 35].

Discussed in further detail earlier in this review, it was fascinating to observe a lack of collateral activity in in all of these studies, suggesting a fundamental difference in the behavior of Cas13 and/or cellular response to initial Cas13 activity in eukaryotic cells. It is important to note that there is a disagreement between which Cas13 subtype is the most efficacious for RNA-targeting activities: one study comparing Cas13b to Cas13a demonstrated that Cas13b was significantly more active for RNA knockdown across a large range of sequence contexts [35]; while on the other hand, a more recent study found that Cas13d was more active than both Cas13a and Cas13b, with Cas13b exhibiting the poorest activity [28]. Future studies will need to reconcile these differences and aim to understand more completely why some Cas13 subtypes are able to more efficiently target a larger RNA sequence space with the human transcriptome. In combination with these studies, future biochemical and structural studies are also required to help us to garner a better understanding of the differences in crRNA-loading and organization, target-RNA binding, and cleavage specificity, as well as intrinsic differences in HEPN nuclease activity, between Cas13 subtypes and homologs. Together, these studies will allow us to understand how these properties may influence Cas13 activity with respect to applications development.

Mutation of the HEPN-nuclease domain results in a Cas13 variant (dCas13) that can specifically bind RNA without cleaving it [29–31]. This has resulted in the development of several additional exciting applications including tracking of bulk movement of RNA transcripts to the cytoplasm and stress granules [34], the use of Cas13b–ADAR2 fusion proteins to enact site-specific adenosine-inosine editing to recode cellular RNAs [35], and most recently the use of Cas13d (both with and without fusion to a hnRNP A1 glycine-rich domain) to promote exon exclusion and effect alternative splicing outcomes [28]. There is no doubt that improvements on these demonstrations will continue to occur, as well as the development of new applications directly observing and/or manipulating various aspects of post-transcriptional regulation (e.g., splicing, polyadenylation, RNA transport, RNA translation) both with allele/SNP-specific resolution and with transcriptome-wide scalability through the use of large crRNA library and orthogonal Cas13 proteins. Given the breakneck speed at which such applications are being developed,

Cas13 has already proven itself to be a more widely applicable platform for RNA-specific applications than its predecessors ushering in a new era of possibilities to observe and manipulate RNA biology for research and therapeutic applications.

Perspectives

Recent bioinformatics approaches have discovered a startling array of new CRISPR–Cas systems, further expanding the known CRISPR "universe" to include Type V and VI systems and a number subtypes within [20, 26–28, 31]. This has expanded not only the "CRISPR–Cas toolbox" for new and exciting molecular biology tools for gene editing and RNA manipulation, but also our understanding of the many ways prokaryotes can protect themselves from mobile nucleic acid invaders. Beyond cataloging new CRISPR–Cas systems, a number of recent studies have highlighted that the mechanisms of CRISPR–Cas-driven immunity are more sophisticated and nuanced than previously appreciated. For example, recent studies highlighted amplification of the antiviral response by Cas10 "second messenger" generation to control Csm6 activity [74, 75] as well as the role of effector proteins such as Csx27/Csx28 [31] and WYL proteins [27] in regulating Cas13 function. Together these studies suggest that prokaryotes have evolved to ways to tune their antiviral response, likely in order to remain permissive to the horizontal gene transfer of potentially beneficial traits.

The studies detailed in this review have rapidly revealed a large amount about how a number of Cas13 subtypes are able to bind and cleave ssRNA in a crRNA-dependent mechanism, and how this behavior can be utilized for RNA diagnostics and RNA manipulation in a range of organisms. Despite these advances, a number of interesting fundamental questions and challenges surrounding Type VI system function still remain. Foremost, a number of questions surround the role and scope of "collateral cleavage" within hosts that harbor Type-VI systems. While it is clear that this non-specific RNase activity is both directly observed biochemically and indirectly observed in the case of *E. coli* heterologous expression experiments, it is not clear how rampant this collateral activity is across the diversity of different bacterial hosts, particularly given the lack of collateral activity observed in all eukaryotes tested to date, and the known differences in RNA degradation mechanisms (and thus response to Cas13 cleavage) not only between eukaryotes and prokaryotes and but also between different bacterial phyla. In the case where collateral cleavage (*versus* specific phage RNA destruction) is a predominate mechanism by which Type VI-containing hosts clear viral infection, it will be also interesting to see whether all RNAs are equally targeted non-specifically to induce global shutdown of cellular metabolism, or there is a somewhat more nuanced response where particularly susceptible RNAs are cleaved initially and result in induction of dormancy/programmed cell death, before the entire RNA pool is depleted (e.g., as is seen with RNase L [79, 81]).

In terms of the mechanistic details of Cas13 function, there are a number of things we do not understand; this includes but is not limited to (a) understanding how the HEPN domains are activated with respect to the target search process, guide organization, and conformational dynamics; (b) how the active site is organized, that is, which active-site residues from HEPN1 *versus* HEPN2 are directly involved in catalysis; (c) the molecular interactions that determine substrate cleavage specificity; and (d) how at the molecular level are accessory proteins such as Csx27, Csx28, and WYL1 able to modulate Cas13 activity, direct or indirectly. A greater understanding of these processes may lead to a rationale for engineering Cas13 proteins with altered HEPN-nuclease substrate preference, greater specificity, and new approaches to further activate or inhibit Cas13 activity.

Finally, it also going to be exciting to see a range of new Cas13-based applications developed. I anticipate we will soon see Cas13 applied in a number of exciting new ways, including but not limited to transcriptome-wide library screens to down-regulated gene expression at the RNA level potentially with more specificity than current RNAi/CRISPRi approaches, modulation of various aspects of RNA metabolism (splicing, transport, polyadenylation, translation, degradation), as well as the observation of these processes with allele specificity using fluorescently-tagged Cas13 proteins.

Acknowledgments

I would like to thank J. Scott Butler, Gavin Knott, and Sam Sternberg for helpful discussions.

Received 9 April 2018;
Received in revised form 8 June 2018;
Accepted 12 June 2018
Available online 22 June 2018

Keywords:
CRISPR–Cas immunity;
Type VI CRISPR–Cas systems;
Cas13;
C2c2;
programmable RNA targeting

Abbreviations used:
CRISPR, clustered regularly interspaced short palindromic repeat; Cas, CRISPR-associated; ssRNA, single-stranded RNA; crRNA, CRISPR RNA; pre-crRNA, precursor crRNA; DR, crRNA direct repeat; PAM, protospacer adjacent motif; PFS, protospacer flanking sequence; nt, nucleotide; bp, base

pair; HEPN, Higher Eukaryotes and Prokaryotes Nucleotide-binding domain; TM, transmembrane-spanning region; WYL, WYL-containing domain; Lba, *Lachnospiraceae bacterium*; Lbu, *Leptotrichia buccalis*; Lsh, *Leptotrichia shahii*; Lwa, *Leptotrichia wadei*; Rsp, *Ruminococcus* species; Pb, *Prevotella buccae*; Psp, *Prevotella* sp. P5–125; Bz, *Bergeyella zoohelcum*; Ur, *Uncultured Ruminoccocus* species; Es, *Eubacterium siraeum*.

References

[1] P. Mohanraju, K.S. Makarova, B. Zetsche, F. Zhang, E.V. Koonin, J. van der Oost, Diverse evolutionary roots and mechanistic variations of the CRISPR–Cas systems, Science 353 (2016), aad5147.

[2] A.V. Wright, J.K. Nunez, J.A. Doudna, Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering, Cell 164 (2016) 29–44.

[3] P.D. Hsu, E.S. Lander, F. Zhang, Development and applications of CRISPR–Cas9 for genome engineering, Cell 157 (2014) 1262–1278.

[4] J.A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR–Cas9, Science 346 (2014).

[5] J. van der Oost, E.R. Westra, R.N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR–Cas systems, Nat. Rev. Microbiol. 12 (2014) 479–492.

[6] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, et al., CRISPR provides acquired resistance against viruses in prokaryotes, Science 315 (2007) 1709–1712.

[7] J.E. Garneau, M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA, Nature 468 (2010) (67–+).

[8] R. Sapranauskas, G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, V. Siksnys, The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*, Nucleic Acids Res. 39 (2011) 9275–9282.

[9] F.J.M. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system, Microbiology - SGM 155 (2009) 733–740.

[10] F.J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, J. Mol. Evol. 60 (2005) 174–182.

[11] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiology 151 (2005) 653–663.

[12] S.J. Brouns, M.M. Jore, M. Lundgren, E.R. Westra, R.J. Slijkhuis, A.P. Snijders, et al., Small CRISPR RNAs guide antiviral defense in prokaryotes, Science 321 (2008) 960–964.

[13] A. Bolotin, B. Quinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology 151 (2005) 2551–2561.

[14] C.R. Hale, P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, et al., RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex, Cell 139 (2009) 945–956.

[15] S.H. Sternberg, H. Richter, E. Charpentier, U. Qimron, Adaptation in CRISPR–Cas systems, Mol. Cell 61 (2016) 797–808.

[16] M.L. Hochstrasser, J.A. Doudna, Cutting it close: CRISPR-associated endoribonuclease structure and function, Trends Biochem. Sci. 40 (2015) 58–66.

[17] K.S. Makarova, D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, et al., Evolution and classification of the CRISPR–Cas systems, Nat. Rev. Microbiol. 9 (2011) 467–477.

[18] K.S. Makarova, F. Zhang, E.V. Koonin, Snapshot: class 1 CRISPR–Cas systems, Cell 168 (2017) 946-e1.

[19] E. Deltcheva, K. Chylinski, C.M. Sharma, K. Gonzales, Y. Chao, Z.A. Pirzada, et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III, Nature 471 (2011) 602–607.

[20] S. Shmakov, O. Abudayyeh Omar, S. Makarova Kira, I. Wolf Yuri, S. Gootenberg Jonathan, E. Semenova, et al., Discovery and functional characterization of diverse class 2 CRISPR–Cas systems, Mol. Cell 60 (2015) 385–397.

[21] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science 337 (2012) 816–821.

[22] G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) E2579–E2586.

[23] K.S. Makarova, F. Zhang, E.V. Koonin, Snapshot: class 2 CRISPR–Cas systems, Cell 168 (2017) 328-e1.

[24] D. Burstein, L.B. Harrington, S.C. Strutt, A.J. Probst, K. Anantharaman, B.C. Thomas, et al., New CRISPR–Cas systems from uncultivated microbes, Nature 542 (2017) 237–241.

[25] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, et al., An updated evolutionary classification of CRISPR–Cas systems, Nat. Rev. Microbiol. 13 (2015) 722–736.

[26] S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, et al., Diversity and evolution of class 2 CRISPR–Cas systems, Nat. Rev. Microbiol. 15 (2017) 169–182.

[27] W.X. Yan, S. Chong, H. Zhang, K.S. Makarova, E.V. Koonin, D.R. Cheng, et al., Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein, Mol. Cell 70 (2) (2018) 327–339.e5.

[28] S. Konermann, P. Lotfy, N.J. Brideau, J. Oki, M.N. Shokhirev, P.D. Hsu, Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors, Cell 173 (2018) (665-76 e14).

[29] O.O. Abudayyeh, J.S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D.B. Cox, et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector, Science 353 (2016) aaf5573.

[30] A. East-Seletsky, M.R. O'Connell, S.C. Knight, D. Burstein, J.H. Cate, R. Tjian, et al., Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection, Nature 538 (2016) 270–273.

[31] A.A. Smargon, D.B. Cox, N.K. Pyzocha, K. Zheng, I.M. Slaymaker, J.S. Gootenberg, et al., Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28, Mol. Cell 65 (2017) (618-30 e7).

[32] A. East-Seletsky, M.R. O'Connell, D. Burstein, G.J. Knott, J.A. Doudna, RNA targeting by functionally orthogonal type VI-A CRISPR–Cas enzymes, Mol. Cell 66 (2017) (373-83.e3).
[33] J.S. Gootenberg, O.O. Abudayyeh, J.W. Lee, P. Essletzbichler, A.J. Dy, J. Joung, et al., Nucleic acid detection with CRISPR–Cas13a/C2c2, Science 356 (2017) 438–442.
[34] O.O. Abudayyeh, J.S. Gootenberg, P. Essletzbichler, S. Han, J. Joung, J.J. Belanto, et al., RNA targeting with CRISPR–Cas13, Nature 550 (2017) 280–284.
[35] D.B.T. Cox, J.S. Gootenberg, O.O. Abudayyeh, B. Franklin, M.J. Kellner, J. Joung, et al., RNA editing with CRISPR–Cas13, Science 358 (2017) 1019–1027.
[36] J.S. Gootenberg, O.O. Abudayyeh, M.J. Kellner, J. Joung, J.J. Collins, F. Zhang, Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6, Science 360 (2018) 439–444.
[37] R. Aman, Z. Ali, H. Butt, A. Mahas, F. Aljedaani, M.Z. Khan, et al., RNA virus interference via CRISPR/Cas13a system in plants, Genome Biol. 19 (2018) 1.
[38] E. Schunder, K. Rydzewski, R. Grunow, K. Heuner, First indication for a functional CRISPR/Cas system in *Francisella tularensis*, Int. J. Med. Microbiol. 303 (2013) 51–60.
[39] G. Vestergaard, R.A. Garrett, S.A. Shah, CRISPR adaptive immune systems of Archaea, RNA Biol. 11 (2014) 156–167.
[40] B. Zetsche, J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, et al., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system, Cell 163 (2015) 759–771.
[41] E.V. Koonin, K.S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR–Cas systems, Curr. Opin. Microbiol. 37 (2017) 67–78.
[42] K.S. Makarova, V. Anantharaman, N.V. Grishin, E.V. Koonin, L. Aravind, CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems, Front. Genet. 5 (2014) 102.
[43] L. Liu, X. Li, J. Wang, M. Wang, P. Chen, M. Yin, et al., Two distant catalytic sites are responsible for C2c2 RNase activities, Cell 168 (2017) (121-34 e12).
[44] L. Liu, X. Li, J. Ma, Z. Li, L. You, J. Wang, et al., The molecular architecture for RNA-guided RNA cleavage by Cas13a, Cell 170 (2017) (714-26 e10).
[45] G.J. Knott, A. East-Seletsky, J.C. Cofsky, J.M. Holton, E. Charles, M.R. O'Connell, et al., Guide-bound structures of an RNA-targeting A-cleaving CRISPR–Cas13a enzyme, Nat. Struct. Mol. Biol. 24 (2017) 825–833.
[46] C. Zhang, S. Konermann, N.J. Brideau, P. Lotfy, S.J. Novick, T. Strutzenberg, et al., Structural basis for the RNA-guided ribonuclease activity of CRISPR–Cas13d, bioRxiv, 2018.
[47] V. Anantharaman, K.S. Makarova, A.M. Burroughs, E.V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing, Biol. Direct 8 (2013) 15.
[48] I. Fonfara, H. Richter, M. Bratovic, A. Le Rhun, E. Charpentier, The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA, Nature 532 (2016) 517–521.
[49] D.C. Swarts, J. van der Oost, M. Jinek, Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR–Cas12a, Mol. Cell 66 (2017) (221-33 e4).
[50] D. Dong, K. Ren, X. Qiu, J. Zheng, M. Guo, X. Guan, et al., The crystal structure of Cpf1 in complex with CRISPR RNA, Nature 532 (2016) 522–526.
[51] A. Hatoum-Aslan, I. Maniv, P. Samai, L.A. Marraffini, Genetic characterization of antiplasmid immunity through a type III-A CRISPR–Cas system, J. Bacteriol. 196 (2014) 310–317.
[52] L.K. Maier, A.E. Stachler, S.J. Saunders, R. Backofen, A. Marchfelder, An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein, J. Biol. Chem. 290 (2015) 4192–4201.
[53] E. Semenova, K. Kuznedelov, K.A. Datsenko, P.M. Boudry, E.E. Savitskaya, S. Medvedeva, et al., The Cas6e ribonuclease is not required for interference and adaptation by the *E. coli* type I-E CRISPR–Cas system, Nucleic Acids Res. 43 (2015) 6049–6061.
[54] S.A. Gorski, J. Vogel, J.A. Doudna, RNA-based recognition and targeting: sowing the seeds of specificity, Nat. Rev. Mol. Cell Biol. 18 (2017) 215–228.
[55] R.N. Jackson, P.B. van Erp, S.H. Sternberg, B. Wiedenheft, Conformational regulation of CRISPR-associated nucleases, Curr. Opin. Microbiol. 37 (2017) 110–119.
[56] W.E. Salomon, S.M. Jolly, M.J. Moore, P.D. Zamore, V. Serebrov, Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides, Cell 162 (2015) 84–95.
[57] L.M. Wee, C.F. Flores-Jasso, W.E. Salomon, P.D. Zamore, Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties, Cell 151 (2012) 1055–1067.
[58] S.H. Sternberg, B. Lafrance, M. Kaplan, J.A. Doudna, Conformational control of DNA target cleavage by CRISPR–Cas9, Nature 527 (2015) 110–113.
[59] A. Tambe, A. East-Seletsky, G.J. Knott, J.A. Doudna, M.R. O'Connell, RNA binding and HEPN nuclease activation are decoupled in CRISPR–Cas13a, bioRxiv, 2017.
[60] L.A. Marraffini, E.J. Sontheimer, Self versus non-self discrimination during CRISPR RNA-directed immunity, Nature 463 (2010) 568–571.
[61] C.R. Hale, S. Majumdar, J. Elmore, N. Pfister, M. Compton, S. Olson, et al., Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs, Mol. Cell 45 (2012) 292–302.
[62] P. Samai, N. Pyenson, W. Jiang, G.W. Goldberg, A. Hatoum-Aslan, L.A. Marraffini, Co-transcriptional DNA and RNA cleavage during type III CRISPR–Cas immunity, Cell 161 (2015) 1164–1174.
[63] G. Tamulaitis, M. Kazlauskiene, E. Manakova, Č. Venclovas, O. Nwokeoji Alison, J. Dickman Mark, et al., Programmable RNA shredding by the type III-A CRISPR–Cas system of *Streptococcus thermophilus*, Mol. Cell 56 (2014) 506–517.
[64] H.J. Staals Raymond, Y. Zhu, W. Taylor David, E. Kornfeld Jack, K. Sharma, A. Barendregt, et al., RNA targeting by the type III-A CRISPR–Cas Csm complex of Thermus thermophilus, Mol. Cell 56 (2014) 518–530.
[65] C.R. Hale, A. Cocozaki, H. Li, R.M. Terns, M.P. Terns, Target RNA capture and cleavage by the Cmr type III-B CRISPR–Cas effector complex, Genes Dev. 28 (2014) 2432–2443.
[66] J. Zhang, C. Rouillon, M. Kerou, J. Reeks, K. Brugger, S. Graham, et al., Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity, Mol. Cell 45 (2012) 303–313.
[67] E. Davidov, G. Kaufmann, RloC: a wobble nucleotide-excising and zinc-responsive bacterial tRNase, Mol. Microbiol. 69 (2008) 1560–1574.

[68] K.P. Lee, M. Dey, D. Neculai, C. Cao, T.E. Dever, F. Sicheri, Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing, Cell 132 (2008) 89–100.
[69] Y. Han, J. Donovan, S. Rath, G. Whitney, A. Chitrakar, A. Korennykh, Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response, Science 343 (2014) 1244–1248.
[70] H. Huang, E. Zeqiraj, B. Dong, B.K. Jha, N.M. Duffy, S. Orlicky, et al., Dimeric structure of pseudokinase RNase L bound to 2-5A reveals a basis for interferon-induced antiviral activity, Mol. Cell 53 (2014) 221–234.
[71] O. Niewoehner, M. Jinek, Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6, RNA 22 (2016) 318–329.
[72] N.F. Sheppard, C.V. Glover III, R.M. Terns, M.P. Terns, The CRISPR-associated Csx1 protein of *Pyrococcus furiosus* is an adenosine-specific endoribonuclease, RNA 22 (2016) 216–224.
[73] Y.K. Kim, Y.G. Kim, B.H. Oh, Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of *Pyrococcus furiosus*, Proteins 81 (2013) 261–270.
[74] M. Kazlauskiene, G. Kostiuk, C. Venclovas, G. Tamulaitis, V. Siksnys, A cyclic oligonucleotide signaling pathway in type III CRISPR–Cas systems, Science 357 (2017) 605–609.
[75] O. Niewoehner, C. Garcia-Doval, J.T. Rostol, C. Berk, F. Schwede, L. Bigler, et al., Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers, Nature 548 (2017) 543–548.
[76] W. Han, S. Pan, B. Lopez-Mendez, G. Montoya, Q. She, Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail, Nucleic Acids Res. 45 (2017) 10740–10750.
[77] W. Yang, Nucleases: diversity of structure, function and mechanism, Q. Rev. Biophys. 44 (2011) 1–93.
[78] E.V. Koonin, F. Zhang, Coupling immunity and programmed cell suicide in prokaryotes: life-or-death choices, BioEssays 39 (2017) 1–9.
[79] J. Donovan, S. Rath, D. Kolet-Mandrikov, A. Korennykh, Rapid RNase L-driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery, RNA 23 (2017) 1660–1671.
[80] S. Rath, J. Donovan, G. Whitney, A. Chitrakar, W. Wang, A. Korennykh, Human RNase L tunes gene expression by selectively destabilizing the microRNA-regulated transcriptome, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15916–15921.
[81] D.A. Cooper, B.K. Jha, R.H. Silverman, J.R. Hesselberth, D. J. Barton, Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs, Nucleic Acids Res. 42 (2014) 5202–5216.
[82] K. Das, T. Acton, Y. Chiang, L. Shih, E. Arnold, G.T. Montelione, Crystal structure of RlmAI: implications for understanding the 23S rRNA G745/G748-methylation at the macrolide antibiotic-binding site, Proc. Natl. Acad. Sci. 101 (2004) 4041–4046.
[83] M. Jinek, F.G. Jiang, D.W. Taylor, S.H. Sternberg, E. Kaya, E. B. Ma, et al., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation, Science 343 (2014) (1215-).
[84] T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I.M. Slaymaker, Y. Li, et al., Crystal structure of Cpf1 in complex with guide RNA and target DNA, Cell 165 (2016) 949–962.
[85] K.S. Makarova, L. Aravind, Y.I. Wolf, E.V. Koonin, Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR–Cas systems, Biol. Direct 6 (2011) 38.
[86] B. Wiedenheft, S.H. Sternberg, J.A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea, Nature 482 (2012) 331–338.
[87] E.V. Koonin, K.S. Makarova, CRISPR–Cas: evolution of an RNA-based adaptive immunity system in prokaryotes, RNA Biol. 10 (2013) 679–686.
[88] J.L. Gardy, F.S. Brinkman, Methods for predicting bacterial protein subcellular localization, Nat. Rev. Microbiol. 4 (2006) 741–751.
[89] C.F. Schuster, R. Bertram, Toxin–antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate, FEMS Microbiol. Lett. 340 (2013) 73–85.
[90] S.A. Shmakov, K.S. Makarova, Y.I. Wolf, K.V. Severinov, E.V. Koonin, Systematic prediction of genes functionally linked to CRISPR–Cas systems by gene neighborhood analysis, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E5307-E16.
[91] S. Hein, I. Scholz, B. Voss, W.R. Hess, Adaptation and modification of three CRISPR loci in two closely related cyanobacteria, RNA Biol. 10 (2013) 852–864.
[92] N.M. Andis, C.W. Sausen, A. Alladin, M.L. Bochman, The WYL domain of the PIF1 helicase from the thermophilic bacterium *Thermotoga elfii* is an accessory single-stranded DNA binding module, Biochemistry 57 (2018) 1108–1118.
[93] J.P. Mackay, J. Font, D.J. Segal, The prospects for designer single-stranded RNA-binding proteins, Nat. Struct. Mol. Biol. 18 (2011) 256–261.
[94] M.R. O'Connell, B.L. Oakes, S.H. Sternberg, A. East-Seletsky, M. Kaplan, J.A. Doudna, Programmable RNA recognition and cleavage by CRISPR/Cas9, Nature 516 (2014) 263–266.
[95] D.A. Nelles, M.Y. Fang, M.R. O'Connell, J.L. Xu, S.J. Markmiller, J.A. Doudna, et al., Programmable RNA tracking in live cells with CRISPR/Cas9, Cell 165 (2016) 488–496.
[96] Y. Liu, Z. Chen, A. He, Y. Zhan, J. Li, L. Liu, et al., Targeting cellular mRNAs translation by CRISPR–Cas9, Sci. Rep. 6 (2016) 29652.
[97] R. Batra, D.A. Nelles, E. Pirie, S.M. Blue, R.J. Marina, H. Wang, et al., Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9, Cell 170 (2017) (899-912 e10).
[98] R. Batra, D.A. Nelles, F. Krach, J.D. Thomas, L. Snjader, S.M. Blue, et al., Reversal of molecular pathology by RNA-targeting Cas9 in a myotonic dystrophy mouse model, bioRxiv, 2017.
[99] S.C. Strutt, R.M. Torrez, E. Kaya, O.A. Negrete, J.A. Doudna, RNA-dependent RNA targeting by CRISPR–Cas9, elife 7 (2018).
[100] G. Dugar, R.T. Leenay, S.K. Eisenbart, T. Bischler, B.U. Aul, C. L. Beisel, et al., CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9, Mol. Cell 69 (2018) (893-905 e7).
[101] A. Lapinaite, J.A. Doudna, J.H.D. Cate, Programmable RNA recognition using a CRISPR-associated argonaute, Proc. Natl. Acad. Sci. 115 (13) (2018) 3368–3373.
[102] B.A. Rousseau, Z. Hou, M.J. Gramelspacher, Y. Zhang, Programmable RNA cleavage and recognition by a natural CRISPR–Cas9 system from *Neisseria meningitidis*, Mol. Cell 69 (2018) (906-14 e4).
[103] D.G. Sashital, Pathogen detection in the CRISPR–Cas era, Genome Med. 10 (2018) 32.