doi:10.1016/j.jmb.2012.06.024 J. Mol. Biol. (2012) 422, 607–616

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

REVIEW

Tuning in to Interference: R-Loops and Cascade Complexes in CRISPR Immunity

Ivana Ivančić-Baće[1], Jamieson AL Howard[2] and Edward L. Bolt[2*]

[1]*Department of Molecular Biology, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia*
[2]*School of Biomedical Sciences, University of Nottingham Medical School, Queens Medical Centre, Nottingham NG7 2UH, UK*

Received 28 May 2012; received in revised form 13 June 2012; accepted 16 June 2012
Available online 25 June 2012

Edited by B. Connolly

Keywords:
CRISPR;
R-loop;
Cascade;
Cas3;
recombination

Stable RNA–DNA hybrids formed by invasion of an RNA strand into duplex DNA, termed R-loops, are notorious for provoking genome instability especially when they arise during transcription. However, in some instances (DNA replication and class switch recombination), R-loops are useful so long as their existence is carefully managed to avoid them persisting. A recent flow of research papers establishes a newly discovered use for R-loops as key intermediates in a prokaryotic immune system called CRISPR (Clustered *R*egularly *I*nterspersed *S*hort *P*alindromic *R*epeats). Structures and mechanism of ribonucleoprotein complexes ("Cascades") that form CRISPR R-loops highlight precision targeting of duplex DNA that has sequence characteristics marking it as foe, enabling nucleolytic destruction of DNA and recycling the Cascade. We review these significant recent breakthroughs in understanding targeting/interference stages of CRISPR immunity and discuss questions arising, including a possible link between targeting and adaptive immunity in prokaryotes.

© 2012 Elsevier Ltd. All rights reserved.

RNA–DNA Hybrids and R-Loops

Double-helical structures of B-form DNA and A-form RNA were reported in the 1950s,[1–4] but to satisfy the central "dogma" of molecular biology ("DNA makes RNA makes protein"), it was apparent that RNA–DNA hybrids might need to form. This was shown to be possible in 1960,[5] and their atomic structures show stable duplex that is closer in dimensions to A-form helix.[6] Physiological relevance of natural RNA–DNA hybrids has been described in several contexts, summarised in Table 1. Most familiar are hybrids formed during transcription and DNA replication. Atomic resolution structures of RNA polymerase transcription complexes show short (12–20 bp) RNA–DNA hybrids.[17–19] The hybrids are transient as structural features of RNA polymerases and proximity of RNA binding proteins (e.g., for translation) steward nascent RNA away from template DNA.[20,21] Transient RNA–DNA hybrids are also characteristic of DNA replication, as Okazaki fragments at replication fork lagging strands[22,23] and in tRNA primed DNA synthesis by a reverse transcriptase first reported in 1970.[24,25]

RNA–DNA hybrids are sometimes stabilised as R-loops, defined as an RNA strand invaded into duplex DNA, base pairing with one DNA strand and displacing the other as a single-stranded DNA (ssDNA) loop (Fig. 1). The name R-loop is used because of the resemblance of these molecules to strand invasion intermediates of homologous recombination called D-loops, which are composed solely of DNA. R-loops themselves are implicated in a specialised type of recombination, site-specific

*Corresponding author. E-mail address: ed.bolt@nottingham.ac.uk.

Abbreviations used: ssDNA, single-stranded DNA; CSR, class switch recombination; crRNA, CRISPR RNA; pre-crRNA, precursor crRNA; PAM, protospacer adjacent motif.

0022-2836/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

Table 1. Physiological relevance of R-loops in biology

Process	RNA–DNA hybrid/R-loop	Details
Transcription	Nascent RNA synthesis from DNA templates"Thread-back" and "extended hybrid" models of R-loop persistence during transcription	7, 8
Epigenetics: methylation	Protection of CpG promoter islands from methylation	9
DNA replication	Priming of lagging strand synthesis.Priming of plasmid (ColE1), viral and mitochondrial replication	10–13
Genome instability	R-loops provoking illegitimate recombination	8, 14
CSR	R-loops at G-rich sequence provoke immunoglobulin diversity	15
Telomere processing	"t-loop" of RNA and G-rich DNA	16
CRISPR targeting	R-loop of crRNA targeting invasive DNA	This work

Table 2. Proteins that prevent, control and reverse R-loop formation

Protein	Activity	Details
RNaseHI	Ribonuclease on RNA strands base paired to DNA	29
Topoisomerase I	Relaxation of negatively supercoiled DNA	30, 31
RecG	Helicase unwinding R-loops	32
Pif1	Helicase unwinding RNA–DNA hybrids	33
Senataxin/Sen1	Helicases unwinding RNA–DNA hybrids	34
Various transcription termination and mRNA processing factors	Bind to nascent RNA to prevent thread-back R-loops	8
Cas3	Helicase unwinding R-loops	35

class switch recombination (CSR) in vertebrate immune systems.[15] Guanine-rich DNA sequences are important for initiating R-loop formation in CSR, and DNA negative supercoiling is another key context that promotes R-loop formation.[7,26] R-loops were first identified at bacterial ColE1 plasmid origins of replication,[27] and similar mechanisms prime replication in mitochondria[10] and viruses.[11] R-loops were also seen in *in vitro* transcription reactions[28] that led to many more studies establishing the importance of RNaseH in removing R-loops after initiation of replication and of DNA topoisomerase I preventing R-loop formation during transcription, summarised in Table 2. The importance of R-loop management in cells is made clear by studies showing mutagenic effects of persistent R-loops through conflict with DNA replication and provoking of illegitimate homologous recombination, reviewed recently.[8]

Here, we focus on a protein–RNA complex that forms R-loops by a precise targeting mechanism guided to duplex DNA. The mechanism is a part of CRISPR/Cas immunity, a newly discovered system in bacteria and archaea that can seek and destroy invasive nucleic acids from virus or plasmid. Emergence of CRISPR/Cas has been striking, not least because its mechanisms are underpinned by adaptive immunity that is expressed in RNA sequences. We highlight some exciting recent work that has provided elegant detail for CRISPR/Cas targeting to DNA, using protein–RNA complexes from bacteria and archaea, and aim to unify this with more general interest in R-loops.[8]

Discovery of CRISPR/Cas RNA-Based Immunity in Bacteria and Archaea

The bedrock of CRISPR/Cas systems is repetitive DNA sequence in prokaryotes called CRISPR (Clustered *R*egularly *I*nterspersed *S*hort *P*alindromic *R*epeats). The name CRISPR was first used in 2002,[36] but *Escherichia coli* CRISPRs were noted earlier as unusual sequence repeats proximal to a gene encoding an alkaline phosphatase isozyme (*iap*).[37] Similar short repetitive sequences under different names were described in bacteria and archaea,[38–40] distinguished by significant differences to other classes of widespread short sequence repeats in bacteria, ERICs and REPs. In a landmark

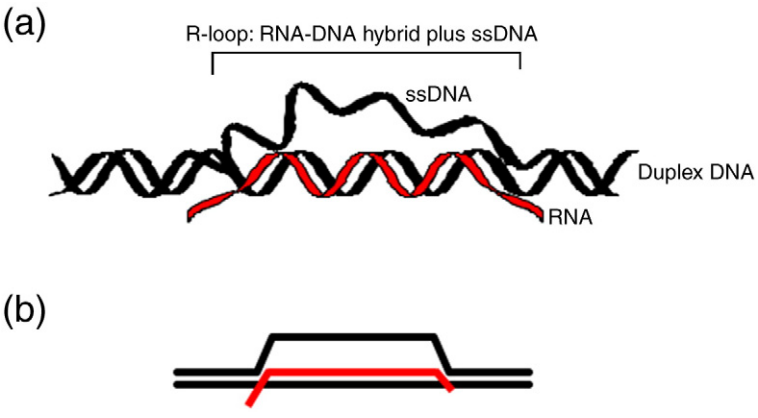

Fig. 1. (a) R-loops are characterised by invasion of an RNA strand into duplex DNA *via* base pairing with one DNA strand thereby displacing the other as ssDNA. For simplicity, R-loops are often drawn as in (b).

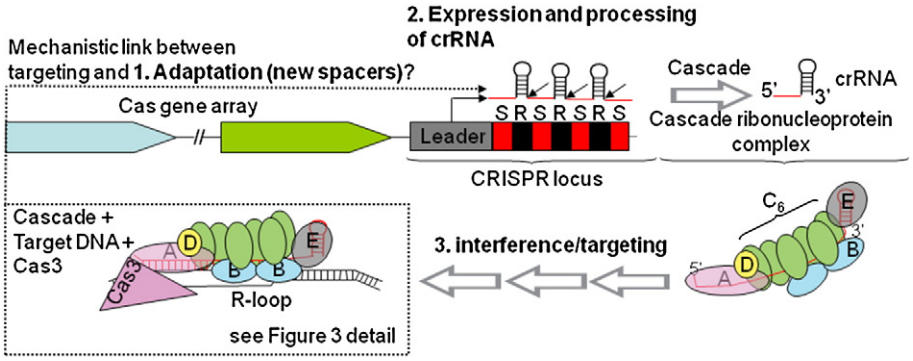

Fig. 2. Major processes in CRISPR systems (adaptation, expression and interference/targeting) summarised with emphasis on Cascade systems that target duplex DNA with crRNA. See the text for details. A CRISPR locus, comprising spacers (S), repeats (R) and a promoter/leader sequence, is the source of RNA message (pre-crRNA) that is processed by Cascade into crRNA to form a ribonucleotide–protein complex. Cascade catalyses targeting of DNA with crRNA, detailed in Fig. 3. Targeting results in R-loop intermediates and recruitment of Cas3 nuclease–helicase, processes that might be linked to generation of DNA fragments (protospacers) for incorporation into CRISPR as spacers by unknown mechanisms.

paper,[36] comparison of these new sequence repeats in multiple species brought on the name CRISPR, defined by two factors: (a) Sequence repeats of 20–30 nt are interspersed with "spacers" of variable non-repeat sequence of 20–75 nt. (b) CRISPR associated (Cas) proteins that are conserved in many organisms with amino acid motifs characteristic of nucleic acid processing enzymes (Fig. 2). Significantly, at about the same time, CRISPR RNA (crRNA) transcripts that were much shorter than full-length CRISPR loci were detected, suggesting endonucleolytic processing of crRNA.[43] At that point, the prevailing mood was that CRISPR/Cas, particularly Cas proteins, cooperated in DNA replication, recombination and repair.[44–46] Observations in bacteria and archaea that the non-repeat spacers within CRISPR matched sequences from plasmids, phages and prophages,[47–49] stimulated a new hypothesis: CRISPR provides defence against viruses in archaea[50] through an immune system based on RNAi (*RNA i*nterference)-like mechanisms.[51] In 2007, experimental manipulation of CRISPR spacer sequences in *Streptococcus thermophilus* DGCC7710, an important bacterium for the dairy industry, was shown to protect cells from decimation by phage, dependent on a then unknown activity of a Cas protein (Cas5, now called Cas9/Csn1).[42]

How Does CRISPR Work? Multifarious Nucleic Acid Processing

CRISPR/Cas systems are genetically diverse and show extensive variation in composition of encoded proteins. *S. thermophilus* DGCC7710 exemplifies this, possessing four CRISPR/Cas systems (CRISPR-1–CRISPR-4) each with different complements of Cas proteins.[52,53] The common theme in each system is the presence of a CRISPR locus, although there is variation in the number of repeats and repeat/spacer sequence composition. However, conservation of underlying principles of CRISPR immunity in different species was shown recently, by introduction of *S. thermophilus* CRISPR-3 into *E. coli* conferring heterologous protection against plasmid and phage.[53]

CRISPR/Cas systems have undergone two iterations of classification based on Cas protein distribution that is subdivided into three types (I, II and III).[54,55] This is reflected in major mechanistic differences in how CRISPR sequences are manipulated for immunity and protection. Overall, in all systems, CRISPR/Cas operates through three major processes, each catalysed by distinct nucleic acid processing enzymes (Fig. 2): (a) *adaptation*, incorporation of a DNA fragment from a nonself invasive element (called a "pre-spacer") into a CRISPR locus as a "spacer" alongside a new CRISPR repeat, and (b) *transcription* and processing of crRNA to arm a protein complex for (c) *interference/targeting* of nucleic acid sequence, called a "protospacer" if the target is DNA, by homologous sequence in crRNA leading to destruction of the invasive plasmid or phage.

Recent reviews detail these processes and regulation of *cas* genes and CRISPR.[52,56–61] In summary, adaptation is poorly understood, but CRISPR loci can acquire new spacer sequence derived from nonself DNA, alongside a new repeat, in response to exposure to phage.[62–64] In the *E. coli* type I CRISPR system, adaptation relies on nucleases Cas1 and Cas2[65,66] through unknown mechanisms in which protospacer is directed for incorporation as a

spacer.[67] Adaptation in the type II CRISPR system (CRISPR-3) of *S. thermophilus* requires Csn2 protein, also by unknown mechanisms.[41,42,68,69]

Transcription of CRISPR from an AT-rich leader sequence generates precursor crRNA (pre-crRNA), which is endonucleolytically shortened into mature crRNA by cleavage in repeat sequences. Two major nuclease groups, Cas6 and Cas9, catalyse crRNA formation, depending on the particular CRISPR/Cas system: using the nomenclature of reference,[55] these are called Cas6,[70,71] Cas6e (previously called CasE and Cse3),[72–74] Cas6f (previously, Csy4)[75] and Cas9–RNase III complex.[76] crRNA is used to target foreign nucleic acid by protein complexes that guide it to complementary nucleotide sequence. In archaeal type III CRISPR systems of *Pyrococcus furiosus* and *Sulfolobus solfataricus*, crRNA is held in a complex called CMR[77–79] (*C*as *m*odule *R*AMP, where RAMP was originally used to mean Repeat-Associated Mysterious Proteins[51]). CMR targets crRNA to other RNA molecules, presumed to be viral mRNA, which are cleaved with mechanistic differences between *P. furiosus* and *S. solfataricus* CMR.[77–79] In the next sections, we focus on CRISPR interference strategies that use crRNA to target duplex DNA in stable DNA–RNA hybrids. Emerging evidence suggests two main routes for this, each requiring different Cas proteins: Cas9 protein typical of type II CRISPR systems or a multi-protein complex called Cascade, synonymous with the interference stage of type I CRISPR systems.

crRNA to Target DNA: Cas9 and Cascade

Cas9

First models of CRISPR/Cas as a prokaryotic immunity system suggested that Cas proteins manipulate crRNA to target invading RNA, similarly to RNAi in eukaryotes.[51] CMR complexes do target RNA, as noted in the previous section. Similarities and distinctions between CRISPR systems and eukaryotic RNAi have been reviewed recently.[61] Other CRISPR/Cas systems use crRNA to target DNA. Before biochemical mechanisms of crRNA–DNA targeting came to light, it was known that spacer sequences transcribed to crRNA in the bacterium *S. thermophilus* match phage DNA sequence from both coding and template strands,[47] suggesting targeting of either strand of phage DNA. This was also inferred in *E. coli*, in experiments using an engineered CRISPR/Cas system to protect against phage λ*vir*,[72] and in *Staphylococcus epidermidis*. In the latter, a nickase gene from a conjugative plasmid could be blocked by a CRISPR system at the level of DNA but not mRNA.[80] Crucially, another study in *S. thermophilus* showed that double-stranded DNA is targeted rapidly (<2 min) by CRISPR/Cas, by a mechanism that required Cas5 (now called Cas9), resulting in nuclease degradation of target DNA.[81] An absolute requirement for *S. thermophilus* Cas9 in CRISPR targeting to plasmid DNA was confirmed later.[53] Cas9 may also have a role in RNaseIII-mediated processing of pre-crRNA to crRNA in some species.[76] Not much is known about the structure or function of Cas9 protein. It is a large protein (>1200 amino acids) with conserved motifs typical of RuvC/RNaseH nucleases that are needed for Cas9 function *in vivo*.[53] Perhaps, Cas9 multitasks, contributing to preparation of crRNA that is retained for targeting and destruction of DNA in a way analogous to Cascade complexes, which we now look at in detail.

Cascade: A crRNA interference complex in type I systems

Cascade (CRISPR-Associated Complex for Antiviral Defence) is a multi-protein ribonucleic acid complex that forms R-loops with complementary DNA sequence, catalysing interference stages of CRISPR immunity in bacteria and archaea. The complex was identified in *E. coli*, in which it remains tightly bound to crRNA that it has formed from pre-crRNA by endonucleolytic cleavage.[72] In the same study an important proof of principle was reported; co-expression of Cascade with a crRNA construct complementary to viral DNA, as well as an additional protein (Cas3), was sufficient to give robust protection of *E. coli* cultures against plaque formation by phage λ*vir*. The use of ectopic expression of these individual components from plasmids in an *E. coli* strain lacking *cas* genes was prescient, side-stepping BaeSR, H-NS/LeuO-mediated transcriptional regulation of *E. coli* CRISPR/*cas* that subsequently came to light.[82–84] The hypothesis that crRNA is targeted to viral DNA of either strand, involving Cascade and Cas3 proteins that are widely distributed in type I CRISPR systems, was established.

E. coli Cascade comprises five proteins, originally called CasABCDE,[72] now Cse1(CasA), Cse2(CasB), Cas7(CasC), Cas5(CasD) and Cas6e(CasE). All of these are required to protect *E. coli* cells against lytic phage in plaque assays.[85] A remarkable sequence of work in the last few years has established how *E. coli* Cascade forms R-loops. These principles probably translate into other Cascade complexes, based on functional studies of a Cas5–Cas7–crRNA complex in *S. solfataricus*,[86] an archaeal species that also processes crRNA within a CMR complex,[79] and studies of *Pseudomonas aeruginosa* PA14 Csy1–Csy4 (now called Csy1, Csy2, Csy3 and Cas6f),[87] which has structural similarities to Cascade from *E. coli*[85] despite there being little sequence similarity between the respective proteins.

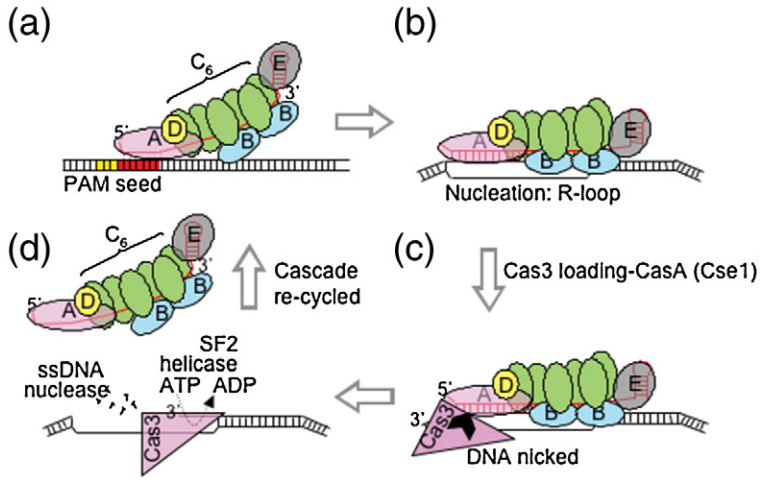

Fig. 3. Life cycle of a bacterial Cascade complex, using *E. coli* as a model system. See the main text for details. For clarity of labelling, proteins of the Cascade complex are labelled CasABCDE, as in Ref. 41, instead of using recently adopted nomenclature, explained in the text (from Ref. 42). (a) CasA (Cse1) is crucial for initial interaction of Cascade with target DNA of complementary seed sequence. (b) This results in some rearrangement of proteins CasB and CasE (Cse2 and Cas6e) and extensive base pairing of crRNA with target DNA and R-loop formation. (c) Cas3 is recruited to targeting complex *via* interaction with CasA and can load onto 3′ ended ssDNA enabling its DNA nicking, nuclease and ATP-dependent 3′-to-5′ translocase activities to degrade the DNA target (d). This may also provide the mechanism for release of Cascade from target DNA with crRNA in tact for subsequent interference reactions.

E. coli Cascade has a stoichiometry of 1:2:6:1:1 (Cse1:Cse2:Cas7:Cas5:Cas6e) in a 405-kDa protein complex bound to crRNA.[85] This crRNA comprises a 32-nt spacer core flanked with an 8-nt 5′ handle and a 21-nt 3′ hairpin loop. The overall morphology of Cascade, described as a seahorse,[85] has a prominent backbone of six Cas7 subunits within which a crRNA cord dictates assembly of Cas7 as a nucleoprotein helix.[88] Filamentous Cas7 backbone is also clear when imaged from *S. solfataricus*.[86] A 5′ end crRNA hook neatly delimits assembly of multiple Cas7 subunits through interaction with the sixth Cas7 subunit.[88] Cse1 is also proximal to the 5′ hook. At the other end of the complex is Cas6e, the Cascade ribonuclease that remains bound to crRNA 3′ end, which is a hairpin loop in *E. coli*.

A Recipe for R-Loops: Cascade, PAM, Seeds and Supercoils

Cascade establishes an R-loop by base pairing crRNA to protospacer DNA that has sequence and structure betraying it as a *bona fide* target (Fig. 3). Cascade and Csy1–Csy4 operate a policy of selective engagement with DNA, triggering R-loop formation when encountering "seed" sequence and, at least in the *E. coli* system, a protospacer adjacent motif (PAM). The seed sequence is probably a common mechanism in all crRNA–DNA pairing through thermodynamic considerations, offering a minimum length requirement for a stable RNA–DNA hybrid forming at the 5′ end of crRNA that displaces an unpaired DNA strand in an R-loop. Similar seeding is responsible for RNA–RNA pairing by Argonautes in eukaryotic RNAi[61] and is important for both Cascade and Csy1–Csy4 complexes.[87,89] Similarities have been noted in structure and function of nucleoprotein filaments formed by archaeal Cas7–crRNA and the bacterial recombinase RecA in complex with ssDNA.[86] RecA filaments catalyse pairing with homologous duplex DNA forming D-loops from seed sequences, and it may be the case that similar forces apply in R-loop formation by Cascades.

In contrast to seeding of R-loop formation, in ways probably common in other biological processes, PAMs can give CRISPR interference sequence specificity. PAMs are typically of 2–3 nt, upstream of protospacer DNA, and they do not match crRNA sequence. Originally, PAMs were recognised as a pattern of nucleotide motifs juxtaposed to protospacer/pre-spacer sequence of invasive DNA in *S. thermophilus* and other CRISPR loci.[47,63,90,91] In *E. coli*, where a trinucleotide PAM is located at positions −3 to −1 relative to crRNA–protospacer base pairing, substitutions to PAM in M13 phage allowed escape from *E. coli* Cascade targeting, measured as efficiency of plaque formation on *E. coli* cell lawns.[89] The presence, absence or manipulation of PAM sequence in plasmid targets resulted in altered efficiency of binding by purified *E. coli* Cascade.[92] Dissociation constants (K_d) were at least 19 times higher for Cascade binding a negatively supercoiled plasmid lacking PAM compared to the same plasmid with a PAM and 3 times higher when PAM sequence was altered to an escape sequence that had been identified in *E. coli* by plaque assays.[89]

It had been noted previously that DNA binding by *E. coli* Cascade relied on its Cse1 subunit[85] and that target binding by Cascade can result in movement of Cse1 (as well as Cse2 and Cse6e) relative to the Cas7 backbone.[88] Mechanistic detail

of Cse1 function is now available from biochemistry of *E. coli* Cascade and the crystal structure of Cse1 from *T. thermophilus*.[93] The key to Cse1 in this instance is a trio of amino acids that includes a phenylalanine forming a loop (L1) that is needed for effective protection by Cascade *in vivo* and is positioned such that it might recognise PAM and destabilise target duplex. Interestingly, *Sulfolobus* Cascade lacks a clear sequence homologue of Cse1 yet PAM sequences are important for targeting reactions in this organism.[86,94] *P. aeruginosa* Csy1–Csy4 lacks Cse1 and shows only sequence-specific DNA binding, requiring the seed DNA sequence. These recent insights into Cascade molecular biology indicate how this nucleoprotein complex sets up an R-loop by distinguishing nonself DNA from self and scanning for seeding sites.

Recent work revealed an elegant solution to how *E. coli* Cascade gains energy for scanning, seeding and conformational movements: negatively supercoiled DNA.[92] The chromosome of *E. coli* is maintained in a negatively supercoiled state. It is well known that hypernegative supercoiling promotes R-loop formation globally in bacteria, which causes cell inviability if supercoils are not relaxed by topoisomerase I.[8] Cascade provides the means to exploit this in a targeted way, once it has established that DNA is nonself by mechanisms described above. All that is then left to do is destroy target DNA. This is less well understood, but recent work has confirmed that interaction of Cascade with Cas3 protein facilitates this.

How to Destroy an R-Loop: Cas3 Nuclease–Helicase

Interplay between Cas3 and Cascade in protection against phage λ*vir* was identified in *E. coli*.[72] Subsequent biochemical studies were above all in other species because of instability of *E. coli* Cas3 for purification unless fused to maltose binding protein[35] or more recently to Cse1.[92] The latter work confirmed that *E. coli* Cas3 is a nuclease active against ssDNA, with both exonuclease and endonuclease activities residing in an HD domain, in agreement with nuclease activities of other Cas3 proteins.[95–98] In addition to HD nuclease activity, Cas3 has superfamily 2 helicase motifs that power ATP-dependent translocation 3′ to 5′ along ssDNA,[98] through R-loops.[35] Cas3 translocation stimulates exonuclease activity,[92,95] consistent with a multitasking machine that can processively degrade the invasive target DNA (Fig. 3). Fusion of *E. coli* Cas3 with Cse1 for *in vitro* studies efficiently targeted Cas3 nuclease to plasmid DNA with PAM and seed sequences and mimics strong evidence for physical interaction of Cse1 with Cas3 in *E. coli*.[92]

Summary Remarks: CRISPR Targeting Linked to Spacer Acquisition?

E. coli Cascade, archaeal Cascade and Csy1–Csy4 complexes all catalyse a core process, DNA-targeted R-loop formation. This can be adorned by mechanistic specificity arising from accessory proteins, Cse1 and unknown partners in archaea. This is parallel in DNA helicase–translocase superfamilies that manipulate diverse nucleic acid structures using specificity inducing accessory domains surrounding a RecA core of ATPase domains. A mechanistic framework for CRISPR/Cas, at least in *E. coli*, is now apparent: Transcriptional activation of CRISPR and Cascade[82–84,99] sets in motion targeting of Cascade ribonucleoprotein complex to invasive DNA, forming a stable R-loop (Fig. 3). Cas3 is recruited *via* Cse1, which places it close to the 3′ tail of target DNA upstream of the crRNA–DNA seed sequence. Cas3 nicks DNA to relax supercoiling for robust degradation of the remaining DNA strand using 3′-to-5′ ATP-dependent translocase–nuclease activities.

Cas3 nuclease activities can be mapped onto crystal structures of Cas3 HD domains (called Cas3″),[95,97] but there is currently no crystal structure of Cas3 superfamily 2 helicase domains. When this information becomes available, it might cast some light on some interesting conserved features at the C-terminus of Cas3 sequences that are not canonical helicase motifs and with no known function at present. To understand how Cas3 is orientated onto target DNA, it will be necessary to understand how it interacts with Cse1: precise loading of Cas3 for translocation 3′ to 5′ along target ssDNA would be an effective means of displacing Cascade–crRNA complex undamaged for recycling to subsequent targets. It is likely that crRNA in R-loops is protected from RNaseH, a potent nuclease against RNA in RNA–DNA hybrids, by its extensive shielding by the Cascade/Csy1–Csy4 backbone. A recent report that *E. coli* Cas3 is a client protein for the Hsp90 protein HtpG[100] offers an intriguing possibility that Cas3, which is not co-transcribed with Cascade, is tethered until recruited by Cse1/Cascade. This might be a way to prevent indiscriminate DNase–helicase activity or interaction with other cellular proteins, but there is no information on handover of Cas3 from HtpG to Cse1.

Does R-loop formation by Cascade and/or nuclease processing by Cas3 promote acquisition of protospacer DNA into CRISPR loci as spacers? It is tempting to speculate that Cascade R-loops might trigger recombination or integration of invasive DNA into CRISPR loci akin to R-loop-directed CSR into repetitive switch (S) regions in mammalian immune systems.[15] S-regions, like CRISPR loci, contain short repeats (25–80 bp) and sequence

motifs that direct recombination events downstream of a promoter. At the time of writing, a feedback loop between CRISPR interference and adaptation has been suggested.[67] Given how little is known about adaptation, presumably involving at least Cas1 and Cas2 proteins, this should prove another interesting mechanism to work out.

Acknowledgements

We are grateful to the Biotechnology and Biological Sciences Research Council for funding a PhD studentship to J.A.L.H. and to the Croatian Ministry of Science, Education and Sports funding through grant 119-1191196-1201 to I.I.-B. Thanks to Simon Cass for critical reading of the manuscript and to Vlado Ivancic-Bace for help with Fig. 2.

References

1. Franklin, R. E. & Gosling, R. G. (1953). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, **172**, 156–157.
2. Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
3. Wilkins, M. H., Seeds, W. E., Stokes, A. R. & Wilson, H. R. (1953). Helical structure of crystalline deoxypentose nucleic acid. *Nature*, **172**, 759–762.
4. Rich, A. & Davies, D. (1956). A new two-stranded helical structure: polyadenylic acid and polyuridylic acid. *J. Am. Chem. Soc.* **78**, 3548–3549.
5. Rich, A. (1960). A hybrid helix containing both deoxyribose and ribose polynucleotides and its relation to the transfer of information between the nucleic acids. *Proc. Natl Acad. Sci. USA*, **46**, 1044–1053.
6. Noy, A., Perez, A., Marquez, M., Luque, F. J. & Orozco, M. (2005). Structure, recognition properties, and flexibility of the DNA·RNA hybrid. *J. Am. Chem. Soc.* **127**, 4910–4920.
7. Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L. & Lieber, M. R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation *in vitro*: a nick can serve as a strong R-loop initiation site. *Mol. Cell. Biol.* **30**, 146–159.
8. Aguilera, A. & Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol. Cell*, **46**, 115–124.
9. Vertino, P. M. & Wade, P. A. (2012). R loops: lassoing DNA methylation at CpGi. *Mol. Cell*, **45**, 708–709.
10. Lee, D. Y. & Clayton, D. A. (1997). RNase mitochondrial RNA processing correctly cleaves a novel R loop at the mitochondrial DNA leading-strand origin of replication. *Genes Dev.* **11**, 582–592.
11. Kreuzer, K. N. & Brister, J. R. (2010). Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol. J.* **7**, 358.
12. Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238.
13. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
14. Li, X. & Manley, J. L. (2006). Cotranscriptional processes and their influence on genome stability. *Genes Dev.* **20**, 1838–1847.
15. Stavnezer, J., Guikema, J. E. & Schrader, C. E. (2008). Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**, 261–292.
16. Feuerhahn, S., Iglesias, N., Panza, A., Porro, A. & Lingner, J. (2010). TERRA biogenesis, turnover and implications for function. *FEBS Lett.* **584**, 3812–3818.
17. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, **292**, 1876–1882.
18. Kettenberger, H., Armache, K. J. & Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, **16**, 955–965.
19. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, **303**, 1014–1016.
20. Naryshkina, T., Kuznedelov, K. & Severinov, K. (2006). The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA–DNA hybrid. *J. Mol. Biol.* **361**, 634–643.
21. Toulokhonov, I. & Landick, R. (2006). The role of the lid element in transcription by *E. coli* RNA polymerase. *J. Mol. Biol.* **361**, 644–658.
22. Ogawa, T. & Okazaki, T. (1980). Discontinuous DNA replication. *Annu. Rev. Biochem.* **49**, 421–457.
23. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.. *Proc. Natl Acad. Sci. USA*, **59**, 598–605.
24. Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226**, 1209–1211.
25. Temin, H. M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226**, 1211–1213.
26. Drolet, M., Bi, X. & Liu, L. F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation *in vitro*. *J. Biol. Chem.* **269**, 2068–2074.
27. Itoh, T. & Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl Acad. Sci. USA*, **77**, 2450–2454.
28. Kadesch, T. R. & Chamberlin, M. J. (1982). Studies of *in vitro* transcription by calf thymus RNA polymerase II using a novel duplex DNA template. *J. Biol. Chem.* **257**, 5286–5295.
29. Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **44**, 1–93.

30. Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* **59**, 723–730.
31. Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A. *et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324.
32. Vincent, S. D., Mahdi, A. A. & Lloyd, R. G. (1996). The RecG branch migration protein of *Escherichia coli* dissociates R-loops. *J. Mol. Biol.* **264**, 713–721.
33. Boule, J. B. & Zakian, V. A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res.* **35**, 5809–5818.
34. Mischo, H. E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A. G., Wei, W., Steinmetz, L. *et al.* (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell*, **41**, 21–32.
35. Howard, J., Delmas, S., Ivancic-Bace, I. & Bolt, E. L. (2011). Helicase dissociation and annealing of RNA–DNA hybrids by *Escherichia coli* Cas3 protein. *Biochem. J.* http://dx.doi.org/10.1042/BJ20110901.
36. Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575.
37. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433.
38. Groenen, P. M., Bunschoten, A. E., van Soolingen, D. & van Embden, J. D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**, 1057–1065.
39. Mojica, F. J., Ferrer, C., Juez, G. & Rodriguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* **17**, 85–93.
40. She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M. & Zillig, W. (1998). Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles*, **2**, 417–425.
41. Nam, K. H., Kurinov, I. & Ke, A. (2011). Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca^{2+}-dependent double-stranded DNA binding activity. *J. Biol. Chem.* **286**, 30759–30768.
42. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S. *et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
43. Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M. *et al.* (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
44. Guy, C. P., Majernik, A. I., Chong, J. P. & Bolt, E. L. (2004). A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system. *Nucleic Acids Res.* **32**, 6176–6186.
45. Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B. & Koonin, E. V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **30**, 482–496.
46. Mojica, F. J., Diez-Villasenor, C., Soria, E. & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246.
47. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
48. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182.
49. Pourcel, C., Salvignol, G. & Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653–663.
50. Lillestol, R. K., Redder, P., Garrett, R. A. & Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea*, **2**, 59–72.
51. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*, **1**, 7.
52. Horvath, P. & Barrangou, R. (2009). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
53. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkr606.
54. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60.
55. Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P. *et al.* (2011). Evolution and classification of the CRISPR–Cas systems. *Nat. Rev., Microbiol.* **9**, 467–477.
56. Deveau, H., Garneau, J. E. & Moineau, S. (2010). CRISPR/Cas system and its role in phage–bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493.
57. Garrett, R. A., Vestergaard, G. & Shah, S. A. (2011). Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.* **19**, 549–556.
58. Karginov, F. V. & Hannon, G. J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell*, **37**, 7–19.

59. Marraffini, L. A. & Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev., Genet.* **11**, 181–190.
60. Raivio, T. (2010). Identifying your enemies—could envelope stress trigger microbial immunity? *Mol. Microbiol.* **79**, 557–561.
61. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331–338.
62. Andersson, A. F. & Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, **320**, 1047–1050.
63. Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P. *et al.* (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400.
64. Tyson, G. W. & Banfield, J. F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–207.
65. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B. *et al.* (2011). A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502.
66. Yosef, I., Goren, M. G. & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gks216.
67. Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**, e35888.
68. Koo, Y., Jung, D. K. & Bae, E. (2011). Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS One*, **7**, e33401.
69. van der Ploeg, J. R. (2009). Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology*, **155**, 1966–1976.
70. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496.
71. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure*, **19**, 257–264.
72. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P. *et al.* (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
73. Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S. & Kuramitsu, S. (2006). Crystal structure of the hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* **15**, 1494–1499.
74. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18**, 688–692.
75. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
76. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A. *et al.* (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
77. Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S. *et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell*, **45**, 292–302.
78. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L. *et al.* (2009). RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell*, **139**, 945–956.
79. Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S. *et al.* (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell*, **45**, 303–313.
80. Marraffini, L. A. & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
81. Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P. *et al.* (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67–71.
82. Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K. H., Bernardis, S., Ke, A. & Delisa, M. P. (2011). Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Mol. Microbiol.* **79**, 584–599.
83. Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. & Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512.
84. Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T. *et al.* (2010). H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393.
85. Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P. *et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536.
86. Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H. *et al.* (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286**, 21643–21656.
87. Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A. *et al.* (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA*, **108**, 10092–10097.
88. Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J., van der Oost, J. *et al.* (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.

89. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B. *et al.* (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA*, **108**, 10098–10103.
90. Horvath, P., Coute-Monvoisin, A. C., Romero, D. A., Boyaval, P., Fremaux, C. & Barrangou, R. (2009). Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* **131**, 62–70.
91. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
92. Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L. *et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.
93. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell*, **46**, 606–615.
94. Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J. V., Jensen, L. R., She, Q. & Garrett, R. A. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49.
95. Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A. & Yakunin, A. F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* **30**, 4616–4627.
96. Han, D. & Krauss, G. (2009). Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776.
97. Mulepati, S. & Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *J. Biol. Chem.* **286**, 31896–31903.
98. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342.
99. Zlatanova, J. & van Holde, K. (1998). Binding to four-way junction DNA: a common property of architectural proteins? *FASEB J.* **12**, 421–431.
100. Yosef, I., Goren, M. G., Kiro, R., Edgar, R. & Qimron, U. (2011). High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc. Natl Acad. Sci. USA*, **108**, 20136–20141.