Published online 20 April 2009 Nucleic Acids Research, 2009, Vol. 37, No. 11 3799–3810
doi:10.1093/nar/gkp228

Crystal structure of the ββα-Me type II restriction endonuclease Hpy99I with target DNA

Monika Sokolowska[1,2], Honorata Czapinska[1,2] and Matthias Bochtler[1,2,3,*]

[1]International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland, [2]Max-Planck-Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr, 108, 01309 Dresden, Germany and [3]Schools of Chemistry and Biosciences, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK

Received January 6, 2009; Revised and Accepted March 25, 2009

ABSTRACT

The ββα-Me restriction endonuclease (REase) Hpy99I recognizes the CGWCG target sequence and cleaves it with unusual stagger (five nucleotide 5′-recessed ends). Here we present the crystal structure of the specific complex of the dimeric enzyme with DNA. The Hpy99I protomer consists of an antiparallel β-barrel and two β4α2 repeats. Each repeat coordinates a structural zinc ion with four cysteine thiolates in two CXXC motifs. The ββα-Me region of the second β4α2 repeat holds the catalytic metal ion (or its sodium surrogate) via Asp148 and Asn165 and activates a water molecule with the general base His149. In the specific complex, Hpy99I forms a ring-like structure around the DNA that contacts DNA bases on the major and minor groove sides via the first and second β4α2 repeats, respectively. Hpy99I interacts with the central base pair of the recognition sequence only on the minor groove side, where A:T resembles T:A and G:C is similar to C:G. The Hpy99I–DNA co-crystal structure provides the first detailed illustration of the ββα-Me site in REases and complements structural information on the use of this active site motif in other groups of endonucleases such as homing endonucleases (e.g. I-PpoI) and Holliday junction resolvases (e.g. T4 endonuclease VII).

INTRODUCTION

Type II restriction endonucleases (REases) can be classified into conventional PD-(D/E)XK, ββα-Me, GIY-YIG, phospholipase-derived and half-pipe endonucleases according to their folds and active site structures (1). Although the ββα-Me enzymes form the second largest group after the PD-(D/E)XK enzymes, structures of REases in this group have not yet been determined. The available models for ββα-Me restrictases (2) are, therefore, all based on structures of ββα-Me endonucleases that are not restriction enzymes (3).

The ββα-Me nucleases occur in all three kingdoms of life. They are defined by a conserved active site β-hairpin (ββ) and α-helix (α) that anchor a catalytic metal ion (Me). The superfamily merges HNH (4) and His-Cys box (5) nucleases that were named earlier for the occurrence of conserved residues in these subgroups (3). Structurally characterized ββα-Me nucleases include unspecific nucleases [e.g. colicins (6–8), NucA (9), Serratia endonuclease (10), periplasmic Vvn endonuclease (11)], homing endonucleases [e.g. I-PpoI (12,13) and I-HmuI (14)] and Holliday junction resolvases [e.g. phage T4 endonuclease VII (15)]. The ββα-Me regions of these enzymes bind one metal ion per subunit with one (e.g. I-PpoI), two (e.g. T4 endonuclease VII) or three (e.g. colicin E7) amino-acid ligands. The ββα-Me region contains a strictly conserved histidine residue that activates a water molecule for incorporation into the DNA substrate (12). The oligomeric state of ββα-Me endonucleases varies. Colicin E9, Vvn and the homing endonuclease I-HmuI are monomeric (11,14,16). In contrast, I-PpoI and the T4 endonuclease VII form the dimers that are expected for enzymes that cut both DNA strands (13,15).

Many ββα-Me nucleases cut both DNA strands with a stagger that leads to 5′-recessed ends in the cleavage products. This is true for most homing endonucleases and with some exceptions (e.g. Eco31I, SapI) also for the ββα-Me REases (17). The rule applies to Hpy99I: this enzyme cleaves at the 3′-ends of its recognition sequence so that cleavage products have five nucleotides long 3′-overhangs. It applies also to ββα-Me REases: MnlI, HphI, MboII and Hin4II all recognize asymmetric sequences and cut them five to eight nucleotides away from the recognition site so that single nucleotide 3′-overhangs result. PacI cleaves inside an eight nucleotide palindromic sequence with a 2-bp stagger. KpnI, NlaIII, SphI and NspI recognize short palindromic sequences and cleave within or immediately adjacent to the recognition

*To whom correspondence should be addressed. Tel: +48 22 5970732 or +44 29 208 70625; Fax: +48 22 5970715; Email: mbochtler@iimcb.gov.pl or bochtlerm@cardiff.ac.uk

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

sequence with 4-bp stagger. In contrast to the ββα-Me REases most PD-(D/E)XK enzymes cut DNA blunt ended or with 5′-overhangs, and only a few exceptional enzymes in this group such as SdaI, BglI and SfiI generate 5′-recessed ends (17–20).

The type II REase Hpy99I from the gastric pathogen *Helicobacter pylori* can be classified as a ββα-Me endonuclease on the basis of statistically significant sequence similarity to the Holliday junction resolvase T4 endonuclease VII, a *bona fide* ββα-Me endonuclease (1). The enzyme recognizes the nearly symmetric (pseudopalindromic) recognition sequence CGWCG (W stands for either A or T, as opposed to S for G or C). The bases at the center of the recognition sequence are not individually recognized, but the enzyme can nevertheless distinguish A:T pairs from G:C pairs with high fidelity (21). Here we present co-crystal structures of Hpy99I with DNA at 1.5 and 1.75 Å resolution that were obtained in the presence of 5 mM $CaCl_2$ and 1 mM EDTA, respectively. Our structures explain the stagger between cutting sites in the two DNA strands, the specificity for the DNA target and the distinction between A:T and G:C pairs (W versus S) at the center of the recognition sequence. They also shed more light on the relationships between ββα-Me REases, homing endonucleases and Holliday junction resolvases.

MATERIALS AND METHODS

Cloning

Codon optimized Hpy99I REase (*hpy99IR*) and Hpy99I MTase (*hpy99IM*) synthetic genes in pBluescript II SK (+) vectors were purchased from Epoch Biolabs, Inc. (Texas, USA). The *hpy99IM* gene was PCR-amplified using primers that were designed to introduce a Shine–Dalgarno sequence eight nucleotides upstream of the ATG start codon and HindIII and BamHI cloning sites upstream and downstream of the gene. Using these sites, the fragment was then placed into the tetracycline resistance region of the pACYC184 (Cm^r) vector. For the expression of the REase, we used pET15bmod (Ap^r), a derivative of pET15b (+) (Ap^r). In this vector, the original EcoRI site has been deleted and a new EcoRI site has been introduced immediately downstream of the codons for the hexa-histidine tag. The *hpy99IR* gene was cloned into this vector using the EcoRI and XhoI restriction sites. The resulting construct coded for the full-length protein with the N-terminal MGHHHHHHEF tag.

Mutagenesis

Hpy99I point mutations D148A, H149A and N165A were introduced by PCR-based site-directed mutagenesis according to the Stratagene protocol with Pfu Plus DNA polymerase (EURx) and confirmed by Sanger dideoxy sequencing. A truncated version of Hpy99I (residues 54–190) was generated by cloning of the corresponding PCR fragment into the expression vector pET15bmod (Ap^r) between the EcoRI and XhoI sites. The resulting construct encoded the Hpy99I fragment starting at residue Ile54 fused to the N-terminal MGHHHHHHEF tag.

REase expression

Expression experiments were done in *Escherichia coli* ER2566 strain (F^- λ^- *fhuA2 [Ion] ompT lacZ::T7geneI gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::*miniTn*10)2 R(zgb-210::*Tn*10)1* (Tet^s) *endA1 [dcm]*). This strain was transformed sequentially with plasmid pACYC184 (Cm^r) bearing the *hpy99IM* gene and then with plasmid pET15bmod (Ap^r) containing the *hpy99IR* gene. Cells were grown in Luria-Bertani (LB) broth medium with appropriate antibiotics at 37°C to OD_{600} 0.7 and induced with 0.5 mM IPTG. After 4 h of induction the cells were harvested by centrifugation and the pellet was stored at –20°C. Wild-type (wt) and mutant Hpy99I proteins were expressed with similar efficiency. The wt and the point mutants were predominantly found in the soluble fraction, but more than 90% of $Hpy99I_{54–190}$ was insoluble.

REase purification

Frozen cells expressing wt Hpy99I REase were thawed and resuspended in buffer A (20 mM Tris/HCl pH 7.6, 500 mM NaCl and 1 mM PMSF). Cells in suspension were opened by sonication and the cell debris was removed by centrifugation at 145 000×*g* for 40 min. Hpy99I was purified by affinity chromatography on Nickel Nitrilotriacetic acid (Ni-NTA) Agarose column (Qiagen). The protein was eluted using an imidazole gradient in buffer B (20 mM Tris/HCl pH 7.6, 200 mM NaCl, 5 mM 2-mercaptoethanol). Fractions containing Hpy99I were combined and concentrated by ultrafiltration using Vivaspin concentrators (10 kDa MWCO). The protein was purified further by size-exclusion chromatography on HiLoad 16/60 Superdex 75 column (GE Healthcare), equilibrated with buffer C (20 mM Tris/HCl pH 7.6, 200 mM NaCl, 1 mM EDTA and 1 mM DTT) prior to co-crystallization in the presence of EDTA or with buffer D (20 mM Tris/HCl pH 7.6, 200 mM NaCl, 5 mM $CaCl_2$ and 1 mM DTT) prior to co-crystallization in the presence of Ca^{2+} ions. Fractions containing Hpy99I REase activity were pooled and concentrated to 13–15 mg/ml. From a 2 l culture, about 9 mg of protein was obtained that appeared homogeneous on Coomassie-stained SDS-PAGE. The mutant proteins were purified according to the same protocol. Point mutants behaved like the wt protein. Some initially soluble $Hpy99I_{54–190}$ precipitated in the dialysis and concentration steps, which reduced the already low yield of this truncated form further.

REase activity assay

DNA cleavage activity was assayed with pUC19 plasmid as a substrate. Cleavage reactions were performed at 37°C with 0.24 pmol (0.20 μg) pUC19 and 16–125 pmol (0.53–5.64 μg) of the protein dimer in 50 mM potassium acetate, 20 mM Tris/acetate pH 7.9. The buffers were supplemented with 10 μM to 10 mM magnesium acetate, 10 μM to 10 mM manganese chloride, 10 μM to 10 mM calcium acetate, 10 μM to 10 mM zinc acetate or 10 mM EDTA and 1 mM of DTT as the reducing agent. As DTT reduced Mn^{2+} and Zn^{2+}, 5 mM 2-mercaptoethanol was

used instead in the reactions with these metal cations. Reactions were performed in 10 µl reaction volume for 1 min and terminated by addition of a stop buffer containing 75 mM EDTA, 50% glycerol, 0.3% SDS and 0.1% bromophenol blue. DNA cleavage products were analyzed electrophoretically on a 1.4% agarose gel. DNA was visualized by ethidium bromide staining and illuminated with UV light.

Crystallization

Oligonucleotides 5′-CT*CGACG*TAGA-3′ and 5′-TA*CGT CG*AGTC-3′ were purchased from Metabion. They were dissolved in 10 mM Tris/HCl pH 7.9 and annealed by heating up to 95°C followed by slow cooling to 4°C to yield a duplex with two nucleotide recessed ends. Hpy99I protein in buffer C or D was concentrated to 13–15 mg/ml and mixed with the oligoduplex at a 1:1.1 molar ratio (protein dimer:duplex DNA). Hpy99I–DNA co-crystals were grown by the vapor-diffusion technique. Initial high-throughput screens were set up at the 200 nl scale using a Cartesian pipetting robot and 96-well Greiner sitting drop plates. Crystallization trials with larger drop volumes were pipetted in CRYSCHEM plates (Hampton Research). Crystals of Hpy99I–DNA complex in the presence of EDTA were obtained with the reservoir containing 0.1 M MES/NaOH pH 6.5, 0.2 M sodium chloride, 0.1 M lithium sulfate and 30% PEG 400. For crystallization in the presence of Ca^{2+}, a reservoir of 0.1 M HEPES/NaOH pH 7.5, 0.1 M sodium chloride and 30% PEG 400 was used. Drops were formed by mixing 2 µl of the protein–DNA solution with 2 µl of the reservoir buffer and equilibrated over the reservoir at 21°C. Needle-shaped crystals of Hpy99I–DNA complex grew within 3 days from both the EDTA and Ca^{2+}-containing solutions. A bromide derivative was obtained by soaking in crystallization buffer supplemented with 1 M sodium bromide for 30 sec prior to flash-cryocooling. All Hpy99I–DNA crystals could be flash-cryocooled directly from the crystallization drop without additional cryoprotection.

Structure determination

All diffraction data were collected at beamline 14.1 of the Berliner Elektronenspeicherring BESSY. Processing and scaling were done with the XDS and XSCALE programs (22). Crystals belonged to space group H32 (R32 in rhombohedral setting) and contained an Hpy99I dimer and one target DNA duplex in the asymmetric unit. The best crystal from the preparation with Ca^{2+} diffracted to 1.5 Å resolution. A high-resolution dataset was measured at wavelength of 0.91841 Å. In order to investigate the anomalous scattering of metal ions, additional diffraction data were collected above the Zn^{2+} edge (in energy) at a wavelength of 1.26516 Å (calculated f'' for Zn 3.78) and below the Zn^{2+} edge (in energy) at a wavelength of 1.31899 Å (calculated f'' for Zn 0.51). The best crystal from the preparation with EDTA diffracted to 1.75 Å and was measured at a wavelength of 0.97982 Å. Another crystal that was grown in the same conditions was soaked with bromide. For this soak, datasets (to 2.2 Å resolution) were collected at the bromide absorption maximum (wavelength 0.92007 Å), inflection point (wavelength 0.92027 Å) and at a remote high energy (wavelength 0.88561 Å). Anomalous and dispersive differences were interpreted with SHELXD (23) run in SAD mode (for peak wavelength only) or MAD mode (for all wavelengths combined). Correlation coefficients and the contrast—a SHELXE measure of map quality after density modification (24)—indicated that solutions were found in all cases. ARP/wARP (25) could interpret these densities for the correct hand, but with an unexpected twist: the most complete model was built when only the SAD data at 0.92007 Å resolution were used to determine phases. Inclusion of the inflection point data had a detrimental effect. In retrospect, this observation could be traced to the presence of four structural Zn^{2+} ions. At 0.92007 Å wavelength, the anomalous signal of the Zn^{2+} ions ($f'' = 2.2$) should be similar to the anomalous signal of bromide ions with partial occupancy ($f'' = 3.8$ for full occupancy). Shifting the wavelength to 0.92027 Å barely affects the scattering of the Zn^{2+} ions, but halves the anomalous signal from bromide ions. Treating these effects correctly turned out to be unnecessary because SAD maps were of sufficient quality for ARP/wARP (25) to deliver a nearly complete model of the Hpy99I dimer without the DNA. A very rough model of the bound DNA in idealized geometry was built with the 3DNA (26) program and then manually adjusted to fit the electron density. The structures were improved with the modeling programs O (27) and COOT (28) and refined with REFMAC (29) and CNS (30). Data collection and refinement parameters are summarized in Table 1. The refined models and corresponding structure factors were deposited in the Protein Data Bank (PDB) with accession codes 3GOX (Hpy99I–DNA complex in the absence of EDTA and presence of Ca^{2+}) and 3FC3 (Hpy99I–DNA complex in the presence of EDTA).

Determination of metal ion identity

The identity of the Zn^{2+} ions was confirmed by anomalous difference Fourier maps for datasets collected above and below the Zn^{2+} edge (peak heights ~60σ at 1.26516 Å, where $f'' = 3.8$ and only ~10σ at 1.31899 Å, where $f'' = 0.5$). The metals in the active centers were readily identified as Na^{+} ions from the crystallization buffer in the crystals grown in the presence of EDTA. This assignment agreed with Na^{+} coordination preferences (31), observed ligand distances close to 2.4 Å and temperature factors above average for the protein even for a third row cation (26 $Å^2$ versus 22 $Å^2$). It was also fairly consistent with valence bond sum calculations (32), which should be applicable at the resolution of the dataset. With Na^{+} parameters, the valence bond sum was calculated to be 1.2 in reasonable agreement with the expected value 1.0.

In the case of the Hpy99I–DNA crystals that were grown in the presence of Ca^{2+} ions (and Na^{+} in the crystallization buffer), the identity of metal ions in the active site was more difficult to determine. Anomalous difference Fourier maps (calculated with model phases)

at low X-ray energies (λ = 1.26516 Å and λ = 1.31899 Å) had significant peaks (between 5σ and 10σ) for most sulfur and phosphorus atoms, but no peaks for the active site metal ions. Therefore, the majority of these ions had to be third row elements (Na^+, Mg^{2+}) rather than fourth row elements (K^+, Ca^{2+}, Mn^{2+}, Zn^{2+}). Temperature factors supported this conclusion. They refined to values close to the protein average for third row cations (20 $Å^2$) and to substantially higher values (25 $Å^2$ and larger) for fourth row cations. Moreover, most of the DNA in the crystals was not cleaved. Therefore, we concluded that Na^+ atoms predominated in the active sites. However, according to the valence bond sum for sodium (1.4 instead of 1.0), the actual ligand distances were on average too short (31), perhaps pointing to the presence of Mg^{2+} ions in a minority of active sites (33). This interpretation would be consistent with residual difference density that suggested cleavage of a fraction of the DNA in the crystals. Unfortunately, a 'pure' Hpy99I–DNA product complex with Mg^{2+} ions in the active sites could not be obtained because crystals that were soaked with Mg^{2+} did not diffract and those that were grown in the presence of Mg^{2+} were too small for crystallographic analysis.

RESULTS

Expression, purification and metal dependence

The Hpy99I REase with an N-terminal hexa-histidine tag and the cognate methyltransferase were co-expressed from synthetic genes in *E. coli* strain ER2566. The REase was purified by a combination of affinity chromatography and size-exclusion chromatography. The activity of Hpy99I was tested in the presence of various divalent metal cations. As a substrate, we used plasmid pUC19, which has five Hpy99I cleavage sites at positions 372, 385, 907, 1701 and 1964. Prior to Hpy99I digestion, the plasmid migrated as a mixture of the supercoiled and nicked forms (Figure 1, lanes labeled 'C'). Two hundred nanograms of plasmid (0.24 pmol, containing 1.20 pmol Hpy99I recognition sites) were partly linearized by 16 pmol of wt Hpy99I and fully digested by at least 31 pmol of wt Hpy99I within 1 min at 37°C, but not by the same amount of several Hpy99I mutants (Figure 1A and B). The Mg^{2+} concentration could be reduced from 10 mM to 20 μM, but a further decrease eventually abolished the activity of the wt enzyme (Figure 1C, lanes 3–6). Mn^{2+} could substitute for Mg^{2+}, albeit at slightly higher concentration (Figure 1C, lanes 7–10). Hpy99I was inactive in the presence of Ca^{2+} and Zn^{2+} (Figure 1C, lanes 11–18) and also in the presence of 10 mM EDTA (Figure 1C, lane 19).

Structure determination

High-throughput crystallization trials with Hpy99I and an 11-mer DNA duplex with two nucleotide recessed ends and the target sequence in the center were set up with either Ca^{2+} ions or EDTA. Crystals were obtained in both cases, although with slightly different buffer compositions. Crystallization was reproducible on a larger scale, and finally led to essentially isomorphous crystals that diffracted to 1.5 Å (in the absence of EDTA) and 1.75 Å (in the presence of EDTA) resolution at a synchrotron beamline. The crystals belonged to space group H32 and contained a single Hpy99I dimer with cognate DNA in the asymmetric unit. For phasing we used a bromide-soaked crystal that had been grown in the presence of EDTA. SAD data were collected at 0.92007 Å wavelength where bromide absorption is maximal. Retrospective analysis showed that structure determination was significantly aided by the anomalous scattering of four Zn^{2+} ions, which bound to the Hpy99I dimer and were retained in the presence of 0.5–1.0 mM EDTA in the purification and crystallization steps. The final quality parameters for the refined structures are summarized in Table 1. The two structures differ slightly in the metal content of the active sites, but are otherwise very similar.

Overall structure

The Hpy99I–DNA co-crystal structures show the specific DNA complex with the enzyme locked onto its target site. In this complex, the Hpy99I dimer forms a tight ring around the DNA (Figure 2 and Supplementary Figure 1). In the orientation of Figure 2, the elongated Hpy99I protomers are predominantly located on the sides of the DNA (Figure 2A). The subunit on the left (colored in Figure 2A) is poised to cleave the scissile phosphate closer to the viewer (Figure 2B). It wraps around the DNA and makes both major and minor groove contacts

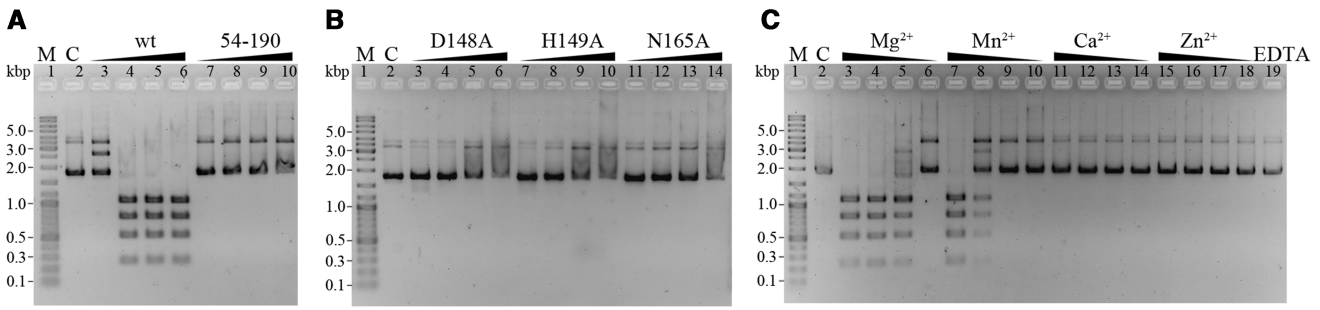

Figure 1. Activity of wild-type and mutant Hpy99I proteins. (**A**) Mixture of supercoiled and nicked pUC19 plasmid DNA (lanes labeled 'C') was digested to fragments of predicted sizes 1094, 794, 522, 263 and 13 bp that are resolved with the exception of the smallest fragment. Digestion experiments were carried out with either (**A**, **B**) increasing amounts of enzyme (16, 31, 62, 125 pmol) and 10 mM Mg^{2+} or (**C**) 31 pmol of wt Hpy99I and decreasing concentrations of divalent metal cations (10 mM, 40 μM, 20 μM, 10 μM).

with the bases above but not below the central base pair (Figure 2C and D). In addition, it contacts the minor groove face of the central base pair (Figure 2E). A region close to the C-terminus of the 'left' subunit crosses over to the right where it contacts the other subunit. The interface between the subunits extends over ~2750 Å^2 [a large value for proteins in this mass range (34)], and their combined interface with the DNA measures ~2200 Å^2.

Hpy99I subunit structure

The Hpy99I protomer consists of an N-terminal β-barrel (residues 1–53), a linker sequence (residues 54–64) and two repeats (residues 65–126 and 127–190) that we term β4α2 repeats because they consist of four β-strands (in two hairpins) and two α-helices each (Figure 2A and 3). The β-barrel has the SH3-domain fold that forms a common structural unit of many proteins, including some that bind nucleic acids such as HIV integrase (35). In the Hpy99I–DNA co-crystal structure, the β-barrel is too far away to interact with the specifically recognized bases, but contacts with a secondary site in longer DNA cannot be ruled out. Deletion of the β-barrel domain (residues 1–53) drastically reduces the solubility of the *E. coli* overexpressed protein and abolishes activity at all tested concentrations (Figure 1A, lanes 7–10). The β-barrel is connected by an 11-residue linker to two β4α2 repeats. Each repeat contains two CXXC motifs that map to the first β-hairpin (HP1) and the first α-helix (H1). The four cysteine residues coordinate a structurally bound Zn^{2+} ion tetrahedrally (Figure 3A–C). Ligand distances and angles suggest that all cysteines are present in their thiolate forms (36). The arrangement is stabilized by polar groups nearby, particularly main chain NH groups as previously reported for other Zn^{2+}-ion-binding sites with four cysteine ligands (36,37). Despite low overall sequence similarity, large regions of the two repeats (residues 65–78, 95–106 and residues 127–140, 159–170) superimpose very well (rmsd for Cα-atoms 1.1 Å (38)). The second repeat encompasses the ββα-Me region that classifies Hpy99I as a ββα-Me endonuclease (Figure 3B).

Table 1. Data collection and refinement statistics

	Hpy99I–DNA in the presence of EDTA (λ = 0.97982 Å)	Hpy99I–DNA in the presence of Ca^{2+} (λ = 0.91841 Å)
Data collection		
Space group	H32	H32
a (Å)	90.1	90.7
b (Å)	90.1	90.7
c (Å)	334.3	334.9
Resolution range (Å)	20.0–1.75	20.0–1.5
Total reflections	205731	303353
Unique reflections	51564	84393
Completeness (%) (last shell)	96.8 (99.1)	99.6 (99.9)
I/σ (last shell)	13.5 (4.5)	16.6 (3.9)
R(sym) (%) (last shell)	6.6 (40.0)	4.0 (33.5)
B(iso) from Wilson (Å^2)	22.6	16.2
Refinement		
Protein atoms excluding H	2974 (3060)[a]	2982 (3102)
DNA atoms excluding H	363 (726)	405 (810)
Solvent molecules	400 (404)	481 (483)
R-factor (%)	18.2	16.9
R-free (%)	20.5	18.6
Rmsd bond lengths (Å)	0.01	0.009
Rmsd angles (°)	1.22	1.2
Ramachandran core region (%)	89.7	89.8
Ramachandran allowed region (%)	9.7	9.9
Ramachandran additionally allowed region (%)	0.6	0.3
Ramachandran disallowed region (%)	0.0	0.0

[a]Alternative conformations counted separately.

DNA structure

In the co-crystal structure, the DNA fits into a positively charged tunnel formed by the Hpy99I dimer (Figure 4A and Supplementary Figure 1). The phosphodiester backbone forms many salt bridges with positively charged residues of the enzyme (Figure 4B). Together with sequence-specific contacts, these interactions contribute to the observed DNA deformations that follow the 2-fold (pseudo)symmetry and were analyzed with the 3DNA program (26). The central base pair of the recognition sequence is unusually open (~12.5°) and close to its neighbors (rise ~2.9 Å). The largest deformation occurs in the step between the inner C:G and outer G:C pair and expands the minor groove. The roll angle for this step is very large (>20°) and the twist angle very small (~25°). Given the large roll, the small twist is not surprising, since the two parameters are usually anticorrelated (39). The deformation is favored by the specific base sequence. The CG step (in 5′–3′ direction) is known for high average positive roll (5.4° ± 5.2°) and the flexibility to accommodate larger than average roll (40). The local deformations of the DNA in complex with Hpy99I affect the global structure of the DNA. The large positive roll angles compress the major and expand the minor groove. A common measure of minor groove width is the distance between phosphorus atoms of the two DNA strands that are separated by 3 bp (41,42). For the Hpy99I-bound DNA, this distance is 18.7 Å, which is almost 7 Å larger than the value for idealized (3DNA generated) B-DNA. The minor groove expansion increases the distance between scissile phosphates from the expected value of ~17 Å for ideal B-DNA to 25.5 Å for the bent DNA in the Hpy99I–DNA complex (Figure 4C).

Specific sequence recognition

In solution, the two subunits of Hpy99I are equivalent, but in the crystal they are placed in different neighborhoods that are not related by crystallographic symmetry and therefore distinguishable. The DNA oligoduplex used for crystallization is asymmetric at the center of the recognition sequence and contains additional asymmetry in the flanking regions (Figure 5A). Therefore, there are two possible binding modes that both contribute to some extent to the electron density for the central base pair. Sequence readout in the Hpy99I–DNA complex can be

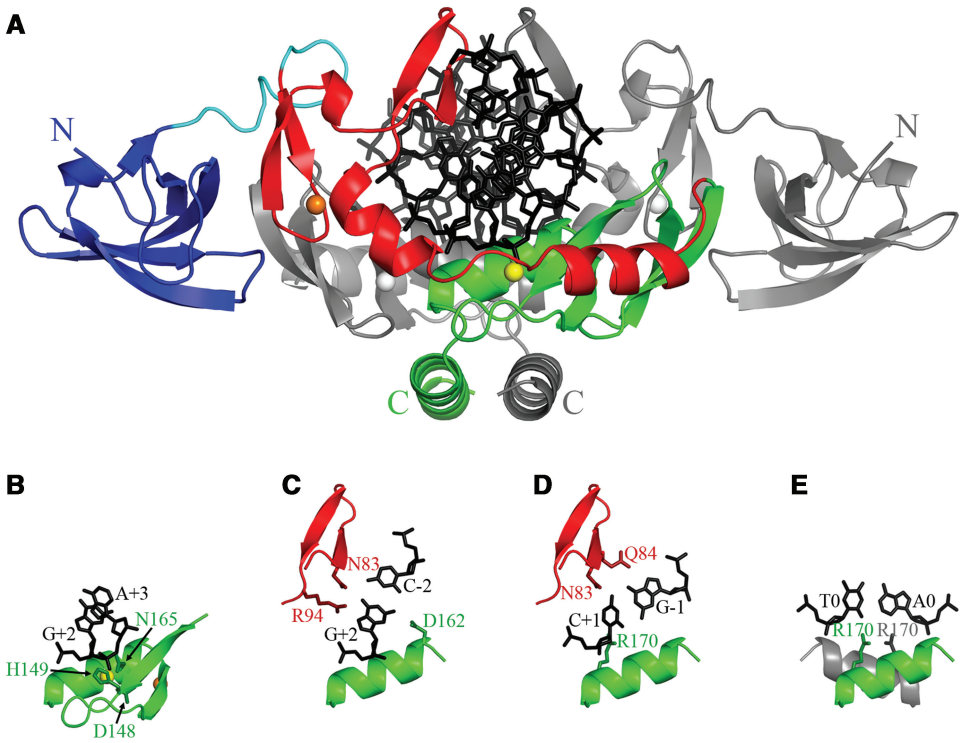

Figure 2. Hpy99I–DNA complex. (**A**) Overall structure of the protein in ribbon and the DNA in all-atom representation. The 2-fold (pseudo)symmetry axis runs vertically. The major and minor groove edges of the central base pair are close to the top and bottom of the figure, respectively. Scissile phosphates and active sites are close to the bottom of the figure. (**B**) Active site and recognition of the (**C**) outer C:G, (**D**) inner G:C and (**E**) central base pairs in the same orientation. One Hpy99I subunit is shown in gray, the other subunit is colored according to region. The N-terminal β-barrel is blue, the linker region cyan, the first β4α2 repeat red, the second β4α2 repeat green, the catalytic metal ion yellow, and the structural Zn^{2+} are orange. One Zn^{2+} ion is hidden behind a helix in *panel* (A), but visible in *panel* (B), where this helix is not shown.

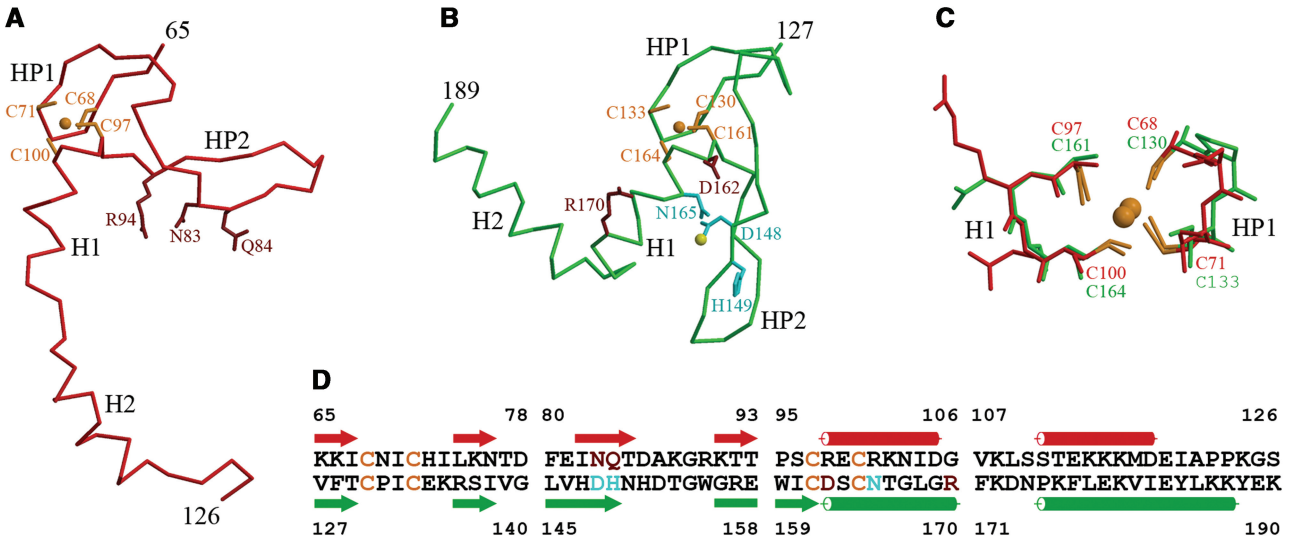

Figure 3. Comparison of the two β4α2 repeats. (**A**) First and (**B**) second β4α2 repeat in Cα-representation in analogous orientation and (**C**) superposition of the Zn^{2+}-binding sites in the two repeats. (**D**) Alignment of selected regions of the two repeats. The first and third blocks of the alignment are based on the spatial correspondence of the Cα atoms and the remaining two blocks were aligned manually. Throughout the figure, the first repeat is red, the second repeat is green, Zn^{2+} ions and coordinating cysteine residues are orange, residues in contact with DNA bases are brown, active-site residues are blue and metal in the active site is yellow. The abbreviations HP1, HP2, H1 and H2 stand for hairpin 1, hairpin 2, helix 1 and helix 2, respectively.

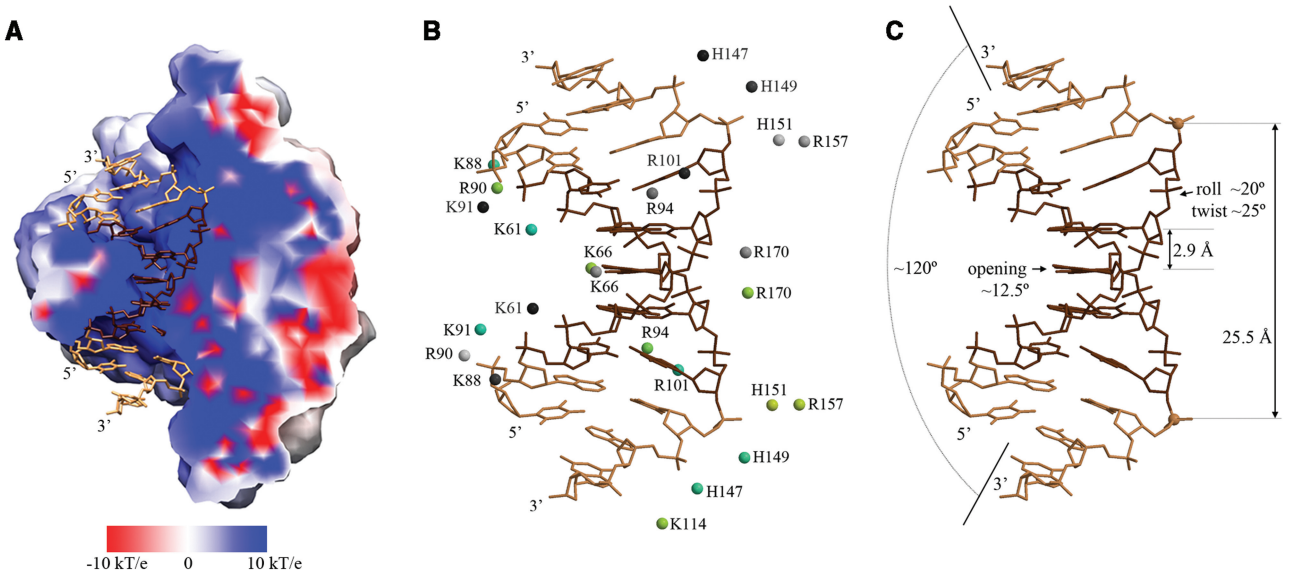

Figure 4. DNA backbone interactions and conformation. The DNA is shown in all-atom representation (light and dark brown for flanking and specifically recognized nucleotides, respectively) with the 2-fold (pseudo)symmetry axis running horizontally. (**A**) Electrostatic potentials for the protein only (with four tetra-thiolate-bound Zn^{2+} ions, two Mg^{2+} ions and half-charged histidine residues) were calculated and mapped to the solvent-accessible protein surface with the DELPHI and CHIMERA programs. (**B**) Salt bridges between DNA backbone and the protein. Basic protein atoms are colored according to their distance to the closest nonbridging phosphate oxygen atom. Colors run from green (2.6 Å) to yellow (5.6 Å) for one Hpy99I subunit and from black (2.7 Å) to light gray (5.9 Å) for the other subunit. (**C**) DNA deformations in the complex. The opening, roll, twist and step values were calculated by the 3DNA program. The value for the total DNA bending angle was estimated by inspection. Scissile phosphates are marked as balls and the distance between them is indicated.

described by simple rules: both protomers contribute to the 'readout' of the central base pair that divides the target DNA into two half-sites: each protomer 'reads' and cleaves only one half-site. The major groove is in contact with the first β4α2 repeat, the minor groove with the second β4α2 repeat (Figure 2B–E).

The outer C:G pair of the recognition sequence is contacted from both sides. On the major groove side, Arg94 and Asn83 donate hydrogen bonds from their side chains to the N7 and O6 atoms of the guanine base, respectively. In addition, the main chain carbonyl atom of Asn83 accepts a hydrogen bond from the exocyclic amino group of the cytosine base of this pair. On the minor groove side, the side chain of Asp162 recognizes the 2-amino group of the guanine base in the central minor groove position (Figure 5B).

The inner G:C pair of the recognition sequence also hydrogen bonds directly to the protein on the major and minor groove sides. On the major groove side Asn83—the residue that interacts also with the outer C:G pair—accepts a hydrogen bond to its Oδ atom from the exocyclic amino group of cytosine. The next residue in the sequence, Gln84, donates a direct hydrogen bond to the O6 atom of guanine. On the minor groove side Arg170 donates a hydrogen bond to the O2 atom of cytosine (Figure 5C).

The central A:T pair of the recognition sequence faces solvent on the major groove side. The Arg170 residues of the two enzyme protomers contact the N3 atom of adenine and the O2 atom of thymine on the minor groove side. These contacts are ideal to recognize either an A:T or a T:A pair because the N3 atom of adenine and the O2 atom of thymine are in almost exactly equivalent positions with respect to the nucleotide glycosidic bonds (43). In contrast, the exocyclic amino group of a G:C or C:G pair in the central minor groove would collide with both Arg170 residues of the Hpy99I enzyme (Figure 5D).

Structural analysis of the active site

The Hpy99I–DNA co-crystals trap the specific complex in a state just 'prior' to cleavage. As in a previous structure of a ββα-Me nuclease with substrate DNA (12), the position of the active site metal ion is taken by a Na^{+} ion from the crystallization buffer (see 'Materials and Methods' section). This assignment is consistent with metal ligand distances, the Na^{+} ion coordination preference (31) and the crystallographic analysis (anomalous signal and temperature factors). The anomalous signal indicates that exogenously added Ca^{2+} does not compete well with the Na^{+} from the crystallization buffer. However, some Mg^{2+} ions appear to be retained and partial DNA cleavage is observed (according to the $f_o - f_c$ difference density) if EDTA is omitted.

The Na^{+} ions that occupy the position of the active site metals in the Hpy99I–DNA crystals are hexa-coordinated by two solvent molecules, two amino acid residues and the proS and (at slightly larger distance) a bridging oxygen atom of the scissile phosphate. The latter oxygen atom would act as a leaving group if doubly charged Mg^{2+} and not singly charged Na^{+} was in the crystal to promote its departure. With the leaving group trapped, a solvent molecule in approximately the right position for in-line attack on the scissile phosphate is kept at a distance that

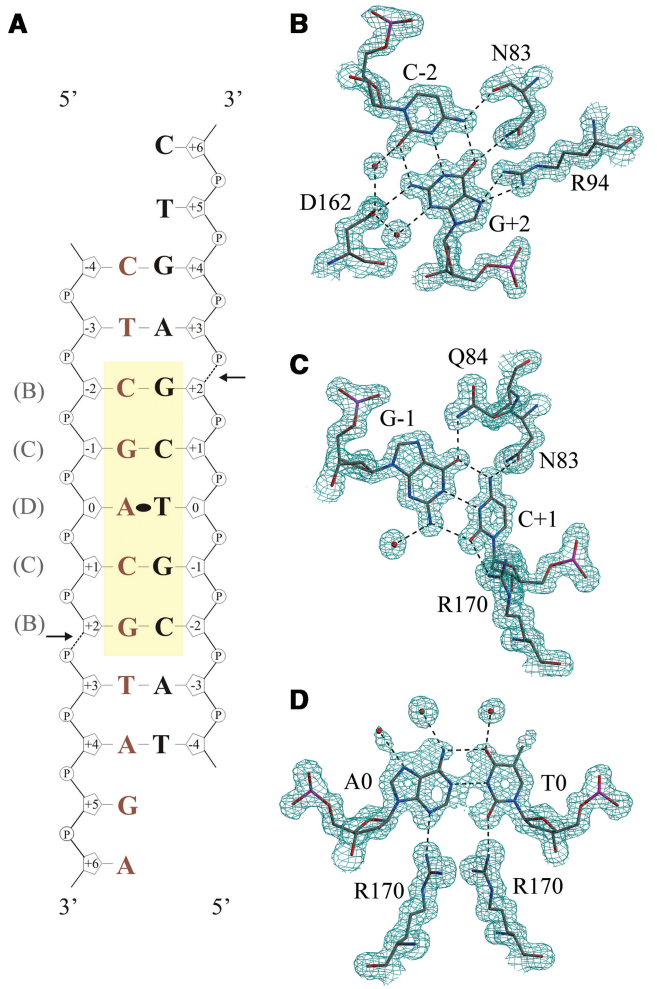

Figure 5. Hydrogen-bonding interactions between Hpy99I and specifically recognized bases. (**A**) Oligoduplex used for crystallization. The specifically recognized bases are highlighted by the yellow background, the 2-fold (pseudo)symmetry axis of the recognition sequence is indicated by the ellipse and the scissile bonds in the phosphodiester backbone by arrows. (**B**–**D**) Hydrogen-bonding interactions between Hpy99I and target DNA. Only one half-site with (**B**) the outer C:G pair, (**C**) the inner G:C pair and (**D**) the central A:T pair is shown. The electron density is the final composite omit map contoured at 1.0σ. It is not clear which of the Arg94 Nη atoms interacts with G+2, therefore both potential hydrogen bonds are shown in (**B**).

exceeds the sum of the van der Waals radii of oxygen and phosphorus. Its proximity to His149 identifies this residue as the likely general base (Figure 6).

Mutagenesis of active site residues

Asp148, His149 and Asn165 were separately mutated to alanines to confirm their role in catalysis. The three point mutants were overproduced like the wt protein and behaved similarly during protein purification. However, all three mutant proteins were inactive at all tested concentrations (Figure 1B). No specific Hpy99I digestion products were observed even when mutated protein amounts and incubation times were 16- and 60-fold higher than necessary for complete DNA cleavage by the wt protein.

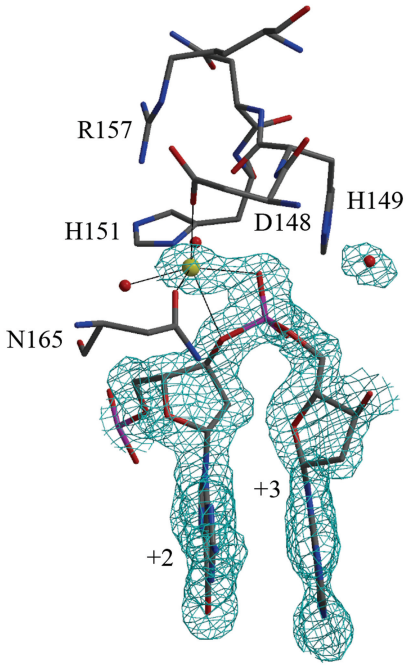

Figure 6. Active site. Residues in or near the active site and the DNA fragment around the scissile phosphoester bond are presented in all-atom representation. The Na^+ ion in the active site is shown in yellow. Gray thin lines indicate its octahedral coordination. Amino-acid residues are color coded according to atom type (black carbon, red oxygen, blue nitrogen, magenta phosphorus). The final composite omit electron density map was contoured at 1.0σ.

DISCUSSION

DNA loading and sliding

DNA digestion by a REase requires the loading of the enzyme on the DNA, scanning of the DNA for target sites and finally cleavage (44). The crystal structures presented in this work show the specific Hpy99I–DNA complex, 'after' the enzyme has locked on its target site, but 'prior' to cleavage (which is prevented by the absence of Mg^{2+} or Mn^{2+} cations). At least at this stage, Hpy99I forms a ring around the DNA (Figure 2 and Supplementary Figure 1). Ring formation facilitates one-dimensional scanning of the DNA, but presents topological problems for Hpy99I loading on circular DNA. Similar digestion rates for circular and linearized plasmid DNA (data not shown) suggest that Hpy99I can 'open up' to dock onto DNA. We also note that surface representations of the specific Hpy99I–DNA complex indicate a very tight tunnel for DNA (Figure 4A and Supplementary Figure 1). For scanning, this complex must 'loosen up', perhaps similar to what has been described for the unspecific and 'hemispecific' DNA complexes of PD-(D/E)XK REases (45,46).

β4α2 repeats

The two β4α2 repeats are not apparent at the sequence level, but readily identified by a structural comparison. As all active site residues are found in the second β4α2 repeat, we looked for their counterparts in the first repeat (Figure 3D). Surprisingly, the active site residues Asp148

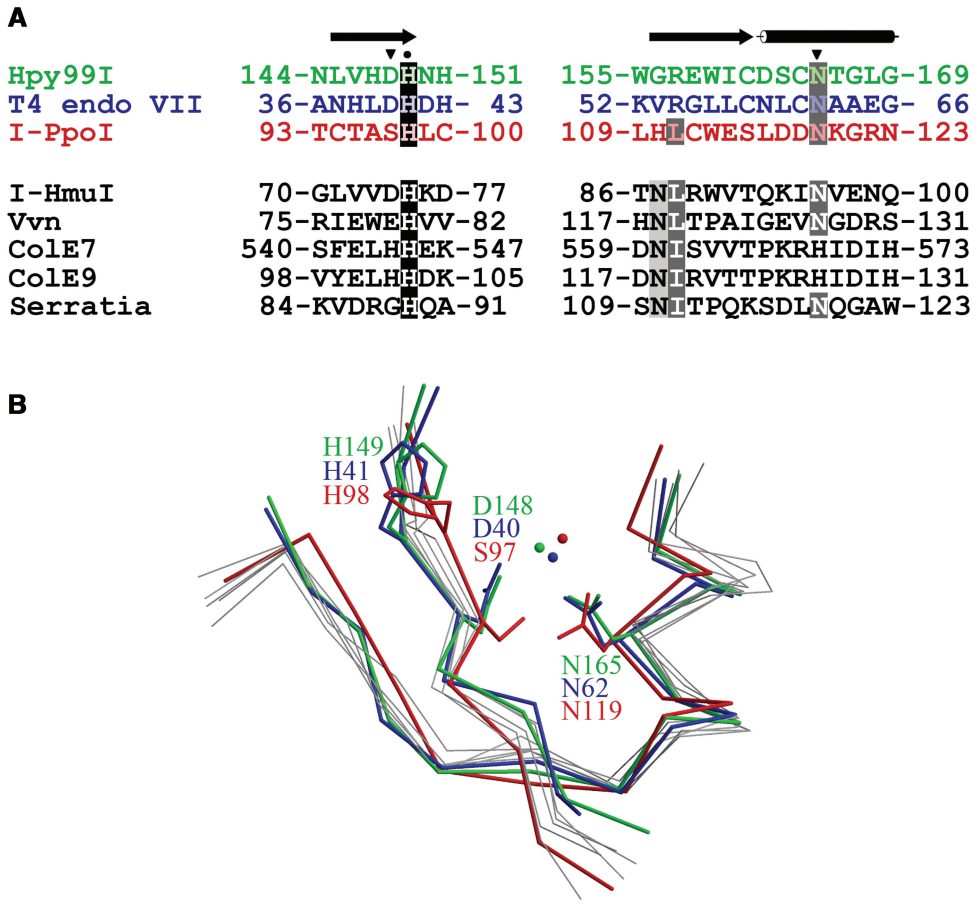

Figure 7. Active site comparison of Hpy99I with other ββα-Me nucleases. (**A**) Structure-based alignment of the active site regions of ββα-Me nucleases. Residue numbering is according to the PDB. The filled circle marks the general base histidine, the inverted filled triangle indicates the metal-chelating residues of Hpy99I. (**B**) Superposition of ββα-Me nuclease Cα-traces. Side chains of active site residues are shown for Hpy99I (Asp148, His149, Asn165), I-PpoI (His98, Asn119) and T4 endonuclease VII (Asp40, His41, Asn62). In I-PpoI, Ser97 replaces a metal ligand and interacts with the metal indirectly via a water molecule. The figure is modeled after Figure 6 of ref. (11) to facilitate comparison.

and His149 align with residues that contact the major groove edge of specifically recognized bases (Asn83 and Gln84) (Figure 3D). Despite this 'coincidence', the two HP2 regions interact nonequivalently with DNA (HP2 of the first repeat inserts deep into the major groove, HP2 of the second repeat only 'touches' the DNA). This shows that the interactions of the Hpy99I repeats with DNA are less stereotypical than the interactions of modular Zn^{2+} finger proteins with consecutive base triplets (47).

The active site

ββα-Me nucleases have been called HNH nucleases to describe nonconsecutive active site residues in some of these enzymes. Hpy99I happens to contain the amino acid sequence 'HDHNH' in positions 147–151. Among these residues, only His149 is referred to by the HNH designation. This residue corresponds to the first and highly conserved 'H' of the HNH motif. It is invariably located at the end of the first β-strand of the ββα-Me region and serves as a general base. Immediately upstream, Hpy99I has the active site metal ligand Asp148. This residue has metal ligand counterparts in some (e.g. Vvn, T4 endonuclease VII, colicin E7), but not other (I-PpoI, Serratia enzyme) ββα-Me nucleases. Due to its poor conservation, this residue is not referred to in the HNH motif designation. The central N of the HNH motif is conserved in I-HmuI, Vvn, the colicins and the Serratia nuclease, but has no counterpart in Hpy99I. The last H of the HNH motif refers to the metal ligand in the middle of the active site α-helix. This residue is a histidine in the colicins, but replaced by asparagine (Asn165) in Hpy99I and also in many other ββα-Me nucleases (Figure 7). Our mutagenesis data for Hpy99I confirm an important functional role for the metal ligands Asp148 and Asn165 and also for the proposed general base residue His149. Inactivation of the counterparts of His149 in I-PpoI, Vvn, colicin E7, colicin E9 and Serratia nuclease also inactivates these enzymes (11,12,48–50).

Unusual stagger of cleavage sites

Hpy99I cleaves target DNA on the 3′-sides of the recognition sequence to fragments that have five nucleotides long 3′-overhangs. This stagger between scissile phosphates is rare and cannot be generated by any other

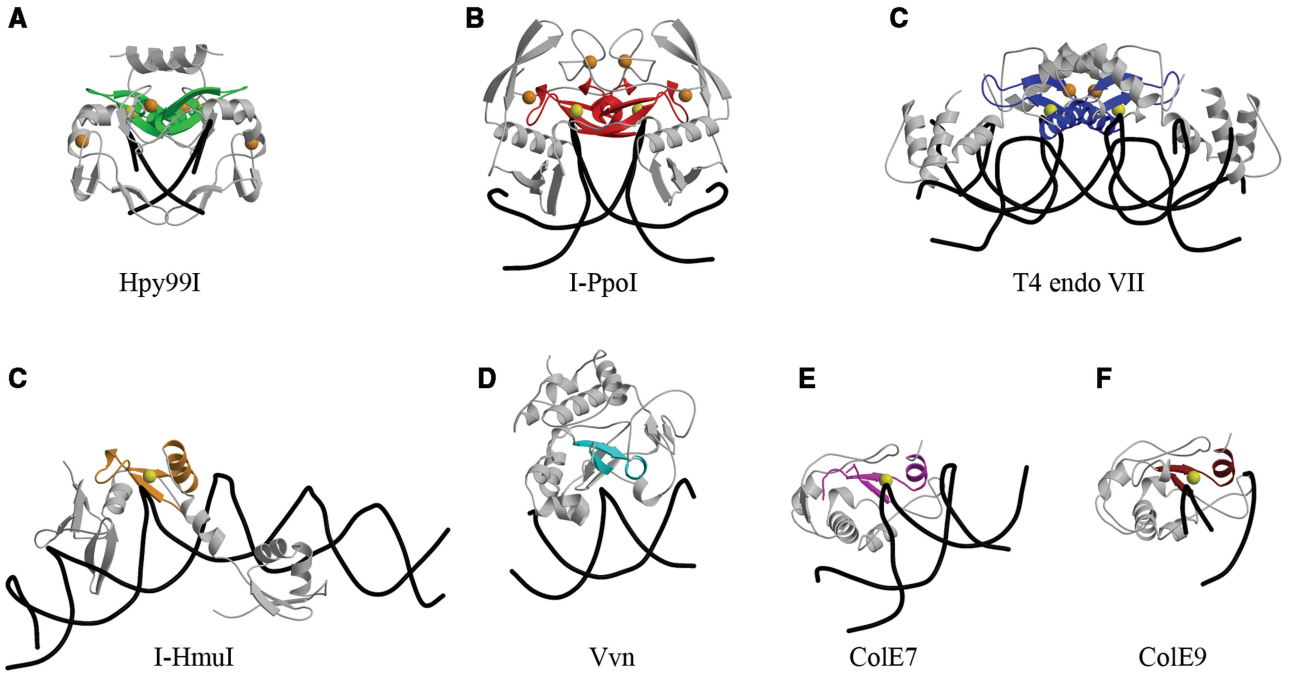

Figure 8. DNA deformation by ββα-Me region minor groove insertion. DNA strands are represented by their smoothed backbones in black; proteins are shown in ribbon representation. The ββα-Me regions are colored. Structural metal ions are represented by orange balls and catalytic metals are shown as yellow balls. The remaining parts of the proteins are shown in gray. In the case of Hpy99I, only the two β4α2 repeats (residues 65–189) are displayed for clarity. Orientations were chosen to demonstrate minor groove insertion of the ββα-Me regions.

REase of known structure. Simple modeling shows that the five-nucleotide stagger between cleavage sites leads to a distance between scissile phosphates of ~17 Å in nondeformed B-DNA, which is typical for many other REase–DNA complexes (51–53). In the Hpy99I–DNA complex, the characteristic α-helices of the ββα-Me regions wedge into the minor groove and expand the distance between scissile phosphates to 25.5 Å. This interaction and the resulting deformation are characteristic for ββα-Me endonuclease DNA complexes (8,11,48,54,55) (Figure 8). LAGLIDADG homing endonucleases that cleave their target sequences with similar stagger deform DNA in the opposite way and reduce the distance between scissile phosphates to 5–8 Å so that the two DNA strands can be cleaved by partially overlapping sites (56).

Distinction of W and S at the center of the recognition sequence

DNA sequences that consist of an odd number of base pairs such as the Hpy99I CGWCG target are at best nearly 2-fold symmetric because Watson–Crick base pairing conflicts with symmetry for the central base pair. Therefore, enzymes that recognize pseudopalindromic (nearly symmetric) target sequences are either unspecific for the bases in the central position of their targets, or they distinguish A:T pairs (W) from G:C pairs (S), irrespective of which base is placed in which strand. Simple considerations show that the distinction between A:T and G:C pairs is difficult on the major groove side but facile on the minor groove side, if the presence or absence of the exocyclic amino group of guanine in the central minor groove position is detected (43,57). This mechanism to distinguish W from S has been suggested for some methyltransferases, albeit without support from a crystal structure (58). The Hpy99I structure in complex with cognate DNA illustrates it and highlights its elegance: swapping of the two DNA strands (exchanging A and T) leaves the two hydrogen bond acceptors in the minor groove in the same position. Symmetric contacts are made by two Arg170 residues of the Hpy99I subunits, but these interactions would be disrupted by the presence of an exocyclic amino group in the central minor groove position (Figure 5D).

A comparison of Hpy99I with I-PpoI and T4 endonuclease VII

Iterative BLAST searches identify a significant sequence similarity between the C-terminal end of Hpy99I and the N-terminal endonuclease domain of T4 endonuclease VII. Here we compare the overall structure of Hpy99I with the structures of I-PpoI as a representative of dimeric ββα-Me homing endonucleases (12,13) and T4 endonuclease VII as a representative of Holliday junction resolvases (15) (Figure 8 and Supplementary Figure 2). The overall structures are widely different but strikingly similar locally. All three enzymes bind structural Zn^{2+} ions (Hpy99I and T4 endonuclease VII even bind a pair in equivalent locations). The ββα-Me active site regions are not only closely related but also similarly dimerize. This could have been predicted for Hpy99I and I-PpoI, because the two enzymes cut DNA with similar stagger. When Hpy99I and T4 endonuclease VII dimers are superimposed, the Hpy99I-bound

DNA strands 'interpolate' between equivalent strands of the Holliday junction (Supplementary Figure 3).

Completing the set of crystal structures for known type II REase superfamilies

Type II REases can be classified into at least five different superfamilies that are known as the PD-(D/E)XK, ββα-Me, GIY-YIG, phospholipase and half-pipe superfamilies. PD-(D/E)XK REases are extremely well-characterized crystallographically: in the REBASE database, there are crystal structures of about 30 different enzymes, often in complex with target DNA (17). To our knowledge, the Hpy99I-DNA co-crystal structure represents the first structure of a ββα-Me type II REase. Structures of BfiI as a representative of the phospholipase group (59) and of PabI as a representative of the half-pipe group (60) have been reported, albeit in both cases without bound DNA. This leaves the group of GIY-YIG REases, with Eco29kI as the best-studied member (61), as the only group of known REases without an experimental structure.

ACCESSION NUMBERS

The sequences described here were deposited in GenBank with the Accession Numbers: 3GOX; 3FC3.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Uwe Mueller for generous allocation of beamtime on MX-beamline-14.1 (BESSY, Berlin) and Dr Georg Zocher for excellent assistance at the beamline.

FUNDING

EMBO/HHMI Young Investigator Award (to M.B); grants from the Polish Ministry of Science and Higher Education (0295/B/PO1/2008/34 to M.B., N N301 028934 to M.S. and PBZ/MEiN/O1/2006/24 to H.C.); and the European Community – Research Infrastructure Action under the FP6 'Structuring the European Research Area' Programme (through the Integrated Infrastructure Initiative Integrating Activity on Synchrotron and Free Electron Laser Science – Contract R II 3-CT-2004-506008). Funding for open access charge: Polish Ministry of Science and Higher Education (0295/B/PO1/2008/34).

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Saravanan,M., Bujnicki,J.M., Cymerman,I.A., Rao,D.N. and Nagaraja,V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.*, **32**, 6129–6135.
3. Kuhlmann,U.C., Moore,G.R., James,R., Kleanthous,C. and Hemmings,A.M. (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.*, **463**, 1–2.
4. Shub,D.A., Goodrich-Blair,H. and Eddy,S.R. (1994) Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.*, **19**, 402–404.
5. Johansen,S., Embley,T.M. and Willassen,N.P. (1993) A family of nuclear homing endonucleases. *Nucleic Acids Res.*, **21**, 4405.
6. Cheng,Y.S., Hsia,K.C., Doudeva,L.G., Chak,K.F. and Yuan,H.S. (2002) The crystal structure of the nuclease domain of colicin E7 suggests a mechanism for binding to double-stranded DNA by the H-N-H endonucleases. *J. Mol. Biol.*, **324**, 227–236.
7. Ko,T.P., Liao,C.C., Ku,W.Y., Chak,K.F. and Yuan,H.S. (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure*, **7**, 91–102.
8. Mate,M.J. and Kleanthous,C. (2004) Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *J. Biol. Chem.*, **279**, 34763–34769.
9. Ghosh,M., Meiss,G., Pingoud,A., London,R.E. and Pedersen,L.C. (2005) Structural insights into the mechanism of nuclease A, a betabeta alpha metal nuclease from Anabaena. *J. Biol. Chem.*, **280**, 27990–27997.
10. Miller,M.D., Cai,J. and Krause,K.L. (1999) The active site of Serratia endonuclease contains a conserved magnesium-water cluster. *J. Mol. Biol.*, **288**, 975–987.
11. Li,C.L., Hor,L.I., Chang,Z.F., Tsai,L.C., Yang,W.Z. and Yuan,H.S. (2003) DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. *EMBO J.*, **22**, 4014–4025.
12. Galburt,E.A., Chevalier,B., Tang,W., Jurica,M.S., Flick,K.E., Monnat,R.J., Jr, and Stoddard,B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat. Struct. Biol.*, **6**, 1096–1099.
13. Flick,K.E., Jurica,M.S., Monnat,R.J. Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96–101.
14. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*, **342**, 43–56.
15. Biertumpfel,C., Yang,W. and Suck,D. (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature*, **449**, 616–620.
16. Pommer,A.J., Wallis,R., Moore,G.R., James,R. and Kleanthous,C. (1998) Enzymological characterization of the nuclease domain from the bacterial toxin colicin E9 from Escherichia coli. *Biochem. J.*, **334 (Pt 2)**, 387–392.
17. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE–enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–270.
18. Tamulaitiene,G., Jakubauskas,A., Urbanke,C., Huber,R., Grazulis,S. and Siksnys,V. (2006) The crystal structure of the rare-cutting restriction enzyme SdaI reveals unexpected domain architecture. *Structure*, **14**, 1389–1400.
19. Newman,M., Lunnen,K., Wilson,G., Greci,J., Schildkraut,I. and Phillips,S.E. (1998) Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.*, **17**, 5466–5476.
20. Vanamee,E.S., Viadiu,H., Kucera,R., Dorner,L., Picone,S., Schildkraut,I. and Aggarwal,A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J.*, **24**, 4198–4208.
21. Kong,H., Lin,L.F., Porter,N., Stickel,S., Byrd,D., Posfai,J. and Roberts,R.J. (2000) Functional analysis of putative restriction-modification system genes in the Helicobacter pylori J99 genome. *Nucleic Acids Res.*, **28**, 3216–3223.
22. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, **26**, 795–800.

23. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1772–1779.
24. Scheldrick,G.M. (2002) Macromolecular phasing with SHELXE. *Zeitschr. Kristallogr./Int. J. Struct. Phys. Chem. Aspects Crystal. Mater.*, **217**, 644–650.
25. Perrakis,A., Morris,R. and Lamzin,V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.*, **6**, 458–463.
26. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
27. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47 (Pt 2)**, 110–119.
28. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
29. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
30. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
31. Harding,M.M. (2002) Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 872–874.
32. Muller,P., Kopke,S. and Sheldrick,G.M. (2003) Is the bond-valence method able to identify metal atoms in protein structures? *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 32–37.
33. Harding,M.M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 401–411.
34. Bahadur,R.P., Chakrabarti,P., Rodier,F. and Janin,J. (2003) Dissecting subunit interfaces in homodimeric proteins. *Proteins*, **53**, 708–719.
35. Eijkelenboom,A.P., Lutzke,R.A., Boelens,R., Plasterk,R.H., Kaptein,R. and Hard,K. (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.*, **2**, 807–810.
36. Simonson,T. and Calimet,N. (2002) Cys(x)His(y)-Zn2+ interactions: thiol vs. thiolate coordination. *Proteins*, **49**, 37–48.
37. Dauter,Z., Wilson,K.S., Sieker,L.C., Meyer,J. and Moulis,J.M. (1997) Atomic resolution (0.94 A) structure of Clostridium acidurici ferredoxin. Detailed geometry of [4Fe-4S] clusters in a protein. *Biochemistry*, **36**, 16065–16073.
38. Collaborative Computational Project Number 4. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.*, **D50**, 760–763.
39. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
40. Packer,M.J., Dauncey,M.P. and Hunter,C.A. (2000) Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.*, **295**, 85–103.
41. Olsen,G.L., Louie,E.A., Drobny,G.P. and Sigurdsson,S.T. (2003) Determination of DNA minor groove width in distamycin-DNA complexes by solid-state NMR. *Nucleic Acids Res.*, **31**, 5084–5089.
42. Laughton,C. and Luisi,B. (1999) The mechanics of minor groove width variation in DNA, and its implications for the accommodation of ligands. *J. Mol. Biol.*, **288**, 953–963.
43. Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
44. Pingoud,A. and Wende,W. (2007) A sliding restriction enzyme pauses. *Structure*, **15**, 391–393.
45. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
46. Townson,S.A., Samuelson,J.C., Bao,Y., Xu,S.Y. and Aggarwal,A.K. (2007) BstYI bound to noncognate DNA reveals a "hemispecific" complex: implications for DNA scanning. *Structure*, **15**, 449–459.
47. Liu,Q., Xia,Z., Zhong,X. and Case,C.C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.*, **277**, 3850–3856.
48. Doudeva,L.G., Huang,H., Hsia,K.C., Shi,Z., Li,C.L., Shen,Y., Cheng,Y.S. and Yuan,H.S. (2006) Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn2+ or inhibitor/Ni2+. *Protein Sci.*, **15**, 269–280.
49. Walker,D.C., Georgiou,T., Pommer,A.J., Walker,D., Moore,G.R., Kleanthous,C. and James,R. (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 Res.*, **30**, 3225–3234.
50. Friedhoff,P., Kolmes,B., Gimadutdinow,O., Wende,W., Krause,K.L. and Pingoud,A. (1996) Analysis of the mechanism of the Serratia nuclease using site-directed mutagenesis. *Nucleic Acids Res.*, **24**, 2632–2639.
51. Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman,S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
52. Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.
53. Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
54. Hsia,K.C., Chak,K.F., Liang,P.H., Cheng,Y.S., Ku,W.Y. and Yuan,H.S. (2004) DNA binding and degradation by the HNH protein ColE7. *Structure*, **12**, 205–214.
55. Hsia,K.C., Li,C.L. and Yuan,H.S. (2005) Structural and functional insight into sugar-nonspecific nucleases in host defense. *Curr. Opin. Struct. Biol.*, **15**, 126–134.
56. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49–95.
57. 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.
58. Kiss,A., Posfai,G., Zsurka,G., Rasko,T. and Venetianer,P. (2001) Role of DNA minor groove interactions in substrate recognition by the M.SinI and M.EcoRII DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.*, **29**, 3188–3194.
59. 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.
60. 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.
61. Ibryashkina,E.M., Zakharova,M.V., Baskunov,V.B., Bogdanova,E.S., Nagornykh,M.O., Den'mukhamedov,M.M., Melnik,B.S., Kolinski,A., Gront,D., Feder,M. *et al.* (2007) Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Struct. Biol.*, **7**, 48.