1-1-2019

Spy-ing on Cas9: Single-molecule tools reveal the enzymology of Cas9

Kelsey Whinn
University of Wollongong, ksw842@uowmail.edu.au

Antoine M. van Oijen
University of Wollongong, vanoijen@uow.edu.au

Harshad Ghodke
University of Wollongong, harshad@uow.edu.au

Spy-ing on Cas9: Single-molecule tools reveal the enzymology of Cas9

Abstract

Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas) systems are an adaptive immune response mechanism in prokaryotes which can target and cleave invading DNA or RNA. The rapid understanding of the type II CRISPR/Cas9 system through biochemical, genetic and structural investigations has contributed to the development of Cas9 for various DNA- and RNA-targeting applications. Recent single-molecule investigations of CRISPR/Cas9 mechanisms have further extended our understanding of target search, binding and cleavage. These investigations are fundamental to the further development of CRISPR/Cas9 tools. This review discusses how single-molecule techniques have illuminated the enzymology of Cas9 endonucleases.

Publication Details

Spy-ing on Cas9: Single-molecule tools reveal the enzymology of Cas9

Kelsey S. Whinn[a,b], Antoine M. van Oijen[a,b]*, Harshad Ghodke[a,b]*,

[a] Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, Australia;

[b]Illawarra Health and Medical Research Institute, Wollongong, Australia

*Corresponding authors: van Oijen, Antoine M (vanoijen@uow.edu.au) & Ghodke, Harshad (harshad@uow.edu.au)

Keywords: Single-molecule, CRISPR/Cas, *Spy*Cas9, *Spy*dCas9

Abbreviations:

CRISPR/Cas	Clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins
*Spy*Cas9	*Streptococcus pyogenes* Cas protein 9
*Spy*dCas9	Catalytically inactivated *Spy*Cas9
gRNA	Guide RNA
sgRNA	Synthetic single gRNA

Abstract

CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein) systems are an adaptive immune response mechanism in prokaryotes that can target and cleave invading DNA or RNA. The rapid understanding of the type II CRISPR/Cas9 system through biochemical, genetic and structural investigations have contributed to the development of Cas9 for various DNA- and RNA-targeting applications. Recent single-molecule investigations of CRISPR/Cas9 mechanisms have further extended our understanding of target search, binding and cleavage. These investigations are fundamental to the further development of CRISPR/Cas9 tools. This review discusses how single-molecule techniques have illuminated the enzymology of Cas9 endonucleases.

1.1   Introduction

In recent years, the ability to target specific DNA sequences has been revolutionised by the discovery of the prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats) systems [1, 2]. CRISPR systems possess a rich and complex biology. Identified as an adaptive immune response mechanism in bacteria and archaea, the CRISPR system allows the host to site-specifically target and destroy

invading DNA or RNA [2-5]. Single-molecule techniques have been successfully used to identify the elemental processes that underlie biological complexity [6-10]. These techniques directly measure the diverse and dynamic behaviours of single molecules without the need for synchronization, thus removing the ensemble averaging of biological states. Further, by permitting observation of transient and rare intermediate states along the reaction pathway, these techniques overcome the detection limits of ensemble-based methods. More recently, an expanding toolkit of single molecule techniques has enabled *in situ* observation of reactions occurring in living organisms, demystifying how individual protein actors work inside the cellular milieu [11-19]. In this report, we discuss the biology of the type II CRISPR/Cas system and the power of single-molecule methods in revealing the fine mechanistic details that are central to it.

2.1 CRISPR Systems

Currently two main classes of CRISPR systems have been described [20]: Class 1 and Class2, with the key difference being the number of protein effectors (CRISPR-associated proteins or 'Cas') required to perform the reaction. Briefly, class 1 systems, containing types I, III and IV, require multi-subunit complexes for RNA-guided surveillance in bacteria and archaea [4, 5, 20, 21]. In contrast, class 2 systems, comprising types II, V and VI, use a single RNA-guided nuclease to target and cleave specific invader DNA or RNA sequences [20, 22-24] (Table 1).

3.1 *Spy*Cas9

The best characterized type II CRISPR system of *Streptococcus pyogenes* is comprised of the Cas9 (*Spy*Cas9) endonuclease (Figure 1) [2, 3, 22, 25]. In this case, CRISPR-mediated adaptive immunity arises from a genetic locus composed of an operon encoding Cas9 and a repeat-spacer array [3]. This array of identical repeat sequences contains unique DNA sequences derived from an infecting virus or plasmid DNA. The acquisition of invader sequences forms the first step of the CRISPR-mediated immunity mechanism (Figure 1) [3]. The repeat-spacer array is then transcribed, producing precursor CRISPR-RNA (pre-crRNA) that is subsequently base paired with a non-coding RNA molecule, *trans*-activating RNA (tracrRNA), and processed to form a mature duplexed crRNA:trcrRNA (gRNA) [26]. Construction of a synthetic single-gRNA (sgRNA) chimera has also been used to direct *Spy*Cas9 to specific sequences [2]. The mature gRNA is then recruited by

*Spy*Cas9 proteins to form mature RNA-guided ribonucleoprotein effectors. These programmed *Spy*Cas9 proteins then survey the cellular environment for the presence of a short NGG motif, termed the protospacer adjacent motif (PAM), and invader sequences that are complementary to the guide RNA [2, 4]. Successful recognition of the complementary sequence is then followed by cleavage of the double-stranded DNA (dsDNA).

3.1.1. Structural biology of *Spy*Cas9

*Spy*Cas9 is a 1368 amino acid, globular protein whose domain architecture is composed of two major lobes; a nuclease (NUC) lobe and a recognition (REC) lobe (Figure 2) [27-30]. The REC lobe contains the REC1 and REC2 domains and a bridge helix. Of these, the REC1 lobe is critical for interactions with the gRNA, whereas the bridge helix is important for interactions with both the target DNA, as well as the gRNA [30]. The NUC lobe consists of the PAM-interacting (PI) domain and two nuclease domains, HNH and RuvC, that cleave DNA targets using one metal and two metal catalytic mechanisms, respectively [30]. Within the NUC lobe, the RuvC domain interfaces with the PI domain, forming a positively-charged surface that interacts with the 3′ tail of the gRNA [30]. We direct readers to reference [31] for a detailed review on the structural biology of Cas9.

3.1.2 Enzymology of *Spy*Cas9

The first step in the formation of the mature effector requires the loading of the crRNA:tracrRNA into the *Spy*Cas9 protein. Cryo-EM studies of apo-*Spy*Cas9 reveal that in the absence of RNA, *Spy*Cas9 is a highly flexible molecule, and large conformational rearrangements are triggered by the binding of RNA (Figure 3a, b) [29]. High-speed fluid atomic force microscopy studies further reveal that these structural transitions occur on the time scale of seconds on *Spy*Cas9 molecules immobilized on mica [32]. In agreement, enhanced MD simulations of apo-*Spy*Cas9 found that conformational rearrangement upon gRNA binding mainly involved the opening of the REC lobe with respect to the NUC lobe [33]. Opening of the bi-lobed structure results in the formation of a positively charged cavity capable of accommodating RNA. Single molecule FRET (smFRET) studies further demonstrate that apo-*Spy*Cas9 undergoes conformational rearrangement upon gRNA binding [34-36]. Binding of sgRNA to apo-*Spy*Cas9 containing FRET-pair labelled REC1 and RuvC domains (approximately 21 Å apart), caused a 56 Å increase in distance

between the REC1 and RuvC domains resulting in a decrease in FRET efficiency [36]. This movement suggested that sgRNA binding triggers a drastic opening of the REC lobe relative to the NUC lobe [36]. Bulk measurements of the kinetics of the conformational change associated with sgRNA binding were described by a single time constant, independent of the sgRNA concentration used [34]. This led the authors to suggest that an initial collision complex (apo-*Spy*Cas9:sgRNA) is formed rapidly, followed by a slower, induced conformational change leading to productive association (*Spy*Cas9:sgRNA). Further, using truncated sgRNA variants, it was found that specific motifs at both ends of the 20-nucleotide sgRNA sequence are required to stabilize the mature *Spy*Cas9:gRNA complex [35]. Collectively, these studies reveal that a complete crRNA:trcrRNA complex (or sgRNA) is required to trigger the flexible rearrangement of the REC lobe, to form a stable and mature *Spy*Cas9:gRNA complex capable of executing target search.

Undoubtedly, the CRISPR/Cas revolution is attributable to the remarkable ability to 'program' and 'target' *Spy*Cas9 to desired genomic loci. How does *Spy*Cas9 discriminate target sites from non-target sites? Single-molecule techniques have been extensively used to describe mechanisms involved in target search [11, 37-43]. Insight into this issue was first provided by Sternberg and co-workers who employed a double-tethered 'DNA curtain' assay to visualize quantum dot labelled nuclease dead *Spy*Cas9 (*Spy*dCas9):gRNA searching and binding to target sequences in λ DNA molecules, that are directionally stretched and tethered in a microfluidic flow cell using TIRF microscopy (Figure 4a-e) [44]. Both wild-type as well as *Spy*dCas9:gRNA constructs employed 3-dimensional diffusion to locate and specifically bind to the target site (Figure 4c). In addition to long-lived binding observed at target sites, *Spy*Cas9:gRNA was also found to bind transiently along the length of the DNA (Figure 4d). Strikingly, the probability of transient binding correlated strongly with the density of PAM sequences along the DNA substrate leading the authors to suggest that specific recognition of PAMs may be a crucial feature of target search by *Spy*Cas9:gRNA (Figure 4e). smFRET studies of dwell times of *Spy*Cas9:gRNA revealed that *Spy*Cas9:gRNA is longer lived on DNA substrates containing multiple PAM sequences [45]. By labelling the tracrRNA and DNA substrate with a FRET pair, this study determined that in conjunction with 3D diffusion, *Spy*Cas9:gRNA employs 1D diffusion to rapidly sample neighbouring PAM

sites. Indeed, measurements of association rates of *Spy*dCas9-YPet:sgRNA in live-cells containing a contiguous array of target sites revealed that it takes *Spy*dCas9:sgRNA approximately 6 hours to locate the target site in *E coli* (Figure 4f, g) [12, 46]. Binding of *Spy*dCas9-Ypet:sgRNA programmed to the *lacO1* operator sequence was triggered by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG), which induced the dissociation of LacI from *lacO1* and allowed *Spy*dCas9-YPet:sgRNA to bind [12] (Figure 4g). Measurements of DNA-bound residence times of *Spy*dCas9-YPet:sgRNA revealed an average lifetime of 30 ms, that is sufficiently fast so that *Spy*dCas9:sgRNA can sample the ~10^6 PAMs present in the *E coli* chromosome [12]. Together, these studies unveil a search mechanism that is exquisitely predicated on the stability of *Spy*Cas9:PAM interactions. The intrinsically weak nature of these interactions is critical for efficient sampling of the genome.

Stable binding is only achieved on genomic loci containing PAM sequences at the correct target site. Following initial recognition, the gRNA invades the dsDNA to form an R-loop [2]. In this state, *Spy*Cas9 wraps around the R-loop while excluding the non-target strand [27-30, 47] (Figure 2c, d). Numerous bulk studies have revealed the requirement of essential sequence complementarity between target DNA and gRNA, termed the seed region, for stable R-loop formation [2, 48, 49]. *In vitro* plasmid cleavage assays revealed that a seed region of at least 13 base pairs between crRNA:DNA proximal to the PAM is required for efficient cleavage, whereas up to six adjacent mismatches at the PAM-distal end are tolerated [2]. In agreement, deep sequencing of *Spy*Cas9:sgRNA targeted loci in human cells showed that the specificity within the seed region ranges from 8-14 bp immediately upstream from the PAM [48]. Stopped-flow measurements of 2-aminopurine fluorescence in the seed region suggest that R-loop formation is the rate-limiting step of *Spy*Cas9 target cleavage [50]. The stepwise formation of *Spy*Cas9:gRNA R-loops has also been observed at the single-molecule level. The use of magnetic tweezers to investigate *Spy*Cas9:gRNA binding revealed that R-loop formation occurs by a directional hybridization between the crRNA and target DNA, beginning at the PAM and extending toward the distal protospacer end [51]. Further, it was found that the PAM sequence primarily affects the association rate of the R-loop, while the base pairs distal to the PAM mainly affect R-loop stability [51]. Magnetic tweezer experiments interrogating the stability of the R-loop showed that PAM distal mismatches cause R-

loop propagation to stall and any further propagation competes with the collapse of the hybridized structure [52]. smFRET investigations of the *Spy*Cas9:gRNA complex on substrates containing varying numbers of mismatches, showed that the stability of the R-loop structure decreases linearly with the number of mismatches proximal to the PAM [53]. Notably, PAM-distal mismatches up to 11 bp in length still allowed the formation of a stable, but inactive complex [53].

Further, smFRET studies have identified two distinct *Spy*Cas9 domain conformations during R-loop propagation [54, 55]. The first of the two states, termed the 'open' conformation, represents a PAM-proximal bound complex that is capable of initiating R-loop formation. The second state, termed 'zipped' conformation, occurs after complete R-loop formation and represents a cleavage-competent state [54]. Recently, analysis of the dwell time of the zipped conformation revealed that the lifetime increases when the base-pairing length between the crRNA:DNA reaches a maximum of 18 bp [55]. The recent combination of optical tweezers with confocal fluorescence microscopy revealed that DNA stretching induced off-target *Spy*Cas9:gRNA binding [56]. smFRET was then employed to interrogate the R-loop formation at these sites using DNA containing bubbles, showing that off-target binding and cleavage occurred in the presence of 10 bp mismatches within the R-loop [56]. Together, these studies indicate that crRNA:DNA complementarity is essential to *Spy*Cas9 stability and cleavage competency. The directional hybridization of the R-loop ensures that when enough mismatches are formed, further propagation is stalled allowing *Spy*Cas9:crRNA to reject the site. Crucially, the cleavage-competent is only accessible when a stable R-loop with the maximum number of base pairs is formed.

*Spy*Cas9 and orthologs induce a blunt double-strand break (DSB), upstream and proximal to the guanine-rich PAM. Cleavage is catalyzed by the two nuclease domains within the NUC lobe, HNH and RuvC, that cleave the target and non-target strands, respectively [2, 28, 30] (Figure 2 a, c, d). Numerous *Spy*Cas9 structures have provided insight into the flexibility and structural rearrangement of the HNH catalytic domain into its active state [27, 29, 30]. smFRET studies showed that the conformational flexibility of the HNH domain directly controls cleavage [35]. Further smFRET investigations of labelled *Spy*Cas9:sgRNA complexes revealed that in the absence of DNA, *Spy*Cas9:sgRNA primarily resides (approximately 50%) in a stable

state termed the R state (referring to the RNA bound *Spy*Cas9 complex) [57]. Introduction of target-containing DNA led to the formation of a new state termed the D state (referring to the correctly base-paired, target DNA bound complex). The *Spy*Cas9:sgRNA complex was found to transiently access a third intermediate state (I state). Strikingly, transitions from R state to the D state necessarily required visits to the I state. Considering that (1) rate of transitions to the D state are ten-fold slower, and (2) the observation that the stable residence in the D state is infrequently encountered on target DNA containing 1-3 bp mismatches, these investigators suggested that the intermediate state serves as a conformational checkpoint gating initial binding and DNA cleavage [57]. Consistent with this hypothesis, investigations of domain flexibility in the presence of mismatched target DNA substrates showed that the HNH domain has a decreased propensity for its cleavage competent state in the presence of mismatched R-loops [58]. Indeed, greater than 4-bp mismatches in the R-loop at the PAM-distal end prevent this domain rearrangement [58]. Thus, the conformational rearrangement of the HNH domain into its cleavage-competent state is only triggered upon confirmation that the correct target site has been bound by the *Spy*Cas9 nuclease. This long-range allosteric communication between the flexible HNH domain and the PAM-distal end of the R-loop act as a final cleavage checkpoint to prevent the degradation of bound off-target sites.

*Spy*Cas9:gRNA complexes remain tightly bound to cleaved DNA products. Observations of quantum-dot labelled *Spy*Cas9 proteins in DNA curtain assays showed that *Spy*Cas9:gRNA remained bound to DNA after cleavage [44] (Figure 4c). Further, plasmid DNA cleavage assays at varying molar ratios of *Spy*Cas9:gRNA and target DNA revealed that *Spy*Cas9:gRNA complexes do not follow Michaelis-Menten kinetics [44]. In agreement, measurements of binding kinetics of *Spy*Cas9:gRNA interactions with substrate DNA performed using bio-layer interferometry revealed that the complex remains bound to cleaved products and displays an identical lifetime to *Spy*dCas9:sgRNA complexes [59]. Additionally, steady-state kinetic analysis of the dissociation of radiolabelled cleavage products exhibited slow multiple-turnover rates for both the HNH (k_{cat} 4.45×10^{-6} s^{-1}) and RuvC domains (k_{cat} 2.1×10^{-6} s^{-1}) [34]. Thus, *Spy*Cas9:gRNA acts as a single-turnover enzyme. Recently, bulk biochemical experiments have shown the ability of RNA polymerase to dislodge *Spy*Cas9 bound to cleavage products [60]. Displacement of

the *Spy*Cas9:sgRNA complex resulted in a multiple-turnover enzyme [60]. Interestingly, displacement of the *Spy*Cas9:sgRNA complex only occurred if the sgRNA had hybridized with the RNAP template strand [60]. The *Spy*Cas9 homolog, *Staphylococcus aureus* (*Sau*Cas9) was recently identified as a multiple-turnover Cas9 enzyme [61]. *In vitro* cleavage reactions analysed by capillary electrophoresis showed that *Sau*Cas9 cleavage resulted in 5-fold more cleavage product formation over 24 hours in comparison to *Spy*Cas9, suggesting that one *Sau*Cas9 complex can cleave multiple DNA substrates [61]. This enhanced rate of turnover is potentially attributable to an enhanced rate of product release [61]. Collectively, these studies indicate that *Spy*Cas9:gRNA complexes remain stably bound to cleavage products and are essentially single-turnover enzymes, unless acted upon by cellular machinery undertaking DNA transcription.

4.1   Conclusions and future perspectives

Single-molecule techniques have been crucial for assembling detailed models for the various kinetic mechanisms involved in effector maturation, target search and discrimination, conformational changes involved in catalysis and product release. Extensive characterisation of these mechanisms has contributed to the continuous adaptation of *Spy*Cas9 for a range molecular biology tools. Whereas *Spy*Cas9 has been extensively studied and remains at the forefront of the revolution in genetic engineering, other class 1 and 2 CRISPR/Cas systems remain relatively under-characterized. The single-molecule approaches used to characterize the enzymology of this exciting new family of proteins represent a broad and powerful toolkit for future investigation of other CRISPR/Cas systems. Ultimately, these approaches will enable the engineering of refined CRISPR/Cas systems for applications in genetic engineering, biotechnology, diagnostics and even fundamental discovery.

Declaration of interests: None.

References
• of special interest

•• of outstanding interest

[1] Ishino Y, Shinagawa H, Makino K *et al.* Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product Journal of Bacteriology 1987; 169:5429-5433.

••[2] Jinek M, Chylinski K, Fonfara I *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity Science 2012; 337:816-821.
> Authors describe the capability of a synthetically derived gRNA (sgRNA) chimera, containing the minimal interactions of a tracrRNA and crRNA, to guide *Spy*Cas9 to cleave target DNA. This gRNA construct has since been used in numerous characterizations of *Spy*Cas9 and in its application as molecular biology tool.

[3] Barrangou R, Fremaux C, Deveau H *et al.* CRISPR provides acquired resistance against viruses in prokaryotes Science 2007; 315:1709-1712.

[4] Brouns S, Jore M, Lundgren M *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes Science 2008; 321:960-964.

[5] Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 2008; 322:1843-1845.

[6] Stracy M, Kapanidis AN. Single-molecule and super-resolution imaging of transcription in living bacteria. Methods 2017; 120:103-114.

[7] Li Y, Schroeder JW, Simmons LA, Biteen JS. Visualizing bacterial DNA replication and repair with molecular resolution. Curr Opin Microbiol 2018; 43:38-45.

[8] Heck C, Michaeli Y, Bald I, Ebenstein Y. Analytical epigenetics: single-molecule optical detection of DNA and histone modifications. Curr Opin Biotechnol 2019; 55:151-158.

[9] Chen H, Larson DR. What have single-molecule studies taught us about gene expression? Genes & Development 2016; 30:1796-1810.

[10] Uphoff S, Sherratt DJ. Single-Molecule Analysis of Bacterial DNA Repair and Mutagenesis. Annu Rev Biophys 2017; 46:411-432.

[11] Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell Science 2007; 316:1191.

••[12] Jones DL, Leroy P, Unoson C *et al.* Kinetics of dCas9 target search in Escherichia coli. Science 2017; 357:1420-1424.
> Authors used single-molecule live-cell imaging to investigate the search mechanisms of SpydCas9. Authors reveal that it takes fluorescently labelled dCas9 approximately 6 hours to find the target sequence, indicating that each potential target within the *E. coli* genome is sampled.

[13] Stracy M, Lesterlin C, Garza de Leon F *et al.* Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. Proc Natl Acad Sci U S A 2015; 112:E4390-4399.

[14] Stracy M, Jaciuk M, Uphoff S *et al.* Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. Nat Commun 2016; 7:12568.

[15] Robinson A, McDonald JP, Caldas VE *et al.* Regulation of Mutagenic DNA Polymerase V Activation in Space and Time. PLoS genetics 2015; 11:e1005482.

[16] Ho HN, van Oijen AM, Ghodke H. The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage. Nat Commun 2018; 9:1570.

[17] Ghodke H, Paudel BP, Lewis JS *et al.* Spatial and temporal organization of RecA in the Escherichia coli DNA-damage response. Elife 2019; 8.

[18] Beattie TR, Kapadia N, Nicolas E *et al.* Frequent exchange of the DNA polymerase during bacterial chromosome replication. Elife 2017; 6.

[19] Uphoff S. Real-time dynamics of mutagenesis reveal the chronology of DNA repair and damage tolerance responses in single cells. Proc Natl Acad Sci U S A 2018; 115:E6516-E6525.

[20] Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol 2017; 37:67-78.

[21] Ozcan A, Pausch P, Linden A *et al.* Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol 2019; 4:89-96.

[22] Sapranauskas R, Gasiunas G, Fremaux C *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res 2011; 39:9275-9282.

[23] Zetsche B, Gootenberg JS, Abudayyeh OO *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 2015; 163:759-771.

[24] Shmakov S, Abudayyeh OO, Makarova KS *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 2015; 60:385-397.

[25] Garneau JE, Dupuis ME, Villion M *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 2010; 468:67-71.

[26] Deltcheva E, Chylinski K, Sharma CM *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011; 471:602-607.

[27] Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 2014; 513:569-573.

[28] Jiang F, Zhou K, Ma L *et al.* A Cas9-guide RNA complex preorganized for target DNA recognition Science 2015; 348:1477-1481.

[29] Jinek M, Jiang F, Taylor DW *et al.* Structures of Cas9 endonuclease reveal RNA-mediated conformational activation Science 2014; 343.

••[30] Nishimasu H, Ran FA, Hsu PD *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014; 156:935-949.

> Extensive characterisation of the multidomain architecture of the Cas9 crystal structure in complex with sgRNA and target DNA, at 2.5 Å resolution.

[31] Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys 2017; 46:505-529.

[32] Shibata M, Nishimasu H, Kodera N *et al.* Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. Nat Commun 2017; 8:1430.

[33] Palermo G, Miao Y, Walker RC *et al.* CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc Natl Acad Sci U S A 2017; 114:7260-7265.

[34] Raper AT, Stephenson AA, Suo Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. J Am Chem Soc 2018; 140:2971-2984.

[35] Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 2015; 527:110-113.

[36] Osuka S, Isomura K, Kajimoto S *et al.* Real-time observation of flexible domain movements in CRISPR-Cas9. EMBO J 2018; 37.

[37] Gorman J, Wang F, Redding S *et al.* Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. Proc Natl Acad Sci U S A 2012; 109:E3074-3083.

[38] Blainey PC, Luo G, Kou SC *et al.* Nonspecifically bound proteins spin while diffusing along DNA. Nat Struct Mol Biol 2009; 16:1224-1229.

[39] Blainey PC, Van Oijen AM, Banerjee A *et al.* A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA PNAS 2006; 103:5752-5757.

[40] Izeddin I, Recamier V, Bosanac L *et al.* Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife 2014; 3.

[41] Liao Y, Schroeder JW, Gao B *et al.* Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. Proc Natl Acad Sci U S A 2015; 112:E6898-6906.

[42] Wang YM, Austin RH, Cox EC. Single molecule measurements of repressor protein 1D diffusion on DNA Phys. Rev. Lett 2006; 97:048302.

[43] Ghodke H, Wang H, Hsieh CL *et al.* Single-molecule analysis reveals human UV-damaged DNA-binding protein (UV-DDB) dimerizes on DNA via multiple kinetic intermediates. Proc Natl Acad Sci U S A 2014; 111:E1862-1871.

[44] Sternberg SH, Redding S, Jinek M *et al.* DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014; 507:62-67.

●[45] Globyte V, Lee SH, Bae T *et al.* CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. EMBO J 2019; 38.

>smFRET investigations of transient Cas9 interactions with target DNA reveal *Spy*Cas9 interacts with PAM sequences weakly. Authors propose a model where lateral diffusion competes with 3D diffusion until both a PAM and target DNA sequence are found.

[46] Hammar P, Wallden M, Fange D *et al.* Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. Nat Genet 2014; 46:405-408.

[47] Jiang F, Taylor DW, Chen J *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage Science 2016; 351:867-871.

[48] Hsu PD, Scott DA, Weinstein JA *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013; 31:827-832.

[49] Mali P, Aach J, Stranges PB *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 2013; 31:833-838.

●[50] Gong S, Yu HH, Johnson KA, Taylor DW. DNA unwinding is the primary determinant of CRISPR-Cas9 activity. Cell Rep 2018; 22:359-371.

>Kinetic analyses of *Spy*Cas9:gRNA binding to DNA reveal the rate-limiting step *Spy*Cas9 cleavage mechanism is R-loop formation.

[51] Szczelkun M, Tikhomirova M, Sinkunas T *et al.* Direct observation of R-loop formation by single RNA-guide Cas9 and Cascade effector complexes PNAS 2014; 111:9798-9803.

[52] Rutkauskas M, Krivoy A, Szczelkun MD *et al.* Single-Molecule Insight Into Target Recognition by CRISPR-Cas Complexes. Methods Enzymol 2017; 582:239-273.

[53] Singh D, Sternberg SH, Fei J *et al.* Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat Commun 2016; 7:12778.

[54] Lim Y, Bak SY, Sung K *et al.* Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. Nat Commun 2016; 7:13350.

●[55] Zeng Y, Cui Y, Zhang Y *et al.* The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. Nucleic Acids Res 2018; 46:350-361.

>smFRET investigations of *Spy*Cas9 R-loop propagation revealed presence of an intermediate state prior to DNA cleavage. Investigations of this 'zipped' conformation revealed that the lifetime of the stably bound complex increases with the base-pairing length of the gRNA:DNA duplex.

[56] Newton MD, Taylor BJ, Driessen RPC *et al.* DNA stretching induces Cas9 off-target activity. Nat Struct Mol Biol 2019; 26:185-192.

●●[57] Dagdas YS, Chen JS, Sternberg SH *et al.* A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv 2017; 3.

>smFRET investigations of the conformational activation of *Spy*Cas9 showed presence of an intermediate checkpoint between DNA binding and cleavage, characterised by the flexible movements of the HNH domain.

●●[58] Yang M, Peng S, Sun R *et al.* The conformational dynamics of Cas9 governing DNA cleavage are revealed by single-molecule FRET. Cell Rep 2018; 22:372-382.

>Single-molecule FRET investigations reveal SpyCas9 cleavage is regulated by a long-range allosteric communication between the HNH domain and PAM-distal end of the R-loop.

[59] Richardson CD, Ray GJ, DeWitt MA *et al.* Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol 2016; 34:339-344.

[60] Clarke R, Heler R, MacDougall MS *et al.* Enhanced bacterial immunity and mammalian genome editing via RNA polymerase-mediated dislodging of Cas9 from double strand DNA breaks 2018.

•[61] Yourik P, Fuchs RT, Mabuchi M *et al. Staphylococcus aureus* Cas9 is a multiple-turnover enzyme RNA 2019; 25:35-44.

    In vitro biochemical and kinetic studies reveal *S. aureus* Cas9 is a multiple turnover enzyme.

Figure Legends:

Figure 1: CRISPR adaptive immunity mechanism. (1) Acquisition; foreign DNA sequence (green) is incorporated as a new spacer within the CRISPR repeat-spacer array. (2) Expression; the repeat-spacer array is transcribed, expressing pre-crRNA, which anneals to tracrRNA (red). Further processing yields mature gRNA duplexes that are recruited by effector proteins (blue). (3) Interference; mature gRNA guides the effector proteins to cleave the complementary sequences of subsequent invading DNA that contain PAM sequences.

Figure 2: Structural biology of *Spy*Cas9 in complex with sgRNA and target DNA. (a) Domain organization of *Spy*Cas9. BH, bridge helix; REC, recognition; NUC, nuclease; PI, PAM-interacting. (b) Schematic representation of the sgRNA hybridized to complementary target DNA sequence. (c) Ribbon representation of the *Spy*Cas9:sgRNA:DNA complex. Red dotted line represents disordered linker. (d) Surface representation of *Spy*Cas9:sgRNA:DNA complex. Dashed yellow circles represent HNH and RuvC active sites, H840A and D10A, respectively. Figure adapted with permission from [30**].

Figure 3: Domain flexibility of *Spy*Cas9. Cryo-EM structures of apo-*Spy*Cas9 (a) (19-Å resolution) and *Spy*Cas9:gRNA (b) (21-Å resolution) reveal flexibility of the bilobed structure. Cartoon representations (left) are shown alongside single-particle EM reconstructions of structures. The smaller NUC lobe (blue) rotates away from the larger REC lobe (gray) upon binding of gRNA. Figure adapted with permission from [29].

Figure 4: Target search and binding of *Spy*Cas9. (a) Schematic of double-tethered DNA curtain assay. λ-DNA substrates (48, 502 bp) are anchored by one end to a lipid bilayer through a biotin-streptavidin linkage. Hydrodynamic force is then used to stretch the DNA molecules until they encounter antibody-coated pentagons, with which they bind to through DIG-labelled ends. (b) WT *Spy*Cas9 or *Spy*dCas9 (top) was programmed with crRNA:tracrRNA targeting one of six specified sites (λ1-6). (c) YOYO1-stained DNA (green) bound by QD-tagged *Spy*Cas9 (magenta) programmed with λ2 gRNA. Binding site of programmed *Spy*Cas9 corresponds to expected target site of λ2 gRNA. *Spy*Cas9 remains bound after cleavage of dsDNA. (d) Kymographs of a single DNA molecule illustrating distinct stable and transient binding events (top). Zoomed in image of transient binding events (bottom). (e) Pearson correlation analysis of PAM distribution and non-target *Spy*Cas9:gRNA binding distribution for λ2 gRNA (blue) and a non-complementary gRNA (green) ($r = 0.59$, $P < 0.05$). Figures adapted with permission from [44]. (f) Schematic of live-cell imaging experimental setup used for kinetic analysis of *Spy*dCas9-Ypet:sgRNA. The microfluidic device (left) contains three ports assigned for medium, running waste and loading waste. The chamber contains three rows, each containing 17 traps. Cells are introduced from the running waste and are caught in the traps. Each trap is a 40 X 40 X 0.9 μm compartment, confined by two rigid walls and two openings. Cells that reach the openings are freed from the traps into the 10 μm deep surrounding (right). Figure adapted with permission from [46] (g) Schematic of single-molecule assay (top panel) where in the absence of IPTG, *lacO1* sites are occupied by LacI, preventing binding of *Spy*dCas9-Ypet:sgRNA (bottom panel). Addition of IPTG dissociates LacI, allowing *Spy*dCas9-Ypet:sgRNA to bind to the *lacO1* target, enabling detection of specific fluorescent spots using exposure times of 5 s. Bottom panel shows fluorescence images obtained before (left) and 10 min after (right) IPTG addition. Scale bar, 2 μm. Figure adapted with permission from [12*].
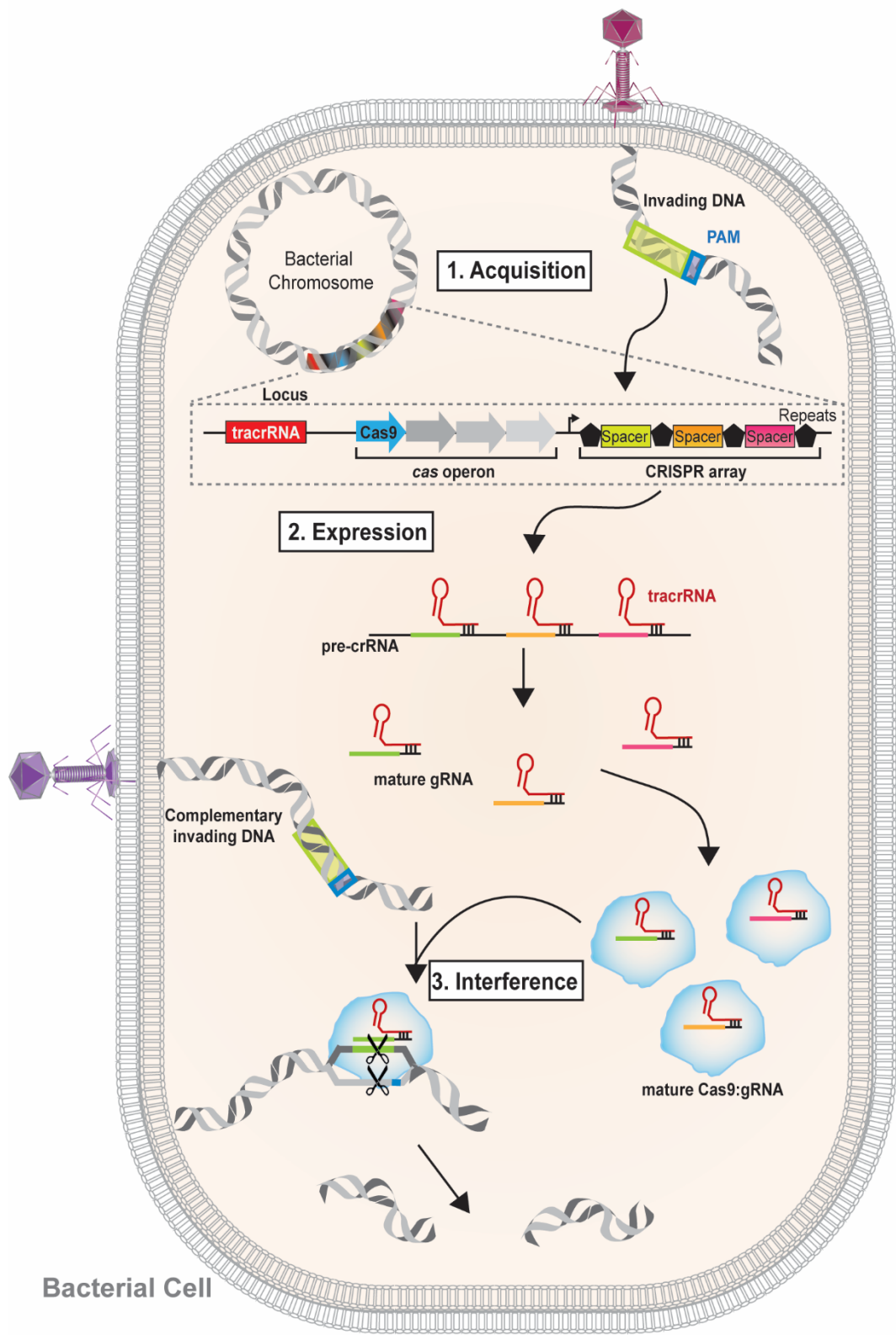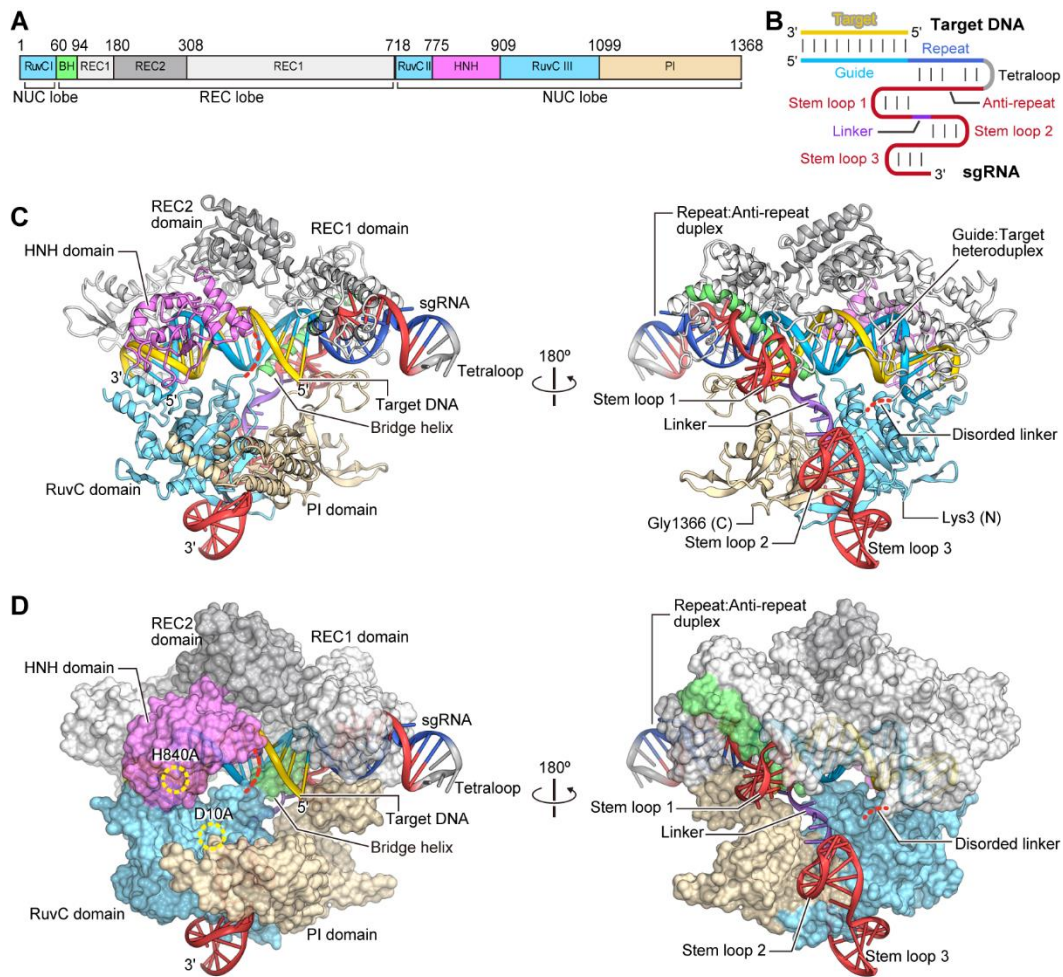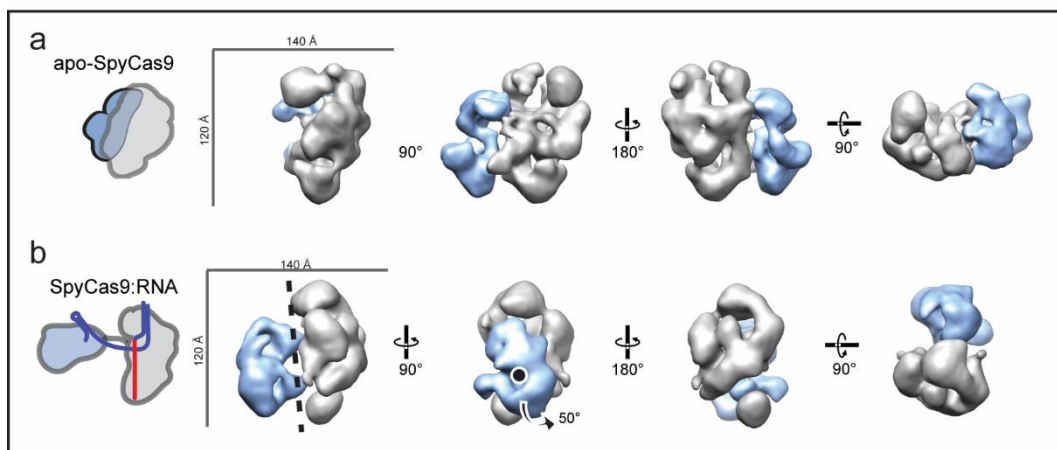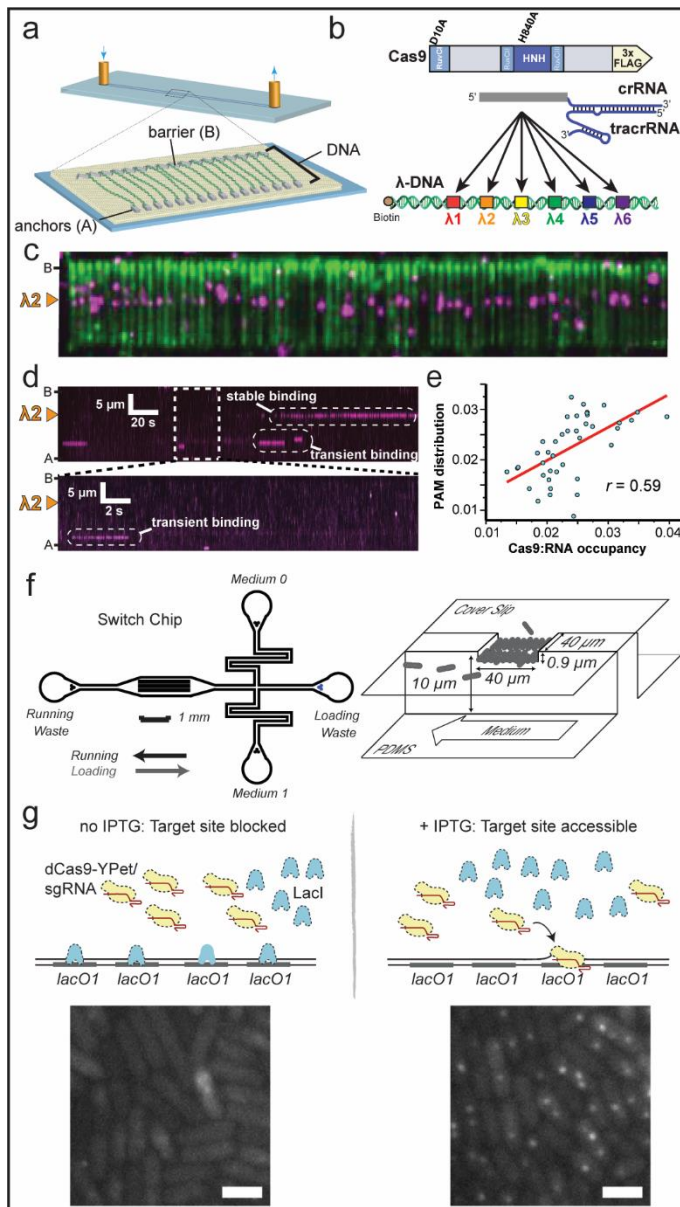
Figure 1

Figure 2

Figure 3

Figure 4