Single-Molecule View of Small RNA-Guided Target Search and Recognition

Globyte, Viktorija; Kim, Sung Hyun; Joo, Chirlmin

DOI
10.1146/annurev-biophys-070317-032923

Publication date
2018

Document Version
Final published version

Published in
Annual Review of Biophysics

Annual Review of Biophysics

Single-Molecule View of Small RNA–Guided Target Search and Recognition

Viktorija Globyte,[1] Sung Hyun Kim,[1,2] and Chirlmin Joo[1]

[1]Kavli Institute of Nanoscience and Department of Bionanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands; email: v.globyte@tudelft.nl, kahutia@snu.ac.kr, c.joo@tudelft.nl

[2]School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea

Annu. Rev. Biophys. 2018. 47:569–93

First published as a Review in Advance on March 29, 2018

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-070317-032923

Keywords

single-molecule, target search, CRISPR, microRNA, RecA

Abstract

Most everyday processes in life involve a necessity for an entity to locate its target. On a cellular level, many proteins have to find their target to perform their function. From gene-expression regulation to DNA repair to host defense, numerous nucleic acid–interacting proteins use distinct target search mechanisms. Several proteins achieve that with the help of short RNA strands known as guides. This review focuses on single-molecule advances studying the target search and recognition mechanism of Argonaute and CRISPR (clustered regularly interspaced short palindromic repeats) systems. We discuss different steps involved in search and recognition, from the initial complex prearrangement into the target-search competent state to the final proofreading steps. We focus on target search mechanisms that range from weak interactions, to one- and three-dimensional diffusion, to conformational proofreading. We compare the mechanisms of Argonaute and CRISPR with a well-studied target search system, RecA.

Contents

1. INTRODUCTION

Target search—from how we use it for various sports and mundane everyday tasks to how predators do to locate their prey—is an essential part of the world surrounding us. In the context of a living organism, an example of the necessity to find a target is the human immune response where the well-being of a person depends on white blood cells finding and destroying their targets. In cells, a well-known example is regulation of gene expression that is accomplished through the sequence-specific binding of transcription factors to DNA activating or inhibiting transcription. Gene expression is also regulated by nucleic acid–guided target search (**Figure 1**).

1.1. Target Search

In this review, we focus on RNA-guided target search.

1.1.1. Small regulatory RNA and DNA. A revolutionary discovery of small noncoding RNAs opened up a new perspective of RNA regulation of gene expression (33). For example, microRNA (miRNA) molecules play a role in translation inhibition and subsequent degradation of messenger RNA molecules in eukaryotes. Small interfering RNA (siRNA) molecules have a similar length to miRNA, but their full complementarity with their targets leads to direct cleavage of the messenger RNA (mRNA). Both miRNA and siRNA associate with proteins belonging to the Argonaute family. These small RNA molecules guide eukaryotic Argonaute to the target site, where recognition occurs via Watson-Crick base pairing between guide and target (**Figure 1**) (**Table 1**). In prokaryotes, small regulatory RNA and DNA molecules not only can regulate gene expression but also can act as a defense mechanism against invading phage genomes and plasmids (**Table 1**). For example, some prokaryotic Argonaute proteins associate with DNA guides to find and destroy complementary DNA target sequences (108).

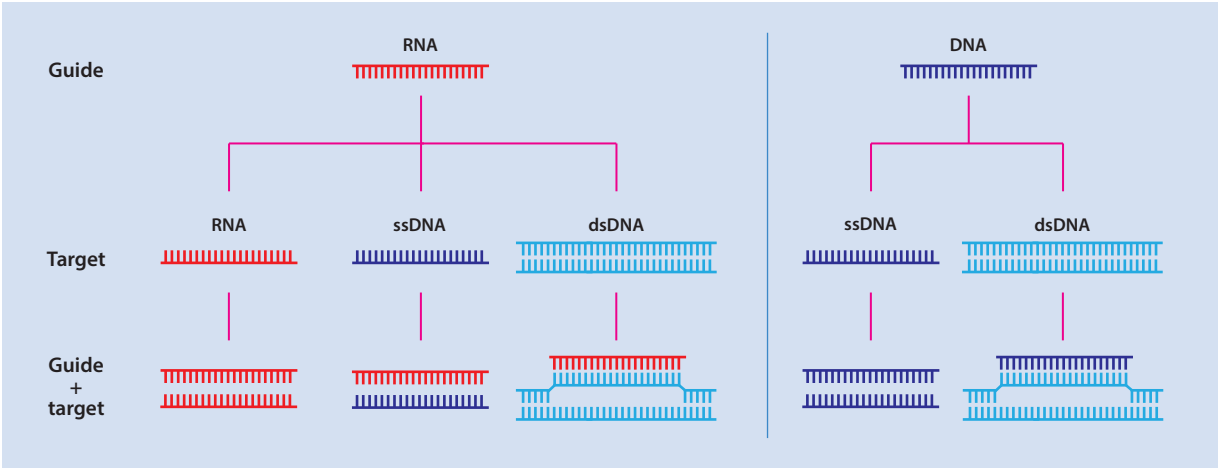


Figure 1

Schematic of different guide molecules and the nucleic acids they can target. Abbreviations: dsDNA, double-stranded DNA; ssDNA, single-stranded DNA.

Another famous example is the CRISPR (clustered regularly interspaced short palindromic repeats) adaptive prokaryotic immune system where CRISPR-associated (Cas) proteins assemble with guide RNAs to find and destroy invaders by cleaving DNA or RNA target sites complementary to the guide. CRISPR immunity consists of multiple stages. Ensuing an infection by mobile genetic elements, short fragments of the invader's DNA are integrated into the CRISPR locus in the host genome as short spacers (6, 11). This first stage of CRISPR immunity is known as the adaptation stage. During this stage, a genetic memory is created that is later used to destroy the invader upon reinfection (11). In the second stage of immunity, transcription of the CRISPR locus and further maturation of the transcript produce short CRISPR RNAs (crRNAs) (21). These crRNAs associate with Cas proteins and destroy the returning invader upon recognition of the protospacer (target) sequence in the third stage, known as the interference stage.

1.1.2. DNA recombination. The target search mechanism based on target-guide pairing is best studied in the RecA/Rad51-mediated homologous recombination process because of the mechanistic information available from two decades of biochemical, structural, and single-molecule studies (15). When double-strand breakage occurs, the damaged DNA should be aligned to its homology for strand exchange reaction to retrieve the lost genetic information (65). The broken end is first coated with RecA in bacteria and with Rad51 in eukaryotes, forming a long, rigid nucleoprotein complex that serves as a homology search and strand exchange machinery. The RecA/Rad51 filament should find the exact matching sequence by scanning the entire genome, which is a tremendous job and should therefore exploit various target search mechanisms. While the focus of this review is RNA-guided target search and recognition, comparison to the well-characterized RecA/Rad51-mediated target search process provides deeper insight into the mechanistic understanding of target search.

1.1.3. Different search modes. Target search is intrinsically a complex process that involves weak interactions and protein conformational changes. Proteins searching for their target have to diffuse through cytosol in three-dimensional (3D) fashion before encountering a DNA/RNA

Table 1 Small regulatory RNA and DNA systems

Guide	RNA					DNA		
Target	RNA		DNA		DNA (cotranscriptional)	DNA		RNA
eAgo/PIWI	Gene silencing by miRNA (68, 91) and piRNA (9, 38) RNA interference by siRNA (33)		–		Heterochromatin formation by siRNA (111) Genome rearrangement by scnRNA (80)	–		–
pAgo	–		DNA interference by diRNA (29, 59, 84)		–	DNA interference by siDNA (108)		RNA interference by siDNA (107)
CRISPR	Cas13a (C2c2)	Gene silencing by crRNA of type VI (1, 30)	Cascade/Cas9/Cas12a (Cpf1)	DNA interference by crRNA of type I (11, 21, 77), II (53, 95), and V (117)	–	–		–
Other systems	Hfq	Gene regulation by sRNA (106, 110)	–		–	RecA/Rad51	Recombination[a] (15)	–

[a]A process does not involve small DNA. This system is covered in this review to discuss previous single-molecule biophysics work on RecA-mediated target search.

Abbreviations: Cas9, CRISPR-associated protein 9; Cpf1, CRISPR-associated endonuclease in Prevotella and Francisella 1; CRISPR, clustered regularly interspaced short palindromic repeats; crRNA, CRISPR RNA; diRNA, DNA-interacting RNA; eAgo, eukaryotic Argonaute; miRNA, microRNA; pAgo, prokaryotic Argonaute; piRNA, Piwi-interacting RNA; PIWI, P-element induced wimpy testis; scnRNA, scan RNA; siDNA, small-interfering DNA; siRNA, small interfering RNA; sRNA, small regulatory RNA.

molecule that they weakly associate with, checking for a target site. Such weak interaction can lead to either quick dissociation or lateral diffusion by sliding or hopping, the latter of which in theory would speed up the target search process. However, there exists a limit beyond which lateral diffusion along the nucleic acid strand would slow down the search process and the protein dissociates. Dissociation again can lead to 3D diffusion or, in some systems, jumping, where the protein can move to another target site that is physically in close proximity but far in sequence, owing to the supercoiling or looped conformation of the nucleic acid. In addition, fast target search, involving mainly weak interactions, and specific target recognition, being mostly stable interactions, require a conformational change in the protein (14, 17, 62, 102). This review aims to give a comprehensive overview of how these different search modes and mechanisms are combined in the target search and recognition of Argonaute, CRISPR/Cas, and RecA proteins at the single-molecule level.

1.2. Single-Molecule Techniques

Single-molecule techniques used in the studies described in this review are total internal reflection fluorescence microscopy (TIRFM) and magnetic tweezers.

1.2.1. Total internal reflection fluorescence microscopy. In TIRFM, only a shallow depth (~100 nm) below the slide surface is illuminated (3). This is achieved by directing the excitation beam at an angle where it undergoes total internal reflection at the interface between the glass slide and solvent in the flow cell. This is perfect for low-light imaging, including single-molecule detection (4). In studies discussed in this review, TIRFM was used in fluorescence, Förster resonance energy transfer (FRET), and DNA curtains assays (41, 43, 44). FRET techniques offer nanometer resolution through resonance energy transfer and allow one to observe processes otherwise impossible because of the physical diffraction limit. DNA curtains involve anchoring a long fluorescently labeled DNA molecule and then stretching it by a laminar flow. DNA curtains can be used to observe long-distance movement that cannot be tracked using techniques such as single-molecule FRET (smFRET). This technique is limited by the physical diffraction limit, and processes that occur on length scales smaller than ~250 nm cannot be observed.

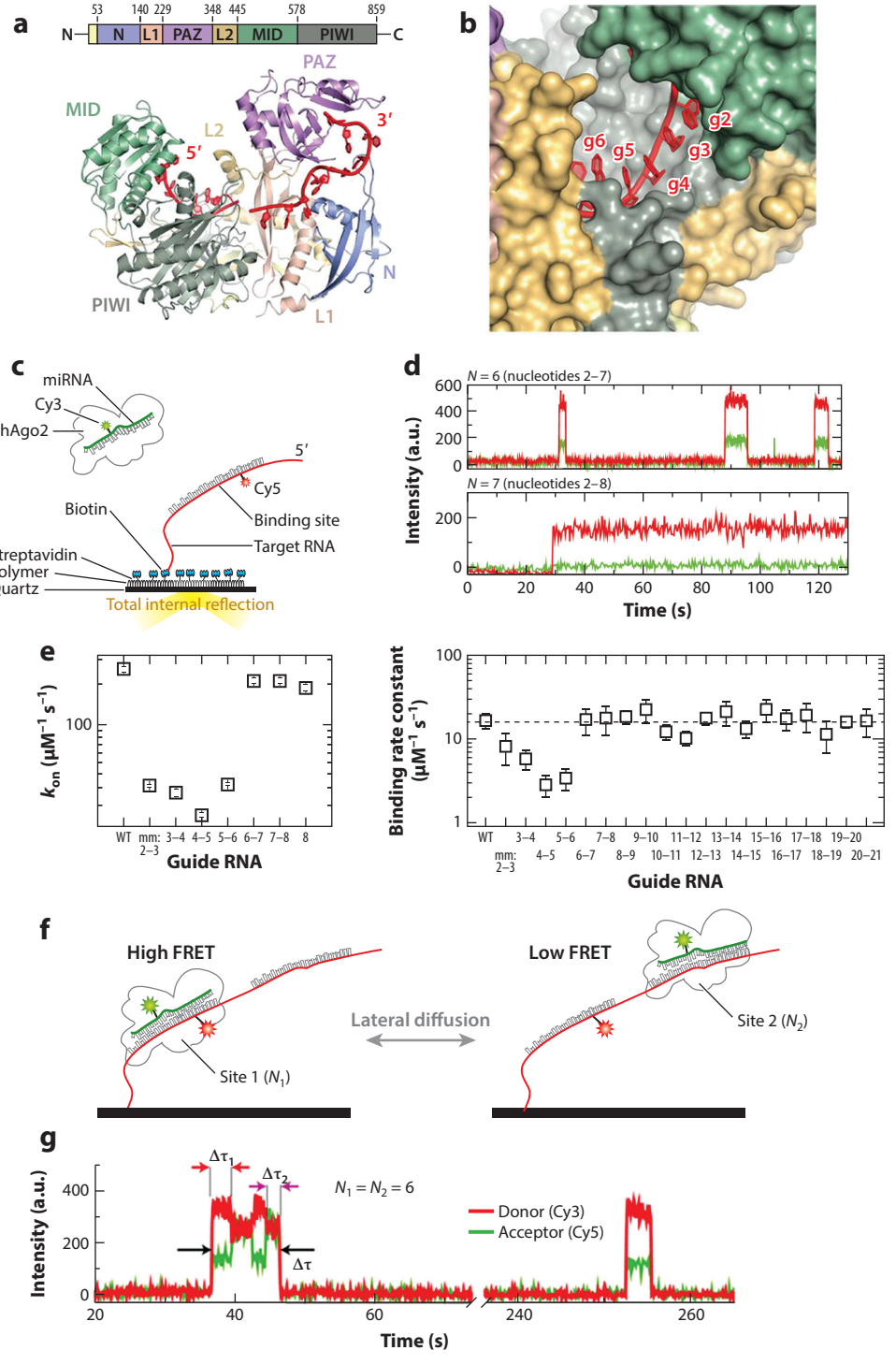
1.2.2. Magnetic tweezers. Magnetic tweezers make use of magnets and a magnetic bead that is attached to a molecule of interest (40). By trapping the bead in the magnetic field, researchers can manipulate the molecule in question, and the force and torque exerted on the molecule can be obtained by measuring the height of the magnetic bead. Magnetic tweezers assays used in studies described in this review make use of supercoiling long DNA molecules. DNA molecules are bound to magnetic beads and immobilized on the surface of the flow cells. When the magnets are turned, torsional stress is applied to the DNA molecule. At low forces, DNA supercoils forming plectonemes, which decreases the extension of the DNA molecule in a symmetric fashion for both positive and negative supercoiling. In such a setting, the position of the bead is very sensitive to even the slightest changes of the length of the DNA molecule that occur by unzipping a portion of the DNA.

2. ARGONAUTE

Argonaute proteins (see the sidebar titled Structure of Argonaute) are highly conserved in all forms of life. Eukaryotic Argonaute proteins play a central role in gene expression through processes referred to as RNA interference, whereas prokaryotic Argonautes participate in host defense via DNA interference (74, 99). In animals, Argonaute proteins loaded with miRNA as guides bind to the 3′ UTR of mRNA and prevent the production of proteins via several pathways that usually involve destabilization of mRNAs, and they require partial complementarity between guiding miRNA and target mRNA (2, 12). Some Argonaute proteins [e.g., human Argonaute 2 (hAGO2)] are able to degrade mRNA without recruiting additional factors (79). This endonucleolytic activity is mediated by siRNAs and requires full complementarity between the target and the guide.

STRUCTURE OF ARGONAUTE

Argonaute proteins have a bilobed architecture with four domains: MID (middle), PIWI (P-element induced wimpy testis), PAZ (PIWI/Argonaute/Zwille), and N-terminal domains (**Figure 2*a***). These domains are highly conserved between eukaryotic and prokaryotic proteins. The MID domain interacts with the 5′ phosphate of the guide (19, 35). The PIWI domain contains an RNase H-like active site and catalyzes the slicing activity (85, 86, 103). The PAZ domain binds the 3′ end of the guide. This interaction protects the guide from being degraded, especially in eukaryotic Argonautes (71, 72). The N-terminal domain is important for target cleavage and the dissociation of the cleavage products (45, 66).



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Target search mechanism of eukaryotic Argonaute. (*a*) Crystal structure of human Argonaute with guide RNA (97). (*b*) Argonaute exposes the first four nucleotides of the seed to the solvent (23). (*c*) Schematic of single-molecule hAGO2 FRET assay (23). (*d*) Representative time traces showing hAGO2 binding to the target with different seed-target complementarity (23). (*e*) Binding rate of mouse (94) and human Argonaute (55) proteins with dinucleotide mismatched guide RNAs. (*f*) Schematic of hAGO2 tandem target FRET assay (23). N_1 and N_2 represent the number of complementary nucleotides at the first and second binding sites, respectively. (*g*) Representative time trace showing hAGO2 transitioning between two FRET states (23). Abbreviations: Cy, Cyanine; FRET, Förster resonance energy transfer; hAGO2, human Argonaute 2; k_{on}, on-rate; L, linker domain; MID, middle domain; miRNA, microRNA; mm, mismatch; nt, nucleotide; PAZ, PIWI/Argonaute/Zwille domain; PIWI, P-element induced wimpy testis domain; τ, dwell time; WT, wild type.

In this section, we focus on animal Argonaute proteins, in particular human and mouse Argonaute, and their target search mechanisms.

2.1. Seed Recognition

Biochemical and bioinformatics studies showed that human Argonaute proteins divide their guide molecules into five distinct domains: 5′ anchor (first nucleotide), seed region (nucleotides 2–8), central region (nucleotides 9–12), the 3′ supplementary region (nucleotides 13–16), and the 3′ tail (nucleotides 17–22) (5, 13, 60, 112). The seed region plays the key role in target search. Structures of human AGO2 have shown that the seed nucleotides 2–6 in the guide are preordered in an A-form helix and exposed to the solvent (96) (**Figure 2*b***). Such preordering helps Argonaute overcome the entropic cost of target base pairing. A single-molecule fluorescence study on mouse AGO2 has shown that Argonaute increases the rate with which RNA binds to its target to levels limited by diffusion, compared to naked guides binding their targets alone, confirming that preordering facilitates seed recognition (94).

The crystal structure of hAGO2 bound to a guide and target proposed a stepwise mechanism for target binding: Pairing of nucleotides 2–5 with the target promotes a conformational change exposing nucleotides 2–8 and 13–16 for target base pairing (97). An smFRET study of hAGO2 has shown that the first three nucleotides in the seed are the most important in determining the binding rate (**Figure 2*c***). The number of consecutive complementary nucleotides does not affect binding rate at all as long as there are no mismatches in the seed. However, the stability of binding is determined by the degree of seed complementarity (23). By varying the number of consecutive complementary nucleotides between guide and target, Chandradoss et al. (23) have shown that binding to the first six nucleotides has a small effect on the dwell time of hAGO2 on the target RNA. Increasing complementarity past seven consecutive nucleotides had a drastic effect with binding events lasting throughout the whole measurement (∼300 s) (**Figure 2*d***). Therefore, hAGO2 uses the first seed nucleotides that are exposed to the solvent to probe potential target sites, and the rest of the seed further stabilizes binding, confirming the stepwise model proposed by previous biochemical and bioinformatics studies.

Other single-molecule fluorescence studies have also shown that seed region is important in achieving fast target search. A study on mouse AGO2 explored the effects of dinucleotide mismatches along the seed region (94). It was shown that the lack of guide and target complementarity within the first six seed nucleotides decreases the binding rate dramatically. A similar result was obtained by an smFRET study on hAGO2 that explored the effect of dinucleotide mismatches along the full guide (55). In particular, both studies have found that mismatches in the middle and 3′ end of the seed reduced binding rate more than mismatches on the 5′ end of the seed (**Figure 2*e***). These

findings are in contrast with Chandradoss et al. (23), who showed that the first three nucleotides in the seed are the most important. This disparity might arise from the fact that, in the two studies described above, the remainder of the guide was fully complementary to the target and could possibly compensate for the mismatches at the beginning of the seed. Despite different findings, all three studies show that the seed region of the guide RNA is crucial for binding rate and stability.

2.2. Lateral Diffusion

Argonaute proteins have to find their target in a large pool of cellular RNAs. Furthermore, miRNA target sites are often found in the 3′ UTR of mRNA, which can be several kilobases long. It is possible that Argonaute target search is facilitated by lateral diffusion as was first hinted at by biochemical studies (5). To test this hypothesis, Chandradoss et al. (23) designed a tandem target assay where two identical target sites were separated by 22 nucleotides on a single RNA strand (**Figure 2*f***). Binding to one side would yield a high FRET value, and binding to the second target site would show low FRET. Using a target of six-nucleotide complementarity with the guide, it was observed that over 70% of binding events showed rapid FRET changes, suggesting that Argonaute is shuttling between the target sites (**Figure 2*g***). It still remains to be determined whether the observed FRET changes are due to sliding (the protein contains constant contact with the RNA), hopping (multiple association and dissociation events are correlated along the contour of the target RNA), or jumping (the protein can jump to another binding site that is physically close but far in sequence).

2.3. Conformational Change

Full pairing of Argonaute guide and target requires a protein conformational change. Eukaryotic and archaeal Argonautes introduce a kink in their guide by an alpha helix, termed helix 7, which would have to shift to accommodate pairing between nucleotides 6 and 7 (96). Indeed, pairing of the first five nucleotides relieves helix 7 and allows further nucleotide pairing (63, 97). However, for further base pairing, the central cleft, where guide and target molecules are accommodated, has to widen, as narrowing of the central cleft restricts pairing past guide nucleotide 8. The conformational changes after the turn of the helix 7 may not be important for silencing activity, but the ensuing movement of the PAZ domain and opening of the channel between PAZ and N-terminal domains are necessary to accommodate full guide-target base pairing for RNA slicing. This conformational proofreading mechanism is further discussed in Section 5.

2.4. Cooperativity

The possibility that neighboring target sites can act cooperatively to retain the AGO-RNA complex on the target RNA has been previously hinted at by biochemical studies, but the mechanism was not clear (20, 42, 93). A possible explanation is provided by the observation of lateral diffusion. The smFRET study using tandem target assay shows a drastic difference between the dwell times on a single target and tandem target with residence times on the tandem target constructs being nearly 10 times higher than those on a single target (23). This result confirmed that the neighboring target sites act synergistically to retain AGO-miRNA on the target strand. Remaining bound to the target RNA for longer could decrease the energetic cost that comes with protein having to change conformation upon associating and dissociating with the target multiple times. In addition, increased dwell time gives more time to recruit other proteins necessary for degradation of mRNA.

3. CRISPR IMMUNITY

Bacteria and archaea use the RNA-mediated adaptive CRISPR/Cas immune system to defend against invading bacteriophages and plasmids (11, 76). Different organisms have evolved distinct CRISPR systems. These systems are grouped into two main classes that are subdivided into six main types (73, 81). The signature of Class 1 CRISPR systems (types I, III, and IV) is the use of a multi-subunit protein complex for target recognition and degradation, whereas Class 2 systems (types II, V, and VI) use a single protein for this task. This section of the review focuses on the two best understood systems: type I [CRISPR/Cascade (CRISPR-associated complex for antiviral defense)] and type II (CRISPR/Cas9).

3.1. Cas9 Protein

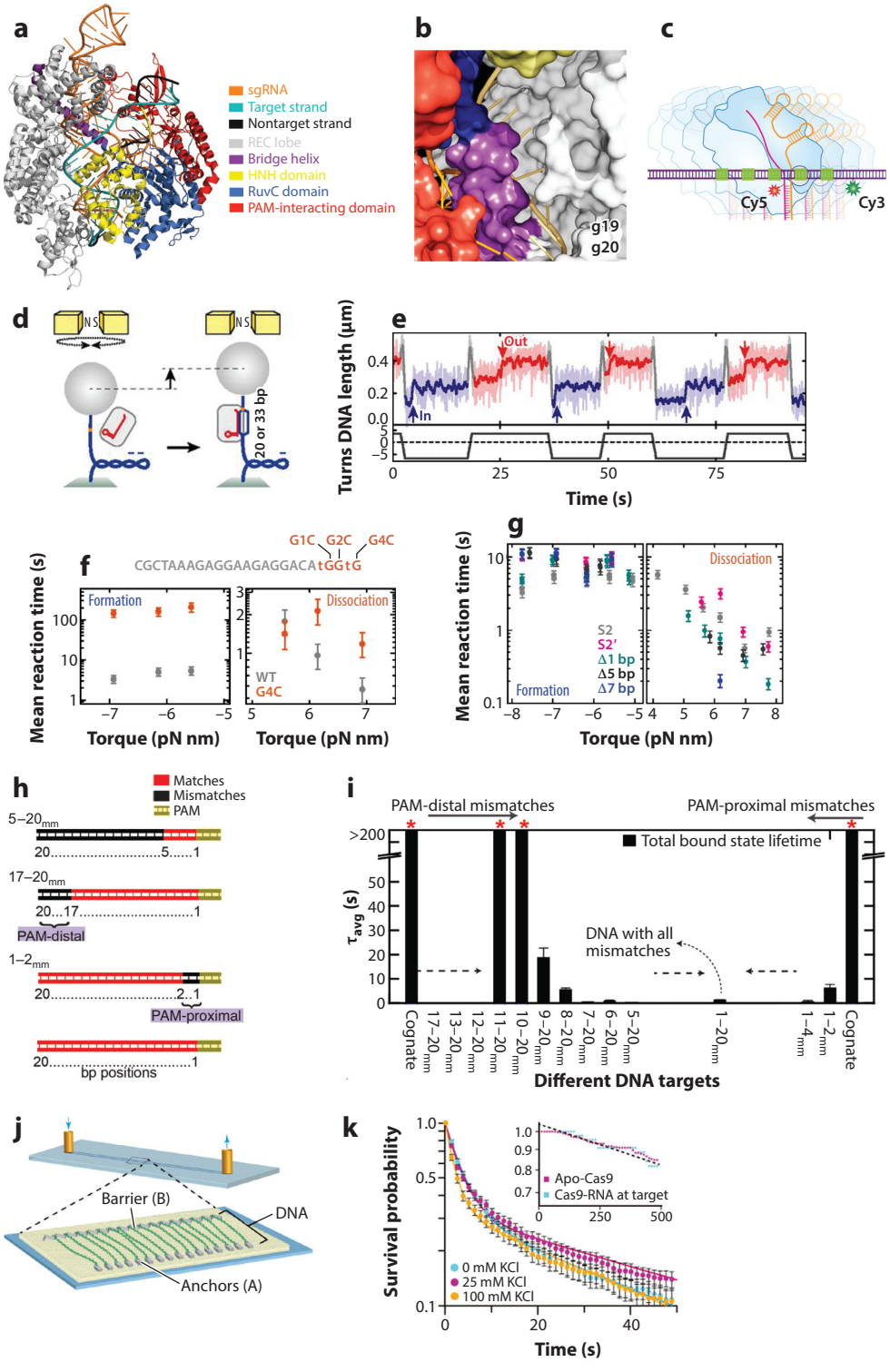
For six years now, type II *Streptococcus pyogenes* Cas9 (SpyCas9) protein has been at the center of attention for genome engineering purposes, owing to its simplicity and programmability (10, 52). This protein is a large endonuclease, consisting of 1,368 amino acids and multiple domains (**Figure 3*a***). SpyCas9 recognizes a 3-nucleotide PAM (protospacer adjacent motif) adjunct to the 3′ end of the 20-nucleotide target sequence and cleaves the target three base pairs downstream from the PAM (see the sidebar titled PAM Recognition). Unlike most other CRISPR systems, SpyCas9 needs two RNA molecules, namely crRNA and *trans*-activating RNA (tracrRNA), to find and destroy the target (37, 53). For genome editing purposes, the two RNA molecules can be fused into one single-guide RNA that maintains full functionality of the effector complex (28, 48). Binding to RNA is crucial for Cas9 targeting, as it enables structural rearrangements necessary to accommodate a DNA target and contains the guide sequence, which is complementary to the target (see the sidebar titled Cas9 Preorganization and Structural Rearrangement for Target Search).

3.1.1. PAM search. PAM recognition is the first step in Cas9 target search and is an intrinsically complex protein-DNA interaction. Binding to the canonical PAM triggers local melting of the DNA at the PAM-adjacent nucleation site (7). This is followed by the directional formation of the RNA-DNA hybrid and the displacement of the nontarget strand (R-loop formation) (50, 92, 109).

Single nucleotide mutations in the PAM are able to slow down or abolish binding and R-loop formation, as shown by magnetic tweezers DNA-supercoiling assays (109). In these experiments,

PAM RECOGNITION

CRISPR systems target specific sequences using Watson-Crick base pairing between guide RNA and target DNA to recognize and cleave the target (36). In addition to the target sequence complementary to the guide RNA, specific Cas proteins involved in DNA interference recognize a PAM (protospacer adjacent motif) sequence as the first step of target search (53, 82). Although different in sequence and placement between different CRISPR systems, a PAM sequence is always present adjacent to the target site. The main role of PAM is to act as an indicator for self–nonself discrimination: The spacer sequences integrated in the host genome are identical to those in the invading DNA; hence the host could recognize and cleave its own DNA, which would be fatal to the cell (78). In contrast, the protospacer sequences in the invader's genome are always flanked by a PAM sequence, which is not integrated in the CRISPR locus. Therefore, upon recognizing the PAM sequence as the first step prior to recognizing the target via Watson-Crick base pairing, the host ensures that the invader is destroyed and the integrity of its own genome remains protected.



(Caption appears on following page)

Figure 3 (*Figure appears on preceding page*)

Target search mechanism of Cas9 protein. (*a*) Crystal structure of SpyCas9 with single-guide RNA and target DNA. (*b*) SpyCas9 preorders the first 10 guide nucleotides in a helical configuration and exposes nucleotides 19–20 to the solvent. (*c*) Schematic of Cas9 smFRET multiple-PAM assay (39). (*d*) Schematic of magnetic tweezers assay (109). (*e*) Time trajectory of the DNA length. Cas9 binds at negative supercoiling, thus increasing DNA length, and dissociates at positive supercoiling also increasing DNA length (109). (*f*) Mean reaction times for R-loop formation and dissociation as a function of torque for WT protospacer and a PAM (G4C) mutant (109). (*g*) Mean reaction times for R-loop formation and dissociation as a function of torque for different protospacer truncations (109). (*h*) Protospacer mutation scheme for smFRET assay (100). (*i*) Dwell times for DNA targets with different lengths and positions of mutations (100). (*j*) Schematic of DNA curtains assay (105). (*k*) Survival probabilities for off-target binding events are represented by a double-exponential decay (105). Abbreviations: bp, base pair; Cas9, CRISPR-associated protein 9; Cy, Cyanine; mm, mismatch; KCl, potassium chloride; N, north magnetic pole; PAM, protospacer adjacent motif; REC, Recognition; S, south magnetic pole; S2, spacer 30 nucleotides long; S2′, spacer 20 nucleotides long; sgRNA, single-guide RNA; smFRET, single-molecule Förster resonance energy transfer; SpyCas9, *Streptococcus pyogenes* Cas9; τ, dwell time; WT, wild type.

performed with *Streptococcus thermophilus* Cas9 (StCas9), DNA was negatively supercoiled to assist R-loop formation (**Figure 3*d*,*e***). When PAM was mutated four nucleotides away from the seed, R-loop formation was still observed but at a much lower rate (**Figure 3*f***). PAM mutations in the positions closer to the seed slowed down R-loop formations even more. Therefore, mutations in the PAM alter R-loop formation by kinetic instability, which renders Cas9 unable to recognize the target and start R-loop formation.

DNA curtains experiments (**Figure 3*j***) showed that PAM recognition involves intrinsically weak interactions (105). While Cas9 remains stably bound to a bona fide target site, only short-lived interactions are observed with off-targets. The off-target binding distribution correlates with the PAM distribution on the lambda DNA, consistent with other studies, showing that

Cas9 PREORGANIZATION AND STRUCTURAL REARRANGEMENT FOR TARGET SEARCH

Apo-Cas9

Apo-Cas9 has a bilobed architecture with one lobe (nuclease lobe) containing the HNH, RuvC, and C-terminal domains and the other (recognition lobe) containing a large helical domain (54). Apo-Cas9 is able to bind DNA; however, it displays no sequence specificity, as shown by DNA curtains assays (105). The nonspecific DNA binding shows strikingly long lifetimes. However, in the presence of heparin or guide RNA as competitors, Apo-Cas9 quickly dissociates from the DNA strand.

Structural Arrangement for PAM and Seed Recognition

TracrRNA activates the Cas9-RNA complex. An important rearrangement upon binding crRNA and tracrRNA occurs in the C-terminal domain, also known as the PAM-interacting domain, which then forms a groove that can accommodate the PAM sequence in its DNA duplex form. Binding to guide RNA therefore enables Cas9 to look for PAM sites in a sequence-specific manner (7). Similar to Argonaute (Ago) systems, the first 10 nucleotides in the seed region of the crRNA are preordered in an A-form helix, with the first nucleotides exposed to the solvent for initial DNA interrogation (**Figure 3*b***). In addition, a kink is introduced into the guide RNA by an insertion of an amino acid (Tyr) between nucleotides 15 and 16, which is relieved upon target binding (51, 54).

Cas9 samples PAM sites before it finds and stably associates with its target site. The short-lived interactions are characterized by a double-exponential decay, indicating that Cas9 has at least two distinct modes when searching for PAM (**Figure 3*k***).

An insight into different binding modes is provided by smFRET experiments that can probe local interactions around PAM sites (39) (**Figure 3*d***). If there are neither PAM nor target sites present on the DNA, Cas9 binding is random and short lived (<0.5 s). However, if there is at least a single PAM site present, Cas9 exhibits two distinct types of behavior: short transient binding to a PAM site (<0.5 s) and more stable binding (~2 s). This implies that, upon binding a PAM site, Cas9 can either dissociate quickly upon failing to form an RNA-DNA duplex or diffuse locally around the PAM, looking for adjacent PAM sites and trying to form an R-loop there. Therefore, Cas9 might use a combination of short-lived 3D diffusion and long-lived 1D diffusion for PAM search.

In vivo, Cas9 has been found to spend approximately a subsecond on PAM sites (64, 56). However, some Cas9 molecules stay stably bound for longer than 5 s, despite the fact that there is no target present, indirectly suggesting that Cas9 may be searching for adjacent PAM sites flanked by a cognate target sequence, potentially using lateral diffusion. Biochemical data have revealed another layer of complexity, showing that Cas9 is able to bind DNA substrates with no target but multiple PAM sites in electrophoretic mobility shift assays (105). This peculiar observation may be explained by local diffusion on the target strand that creates a synergistic effect between neighboring PAM sites.

3.1.2. Seed recognition. The next step after recognizing the correct PAM is the recognition of a seed sequence on the target DNA. This sequence is the first 8–12 nucleotides downstream from the PAM. Recognition of the seed via Watson-Crick base pairing between guide RNA and target DNA is crucial for stable binding. A magnetic tweezers study investigated the effect of protospacer truncations on the stability of the RNA-DNA R-loop using StCas9 (109). For 1– or 5–base pair (bp) truncations from the PAM-distal end, R-loop stability was slightly reduced. R-loops were detected for truncations up to 7 bp with little change in the association rate but were not detected for 9-bp truncations (**Figure 3*g***). R-loops of 11 bp or shorter were not formed, revealing directional R-loop formation.

Additional evidence for the directionality of R-loop formation by Cas9 has been shown by an smFRET study (70). Here, immobilized DNA is labeled with a donor dye at the PAM-distal end and RNA is labeled at the 5′ end with an acceptor dye. Upon full complementarity, a high FRET state was observed. However, it was also possible to capture an intermediate FRET state that corresponds to only PAM-proximal base pairing before a high FRET state was reached. These findings show substeps in guide-target pairing and confirm the directionality of R-loop formation between RNA and DNA.

Another smFRET study has explored the effects that both PAM-proximal and PAM-distal mismatches have on Cas9 protospacer binding (100) (**Figure 3*b***). Even 2-bp PAM proximal mismatches are able to severely decrease binding, and 4-bp PAM-proximal mismatches decrease binding to the levels of fully mismatched targets. In contrast, PAM-distal mismatches are much better tolerated, with up to 12 mismatches showing the binding stability as the cognate target (**Figure 3*i***). Together, these results underscore the importance of the seed region, which in this study has been shown to be eight nucleotides. This shows that early mismatched regions are able to stop R-loop formation and abolish binding regardless of the sequence downstream of the mismatched region.

3.1.3. Final stages of target recognition. Initial target search of Cas9 is a complex process involving multiple binding modes. Although finding the seed is enough for stable binding, Cas9 cleavage requires more stringent Watson-Crick base pairing. If a target is extensive enough for stable binding but not extensive enough for cleavage, Cas9 undergoes dynamic conformational changes. Single-molecule and bulk FRET experiments where the two nuclease domains, HNH and RuvC, were labeled have shown that the movement of the HNH domain from the PAM-distal end to the cleavage site is possible only when the complementarity between guide RNA and target DNA is no less than 18 nucleotides (25, 104). Four PAM-distal mismatches are enough to stop Cas9 from reaching the final conformation state, leaving Cas9 transitioning between the initial and intermediate states.

3.2. Cascade Protein Complex

A majority of CRISPR systems found in nature belong to the type I CRISPR family. These types of CRISPR systems use a multi-subunit protein complex, Cascade, for the recognition of invading foreign DNA (21). Cascade alone is unable to degrade target DNA and instead relies on the recruitment of a Cas3 endonuclease (83, 101). Cascade complexes play a role not only in interference but also in a process called priming during which CRISPR memory is rapidly updated to fight escape mutants (32).

3.2.1. Structural arrangements enable target search. *Escherichia coli* Cascade consists of 11 Cas proteins (one copy of Cse1, two copies of Cse2, six copies of Cas7, one copy of Cas5e, and one copy of Cas6e) and adopts a sea horse–like structure (49, 118) (**Figure 4*a*,*b***). Its crRNA has a less intricate architecture than that of Cas9, consisting only of a 3′-stem loop and a protospacer region. Similar to Cas9, crRNA in Cascade is preordered in a pseudo A-form helical configuration. The spacer sequence is divided into segments by the flipping out of every sixth nucleotide. The nucleotides within the first two segments are crucial to target binding and are defined as the seed (98). Mismatches in the other segments are much better tolerated and can still lead to successful interference.

3.2.2. PAM search. Cascade PAM recognition is more promiscuous with at least five interfering PAM sequences identified for *E. coli* Cascade (47, 83, 113). The PAM is recognized by Cse1 in double-stranded form, from the minor-groove site. Such mode of the minor-groove recognition indicates that mutated PAMs can be tolerated as long as target sequence is optimal (46). This is supported by a DNA curtains study where Cascade was able to bind a fully matching protospacer that was lacking a PAM but with a much reduced binding rate (90) (**Figure 4*c***).

3.2.3. Directional R-loop formation. Similar to Cas9, upon binding a PAM site, Cascade interrogates the target DNA and forms an RNA-DNA R-loop in a directional fashion as shown by magnetic tweezers experiments (92). Point mutations in the seed region required higher negative supercoiling for R-loop formation than PAM-distal mutations. This agrees with the CHAMP (chip-hybridized association-mapping platform) assay, where it was found that the farther the point mutations are from the PAM, the more tolerated they are (57). In addition, in magnetic tweezers experiments, Cascade stalling at mutations was observed, which needed appropriate supercoiling (approximately fourfold higher for mutations close to PAM compared to wild type) to overcome the intermediate state for full R-loop formation (**Figure 4*d***). Reversible binding

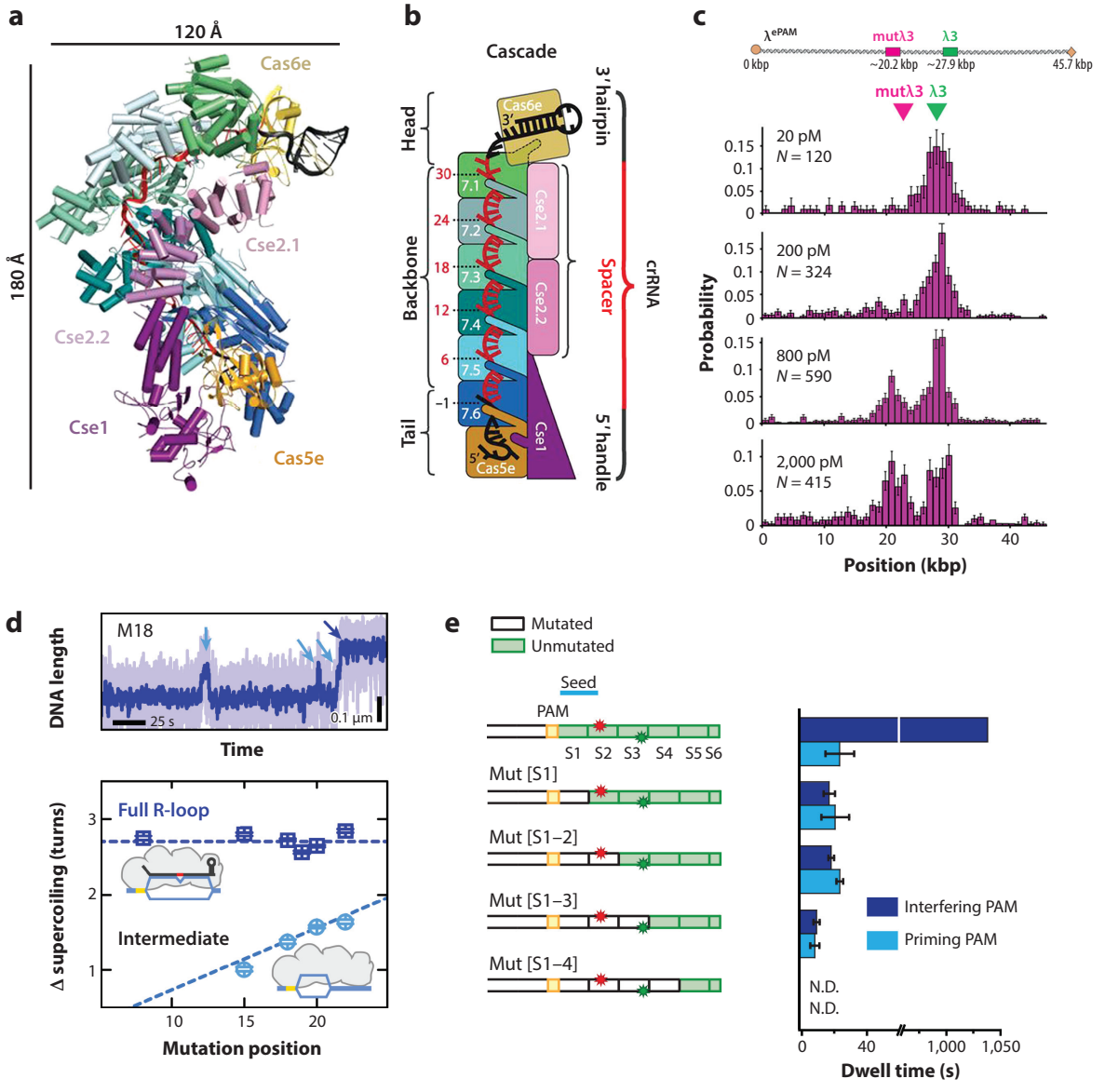


Figure 4

Target search of Cascade protein complex. (*a*) Crystal structure of type I-E *Escherichia coli* Cascade with guide RNA (49). (*b*) Schematic of Cascade (49). (*c*) Histogram showing binding probability to cognate target and a protospacer lacking a PAM at different Cascade concentrations in DNA curtains experiments (90). (*d*) R-loop formation trajectory protospacer with a point mutation at position 18 and mean supercoiling changes associated with full and intermediate R-loop formation (92). (*e*) Schematic of protospacer mutations for smFRET assay and dwell times for each construct (18). Abbreviations: Cascade, CRISPR-associated complex for antiviral defense; crRNA, CRISPR RNA; ePAM, escape protospacer adjacent motif; kbp, kilobase pair; λ, lambda phage DNA; Mut, Mutation; N.D., no data; PAM, protospacer adjacent motif; S, segment; smFRET; single-molecule Förster resonance energy transfer.

events were also observed with DNA curtains, suggesting Cascade might make multiple attempts before it stably engages with the protospacer.

3.2.4. Nondirectional binding. Despite magnetic tweezers assays suggesting only directional R-loop formation, an smFRET study has shown that Cascade is able to bind DNA in a sequence-specific but nondirectional manner (18). Targets with a cognate PAM and fully complementary DNA sequence exhibited two types of binding events—long events characterized by an initial high FRET that soon transitioned to a low FRET state, and short-lived events exhibiting low or mid FRET state, corresponding to partially unwound DNA. Targets with a mutated seed region showed only the second type of events (**Figure 4*e***). This is in contrast to Cas9, where seed mutations completely abolish target binding. It was confirmed in vivo that the constructs that exhibited noncanonical binding modes triggered the priming response, which allows CRISPR memory to be rapidly updated.

3.2.5. Cas3 recruitment. Cascade does not degrade the target itself but rather recruits the Cas3 nuclease (101). In the magnetic tweezers experiments, it was shown that R-loop locking is required for the recruitment of Cas3, regardless of any mutations in the protospacer. However, mutations of the PAM significantly affected Cas3 cleavage, even if the R-loop was fully formed and was in its locked state, implying a dual signaling mechanism upon target recognition. Consistent with these findings, DNA curtains assay has shown that Cascade bound to a target flanked by a PAM could readily recruit Cas3 nuclease for DNA degradation. However, at PAM-lacking sites, Cascade could not directly recruit Cas3. Finally, the CHAMP assay also suggested that Cas3 is recruited in a DNA sequence–dependent manner.

4. RecA-MEDIATED DNA RECOMBINATION

Cellular DNA experiences damages from harsh environments. The double-strand break (DSB) is a lethal DNA damage. A DSB site is recognized by protein complexes such as RecBCD, RecFOR, and BRCA2, which process the damaged DNA to make a long single-stranded overhang (**Figure 5*a***) (27). RecA/Rad51 proteins are loaded onto the single-stranded overhang, forming a several-kilobase-long nucleoprotein filament (8). The single-stranded DNA (ssDNA) within the filament is linearly stretched such that the Watson-Crick pairing sides of the bases are exposed to the solvent. This structural arrangement is utilized for homology search and strand exchange reaction (**Figure 5*b***) (24). All these steps need to be processed quickly and accurately against the genome consisting of millions to billions of bases within the time window of DNA replication in the cell cycle (typically 1–2 h) (69). In this section, single-molecule studies on the homology search and target recognition mediated by RecA/Rad51 are covered.

4.1. Structural Platform for Homology Search

RecA forms a helical filament on ssDNA (**Figure 5*b***) (24, 31). X-ray crystallography revealed that the bases within a filament are grouped into three bases (designated as a triplet), each occupied by a single RecA monomer (24). The triplet maintains the B-form-like structure. Binding to a homologous sequence triggers slight structural rearrangement of the triplet, which signals RecA to check for correct pairing. Growing evidence supports the hypothesis that the triplet is the fundamental unit of base pair recognition (67, 87) as well as of strand exchange reaction (88). It has further been shown that RecA controls the way grouping bases precisely locate the recombination hot spot, Chi. (61).

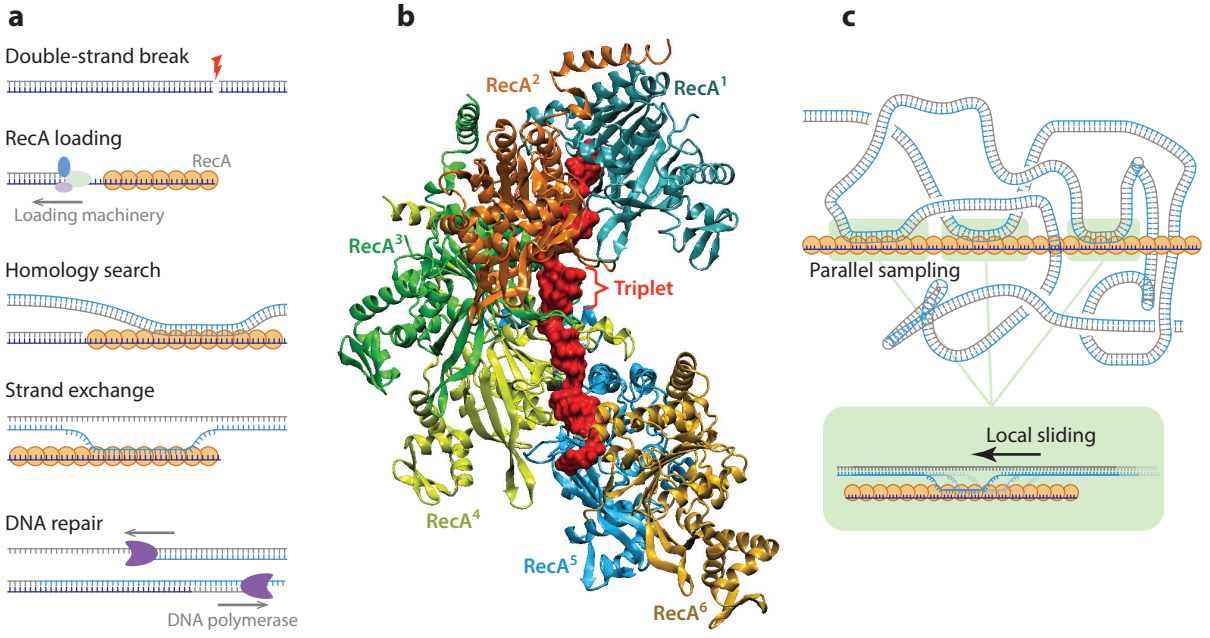


Figure 5

Homology searching mechanism in RecA-mediated recombination DNA repair. (*a*) Schematic of RecA-mediated homology search and strand exchange reaction process. (*b*) Crystal structure of RecA filament (each monomer in different color) with embedded single-stranded DNA (*red*). Six RecA monomers that make one helical turn within the filament are shown. (*c*) One-dimensional and three-dimensional search mechanisms of RecA filament. Multiple samplings of homology take place at different regions of the filament via intersegmental transfer (highlighted with *green boxes*). Each sampling contact may locally slide along the double-stranded DNA to facilitate the search process (*large green box*).

4.2. One-Dimensional Versus Three-Dimensional Search Mechanisms

Advances in single-molecule methods enabled real-time observation of homology search, pairing, and the strand exchange process (**Figure 5*c***) (26, 34, 67, 75, 87–89). Direct observation of RecA filament sliding along double-stranded DNA (dsDNA) has been reported by Ragunathan et al. (89). By using smFRET, the authors found that the RecA filament initially binds to the dsDNA at random positions and slides along the dsDNA for homology search. The diffusion coefficient of this 1D sliding was 8,000 bp^2/s, allowing two-orders-of-magnitude-faster homology search compared to a simple 3D diffusion model. They estimated sliding could occur up to several hundred bases before dissociation. A follow-up work of DNA curtains experiments by Qi et al. (87) indicated the lack of long-distance lateral diffusion and thus suggests that the lateral diffusion distance is shorter than the diffraction limit (~250 nucleotides).

In the same year, experimental evidence of facilitated 3D search was reported (34). By using optical tweezers combined with fluorescence imaging within a microfluidic device, Forget & Kowalczykowski (34) have found that the search and pairing process slows down when the dsDNA is linearly stretched, implying that 3D conformation of the dsDNA target is crucial in the homology search. Moreover, transient loop formation of the dsDNA at the off-target site has also been observed in support of the intersegmental transfer, a process that a filament transposes via simultaneous binding to more than two different regions of the dsDNA. This intersegmental transfer allows parallel homology sampling at multiple regions of the dsDNA (16).

4.3. R-Loop Stability

RecA/Rad51 recombinases rely on Watson-Crick base pairing to find homology. By measuring the kinetic lifetimes of a transiently bound homology search intermediate, Qi and colleagues (67, 87) have found that microhomologies shorter than eight nucleotides cannot make a stable R-loop for further homology recognition. In contrast, Ragunathan et al. (89) have observed transient homology pairing as short as six nucleotides. The disagreement in numbers may be due to the location of the homology sequence; Ragunathan et al. used a dsDNA with homology at one end of the dsDNA, whereas Qi et al. placed a homology sequence in the middle of the dsDNA. Nevertheless, these studies show that RecA uses a length-based discrimination mechanism to find homology.

5. INTEGRATED VIEW ON TARGET RECOGNITION

Target search and recognition consist of different search modes and stages. This section provides an integrated view of target search and recognition processes.

5.1. Modes of Target Search

Target search is an essential part of the functioning of many different proteins. Despite differences in function, any target search and recognition should be both rapid and specific. The optimum way to achieve this is to use a combination of 3D diffusion and 1D diffusion while minimizing time spent on off-targets. See **Table 2** for a summary.

Human AGO2 of the Argonaute protein family achieves this by exposing only the first few nucleotides of the seed region and using them to probe potential target sequences. An smFRET study has shown that exposing the first three nucleotides facilitates target search by lateral diffusion in which the hAGO2-miRNA does not dissociate from the RNA strand even if a fully matching target is not found but diffuses laterally to a neighboring target site (23). Adjacent target sites

Table 2 Single-molecule studies of Argonaute, Cas9, Cascade, and RecA

Searcher protein	Motif for initial recognition	Requirements for stable binding	Requirements for competent state	Conformational change upon (. . .)	Facilitated target search
AGO	First 3 nt of seed (23, 94)	Seed pairing (7 nt) (23)	Full-length guide pairing (~22 nt) (55, 115)	Seed recognition (63) Full-length target recognition (58, 116)	Lateral diffusion (>10 nt) (23)
Cas9	PAM (3 nt) (39, 105)	PAM (100, 105, 109) Seed pairing (8 nt) (100) Seed pairing (13 nt) (109)	PAM (70) Near-full-length guide pairing (18–20 nt) (70)	Seed recognition (25) Full-length target recognition (25)	Lateral diffusion (>10 nt) (39)
Cascade	PAM (3 nt) (90)	Near-full-length guide pairing (28–32 nt) (92, 109)	PAM (90, 92, 109) Near-full-length guide (28–32 nt) (92, 109)	Near-full-length target recognition (18, 114[a])	Lateral diffusion (>100 bp) (22)
RecA	5–7 nt (89)	8-nt pairing (67, 87)	15-nt pairing (88)	Not reported	Lateral diffusion (>10 bp) (89)

[a]A bulk, not single-molecule, Förster resonance energy transfer study.

Abbreviations: AGO, Argonaute; bp, base pairs; Cas9, CRISPR-associated protein 9; Cascade, CRISPR-associated complex for antiviral defense; nt, nucleotide(s); PAM, protospacer adjacent motif.

also act synergistically to keep the protein from dissociating, thereby increasing the probability of finding a cognate target nearby.

Similar to Argonaute, Cas9 uses a mixture of 3D and lateral diffusion to find its target. First, the protein weakly associates with a PAM site, interrogating the adjacent sequence for complementarity, as shown in DNA curtains experiments (105). An smFRET study has additionally shown that Cas9 has two binding modes, one being a specific PAM and guide-target-mediated interaction and another, termed sampling mode, being a search mode that does not involve RNA-DNA base pairing and is likely protein sampling the DNA for a PAM site—with one of the ways possibly being lateral diffusion (100). Another smFRET study has directly shown that Cas9 is able to transition without dissociation between adjacent PAMs or PAMs with partial target sites via lateral diffusion (39). The CRISPR type I Cascade protein complex has been shown to use 3D diffusion to find its target (90). Recently, it has also been shown that the Cascade protein complex from a thermophilic organism (*Thermobifida fusca*) also uses 1D diffusion during its target search (22).

RecA also takes advantage of the synergetic acceleration in the target search process by combining both 1D and 3D search mechanisms (**Figure 5*c***). Similar to the Argonaute and CRISPR systems, RecA filament slides locally along dsDNA (89). During this short-range 1D search, RecA recognizes only the homologous sequences longer than a certain length (6–8 nucleotides) (87, 89). Ignoring short matches effectively decreases the search complexity because the number of the matching sequences in the genome decreases exponentially with longer pairing size (67, 87). Argonaute shares this pairing size–based discrimination to reduce the search complexity (23), while PAM recognition of Cascade/Cas9 proteins is another effective way to reject large numbers of off-targets from endogenous sequences. In 3D search mechanisms, a notable difference of RecA compared to Argonaute and CRISPR systems is the length differences of the guide DNA/RNA: several kilobases for RecA versus ~20–30 nucleotides for Argonaute and CRISPR. Extremely longer length of the RecA filament enables intersegmental transfer by simultaneous binding to multiple regions of the genomic DNA, allowing parallel homology sampling (34). No such parallel sampling has been reported for Argonaute and CRISPR systems, yet jumping to a faraway region of the strand that is in close proximity in space might still be an effective way to facilitate the search process when combined with the local 1D sliding.

5.2. Mechanism of Kinetic Proofreading

Initial weak interactions with the subseed region in the case of Argonaute and with PAM in the case of CRISPR proteins aid in fast dismissal of off-target sites. When the initial target recognition step is successful, further proofreading occurs via intricate protein conformational changes.

In the case of human Argonaute, the first conformational change the protein undergoes under binding to a cognate target is the release of the kink between guide nucleotides 6 and 7, which allows for further guide-target base pairing (63, 97). This kink is released by rotating the alpha helix 7 by 4 Å, and this release is possible only if guide nucleotides 2–5 are fully complementary to the target. Such conformational rearrangement allows for full seed base pairing and sharply increases affinity (human AGO2 residence times on its target). Seed base pairing, however, is not enough to trigger the catalytic activity of this protein. Binding to the seed widens the channel between the PAZ and N-terminus domains, which in turn allows for the disordered supplemental region of the guide RNA (nucleotides 13–16) to adopt an A-helical conformation (97). Such preordering would decrease the entropic cost of target pairing even further. Base pairing in the mid-region has also been shown to be crucial for target cleavage, as shown by biochemical and single-molecule studies (55).

Despite full complementarity, Argonaute is unable to cleave some targets (55). A possible explanation could be that certain sequences are unable to trigger the final conformational change

in the middle region that would position the catalytic residues next to the cleavage site. It is also suggested that off-target site rejection is assisted by the interactions with the 3′ end of the guide and the PAZ domain of the protein. Single-molecule FRET studies have shown that modifications at this end of the guide slow down the protein dissociation from the target, which could lead to potential cleavage of off-target sites (58). Further single-molecule and structural studies will reveal the full conformational proofreading mechanism of hAGO2 even further and answer lingering questions such as why some sequences cannot be cleaved despite full complementarity.

The Cas9 protein also undergoes extensive conformational changes throughout its target search. The major conformational change occurs upon binding guide RNA, which enables Cas9 to search for PAM in a sequence-specific manner (see the sidebar titled Cas9 Preorganization and Structural Rearrangement for Target Search) (7). Upon binding to the correct PAM, Cas9 bends the DNA so that the duplex could be unzipped and interrogated. The first 10 nucleotides of the seed are preordered in an A-helical configuration, thus prepaying the entropic cost for target base pairing (51). As in the case with Argonaute, binding to the seed is enough to stabilize the Cas9-RNA-DNA complex (100). However, biochemical data have shown that cleavage for such targets is very inefficient. Further structural and single-molecule studies have revealed that Cas9 undergoes another conformational change as the complementarity between guide and target increases (25, 50). Initially, the HNH domain that cleaves the target strand is positioned at the PAM-distal end, far from the cleavage site. However, when a full target has been found, the HNH domain moves to the cleavage site, thus achieving a catalytically active conformation. Single-molecule and bulk FRET experiments showed that at least 18 of 20 nucleotides between the target and guide have to be complementary for Cas9 to achieve target cleavage (25, 104). As the number of mismatches is increased, the HNH domain is unable to pass through an intermediate conformation and cleave the target strand. The nontarget strand is cleaved by the RuvC domain, which is initially already positioned close to the cleavage site. However, without the movement of HNH domain, the nontarget strand also cannot be cleaved, indicating a signaling mechanism between the two domains as the final checkpoint before target cleavage.

The Cascade protein complex does not have a prominent conformational proofreading mechanism for binding. This could be a potential explanation as to why it can bind targets without a PAM or with significant mismatches in the seed. However, binding to a fully complementary target "locks" Cascade and stabilizes the R-loop (92, 109). It has been shown that this locking and the presence of a correct PAM sequence is required to recruit the Cas3 nuclease for target degradation (90, 92). Binding to partial targets instead triggers a priming response, where CRISPR memory is rapidly updated to fight escape mutants (18). In this response, it is likely that the Cas1-Cas2 protein complex, which is responsible for the spacer integration in the CRISPR locus, is necessary to recruit Cas3 (57, 90).

6. CONCLUDING REMARKS

Protein target search is a complicated process involving different search modes comprising weak interactions and protein conformational changes. Despite the difference in function, target search mechanisms of proteins from different families like Argonaute, CRISPR, and RecA share a lot of similarities. To begin with, these proteins recognize a short nucleic acid sequence as an initial recognition step: Argonaute recognizes the first 3 nucleotides of the seed, CRISPR/Cas proteins recognize a PAM sequence, and RecA recognizes the first 5–7 nucleotides of its target (23, 82, 89, 90, 105) (**Figure 6**; **Table 2**). Furthermore, all described proteins use a mixture of 3D and 1D diffusion to efficiently locate their targets (22, 23, 39, 89, 90, 105). Strikingly, all proteins can laterally diffuse approximately 10 bp in length. Another similarity is that Argonaute, Cas9, and RecA

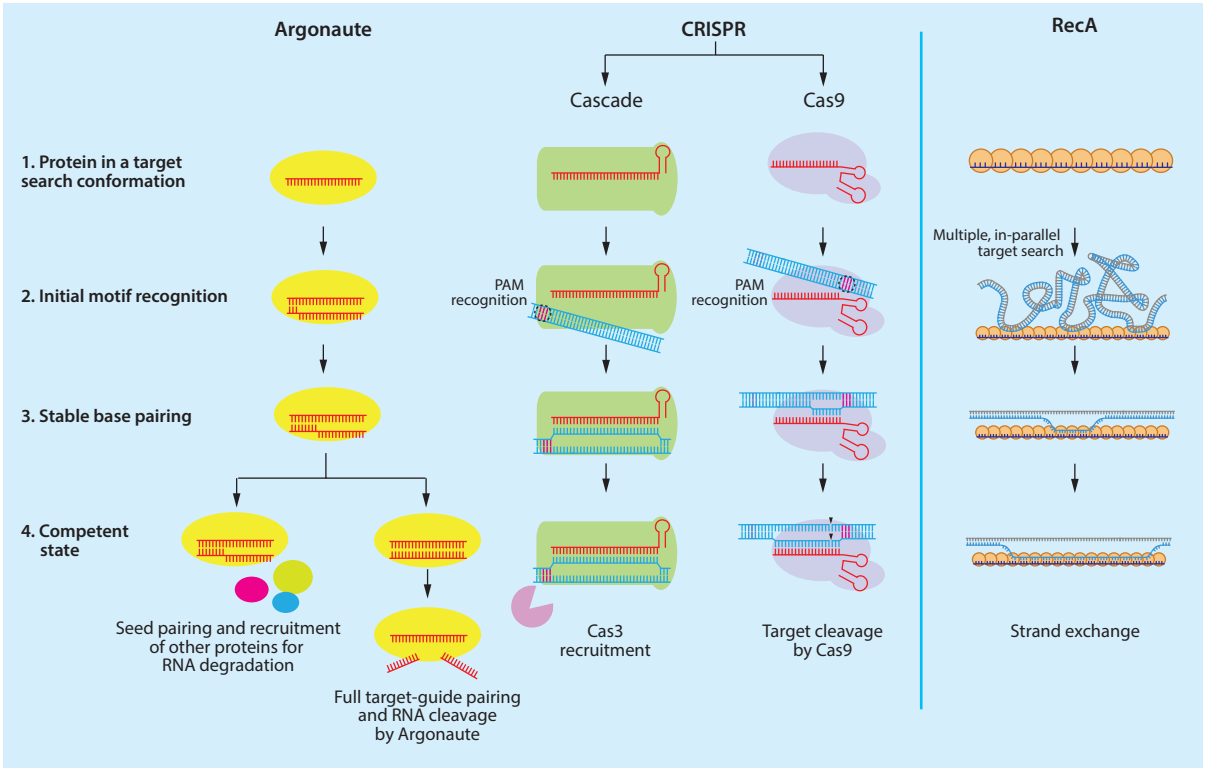


Figure 6

Summary of target search of Argonaute, Cas9, Cascade, and RecA proteins. Abbreviations: Cas, CRISPR-associated; Cascade, CRISPR-associated complex for antiviral defense; CRISPR, clustered regularly interspaced short palindromic repeats; PAM, protospacer adjacent motif.

proteins do not require full target complementarity for stable base pairing, but instead, binding is stabilized by the pairing of the first 7–12 nucleotides, depending on the protein (23, 67, 87, 100) (**Figure 6**). In all described systems, the competent state requires more extensive base pairing, with Argonaute and CRISPR proteins requiring the full length and near full length of the guide, respectively, and RecA requiring 15 nucleotides (25, 55, 70, 88, 115). In addition, Argonaute and CRISPR proteins undergo conformational changes during different stages of target recognition (18, 25, 58, 63, 114, 116) (**Figure 6**). Conformational changes of RecA upon target recognition have not been directly observed. All in all, despite subtle differences, these different protein families make use of the same core principles to find their targets in a fast and efficient manner.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to Thijs Cui and Misha Klein for critical reading. C.J. was funded by the European Research Council under the European Union's Seventh Framework Program (FP7/2007–2013),

ERC grant agreement no. 309509, a Netherlands Organisation for Scientific Research (NWO) VIDI grant (864.11.005), and the NWO Frontiers of Nanoscience program. S.H.K. was funded by the National Creative Research Initiative Program of the National Research Foundation (Center for Single-Molecule Systems Biology, NRF-2011-0018352).

LITERATURE CITED

1. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, et al. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353:aaf5573
2. Ambros V. 2004. The functions of animal microRNAs. *Nature* 431:350–55
3. Ambrose EJ. 1956. A surface contact microscope for the study of cell movements. *Nature* 178:1194
4. Ambrose WP, Goodwin PM, Nolan JP. 1999. Single-molecule detection with total internal reflection excitation: comparing signal-to-background and total signals in different geometries. *Cytometry* 36:224–31
5. Ameres SL, Martinez J, Schroeder R. 2007. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* 130:101–12
6. Amitai G, Sorek R. 2016. CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14:67–76
7. Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513:569–73
8. Anderson DG, Kowalczykowski SC. 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell* 90:77–86
9. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442:203–7
10. Barrangou R, Doudna JA. 2016. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* 34:933–41
11. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–12
12. Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–97
13. Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–33
14. Bauer M, Metzler R. 2012. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* 102:2321–30
15. Bell JC, Kowalczykowski SC. 2016. Mechanics and single-molecule interrogation of DNA recombination. *Annu. Rev. Biochem.* 85:193–226
16. Bell JC, Kowalczykowski SC. 2016. RecA: regulation and mechanism of a molecular search engine. *Trends Biochem. Sci.* 41:491–507
17. Berg OG, Winter RB, von Hippel PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–48
18. Blosser TR, Loeff L, Westra ER, Vlot M, Kunne T, et al. 2015. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* 58:60–70
19. Boland A, Tritschler F, Heimstadt S, Izaurralde E, Weichenrieder O. 2010. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep.* 11:522–27
20. Broderick JA, Salomon WE, Ryder SP, Aronin N, Zamore PD. 2011. Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *RNA* 17:1858–69
21. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, et al. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–64
22. Brown MW, Dillard KE, Xiao Y, Dolan AE, Hernandez ET, et al. 2017. Assembly and translocation of a CRISPR-Cas primed acquisition complex. bioRxiv 208058. **https://www.biorxiv.org/content/early/2017/10/26/208058**
23. Chandradoss SD, Schirle NT, Szczepaniak M, MacRae IJ, Joo C. 2015. A dynamic search process underlies microRNA targeting. *Cell* 162:96–107

24. Chen Z, Yang H, Pavletich NP. 2008. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* 453:489–94
25. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* 3:eaao0027
26. De Vlaminck I, van Loenhout MT, Zweifel L, den Blanken J, Hooning K, et al. 2012. Mechanism of homology recognition in DNA recombination from dual-molecule experiments. *Mol. Cell* 46:616–24
27. Dillingham MS, Kowalczykowski SC. 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* 72:642–71
28. Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096
29. Doxzen KW, Doudna JA. 2017. DNA recognition by an RNA-guided bacterial Argonaute. *PLOS ONE* 12:e0177097
30. East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JH, et al. 2016. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538:270–73
31. Egelman EH, Yu X. 1989. The location of DNA in RecA-DNA helical filaments. *Science* 245:404–7
32. Fineran PC, Gerritzen MJ, Suarez-Diez M, Kunne T, Boekhorst J, et al. 2014. Degenerate target sites mediate rapid primed CRISPR adaptation. *PNAS* 111:E1629–38
33. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–11
34. Forget AL, Kowalczykowski SC. 2012. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. *Nature* 482:423–27
35. Frank F, Sonenberg N, Nagar B. 2010. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465:818–22
36. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71
37. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS* 109:15539–40
38. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202
39. Globyte V, Lee SH, Bae T, Kim J-S, Joo C. 2018. CRISPR Cas9 searches for a protospacer adjacent motif by one-dimensional diffusion. bioRxiv 264879. **https://doi.org/10.1101/264879**
40. Gosse C, Croquette V. 2002. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys. J.* 82:3314–29
41. Greene EC, Wind S, Fazio T, Gorman J, Visnapuu ML. 2010. DNA curtains for high-throughput single-molecule optical imaging. *Methods Enzymol.* 472:293–315
42. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27:91–105
43. Ha T. 2001. Single-molecule fluorescence methods for the study of nucleic acids. *Curr. Opin. Struct. Biol.* 11:287–92
44. Ha T. 2001. Single-molecule fluorescence resonance energy transfer. *Methods* 25:78–86
45. Hauptmann J, Dueck A, Harlander S, Pfaff J, Merkl R, Meister G. 2013. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat. Struct. Mol. Biol.* 20:814–17
46. Hayes RP, Xiao Y, Ding F, van Erp PB, Rajashankar K, et al. 2016. Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530:499–503
47. Hochstrasser ML, Taylor DW, Bhat P, Guegler CK, Sternberg SH, et al. 2014. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *PNAS* 111:6618–23
48. Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262–78
49. Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, et al. 2014. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* 345:1473–79
50. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, et al. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351:867–71

51. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* 348:1477–81
52. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31:233–39
53. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–21
54. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343:1247997
55. Jo MH, Shin S, Jung S-R, Kim E, Song J-J, Hohng S. 2015. Human Argonaute 2 has diverse reaction pathways on target RNAs. *Mol. Cell* 59:117–24
56. Jones DL, Leroy P, Unoson C, Fange D, Ćurić V, et al. 2017. Kinetics of dCas9 target search in *Escherichia coli*. *Science* 357:1420–24
57. Jung C, Hawkins JA, Jones SK Jr., Xiao Y, Rybarski JR, et al. 2017. Massively parallel biophysical analysis of CRISPR-Cas complexes on next generation sequencing chips. *Cell* 170:35–47.e13
58. Jung S-R, Kim E, Hwang W, Shin S, Song J-J, Hohng S. 2013. Dynamic anchoring of the 3′-end of the guide strand controls the target dissociation of Argonaute–guide complex. *J. Am. Chem. Soc.* 135:16865–71
59. Kaya E, Doxzen KW, Knoll KR, Wilson RC, Strutt SC, et al. 2016. A bacterial Argonaute with non-canonical guide RNA specificity. *PNAS* 113:4057–62
60. Khorshid M, Hausser J, Zavolan M, van Nimwegen E. 2013. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat. Methods* 10:253–55
61. Kim SH, Ahn T, Cui TJ, Chauhan S, Sung J, et al. 2017. RecA filament maintains structural integrity using ATP-driven internal dynamics. *Sci Adv.* 3:e1700676
62. Klein M, Chandradoss SD, Depken M, Joo C. 2017. Why Argonaute is needed to make microRNA target search fast and reliable. *Semin. Cell Dev. Biol.* 65:20–28
63. Klum SM, Chandradoss SD, Schirle NT, Joo C, MacRae IJ. 2018. Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. *EMBO J.* 3:321–45
64. Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, et al. 2015. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* 350:823–26
65. Kowalczykowski SC. 2015. An overview of the molecular mechanisms of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* 7:a016410
66. Kwak PB, Tomari Y. 2012. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* 19:145–51
67. Lee JY, Terakawa T, Qi Z, Steinfeld JB, Redding S, et al. 2015. Base triplet stepping by the Rad51/RecA family of recombinases. *Science* 349:977–81
68. Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–54
69. Lesterlin C, Ball G, Schermelleh L, Sherratt DJ. 2014. RecA bundles mediate homology pairing between distant sisters during DNA break repair. *Nature* 506:249–53
70. Lim Y, Bak SY, Sung K, Jeong E, Lee SH, et al. 2016. Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* 7:13350
71. Lingel A, Simon B, Izaurralde E, Sattler M. 2004. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* 11:576–77
72. Ma JB, Ye K, Patel DJ. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429:318–22
73. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, et al. 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13:722–36
74. Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41:4360–77
75. Mani A, Braslavsky I, Arbel-Goren R, Stavans J. 2010. Caught in the act: the lifetime of synaptic intermediates during the search for homology on DNA. *Nucleic Acids Res.* 38:2036–43
76. Marraffini LA. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* 526:55–61

77. Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–45
78. Marraffini LA, Sontheimer EJ. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463:568–71
79. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15:185–97
80. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. 2002. Analysis of a *piwi*-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* 110:689–99
81. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353:aad5147
82. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–40
83. Mulepati S, Bailey S. 2013. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* 288:22184–92
84. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. 2013. Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* 51:594–605
85. Parker JS. 2010. How to slice: snapshots of Argonaute in action. *Silence* 1:3
86. Parker JS, Roe SM, Barford D. 2004. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* 23:4727–37
87. Qi Z, Redding S, Lee JY, Gibb B, Kwon Y, et al. 2015. DNA sequence alignment by microhomology sampling during homologous recombination. *Cell* 160:856–69
88. Ragunathan K, Joo C, Ha T. 2011. Real-time observation of strand exchange reaction with high spatiotemporal resolution. *Structure* 19:1064–73
89. Ragunathan K, Liu C, Ha T. 2012. RecA filament sliding on DNA facilitates homology search. *eLife* 1:e00067
90. Redding S, Sternberg SH, Marshall M, Gibb B, Bhat P, et al. 2015. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* 163:854–65
91. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403:901–6
92. Rutkauskas M, Sinkunas T, Songailiene I, Tikhomirova MS, Siksnys V, Seidel R. 2015. Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* 10:1534–43
93. Saetrom P, Heale BS, Snove O Jr., Aagaard L, Alluin J, Rossi JJ. 2007. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* 35:2333–42
94. Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. 2015. Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell* 162:84–95
95. 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.* 39:9275–82
96. Schirle NT, MacRae IJ. 2012. The crystal structure of human Argonaute2. *Science* 336:1037–40
97. Schirle NT, Sheu-Gruttadauria J, MacRae IJ. 2014. Structural basis for microRNA targeting. *Science* 346:608–13
98. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, et al. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *PNAS* 108:10098–103
99. Shabalina SA, Koonin EV. 2008. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* 23:578–87
100. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T. 2016. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* 7:12778
101. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, et al. 2013. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* 32:385–94
102. Slutsky M, Mirny LA. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87:4021–35

103. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305:1434–37
104. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527:110–13
105. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507:62–67
106. Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* 43:880–91
107. Sunghyeok Y, Taegeun B, Kyoungmi K, Omer H, Hwan LS, et al. 2017. DNA-dependent RNA cleavage by the Natronobacterium gregoryi Argonaute. bioRxiv 101923. **https://doi.org/10.1101/101923**
108. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, et al. 2014. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507:258–61
109. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, et al. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *PNAS* 111:9798–803
110. Updegrove TB, Zhang A, Storz G. 2016. Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* 30:133–38
111. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, et al. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303:672–76
112. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. 2012. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151:1055–67
113. Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, 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
114. Xue C, Whitis NR, Sashital DG. 2016. Conformational control of Cascade interference and priming activities in CRISPR immunity. *Mol. Cell* 64:826–34
115. Yao C, Sasaki HM, Ueda T, Tomari Y, Tadakuma H. 2015. Single-molecule analysis of the target cleavage reaction by the *Drosophila* RNAi enzyme complex. *Mol. Cell* 59:125–32
116. Zander A, Holzmeister P, Klose D, Tinnefeld P, Grohmann D. 2014. Single-molecule FRET supports the two-state model of Argonaute action. *RNA Biol.* 11:45–56
117. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–71
118. Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, et al. 2014. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* 515:147–50

Annual Review of Biophysics
Volume 47, 2018

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biophysics* articles may be found at http://www.annualreviews.org/errata/biophys