744–754 Nucleic Acids Research, 2011, Vol. 39, No. 2 Published online 21 September 2010
doi:10.1093/nar/gkq834

DNA intercalation without flipping in the specific ThaI–DNA complex

Malgorzata Firczuk[1,2], Marek Wojciechowski[1,3], Honorata Czapinska[1,3] and Matthias Bochtler[1,3,4,*]

[1]International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland, [2]Department of Immunology, Centre of Biostructure Research, Medical University of Warsaw, Banacha 1A, F building, 02-097 Warsaw, Poland, [3]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01309 Dresden, Germany and [4]Schools of Chemistry and Biosciences, Main Building, Park Place, Cardiff University, Cardiff CF10 3AT, UK

Received June 17, 2010; Revised August 23, 2010; Accepted August 27, 2010

ABSTRACT

The PD-(D/E)XK type II restriction endonuclease ThaI cuts the target sequence CG/CG with blunt ends. Here, we report the 1.3 Å resolution structure of the enzyme in complex with substrate DNA and a sodium or calcium ion taking the place of a catalytic magnesium ion. The structure identifies Glu54, Asp82 and Lys93 as the active site residues. This agrees with earlier bioinformatic predictions and implies that the PD and (D/E)XK motifs in the sequence are incidental. DNA recognition is very unusual: the two Met47 residues of the ThaI dimer intercalate symmetrically into the CG steps of the target sequence. They approach the DNA from the minor groove side and penetrate the base stack entirely. The DNA accommodates the intercalating residues without nucleotide flipping by a doubling of the CG step rise to twice its usual value, which is accompanied by drastic unwinding. Displacement of the Met47 side chains from the base pair midlines toward the downstream CG steps leads to large and compensating tilts of the first and second CG steps. DNA intercalation by ThaI is unlike intercalation by HincII, HinP1I or proteins that bend or repair DNA.

INTRODUCTION

PD-(D/E)XK enzymes form the largest group of type II restriction endonucleases (1). Most members of this group are dimers or assemblies of dimers that recognize fully or nearly (pseudo) palindromic recognition sequences and can cut both DNA strands in a single binding event (2). Some rare monomeric REases require two sequential binding events for this task (3–5). The PD-(D/E)XK restriction endonucleases share a conserved core region that was originally described as a five stranded β-sheet sandwiched with two α-helices (6), but is now known to be more variable (7). Type II PD-(D/E)XK restriction endonucleases typically bind one or two Mg^{2+} ions, but structures with three metal ions in the active site have also been reported (2,8). The two metal ions play distinctly different roles (9). In the pre-reactive complex, metal ion A positions a solvent molecule for in-line nucleophilic attack and helps to deprotonate it (2,10) (Figure 1). Both metal ions (A and B) coordinate the proS oxygen atom of the phosphate at the site of cleavage, presumably to stabilize the (extra) negative charge in the transition state (11). In many structures, metal ion B is in contact with the 3′-oxygen atom of this phosphate. This direct (3,4) or water-mediated (2) interaction facilitates the departure of the leaving group (11). In PD-(D/E)XK restriction endonucleases, metal ion A is conserved and metal B 'optional'. However, metal ion B and not A is analogous to the single metal ion in the ββα-Me (HNH and His-Cys box) (12,13), GIY-YIG (14) and HUH (9) nuclease families.

In 'canonical' PD-(D/E)XK REases, the catalytic residues are anchored within or close to the conserved secondary structure elements of the core region (15). A glutamate residue positioned in the middle of the α-helix is not represented in the PD-(D/E)XK motif. Its acidic side chain holds a solvent molecule that belongs to the coordination sphere of metal A, but is not incorporated into product. In

*To whom correspondence should be addressed. Tel: +0048 22 5970732/+0044 29 208 70625; Fax: +0048 22 5970715; Email: MBochtler@iimcb.gov.pl; BochtlerM@Cardiff.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

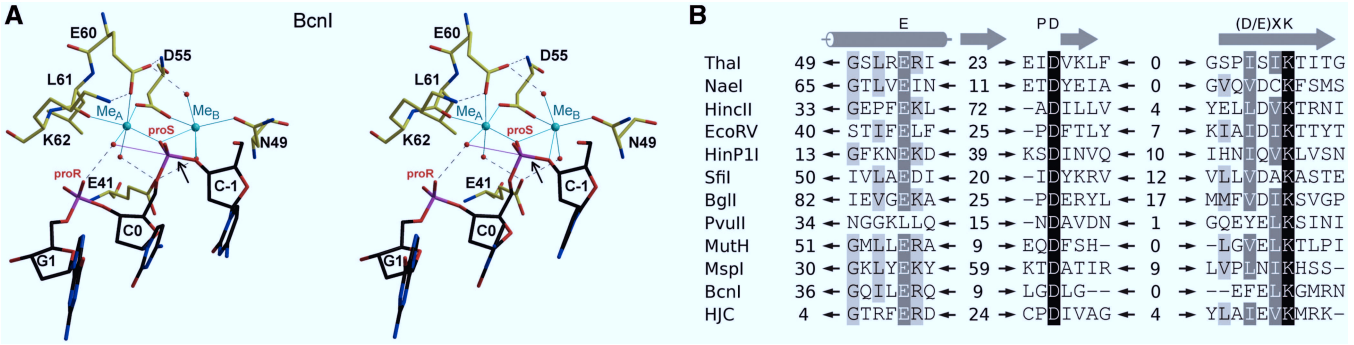

Figure 1. ThaI versus other PD-(D/E)XK restriction endonucleases: (**A**) All-atom stereo representation of a 'typical' PD-(D/E)XK REase active site with two bound metal ions. The coordinates were taken from the ternary complex of BcnI with target DNA and Na^+ ions (PDB code 2ODI). The arrow marks the scissile bond. (**B**) Structure-based alignment of ThaI, other PD-(D/E)XK restriction endonucleases, DNA mismatch repair nickase MutH and a Holliday junction endonuclease of the same catalytic family. The alignment was generated automatically (with slight manual adjustments) using the ThaI coordinates and presents sequences in the order of overall structural similarity to ThaI. The structural alignment confirms the conclusions of an earlier sequence alignment (1).

some structures, the glutamate also binds another solvent molecule linked to metal B (4). An aspartate upstream of the second β-strand coordinates both metal ions directly (2). In some of the originally studied restriction endonucleases, this residue is preceded by a proline, which has given rise to the PD designation in the PD-(D/E)XK acronym. The (D/E)XK motif is located in the middle of the third β-strand. The aspartate or glutamate in this motif coordinates metal ion A. The lysine is thought to either help to stabilize the extra negative charge in the transition state, or alternatively to act as a base with respect to the attacking water molecule (2,10). The 'canonical' arrangement of active site residues in the PD-(D/E)XK nuclease family is not invariable. For example, in PvuII REase (16), the glutamate in the first α-helix is missing. The lysine residue of the (D/E)XK motif can be replaced by a glutamate [e.g. in BamHI (17)] or a glutamine [e.g. in BglII (18) and NotI (19)]. In EcoO109I (20) and the NgoMIV/Ecl18kI (21,22) restriction endonuclease group, the lysine of the (D/E)XK motif is present, but the place of the (D/E) residue is taken by a serine (Figure 1).

Target sequence recognition elements of 'canonical' PD-(D/E)XK restriction endonucleases are more variable than the catalytic core. As REase–DNA interactions are very specific and often extremely tight, major DNA distortions are not uncommon. In many instances, it is thought that the enzymes explore the intrinsic, sequence-dependent DNA deformability and thus add an 'indirect' contribution to 'direct readout' (via hydrogen bonding and van der Waals interactions) (12,23–29). The DNA distortions that are induced or trapped by restriction endonucleases can be very drastic. The Ecl18kI group of REases can flip nucleotides and 'compress' the DNA to fill the void left behind by the flipped bases. The net effect is a shift of register between cleavage sites by 1 bp (22,30,31). EcoRV is well known for its ability to introduce a drastic bend into the DNA (32,33). HinP1I kinks DNA by intercalating a phenylalanine residue (34). HincII deforms target DNA and induces cross-strand purine stacking at the central pyrimidine step (27). In the specific HincII–DNA complex, two glutamine residues of the enzyme dimer are inserted into the DNA base stack from the major groove side (26). So far, HinP1I and HincII are the only restriction endonucleases known to insert amino acids into the base stack. In both cases, intercalation is not within the target sequence, but between the outermost specifically recognized base pair and the flanking sequence, either on one or both sides (26,34). Very recently, dramatic DNA distortions have been reported for the complex of the ββα-Me restriction endonuclease PacI with DNA (13). This enzyme disrupts the base pairs in its target sequence so that two bases on each strand are unpaired, four are engaged in non-canonical A:A and T:T base pairs, and the remaining bases are matched with new Watson–Crick partners.

The restriction endonuclease ThaI occurs naturally in an ATCC 25905 strain of *Thermoplasma acidophilum*, which grows at a high temperature (60°C) and a very acidic pH (0.7–2.0) (35). The enzyme is a blunt-end cutter specific for the CG/CG sequence ('/' marks the cleavage site). Extensive sequence comparisons classify ThaI as a PD-(D/E)XK restriction endonuclease and predict an active site built by residues Glu54, Asp82 and Lys93 (1). The 'PD' Asp82 is not preceded by the usual proline, but by an isoleucine residue like in BamHI and BglII (11). Lys93 is the lysine of the (D/E)XK motif. The place of the acidic residue in the motif is taken by a serine, like in the NgoMIV/Ecl18kI group of REases. The ThaI amino acid sequence contains one PD (residues 144–145) and two (D/E)XK (residues 82–84 and 195–197) fragments, but the match to the 'canonical' motifs is thought to be incidental. In *T. acidophilum*, the host DNA is protected by an accompanying N4-methyltransferase (36). To our knowledge, it has not been determined whether the methyltransferase modifies the first or second cytosine in the sequence. *In vitro* experiments demonstrate that several patterns of physiologically irrelevant C5 methylation protect DNA against ThaI cleavage (37), Therefore, the enzyme could be useful to analyze eukaryotic CG methylation.

We have cloned the ThaI REase and overproduced it in cells that were protected against its activity by

co-expression of the cognate methyltransferase. Here, we report the 1.3 Å resolution crystal structure of ThaI in complex with uncleaved DNA. Our structure confirms the bioinformatics predictions (1) and shows that the enzyme inserts a pair of amino acids into the target DNA in a manner that is very different from what is seen in the HincII and HinP1I–DNA complexes (26,34) and also quite unlike other instances of intercalation in protein–DNA complexes (38).

MATERIALS AND METHODS

General methods

Buffers were titrated and pH values measured at room temperature.

Cloning and expression

The ThaI REase (*thaIR*) and MTase (*thaIM*) genes were amplified by PCR from *Thermoplasma acidophilum* strain ATCC 25905. The *thaIM* coding sequence was amplified with primers to introduce a Shine–Dalgarno sequence 8-nt upstream of the ATG start codon and cloned via HindIII and SalI under the control of the tetracycline promoter of the pACYC184 (Cm^r) vector. The *thaIR* coding sequence was cloned with EcoRI and XhoI into the pET15bmod (Ap^r) vector, which led to a construct that coded for ThaI REase with N-terminal, vector-derived histidine tag (sequence MGHHHHHHEF). The starter methionine of the ThaI open reading frame was missing in the sequence.

Expression and purification

The *Escherichia coli* strain ER2566 (NEB) was sequentially transformed with pACYC184 (Cm^r) plasmid bearing the *thaIM* gene and then with pET15bmod (Ap^r) plasmid containing the *thaIR* gene. Cells were grown in LB medium with appropriate antibiotics at 37°C to OD_{600} 1.0, induced with 0.5 mM IPTG, shifted to 29°C and harvested 16 h later. ThaI was purified by affinity chromatography on Ni-NTA Agarose column (Qiagen). The protein was eluted using an imidazole gradient in buffer A (20 mM Tris/HCl pH 7.6, 200 mM NaCl, 10 mM $MgCl_2$). Fractions containing ThaI were collected and concentrated by ultrafiltration. The enzyme was further purified by size-exclusion chromatography on Superdex75 10/300GL column (GE Healthcare) equilibrated with buffer B (10 mM Tris/HCl pH 7.6, 50 mM KCl, 10 mM $CaCl_2$, 1 mM DTT). ThaI containing fractions were pooled and used for protein–DNA complex preparation.

11mer single stranded DNA fragments of sequences 5′-GGTA*CGCG*ATG-3′ and 5′-CCAT*CGCG*TAC-3′ were purchased from oligo.pl. The oligonucleotides were dissolved in 10 mM Tris/HCl pH 7.5 and annealed overnight by heating to 95°C and slow cooling to 4°C. ThaI was mixed in 1:1.5 or 1:1.2 molar ratio with DNA (protein dimer:dsDNA), incubated for 10 min at room temperature, concentrated and subjected to size-exclusion chromatography on a Superdex75 10/300GL column (GE Healthcare), equilibrated with buffer B. The fractions containing the enzyme–DNA complex were collected and concentrated by ultrafiltration.

ThaI variant activity assays

Fifty milliliter of terrific broth supplemented with 1 mM $MgCl_2$; 0.2% lactose; 0.05% glucose; 100 µg/ml ampicillin; 34 µg/ml chloramphenicol was inoculated with a single colony of *E. coli* ER2566 cells harboring pACYC184_*thaIM* and pET15bmod_*thaIR* (wild-type or mutant) plasmids. Cells were grown for 24 h at 37°C. Later, 1 ml aliquots were collected and centrifuged (30 s, 16 000*g*). Cell pellets were suspended in 0.8 ml of ThaI activity buffer C (20 mM Tris/HCl pH 7.5, 50 mM KCl, 10 mM $MgCl_2$) and disrupted by sonication (2×10 s, 90 W). Lysates were incubated for 30 min at 60°C and clarified by centrifugation (15 min, 16 000*g*). Concentration of ThaI variants was assessed densitometrically from Coomassie-stained SDS–PAGE using Fluor-S Multimager and QuantityOne software (Bio-Rad). Lysates were diluted to obtain an ~5 ng/µl enzyme concentration. In order to determine the activity of each ThaI variant, 600 ng of pUC19 plasmid was digested with 5 ng of the enzyme for 1 h at 60°C in buffer C. Reactions were analyzed by electrophoresis using 1% agarose gels in TBE buffer (90 mM Tris/HCl, 90 mM boric acid, 2 mM EDTA).

Crystallization

A stoichiometric mixture of ThaI with DNA (8 mg/ml of protein dimer and an equimolar amount of dsDNA) was used for crystallization trials. Initial sitting drop experiments (1 : 1 protein–DNA solution and reservoir buffer) were set up with the help of a Cartesian robot in 96-well Greiner plates. Starting crystallization conditions were optimized on a larger scale (4-µl sitting drops) using CRYSCHEM plates (Hampton Research). The best diffraction was obtained for a crystal grown at 21°C from mother liquor that initially contained a 1:1 mixture of protein and reservoir buffer (0.1 M sodium acetate pH 4.5, 200 mM $(NH_4)_2SO_4$, 30% PEG400 and 1.5% 1,2,3-heptantriol). Crystals were cryocooled directly from the mother liquor.

Structure determination

Diffraction data to 1.3 Å resolution were recorded at beamline BW6 at DESY (Hamburg). Crystals were assigned to space group P2(1) with cell constants 39.3 Å, 72.9 Å, 88.8 Å, $\beta = 102°$. The presence of DNA in the crystals was predicted with the program DIBER (39), which also suggested that the DNA was oriented along the direction of the $a = 39.3$ Å unit cell edge, in agreement with the expected length of the 11-mer straight B-DNA. The ThaI–DNA co-crystal structure was solved by the SIRAS method. A conventional derivative was obtained by soaking of a native crystal for 4 h with ~1.5 mM uranyl acetate in the reservoir buffer. Data were collected on a rotating anode X-ray generator (Rigaku RU300, no microfocus) with long (10 min per 0.5°) exposures. Both native and derivative data were processed with iMOSFLM (40) and scaled with SCALA (41). The success of the

soaking procedure was confirmed by an inspection of the isomorphous and anomalous difference Fourier maps. The data were interpreted by the program SHARP (42) in terms of a four uranyl ion substructure. After solvent flattening, phases were sufficiently accurate for ARP/wARP (43). The program could build the almost complete ThaI dimer and assigned the amino acid sequence automatically. However, visual inspection showed that it had not interpreted two linker regions in the electron density and therefore missed the cross-over of the C-terminal ThaI ends. An oligoduplex with the correct sequence in standard B-DNA geometry was generated with the program 3DNA (44), and manually adjusted to fit the highly distorted DNA in the structure. The model building added an unexpected twist to the match between the unit cell repeat in the a-direction and the length of an 11-mer oligoduplex. In the crystal, DNA 'rods' were indeed formed. However, protein methionine intercalation shifted the register by 1 bp. This was compatible with the unit cell repeat because the DNA overhangs did not form the expected Watson–Crick base pairs. The programs COOT (45), REFMAC (46) and CNS (47) were used for refinement. The data collection statistics and quality parameters of the deposited structure (PDB code 3NDH) are summarized in Supplementary Table S1.

RESULTS

Overexpression, purification and activity assays

N-terminally histidine tagged ThaI restriction endonuclease was overexpressed and purified by affinity and size-exclusion chromatography. Gel filtration experiments suggested that the enzyme forms a dimer like the majority of other PD-(D/E)XK restriction endonucleases (Supplementary Figure S1). ThaI activity was assayed with plasmid pUC19 as a substrate at 60°C (the typical growth temperature of the source organism), and at 37°C (Supplementary Figure S2A). At neutral pH, ThaI binds DNA in the presence of Mg^{2+} or Ca^{2+}, but not in the absence of divalent metal cations (Supplementary Figure S1). At this pH, ThaI is active in the presence of Mg^{2+}, but not Ca^{2+} ions. At acidic pH, ThaI binds DNA also in the absence of divalent metal cations, but the cleavage activity is greatly diminished, even in the Mg^{2+} containing buffer (Supplementary Figure S2B).

Crystallization and structure determination

ThaI–DNA crystals were obtained at acidic pH (4.5–4.8). For the complex formation, the 11-mer oligonucleotides (5′-GGTA*CGCG*ATG-3′ and 5′-CCAT*CGCG*TAC-3′) were annealed to form double stranded DNA with one nucleotide 5′-overhangs. Co-crystals of this oligoduplex and ThaI belonged to space group P2(1) and diffracted to 1.3 Å resolution at BW6 beamline (DESY, Hamburg). They contained one enzyme dimer in complex with a single dsDNA molecule in the asymmetric unit. Attempts to solve the ThaI–DNA structure by molecular replacement were unsuccessful. Therefore, experimental phases were eventually obtained with a uranyl acetate derivative.

Overall structure

An overview of the ThaI–DNA complex is presented in Figure 2. Several features of the complex are immediately recognizable. The bulk of ThaI is located on the minor groove side of the specifically recognized bases. This is also the region of the most extensive dimerization contacts, many of which are mediated by the domain-swapped C-terminal helices. Some contacts are also made on the major groove side, but these are less extensive. The ThaI dimer completely embraces the DNA in the specific complex, which is not compatible

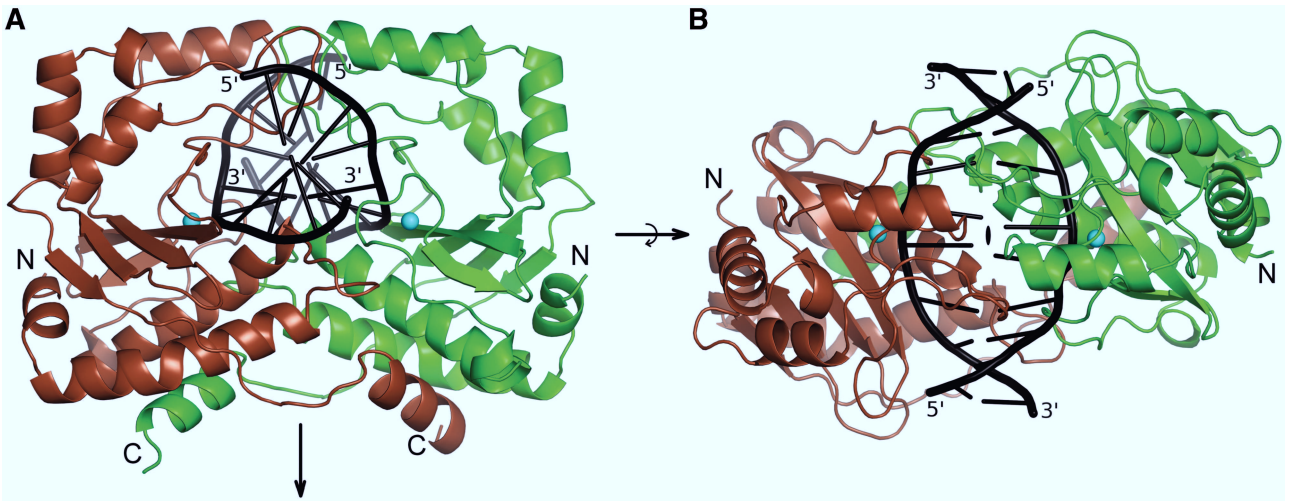

Figure 2. Overview of the ThaI–DNA complex: the two subunits of ThaI are shown in ribbon representation in green and brown color, respectively. The DNA is represented by its smoothed backbone with sticks for the bases. (**A**) View along the DNA. The major grooves of the central two base pairs point upwards. The 2-fold symmetry axis (arrow) of the recognition sequence and of the ThaI dimer runs vertically. In the vicinity of the central base pairs, the DNA backbone is oriented toward the viewer on the left and away from the viewer on the right side (in the 5′–3′ direction). (**B**) View toward the major groove of the central DNA base pairs. The two views differ by a 90° rotation about the horizontal axis.

with substrate loading. As ThaI can cleave circular DNA, the enzyme must open up when it first encounters the substrate. This is not unusual and has been described for many other restriction endonucleases (48). As expected, the core regions of ThaI and other blunt-end cutters are arranged similarly with respect to the DNA. The correspondence is particularly striking for NaeI (49,50). This blunt-end cutting restriction endonuclease also scores highest in a DALI search (51) of the Protein Data Bank (PDB) that measures only protein similarity, but does not take the dimerization and DNA binding mode into account (DALI *Z*-score 9.6, Figure 1B and Supplementary Figure S3).

Active site

The structure of the ThaI–DNA complex around the scissile phosphoester bond confirms the predicted active site residues (1) (Figure 3A and B). They are found in the expected secondary structure contexts (15) and are appropriately located with respect to the substrate DNA to play the anticipated catalytic roles (2). Glu54 is positioned in the middle of the conserved α-helix. It binds a solvent molecule that lies in the coordination sphere of metal ion A, but is not the nucleophile. Asp82 is the 'PD' aspartate, which also interacts with metal A and would most likely act as the bridging ligand between the two metal ions if both were present in the active site. Lys93 is the '(D/E)XK' lysine that anchors the nucleophilic solvent molecule. However, serine instead of aspartate or glutamate is present in the sequence two residues upstream of the catalytic lysine. The side chain of Ser91 hydrogen bonds to a water molecule in the coordination sphere of metal A and thus it seems to act as an 'indirect' ligand of this metal ion (Figure 3). Thus, the serine could functionally replace the acidic residue. Alternatively, the low pH in the crystals could have induced a catalytically non-relevant conformation and might have displaced an acidic residue recruited from elsewhere in the sequence. Glu80 is directed toward solvent and might rearrange at neutral pH, but mutagenesis data (see below) argue against this possibility. In the ThaI–DNA co-crystal, the Nε atom of Trp107 donates a hydrogen bond to the proR non-bridging oxygen atom of the cleavage site phosphate. At least in an associative mechanism, this interaction could stabilize the (extra) negative charge in the transition state and thus play a role in catalysis. However, the hydrogen bond is unusual because the proR oxygen atom is in contact with solvent in many PD-(D/E)XK restriction endonuclease–DNA structures, and comes close to the active site lysine in some (22,26).

Metal ions and waters in the active site

ThaI–DNA crystals were obtained at acidic pH where the enzyme activity is only residual. Moreover, we used Ca^{2+} instead of Mg^{2+} in the crystallization buffer so that the DNA cleavage is further suppressed. As expected in these conditions, the DNA in the structure is not hydrolyzed, and we observe robust electron density for the scissile phosphoester bond. A water molecule is positioned in-line with the scissile bond, but too far away from the phosphorus atom for productive nucleophilic attack (3.3 Å, Figure 3). This solvent molecule is part of the almost perfect octahedral coordination sphere of metal ion A. The robust electron density for this metal ion is likely due to a bound Na^{+} or Ca^{2+} ion, which are both present in the buffer and have previously been found in restriction and homing endonuclease active sites (4,12,20,52). There is no electron density for metal ion B in the ThaI–DNA complex. The Asp82 carboxyl oxygen atom that would normally coordinate this metal ion comes

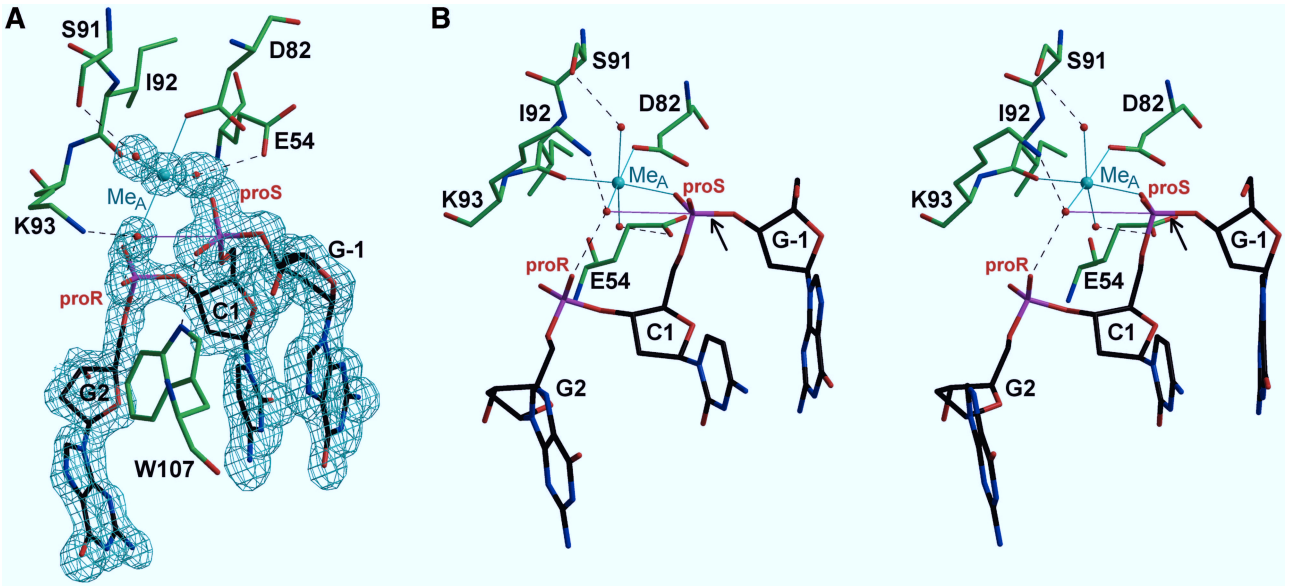

Figure 3. ThaI active site: (**A**) Mono representation of the ThaI active site showing the final composite omit electron density contoured at 1.5σ for the DNA, the metal ion and its coordinating water molecules. (**B**) Stereo representation of the active site highlighting the catalytic residues. In both panels, the interactions of the metal ion A (cyan) are shown as thin continuous lines and selected hydrogen bonds as dashed lines. The arrow in panel B marks the scissile bond. The orientation of panel B is consistent with Figure 1A.

instead very close to the proS oxygen atom of the phosphate at the site of cleavage. Such proximity would normally be unfavorable, but is possible in this case because the aspartic acid and the phosphate can share a proton in the acidic pH of the crystallization drop.

DNA recognition

Each ThaI subunit cleaves one DNA strand and contacts the bases of this strand on the 3′-side of the scissile phosphoester bond and their Watson–Crick partners. Direct hydrogen bonding between the enzyme and the DNA bases on the minor groove side is limited to a single interaction from the N2 atom of the inner guanine to the Oε atom of Glu48. All other direct interactions between the DNA bases and ThaI endonuclease occur on the major groove side. The inner guanine accepts two hydrogen bonds to its N7 and O6 atoms from the guanidino group of Arg171. The outer guanine interacts with the ε-amino group of Lys104. Its complementary cytosine donates a hydrogen bond to the main chain oxygen atom of Asn169. The ThaI structure strongly suggests that van der Waals contacts and edge-face interactions play a role in sequence selectivity. Contacts to Trp107 should limit the tolerance to methylation of the inner cytosine. N4 methylation might be acceptable (3.3 Å distance to Trp107 Cβ), but C5 methylation would lead to a clash with the indole ring. Clashes with the carbonyl oxygen atom of Asn169 should result from either N4 or C5 methylation of the outer cytosine, unless steric conflicts can be avoided by an adaptive fit (Figure 4).

Methionine intercalation

The most remarkable feature of the ThaI–DNA co-crystal structure is the 'deep' intercalation of the symmetric Met47 residues and the resulting unstacking and unwinding of the DNA (Figure 5A). Each methionine approaches the DNA from the minor groove side and inserts between the C and G bases of the downstream CG step of the recognition sequence. Although van der Waals contacts between the two bases are entirely replaced by hydrophobic interactions with the methionine side chain, their Watson–Crick hydrogen bonding is not disrupted. Analysis with the 3DNA software (44) shows that the intercalation increases the rise between base pairs to about 7 Å or approximately twice its usual value (Figure 5B). Phosphorus–phosphorus (P_n–P_{n+1}) distances in the DNA backbone are only mildly altered (values range from 5.6 to 7.0 Å). Instead, the extra height of the two CG steps comes at the expense of the twist, which is reduced from its usual value of about 36° (360°/10) to between 10 and 15°. A view toward the major groove shows that the inner base pairs of the recognition sequence are strongly tilted (Figure 5). According to the 3DNA software (44), the first CG step has a negative tilt of about −12°, which results in the oblique orientation of the following base pairs. The central GC step is characterized by a tilt close to 0°, reflecting the nearly parallel arrangement of the middle bases. Finally, the second CG step has a positive tilt of about 15° which restores the standard orientation of the downstream base pairs. A side view of the DNA indicates a bend at the center of the recognition sequence which is primarily due to the positive, ∼12° roll of the central GC step into the major groove (Table 1). The 3DNA program also indicates that the propeller twist is positive for the specifically recognized sequence, and (as expected for the standard B-DNA) negative for most of the flanking base pairs.

Biochemical verification of the structural observations

As the ThaI–DNA co-crystal was obtained in conditions that do not support DNA cleavage, we decided to verify the structure-based conclusions biochemically. Key residues of the enzyme were individually substituted with alanine. The mutant proteins were tested for activity against pUC19 plasmid under conditions that are optimal for wild-type ThaI (Figure 6). A loading control confirmed that equal amounts of the enzyme were used in all digestions (Figure 6A). Further controls confirmed that the plasmid was uncleaved in the absence of a cell lysate (Figure 6B, lane 2) and in the presence of a mock lysate of cells lacking ThaI expression (Figure 6B, lane 3).

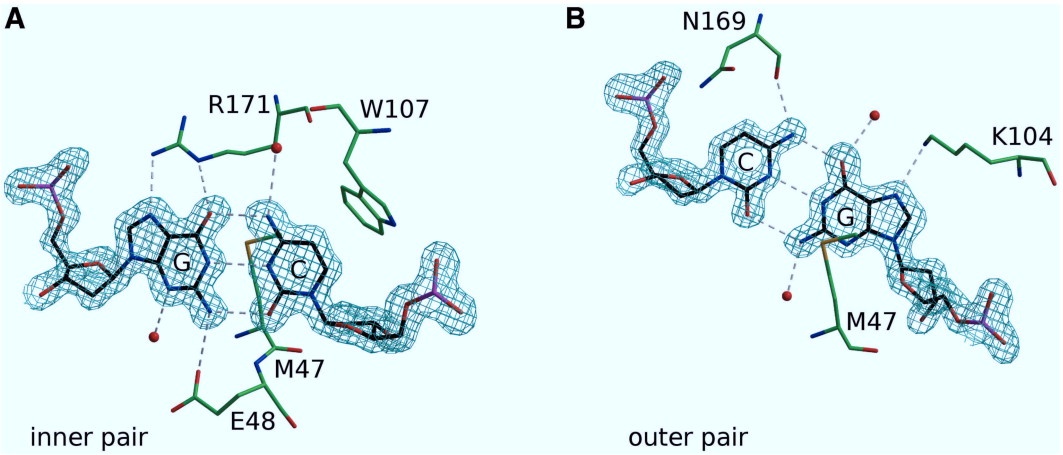

Figure 4. DNA sequence recognition by ThaI: hydrogen bonding interactions between ThaI and the specifically recognized bases of the inner (**A**) and outer (**B**) pairs are shown as dashed lines. The intercalating Met47 and wedging Trp107 are also included in the figure. The electron density is a final composite omit map contoured at 1.5σ.

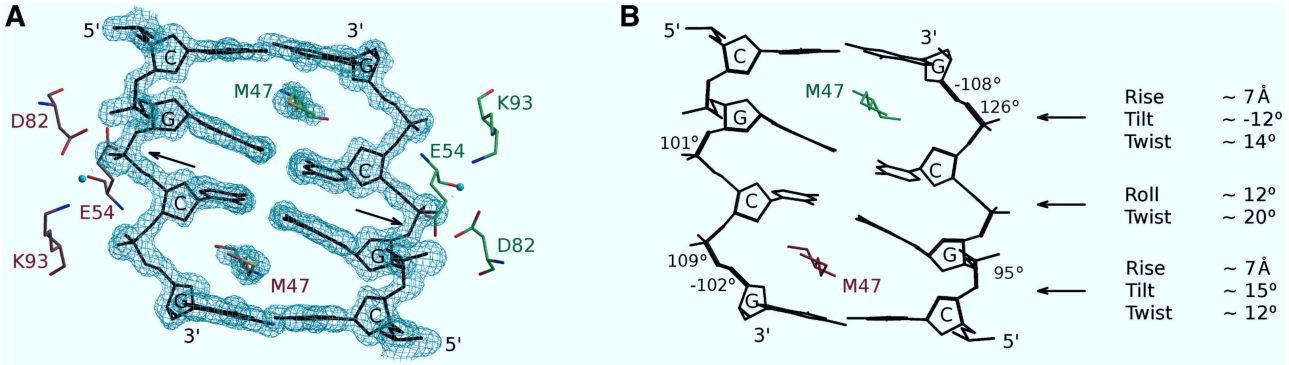

Figure 5. DNA distortions in the ThaI–DNA complex: (**A**) the specifically recognized bases and key amino acids are shown in all-atom representation. The metal ions A in the active sites are represented by cyan balls. Final composite omit electron density map for the DNA and intercalating methionines was contoured at 1.5σ. (**B**) All-atom representation of the DNA and intercalating methionines. The thickness of the DNA backbone corresponds to the deviations of the torsion angles from their values in idealized regular B-DNA. The differences of >90° are stated explicitly. The most characteristic geometry distortion parameters for the DNA base pair steps are indicated.

Table 1. DNA distortion in complex with ThaI restriction endonuclease: all parameters were calculated with the 3DNA software (44). The steps of the ThaI recognition sequence are highlighted in bold

Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
GT/AC	0	0.2	3.3	5.2	−0.1	34.9
TA/TA	−0.4	−0.1	3.4	−5.8	−2.1	38.8
AC/GT	0.6	−0.5	3.0	6.3	4.3	31.1
CG/CG	**−2.3**	**−0.9**	**7.1**	**−12.0**	**−1.5**	**14.3**
GC/GC	**0**	**−1.5**	**2.8**	**−2.1**	**12.0**	**20.1**
CG/CG	**2.1**	**−1.1**	**6.9**	**15.2**	**−2.5**	**12.0**
GA/TC	0	−0.4	3.0	−6.8	2.1	32.6
AT/AT	0.7	−1.0	3.1	−1.4	−0.2	30.7
TG/CA	−1.6	−1.2	3.9	−8.3	5.5	35.0

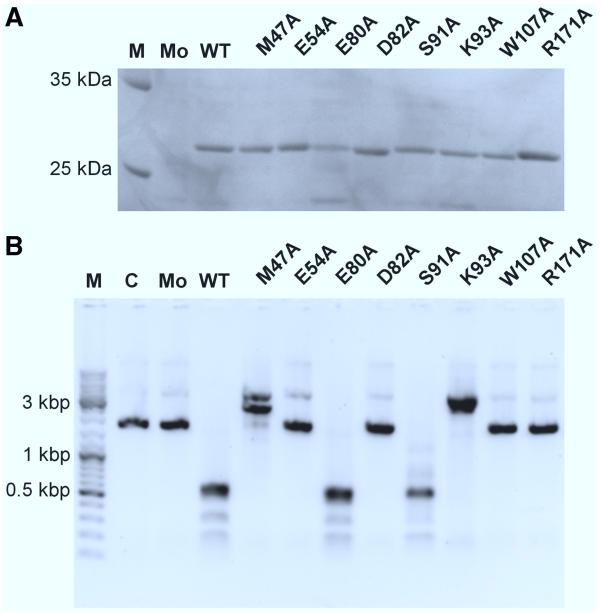

Figure 6. Enzymatic activity of the ThaI variants: (**A**) Coomassie stained SDS–PAGE gel proving that equal amounts (5 ng) of ThaI and its variants were used in the assay. (**B**) Electrophoresis of 600 ng pUC19 plasmid alone (control, 'C'), incubated with *E. coli* lysate of cells not overexpressing ThaI (mock, 'Mo'), or overexpressing ThaI or its variants (lanes 4–12). 'M' stands for the protein marker in panel A and for the DNA marker in panel B.

In contrast, robust substrate digestion was observed with the wild-type enzyme (Figure 6B, lane 4). Mutation of the intercalating Met47 to alanine reduced the activity, but some linearization of the plasmid was still observed (Figure 6B, lane 5). Mutation of the Glu54 and Asp82 catalytic residues inactivated the enzyme (Figure 6B, lanes 6 and 8). Mutation of the active site Lys93 had a less drastic effect (Figure 6B, lane 10). Glu80 was irrelevant for activity, and surprisingly the '(D/E)' serine residue (Ser91) could also be substituted with alanine without much effect (Figure 6B, lanes 7 and 9). In contrast, the replacement of the active site Trp107 or of the sequence reading Arg171 with alanine rendered the enzyme inactive (Figure 6B, lanes 11 and 12).

The partial activity after the substitution of Met47 with Ala was unexpected. Therefore, we randomized the bases of the methionine coding triplet. Among the 29 sequenced clones, 26 had only the intended mutations and included all but five possible Met replacements (the Asp, Ile, Lys, His and Gln variants were missing). A loading control confirmed that equal amounts of ThaI and its variants were used in the assay (Figure 7A). Digestion experiments indicated that all variants were much less active than the wild-type enzyme (Figure 7B). Only the Ser47 variant had significant nucleolytic activity. The previously tested Ala47 and the new Thr47 and Leu47 variants were less active than the Ser47 variant, but had somewhat higher residual activity than the other variants. ThaI with a cysteine residue at position 47 did not cleave DNA in the absence of any reducing agent, but had comparable activity to the Ala47, Thr47 and Leu47 variants when DTT was added to the digestion mix (data not shown).

DISCUSSION

Unusual active site

The ThaI active site is uncommon in at least two respects. First, the acidic residue in the (D/E)XK motif is replaced

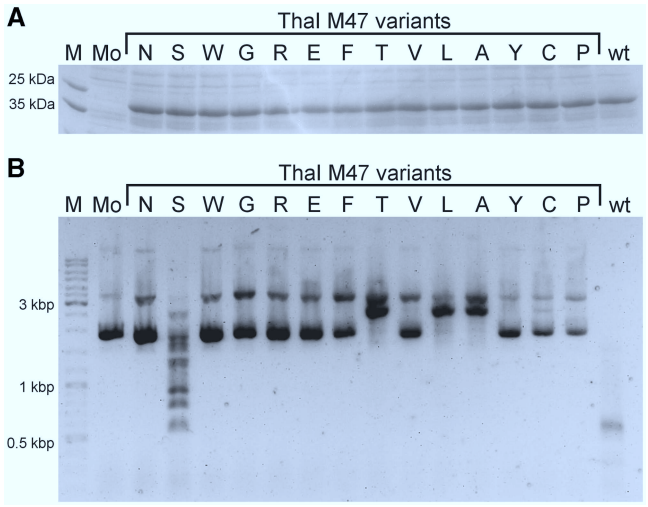

Figure 7. Enzymatic activity of the ThaI and its variants with substitutions of the DNA intercalating Met47: (**A**) Coomassie stained SDS–PAGE gel proving that equal amounts (5 ng) of ThaI and its variants were used in the assay. (**B**) Electrophoresis of 600 ng pUC19 plasmid incubated with *E. coli* lysate of cells not overexpressing ThaI (mock, 'Mo'), or overexpressing ThaI or its variants (lanes 3–17). 'M' stands for the protein marker in panel A and for the DNA marker in panel B.

by a serine. The absence of a 'canonical' active site residue has occasionally been found to be compatible with activity. The best example is PvuII, which lacks the acidic residue in the conserved α-helix (16). More often, the absence of a catalytic residue from its standard place indicates that its structural equivalent is located somewhere else in sequence (30,53). The replacement of (D/E)XK with SXK has precedents in EcoO109I and in the Ecl18kI group of REases (20,22,30). For Cfr10I in the latter group, it has been shown that the mutation of the serine residue to alanine reduces, but does not abolish activity (54). This finding is in line with our experimental result that Ser91 in ThaI is not essential for catalysis.

Second, the tryptophan residue is present in the vicinity of the active site and has a possible catalytic role. However, we cannot strictly exclude that the hydrogen bonding contact with the scissile bond phosphate is an artifact, because the pH in the crystals is too low for activity. In this case, the lack of activity of the W107A variant at permissive pH could be attributed to a structural perturbation accompanying the loss of the indole ring. Likewise, the lack of metal B in our ThaI–DNA co-crystal structure could result from the low pH of the crystallization buffer. We note that the question of one versus two or even three metal ions in PD-(D/E)XK restriction endonucleases has never been fully resolved. Like the quest for the precise role of the individual metal ions, it might not have a universally applicable answer (2,8,10,55–57).

ThaI versus HincII and HinP1I intercalation

HincII and HinP1I are the only restriction endonucleases that have previously been shown to insert amino acids into the base stack (26,34). ThaI and HincII are both blunt-end cutters and share a dimer arrangement that reflects the 2-fold symmetry of their target sequences (Supplementary Figure S3). Intercalation also follows the symmetry and therefore occurs in duplicate in both enzyme–DNA complexes. Nevertheless, the details are very different. ThaI inserts methionines, HincII glutamines. ThaI introduces its amino acids from the minor groove side within the recognition sequence, HincII from the major groove side between the outermost specifically recognized and the flanking base pairs. ThaI intercalation increases the rise between base pairs and unwinds DNA (reduces twist), but induces only a slight roll into the major groove. HincII increases the rise too, but at much less cost to the twist. There is no counterpart in the ThaI–DNA complex for the characteristic cross-strand stacking at the center of the HincII bound DNA. Finally, the inserted residues in ThaI (Met47) and HincII (Gln138) are anchored in non-equivalent positions of the fold. Therefore, the two intercalation cases have almost certainly evolved independently.

The comparison of ThaI with HinP1I is less straightforward. HinP1I is not a blunt-end cutter and has a very different architecture. The enzyme is a dimer, but the dimer axis does not coincide with the 2-fold axis of the bound DNA. Instead, each of the back-to-back interacting protomers binds a complete recognition sequence and thus, a double strand cut requires two sequential binding events. The interactions of HinP1I with a single palindromic target site are therefore not symmetric and do not occur in duplicate. In the HinP1I complex, a phenylalanine side chain is intercalated into the major groove, although less deeply than ThaI methionines or HincII glutamines. HinP1I introduces a 60° kink in the DNA and brings about much larger long-distance changes than ThaI. In summary, we conclude that the ThaI–DNA intercalation complex is unlike the HincII–DNA and HinP1I–DNA intercalation complexes.

Protein intercalation in intact DNA for bending or kinking

Proteins that bend or kink DNA have frequently been reported to insert one or several residues into the base stack (38,58). Examples in bacteria include the eubacterial integration host factor (IHF) (59), histone-like (HU) protein (60), Hbb protein (61), purine repressor purR (62) and the archaebacterial chromosomal protein Sac7d (63). Examples in eukaryotes comprise HMG group proteins SRY (58), Sox2 (64), Sox17 (65), the nucleosome binding and remodeling protein NHP6A (66) and the TATA box binding protein (TBP) (67,68). In all cases, the DNA intercalation is from the minor groove side, and involves either one or two amino acids per site. Protein residues are 'wedged' into the minor groove, but do not penetrate the base stack. This mode of intercalation expands the minor groove and compresses the major groove, which either induces or stabilizes a drastic DNA bend or kink. The identity of the inserted residues varies. In several cases methionine is involved (Sac7d, Sox2, Sox17, NHP6A). Nevertheless, the intercalation by ThaI and DNA bending proteins is dissimilar, because the ThaI Met47 penetrates the base stack

entirely, which shifts the DNA register, but does not induce a major bend or kink.

Protein intercalation in intact DNA to provide access

DNA methyltransferases form a large group of proteins that insert amino acid residues into intact DNA. In the 'classic' M.HhaI–DNA complex, a glutamine and a serine (from the major and minor groove side, respectively) displace the substrate base from the DNA stack into a catalytic pocket of the enzyme (69). The intercalation by methyltransferases is very different from the one observed for ThaI. In the former, the inserted amino acids extrude a base and fill the resulting void next to the estranged base. In contrast, the space for methionine intercalation in the ThaI–DNA complex is mostly created by a register shift in the absence of any nucleotide flipping.

Examples of intercalation in damaged DNA

Intercalation is rare in proteins that deal with intact DNA, but common in enzymes that process damaged DNA. Intercalation without flipping has been reported for very short patch repair endonuclease (Vsr) (70) and for the mismatch repair protein MutS (71,72). It has also been seen in the 'scanning' complex of the base excision repair enzyme MutM (73,74). Once the lesion has been found, base excision and direct alkylation repair proteins displace the damaged base and fill the void next to the estranged one (75–78). Nucleotide excision repair enzymes DDB1–DDB2 and Rad4–Rad23 insert amino acids or even a complete hairpin to extrude photoproducts from their substrates (79,80). Finally, photolyase displaces the base opposite to damage to make space for the repair of the photoproduct within the DNA stack (81).

ThaI versus other known examples of intercalation

DNA intercalation is common as a means to detect and repair damaged DNA. It can also occur in intact DNA in the scanning complexes of the repair pathways. 'Deep' intercalation into undamaged DNA is the rule for enzymes that require access to bases modify them chemically. 'Shallow' intercalation (from the minor groove side) into intact DNA is typical for proteins that bend or kink DNA. Thus, the ThaI–DNA complex that we have described in this work adds another prototype for DNA intercalation to the collection.

ACCESSION NUMBER

3NDH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful for access to the DESY (Hamburg, Germany), BESSY (Berlin, Germany) and DIAMOND (Didcot, United Kingdom) protein crystallography beamlines. In particular, we would like to thank Prof H.B. (BW6, DORIS, DESY) for allocation of beamtime at short notice.

FUNDING

EMBO/HHMI Young Investigator Award (to M.B.); a Foundation for Polish Science START Scholarship (to M.F.); EC FP7 grant 'Proteins in Health and Disease' (HEALTH-PROT, GA No 229676). M.F. is a member of the TEAM Program co-financed by the Foundation for Polish Science and the EU European Regional Development Fund. Funding for open access charge: EMBO/HHMI Young Investigator Award (to M.B.).

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. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
3. 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.
4. 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.
5. Xu,Q.S., Kucera,R.B., Roberts,R.J. and Guo,H.C. (2004) An asymmetric complex of restriction endonuclease MspI on its palindromic DNA recognition site. *Structure*, **12**, 1741–1747.
6. Venclovas,C., Timinskas,A. and Siksnys,V. (1994) Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins*, **20**, 279–282.
7. Niv,M.Y., Ripoll,D.R., Vila,J.A., Liwo,A., Vanamee,E.S., Aggarwal,A.K., Weinstein,H. and Scheraga,H.A. (2007) Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res.*, **35**, 2227–2237.
8. Pingoud,V., Wende,W., Friedhoff,P., Reuter,M., Alves,J., Jeltsch,A., Mones,L., Fuxreiter,M. and Pingoud,A. (2009) On the divalent metal ion dependence of DNA cleavage by restriction endonucleases of the EcoRI family. *J. Mol. Biol.*, **393**, 140–160.
9. Yang,W. (2008) An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.*, **15**, 1228–1231.
10. Xie,F., Briggs,J.M. and Dupureur,C.M. (2010) Nucleophile activation in PD...(D/E)xK metallonucleases: an experimental and computational pK(a) study. *J. Inorg. Biochem.*
11. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
12. 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.
13. 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.
14. Truglio,J.J., Rhau,B., Croteau,D.L., Wang,L., Skorvaga,M., Karakas,E., DellaVecchia,M.J., Wang,H., Van Houten,B. and Kisker,C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J.*, **24**, 885–894.

15. 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.
16. Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.
17. 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.
18. Lukacs,C.M., Kucera,R., Schildkraut,I. and Aggarwal,A.K. (2000) Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 A resolution. *Nat. Struct. Biol.*, **7**, 134–140.
19. Lambert,A.R., Sussman,D., Shen,B., Maunus,R., Nix,J., Samuelson,J., Xu,S.Y. and Stoddard,B.L. (2008) Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure*, **16**, 558–569.
20. Hashimoto,H., Shimizu,T., Imasaki,T., Kato,M., Shichijo,N., Kita,K. and Sato,M. (2005) Crystal structures of type II restriction endonuclease EcoO109I and its complex with cognate DNA. *J. Biol. Chem.*, **280**, 5605–5610.
21. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
22. 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.
23. Wenz,C., Jeltsch,A. and Pingoud,A. (1996) Probing the indirect readout of the restriction enzyme EcoRV. Mutational analysis of contacts to the DNA backbone. *J. Biol. Chem.*, **271**, 5565–5573.
24. 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.
25. 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.
26. 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.
27. Little,E.J., Babic,A.C. and Horton,N.C. (2008) Early interrogation and recognition of DNA sequence by indirect readout. *Structure*, **16**, 1828–1837.
28. Babic,A.C., Little,E.J., Manohar,V.M., Bitinaite,J. and Horton,N.C. (2008) DNA distortion and specificity in a sequence-specific endonuclease. *J. Mol. Biol.*, **383**, 186–204.
29. Joshi,H.K., Etzkorn,C., Chatwell,L., Bitinaite,J. and Horton,N.C. (2006) Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism. *J. Biol. Chem.*, **281**, 23852–23869.
30. 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.
31. Golovenko,D., Manakova,E., Tamulaitiene,G., Grazulis,S. and Siksnys,V. (2009) Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII. *Nucleic Acids Res.*, **37**, 6613–6624.
32. Kostrewa,D. and Winkler,F.K. (1995) Mg^{2+} binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 A resolution. *Biochemistry*, **34**, 683–696.
33. Horton,N.C. and Perona,J.J. (1998) Role of protein-induced bending in the specificity of DNA recognition: crystal structure of EcoRV endonuclease complexed with d(AAAGAT) + d(ATCTT). *J. Mol. Biol.*, **277**, 779–787.
34. Horton,J.R., Zhang,X., Maunus,R., Yang,Z., Wilson,G.G., Roberts,R.J. and Cheng,X. (2006) DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.*, **34**, 939–948.
35. McConnell,D.J., Searcy,D.G. and Sutcliffe,J.G. (1978) A restriction enzyme Tha I from the thermophilic mycoplasma Thermoplasma acidophilum. *Nucleic Acids Res.*, **5**, 1729–1739.
36. Ruepp,A., Graml,W., Santos-Martinez,M.L., Koretke,K.K., Volker,C., Mewes,H.W., Frishman,D., Stocker,S., Lupas,A.N. and Baumeister,W. (2000) The genome sequence of the thermoacidophilic scavenger Thermoplasma acidophilum. *Nature*, **407**, 508–513.
37. Strobl,J.S. and Thompson,E.B. (1984) Methylation of either cytosine in the recognition sequence CGCG inhibits ThaI cleavage of DNA. *Nucleic Acids Res.*, **12**, 8073–8083.
38. Werner,M.H., Gronenborn,A.M. and Clore,G.M. (1996) Intercalation, DNA kinking, and the control of transcription. *Science*, **271**, 778–784.
39. Chojnowski,G. and Bochtler,M. (2010) DIBER: protein, DNA or both? *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 643–653.
40. Leslie,A.W.G. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
41. Collaborative Computational Project Number 4. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
42. Bricogne,G., Vonrhein,W., Paciorek,C., Flensburg,M., Schiltz,E., Blanc,P., Rovesi,R. and Evans,M.G. (2002) Enhancements in AUTOSHARP and SHARP with applications to difficult phasing problems. *Acta Cryst.*, **A58(Suppl.)**, C239.
43. 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.
44. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
45. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
46. 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.
47. Brunger,A.T. (2007) Version 1.2 of the crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
48. Pingoud,A. and Wende,W. (2007) A sliding restriction enzyme pauses. *Structure*, **15**, 391–393.
49. Huai,Q., Colandene,J.D., Topal,M.D. and Ke,H. (2001) Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat. Struct. Biol.*, **8**, 665–669.
50. Huai,Q., Colandene,J.D., Chen,Y., Luo,F., Zhao,Y., Topal,M.D. and Ke,H. (2000) Crystal structure of NaeI-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J.*, **19**, 3110–3118.
51. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38(Suppl.)**, W545–W549.
52. 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.
53. Bujnicki,J.M. and Rychlewski,L. (2001) Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene*, **267**, 183–191.
54. Skirgaila,R., Grazulis,S., Bozic,D., Huber,R. and Siksnys,V. (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol.*, **279**, 473–481.
55. Horton,N.C., Newberry,K.J. and Perona,J.J. (1998) Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl Acad. Sci. USA*, **95**, 13489–13494.
56. Imhof,P., Fischer,S. and Smith,J.C. (2009) Catalytic mechanism of DNA backbone cleavage by the restriction enzyme EcoRV: a quantum mechanical/molecular mechanical analysis. *Biochemistry*, **48**, 9061–9075.

57. Mones,L., Simon,I. and Fuxreiter,M. (2007) Metal-binding sites at the active site of restriction endonuclease BamHI can conform to a one-ion mechanism. *Biol. Chem.*, **388**, 73–78.
58. Werner,M.H., Huth,J.R., Gronenborn,A.M. and Clore,G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705–714.
59. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
60. Swinger,K.K., Lemberg,K.M., Zhang,Y. and Rice,P.A. (2003) Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J.*, **22**, 3749–3760.
61. Mouw,K.W. and Rice,P.A. (2007) Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending protein Hbb. *Mol. Microbiol.*, **63**, 1319–1330.
62. Schumacher,M.A., Choi,K.Y., Zalkin,H. and Brennan,R.G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, **266**, 763–770.
63. Robinson,H., Gao,Y.G., McCrary,B.S., Edmondson,S.P., Shriver,J.W. and Wang,A.H. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, **392**, 202–205.
64. Remenyi,A., Lins,K., Nissen,L.J., Reinbold,R., Scholer,H.R. and Wilmanns,M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.*, **17**, 2048–2059.
65. Palasingam,P., Jauch,R., Ng,C.K. and Kolatkar,P.R. (2009) The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J. Mol. Biol.*, **388**, 619–630.
66. Masse,J.E., Wong,B., Yen,Y.M., Allain,F.H., Johnson,R.C. and Feigon,J. (2002) The S. cerevisiae architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. *J. Mol. Biol.*, **323**, 263–284.
67. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
68. Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.
69. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
70. Tsutakawa,S.E., Jingami,H. and Morikawa,K. (1999) Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell*, **99**, 615–623.
71. Obmolova,G., Ban,C., Hsieh,P. and Yang,W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703–710.
72. Lamers,M.H., Perrakis,A., Enzlin,J.H., Winterwerp,H.H., de Wind,N. and Sixma,T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, **407**, 711–717.
73. Banerjee,A., Santos,W.L. and Verdine,G.L. (2006) Structure of a DNA glycosylase searching for lesions. *Science*, **311**, 1153–1157.
74. Qi,Y., Spong,M.C., Nam,K., Banerjee,A., Jiralerspong,S., Karplus,M. and Verdine,G.L. (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature*, **462**, 762–766.
75. Fromme,J.C., Banerjee,A., Huang,S.J. and Verdine,G.L. (2004) Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature*, **427**, 652–656.
76. Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*, **384**, 87–92.
77. Barrett,T.E., Savva,R., Panayotou,G., Barlow,T., Brown,T., Jiricny,J. and Pearl,L.H. (1998) Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, **92**, 117–129.
78. Yang,C.G., Yi,C., Duguid,E.M., Sullivan,C.T., Jian,X., Rice,P.A. and He,C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961–965.
79. Scrima,A., Konickova,R., Czyzewski,B.K., Kawasaki,Y., Jeffrey,P.D., Groisman,R., Nakatani,Y., Iwai,S., Pavletich,N.P. and Thoma,N.H. (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, **135**, 1213–1223.
80. Min,J.H. and Pavletich,N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570–575.
81. Mees,A., Klar,T., Gnau,P., Hennecke,U., Eker,A.P., Carell,T. and Essen,L.O. (2004) Crystal structure of a photolyase bound to a CPD-like DNA lesion after *in situ* repair. *Science*, **306**, 1789–1793.