Published online 18 May 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 15* ***7563–7572***
doi:10.1093/nar/gks428

Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI

Wojciech Siwek[1,2], Honorata Czapinska[2], Matthias Bochtler[2,3,4,*], Janusz M. Bujnicki[1,5,*] and Krzysztof Skowronek[1,*]

[1]Laboratory of Bioinformatics and Protein Engineering, [2]Laboratory of Structural Biology, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, [3]Institute of Biochemistry and Biophysics PAS, Pawinskiego 5a, 02-106 Warsaw, Poland, [4]Schools of Chemistry and Biosciences, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK and [5]Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland

Received January 27, 2012; Revised April 23, 2012; Accepted April 24, 2012

ABSTRACT

DNA methylation-dependent restriction enzymes have many applications in genetic engineering and in the analysis of the epigenetic state of eukaryotic genomes. Nevertheless, high-resolution structures have not yet been reported, and therefore mechanisms of DNA methylation-dependent cleavage are not understood. Here, we present a biochemical analysis and high-resolution DNA co-crystal structure of the N^6-methyladenine (m6A)-dependent restriction enzyme R.DpnI. Our data show that R.DpnI consists of an N-terminal catalytic PD-(D/E)XK domain and a C-terminal winged helix (wH) domain. Surprisingly, both domains bind DNA in a sequence- and methylation-sensitive manner. The crystal contains R.DpnI with fully methylated target DNA bound to the wH domain, but distant from the catalytic domain. Independent readout of DNA sequence and methylation by the two domains might contribute to R.DpnI specificity or could help the monomeric enzyme to cut the second strand after introducing a nick.

INTRODUCTION

Restriction modification (RM) systems are composed of a restriction endonuclease and a DNA methyltransferase, and provide bacteria with protection against non-self DNA (1). In most cases, the self DNA is marked by methylation, introduced in a sequence-specific manner by the methyltransferase. Non-self DNA, which lacks the methylation mark, is cleaved by the endonuclease. However, the rule that self, but not non-self DNA is modified, is not universal. Presumably, under the pressure from RM systems, some phages have evolved to incorporate a wide variety of modified DNA bases into their genomes. In a further escalation of the 'arms race' between bacteria and their pathogens, this has then led to the evolution of modification-dependent restriction enzymes, which treat unmodified DNA as 'self' and modified DNA as 'non-self'.

Modification-dependent restriction enzymes are a very diverse group, and in the current classification of nucleases fall either into the type IIM (where 'M' stands for modification-dependent) or type IV class, depending on subunit composition, biochemical properties and specificity of cleavage (2). Type IV McrA belongs to the ββα-Me (or HNH) superfamily of nucleases, and is, at least *in vivo*, specific for the Ym5CGR target sequence (where Y is C or T and R is G or A) (3). Type IV McrBC is an oligomeric protein complex, which recognizes pairs of Rm5C sites separated by 40–2000 bp of non-specific DNA. The enzyme consists of the GTPase McrB, which is also required for the recognition of methylation, and the McrC subunit that harbors a PD-(D/E)XK motif and catalyzes DNA cleavage (4,5). Type IV Mrr is a cryptic endonuclease thought to act on modified adenines, whose activity appears to be triggered by some forms of stress (6). The enzyme causes genotoxicity when expressed in bacteria containing certain type III DNA m6A methyltransferases. A strong anti-correlation between close homologues of Mrr and these methyltransferases has been observed in proteobacterial genomes (7). Some enzymes related to Mrr, including the recently characterized R.MspJI, have been shown to cleave DNA

*To whom correspondence should be addressed. Tel: +48 22 5970 732; Fax: +48 22 5970 715; Email: mbochtler@iimcb.gov.pl
Correspondence may also be addressed to Janusz Bujnicki. Tel: +48 22 5970 750; Fax: +48 22 5970 715; Email: iamb@genesilico.pl
Correspondence may also be addressed to Krzysztof Skowronek. Tel: +48 22 5970 753; Fax: +48 22 5970 715; Email: krzysiek@genesilico.pl

that is methylated on cytosines (8). There are also many other methylation-specific activities (e.g. R.AoxI, R.BisI, R.BlsI, R.GlaI, R.GluI, R.KroI and R.PscI), which have not yet been ascribed to molecularly defined complexes and their domain composition and mechanism of cleavage remains unknown.

R.DpnI is the best known modification-dependent restriction endonuclease that recognizes methylated adenines rather than cytosines. The enzyme was originally isolated from *Streptococcus* (formerly *Diplococcus*) *pneumoniae* (9). *In vivo*, it protects bacteria against phages that have been propagated on Dam^+ hosts. This is possible thanks to its specificity for Gm6ATC, the product of Dam methylation. *In vitro*, R.DpnI cleaves GATC sites efficiently when these are fully methylated (i.e. in both strands) and also hemi-methylated, when the enzyme is used in high concentration. It is a blunt end cutter that cleaves DNA in the middle of the recognition sequence (10–12), rather than distantly from it like many other modification-dependent restriction endonucleases. Due to this methylation dependence, R.DpnI has been classified as a type IIM enzyme. Unlike most other type II restriction endonucleases, it is a monomer in solution (10). Due to its unusual properties, it is widely used in biotechnological applications. A standard site-directed mutagenesis strategy (Stratagene) exploits the enzyme to remove a non-mutated template that has been propagated in a Dam^+ host, but not a PCR product carrying the mutation. The preference of R.DpnI for fully methylated over hemi-methylated sites has been used to create 'ultra-rare' cutting enzymes by combining it with suitably chosen m6A methyltransferases (13–15). R.DpnI is also useful for studies of bacterial DNA replication, because digestion with low amounts of the enzyme cleaves all but freshly replicated DNA (12).

Despite the widespread use of R.DpnI in biotechnology and the interest in other modification-dependent enzymes, which could be applied to studies of epigenetic modifications of eukaryotic DNA, the present understanding of this group of enzymes is very limited. In particular, crystal or NMR structures have not been reported, and the mechanisms of modification-dependent cleavage remain mysterious. In a first step towards a better understanding of methylation dependent cleavage, we present a detailed biochemical characterization of R.DpnI and its co-crystal structure with fully methylated substrate DNA.

MATERIALS AND METHODS

DNA manipulations

The *dpnI* gene as well as separate domain genes were cloned by standard techniques into the pET28 vector with a PreScission cleavable N-terminal His-tag (16). The QuickChange method (Stratagene) was used to introduce key residue substitutions into the R.DpnI coding sequence in the expression construct.

Protein preparation

ER2925 *Escherichia coli* strain (New England Biolabs) lysogenized with DE3 element using λDE3 Lysogenization Kit (Novagene) was used as an expression host. The expression was induced with 0.2 mM IPTG and carried out overnight at 25°C. After centrifugation, cell pellets were resuspended in buffer A: 15% (v/v) glycerol, 50 mM PIPES (pH 7), 300 mM NaCl, 2.5 mM imidazole, 10 mM β-mercaptoethanol. Cells were disrupted in French press at 20 000 psi and the lysate was cleared by 30 min centrifugation at 15000×*g*. The supernatant was applied on a His-Select Nickel Affinity Gel (Sigma) equilibrated with buffer A. The resin was washed with buffer B: 15 % (v/v) glycerol, 50 mM PIPES (pH 7), 2 M NaCl, 2.5 mM imidazole, 1% igepal, 10 mM β-mercaptoethanol and next with buffer C: 15% (v/v) glycerol, 50 mM PIPES (pH 7), 300 mM NaCl, 5 mM imidazole, 10 mM β-mercaptoethanol. R.DpnI was eluted with buffer D: 15 % (v/v) glycerol, 50 mM PIPES (pH 7), 300 mM NaCl, 250 mM imidazole, 10 mM β-mercaptoethanol. For His-tag cleavage, the protein was 5-fold diluted with buffer E: 15% (v/v) glycerol, 50 mM PIPES (pH 7), 50 mM NaCl, 10 mM β-mercaptoethanol. The PreScission protease cleavage was done overnight at 4°C. The protease was removed on a MonoS column (GE Healthcare) equilibrated with buffer E. R.DpnI was eluted with 50–1000 mM gradient of NaCl in buffer E. Enzyme containing fractions were pooled together and the buffer was exchanged on a PD-10 column (GE Healthcare) to buffer F: 15% (v/v) glycerol, 50 mM PIPES (pH 7), 150 mM NaCl, 10 mM β-mercaptoethanol, 50 mM arginine and 50 mM glutamic acid. For crystallization trials, buffer F was supplemented with 5 mM $CaCl_2$ (buffer G).

Cleavage assays

The cleavage assays were performed for 2 h at 37°C in 20 µl reaction volume with the standard Tango buffer (Fermentas). The concentration of R.DpnI was 37.5 nM. Due to the low yield of the PD-(D/E)XK domain expression, its concentration could not be determined. Dam methylated pBR322 plasmid DNA in the final concentration of 5.6 nM (GATC sites concentration: 123.2 nM) was used as a substrate unless stated otherwise. The non-methylated plasmids were isolated from a Dam^- host. The Hia5-methylated pBR322 plasmid was isolated from a Dam^+ host and additionally *in vitro* methylated by Hia5 methyltransferase (a kind gift of M. Radlinska).

Cleavage of substrates with one and two GATC sites

The assay was performed with dsDNA with either two GATC sites (distance between sites 141 bp, flanking sequences 460 and 272 bp) or with either of the sites separately mutated to a GGCC sequence. The substrates were prepared from the pUC18 plasmid methylated *in vivo* in Dam^+ *E. coli* cells. The dsDNA fragments were cut out of the plasmid and its derivatives with R.SapI and R.SspI restriction endonucleases prior to cleavage with R.DpnI. The reaction products were resolved by agarose gel electrophoresis. The appropriate fragments were cut out of

the gel and purified by Gel Out kit (A&A Biotechnology). The reaction progress was measured by monitoring the decrease of the substrate at indicated time points for 2 h (Supplementary Figure S1). R.DpnI and dsDNA were used in the concentrations of 37.5 pM and 9.25 nM, respectively.

Nitrocellulose filter binding assays

The relative binding affinities were determined in a filter binding assay (17). Substrate oligonucleotides, including the methylated ones (Metabion) were radiolabeled with [γ-^{33}P]ATP (Hartmann Analytic) and T4 polynucleotide kinase (Fermentas). Mixtures (200 µl total volume in buffer F containing double-stranded oligonucleotide (15 nM) and R.DpnI (500 nM) were incubated for 30 min at 25°C. Reaction mixtures were filtered through a 0.22 -µm nitrocellulose filter (Whatman) in a Dot-Blot apparatus (Bio-Rad). Each well was washed three times with 200 µl of binding buffer. Dried filters were exposed to a phosphoimager screen overnight. Images were scanned on a Storm Phosphoimager (GE Healthcare) and retained radioactivity was quantified using ImageQuant software (GE Healthcare).

Crystallization

The DNA duplex for the crystallization experiments was obtained by annealing 5′-CTG**Gm6ATC**CAG-3′ oligonucleotide with itself by slow cool-down from 95°C to 4°C during 30 min. The R.DpnI mediated cleavage of the short oligonucleotide was confirmed electrophoretically (Supplementary Figure S2). The protein in 10 mg/ml concentration was provided in buffer G and mixed in 1:1 molar ratio with target DNA duplex also in buffer G to form a tight complex. Crystal growth was initiated by mixing the protein-DNA solution in 3:1 ratio with crystallization buffer B (200 mM potassium sulfate, 100 mM betaine, 20% w/v PEG 3350, pH 6.8). Crystal growth occurred at 20°C over a period of ~30 days. Prior to data collection, crystals were transferred to a 3:1 mix of crystallization buffer B and 100% glycerol and flash-cryocooled to 100 K.

Structure determination and refinement

High resolution (2.05 Å) and multiple anomalous diffraction data (to about 2.9 Å resolution, collected at 1.2833 and 1.2835 Å wavelengths) were collected at the Diamond Light Source synchrotron. The SHELXC program (18) confirmed the presence of anomalous signal. SHELXD (19) was used to interpret the anomalous signal in terms of a heavy atom structure with three anomalous scatterers that were later confirmed to coincide with two Zn^{2+} binding sites and one additional site on the symmetry axis. The original MAD density was subjected to cyclic 2-fold averaging and solvent flattening, taking care to avoid excessive modification of the phases, particularly for the high-resolution reflections. The resulting density was then of sufficient quality for the ARP/wARP program to deliver a 75% complete model of the protein and some DNA fragments (20). The model was improved by mapping the best fragments of both protein monomers onto the local symmetry mates. DNA was mostly built manually using the COOT program (21). Refinement was carried out with the REFMAC (22) and CNS (23) programs. The resulting quality indicators are summarized in Supplementary Table S1 and appear satisfactory for a structure of about 2 Å resolution. Atomic coordinates of the refined model and corresponding structure factors have been deposited to Protein Data Bank with the accession code 4ESJ.

RESULTS

Expression of a stable and active variant of R.DpnI

R.DpnI is DNA methylation-dependent, and thus its expression should not, in general, be toxic in the *E. coli* Dam$^-$ strain ER2925. A lysogenized version of the strain that expresses T7 polymerase was used to overproduce the enzyme under the control of the T7 promoter. As the R.DpnI amino acid sequence contained predicted thrombin and TEV cleavage sites, we expressed a variant with PreScission cleavable N-terminal histidine tag, which was soluble, but prone to degradation. Mass spectrometry was used to detect the cleavage site between R134 and R135. This suggested that R.DpnI was hydrolyzed by the *E. coli* OmpT protease. In order to suppress the cleavage, we replaced the less conserved R134 with alanine and asparagine (the latter occurs naturally in R.DpnI homologues). The asparagine, but not the alanine mutation was well tolerated. Therefore, all further experiments were carried out with this variant of the protein.

Domain organization and classification

R.DpnI had been grouped with restriction endonucleases that are difficult to classify into the major phylogenetic classes (24). However, a closer inspection indicated that the enzyme is a two-domain protein with a predicted N-terminal PD-(D/E)XK domain and C-terminal HTH domain. PD-(D/E)XK enzymes tend to be very selective about metal co-factors and typically accept only either Mg^{2+} or Mn^{2+}. In contrast, both $\beta\beta\alpha$-Me (HNH) and GIY-YIG endonucleases are more promiscuous in their requirement for divalent metal ions (25,26) and phospholipase-like enzymes (such as R.BfiI) do not require a metal ion co-factor altogether (27). We therefore tested the activity of R.DpnI in the presence of either Mg^{2+}, Ca^{2+} or Mn^{2+}. Like other PD-(D/E)XK superfamily enzymes (28), R.DpnI was active in the presence of Mg^{2+} or Mn^{2+}, but not Ca^{2+} ions (Figure 1A).

Biochemical validation of the active site prediction

Bioinformatic analysis suggested that D53, E64 and K66 were part of the R.DpnI PD-(D/E)XK motif. Despite the name, the proline is frequently not conserved in the PD-(D/E)XK family (24). Its role in R.DpnI could either be played by the A52 just upstream of D53, or by a nearby proline (P50). In order to verify the assignment of candidate 'signature' residues, P50, D53, E64, L65 and K66 were all individually substituted with alanines. P50

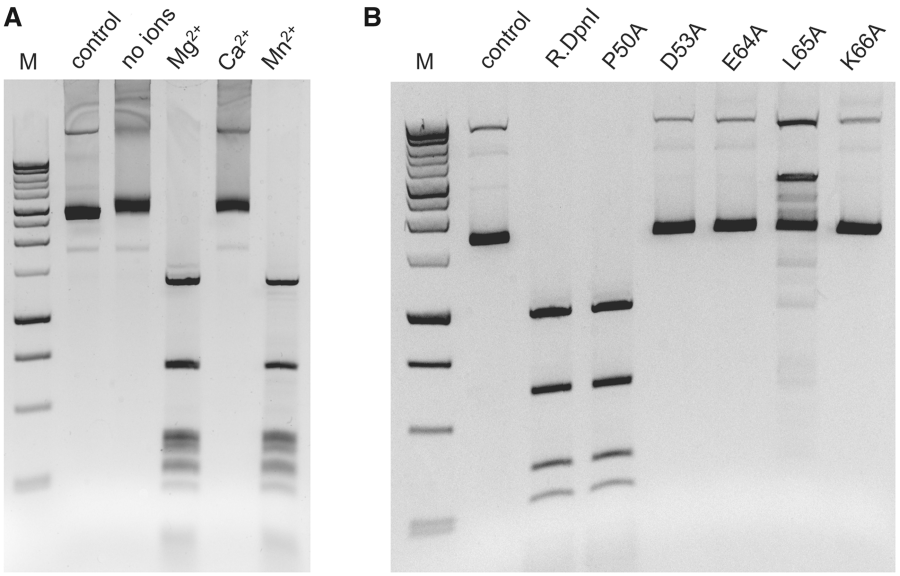

Figure 1. Biochemical validation of R.DpnI active site prediction. (**A**) Divalent metal ion dependence of the plasmid DNA cleavage by R.DpnI. The reaction buffer contained: no divalent metal ions, 10 mM $MgCl_2$, 10 mM $CaCl_2$ or 10 mM $MnCl_2$. (**B**) Cleavage activity of R.DpnI variants with single amino acid substitutions within the active site. M: molecular mass marker, control: non-cleaved substrate.

turned out to be unimportant for the catalytic activity. The L65A variant nicked DNA at the tested concentration and probably also cleaved the plasmid at a very slow rate. All other changes essentially abolished activity of R.DpnI as expected (Figure 1B).

Determination of sequence specificity

R.DpnI is known to cleave Gm6ATC, but not non-methylated target sites. We noticed that 'star' activity of R.DpnI against other adenine-methylated sequences would not compromise a Dam^- host (without other adenine methyltransferases) and would therefore not be selected against. In order to test for such activity, we took advantage of the recent discovery of the fairly unspecific adenine methyltransferase Hia5, which methylates the majority of adenine residues in a DNA template (29). We digested either Dam or Dam and Hia5-methylated pBR322 plasmid with R.DpnI. At high concentration of the endonuclease, the Hia5-methylated DNA was cleaved at a larger number of sites than the Dam-methylated DNA. This indicates that R.DpnI can cleave DNA with methylated adenine in contexts other than Gm6ATC (Figure 2A). The activity was specific for R.DpnI and not due to a contaminating m6A-dependent nuclease from the *E. coli* lysate, because it was abolished by amino acid replacements in the R.DpnI active site (Supplementary Figure S3). In order to characterize the star activity more closely, we used R.DpnI to digest pBR322 plasmid DNA that was methylated by either Dam alone or Dam and Hia5. The digestion fragments were ligated into a blunt end recipient vector (pUC18). Sequencing of 61 independent clones confirmed that R.DpnI cleaved Dam-modified DNA only at (methylated) GATC sites. In Dam and Hia5-modified DNA, (methylated) GATC was the most frequently cleaved site, but many other sites, primarily those with only one non-cognate base, were also found (Figure 2B and Supplementary Table S2). We note that m6A methylation was not explicitly tested in this assay (any cleavage at non-modified sites does contribute to the outcome of this experiment) and that star activity leading to sticky end products would not be detected.

Determination of binding affinity

It is known that R.DpnI cleaves fully methylated DNA much more efficiently than hemi-methylated DNA (12), but it is not clear whether this effect is purely dependent on catalytic rates, or whether binding effects also play a role. We therefore assayed R.DpnI in a filter-binding assay against fully methylated (mDNA), hemi-methylated (hmDNA) or non-methylated oligo DNA (nmDNA), as well as an additional set of oligos with scrambled sequences (smDNA, ssmDNA and snmDNA) in the presence of Ca^{2+} ions (Figure 3 and Supplementary Table S3A). We conclude from the results of these experiments that both methyl groups and the correct sequence context are important for tight DNA binding by the enzyme.

Crystallization and structure determination

In order to understand the structural basis of the R.DpnI properties, the protein was crystallized in complex with a stoichiometric amount of fully methylated substrate DNA. The latter was obtained by self-annealing of the palindromic 5′-CTG**Gm6ATC**CAG-3′ oligonucleotide. In order to prevent cleavage of the substrate DNA, crystallization drops contained Ca^{2+} instead of Mg^{2+} ions. Crystals grew in space group *C*2 and diffracted to 2.05 Å resolution at the Diamond Light Source. Due to the presence of two CXXC motifs in the R.DpnI sequence, it was anticipated that the enzyme should bind a zinc ion.

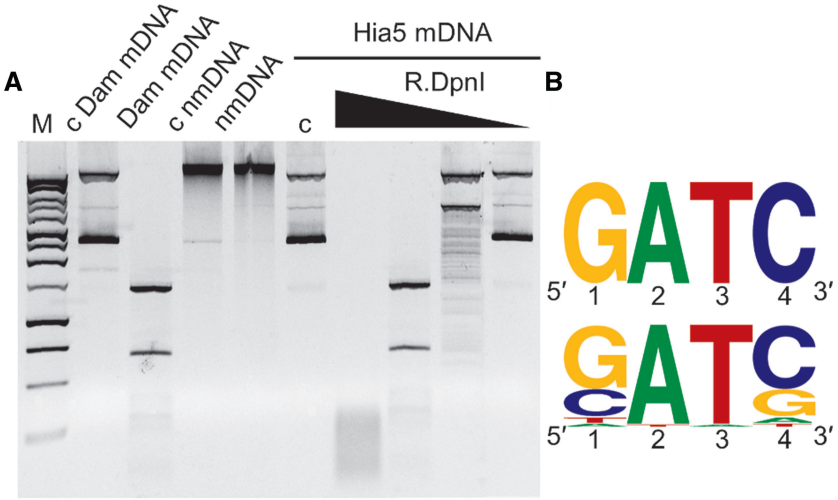

Figure 2. Determination of R.DpnI sequence selectivity. (**A**) R.DpnI mediated cleavage of Dam and Hia5-methylated plasmid DNA in comparison to Dam-methylated and non-methylated DNA. R.DpnI concentration: 37.5 nM, pBR322 plasmid concentration: 5.6 nM, GATC sites concentration: 123.2 nM and SATS sites concentration: 537.6 nM. The wedge indicates decreasing concentration of R.DpnI (37.5 nM, 3.75 nM, 375 pM and 37.5 pM). M: molecular mass marker, c: non-cleaved DNA, Dam mDNA: Dam-methylated plasmid, nmDNA: non-methylated plasmid (isolated from a Dam$^-$ *E. coli* strain), Hia5 mDNA: Hia5-methylated plasmid isolated from a Dam$^+$ *E. coli* strain. (**B**) Sequence logo of R.DpnI selectivity on Dam (upper logo) and Dam and Hia5-methylated DNA (lower logo). pBR322 Dam or Dam$^+$ Hia5-methylated plasmid was cleaved with R.DpnI and the resulting fragments were cloned to pUC18. A total of 61 independent clones were sequenced. Logos were generated by the WebLogo program (48).

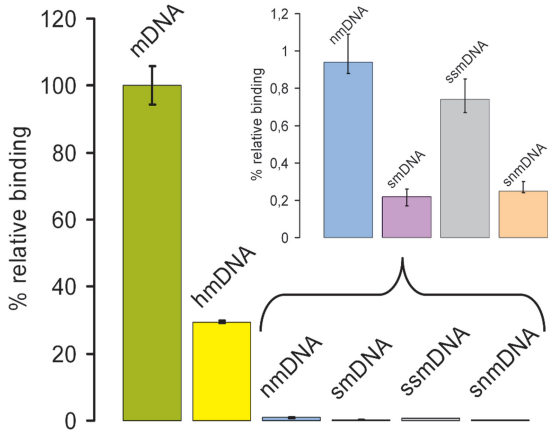

Figure 3. Binding of R.DpnI to different 21-mer oligo DNAs. Abbreviations for oligonucleotide duplexes: mDNA: fully methylated Gm6ATC, hmDNA: hemi-methylated Gm6ATC, nmDNA: non-methylated GATC, smDNA: methylated scrambled sequence (CTm6AG), ssmDNA: methylated sequence with scrambled positions 1 and 4 and unchanged m6A position (Cm6ATG). snmDNA: non-methylated scrambled sequence (CTAG). For full oligo sequences see Supplementary Table S4.

Although Zn^{2+} ions were not added to any of the buffers during protein purification and crystallization, the metal remained bound to R.DpnI according to the X-ray fluorescence signal and could be exploited to phase the diffraction data.

Gross structure of the R.DpnI–DNA complex

Crystals contain two copies of monomeric R.DpnI in complex with target DNA in the asymmetric unit. The two monomers are not related by a 2-fold axis with each other or with symmetry mates, therefore it can be rigorously excluded that a physiologically relevant dimer is formed in the crystals. R.DpnI activity as a monomer is also supported by the observation that in digests with this enzyme, nicks accumulate before double-strand breaks are formed (Supplementary Figure S4). The two molecules of R.DpnI with their bound DNA duplexes are nearly superimposable. Therefore, it suffices to discuss one monomer only. As predicted by the bioinformatic analysis, each R.DpnI chain consists of an N-terminal catalytic PD-(D/E)XK domain (residues 1–182) and a C-terminal domain (residues 183–254). Each DNA duplex is bound closely to a single C-terminal domain, but is distant to the PD-(D/E)XK domain of the same protein molecule. Therefore, the DNA complex corresponds to the DNA bound state of the C-terminal part of the enzyme but does not capture catalytically productive complex of its N-terminal part (Figure 4A).

Structure of the N-terminal PD-(D/E)XK domain

The R.DpnI N-terminal catalytic domain has the typical PD-(D/E)XK-fold. DALI structure comparisons with other structures in the Protein Data Bank identified R.PvuII (30) and R.ThaI (31) as most similar PD-(D/E)XK endonucleases. The superpositions identify residue E25 of R.DpnI as the conserved active site glutamate in the helix upstream of the PD-(D/E)XK β-sheet (Supplementary Figure S5). Structure based assignment of the other catalytic residues is consistent with the bioinformatic analysis and with the results of the site-directed mutagenesis experiments (Figure 1B). R.DpnI catalytic domain possesses an unusual feature, a Zn^{2+} ion, which is liganded by four cysteine residues. Two of them are located between the catalytic glutamate (E25) and the 'PD' aspartate (D53), the other two are in a hairpin between the 'D' (D53) and the 'EXK' (E64 to K66).

Structure of the C-terminal domain in complex with target DNA

The C-terminal domain contains the predicted HTH motif and is best described as a winged helix (wH). It consists of 3 α-helices and a 3-stranded β-sheet. The most C-terminal helix ('recognition helix') inserts directly into the major groove of the target DNA, where the two m6A methyl groups are located in close proximity to each other. Both of them are buried in the pockets of the wH domain. The walls of the pockets are built from the main chain of the recognition helix and the hydrophobic parts of its polar side chains (E228, R231 and Q232 for one and E228, K229 and Q232 for the other methyl group; Figure 4B). Hydrogen bonding interactions between the wH domain and the target sequence are focused on the purine bases and except for a single one donated by the side chain of Q232 do not involve any of the pyrimidines (Figure 4C). The biological relevance of the crystallographically observed binding mode is supported by the structure comparisons. Many winged helix domains, including some with remarkably high similarity to the one of R.DpnI, bind to the major groove of the DNA, although the precise orientation of the oligoduplex might vary. The superposition of protein and DNA is particularly striking, e.g. for the comparison of R.DpnI and human Foxo3A DNA-binding domain (PDB code 2UZK, DOI:10.2210/pdb2uzk/pdb; Supplementary Figure S6).

Relevance of the wH binding mode to DNA

In order to further verify the relevance of the crystallographically observed DNA binding mode in solution, the R.DpnI wH domain (residues 181-254) was expressed in isolation in *E. coli*. The domain bound to oligo DNA containing fully methylated Gm6ATC sequence, although not as tightly as the entire enzyme. The binding was specific, because it was reduced by the replacement of a key residue in the wH domain (R231) with alanine, both in the context of the full-length enzyme and of the wH domain alone (Figure 5 and Supplementary Table S3). Moreover, it was sensitive to the DNA sequence and methylation state. The wH domain bound fully methylated DNA much more effectively than hemi-methylated or non-methylated DNA. A scrambled sequence was bound with considerable affinity only if the m6As were staggered in the correct way, despite overall weaker binding. The specificity of the

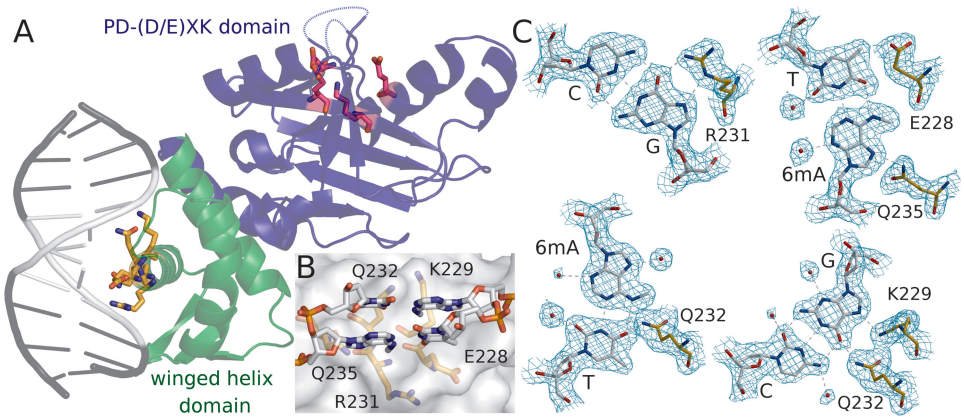

Figure 4. Crystal structure of the R.DpnI-DNA complex. (**A**) Overview of the R.DpnI structure. The active site residues of the catalytic domain (blue) are indicated in magenta and the residues responsible for sequence recognition by the wH domain (green) are in yellow. Two loops that could not be reliably traced in the electron density (residues 46–49 and 131–136) were marked by the dotted lines. (**B**) Detail of the binding of the central m6A-T pairs to the winged helix domain. (**C**) Sequence recognition by the winged helix domain. The composite omit map was contoured at 1 σ.

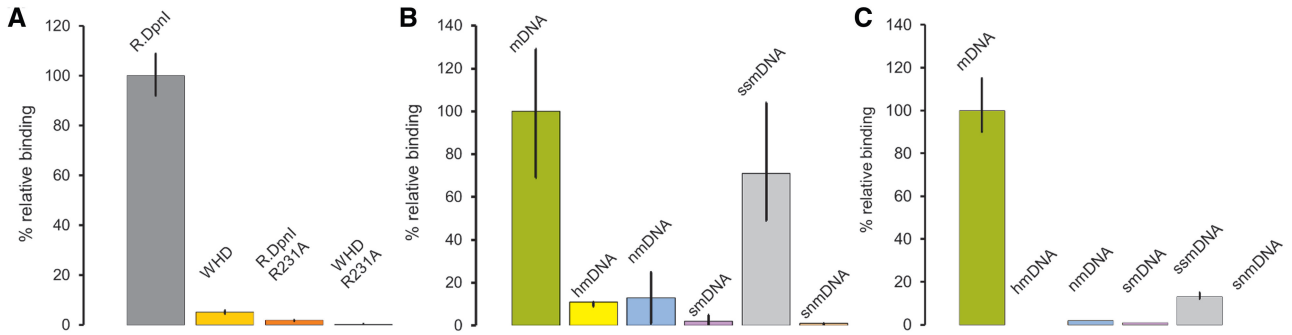

Figure 5. Binding assay of R.DpnI and its variants to different oligo DNAs. (**A**) Relative binding of different R.DpnI variants to methylated oligo DNA (mDNA). (**B**) Relative binding of the R.DpnI wH domain to different oligo DNAs. (**C**) Relative binding of the R.DpnI R231A variant to different oligo DNAs. The oligo coding is analogous as in Figure 3 (for data quantification see Supplementary Table S3, for sequences see Supplementary Table S4).

wH domain for the Gm6ATC target sequence was further confirmed by a competition experiment. When used in excess, it protected DNA against digestion with R.DpnI, but not R.AluI, which recognizes a scrambled version of the sequence with non-methylated adenine (AGCT), also cuts with blunt ends (Supplementary Figure S7). We concluded from these data that the binding mode of the wH domain to DNA observed in the crystal was biologically relevant and that the wH domain probes both DNA methylation state and its sequence.

Effects of changes in the wH domain on the activity of full length R.DpnI

In order to confirm the role of the wH domain in the context of full length R.DpnI, we selected residues within 5 Å spheres around the two m6A methyl groups (E228, K229, R231, Q232 and Q235) and substituted them one by one to alanines. Although the crystal structure indicates that the five residues are involved in the formation of the hydrophobic pockets as well as in direct hydrogen bonds with the bases, we did not see strong effects on the activity of R.DpnI against a plasmid containing fully methylated Gm6ATC sites (Supplementary Figure S8). However, in contrast to the wild-type enzyme, some of the variants did not cleave hemi-methylated DNA under our assay condition (Figure 6). The effects of alterations to the wH domain in the context of the full-length enzyme were confirmed in a filter binding assay. The full-length enzyme with alanine instead of R231 still bound well to methylated DNA, but not to hemi-methylated or non-methylated DNA or to other control substrates (Figure 5C). Together, the effects of substitutions in the wH domain suggested that the catalytic domain may have sequence and methylation specificity of its own.

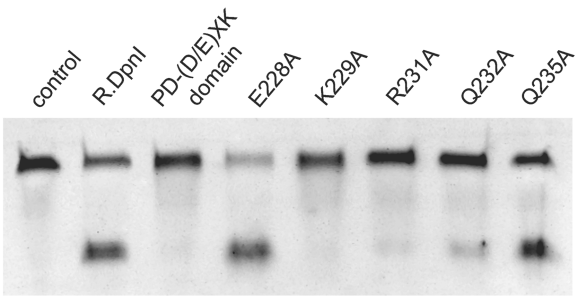

Figure 6. Hemi-methylated DNA (cuthmDNA) cleavage by R.DpnI variants with alanine substitutions in the wH domain. Control: non-cleaved substrate. The sequence of the used oligo DNA is available in Supplementary Table S4. R.DpnI concentration: 37.5 nM. dsDNA concentration: 1 μM.

Expression and properties of the catalytic domain

The PD-(D/E)XK domain of R.DpnI could be produced in isolation (using the same expression system as for the full length protein), but the yield was very low. Therefore, the experiments with this domain were carried out in crude cell extracts. The catalytic domain in isolation cleaved Dam-methylated, but not non-methylated DNA (Figure 7A). Hypermethylation of Dam-methylated DNA with Hia5 did not change the pattern of digestion fragments that were obtained after incubation with the catalytic domain (Figure 7B). Star activity was not observed, but this might have been due to the limited amounts of PD-(D/E)XK domain that we could express. We conclude from these experiments that the R.DpnI catalytic domain on its own cleaves DNA in a methylation- and sequence-specific manner.

A model for the interaction of the catalytic domain with DNA

As we could not produce enough of the catalytic domain to attempt co-crystallization with target DNA, we had to resort to molecular modeling to investigate the cleavage complex. Automatic structure comparisons with models in the Protein Data Bank indicated that the catalytic domain of R.DpnI was most similar to the R.PvuII monomer (PDB accession code 1F0O) (30). Therefore, we considered R.PvuII for our modeling, but confirmed that this particular choice did not influence our conclusions

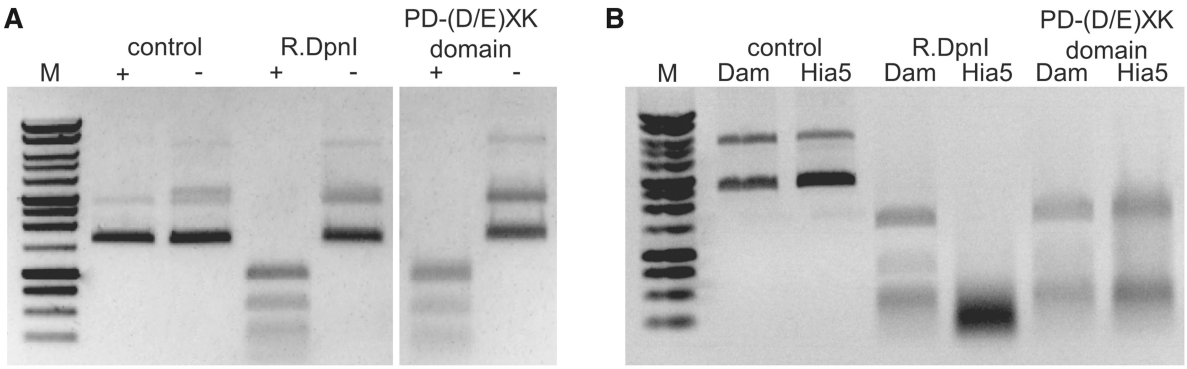

Figure 7. Properties of the PD-(D/E)XK domain. (**A**) Cleavage of Dam-methylated and non-methylated pUC18 DNA by R.DpnI and its catalytic domain. (**B**) Cleavage of Dam and Dam and Hia5-methylated pBR322 DNA by R.DpnI and its catalytic domain. M: molecular mass marker, control: non-cleaved substrate, '+': Dam-methylated DNA, '−': non-methylated DNA, Dam: Dam-methylated plasmid, Hia5: Hia5-methylated plasmid isolated from a Dam^+ *E. coli* strain. R.DpnI concentration: 37.5 nM, pUC18 plasmid concentration: 9.0 nM, pUC18 GATC sites concentration: 135 nM, pBR322 plasmid concentration: 5.6 nM, pBR322 GATC sites concentration: 123.2 nM and pBR322 SATS sites concentration: 537.6 nM.

in a major way. In a first step, we mapped the monomeric R.DpnI onto the two subunits of the R.PvuII dimer. In the second step, we then overlaid this complex onto a DNA duplex in the crystal, choosing the register so that the cleavage sites were superimposed. Two possible binding modes result, which reflect the need for two docking events to make a double strand cut with a single active site only (Supplementary Figure S9). The modeling predicts that the catalytic domain of R.DpnI approaches the DNA from the minor groove side, but almost wraps around it to make some contacts from the major groove side as well, particularly at the ends and flanks of the recognition sequence. The described modeling is too inaccurate to pinpoint sequence- or modification-specific contacts of the R.DpnI catalytic domain with DNA. However, the model does identify candidate residues that are located at the interface (Q18, R21, R126 and W138) or might take part in interactions if the 129–137 loop reorders in a productive complex (R135). We have substituted all these residues separately to alanines, and characterized the catalytic efficacy of the resulting R.DpnI variants (Supplementary Figure S10). All amino acid exchanges reduced the activity of the enzyme. However, further cleavage experiments with an excess of the mutant proteins also indicated that none of the exchanges abolished activity completely. The Q18A and R135A substitutions were tolerated best, suggesting that these residues play a lesser role in DNA binding. Together, the results of the mutagenesis experiments are consistent at least qualitatively with the model for the complex of the catalytic domain of R.DpnI with DNA.

Dependence of cleavage efficiency on the number of GATC sites

The presence of two binding sites for the same target sequence in a restriction endonuclease usually indicates that the enzyme behaves as a type IIE enzyme, i.e. that binding of the enzyme to an effector site is required to license cleavage of the substrate site (32). In order to probe the dependency of R.DpnI catalytic efficiency on the presence of effector sites, we compared DNA cleavage rates for substrates with either one or two target sequences combined in the same DNA. For an enzyme that is not influenced by remote sites, one would expect that the rate of first cleavage in the substrate should not be higher than the sum of rates for the substrates with only one site. This is not observed and instead, the data indicate that the presence of a second site within one DNA fragment influences R.DpnI cleavage efficiency. Residual cleavage of substrates with only one site could either indicate activation by another substrate *in trans* or a deviation from the ideal type IIE behavior (Supplementary Figure S1).

Catalytic and wH domains *in trans*

For type IIE enzymes, the catalytic and effector domains cooperate for DNA cleavage. The kinetic data suggest that R.DpnI behaves in some ways like a type IIE enzyme and that therefore the wH and catalytic domains of this enzyme should also act synergistically. Does this still work when the two domains are provided separately? In order to test this, we pre-incubated substrate DNA with an excess of R.DpnI wH domain and then attempted to cleave this DNA with either full-length enzyme or its catalytic domain, or a control restriction endonuclease recognizing a different site of the same size. Our data indicate that the wH domain in isolation prevents cleavage by R.DpnI or its catalytic domain, but lowers cleavage by the control endonuclease only slightly. This effect is most likely due to the competition of the winged helix domain with R.DpnI or its catalytic domain for target sites on the DNA (Figure 8 and Supplementary Figure S7).

DISCUSSION

Our analysis revealed that R.DpnI is a monomeric restriction endonuclease with an N-terminal catalytic PD-(D/E)XK domain and a C-terminal winged helix domain. Both domains are independently able to bind DNA in a sequence- and modification-specific manner. This study defines the binding interface between the DNA and wH domain with the help of a crystal structure. It was, so far, not possible to obtain a crystal structure with the DNA bound to the catalytic domain. Therefore, this interaction was modeled based on the structural similarity of R.DpnI to other PD-(D/E)XK restriction endonucleases, which have been crystallized with their DNA substrates. The resulting model is not sufficiently accurate to reveal the atomic details of sequence- or modification-specific recognition by the catalytic domain. However, characterization of R.DpnI variants with amino acid residues in the predicted binding interface altered by site-directed mutagenesis revealed that these residues are indeed important for the enzyme activity (Supplementary Figures S9 and S10).

The finding that R.DpnI has two independent binding sites for its target sequence can be interpreted in at least two ways. First, the structural data do not absolutely

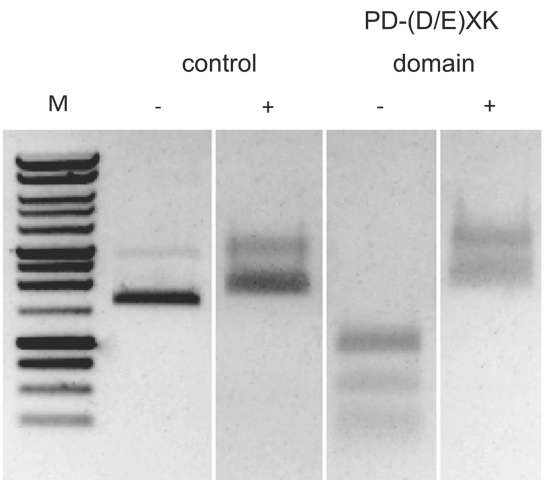

Figure 8. Pre-incubation of plasmid DNA with the R.DpnI wH domain. M: molecular mass marker, '−': cleavage reaction without pre-incubation with the wH domain, '+': cleavage reaction after pre-incubation with the wH domain (5-fold molar excess of wH domain over Gm6ATC sites was used).

exclude that catalytic and wH domain bind to the same instance of one target site. In models that position both domains on the same DNA (Supplementary Figure S9), the resulting clashes between the domains are mild, and the distances of ~30 and 40 Å between the C-terminus of the catalytic domain and the N-terminal end of the wH domain (for the two alternative binding modes) could be bridged by local refolding. However, the finding that R.DpnI behaves kinetically like a type IIE enzyme makes it more likely that the wH and catalytic domains bind to two different instances of the target sequence. This model avoids the need for local rearrangements in the protein structure, and has several other attractive features. First, the need for binding of two copies of the target sequence should enhance fidelity. Thus, the model helps to explain how relatively weak attractive van der Waals interactions between the m6A methyl groups and the protein can stringently license the enzymatic reaction without violation of thermodynamic principles. Second, for a monomeric restriction endonuclease, which can introduce double-strand breaks in DNA only by sequential nicking, anchoring of the enzyme to another site in the vicinity could be a good way to ensure that it completes the cleavage of a nicked substrate. Recently, bioinformatic analyses supported by mutagenesis showed that the modification-dependent enzyme Mrr is also composed of a wH and PD-(D/E)XK domains (33). It will be interesting to determine whether Mrr and related modification-dependent restriction enzymes will be similar to R.DpnI in terms of the dependence on effector sites.

Modification dependence, a new theme in restriction endonuclease structural research

Early structural studies of type II restriction endonucleases were concentrated on the very widespread PD-(D/E)XK enzymes and the mechanisms that explained the exquisite specificity of these enzymes, mostly hydrogen bonds and shape selection (34). Subsequent studies have added to our understanding of DNA recognition by indirect readout (35), semi-degenerate recognition (36,37), and to unusual modes of DNA recognition by nucleotide flipping (38,39), protein intercalation into DNA (31,40) or disruption of the Watson–Crick pairing (41). More recently, there has been much interest in the crystallographic elucidation of structures of restriction endonucleases that are phylogenetically unrelated to the PD-(D/E)XK enzymes. This has led to the solution of crystal structures of ββα-Me (or HNH) (41,42), GIY-YIG (43,44), half-pipe (45) and phospholipase-like (46) restriction enzymes. The topic of modification-dependent restriction endonucleases poses an interesting puzzle: how can DNA cleavage depend on the presence of a modification? As attractive van der Waals interactions are much weaker than repulsive ones (47), this question is much harder to answer than the reverse one how DNA cleavage can be prevented. Surprisingly, this question has been largely neglected, particularly by the crystallographic community. Therefore, we anticipate that much remains to be discovered in the area of modification-dependent restriction endonucleases.

ACCESSION NUMBERS

4ESJ.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–10.

ACKNOWLEDGEMENTS

Dr Monika Radlinska (Warsaw University) kindly provided the Hia5-methylated plasmid for specificity studies. Prof. Thomas Sorensen and Dr James Sandy [Diamond Light Source (DLS), Didcot, UK] and Prof. Hans Bartunik (Max-Planck-Arbeitsgruppen fuer Strukturelle Molekularbiologie, Hamburg) granted beamtime and assisted during data collection at the IO2 beamline of DLS, UK and BW6 beamline of DESY, Germany.

FUNDING

The Polish Ministry of Science and Higher Education [MNiSW, 188/N-DFG/2008/0]. W.S. was funded by MNiSW [N N401 585738]. K.S. and J.M.B. were funded by the European Research Council to J.M.B. [StG 261351]. J.M.B. has been also supported by the 'Ideas for Poland' fellowship from the Foundation for Polish Science (FNP). H.C. and M.B. are grateful for a TEAM grant from the FNP and the EU European Regional Development Fund to M.B. [TEAM/2010-6/1]. M.B. also acknowledges funding from a MNiSW grant [0295/B/PO1/2008/34] and in-house funding from Cardiff University. Funding for open access charge: TEAM/2010-6/1.

Conflict of interest statement. None declared.

REFERENCES

1. Pingoud,A.M. (ed.), (2004) *Restriction endonucleases*. Springer, Berlin.
2. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., Dryden,D.T.F., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
3. Raleigh,E.A. and Wilson,G. (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl Acad. Sci. USA*, **83**, 9070–9074.
4. Pieper,U. and Pingoud,A. (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from Escherichia coli. *Biochemistry*, **41**, 5236–5244.
5. Bujnicki,J.M. and Rychlewski,L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol.*, **3**, 69–72.
6. Aertsen,A. and Michiels,C.W. (2005) Mrr instigates the SOS response after high pressure stress in Escherichia coli. *Mol. Microbiol.*, **58**, 1381–1391.
7. Tesfazgi Mebrhatu,M., Wywial,E., Ghosh,A., Michiels,C.W., Lindner,A.B., Taddei,F., Bujnicki,J.M., Van Melderen,L. and

Aertsen,A. (2011) Evidence for an evolutionary antagonism between Mrr and Type III modification systems. *Nucleic Acids Res.*, **39**, 5991–6001.
8. Zheng,Y., Cohen-Karni,D., Xu,D., Chin,H.G., Wilson,G., Pradhan,S. and Roberts,R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
9. Lacks,S. and Greenberg,B. (1975) A deoxyribonuclease of Diplococcus pneumoniae specific for methylated DNA. *J. Biol. Chem.*, **250**, 4060–4066.
10. de la Campa,A.G., Springhorn,S.S., Kale,P. and Lacks,S.A. (1988) Proteins encoded by the DpnI restriction gene cassette. Hyperproduction and characterization of the DpnI endonuclease. *J. Biol. Chem.*, **263**, 14696–14702.
11. Lacks,S.A., Mannarelli,B.M., Springhorn,S.S. and Greenberg,B. (1986) Genetic basis of the complementary DpnI and DpnII restriction systems of S. pneumoniae: an intercellular cassette mechanism. *Cell*, **46**, 993–1000.
12. Lu,L., Patel,H. and Bissler,J.J. (2002) Optimizing DpnI digestion conditions to detect replicated DNA. *Biotechniques*, **33**, 316–318.
13. Wilson,W.W., Mebane,E.W. and Hoffman,R.M. (1993) Creation of ultra-rare restriction sites in intact eucaryotic chromosomes mediated by bacterial methylases: an approach to sequencing and analyzing tumor and normal genomes. *Anticancer Res.*, **13**, 17–20.
14. Weil,M.D. and McClelland,M. (1989) Enzymatic cleavage of a bacterial genome at a 10-base-pair recognition site. *Proc. Natl Acad. Sci. USA*, **86**, 51–55.
15. Waterbury,P.G., Rehfuss,R.P., Carroll,W.T., Smardon,A.M., Faldasz,B.D., Huckaby,C.S. and Lane,M.J. (1989) Specific cleavage of the yeast genome at 5′-ATCGATCGAT-3′. *Nucleic Acids Res.*, **17**, 9493.
16. Sambrook,J., MacCallum,P. and Russell,D. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
17. Conlan,L.H. and Dupureur,C.M. (2002) Dissecting the metal ion dependence of DNA binding by PvuII endonuclease. *Biochemistry*, **41**, 1335–1342.
18. Sheldrick,G.M. (2002) Macromolecular phasing with SHELXE. *Zeitschrift für. Kristallographie*, **217**, 644–650.
19. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1772–1779.
20. Morris,R.J., Perrakis,A. and Lamzin,V.S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.*, **374**, 229–244.
21. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
22. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. **D53**, 240–255.
23. Brunger,A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
24. Orlowski,J. and Bujnicki,J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552–3569.
25. Sui,M.J., Tsai,L.C., Hsia,K.C., Doudeva,L.G., Ku,W.Y., Han,G.W. and Yuan,H.S. (2002) Metal ions and phosphate binding in the H-N-H motif: crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. *Protein Sci.*, **11**, 2947–2957.
26. Kleinstiver,B.P., Berube-Janzen,W., Fernandes,A.D. and Edgell,D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS One*, **6**, e23804.
27. Sapranauskas,R., Sasnauskas,G., Lagunavicius,A., Vilkaitis,G., Lubys,A. and Siksnys,V. (2000) Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of Salmonella typhimurium. *J. Biol. Chem.*, **275**, 30878–30885.
28. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
29. Drozdz,M., Piekarowicz,A., Bujnicki,J.M. and Radlinska,M. (2012) Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res.*, **40**, 2119–2130.
30. Horton,J.R. and Cheng,X. (2000) PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.*, **300**, 1049–1056.
31. Firczuk,M., Wojciechowski,M., Czapinska,H. and Bochtler,M. (2011) DNA intercalation without flipping in the specific ThaI-DNA complex. *Nucleic Acids Res.*, **39**, 744–754.
32. Yang,C.C., Baxter,B.K. and Topal,M.D. (1994) DNA cleavage by NaeI: protein purification, rate-limiting step, and accuracy. *Biochemistry*, **33**, 14918–14925.
33. Orlowski,J., Mebrhatu,M.T., Michiels,C.W., Bujnicki,J.M. and Aertsen,A. (2008) Mutational analysis and a structural model of methyl-directed restriction enzyme Mrr. *Biochem. Biophys. Res. Commun.*, **377**, 862–866.
34. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
35. Martin,A.M., Sam,M.D., Reich,N.O. and Perona,J.J. (1999) Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nat. Struct. Biol.*, **6**, 269–277.
36. Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res.*, **35**, 2035–2046.
37. Sokolowska,M., Kaus-Drobek,M., Czapinska,H., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Monomeric restriction endonuclease BcnI in the apo form and in an asymmetric complex with target DNA. *J. Mol. Biol.*, **369**, 722–734.
38. Bochtler,M., Szczepanowski,R.H., Tamulaitis,G., Grazulis,S., Czapinska,H., Manakova,E. and Siksnys,V. (2006) Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J.*, **25**, 2219–2229.
39. Szczepanowski,R.H., Carpenter,M.A., Czapinska,H., Zaremba,M., Tamulaitis,G., Siksnys,V., Bhagwat,A.S. and Bochtler,M. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res.*, **36**, 6109–6117.
40. Horton,N.C., Dorner,L.F. and Perona,J.J. (2002) Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. *Nat. Struct. Biol.*, **9**, 42–47.
41. Shen,B.W., Heiter,D.F., Chan,S.H., Wang,H., Xu,S.Y., Morgan,R.D., Wilson,G.G. and Stoddard,B.L. (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI. *Structure*, **18**, 734–743.
42. Sokolowska,M., Czapinska,H. and Bochtler,M. (2009) Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.*, **37**, 3799–3810.
43. Sokolowska,M., Czapinska,H. and Bochtler,M. (2011) Hpy188I-DNA pre- and post-cleavage complexes: snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.*, **39**, 1554–1564.
44. Mak,A.N., Lambert,A.R. and Stoddard,B.L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure*, **18**, 1321–1331.
45. Miyazono,K., Watanabe,M., Kosinski,J., Ishikawa,K., Kamo,M., Sawasaki,T., Nagata,K., Bujnicki,J.M., Endo,Y., Tanokura,M. *et al.* (2007) Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res.*, **35**, 1908–1918.
46. Grazulis,S., Manakova,E., Roessle,M., Bochtler,M., Tamulaitiene,G., Huber,R. and Siksnys,V. (2005) Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl Acad. Sci. USA*, **102**, 15797–15802.
47. Slater,J.C. and Kirkwood,J.G. (1931) The van der Waals forces in gases. *Phys. Rev.*, **37**, 682–697.
48. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.