Molecular Microbiology (2018) 110(3), 484–497

doi:10.1111/mmi.14118
First Published online 5 October 2018

Identification of a conserved DNA sulfur recognition domain by characterizing the phosphorothioate-specific endonuclease SprMcrA from *Streptomyces pristinaespiralis*

Hao Yu,[1] Guang Liu,[1] Gong Zhao,[1] Wenyue Hu,[1] Geng Wu,[1] Zixin Deng[1] and Xinyi He [1]*

State Key Laboratory of Microbial Metabolism, Joint International Laboratory on Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China.

mutation of the central residue in HRH resulted in 30-fold enhancement in cleavage activity of phosphorothioate DNA and altered the cleavage efficiency at some sites, whereas mutation of both His residues abolished restriction activity. This is the first report of a recognition domain for phosphorothioate DNA and phosphorothioate-dependent and sequence-specific restriction activity.

Summary

Streptomyces **species have been valuable models for understanding the phenomenon of DNA phosphorothioation in which sulfur replaces a non-bridging oxygen in the phosphate backbone of DNA. We previously reported that the restriction endonuclease ScoMcrA from *Streptomyces coelicolor* cleaves phosphorothioate DNA and Dcm-methylated DNA at sites 16–28 nucleotides away from the modification sites. However, cleavage of modified DNA by ScoMcrA is always incomplete and accompanied by severe promiscuous activity on unmodified DNA. These features complicate the studies of recognition and cleavage of phosphorothioate DNA. For these reasons, we here characterized SprMcrA from *Streptomyces pristinaespiralis*, a much smaller homolog of ScoMcrA with a rare HRH motif, a variant of the HNH motif that forms the catalytic center of these endonucleases. The sulfur-binding domain of SprMcrA and its phosphorothioation recognition site were determined. Compared to ScoMcrA, SprMcrA has higher specificity in discerning phosphorothioate DNA from unmodified DNA, and this enzyme generally cuts both strands at a distance of 11–14 nucleotides from the 5′ side of the recognition site. The HRH/HNH motif has its own sequence specificity in DNA hydrolysis, leading to failure of cleavage at some phosphorothioated sites. An R248N**

Accepted 1 September 2018. *For correspondence. E-mail xyhe@sjtu.edu.cn; Tel. (+86) 2162932943 extension 2131; Fax (86)2162932418.

Introduction

DNA modifications are widely present in many forms of life. One of the most common epigenetic modifications in eukaryotes is 5-methyl cytosine (5mC), which is implicated in repressed chromatin state, inhibition of transcription and genome stability (Bird, 2002; Tost, 2010). In prokaryotes, DNA modification is often coupled with a restriction endonuclease (REase) to form a restriction and modification system (R–M). For Type I-III R–M systems, DNA modification generally protects the target sequence from cleavage by the cognate restriction enzyme in the host while the REases destroy DNA of invading phages or mobile genetic elements that lack modification or have different modification patterns. By contrast, modification-dependent REases cut DNA only when it is modified. This family of REases currently consists of type IIM and type IV categories. Type IIM REase cleaves DNA at a constant distance away from the modification site, i.e. MspJI family proteins (Cohen-Karni *et al*., 2011; Horton *et al*., 2014), or exactly at the modification site, i.e. DpnI (Siwek *et al*., 2012). By contrast, Type IV REase cuts DNA at a variable distance from the modification site i.e. McrBC (Sutherland *et al*., 1992), PvuRts1I (Szwagierczak *et al*., 2011) and ScoMcrA (Liu *et al*., 2010).

In addition to DNA methylation and its derivatives of hypermodified bases (Swinton *et al*., 1983; Borst and Sabatini, 2008), a large number of bacterial genomes (> 1349 strains) (Tong *et al*., 2018) encode phosphorothioate (PT) modification systems, which are generally encoded by gene clusters consisting of the five genes *dndA* through *dndE* (Zhou *et al*., 2005). Most of the *dnd*

gene clusters are located in mobile genetic elements (He *et al*., 2007; Ou *et al*., 2009; Wang *et al*., 2011), very similar to the predicted mobility of typical DNA restriction-modification systems (Furuta *et al*., 2010). A few bacterial PT modification systems modify the DNA backbone by replacing a non-bridging oxygen in the phosphate with a sulfur atom in a DNA sequence-specific way (Wang *et al*., 2007; Cao *et al*., 2014). In bacteria, the phosphate–sulfur bond in the PT linkage adopts the R_P configuration rather than its stereoisomer of S_P [11]. Four types of consensus sequences for PT modification have been determined in bacteria, namely, G_{PS}GCC from *Streptomyces lividans* 66, G_{PS}AAC/G_{PS}TTC from *Escherichia coli* B7A and *Salmonella enterica* 87, G_{PS}ATC from *Bermanella marisrubri* RED65 and C_{PS}CA from *Vibrio cyclitrophicus* FF75 (Wang *et al*., 2011). However, genomic mapping revealed that target sequences on genomes or plasmids were not equally phosphorothioated by Dnd systems (Cao *et al*., 2014). Interestingly, DNA PT-modification has been implicated in conferring oxidation resistance to the host bacteria (Xie *et al*., 2012; Kellner *et al*., 2017) and influencing the global transcriptional response (Gan *et al*., 2014).

DNA modification can serve as a signal for recruiting DNA binding proteins that function as 'readers' for these epigenetic modifications in varied DNA sequences. The SET and RING-associated (SRA) domain is specific for 5mC binding, where it 'flips out' 5mC from the DNA helix and positions it in an accommodation pocket (Arita *et al*., 2008; Avvakumov *et al*., 2008; Hashimoto *et al*., 2008). SRA domains in eukaryotes are usually fused or associated with other functional domains, thus linking varied cellular processes with DNA 5mC methylation (Hashimoto *et al*., 2009; Rajakumara *et al*., 2011). In prokaryotes, SRA or SRA-like domains are often fused or associated with DNA catalytic motifs of REases (Iyer *et al*., 2011), to constitute DNA modification-dependent REases.

Given the existence of a sulfur-modified DNA backbone in many bacteria, it could be hypothesized that there are modification-dependent REases for recognition of PT DNA modification. In this regard, we previously reported the serendipitous identification of the type IV REase ScoMcrA from *Streptomyces coelicolor* A(3)2, which blocks the establishment of *dnd*-mediated DNA phosphorothioation and *dcm*-mediated 5mC modification in the same bacterial cell (Liu *et al*., 2010); therefore, ScoMcrA was postulated to contain recognition domains for the two different types of DNA modification (Loenen and Raleigh, 2014). Although ScoMcrA cleaves 5mC DNA generated by Dcm, no SRA domains specific for 5mC recognition could be identified by using ScoMcrA as the query. The PT recognition domain on ScoMcrA also could not be discerned simply based on sequence comparison.

ScoMcrA was named based on the presence of an EcoKMcrA-like HNH motif in its C-terminus (Roberts *et al*., 2015). The HNH motif is characterized by the presence of two highly conserved His residues and usually one Asn residue, which are located in a spatially conserved ββα-metal finger fold (Roberts *et al*., 2015). The first His acts as a general base to activate a water molecule for nucleophilic attack of the scissile phosphodiester bond. For hydrolysis of DNA, a divalent metal ion is bound in the active center by interaction with the side chain of the second His residue, as well as the residue immediately N-terminal to the first His residue. The bound metal can stabilize the transition state and coordinate a water molecule to protonate the leaving group (Pommer *et al*., 1999).

Some HNH REases exhibit promiscuous cleavage activity under altered conditions, such as low ionic strength, elevated pH, or when Mg^{2+} is substituted by Mn^{2+} (Wei *et al*., 2008). ScoMcrA also requires the divalent metal ion Mn^{2+} to activate its cleavage activity and a pH of 9 to achieve optimal efficiency. However, the DNA modification-dependent cleavage activity is relatively weak even under optimal conditions as evidenced by the persistently incomplete digestion of modified DNA substrates (Liu *et al*., 2010). ScoMcrA makes multiple cleavages at sites 16–28 nucleotides away from the palindromic modification sequence G_{PS}GCC (Liu *et al*., 2010), and therefore it is very difficult to determine if a DNA molecule is cut once or more. Moreover, ScoMcrA displays severe promiscuous cleavage activity on unmodified DNA if the incubation time is extended beyond 5 min at some protein concentrations (Liu *et al*., 2010).

In order to avoid the disadvantages of ScoMcrA for analyzing DNA cleavage activity, we here characterized SprMcrA, a ScoMcrA homolog of much smaller size from *Streptomyces pristinaespiralis,* and determined its recognition domain for PT-modified DNA. Compared with ScoMcrA, SprMcrA displays a higher cleavage specificity for PT-modified versus unmodified DNA. SprMcrA also uses a C-terminal HNH motif variant, HRH, to cleave DNA mainly at a distance of N13/N14 or N11/N12 on the 5′ side of the PT linkage. Weak nickase activity and flexible cleavage of PT DNA at several sites by SprMcrA were also observed. In addition, we provide evidence that the HRH (or HNH) motif has recognition sequence specificity in DNA hydrolysis.

Results

Identification of a sulfur-binding domain (SBD) in ScoMcrA family proteins

Using ScoMcrA as the query, a large group of homologs with significant similarity (*E*-value of < $8e^{-21}$ for the 100th homolog) can be easily identified from bacterial genomes. A conserved domain of unknown function,

which spans from 105 to 270 amino acid (aa) residues N-terminal to the HNH motif in ScoMcrA, was identified in the related proteins. The homologs were arbitrarily classified into three groups based on the location of this unknown domain in each protein (Fig. 1B, and Supplementary Sequence Data). ScoMcrA represents the longest protein. The sequence of ScoMcrA does not contain a region with significant similarity to the known SRA domain, which is the dominant 'reader' responsible for 5mC, and the recognition domain for PT-DNA was also not determined in ScoMcrA.

SprMcrA from *S. pristinaespiralis* represents the third group of homologs, which lacks the N-terminal region present in the other two groups (Fig. 1B). SprMcrA is missing 159 aa N-terminal to the HNH motif (Supplementary Sequence Data) and was chosen to study its specificity and activity on modified DNA. The ability of SprMcrA to restrict plasmids bearing *dnd* genes of *E. coli* B7A or genes for 5mC modification was examined using control BL21(DE3) cells and BL21(DE3) cells expressing *sprMcrA* (Fig. 1C). The cells expressing *sprMcrA* exhibited 1000-fold greater restriction activity against the establishment of PT modification by the *dnd* gene cluster from *E. coli* B7A but showed no increase in restriction activity to the establishment of 5mC modification (Fig. 1C). These findings suggest that the unknown domain shared by SprMcrA and ScoMcrA (Fig. 1B, dot pattern region) is the PT-DNA binding domain (SBD) and also imply that the 159-aa region missing from SprMcrA governs the specific recognition of 5mC. The approximate regions for the SBD and the 5mC recognition domains are shown, respectively, as pink and yellow text in the Supplementary Sequence Data.

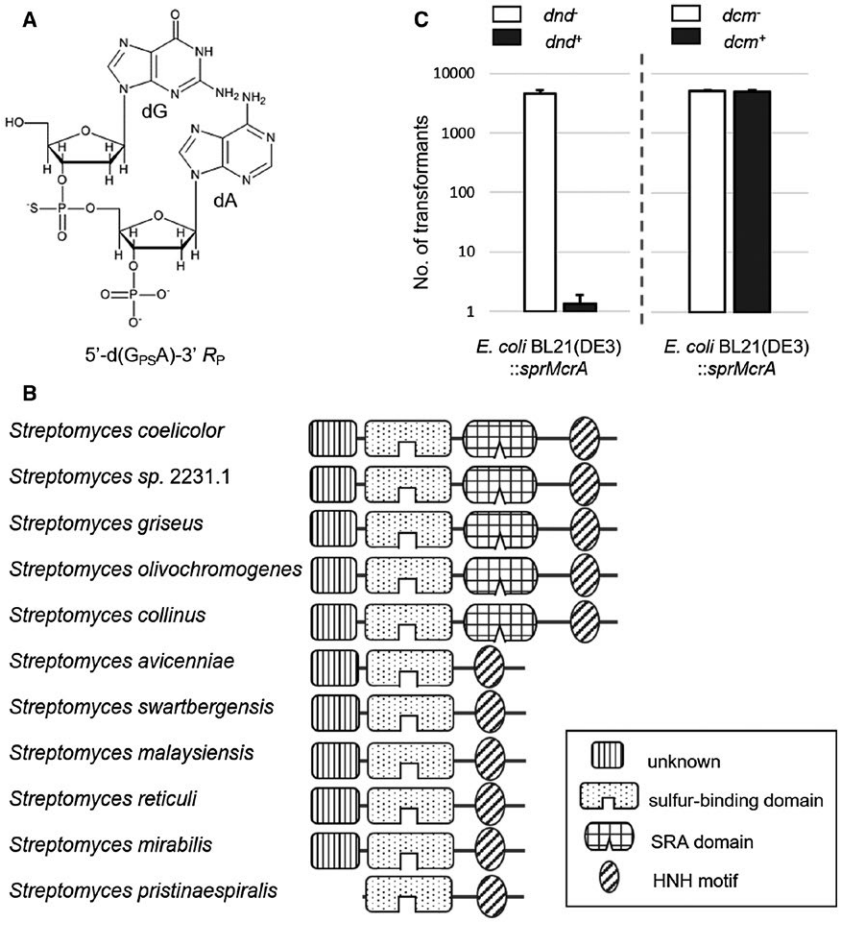

Fig. 1. A. The chemical structure of DNA sulfur modification which is sequence-selective and stereo-specific phosphorothioation of the DNA backbone.
B. Domain organization of ScoMcrA homologs from *Streptomyces*. Group one has a leader peptide, SBD, SRA-like and HNH domain; group two lacks the SRA-like domain; and group three lacks the head and SRA-like domain. Protein accession numbers and aa sequences are included in Supplementary Materials.
C. *sprMcrA* expression in *Escherichia coli* BL21(DE3) restricts the transfer of the *dnd* gene cluster of *E. coli* B7A but not of the *dcm* gene. Frequencies of transformation into the *sprMcrA*-expressing host are shown for (left side) the empty pACYCDute vector (PT^-, white bar) and pYH9 harboring the *dnd* of *E. coli* B7A (PT^+, black bar) and (right side) the pACYCDute vector (dcm^-, white bar) and pYH10 harboring the *dcm* gene (dcm^+, black bar).

To help identify the potential sulfur recognition motif in the SBD domain, a multiple alignment of the 100 closest homologs to SprMcrA$_{1–165}$ was prepared, revealing a motif composed of P$_{79}$-L$_{83}$-W$_{89}$, with each residue showing 94~100% conservation among the 100 homologs (Fig. S1). Mutation of either of L83A or W89A on SBD and full-length SprMcrA leads to very low expression of the mutant protein (<0.5 mg L^{-1}, Fig. S2A), implying that these two 100% conserved residues might play a role to keep the integrity of the protein. By contrast, P79A is expressed at a comparable level to wild-type protein (Fig. S2A, upper panel). A fluorescence polarization assay revealed that a mutant SprMcrA protein containing the P79A mutation shows a 14.5-fold decreased affinity for PT-DNA when compared with the wild-type protein (Fig. S2B). However, mutation of L83A and W89A leads to complete no expression (Fig. S2A). However, evidence for direct association of the motif residues with the DNA sulfur will require protein structural analysis.

DNA sequence and stereospecificity in recognition of PT-DNA

Gel-filtration analysis showed that purified SprMcrA forms a homodimer in solution (Fig. S3). To demonstrate the ability of SprMcrA to bind PT-DNA, we synthesized and generated eighteen 10-bp DNA oligonucleotides, each of which contained a central 3 or 4-nt consensus sequence representing the four types of natural DNA PT modification sites in bacteria (Table S1) (Cao *et al.*, 2014). According to the results (Fig. 2A), unmodified DNA or PT-DNA oligonucleotides of the S_P stereoisomer

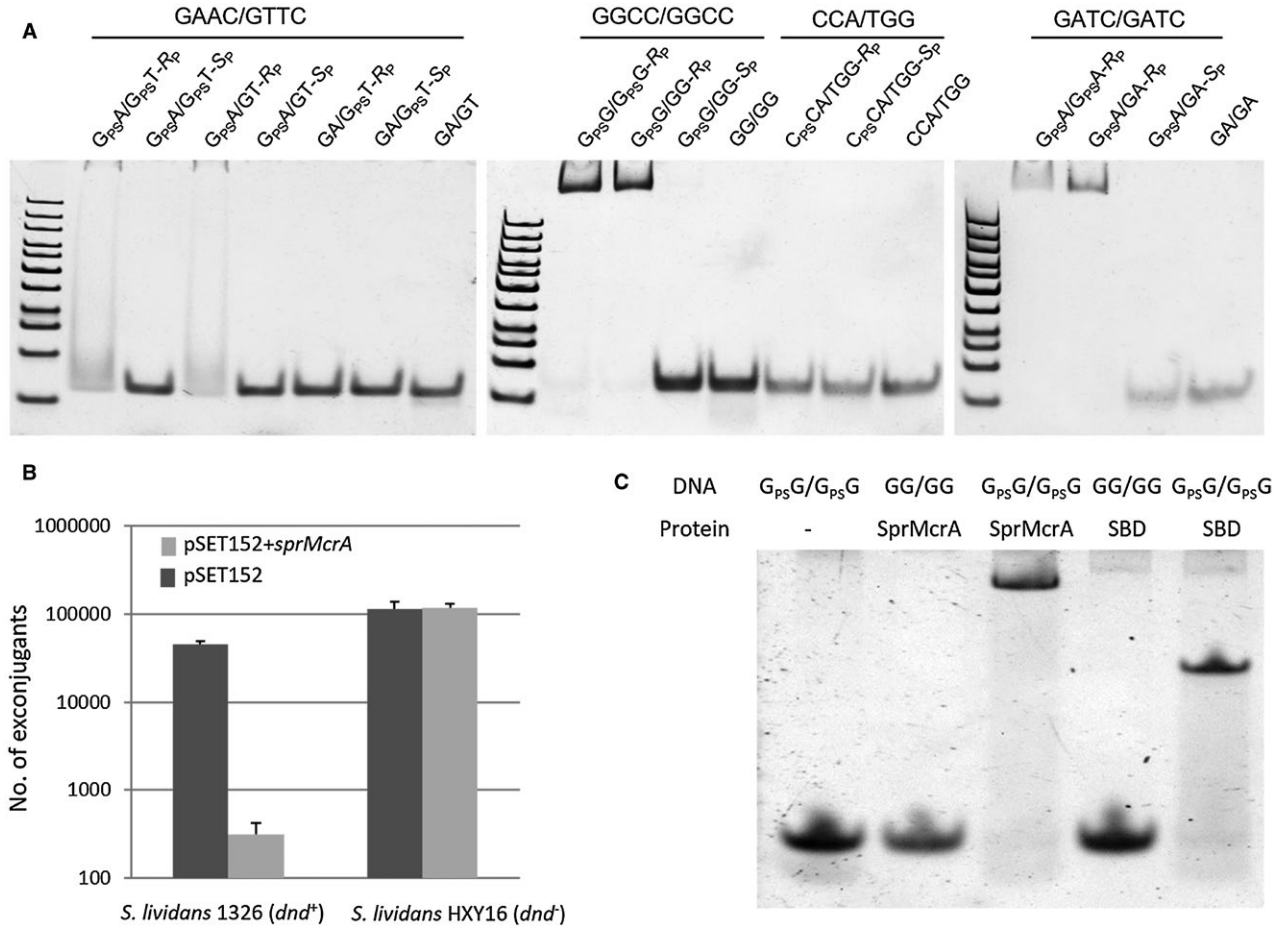

Fig. 2. A. EMSA analysis with SprMcrA and PT-DNA duplexes. Oligonucleotides of 10 nt in length with the PT modifications of G$_{PS}$GCC, G$_{PS}$AAC, G$_{PS}$TTC, G$_{PS}$ATC or C$_{PS}$CA were synthesized and annealed to generate fully, hemi- and non-PT duplexes. SprMcrA shows affinity for G$_{PS}$GCC, G$_{PS}$ATC and G$_{PS}$AAC, but not for G$_{PS}$TTC and C$_{PS}$CA.
B. Uptake efficiency of SprMcrA by the PT$^+$ *Streptomyces* host decreases by > 10^3-fold compared with PT$^-$ host. A plasmid carrying the *sprMcrA* gene with its own promoter was introduced via conjugation into *S. lividans* 1326 (PT$^+$) and *S. lividans* HXY16 (lacking the genomic island that contains the *dnd* cluster, PT$^-$). The restriction activity on SprMcrA by the host was calculated by comparing the exconjugant numbers for pSET152 + *sprMcrA* and pSET152. The PT modification site is GpsGCC in *S. lividans* 1326. The cloning vector pSET152 without *sprMcrA* served as a control.
C. EMSA analysis of SprMcrA and its SBD with PT-DNA duplexes. SBD (1–165 aa) shows an affinity similar to that of full-length SprMcrA for 10bp PT-DNA duplexes with the G$_{PS}$GCC modification.

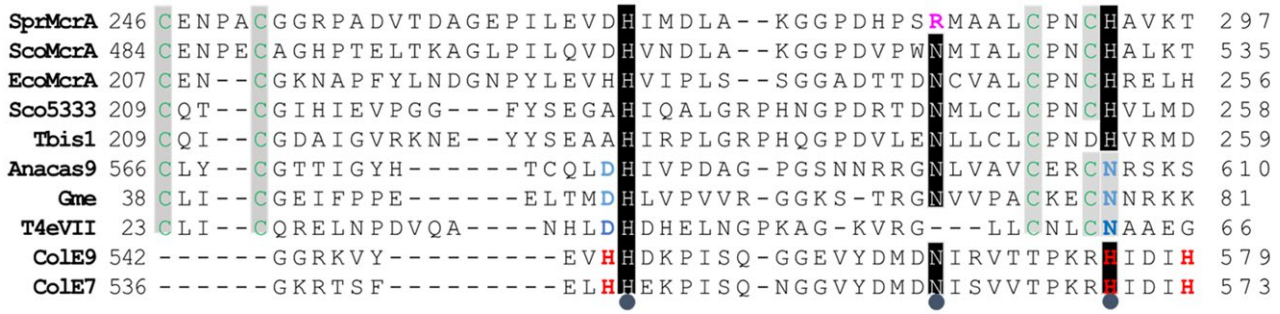

Fig. 3. Multiple sequence alignment of several known HNH motifs of restriction endonucleases. Residues in green designate the CCCC or CCCH type of zinc finger motif. Filled black circles represent the conserved HNH motifs. The D and N marked in blue are active sites of the HNH motifs in AnaCas9, Gme and T4eVII that bind to divalent ions. A red 'H' designates the divalent ion-binding residues of the HNH motifs in McrA, ColE9 and ColE7. [Colour figure can be viewed at wileyonlinelibrary.com]

were clearly not associated with SprMcrA as evidenced by the lack of shifting. Like ScoMcrA, SprMcrA recognizes PT-DNA oligonucleotides of the R_P stereoisomer, which is the form adopted by all natural DNA PT modifications *in vivo*. SprMcrA bound to G_{PS}AAC/G_{PS}TTC, G_{PS}ATC/G_{PS}ATC and G_{PS}GCC/G_{PS}GCC (Fig. 2A), and to hemi-PT DNA (G_{PS}AAC/GTTC, G_{PS}ATC/GATC and G_{PS}GCC/GGCC) except for GAAC/G_{PS}TTC and C_{PS}CA/TGG.

Consistently, an *in vivo* restriction activity assay showed that SprMcrA strongly blocked the uptake of the *dnd* gene cluster that governs PT modification at the sites G_{PS}AAC/G_{PS}TTC (Fig. 1C) and G_{PS}GCC/G_{PS}GCC (Fig. 2B). It seems that SprMcrA recognizes G_{PS}R (R = purine, A or G) on just one DNA strand. In parallel, the SBD domain of SprMcrA, ranging from aa positions 1-165, was purified and assayed for its association with the PT-DNA oligonucleotide G_{PS}GCC/G_{PS}GCC. SBD specifically bound to PT-DNA (Fig. 2C), further supporting SBD as being the PT-DNA recognition domain.

Cofactor requirement and optimal pH for SprMcrA cleavage of PT-DNA

Instead of a typical HNH motif, SprMcrA has HRH at its C-terminus, and the location of the two His residues aligns well with their equivalents in other functional homologs (Fig. 3). In general, for a given HNH REase, a divalent metal ion is coordinated as cofactor in the catalytic center for DNA hydrolysis (Galburt *et al*., 1999; Pommer *et al*., 1999; Drouin *et al*., 2000; Keeble *et al*., 2002; Ku *et al*., 2002; Kriukiene, 2006; Chan *et al*., 2010; Vasu *et al*., 2013). Accordingly, purified SprMcrA displayed no DNA cleavage activity in the absence of metal ion. Seven divalent metal ions of varied concentrations were tested in SprMcrA digestion of PT-modified genomic DNA from *E. coli* B7A, and 1–10 mM Mn^{2+} had an obviously stimulating effect on DNA hydrolysis (Fig. S4A), generating DNA fragments mainly above 1 kb in size. Co^{2+} activated partial cleavage, but cleavage was not evident with the other metal ions (Fig. S4A).

SprMcrA contains a zinc-specific finger that is composed of a rare CxxxxC motif, rather than the usual CxxC and this motif is located 18 aa residues N-terminal to the first His of the HNH motif. A CxxC motif is located immediately next to the second His (Fig. 3). Given these features, we postulated that the HNH motif of SprMcrA might coordinate two metal ions in a way similar to that of PacI (Shen *et al*., 2010), T4eVII (Raaijmakers *et al*., 1999) and KpnI (Saravanan *et al*., 2007). We, thus, tested the effect of the simultaneous presence of Zn^{2+} and Mn^{2+} on the cleavage of sulfur-modified pUC18 DNA and found that equimolar Zn^{2+} can completely suppress DNA hydrolysis stimulated by Mn^{2+} (Fig. S4B). For DNA hydrolysis, both His residues are deprotonated to perform their functions (Eastberg *et al*., 2007), and thus DNA cleavage efficiency is expected to vary with pH. In the case of SprMcrA, optimal DNA cleavage activity was observed at pH 6.5, with a complete loss of activity at pH values above 8 (Fig. S5). In contrast, some HNH REases, such as ScoMcrA (Liu *et al*., 2010) and ColE9 (Pommer *et al*., 2001), showed optimal activity at pH of 8.5–9.0 and pH 9 respectively. It is unclear why SprMcrA requires an acidic pH value that would not seem to favor deprotonation of the imidazole ring of H270, but structural analyses may shed light on the different pH requirements of SprMcrA and ScoMcrA.

High specificity and low star activity of SprMcrA for PT-DNA

We previously reported that ScoMcrA displayed low toxicity to host cells without 5mC or PT modification and that, compared to control cells, host cells expressing ScoMcrA had >1000-fold stronger restriction activity against the uptake of *dcm* or *dnd* genes. However, ScoMcrA at a concentration of 25 ng μl^{-1} (equal to 380 nM) cleaved unmodified DNA nonspecifically when the incubation time was extended beyond 5 min (Liu

et al., 2010). To measure the specificity of SprMcrA on PT-modified DNA, PT-modified and unmodified linearized pUC18 DNA was incubated with increasing concentrations of SprMcrA for 1 h (Fig. S6). With PT-DNA, 40 nM SprMcrA gave rise to distinct fragments that remained unchanged until the concentration was increased to 640 nM, when low star activity was detected. In terms of the REase fidelity index, SprMcrA displayed 16-fold (FI: 640/40 = 16) specificity (Wei *et al*., 2008), determined by comparing the concentrations leading to onset of cleavage of unmodified and modified DNA (Fig. S6).

SprMcrA was also measured for its nickase activity as a number of HNH endonucleases, i.e. Gme_0936 (Xu *et al*., 2013) and Sco5333 (Han *et al*., 2015), make a single cut on one DNA strand. At concentrations of 20 nM–320 nM, SprMcrA showed very weak DNA nicking activity (Fig. S7) as evidenced by a slight increase in the intensity of the band corresponding to the open circular form of pUC18. By comparison, double-strand cleavage activity was more obvious as there was a gradual accumulation of linear PT-pUC18 with increasing concentrations of SprMcrA up to 320 nM (Fig. S7).

Determination of SprMcrA cleavage sites

For a modification-dependent REase, the position of the cleavage site with respect to the modification site is a characteristic feature. To determine the cleavage site for SprMcrA, we choose pUC18 as the substrate as it contains 18 sites of GAAC/GTTC or GTTC/GAAC for *in vivo* PT modification, with four of them, at nt 804, 1479, 1676 and 2584 (numbering based on GenBank accession number L09136 DNA sequence), known to be unmodified or modified at very low frequency based on the mapping statistics of PT sites of pBluescript SK$^+$ (Cao *et al*., 2014). Ten regions of DNA fragments were individually purified from SprMcrA-digested PT-pUC18 (Fig. S8) and subjected to runoff sequencing (Fig. S9; for primers used, see Table S3). For eight free-standing modification sites (Fig. 4), the staggered dsDNA cleavage occurred on the 5′ side of G_{PS}AAC and generated a one-nucleotide 3′ overhang on the strand of G_{PS}TTC. The distances between the cleavage site and the G_{PS}A linkage fell mainly into two ranges, i.e. N13/N14 or N11/N12. However, an exception was the free-standing site 1994, for which cleavage occurred N19/N20 and N18/

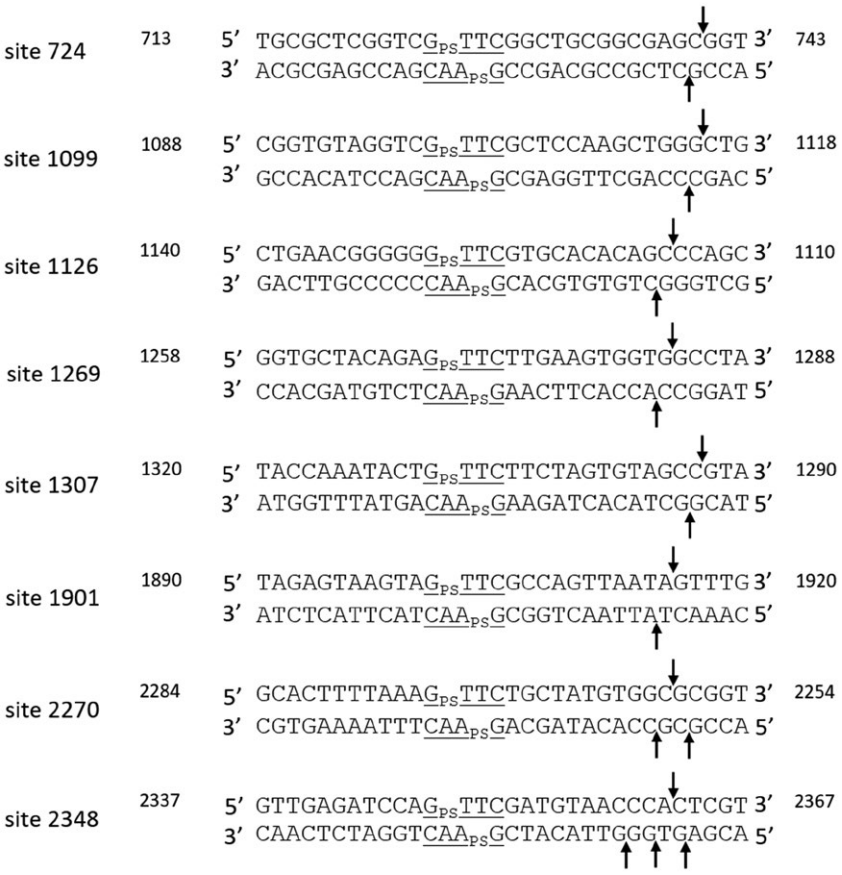

Fig. 4. Characterization of the cleavage sites of PT-pUC18 by SprMcrA. Cleavage sites were detected by the runoff sequencing method. The modified sites were named according to the location of the modified nucleotide on the plasmid. The preferred phosphorothioated sequence G_{PS}AAC/G_{PS}TTC is underlined. Vertical arrows indicate the cleavage sites near the phosphodiester linkage detected by sequencing.

N19 from the 5′ side from G_{PS}TTC. In addition to the four unmodified sites, no cleavage was detected in the vicinity of the sites 837 and 994. PT-DNA duplexes of the same DNA sequences as these undigested sites were synthesized as substrates. These PT-DNA oligonucleotides were all shifted by SprMcrA in EMSAs (Fig. S10A). However, no cleavage of these PT-DNA oligonucleotides by SprMcrA was observed (Fig. S10B), demonstrating that binding to PT-DNA by SBD does not necessarily elicit DNA cleavage and implying that the catalytic HNH motif has its own sequence specificity in DNA cleavage.

Two regions on pUC18, spanning from nt 1070 to 1150 and from nt 2255 to 2371, have pairs of head-to-head or head-to-tail PT sites and in these regions, orientation and distance of the cleavage site to the PT site did not match with the identified pattern for the eight free-standing PT sites (Fig. S11). Cleavage displayed versatility with respect to PT sites, for example, cleavages between site

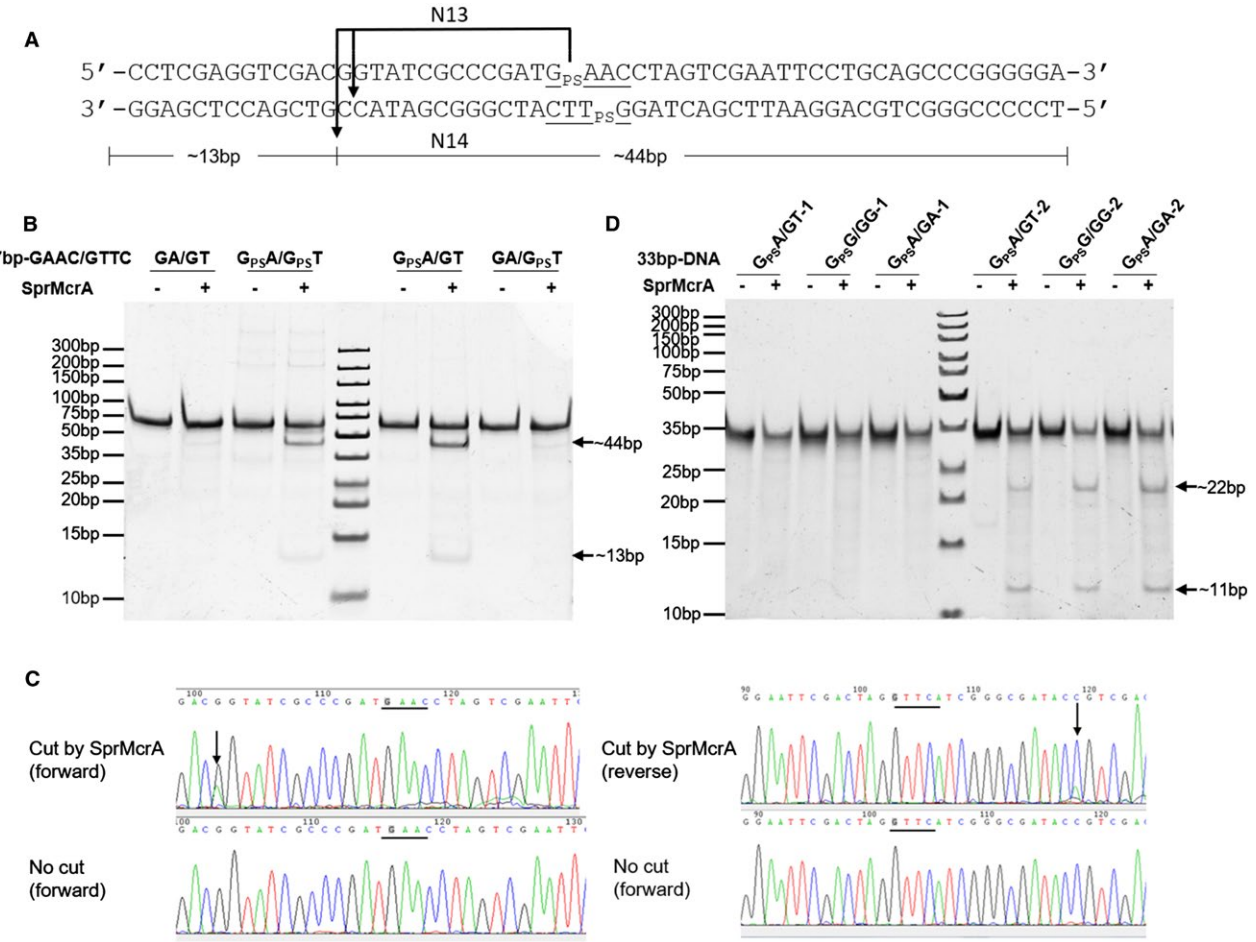

Fig. 5. *In vitro* cleavage of synthetic PT-DNA duplexes by SprMcrA.
A. Sequence of the 57-bp PT-DNA duplex used for determination of the cleavage position. The sequence G_{PS}AAC/G_{PS}TTC underlined denotes the PT sites that are the same as the consensus sequence determined in *Escherichia coli* B7A. Cleavage at the position marked by arrows will result in two staggered fragments of ~44 bp (N43/N44) and ~13 bp (N14/N13).
B. Cleavage of the 57-bp PT-DNA duplex by SprMcrA. The oligonucleotides were phosphorothioated at sites G_{PS}AAC/G_{PS}TTC, G_{PS}AAC/GTTC, or GAAC/G_{PS}TTC and GAAC/GTTC was used as a control. Two fragments of about 44 bp and 13 bp were generated in the cleavage of G_{PS}AAC/G_{PS}TTC and G_{PS}AAC/GTTC. The cleavage position is about N13/N14 away from the G_{PT}A linkage.
C. Runoff sequencing of the cleavage site on the 57-bp duplexes. Modified sites are underlined beneath the sequences. Double peaks (A/C, A/G) marked by a black arrow indicate a nick on the template strand as the extrinsic A was added by Taq DNA polymerase. Undigested plasmid was used as a control in sequencing.
D. Cleavage of 33-bp hemi-PT duplexes with G_{PS}AAC/GTTC, G_{PS}GCC/GGCC and G_{PS}ATC/GATC sites. Two sets of double-stranded 33-bp hemi-PT duplexes (Table S1) were used to determine the orientation of the cleavage site. The PT linkage is 6 nt away from the 5′ end of the PT-DNA duplexes G_{PS}AAC-33bp-1, G_{PS}GCC-33bp-1 and G_{PS}ATC-33bp-1 (left side of gel). The PT linkage is 25 nt away from the 5′ end of DNA duplexes G_{PS}AAC-33bp-2, G_{PS}GCC-33bp-2 and G_{PS}ATC-33bp-2 (right side of gel). Cleavage only occurred on G_{PS}AAC-33bp-2, G_{PS}GCC-33bp-2 and G_{PS}ATC-33bp-2 at a distance of N12/N11 or N21/N22 base pairs from the respective GpsG or GpsA linkage on the bottom strand. [Colour figure can be viewed at wileyonlinelibrary.com]

2270 and 2300 is not on the 5′ side of G_{PS}AAC. Cleavage occurred at a distance of N22/N23 from the 5′ side of G_{PS}AAC site 2300, much longer than N11/N12 or N13/N14. In addition, four distinct nick sites were detected in the region spanning from 1070 to 1150 in the bottom strand (Fig. S11), in accordance with the result for the nickase activity assay (Fig. S7).

We postulated that *in vivo* DNA PT modification efficiency in different DNA sequence contexts, resulting in either fully or hemi-PT-modified DNA, may affect cleavage by SprMcrA. Therefore, four 57-bp DNA duplexes of the same sequence with full- or hemi-PT modifications at the central GAAC/GTTC site (Fig. 5A) were evaluated as cleavage substrates for SprMcrA. Two DNA fragments of ~13 bp and ~44 bp were generated by SprMcrA digestion of the 57-bp DNA duplexes containing G_{PS}AAC/G_{PS}TTC and G_{PS}AAC/GTTC (Fig. 5B). The 57-bp fragment was cloned into pUC18 and transformed into *E. coli* B7A for runoff sequencing; the cleavage site was determined to be N13/N14 from the 5′ side of G_{PS}AAC (Fig. 5C). We then replaced G_{PS}AAC with G_{PS}GCC or G_{PS}ATC in 33-bp DNA duplexes with a sequence identical to that of the 5′ terminus of the 57-bp DNA duplex. Cleavage of the three PT-DNA oligonucleotides by SprMcrA showed the same pattern (Fig. 5D), implying that different PT modification sites in this particular sequence do not alter the cleavage sites for SprMcrA.

SprMcrA variant R284N has increased cleavage activity and altered sequence specificity

SprMcrA is annotated as a HNH nuclease, a type of nuclease characterized by a conserved motif composed of the three catalytic residues His-Asn-His or His-Asn-Asn. However, the second residue of this motif in SprMcrA is Arg instead of Asn, whereas most homologs of SprMcrA contain a glutamine (Gln) instead of Asn at this position (Fig. S1). To investigate the importance of this motif to its activity, three variants of SprMcrA with the mutations H270A, R284A or H293A were constructed by site-directed mutagenesis, and plasmids expressing these mutated proteins were transformed into *E. coli* BL21(DE3) to compare their restriction activity on an incoming *dnd* gene cluster (Fig. S12). The R284A mutation had minimal effect as compared to wild-type SprMcrA, whereas variants H270A and H293A abolished the *in vivo* restriction activity of SprMcrA. Additionally, when the Arg residue was changed to Asn, PT-DNA cleavage activity of $SprMcrA_{R284N}$ increased by about 30-fold compared to the activity of SprMcrA (Fig. 6A). Interestingly, the R284N mutation slightly altered the cleavage specificity as evidenced by the change in cleaved bands. For example, band 1 (indicated by white arrow, Fig. 6A), corresponding to cleavage at site 724, disappeared with the R284N mutation. To confirm this finding, we synthesized a 33-bp PT-DNA oligonucleotide (Table S1) that included the sequence flanking site 724 to use as substrate for wild-type SprMcrA and the mutant protein R284N. Consistent with the disappearance of the band corresponding to cleavage at site 724 in the digestion of PT-pUC18 by $SprMcrA_{R284N}$, the R284N mutation of the HNH motif abolished the cleavage activity at site 724 that was present with wild-type SprMcrA (Fig. 6B).

Discussion

The type IV HNH REase ScoMcrA was postulated to contain two DNA recognition domains as it cuts DNA in a PT- and 5mC-dependent manner. However, the

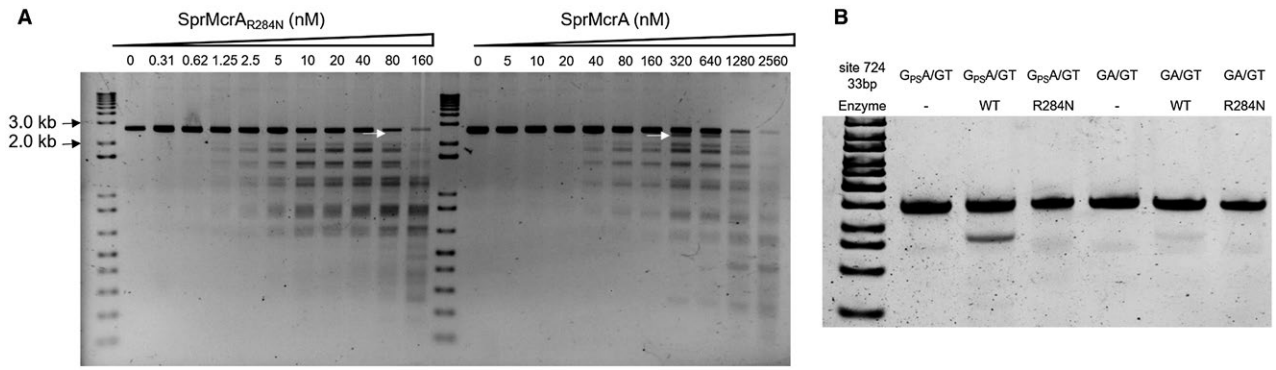

Fig. 6. Comparison of SprMcrA-mediated and mutant $SprMcrA_{R284N}$-mediated *in vitro* cleavage of phosphorothioated plasmid DNA.
A. PT-pUC18 plasmid DNA isolated from *Escherichia coli* B7A (*dnd*$^+$) was linearized by EcoRI and used as the substrate with increasing concentrations of SprMcrA and mutant protein $SprMcrA_{R284N}$. Incubation was performed for 1 h at 30°C. The white arrow marks the position of the band missing in the digestion with mutated protein. This band corresponds to cleavage at site 724. Another band of 288 bp resulting from cleavage at site 724 is too faint to detect.
B. Cleavage of 33-bp PT-duplexes containing the same sequence of site 724. PT and non-PT duplexes of site 724 were synthesized and incubated in the reaction buffer with wild-type SprMcrA or $SprMcrA_{R284N}$. Cleavage only occurred on PT-DNA duplexes incubated with wild-type SprMcrA.

domain in charge of recognizing the PT modification had not been characterized. Identifying the domain governing 5mC recognition in ScoMcrA is also of interest as there is no detectable homology to the known SRA domain. For these purposes, SprMcrA, a much simpler homolog (326 aa) of ScoMcrA, was selected to study its activity and specificity. *In vivo,* SprMcrA displayed restriction activity that prevented the establishment of PT modification but not the establishment of 5mC modification by Dcm. This PT-specific restriction activity enabled the identification of the N-terminal region of SprMcrA as the recognition domain for PT-DNA, here named as SBD, to look into the distribution of SBD in bacteria and see if SBD is exclusively linked with HNH motif. Homologs to SBD from *S. pristinaespiralis* with a homology of *E*-value <0.001 was compared and analyzed. Homologs containing SBD domain are not only confined in Streptomyces but distributed in other genera, such as Nocardia, Actinomadura, Thermomonospora, Nonomuraea, Actinocorallia, Herbidospora, Kitasatospora, Bacteroidetes, Cryptosporangium, Fabibacter, Rubricoccus and Endozoicomonas. Among these homologs, 4 out of 83 SBDs are not linked with HNH motif, indicated by the green rectangles (Fig. S1). Although WP_062430263.1 was annotated as HNH endonuclease, but no HNH motif can be detected when blasted against swissprot. One short portion immediately downstream SBD displaying homology to U6 snRNA-specific terminal uridylyltransferase 1; a HNHc motif ranging from residues 350 to 410 in WP_051759891.1 was detected despite that there is no HNH motif at equivalent sites to those typical SBD-linked HNH motif, a short region of this protein N-terminal SBD showed similarity to Glucose-1-phosphate adenylyltransferase small subunit; WP_018680415.1 contains a DUF3883 domain, which is usually found in restriction endonucleases with active PD-(D/E)XK motif, characteristic of the largest family of endonuclease. The fourth one, WP_096496360.1, doesn't encode any endonuclease motif, but there is a region N-terminal SBD showing similarity to large-conductance mechanosensitive channel. Despite that SBD domain is nearly linked with HNH motif in the homologs, but some exceptions as mentioned above implied that SBD might be involved in cellular processes related to metabolism of PT-DNA other than PT-dependent DNA cleavage. For these proteins, a variety of sulfur-containing molecules might be tested as the substrates to probe the recognition by SBD homologs.

Sulfur-binding domains of ScoMcrA and SprMcrA both associate with the PT R_P linkage but not with the S_P stereoisomer. The protein structure of SBD should clarify the basis for this preference. In contrast, a continuous region of 159 aa with a distinct boundary between the HNH motif and SBD in ScoMcrA is absent from SprMcrA. We speculate that this region is the 5mC recognition domain (Fig. 1, and Supplementary Data) as ScoMcrA cuts 5mC DNA mainly at a distance of N14/N15 away from C^mCTGG, but it cuts PT DNA at a varied but longer distance of N18–N24 away from the PT linkage, and thus it is more likely that the PT recognition domain is located farther away from the HNH motif than from the 5mC recognition domain. In accordance with this, SprMcrA, which lacks this 159-aa region, does not restrict uptake of the *dcm* gene, which confers the 5mC modification. Structural studies on both recognition domains are underway.

We previously reported that ScoMcrA showed star activity in cleaving unmodified DNA nonspecifically if the incubation time was extended over 5 min at a protein concentration of 25 ng μl^{-1} (equal to 380 nM). By contrast, 640 nM SprMcrA did not cleave unmodified DNA even after 1 h, whereas we detected cleavage of PT-DNA beginning at a protein concentration of 40 nM. Two possible reasons may account for this significant difference in specificity. In contrast to SprMcrA, ScoMcrA contains the postulated 159-aa 5mC recognition domain upstream of the HNH motif, and the allosteric effect generated upon the binding of SBD to the PT linkage might be attenuated spatially by the presence of this additional domain in the process of transduction from SBD to HNH. Moreover, as a DNA recognition domain, SRA not only specifically interacts with 5mC but also makes extensive contacts with base, ribose, or phosphate nearby, which may also occur with the SBD domain and the 5mC recognition domain. Therefore, the combined nonspecific affinity between these DNA moieties and the two DNA recognition domains might decrease the proportion of specific affinity stemming from binding to the PT linkage and cause low specificity in discriminating PT from unmodified DNA. Differences in the binding affinity for PT-DNA of the two SBD domains may also play a role in determining star activity.

Although the HNH motif is named after the three most conserved aa residues, a number of variants in which the second His residue was changed to Asn were reported for AnaCas9, Gme, T4eVII and I-HumI (Fig. 3). The Asn residue of the HNH motif stabilizes the position of the two β-strands relative to one another but can be replaced by alternative modes of structural stabilization. No Asn or residue equivalent to Asn is present at this site in T4eVII, for which the overall orientation of the HNH domain is the same as with the typical motif, but its cleavage activity is preserved (Eastberg *et al.*, 2007). In this study, we reported that SprMcrA has the HNH variant HRH, possibly resulting from mutation of Gln (CAG) into Arg (CGG) as closest homologs of SprMcrA have Gln as the central residue of the motif (Fig. S1). The side chains of Gln and Asn are similar but differ in length by one carbon atom.

Asn is located in a loop region that is linked with two antiparallel β-sheets and is far away from the scissile phosphate of DNA and not directly involved in the coordination of the metal ions, but N560D of ColE7 (Huang and Yuan, 2007) and N118A of colicin E9 (Walker *et al*., 2002) both decreased the activity of these HNH REases. However, N118A mutation of Colicin E9 did not affect its ability to chelate the metal ions. Both the side chain of N560 of ColE7 and N118 of Colicin E9 bond to amino acid residues within the loop connecting two antiparallel β sheets, thus constrain the loop in the metal finger structure and keep the general base His and scissile phosphate in the correct position for DNA hydrolysis. The structures for $HR_{284}H$ and $HN_{284}H$ motif were modeled and superimposed to each other (Fig. S13A and B), the orientation of R284 and N284 is different, with the latter more akin to N560 of Colicin E9 (Fig. S13C). The modeled bonding network is predicted for $HR_{284}H$ and $HN_{284}H$, respectively, and compared with that for N560A (Fig. S13D). As mutation of N560 into A, D or H, respectively, decreased the activity of ColE7, implying that N rather than any of other residues at this site is most beneficial to the endonuclease activity. Accordingly, it is easy to understand that the endonuclease activity increases when R284 was conversely mutated to N284 of SprMcrA.

The fact that R284A has no effect on activity argues that the SprMcrA fold is stabilized in a different manner and that the R284N increases activity by further stabilizing the structure in a manner more akin to other HNH nucleases. The mutation of Q to R might result from a spontaneous point mutation of CAG to CGG during evolution. It has been proposed that the biological significance of this mutation is the alleviation of cell toxicity; however, induction of SprMcrA in BL21(DE3) had undetectable effects on the growth rate and cell biomass of the host strain (Fig. S14). Compared with wild-type protein, we observed the loss of cleavage at site 724 with the R284N mutation as well as changes in cleavage efficiency at a number of modification sites (Fig. 5A and B). The bacterium might adopt spontaneous mutation at this site to regulate the cleavage activity and sequence specificity of HNH REase to cope with varied foreign DNA.

SprMcrA use its N-terminal DNA recognition domain to bind to the PT site and the C-terminal HNH to cut in the vicinity of this site. Staggered cut of double-stranded PT-DNA by SprMcrA are accompanied with weak nicking activity (Fig. S7), and consistently some nicking sites (Fig. S11) were detected in the determination of cleavage sites. This is akin to the activity of FokI, one monomer of which was proposed to bind to the target site by its DNA recognition domain and to make a nick 13 bases downstream with the catalytic domain to generate a bottom strand nicked intermediate(Wah *et al*., 1998). For hydrolysis of double strands, another monomer is recruited and dimerizes with the initial monomer to make a proximal cut 9 bases away in the top strand (Wah *et al*., 1997; Bitinaite *et al*., 1998). The dimeric SprMcrA may use the same scenario as that of FoKI to generate nick and staggered cut of double strands.

SprMcrA cuts PT DNA at two distinct distances from the PT site, but this is not due to selective binding to $G_{PS}T$ or $G_{PS}A$ as $G_{PS}TTC$ in at least three different sequence contexts cannot be bound and cut by SprMcrA (Fig. S15). Instead, the two different cutting sites are likely related to the sequence specificity of the C-terminal HNH motif as well as the DNA context between modification and cleavage sites. We make this hypothesis based on three pieces of evidence: (i) SprMcrA can bind to six PT sites (Fig. S10), but failed to cut in the vicinity of these sites; (ii) we detected nick sites 11 and 13 bases on the 5′ side of site 2270, and nick sites 9, 11 and 13 bases on the 5′ side of site 2348 in the strand of $G_{PS}AAC$ and the remaining six PT sites (Fig. S11) by chance are either 11 or 13 bases on the 5′ side of $G_{PS}AAC$; (iii) mutation of R284N, changing the catalytic center HRH into HNH, abolished cleavage at site 724 of pUC18 (Fig. 6), a result that also demonstrated that HNH in part determines the sequence specificity for SprMcrA. Several HNH-domain proteins recognize and cleave specific sequences, such as I-HmuI (Shen *et al*., 2004) and PacI (Shen *et al*., 2010). HNH endonucleases from phage and prophages have strand-specific and site-specific cleavage or nicking activity with 3- to 7-bp specificity (Xu and Gupta, 2013). More studies on the sequence specificity of the HNH_{Spr} domain and SBD_{Spr} are needed to understand the mechanism by which these domains function *in vivo*. Furthermore, we characterized SprMcrA with better performance in discerning PT-DNA from unmodified DNA than ScoMcrA and R284N mutant with 30-fold high activity than wild-type SprMcrA. This may lead to development of a PT-DNA-specific restriction endonuclease.

Experimental procedures

Preparation and purification of R_P-phosphorothioate DNA oligonucleotides

The PT-DNA oligonucleotides were chemically synthesized and PAGE-purified by Genewiz, Inc. The R_P and S_P stereoisomers were separated by anion-exchange HPLC with a DNA Pac PA-100 analytical column (Thermo Fisher) on an Agilent 1260 Infinity series system at a flow rate of 1 ml min^{-1} with the following parameters: column temperature, 50°C; solvent A consisting of 10 mM Tris–HCl, pH 8.0; solvent B consisting of 10 mM Tris–HCl, pH 8.0, 1 M NaCl; gradient of 5% B for 5 min, 20% B to 50% B over 30 min, followed by 100% B for 10 min; detection by UV absorbance at 260 nm. The eluent was desalted with a Copure C18 column (Biocomma), dried on an RVC 2-25 rotational vacuum concentrator (Martin Christ GmbH) and dissolved with distilled, deionized water. The concentration of oligonucleotides was determined by

spectrophotometric measurement on a NanoDrop 2000 spectrophotometer (Thermo Fisher), and double-stranded (ds) DNA was prepared by mixing equimolar concentrations of complementary oligonucleotides, followed by heating to 95°C for 2 min and gradual cooling. To remove the unmatched single-stranded DNA, the annealed DNA mixture was further separated using the HPLC system and conditions described above but with a changed gradient of 10% B to 70% B over 40 min. The eluent was desalted, dried and then dissolved as described above. All purified DNA oligonucleotides are listed in Table S1. PT sites are underlined. The shaded DNA sequences are the same except for the PT sites.

Construction of protein expression vector and site-directed mutagenesis

The *sprMcrA* gene was amplified by PCR, with primers SprMcrA-F (5′-GGACCATATGCCGCTCACCGACACAGACCGGT-3′) and SprMcrA-R (5′- GGAATTCTTAGCCGTCGG CCGCGCCC CGCATCCGC-3′), from the genomic DNA of *S. pristinaespiralis* [a gift from Yinhua Lu at the Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences] and cloned into the NdeI and EcoRI sites of plasmid pET28a (Novagen) with a 6×His tag at the N-terminus of the polypeptide. N-terminally His-tagged $SprMcrA_{H270A}$, $SprMcrA_{R284A}$, $SprMcrA_{R284N}$ and $SprMcrA_{H293A}$ mutant variants were constructed by whole-plasmid PCR and DpnI digestion (Zheng *et al.*, 2004). All primers for site-directed mutagenesis are listed in Table S2. All gene insertions and their mutants were cloned in *E. coli* strain DH10b, and point mutations were confirmed by DNA sequencing at Majorbio Co. Ltd., Shanghai. The proteins were expressed and purified from *E. coli* strain BL21(DE3).

Electrophoretic mobility shift assay (EMSA)

One EMSA reaction contained 40 ng DNA (equivalent to 600 nM of a 10-bp oligonucleotide or 190 nM of a 33-bp oligonucleotide) and protein at a concentration twofold or fourfold that of the DNA (molar ratio) in 10 µl binding buffer (20 mM Tris–HCl, pH 8.0, 100 mM NaCl and 5% glycerol). After incubation at room temperature for 5 min, the reaction mixtures were loaded onto 12% non-denaturing polyacrylamide gels (acrylamide:bis-acrylamide ratio of 79:1, w/w) and electrophoresed in 0.5× TBE buffer at 15 mA for 30 min.

Measuring transformation efficiency

The expression vector for SprMcrA was introduced into *E. coli* BL21(DE3), and competent cells of the resulting strain were prepared using the standard calcium chloride protocol. For the transformation efficiency assay, 100 ng of pYH9 (carrying *dnd* for PT modification), pYH10 (carrying *dcm* for 5mC modification at C^mCWGG), or the control vector pACYCDute were introduced into the competent cells (10^7 per transformation). The number of *E. coli* colony-forming units in each experiment was determined by making serial dilutions and plating on LB agar plates supplemented with appropriate antibiotics. Each experiment was repeated three times.

Protein expression and purification

A 10 ml culture grown overnight from a single colony was inoculated into 1 l of Luria broth medium supplied with 50 mg ml^{-1} kanamycin. The culture was incubated at 37°C to an OD_{600} of 0.6–0.8 and then induced by the addition of 0.2 mM isopropyl-D-1-thiogalactopyranoside (IPTG) for another 20 h at 16°C. The cells were harvested and resuspended in 20 ml binding buffer [20 mM 2-(N-morpholino) ethanesulfonic acid (MES), pH 6.8, 50 mM imidazole and 300 mM NaCl] and lysed by sonication at 4°C. After centrifugation at 16000*g* for 60 min at 4°C, the supernatant was applied to a 2 ml Ni-NTA column (GE Healthcare) pre-equilibrated with binding buffer. The Ni-NTA column was eluted with 10 ml elution buffer (20 mM MES, pH 6.8, 300 mM imidazole and 300 mM NaCl) after washing. The eluted His_6-tagged protein was further purified with a HiTrap Heparin HP affinity chromatography column (GE Healthcare), with elution using a salt gradient of 0.1 M NaCl to 1 M NaCl in 20 column volume with 2 ml min^{-1} flow rate and then a Superdex200 10/300 GL gel filtration chromatography column (GE Healthcare) equilibrated with 10 mM Tris–HCl, pH 8.0, 100 mM NaCl and 1 mM DTT, using an AKTA FPLC system (GE Healthcare). The peak fractions were combined and concentrated to 10 mg ml^{-1}. Purified proteins were visualized by Coomassie-stained 15% SDS–PAGE analysis.

DNA cleavage assay by purified SprMcrA

The 57-bp and 33-bp PT DNA oligonucleotides were generated to determine the cleavage sites of SprMcrA. The fully, hemi- and non-PT DNA duplexes used in this study are listed in Table S1. For the oligonucleotide cleavage assay, 0.2 µM DNA in each reaction was incubated with 0.2 µM SprMcrA in 20 mM MES, pH 6.5, 50 mM NaCl, 1 mM $MnCl_2$ and 1 mM DTT in a volume of 20 µl at 30°C for 1 h. The reactions were followed by digestion with 2 units of proteinase K (Roche) and examined by 15% native PAGE (acrylamide:bis-acrylamide ratio of 19:1) with a control oligonucleotide of appropriate length synthesized by Genewiz. Inc. The Ultra Low Range DNA Ladder (Thermo Fisher) was used as a molecular size marker.

To determine the cleavage sites on PT-pUC18 isolated from *E. coli* B7A (*dnd*$^+$), pUC18 linearized by EcoRI was digested with SprMcrA in a reaction buffer containing 20 mM MES, pH 6.5, 50 mM NaCl, 1 mM $MnCl_2$ and 1 mM DTT at 30°C for 1 h. The reactions were stopped by proteinase K digestion and examined by electrophoresis on a 2% agarose gel.

Runoff sequencing to determine cleavage sites in PT-DNA oligonucleotides and PT-pUC18

Linearized PT-pUC18 was incubated with SprMcrA followed by digestion of proteinase K. Except for the band corresponding to linear pUC18 not digested by SprMcrA, DNA fragments of similar sizes were recovered as individual gel sections. To avoid omitting DNA fragments < 500-bp in size, we used EcoRI- and NdeI-linearized PT-pUC18 DNA as substrates for SprMcrA cleavage and DNA recovery. Ten

gel sections of DNA fragments were finally purified for runoff sequencing with a series of primers near the GAAC/GTTC binding sites (Table S3). The double peaks (A/C, A/G, A/T) or enhanced A peaks in the sequencing results indicated a nick in the template strand as the extrinsic A was added by Taq DNA polymerase through its template-independent terminal nucleotide transferase activity. Undigested pUC18 was used as a sequencing control.

Protein sequence alignment and structure analysis

The protein sequences of SprMcrA and its homologs were aligned by ClustalW software. The HNH motif tertiary structures of SprMcrA and SprMcrA$_{R284N}$ were predicted using the online server iTASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), and the model with the highest C-score was selected. Protein structure superposition was accomplished using PyMOL software (https://www.pymol.org).

Gel filtration to determine the absolute molecular weight of SprMcrA

For SprMcrA gel filtration, the Superdex 200 10/300 GL column (GE Healthcare) was equilibrated with 10 mM MES, pH 6.8, 300 mM NaCl at 1 ml min^{-1} through an ÄKTA FPLC (GE Healthcare). Next, 500 µl of 1 mg ml^{-1} SprMcrA and 1 mg ml^{-1} chymotrypsinogen A (25 kD), ovalbumin (43 kD) and albumin (66 kD) were loaded onto the column sequentially prior to elution at a flow rate of 1 ml min^{-1}.

Fluorescence polarization assay for analysis of DNA binding

For probes, 5′-FAM-labeled PT-duplexes were synthesized and purified, with only one strand labeled on each probe. Twofold serially diluted protein solutions (5 µM starting concentration, 16 dilutions) were mixed with a final concentration of DNA probe of 5 nM in a Corning 3575 plate, using binding buffer composed of 20 mM MES, pH 6.8, 5% glycerol, 50 mM NaCl and 1 mM DTT. Reactions were incubated for 10 min at room temperature, and fluorescence polarization was measured at room temperature on a SpectraMax I3× (Molecular Devices) using 485/20 nm and 528/20 nm filters for emission and excitation respectively. The dissociation constants (K_D) were calculated using GraphPad Prism software (version 6.0) by fitting the experimental data to the following equation: [mP] = [maximum mP][C]/(KD + [C]) + [baseline mP], and then the curve was replotted using % of saturation calculated as ([mP] – [baseline mP])/([maximum mP] – [baseline mP]), where mP is millipolarization and [C] is protein concentration.

Ion requirement and pH optimization for PT-DNA cleavage by purified SprMcrA

Escherichia coli B7A bears a chromosomal *dnd* gene cluster that generates PT modification of DNA at G$_{PS}$AAC/G$_{PS}$TTC sites with varied efficiency (Wang *et al.*, 2011). The genomic DNA of *E. coli* B7A (PT$^+$) and *E. coli* DH10b (PT$^-$) prepared by QIAamp DNA Mini Kit (Qiagen) were measured as substrate and control, respectively, for SprMcrA and its mutants. To test the ion requirement for DNA cleavage activity, purified SprMcrA (400 ng) and DNA substrate (200 ng) were incubated at 30°C for 1 h in a total 20 µl reaction system (20 mM Tris–Cl, pH 7.5, 50 m NaCl) supplemented with different metal ions (Co^{2+}, Cu^{2+}, Mg^{2+}, Mn^{2+}, Ni^{2+}, Ca^{2+} or Zn^{2+} at a concentration of 1 µM–10 mM) in each digestion. The optimal pH was determined in a similar manner using equivalent buffers of pH 5 to pH 10, supplemented with 1 mM Mn^{2+}. The reactions were digested with 5 units of proteinase K (Roche) to remove SprMcrA prior to separation by electrophoresis in 0.8% agarose gel.

Growth curves of E. coli hosts expressing SprMcrA and SprMcrA$_{R284N}$

Escherichia coli BL21(DE3) strains expressing SprMcrA or SprMcrA$_{R284N}$ were cultured overnight and measured for cell density by OD_{600} value using the Microplate Reader (BioTeK Syngery2). The same volume of cells of each strain was inoculated into 200 µL LB at a ratio of V/V = 1:100 into microplate wells supplied with 100 µg ml^{-1} ampicillin, followed by incubation at 37°C in the microplate reader with shaking, with OD_{600} measurements taken once per 30 min. Each experiment was repeated three times.

Acknowledgements

We are grateful to Dr. Yin-Hua Lu (Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences) for providing genomic DNA of *Streptomyces pristinaespiralis*. This work was supported by grants from the National Natural Science Foundation of China (grant numbers 31670034, 31470195, and 21661140002), and in part by the Special Fund for Science and Technology Innovation of Shanghai Jiao Tong University awarded to X.H.

Author contributions

HY have made major contributions to design of the study, acquisition of the data, analysis of the data and writing of the manuscript. XYH have made major contributions to conception of the study and writing of the manuscript. GL, GZ, WYH and GW have made contributions to acquisition and analysis of the data. ZXD have made contributions to conception of the study.

References

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. and Shirakawa, M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C. *et al.* (2008) Structural basis for recognition

of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes & Development*, **16**, 6–21.
Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10570–10575.
Borst, P. and Sabatini, R. (2008) Base J: discovery, biosynthesis, and possible functions. *Annual Review of Microbiology*, **62**, 235–251.
Cao, B., Chen, C., DeMott, M.S., Cheng, Q., Clark, T.A., Xiong, X. *et al.* (2014) Genomic mapping of phosphorothioates reveals partial modification of short consensus sequences. *Nature Communications*, **5**, 3951.
Chan, S.H., Opitz, L., Higgins, L., O'Loane, D. and Xu, S.Y. (2010) Cofactor requirement of HpyAV restriction endonuclease. *PLoS One*, **5**, e9071.
Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J. *et al.* (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11040–11045.
Drouin, M., Lucas, P., Otis, C., Lemieux, C. and Turmel, M. (2000) Biochemical characterization of I-CmoeI reveals that this H-N-H homing endonuclease shares functional similarities with H-N-H colicins. *Nucleic Acids Research*, **28**, 4566–4572.
Eastberg, J.H., Eklund, J., Monnat, R. Jr. and Stoddard, B.L. (2007) Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry*, **46**, 7215–7225.
Furuta, Y., Abe, K. and Kobayashi, I. (2010) Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. *Nucleic Acids Research*, **38**, 2428–2443.
Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J. Jr. *et al.* (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nature Structural Biology*, **6**, 1096–1099.
Gan, R., Wu, X., He, W., Liu, Z., Wu, S., Chen, C. *et al.* (2014) DNA phosphorothioate modifications influence the global transcriptional response and protect DNA from double-stranded breaks. *Scientific Reports*, **4**, 6642.
Han, T., Yamada-Mabuchi, M., Zhao, G., Li, L., Liu, G., Ou, H.Y. *et al.* (2015) Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins. *Nucleic Acids Research*, **43**, 1147–1159.
Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E. and Cheng, X. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, **455**, 826–829.
Hashimoto, H., Horton, J.R., Zhang, X. and Cheng, X. (2009) UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics*, **4**, 8–14.
He, X., Ou, H.Y., Yu, Q., Zhou, X., Wu, J., Liang, J. *et al.* (2007) Analysis of a genomic island housing genes for DNA S-modification system in Streptomyces lividans 66 and its counterparts in other distantly related bacteria. *Molecular Microbiology*, **65**, 1034–1048.
Horton, J.R., Wang, H., Mabuchi, M.Y., Zhang, X., Roberts, R.J., Zheng, Y. *et al.* (2014) Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Research*, **42**, 12092–12101.
Huang, H. and Yuan, H.S. (2007) The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *Journal of Molecular Biology*, **368**, 812–821.
Iyer, L.M., Abhiman, S. and Aravind, L. (2011) Natural history of eukaryotic DNA methylation systems. *Progress in Molecular Biology and Translational Science*, **101**, 25–104.
Keeble, A.H., Hemmings, A.M., James, R., Moore, G.R. and Kleanthous, C. (2002) Multistep binding of transition metals to the H-N-H endonuclease toxin colicin E9. *Biochemistry*, **41**, 10234–10244.
Kellner, S., DeMott, M.S., Cheng, C.P., Russell, B.S., Cao, B., You, D. *et al.* (2017) Oxidation of phosphorothioate DNA modifications leads to lethal genomic instability. *Nature Chemical Biology*, **13**, 888–894.
Kriukiene, E. (2006) Domain organization and metal ion requirement of the Type IIS restriction endonuclease MnlI. *FEBS Letters*, **580**, 6115–6122.
Ku, W.Y., Liu, Y.W., Hsu, Y.C., Liao, C.C., Liang, P.H., Yuan, H.S. *et al.* (2002) The zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis. *Nucleic Acids Research*, **30**, 1670–1678.
Liu, G., Ou, H.Y., Wang, T., Li, L., Tan, H., Zhou, X. *et al.* (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genetics*, **6**, e1001253.
Loenen, W.A. and Raleigh, E.A. (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Research*, **42**, 56–69.
Ou, H.Y., He, X., Shao, Y., Tai, C., Rajakumar, K. and Deng, Z. (2009) dndDB: a database focused on phosphorothioation of the DNA backbone. *PLoS One*, **4**, e5132.
Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C. *et al.* (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *Journal of Molecular Biology*, **314**, 735–749.
Pommer, A.J., Kuhlmann, U.C., Cooper, A., Hemmings, A.M., Moore, G.R., James, R. *et al.* (1999) Homing in on the role of transition metals in the HNH motif of colicin endonucleases. *Journal of Biological Chemistry*, **274**, 27153–27160.
Raaijmakers, H., Vix, O., Toro, I., Golz, S., Kemper, B. and Suck, D. (1999) X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. *The EMBO Journal*, **18**, 1447–1458.
Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D. *et al.* (2011) A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. *Genes & Development*, **25**, 137–152.
Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2015) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*, **43**, D298–299.

Saravanan, M., Vasu, K., Ghosh, S. and Nagaraja, V. (2007) Dual role for Zn2+ in maintaining structural integrity and inducing DNA sequence specificity in a promiscuous endonuclease. *Journal of Biological Chemistry*, **282**, 32320–32326.

Shen, B.W., Heiter, D.F., Chan, S.H., Wang, H., Xu, S.Y., Morgan, R.D. *et al.* (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI. *Structure*, **18**, 734–743.

Shen, B.W., Landthaler, M., Shub, D.A. and Stoddard, B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *Journal of Molecular Biology*, **342**, 43–56.

Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J.M. and Skowronek, K. (2012) Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Research*, **40**, 7563–7572.

Sutherland, E., Coe, L. and Raleigh, E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of Molecular Biology*, **225**, 327–348.

Swinton, D., Hattman, S., Crain, P.F., Cheng, C.S., Smith, D.L. and McCloskey, J.A. (1983) Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 7400–7404.

Szwagierczak, A., Brachmann, A., Schmidt, C.S., Bultmann, S., Leonhardt, H. and Spada, F. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Research*, **39**, 5149–5156.

Tong, T., Chen, S., Wang, L., Tang, Y., Ryu, J.Y., Jiang, S. *et al.* (2018) Occurrence, evolution, and functions of DNA phosphorothioate epigenetics in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(13), E2988–E2996.

Tost, J. (2010) DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Molecular Biotechnology*, **44**, 71–81.

Vasu, K., Nagamalleswari, E., Zahran, M., Imhof, P., Xu, S.Y., Zhu, Z. *et al.* (2013) Increasing cleavage specificity and activity of restriction endonuclease KpnI. *Nucleic Acids Research*, **41**, 9812–9824.

Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 10564–10569.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.

Walker, D.C., Georgiou, T., Pommer, A.J., Walker, D., Moore, G.R., Kleanthous, C. *et al.* (2002) Mutagenic scan of the H-N-H motif of colicin E9: implications for the mechanistic enzymology of colicins, homing enzymes and apoptotic endonucleases. *Nucleic Acids Research*, **30**, 3225–3234.

Wang, L., Chen, S., Vergin, K.L., Giovannoni, S.J., Chan, S.W., DeMott, M.S. *et al.* (2011) DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 2963–2968.

Wang, L., Chen, S., Xu, T., Taghizadeh, K., Wishnok, J.S., Zhou, X. *et al.* (2007) Phosphorothioation of DNA in bacteria by dnd genes. *Nature Chemical Biology*, **3**, 709–710.

Wei, H., Therrien, C., Blanchard, A., Guan, S. and Zhu, Z. (2008) The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Research*, **36**, e50.

Xie, X., Liang, J., Pu, T., Xu, F., Yao, F., Yang, Y. *et al.* (2012) Phosphorothioate DNA as an antioxidant in bacteria. *Nucleic Acids Research*, **40**, 9115–9124.

Xu, S.Y. and Gupta, Y.K. (2013) Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Research*, **41**, 378–390.

Xu, S.Y., Kuzin, A.P., Seetharaman, J., Gutjahr, A., Chan, S.H., Chen, Y. *et al.* (2013) Structure determination and biochemical characterization of a putative HNH endonuclease from Geobacter metallireducens GS-15. *PLoS One*, **8**, e72114.

Zheng, L., Baumann, U. and Reymond, J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, **32**, e115.

Zhou, X., He, X., Liang, J., Li, A., Xu, T., Kieser, T. *et al.* (2005) A novel DNA modification by sulphur. *Molecular Microbiology*, **57**, 1428–1438.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.