RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2016, **6**, 33514

An insight into the protospacer adjacent motif of *Streptococcus pyogenes* Cas9 with artificially stimulated RNA-guided-Cas9 DNA cleavage flexibility†

Yujie Geng, Zixin Deng and Yuhui Sun*

The CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) protein, Cas9, is a RNA-guided endonuclease that uses RNA–DNA base pairing to recognize and cleave double-stranded DNA (dsDNA) with a protospacer adjacent motif (PAM). It is widely accepted that the most commonly used *Streptococcus pyogenes* Cas9 (SpyCas9) protein recognizes a canonical 5′-NGG-3′ sequence in the PAM. In this study, we discovered another critical characteristic required for SpyCas9 cleavage *i.e.* the interspace between the protospacer and NGG. The results generated from DNA cleavage assays showed that both interspace length and the presence of a GG dinucleotide (particularly the upstream guanosine) are critical components in permitting SpyCas9-mediated cleavage. Interestingly, the interspace length significantly affects the selection of SpyCas9 cleavage sites on the non-complementary strand. Additionally, the complementary strand cleavage site is determined by the location of the single-molecular guide RNA (sgRNA). This indicates that PAM and sgRNA play different roles in determining the SpyCas9 specific cleavage site. Importantly, we also revealed for the first time that *in vitro* annealing of dsDNA with exogenous PAM-presenting oligonucleotides (PAMmers) stimulated SpyCas9 cleavage of target dsDNA without PAM. This study pertaining to PAM and SpyCas9 is expected to improve our understanding of SpyCas9 with an associated impact on related bioengineering capabilities.

Received 30th January 2016
Accepted 24th March 2016

DOI: 10.1039/c6ra02774a

www.rsc.org/advances

Introduction

The CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) protein, Cas9, has been widely used in genome editing[1–4] and gene regulation.[5–11] Cas9 uses HNH and RuvC nuclease domains to cleave complementary (target) and non-complementary (non-target) strands, respectively, yielding a double-strand DNA break (DSB).[12] Specific recognition and cleavage of target DNA by Cas9 requires the presence of a complementary dual crRNA–tracrRNA guide or a chimeric single-molecule guide RNA (sgRNA) along with a protospacer adjacent motif (PAM).[12] PAMs identified from different CRISPR–Cas systems and organisms have been shown to vary in sequence, length and location. Up until now, three different types of CRISPR–Cas system have been identified. In the type II CRISPR–Cas system, programmed DNA cleavage requires the fewest components and thus is suitable for bioengineering. PAMs from type II system are located downstream of the protospacer.[13] The PAM sequence for the most commonly used *Streptococcus pyogenes* Cas9 (SpyCas9) is widely accepted to be 5′-NGG-3′. Previous analyses incorporating the elucidation of crystal structures and electron microscopic reconstructions of SpyCas9 and its RNA- and DNA-bound complexes revealed a conformation flexibility associated with SpyCas9.[14,15] It would appear that SpyCas9 adopts an auto-inhibited conformation in the *apo* state. Conversely, the binding of guide RNA results in a conformational rearrangement that results in the formation of a central channel for target DNA binding.[14] The crystal structure of SpyCas9 complexed with an 83 nucleotide (nt) sgRNA and a partially duplexed target DNA structure containing a 5′-TGG-3′ PAM sequence, indicated that PAM recognition has a dual function. It serves as a critical determinant of initial target DNA binding while also constituting a licensing element in subsequent strand separation and guide-RNA-target-DNA hybridization.[16] After PAM recognition, sufficient sgRNA–DNA complementarity enables DNA double strand cleavage by driving concerted conformational changes in HNH and RuvC nuclease domains.[17] These studies highlighted the importance of RNA binding and PAM recognition in optimal SpyCas9 functioning.

SpyCas9 predominantly cleaves DNA at a 3 base pair (bp) site located upstream of the PAM,[12,15] resulting in a blunt end. However, the reason for this cleavage, where SpyCas9 cleaves

Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Wuhan University), Ministry of Education, and Wuhan University School of Pharmaceutical Sciences, Wuhan 430071, People's Republic of China. E-mail: yhsun@whu.edu.cn

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ra02774a

the two DNA strands at the same site, is not entirely understood. In both structural complexes of SpyCas9–sgRNA (either with PAM-containing target DNA or single-stranded target DNA), the HNH domain is located away from the scissile phosphate of the complementary strand.[15,16] This suggests that this conformation differs from the authentic SpyCas9 cleavage conformation. Furthermore, there are key structural dissimilarities between the SpyCas9 RuvC/HNH domain and similar RuvC/HNH nuclease domains.[15] This makes it difficult to elucidate the SpyCas9 RuvC/HNH domain following analysis of other RuvC/HNH nucleases. In this study, we altered the length of interspace between NGG and protospacer from upstream (−6 bp) of the native gapless site (0) to downstream (+6 bp) of the same site by changing the relative position of the 5′-NGG-3′ PAM sequence and the 20 nt guide sequence of sgRNA. This allowed us to investigate the influence of this sequence on SpyCas9 activity. The cleavage assay showed that SpyCas9 is capable of cleaving DNA containing an interspace length of −1 bp (GG), 0 bp (NGG), +1 bp (NNGG), +2 bp (NNNGG), and even +3 bp (NNNNGG), demonstrating a critical role for interspace length in relation to SpyCas9 cleavage. We next determined the cleavage sites of complementary and non-complementary strands for all cleavage products following end-sequencing. We found that the cleavage site in the complementary strand was invariably located on the sgRNA guide sequence regardless of the interspace length. In contrast, the non-complementary strand interspace length greatly affected cleavage frequency and the position associated with the cleavage.

When the crystal structure of SpyCas9 (complexed with a sgRNA and a PAM-containing target DNA) was analyzed, the GG dinucleotide appeared to be recognized by two conserved arginine residues of SpyCas9.[16] This seems to be a prerequisite for DNA strand separation and sgRNA-target DNA hybridization. Our *in vitro* assay showed the apparent difference in efficacy associated with each guanosine (G) nucleotide in the 5′-G_1G_2-3′ dinucleotide sequence in relation to SpyCas9 function at a high concentration of Cas9–sgRNA complex. Interestingly, substitutions of G_1 with a thymine nucleotide (T) or cytosine nucleotide (C) completely abolished the cleavage activity of SpyCas9, while substitutions of G_2 with T or C yielded modest reductions in cleavage activity. Notably, our study also demonstrated that the effect of the GG dinucleotide in relation to SpyCas9-mediated cleavage is partially recovered by DNA heating denaturation and *in vitro* annealing with artificial PAM-presenting oligonucleotides (PAMmer). This permits subsequent SpyCas9-mediated cleavage of DNA in the absence of native PAM sequence.

Experimental

Bacterial strains and culture conditions

E. coli DH10B was used as a bacterial host during cloning. *E. coli* Rosetta (DE3) was used for protein expression. Both of the *E. coli* strains were cultured in 2× TY broth (tryptone 1.6%, yeast extract 1%, NaCl 0.5%) and 2× TY agar at 37 °C with appropriate antibiotic selection (100 μg ml^{-1} ampicillin and 50 μg ml^{-1} kanamycin).

Expression and purification of SpyCas9

SpyCas9 was expressed and purified essentially as previously described.[12] Briefly, His-tagged SpyCas9 was expressed in *Escherichia coli* Rosetta (DE3). This strain was grown overnight in 2× TY medium at 18 °C following induction with 0.2 mM IPTG. Cells were lysed in 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM TCEP. The clarified lysate was applied to a 3 ml Ni-NTA (GE Life Sciences) affinity column. The column was washed with wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 50 mM imidazole pH 8.0) and eluted with elution buffer (20 mM Tris pH 8.0, 250 mM NaCl, 250 mM imidazole pH 8.0). The protein was further purified using a PD-10 Desalting Column (GE Life Sciences), and eluted with store buffer (10 mM Tris pH 7.4, 500 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol). Eluted protein was concentrated to ∼4 mg ml^{-1} using Amicon® Ultra Centrifugal Filters (Millipore) and stored at −25 °C.

In vitro transcription and purification of sgRNA

sgRNAs were prepared by *in vitro* transcription using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific). The DNA template for *in vitro* transcription was generated using PCR with the primers listed in Table S1 in ESI.† PCR was performed with denaturation at 95 °C for 3 min, followed by 30 cycles of 95 °C for 30 s, 68 °C for 15 s, and 72 °C for 10 min. sgRNAs were pre-annealed prior to the reaction by heating to 95 °C for 30 s, and slowly cooling down to room temperature.

DNA manipulation and site-directed mutagenesis

DNA manipulations including DNA preparation, digestion, purification, and agarose gel electrophoresis were performed using standard techniques. All site-directed mutagenesis[18] reactions were performed using pUC18 as the template with the primers listed in Table S1 in ESI.† Restriction endonucleases and Phusion High-Fidelity Master Mix with GC-buffer were purchased from NEB. DNA purification and plasmid preparation kits were purchased from TIANGEN. Oligonucleotide primer synthesis and DNA sequencing were performed at GenScript and Tsingke.

DNA cleavage assay

Cas9–sgRNA complexes were reconstituted prior to the cleavage assay by pre-incubating SpyCas9 (250 nM) and sgRNA (250 nM) at 37 °C for 10 min in a reaction buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.1 mM EDTA, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), and 5% glycerol). 300 ng of circular or *Ssp*I-linearized pUC18 plasmid were added to the reaction mix and the cleavage reactions (20 μl in total volume) were performed at 37 °C for 1 h. When end-sequencing was required for cleavage products, the reaction volume was increased to 50 μl with a proportional increase in the quantity of SpyCas9, sgRNA, and reaction buffer. Upon introduction if PAMmers, the DNA substrate (*Ssp*I-linearized pUC18 plasmid, 400 ng) and PAMmer (1.2 μM) were pre-incubated at 95 °C for 3 min and 58 °C for 10

min prior to the addition of the SpyCas9–sgRNA complex. Reactions were terminated by addition of 1 μl of 20 mg ml^{-1} proteinase K to the 20 μl of reaction, then incubated at 55 °C for 1 h. Cleavage products were resolved by gel electrophoresis using 0.8% agarose gel and were visualized by ethidium bromide staining.

Results and discussion

The length of interspace between the protospacer and NGG affects both SpyCas9 cleavage activity and cleavage position

It has been well documented that the HNH and RuvC nuclease domains of SpyCas9 are used for double-stranded DNA cleavage at a position 3 bp upstream of the PAM sequence, yielding a blunt end.[12,15] Both sgRNA binding and GG dinucleotide recognition are crucial for Cas9-mediaed DNA cleavage.[12] However, there has been no research conducted into how the locations of the sgRNA guide sequence and NGG affecting cleavage site choosing by SpyCas9. In order to analyze this, we attempted to use both supercoiled circular and linear pUC18 plasmid DNA as templates and artificially reposition the native NGG from 5 bp downstream (+1 to +5) of a randomly selected locus to 3 bp upstream (−1 to −3) of the same locus at protospacer-1 through site-directed mutation (Fig. 1A). We subsequently conducted the cleavage assay using DNA substrates containing the repositioned PAMs (see Experimental). Surprisingly, the cleavage result revealed the substantial cleavage of both supercoiled circular and linearized plasmid DNA created by *Ssp*I containing PAM_{-1}, and PAM_{+2} besides PAM_{+1} (Fig. 1B and C), which has been identified as NNGGN *in vivo*.[19,20] When native PAM_0 was used as a positive control, no comparatively significant cleavage was observed for other PAM mutations. Similar cleavage products were also observed when PAM_{-1}, PAM_{+1}, and PAM_{+2} from another two randomly selected locus protospacer-2 and -3 in both supercoiled circular and linearized plasmid DNA were analyzed (Fig. S1 and S2 in ESI†). Furthermore, similar but less cleavage can be seen when the interspace was even elongated to 3 bp (PAM_{+3}) (Fig. 1B and C and S1 and S2 in ESI†). This observation demonstrated that the interspace between the PAM and protospacer is also critical for SpyCas9 cleavage and the range can cover at least from 1 bp overlap (−1) to 2 bp (+2) even more distance suggesting a variable flexibility and inclusivity of SpyCas9. Previous structural and biochemical observations demonstrated that PAM recognition acts as a critical determinant in initial target DNA binding. This recognition is also important in subsequent strand separation and sgRNA-target DNA hybridization. PAM recognition and the subsequent interaction with SpyCas9 result in local duplex melting. This facilitates the transition of the SpyCas9–RNA complex to the SpyCas9–RNA–DNA ternary complex.[14–16,21] We believe that the critical role of interspace between target sequence and NGG may dictate the number of bases that require unwinding only *via* interaction between PAM and SpyCas9 prior to RNA–DNA heteroduplex formation (Fig. 2). The results of our DNA-cleavage assay demonstrate that the interspace can span a distance of up to five base pairs. This is similar to results obtained for *Streptococcus thermophilus* LMG18311.[13] Our analysis demonstrated that SpyCas9–PAM interaction could result in at least four base pairs unwinding. The number of bases involved in the initial pairing between guide sgRNA sequence and target DNA is likely to affect DNA–RNA heteroduplex formation. This effect may result in a variation in cleavage efficiency of DNA containing different interspace lengths.

Interestingly, when PAM_{+3} in protospacer-1 was analyzed, substantial cleavage was only observed in circular plasmids, suggesting that DNA configuration can also affect SpyCas9 cleavage. This is in agreement with observations made following SpyCas9 mutation, where Arg 1333 and Arg 1335 were replaced with Ala, resulting in a difference in cleavage activity associated with circular and linear DNA.[16] The additional cleavage of the circular DNA is most likely due to the occurrence of supercoiling, allowing R-loop formation at mismatched PAMs. SpyCas9 originates from the type II CRISPR–Cas system, which is a naturally occurring adaptive microbial immune system that facilitates defense against invading phage and other mobile genetic elements.[22–25] Under normal conditions, it would appear that supercoiled DNA is a more suitable template for SpyCas9.

It has been reported that SpyCas9 is a nuclease with strong conformational flexibility, which maintains an auto-inhibited conformation in the *apo* state. RNA loading rearranges the two lobes of SpyCas9 to form a central channel for DNA binding. DNA binding, especially the interaction between PAM and the PAM-interaction CTD domain, induces a further structural rearrangement.[14–16] A double-tethered DNA curtains assay showed that the association between SpyCas9–RNA and target sites is directed through three-dimensional collisions, and not through facilitated diffusion processes including one-dimensional sliding, and/or hopping.[21] Therefore, we believe that effective cleavage of target DNA by SpyCas9 depends on whether the target DNA, sgRNA and SpyCas9 ternary complex adopts the appropriate formation, rather than a complete reliance on the DNA sequence composition. In other words, a perfectly matched protospacer and canonical PAM sequence are sufficient, but not necessary for SpyCas9 cleavage. This may be an alternative reason for the presence of an unpredictable off-target site. In some cases, the target DNA is still subjected to cleavage by SpyCas9 even when the sequence does not fully match the 20 nt guide sequence of the sgRNA.[20,26,27] As a result, we propose that to facilitate more accurate predictions of off-target site locations, considerations associated with DNA, sgRNA and SpyCas9 ternary complex conformations are required in conjunction with a better understanding of the conformational features of the SpyCas9–sgRNA-cleaved DNA ternary complex.

To investigate the exact cleavage site associated with the altered PAMs following cleavage with SpyCas9, we next recovered all available cleavage products following separation using agarose gel electrophoresis. The isolated product termini were subsequently end-sequenced. Notably, we found that the cleavage site of the complementary strand was always 3 nt upstream of the 3′ end of the 20 nt sgRNA guide sequence (Fig. 1 and S1 and S2 in ESI†). This is consistent with previous

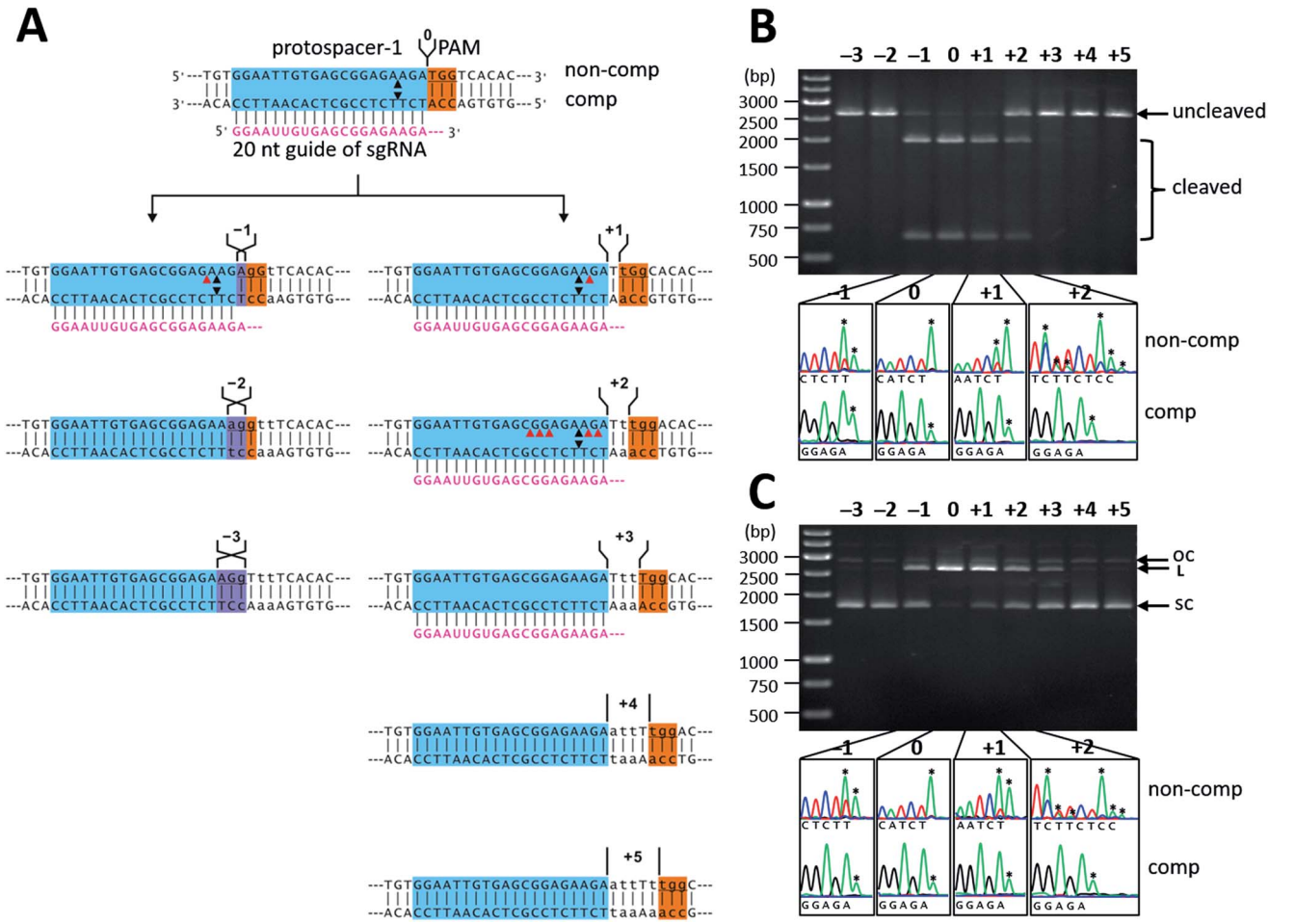

Fig. 1 Impact of length of the interspace between protospacer-1 and NGG on SpyCas9 cleavage. (A) Schematic representation of dsDNA with interspace of variable length. The protospacer region, the NGG region (underlined) and the overlapping region of NGG and protospacer are highlighted in blue, orange, and purple, respectively. sgRNA guide sequence is coloured red. The black and red triangles indicate known cleavage sites associated with SpyCas9 and new cleavage sites observed as a result of this study, respectively. The number with/without + and − indicates the number of base pairs associated with the interspace or overlapping region between the protospacer and NGG, respectively. (B) and (C) show cleavage results associated with the dsDNA represented in (A) and end-sequencing analysis of cleavage products associated with the *Ssp*I-linearized plasmid and circular plasmid, respectively. The 3′ terminal A overhang which is an artifact of the sequencing reaction is represented using an asterisk. The following abbreviations/acronyms were used: OC, open circular DNA; L, linear DNA; SC, supercoiled DNA; comp, complementary strand; non-comp, non-complementary strand.

observations in *Streptococcus thermophilus* LMG18311,[13] suggesting that Cas9 orthologs require identical complementary strand cleavage sites. The cleavage site of the non-complementary strand is altered when the NGG sequence is re-positioned. Interestingly, the end-sequencing results revealed that there was a single cleavage site consistently located 3 nt upstream of NGG. Additional cleavage sites were observed upstream of NGG, as far as 11 nt from the NGG sequence (Fig. 1 and S1 and S2 in ESI†). Additionally, the newly discovered cleavage sites associated with PAM_{+2} from protospacer-1 and -2 were divided into two groups based on the cleavage site location. These groups were separated by three to four uncleaved nucleotides. This indicated that the mechanism associated with cleavage of non-complementary strands is more complicated than that of complementary strands.

Furthermore, we have observed that a cleavage site corresponding to the invariable cleavage site of complementary strands is always found on the non-complementary strand (upper black triangle in Fig. 1 and S1 and S2 in ESI†), suggesting that sgRNA is also involved in cleavage of the non-complementary strand. Taken together, these results suggest that the cleavage site of the complementary strand is controlled by the position of the sgRNA guide sequence, while the cleavage site of the non-complementary strand is determined by both sgRNA and NGG. This implies that base pairing between the guide sequence of the sgRNA and the complementary DNA strand results in the deposition of the scissile phosphate group near the HNH domain. In addition, interaction between the PAM-interacting carboxyl-terminal domain (CTD) of SpyCas9 and the GG dinucleotide, together with positioning of sgRNA, results in the deposition of the scissile phosphate of the non-complementary strand near the RuvC domain. This demonstrates that sgRNA and PAM have different roles in determining the cleavage site of DNA double strands.

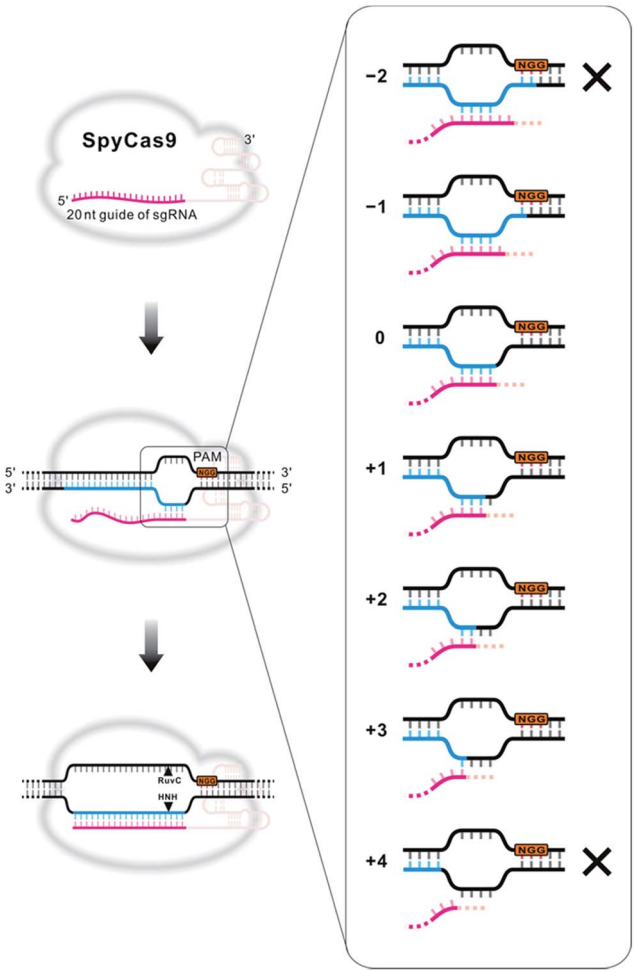

Fig. 2 Model for SpyCas9–sgRNA-mediated targeting and cleavage of dsDNA with variable interspace regions. sgRNA binding to SpyCas9 stimulates preorganization of SpyCas9 for PAM recognition. PAM interaction with SpyCas9–sgRNA initiates (a minimum of) 3–4 bp dsDNA melting immediately upstream of the PAM. The local strand separation reaction enables base pairing between the sgRNA guide and the target sequence. This results in the binding of sgRNA to the target sequence, thereby initiating R-loop formation. dsDNA with an interspace of −1, 0, +1, +2, and +3 bp can facilitate substantial sgRNA binding (with variations in the binding affinities). Base pairing between the guide sequence of sgRNA and the target DNA sequence in the complementary strand subsequently promotes further stepwise destabilization of the target DNA duplex. This also permits R-loop expansion *via* sequential unwinding. The complete R-loop serves as a signal for subsequent dsDNA cleavage by the HNH and RuvC nuclease domains of SpyCas9. The guide sequence of sgRNA is coloured red and the target sequence in the complementary strand is coloured blue. The number with/without + and − indicates the number of base pairs associated with the interspace or overlapping region between the protospacer and NGG, respectively.

Previous observations suggested that the RuvC domain of Cas9 from *Streptococcus thermophilus* LMG18311 is involved in cleaving the non-complementary strand at a site that is located at a specific distance from the conserved sequence of PAM.[13] However, our results demonstrated that both sgRNA binding and the positioning of the NGG sequence could affect the cleavage site of the non-complementary strand. This suggests that SpyCas9 uses a more complex mechanism in the recognition of the cut site of the non-complementary strand.

The guanosine G_1 of 5′-NG_1G_2-3′ in PAM is sufficient but not necessary for SpyCas9 function

Several biochemical studies have shown that mutations in PAM sequence greatly compromise SpyCas9 cleavage.[12,19,20,28] PAM recognition coincides with initial destabilization of the adjacent sequence and directional target DNA unwinding from the PAM-proximal end, thereby resulting in the formation of a RNA–DNA heteroduplex.[16,21] The crystal structure of the SpyCas9–sgRNA–PAM DNA duplex indicates that the GG dinucleotide in the PAM sequence is recognized following base-specific hydrogen-bond interactions with Arg 1333 and Arg 1335.[16] In order to evaluate each of the two guanosines in the GG dinucleotide associated with the PAM sequence, we substituted each G with A, T, or C, respectively, by site-directed mutagenesis and subsequently investigated the cleavage activity of SpyCas9 in conjunction with the mutated PAM sequences at a higher Cas9–sgRNA complex concentration (~250 nM) compared with the reported 0.5 nM and 50 nM.[29] In agreement with previous studies, our results showed that the single-base transition of G to A had the least effect on SpyCas9 cleavage.[19,20,28–30] It is likely that transversion variants resulted in more significant effects than transitions because both G and A are purine bases. SpyCas9 interacts with the GG dinucleotide *via* hydrogen-bonding with a purine base. Surprisingly, we observed that when the first guanine nucleotide (G_1 in the 5′-NG_1G_2-3′ PAM sequence) was substituted with a pyrimidine (T or C), the cleavage activity of Cas9 was fully abolished. However, when, G_2 was substituted with a pyrimidine (T or C), there was still some partial cleavage activity (Fig. 3A and S3 in ESI†). This suggests that SpyCas9 can catalyze cleavage of dsDNA as long as G_1 is present in the PAM sequence. One conclusion that can be drawn from this analysis is that G_1 (5′-NG_1G_2-3′) is sufficient but not necessary for SpyCas9 cleavage activity at a high concentration of Cas9–sgRNA complex. This would also suggest that G_1 plays a more important role than G_2 in facilitating the activity of SpyCas9. This might be due to the fact that C_1 (5′-C_2C_1N-3′) in the complementary strand is involved in the interaction with the phosphate lock loop through contact with Lys 1107. This contact requires a pyrimidine to be present at this position.[16] In addition to analyzing the potential necessity associated with the presence of each guanosine in PAM, we also tested SpyCas9 cleavage activity after both guanosines had been replaced by non-guanine nucleotides. The results demonstrated that there was no activity detected (apart from some very weak cleavage products when TGG was substituted with TAA) (Fig. 3B). This would suggest that the presence of the combination of common purine bases (G and A) allows for the partial replacement of G with A during recognition by SpyCas9.

Specificity is of vital importance for all genome-editing tools, especially in applications that apply to human therapeutics and food sources. Extensive efforts have been made to characterize the off-target sites associated with SpyCas9. However, previous studies and sgRNA design platforms have tended to focus on

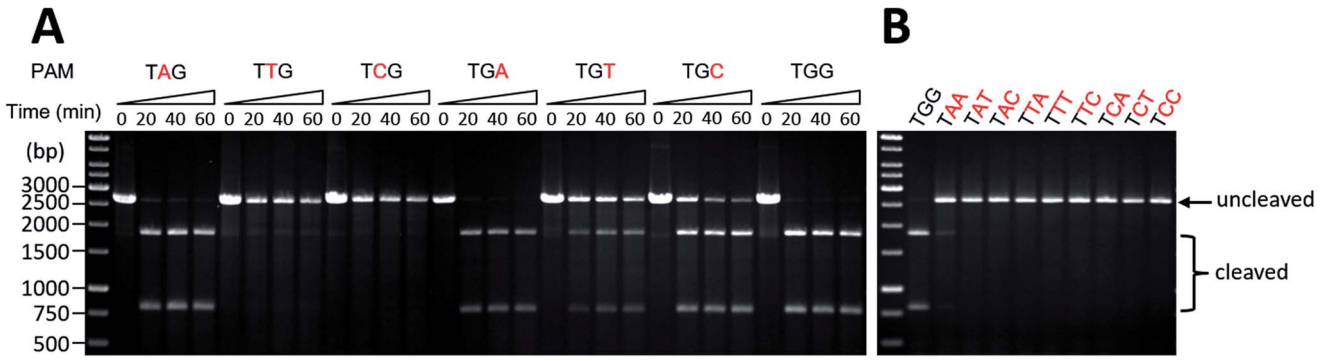

Fig. 3 The influence of single (A) and double (B) guanosine mutations in the NGG sequence of PAM on SpyCas9-mediated cleavage using *Ssp*I-linearized plasmid on protospacer-2. The mutated nucleotides are indicated in red.

the homology between the target DNA sequence and guide RNA.[26,31] In this study, we found that SpyCas9 facilitates substantial cleavage of target DNA containing variant PAMs. This occurs in the presence of interspace regions composed of different lengths or mutations in the GG dinucleotide of the PAM sequence.

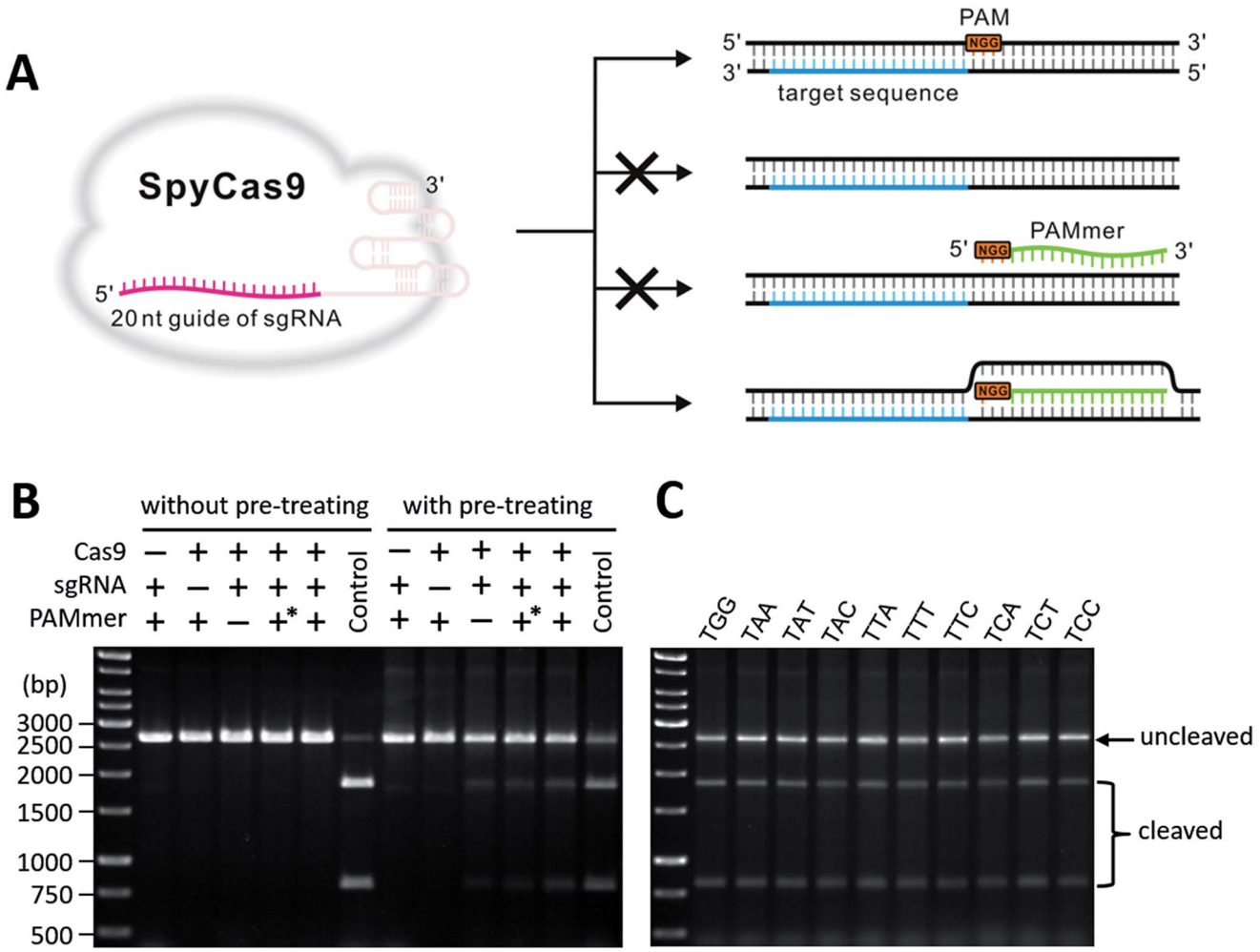

Fig. 4 *In vitro* annealing of dsDNA lacking a NGG PAM with a PAMmer enables SpyCas9 to target and cleave non-PAM sites. (A) Schematic diagram of the approach designed to target and cleave non-PAM sites of dsDNA with *in vitro* annealing and introduction of a PAMmer. The 20 nt guide sequence of sgRNA is coloured red, the target sequence in the complementary strand is blue and PAMmer is shown in green. (B) The SpyCas9 cleavage assay using a *Ssp*I-linearized plasmid harboring a TTC sequence in PAM and using different conditions. The plasmid containing the TGG sequence in PAM was used as a control. A PAMmer without TGG is indicated by +*. (C) SpyCas9 cleavage assay with *Ssp*I-linearized plasmid harboring all non-guanosine PAMs following *in vitro* annealing with PAMmer.

In vitro annealing with PAMmer facilitates SpyCas9 cleavage of target dsDNA lacking PAM

SpyCas9 is a RNA-guided endonuclease that uses RNA–DNA base-pairing to target dsDNA. Previous studies have shown that PAM recognition is required for SpyCas9–RNA-mediated binding and cleavage of DNA.[21] Interestingly, results generated following a substitution analysis demonstrated the importance of the positioning of guanosine in PAM in relation to SpyCas9 activity. As a result, SpyCas9 does not catalyze a reaction with target DNA in the absence of NGG. Indeed, this requirement may impede the extensive application of SpyCas9 in bioengineering. One means by which this obstacle might be addressed is through the introduction of an exogenous oligonucleotide containing the NGG sequence. This would facilitate the introduction of a sequence mimicking PAM allowing recognition and unwinding reactions that are normally conducted following the presence of the NGG sequence. Prior studies have shown that DNA substrates containing a cognate PAM that mismatch with the complementary strand are cleaved just as efficiently as fully matched PAM sequences, provided that there is a GG dinucleotide in the non-complementary strand.[12] In addition, single-strand DNA[21] and single-strand RNA targets[32] can be activated for cleavage by a separate PAMmer. We tested our hypothesis that dsDNA lacking an appropriate PAM sequence could be cleaved by SpyCas9 following the introduction of an exogenous PAMmer. To carry out this analysis, we performed a SpyCas9 cleavage assay on the *Ssp*I-linearized pUC18 plasmid containing PAM without the NGG sequence, following the addition of an exogenous PAMmer. In opposition to our expectations, the dsDNA that lacked the NGG sequence still harbored resistance to SpyCas9 cleavage, as was the case with the negative controls (Fig. 4B). This result was likely due to the shortage of template DNA base-pairing with PAMmer. To allow the unwinding of DNA double strands, we pre-treated template dsDNA and PAMmer together using a heat denaturation step. This resulted in the separation of the DNA double strands. The complementary strand was then annealed to PAMmer DNA before being mixed with SpyCas9–sgRNA. This procedure resulted in a hybrid dsDNA molecule with a non-congenital NGG PAM just downstream of the target protospacer. This, in theory, would permit SpyCas9 to perform its normal cleavage reaction. Significantly, an obvious cleavage was observed in the assay following the pre-treatment procedure (Fig. 4B). The results suggested that the pre-treatment procedure was effective in facilitating SpyCas9 functioning. However, weak cleavage was also observed in the absence of NGG in the PAMmer sequence and the same observation was also made even when no PAMmer was present (Fig. 4B). This indicated that the introduction of a PAMmer promoted SpyCas9 cleavage but was not an obligate factor in the reaction. The result demonstrated that pre-treatment could also result in weak destabilization of dsDNA and the NGG in PAMmer could act as an allosteric regulator of the Cas9–RNA nuclease. We suspect that the location of NGG in PAMmer would affect the cleavage efficiency, so we conduct the cleavage assay with NGG locating at 5′,3′ and in middle of PAMmer, even with extended PAMmer. As shown in Fig. 5, PAMmer with NGG locating at 3′ end has an obvious weak cleavage while others have similar cleavage. We also tested whether PAMmer binds non-complementary strand would also work well, the cleavage result shows a poor difference between PAMmers binding to two DNA strands.

We next sought to apply the strategy of pre-treatment and PAMmer binding to complementary strand with NGG locating at 5′ end to other DNA sequences lacking the NGG PAM sequence. We subsequently tested SpyCas9 cleavage taking the sequences in Fig. 3B as a control. The results demonstrated that dsDNA sequences that were not previously cleaved by SpyCas9 due to the absence of an appropriate PAM (Fig. 3B) were all cleaved following the strategy of pre-treatment and the introduction of PAMmer (Fig. 4C). This observation suggested that the effect of PAM recognition could be partially recovered by appropriate artificial measures.

As a programmable RNA-directed DNA endonuclease, SpyCas9 recognizes target sequences using a 20 nt guide sequence located at the 5′ terminus of the sgRNA and an appropriate PAM sequence. SpyCas9 can be effectively directed to its target sequence with PAM by designing the guide sequence of the sgRNA. It is therefore not surprising that SpyCas9 has become more preferable than restriction enzymes for certain applications due to lower restriction and higher specificity. Promisingly, SpyCas9 has also been recently employed as a programmable molecular tool for *in vitro* DNA manipulations.[33–35] Since the restriction of Cas9 is defined by the specific PAM, the ability to alter PAM specificities potentially expands Cas9-related bioengineering applications. Some efforts have been made to modify SpyCas9 activity to facilitate recognition of alternative PAM sequences through bacterial selection-based directed evolution and combinatorial design.[36] In this study, we observed that when the concentration of Cas9–sgRNA complex increased to ~250 nM, some previously undescribed PAM sequences such as NGT, NGC, NNNGG, and NNNNGG

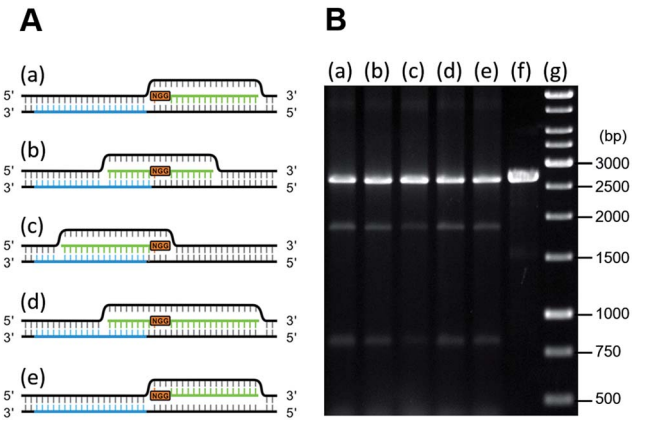

Fig. 5 Impact of NGG location in PAMmer and binding strand of PAMmer on SpyCas9 cleavage. (A) Schematic diagram of NGG locating at 5′,3′ end, in the middle of PAMmer, extended PAMmer and PAMmer binding to non-complementary strand. The target sequence in the complementary strand is blue and PAMmer is shown in green. (B) The SpyCas9 cleavage assay using a *Ssp*I-linearized plasmid harboring a TTC sequence in PAM and with PAMmer shown in (A).

mediated substantial Cas9 cleavage. This makes SpyCas9 a more convenient tool for *in vitro* use. Importantly, the introduction of a pre-treatment step along with the use of a PAMmer, facilitated the use of SpyCas9 in association with sites lacking orthodox PAM sequences, might shed light on artificial imitation of SpyCas9-mediated DNA cleavage. However, this strategy demonstrates that PAM effect can be replaced to some extent, it may be worrisome that some factors that destabilize dsDNA in cells can also increase the risk of off-target cleavage.

Conclusions

The recent developed CRISPR–Cas9 has provided extraordinary convenience for gene editing and regulation. Observations generated from cleavage assays and crystal structure of SpyCas9 have revealed the importance of guide RNA binding and PAM recognition for SpyCas9-mediated DNA cleavage. Here, we found that the cleavage site of complementary strand is completely determined by the location of guide RNA while the location of guide RNA and NGG PAM sequence jointly determine the cleavage site of non-complementary strand.

On the other hand, as a programmable RNA-directed DNA endonuclease, SpyCas9 can be effectively directed to its target sequence with PAM by designing the guide sequence of the sgRNA. Since the restriction of Cas9 is defined by the specific PAM, the ability to alter PAM specificities potentially expands Cas9-related bioengineering applications. We revealed for the first time that *in vitro* annealing of dsDNA with exogenous PAMmer stimulated SpyCas9 cleavage of target dsDNA without PAM, and this might open the way for enhancing its DNA editing abilities.

Author contributions

Y. S. and Y. G. conceived the experiments. Y. G. conducted the experiments. Y. G., Y. S. and Z. D. analyzed the results. Y. S. and Y. G. prepared the manuscript.

Acknowledgements

This work was supported by a grant from 973 Program (2012CB721005) from the Ministry of Science and Technology of China, and in part by the National Science Foundation of China (31270120) and the Translational Medical Research Fund of Wuhan University School of Medicine.

Notes and references

1 P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church, *Science*, 2013, **339**, 823–826.
2 L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang, *Science*, 2013, **339**, 819–823.
3 H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch, *Cell*, 2013, **153**, 910–918.
4 J. F. Li, J. E. Norville, J. Aach, M. McCormack, D. Zhang, J. Bush, G. M. Church and J. Sheen, *Nat. Biotechnol.*, 2013, **31**, 688–691.
5 L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin and W. A. Lim, *Cell*, 2013, **152**, 1173–1183.
6 D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang and L. A. Marraffini, *Nucleic Acids Res.*, 2013, **41**, 7429–7437.
7 L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman and L. S. Qi, *Cell*, 2013, **154**, 442–451.
8 M. L. Maeder, S. J. Linder, V. M. Cascio, Y. Fu, Q. H. Ho and J. K. Joung, *Nat. Methods*, 2013, **10**, 977–979.
9 P. Perez-Pinera, D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi, L. R. Polstein, P. I. Thakore, K. A. Glass, D. G. Ousterout, K. W. Leong, F. Guilak, G. E. Crawford, T. E. Reddy and C. A. Gersbach, *Nat. Methods*, 2013, **10**, 973–976.
10 P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang and G. M. Church, *Nat. Biotechnol.*, 2013, **31**, 833–838.
11 K. M. Esvelt, P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung and G. M. Church, *Nat. Methods*, 2013, **10**, 1116–1121.
12 M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science*, 2012, **337**, 816–821.
13 H. Chen, J. Choi and S. Bailey, *J. Biol. Chem.*, 2014, **289**, 13284–13294.
14 M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales and J. A. Doudna, *Science*, 2014, **343**, 1247997.
15 H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang and O. Nureki, *Cell*, 2014, **156**, 935–949.
16 C. Anders, O. Niewoehner, A. Duerst and M. Jinek, *Nature*, 2014, **513**, 569–573.
17 S. H. Sternberg, B. LaFrance, M. Kaplan and J. A. Doudna, *Nature*, 2015, **527**, 110–113.
18 L. Zheng, U. Baumann and J. L. Reymond, *Nucleic Acids Res.*, 2004, **32**, e115.
19 W. Jiang, D. Bikard, D. Cox, F. Zhang and L. A. Marraffini, *Nat. Biotechnol.*, 2013, **31**, 233–239.
20 P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao and F. Zhang, *Nat. Biotechnol.*, 2013, **31**, 827–832.
21 S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene and J. A. Doudna, *Nature*, 2014, **507**, 62–67.
22 R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath, *Science*, 2007, **315**, 1709–1712.
23 S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin and J. van der Oost, *Science*, 2008, **321**, 960–964.

24 J. E. Garneau, M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan and S. Moineau, *Nature*, 2010, **468**, 67–71.
25 B. Wiedenheft, S. H. Sternberg and J. A. Doudna, *Nature*, 2012, **482**, 331–338.
26 Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung and J. D. Sander, *Nat. Biotechnol.*, 2013, **31**, 822–826.
27 V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna and D. R. Liu, *Nat. Biotechnol.*, 2013, **31**, 839–843.
28 B. X. Fu, L. L. Hansen, K. L. Artiles, M. L. Nonet and A. Z. Fire, *Nucleic Acids Res.*, 2014, **42**, 13778–13787.
29 T. Karvelis, G. Gasiunas, J. Young, G. Bigelyte, A. Silanskas, M. Cigan and V. Siksnys, *Genome Biol.*, 2015, **16**, 253.
30 Y. Zhang, X. Ge, F. Yang, L. Zhang, J. Zheng, X. Tan, Z. B. Jin, J. Qu and F. Gu, *Sci. Rep.*, 2014, **4**, 5405.
31 C. E. Vejnar, M. A. Moreno-Mateos, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha and A. J. Giraldez, *Nat. Methods*, 2015, **12**, 982–988.
32 M. R. O'Connell, B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan and J. A. Doudna, *Nature*, 2014, **516**, 263–266.
33 W. Jiang, X. Zhao, T. Gabrieli, C. Lou, Y. Ebenstein and T. F. Zhu, *Nat. Commun.*, 2015, **6**, 8101.
34 N. C. Lee, V. Larionov and N. Kouprina, *Nucleic Acids Res.*, 2015, **43**, e55.
35 J. W. Wang, A. Wang, K. Li, B. Wang, S. Jin, M. Reiser and R. F. Lockey, *BioTechniques*, 2015, **58**, 161–170.
36 B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng, A. P. Gonzales, Z. Li, R. T. Peterson, J. R. Yeh, M. J. Aryee and J. K. Joung, *Nature*, 2015, **523**, 481–485.

