

Restriction endonuclease BpuJI specific for the 5'-CCCGT sequence is related to the archaeal Holliday junction resolvase family

Rasa Sukackaite¹, Arunas Lagunavicius¹, Kornelijus Stankevicius¹,
Claus Urbanke², Česlovas Venclovas¹ and Virginijus Siksnys^{1,*}

¹Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania and ²Strukturanalyse, Medizinische Hochschule Hannover, Carl Neuberg Strasse 1, D-30632 Hannover, Germany

Received February 1, 2007; Revised February 15, 2007; Accepted March 5, 2007

ABSTRACT

Type IIS restriction endonucleases (REases) recognize asymmetric DNA sequences and cleave both DNA strands at fixed positions downstream of the recognition site. REase BpuJI recognizes the asymmetric sequence 5'-CCCGT, however it cuts at multiple sites in the vicinity of the target sequence. We show that BpuJI is a dimer, which has two DNA binding surfaces and displays optimal catalytic activity when bound to two recognition sites. BpuJI is cleaved by chymotrypsin into an N-terminal domain (NTD), which lacks catalytic activity but binds specifically to the recognition sequence as a monomer, and a C-terminal domain (CTD), which forms a dimer with non-specific nuclease activity. Fold recognition approach reveals that the CTD of BpuJI is structurally related to archaeal Holliday junction resolvases (AHJR). We demonstrate that the isolated catalytic CTD of BpuJI possesses end-directed nuclease activity and preferentially cuts 3 nt from the 3'-terminus of blunt-ended DNA. The nuclease activity of the CTD is repressed in the apo-enzyme and becomes activated upon specific DNA binding by the NTDs. This leads to a complicated pattern of specific DNA cleavage in the vicinity of the target site. Bioinformatics analysis identifies the AHJR-like domain in the putative Type III enzymes and functionally uncharacterized proteins.

INTRODUCTION

In a typical Type II restriction–modification (R-M) system, two enzymes work together to protect host cell

DNA and exclude invading exogenous DNA. A restriction endonuclease (REase) recognizes a specific DNA sequence and cleaves it if it is not modified. The associated DNA methyltransferase (MTase) recognizes the same sequence and confers host DNA protection from cleavage by methylation of adenine or cytosine bases within the same recognition sequence. Host DNA is normally methylated by the MTase following replication, whereas invading non-self DNA is not, and therefore it is cleaved by REase. There are four main groups of R-M systems called Types I, II, III and IV (1).

Type II R-M systems constitute the largest fraction that has been characterized biochemically (2). Most often Type II R-M systems contain separate REase and MTase enzymes. Genes of Type II R-M enzymes are usually located in the close vicinity or even overlap. The MTases share several conserved amino acid motifs, but REase proteins show high sequence diversity and different behaviours categorized into 11 overlapping subtypes (1).

Orthodox Type II enzymes like EcoRI or BamHI recognize palindromic sequences of 4–8 bp in length, and cleave within these sequences to generate a defined restriction pattern of products that have both a 5'-PO₄ and a 3'-OH termini (3). Some recognize discontinuous palindromes, interrupted by a segment of specified length but unspecified sequence. The DNA fragments produced upon cleavage by orthodox Type II REases have 'blunt' or 'sticky' ends with 3'- or 5'-overhangs of fixed length of up to nine nucleotides. The most common cleavage products are 4 nt 5'-overhangs or blunt ends (2). Orthodox Type II enzymes often are homodimers, and each monomer harbours a single active site. Different DNA cleavage patterns most often result from changes in the dimerization mode that alter the distance between the active sites and hence the number of base pairs interspaced between scissile phosphates (3). Alternatively, DNA deformation and nucleotide flips may

*To whom correspondence should be addressed. Email: siksnys@ibt.lt
Present address:

Arunas Lagunavicius, Fermentas UAB, Graičiūno 8, LT-02241 Vilnius, Lithuania

be employed by Type II REases to generate a distinct cleavage pattern (4).

Type IIS REases that interact with the asymmetric sequence typically cut DNA in a strictly fixed location outside its recognition site to produce 'blunt' or 'sticky' ends with 3'- or 5'-overhangs (2). For example, FokI, an archetypal Type IIS restriction enzyme, recognizes the sequence 5'-GGATG and cuts top and bottom strands 9 and 13 nt downstream of this site, respectively (5). In solution FokI exists as a monomer, which has a bipartite structure. The N-terminal domain (NTD) is responsible for DNA binding while the catalytic C-terminal domain (CTD) contains a single active site (6,7). To cleave both strands at one site, two catalytic domains have to associate to give a dimer with two active sites (8,9) Dimerization of the catalytic domains fixes the distance between the active sites and determines the cleavage pattern of FokI. Many other Type IIS enzymes follow the same pattern (10). Type IIS enzyme HphI, however, cuts two DNA strands either 8 and 7 nt away or 9 and 8 nt away (11). The metal-independent Type IIS REase BfiI usually cuts bottom and top DNA strands 4 and 5 nt downstream of the target site 5'-ACTGGG, respectively, however additional cleavages of the top strand were also detected 6 and 7 nt away from the site (12). Thus, both HphI and BfiI show some variations in the cleavage site, however HphI maintains the cleavage register on top and bottom strands, whereas BfiI action on the top strand seems to be detached from that on the bottom. Interestingly, the 'chimeric' FokI variants produced by linking various DNA binding domains to the cleavage domain of FokI yield multiple cuts on both DNA strands (13,14).

Type I, III and IV restriction enzymes generally cleave DNA at variable positions, however they are multifunctional proteins which use translocation mechanisms and require ATP or GTP to cut DNA at sites distant from their binding sites (15,16). Some Type IIB restriction enzymes like AtoI (17) display variable cleavage pattern on some sequences containing the recognition site, however they require AdoMet and Mg^{2+} for DNA cleavage. Type II REases, which need Mg^{2+} ions as a single reaction cofactor, generally cleave the DNA at a defined location within or near the target sequence.

Here we show that Type II REase BpuJI, specific for the asymmetric sequence 5'-CCCGT, introduces multiple cuts in both DNA strands at both sides of the recognition site. We provide experimental evidence that BpuJI is comprised of two domains with separate functions: the NTD is involved in specific DNA binding while the CTD is responsible for cleavage. We show that the isolated catalytic domain of BpuJI is an active stand-alone nuclease, which possesses a non-specific end-directed nuclease activity and shows a preference for the blunt-ended DNA. The unusual cleavage pattern characteristic for the wild type (wt) BpuJI results from the site-specific and end-directed nuclease activity of the CTD manifested upon the NTD binding to the recognition site. Bioinformatics analysis supported by mutational studies indicates that the catalytic domain of BpuJI belongs to the family of archaeal Holliday junction resolvases (AHJR).

MATERIALS AND METHODS

Proteins

Cloning of BpuJI R-M system will be described elsewhere. BpuJI REase was purified from *Escherichia coli* ER2267 cells, bearing the pACYC184 derivative with the inserted *bpuJIM1* and *bpuJIM2* genes and the pAL4A derivative with inserted *bpuJIR* gene. The plasmids were a gift from Fermentas, Vilnius, Lithuania. The cells were harvested by centrifugation, suspended in buffer A [10 mM potassium phosphate (pH 7.0), 0.2 M KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, 0.025% Triton X-100] and sonicated. After removing of the insoluble material by centrifugation, the crude extract was subjected to chromatography on heparin-sepharose, blue-sepharose and AH-sepharose (all from GE Healthcare, Uppsala, Sweden) columns using KCl gradient in buffer A. Fractions containing BpuJI were pooled and dialysed against storage buffer [10 mM Tris-HCl (pH 7.5 at 25°C), 0.2 M KCl, 0.1 mM EDTA, 1 mM DTT, 0.025% Triton X-100, 50% (v/v) glycerol]. Protein concentration was determined from A_{280} readings using an extinction coefficient of $57\,560\text{ M}^{-1}\text{ cm}^{-1}$ and expressed in terms of monomer.

Chymotrypsin was purchased from Sigma, Munich, Germany, other enzymes used were from Fermentas.

DNA

Phage λ , phage Φ X174 and pUC57 DNAs were purchased from Fermentas. Specific (203 bp) and non-specific (179 bp) PCR fragments were obtained using primers I and II (Supplementary Table 1), *Pfu* DNA polymerase, and pUC57 or pUC19 DNA (Fermentas) as a template, respectively. Primers III, IV and pUC57 template were used to synthesize specific PCR fragment with different flanking sequences. Before PCR reactions one of the primers was 5'-labelled with $[\gamma\text{-}^{33}\text{P}]\text{ATP}$ and T4 polynucleotide kinase. The purification of PCR fragments was performed by S-400 columns (GE Healthcare).

Oligonucleotide duplexes (Table 1) were obtained by annealing oligonucleotides synthesized at Metabion, Martinsried, Germany. Where indicated, oligonucleotides were 5'-labelled with $[\gamma\text{-}^{33}\text{P}]\text{ATP}$ and T4 polynucleotide kinase or 3'-labelled with $[\alpha\text{-}^{32}\text{P}]\text{ddATP}$ and terminal polynucleotidyl transferase.

Plasmid and λ DNA cleavage assay

Cleavage was carried out in reaction buffer [50 mM Tris-HCl (pH 7.5 at 25°C), 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml of BSA], reactions contained 20 $\mu\text{g/ml}$ DNA and 100 nM BpuJI. Aliquots were removed at fixed time intervals and mixed with one-third volume of loading dye solution containing 0.1% SDS. The samples were heated at 65°C for 20 min and analysed by electrophoresis through agarose.

To determine cleavage positions the reaction was quenched with phenol/chloroform, DNA was precipitated from aqueous phase and used as template for run-off sequencing. The BigDye[®] Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster

Table 1. Oligonucleotide duplexes

Duplex	Sequence
16/16(SP)	5'TCGGTACCCGTGGATC 3'AGCCATGGGCACCTAG
16/16	5'AGCGTAGCACTGGGCT 3'TCGCATCGTGACCCGA
26/26(SP)	5'GAGCTCGGTACCCGTGGATCCTCTAG 3' TCGAGCCATGGGCACCTAGGAGATCT
37/37	5'GTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGT 3' ACTTAAGCTCGAGCCATGGGCCCTAGGAGATCTCAT
37/25	5'GTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGT 3' ACTTAAGCTCGAGCCATGGGCCCT
25/37	5'GTGAATTCGAGCTCGGTACCCGGGG 3' ACTTAAGCTCGAGCCATGGGCCCTAGGAGATCTCAT
25/25	5'GTGAATTCGAGCTCGGTACCCGGGG 3' ACTTAAGCTCGAGCCATGGGCCCT

City, CA, USA), and primers V and VI (Supplementary Table 1) were used in sequencing reactions.

PCR fragment cleavage assay

BpuJI cleavage of PCR fragments P³³-labelled at 5'-terminus of the top or bottom strand was carried out in the reaction buffer (see above) at 25°C. Reaction mixes contained 1–2 nM of DNA fragment and 0.2–3 μM of BpuJI. In the oligonucleotide activation experiments, the equimolar amount of the specific 16/16(SP) duplex was added into the reaction mixture. Aliquots were removed at timed intervals, reaction was quenched with phenol/chloroform and DNA was precipitated from the aqueous phase. The samples were analysed in the standard sequencing gels alongside with the reaction products of dideoxy sequencing reactions obtained with the same end-labelled primer, template and Cycle Reader DNA sequencing kit (Fermentas). Radiolabelled DNA was detected and quantified using a Cyclone Storage Phosphor System with OptiQuant 3.0 software (Perkin-Elmer, Wellesley, MA, USA). To map cleavage sites on the complementary strand, some DNA samples were split into two, one-half was treated with T4 DNA polymerase in the presence of dNTP and analysed in the sequencing gel as described above.

Gel mobility shift assay for DNA binding

The ³³P-labelled oligoduplex at 2 nM concentration was incubated with 1–400 nM of BpuJI for 10 min at room temperature in a binding buffer [30 mM MES (pH 6.5), 30 mM His, 10% (v/v) glycerol, 0.2 mg/ml of BSA]. To monitor formation of the synaptic complexes, 2 nM of the radiolabelled DNA and unlabelled oligoduplexes 16/16(SP) and 26/26(SP) were added to the binding mixes keeping fixed 1 : 1 total DNA : protein monomer ratio. Samples were loaded onto 8% polyacrylamide gels and run in buffer containing 30 mM MES (pH 6.5), 30 mM His for 2–4 h at 6 V/cm. After electrophoresis gels were analysed using a Cyclone Storage Phosphor System, as above.

Sedimentation equilibrium

Sedimentation equilibrium experiments were performed with a Beckman/Coulter XLA analytical ultracentrifuge, in six-channel centrepieces allowing simultaneous investigation of 24 samples in a single run. Runs were done at 20°C and 12 000 rpm, absorbance readings at 280 nm were taken in 1 hour intervals. Samples at initial loading concentrations 1.6, 3.2 and 6.4 μM of protein in a buffer containing 10 mM Tris-HCl (pH 7.5), 0.2 M KCl, 0.1 mM EDTA, 0.1 mM DTT were spun for at least 18 h. It was assumed that equilibrium had been reached when the measured concentration profiles did not change measurably over the following 12 h. For signal smoothing, all profiles measured during this 12 h period at equilibrium were averaged. Apparent molar masses were calculated from the averaged profiles as described earlier, using a partial specific volume of $7.386 \times 10^{-4} \text{ m}^3 \text{ kg}^{-1}$. The latter was calculated from the amino acid composition using the values tabulated in SEDNTERP (<http://www.rasmb.bbri.org>).

Limited proteolysis

The digestion of BpuJI by chymotrypsin was carried out in proteolysis buffer [10 mM potassium phosphate (pH 7.4), 100 mM KCl] for 1 h at 25°C and terminated by adding phenylmethylsulphonyl fluoride (PMSF) to a final concentration of 1 mM. To isolate the N-terminal fragment, 0.4 mg/ml of BpuJI was digested with 0.5 μg/ml of chymotrypsin. The sample was loaded onto a Superdex 75 HR column (GE Healthcare) and developed with proteolysis buffer. To isolate the C-terminal fragment, proteolysis mixture, containing 0.4 mg/ml of BpuJI and 0.025 mg/ml of chymotrypsin, was applied to the heparin-sepharose column, washed with buffer A (see above) and eluted with KCl gradient. The pooled fractions containing BpuJI proteolytic fragments were dialysed against the storage buffer (see above), concentrations were determined by Bradford assay with NanoQuant kit (Roth, Karlsruhe, Germany) and expressed in terms of monomer. N-terminal sequencing was performed on

Applied Biosystems 477A sequencer in the University of Bristol Proteomics Facility.

Gel-filtration

Gel-filtration of BpuJI and its proteolytic fragments was carried out on TSK-GEL HPLC column SuperSW 2000 (Tosoh Bioscience, Montgomeryville, PA, USA) in a buffer containing 0.1 M sodium phosphate (pH 6.7) and 0.1 M sodium sulphate. The samples contained 1.5 μ M of protein in the same buffer. Elution from the column was monitored by measuring absorbance at 215 nm.

Gel filtration of protein–DNA complexes was carried out on Superdex 200 HR column (GE Healthcare) pre-equilibrated with buffer containing 50 mM Tris–HCl (pH 7.5), 0.1 M KCl and 2 mM CaCl₂. The samples were prepared in the same buffer as above and contained 3 μ M of protein and 3 μ M of 16/16(SP). The control samples were prepared without DNA or protein. Elution from the column was monitored by measuring absorbance at 280 nm.

The calibration curves were generated by measuring the elution volumes of a series of standard proteins (GE Healthcare). The molecular masses were calculated by interpolating elution volumes onto the calibration curve.

Oligonucleotide duplex cleavage assay

Cleavage of radiolabelled oligonucleotide duplexes was assessed by adding isolated CTD (final concentration 100 nM) to 2 nM of duplex in reaction buffer (see above) at 25°C. Aliquots were removed at timed intervals, mixed with loading dye [95% (v/v) formamide, 20 mM EDTA (pH 8.0), 0.01% bromophenol blue], heated at 95°C for 5 min and analysed by denaturing polyacrylamide/urea/TBE gels. Cleavage products of labelled 37/37 oligoduplex by XbaI, Acc65I, EcoRI, BamHI, Ecl136II and S1 nuclease were used as size markers.

Bioinformatics analysis

Fold recognition (an attempt to match sequence with known structures) was performed using HHpred server (<http://protevo.eb.tuebingen.mpg.de/hhpred>) (18), which is based on a comparison of profile Hidden Markov Models (HMMs) (19). To ensure consistency of the results in addition to the BpuJI CTD several other homologous sequences were analysed as well. Detection of sequence homologs of the BpuJI CTD was performed with saturated BLAST (20), which is based on multiple intermediate (transitive) sequence searches. Each search consisted of PSI-BLAST (21) run of up to eight iterations using low complexity filtering and a conservative *E*-value cutoff (10^{-4}) for sequence inclusion into the profile. Multiple sequence alignment was constructed using combination of PCMA (22) and ProbCons (23) followed by some manual adjustments to improve the hydrophobic packing of the core structure of the PD-(D/E)XK fold. The identification of additional conserved domains in the BpuJI CTD homologs was done using reverse-position-specific BLAST (RPS-BLAST) searches against the Conserved Domain Database (CDD) (24). For the

phylogenetic distribution analysis of BpuJI CTD-like domain in eukaryotes, additional BLAST searches were performed against translated NCBI databases of complete and incomplete genomes as well as EST databases.

Site-directed mutagenesis

Site-directed mutagenesis was performed by megaprimer PCR method (25). The mutations were confirmed by sequencing and mutant proteins purified following the same procedure described for the wt protein.

RESULTS

Preliminary characterization of the BpuJI REase

According to the REBASE database (2) the BpuJI REase is specific for the 5'-CCCGT sequence. The exact cleavage position of BpuJI is unknown, however the enzyme is assigned to the Type IIS subtype assuming that it cuts at least one strand of the DNA duplex outside of the asymmetric recognition sequence. Typically, Type IIS R-M systems have one REase gene and two MTase genes, one to modify each strand of the asymmetric recognition sequence. The BpuJI R-M system demonstrates similar architecture and contains a single *bpuJIR* gene for restriction enzyme flanked by two genes *bpuJIM1* and *bpuJIM2* encoding 5-methylcytosine MTases (K. Stankevicius, unpublished data).

Like most of the Type II restriction enzymes BpuJI requires Mg²⁺-ions for DNA cleavage. Phage lambda DNA contains 103 5'-CCCGT sequences and gives a complex cleavage pattern upon BpuJI treatment. On the other hand, pUC57 plasmid and phage Φ X174 DNA that have six and two recognition sites, respectively, give a distinct DNA fragmentation that corresponds to the BpuJI cleavage at 5'-CCCGT sequences (Figure 1). To determine the BpuJI cleavage site the linear products obtained upon Φ X174 DNA cleavage were subjected to run-off sequencing, however no clear termination was observed (data not shown).

PCR fragment cleavage by BpuJI

To map BpuJI strand breaks we employed the 203 bp DNA fragment containing a single 5'-CCCGT sequence positioned approximately in the middle. The fragment was obtained by PCR using a pair of primers flanking the 5'-CCCGT sequence in the multiple cloning sites region of pUC57. BpuJI cuts this fragment rather slowly (data not shown) therefore high concentrations of enzyme were used to achieve complete cleavage.

To determine the cleavage position the 1 nM of 203 bp PCR fragment P³³-labelled at 5'-terminus either at the top or bottom strand was incubated with 3 μ M of BpuJI, aliquots were removed at different time intervals and sizes of the resulting fragments were analysed in the sequencing gel. We found that BpuJI reaction on the top DNA strand yields multiple cuts located at varying distances on both sides of the recognition sequence (Figure 2A). Moreover, the short time (5 and 20 min) and the prolonged (2 and 5 h) reactions generate DNA fragments

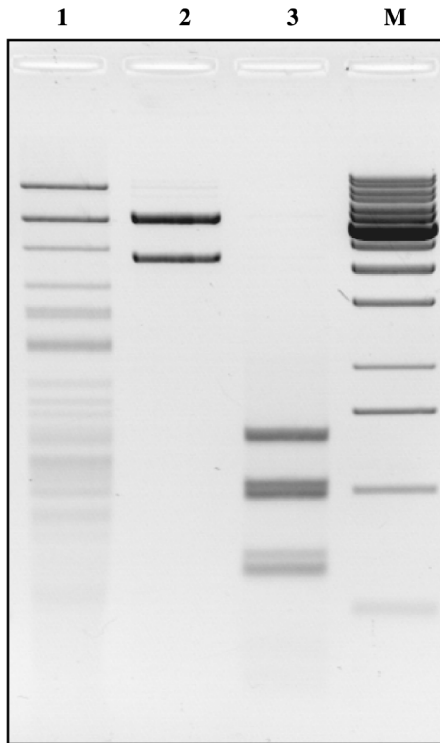


Figure 1. BpuJI digests of λ (lane 1), ϕ X174 (lane 2) and pUC57 (lane 3) DNAs. Total 20 μ g/ml of DNA was incubated with 100 nM of BpuJI for 1 h at 37°C, products were analysed by electrophoresis through 1% agarose. Lane M—DNA size marker.

of different length. The short time incubation yields a mixture of fragments that correspond to the top strand cleavage from 2 to 25 nt downstream the recognition site. The major products (corresponding to the strongest bands in the gel match to the BpuJI cut 2 and 11 nt downstream of the target site. The predominant product resulting from the prolonged incubations correspond to the BpuJI cut 18 nt upstream of the recognition site. Thus BpuJI cleaves the top strand on both sides of the recognition site but cuts first downstream of the target.

Mapping of the BpuJI breaks on the bottom strand also reveals cleavage at multiple sites (Figure 2B). The short time incubation (5 and 20 min) results in the bottom strand cleavage 6, 7, 13, 15, 16, 19 and 23 nt downstream of the recognition site. However, prolonged incubations generate a set of shorter products that correspond to the BpuJI cut 19 and 23 nt downstream of the target. Prolonged incubations reveal BpuJI induced breaks on the top strand upstream of the recognition site. To find out if cleavage of the bottom strand occurs on both sides of the recognition sequence, we mapped the BpuJI breaks on the bottom strand using a method based on the ability of T4 DNA polymerase to process differently DNA ends depending on the cleavage pattern. If cleavage generates a 5'-overhang, then the top strand will be extended by polymerase; if cleavage generates a 3'-overhang, then the top strand will be digested back; if cleavage generates a blunt end, then no change in length will be observed. When the fragment, 5'-labelled in top strand, was cleaved

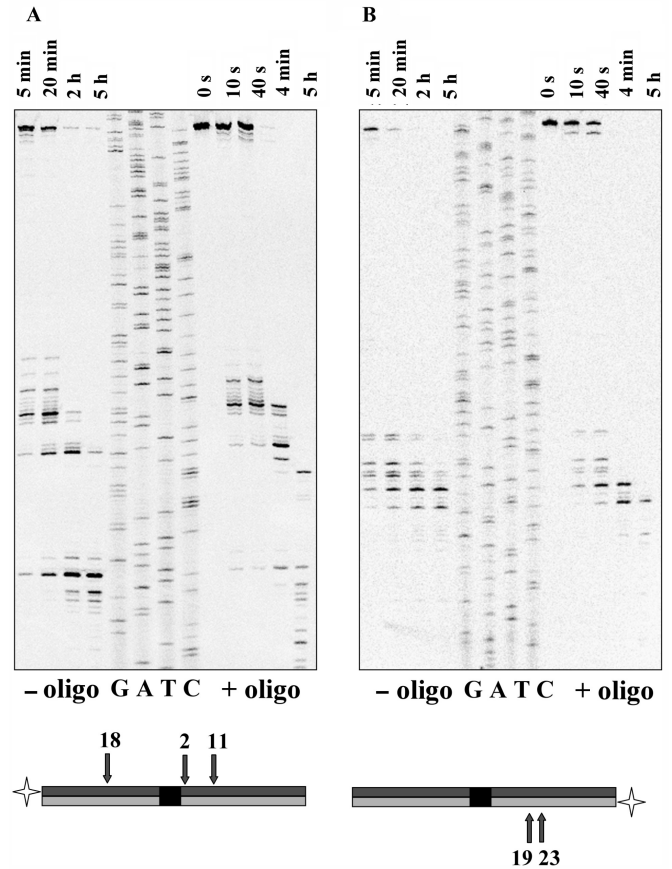


Figure 2. Cleavage of the specific PCR fragment by BpuJI. Total of 1 nM of the PCR fragment, 5'-labelled in top (A) or bottom (B) strand, was incubated with 3 μ M of BpuJI in the presence or absence of the 16/16(SP) oligoduplex. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through the standard sequencing gel. Lanes G, A, T, C are sequence ladders. In cartoons below the gels the BpuJI recognition sequence is indicated by a rectangle, 33 P is indicated by a star. The arrows show the major cleavage positions, numbers above the arrows indicate the distance between the recognition sequence and the position of cleavage.

with BpuJI and afterwards treated with T4 DNA polymerase, along with fragment corresponding to bottom strand cleavage downstream the recognition sequence, short fragments were observed indicating that complementary strand is also cleaved 13 and 16 nt upstream the recognition site (Supplementary Figure 1A). A similar variable cleavage pattern was observed at another BpuJI site in pUC57 flanked by different sequences, however the distances of major cuts in respect to the recognition sequence varied at different sites (data not shown).

BpuJI needs two recognition sites for its optimal activity

Many Type IIS REases display their maximal catalytic activities when bound to the two copies of their recognition sequences (10,26,27). Cleavage of the single site DNA by enzyme interacting with two recognition sites often can be stimulated by the addition of an oligoduplex that carries its recognition sequence (27–29). In these cases, the oligonucleotide provides *in trans* a second site

for the enzyme. In order to find out if BpuJI needs two sites for its optimal catalytic activity we have compared BpuJI cleavage rate of 203 bp PCR fragment in the presence or absence of a cognate oligoduplex. We have found that addition of the 16/16(SP) oligoduplex carrying the BpuJI recognition sequence considerably accelerated cleavage of the 203 bp PCR fragment. Indeed, the cleavage pattern of the top and bottom strands of 203 bp fragment obtained in the presence of cognate 16/16(SP) oligoduplex after the 40 s incubation was very similar to the pattern obtained in the absence of oligonucleotide after 2 h incubation (Figure 2). Oligoduplex stimulated BpuJI cleavage of the 203 bp fragment, however it did not change the cleavage positions: both strands are still cut at multiple sites. Non-specific oligonucleotide had no effect on the cleavage rate (data not shown).

BpuJI binds two DNA copies simultaneously

DNA binding properties of BpuJI were studied by gel shift analysis using ^{33}P -labelled 16/16(SP) and 16/16 oligonucleotide duplexes. To avoid cleavage, binding experiments were performed in the absence of Mg^{2+} ions which are necessary for catalysis. BpuJI binding to the DNA duplex 16/16(SP) containing single recognition site yielded two different complexes, named 1 and 2 (Figure 3A). No binding was observed under the same conditions using non-cognate 16/16 oligoduplex, confirming that complexes 1 and 2 result due to the specific BpuJI binding. The titration of a fixed amount of the 16/16(SP) duplex with increasing concentrations of BpuJI yielded first, at low concentrations of protein, the form with the higher mobility, complex 1; then at increased protein concentrations, a less mobile species, complex 2. Since 'oligonucleotide activation' of the BpuJI cleavage suggests that it can interact with two copies of the recognition sequence, the two complexes could represent BpuJI protein bound to one and two DNA molecules, as was demonstrated previously for REases SfiI and BfiI (27,30).

To find out if BpuJI protein can concurrently bind to two copies of specific DNA, oligoduplexes of different lengths were employed. BpuJI was added to the separate tubes containing mixtures of oligonucleotide duplexes 16/16(SP) and 26/26(SP) at different ratios but keeping the fixed 1 : 1 total DNA : protein monomer ratio. Traces of ^{33}P -labelled 16/16(SP) and 26/26(SP) were added into the binding mixtures for detection purpose. Gel shift analysis revealed that BpuJI when bound either to the 16/16(SP) or 26/26(SP) oligoduplex forms a single complex, which moves in the gel with characteristic electrophoretic mobility (Figure 3B). Incubation of the ^{33}P -labelled 16/16(SP) duplex with unlabelled cognate 26/26(SP) oligonucleotide resulted in the formation of the novel complex with intermediate electrophoretic mobility. The yield of this complex varied depending on the ratio of unlabelled 16/16(SP) and 26/26(SP) duplexes. This intermediate complex must represent BpuJI bound to both ^{33}P -labelled 16/16(SP) and unlabelled 26/26(SP) duplexes and provides a direct evidence of BpuJI binding of two recognition sites at the same time.

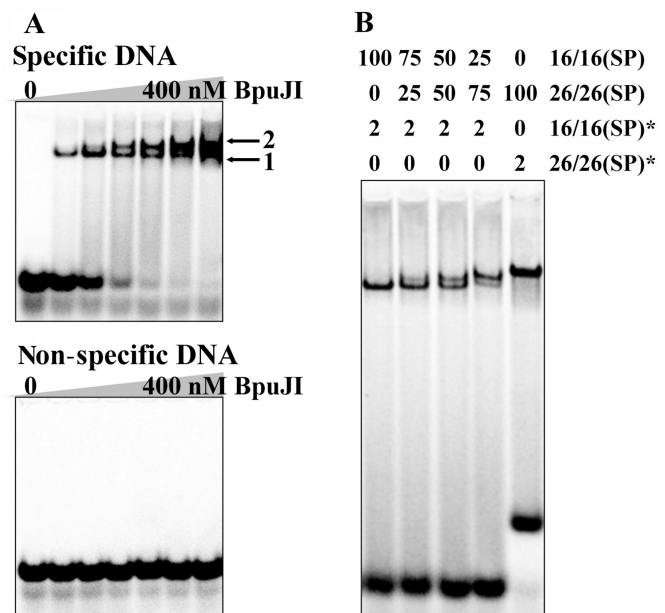


Figure 3. DNA binding by the BpuJI REase. (A) Gel shift analysis of BpuJI binding. Binding reactions contained 2 nM of ^{33}P -labelled specific 16/16(SP) (upper gel) or non-specific 16/16 (lower gel) duplex and 0, 8, 16, 40, 80, 160 and 400 nM of BpuJI in binding buffer. The samples were subjected to PAGE for 2 h, as described in Materials and Methods. (B) Analysis of the BpuJI–DNA complex. The reactions contained 100 nM of BpuJI, ^{33}P -labelled (the label is indicated by a star) and the unlabelled specific 16/16(SP) and 26/26(SP) duplexes. The concentrations of duplexes (in nM) are indicated above the relevant lanes. The samples were subjected to PAGE for 4 h, as described in 'Materials and methods'.

The 3'-end-directed nucleolytic activity

Alongside to the cleavage in the vicinity of the recognition site 5'-CCCGT, BpuJI treatment of the 5'-labelled 203 bp PCR fragment in the presence of the cognate oligonucleotide yielded extra products that correspond to the cleavage occurring few nucleotides away from the 3'-end of the fragment (Figure 2). These products are formed upon BpuJI action on the top DNA strand also in the absence of the activating oligonucleotide, albeit at a much slower rate.

To examine if BpuJI possesses similar activity on the non-specific DNA, the radioactively labelled 179 bp PCR fragment lacking the recognition site was incubated with BpuJI in the absence or presence of the cognate 16 bp oligoduplex. In the absence of the activating oligonucleotide no cleavage of top strand occurred after 5 and 20 min, however prolonged incubation yielded weak bands that correspond to the cut few nucleotides away from the 3'-end of the fragment (Figure 4A). A similar pattern was observed on the bottom strand (Figure 4B). Analysis of the products resulting from BpuJI cleavage of non-specific fragment in a sequencing gel revealed that the top strand was cut 3 and 6 nt from the 3'-end and 4 nt from the 3'-end on the bottom strand suggesting that BpuJI possesses end-directed nuclease activity. When products of non-specific nuclease activity were treated with T4 DNA polymerase, the full-length strand was synthesized confirming that

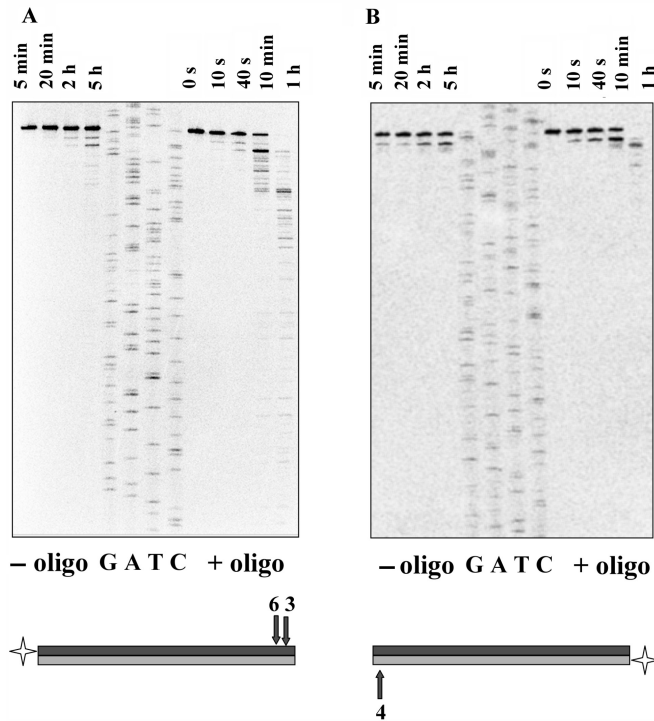


Figure 4. Cleavage of the non-specific PCR fragment by BpuJI. About 1 nM of the PCR fragment, 5'-labelled in top (A) or bottom (B) strand, was incubated with 3 μ M of BpuJI in the presence or absence of the 16/16(SP) oligoduplex. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through the standard sequencing gel. Lanes G, A, T, C are sequence ladders. In cartoons below the gels 33 P is indicated by a star, the arrows show the positions of the initial cleavage, numbers above the arrows indicate the distance from 3'-terminus.

cleavage occurred from the 3'-terminus of the labelled strand and not from 5'-terminus of the complementary strand (Supplementary Figure 1B).

Addition of the cognate oligonucleotide strongly stimulated the 3'-end directed activity. Indeed, after 10 min incubation, most of the 179 bp PCR fragment was cut into shorter fragments producing a DNA ladder (Figure 4). Cleavage of the bottom strand followed a similar pattern. Hence, in addition to sequence-directed nuclease activity, BpuJI also possesses end-directed nuclease activity, which is activated by binding to the recognition site. The two activities lead to complicated pattern of specific DNA cleavage and also to the end-directed cleavage of DNA without recognition site, which is very pronounced in the presence of specific DNA.

BpuJI is a dimer in solution and has a modular structure

The amino acid sequence predicts the BpuJI subunit molecular mass of 53.9 kDa. Sedimentation equilibrium at 12 000 rpm for BpuJI yielded molecular mass of 109 kDa (Supplementary Figure 2), which corresponds to the BpuJI dimer. Most of the Type IIS REases like FokI are monomers in solution (6,31,32) that dimerize upon DNA binding (8,10,33). Thus, in contrast to many Type IIS

enzymes, which are monomers, BpuJI is a dimer in solution.

To examine the structural organization of BpuJI in more detail, the protein was subjected to a limited proteolysis. Two proteolytic fragments, with apparent molecular masses of \sim 35 and \sim 20 kDa (Supplementary Figure 3A) were isolated from chymotrypsin digests and their N-terminal amino acids were determined by micro-sequencing. The N-terminal amino acid sequence of \sim 35 kDa fragment matched exactly the sequence at the N-terminus of BpuJI, while the N-terminal sequence of the \sim 20 kDa fragment started from the T303 residue in the intact protein. The N-terminal fragment was less protease-resistant than the C-terminal proteolytic fragment, however both fragments resisted further proteolysis in the presence of the cognate 16/16(SP) oligoduplex (data not shown).

Apparent mass values for the N-terminal and the C-terminal proteolytic fragments estimated from their elution volumes during size-exclusion chromatography (Supplementary Figure 3B) were \sim 38 and \sim 49 kDa, respectively. These values suggest that the N-terminal fragment is a monomer in solution while the C-terminal fragment is a dimer. Hence, BpuJI appears to be organized into the NTD and CTD structural domains with dimerization interface formed by the CTDs.

Functional analysis of the BpuJI domains

The isolated proteolytic fragments were subjected to DNA cleavage and binding assays to analyse the function of the individual BpuJI domains. DNA cleavage was examined on the phage λ DNA, which has multiple BpuJI recognition sites. In contrast to the intact BpuJI, which cleaves DNA into discrete set of fragments, the isolated CTD gives a smear characteristic for the non-specific nuclease (Figure 5A). The isolated NTD has no detectable nuclease activity on λ DNA. In gel shift assays the isolated NTD of BpuJI bound readily to the cognate 16/16(SP) oligoduplex, but not to the DNA lacking the recognition site (Figure 5B) while CTD showed no binding to either oligoduplex under the same experimental conditions. These results indicate that BpuJI domains have distinct functions: the NTD recognizes the 5'-CCCGT sequence and the CTD possesses the nuclease activity.

The N-terminal domain binds a single DNA copy as a monomer

According to the size exclusion chromatography experiments, the NTD of BpuJI is a monomer in solution. However, one cannot exclude that dimerization of the BpuJI NTD is required for cognate DNA binding. For example, NTDs of SdaI REase dimerize to bind a single copy of cognate DNA (34). Therefore, the stoichiometry of the BpuJI NTD binding to the cognate oligoduplex was analysed by the size exclusion chromatography. In the absence of DNA the isolated NTD eluted at a volume that corresponds to an apparent M_w of \sim 40 kDa, while 16/16(SP) duplex eluted with an apparent M_w of \sim 27 kDa (Supplementary Figure 4). The latter value is 2.5 times higher than the actual molecular mass of 16 bp oligoduplex (10.7 kDa), most likely due to cylindrical

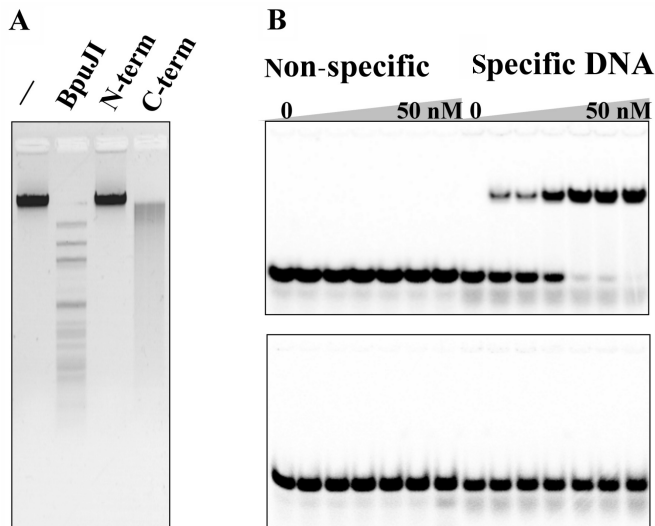


Figure 5. Functional analysis of the BpuJI domains. (A) λ DNA cleavage assay. Reactions contained 20 μ g/ml of DNA and 25 nM of BpuJI N- or C-terminal domain, respectively. After 1 h incubation at 37°C, products were analysed by electrophoresis through 1% agarose. (B) DNA binding assay. Total of 2 nM of 33 P-labelled 16/16(SP) or 16/16 duplex was incubated with increasing amounts (0, 1, 2, 5, 10, 20 and 50 nM) of the N-terminal (upper gel) or the C-terminal (lower gel) domains. The samples were subjected to PAGE for 2 h, as described in Materials and Methods.

shape and much higher frictional ratio compared to spherically shaped marker proteins. When 16/16(SP) duplex was added to the NTD at protein : DNA molar ratio 1:1, the complex eluted from gel filtration column as a single peak. The apparent mass value for this complex was ~54 kDa, close to that expected for monomer bound to DNA duplex. Thus, the NTD of BpuJI domain binds DNA as a monomer.

The C-terminal domain possesses the 3'-end-directed nucleolytic activity

The nuclease activity of isolated BpuJI CTD was further investigated using 5'-labelled 179 bp PCR fragment (Figure 6A). The CTD of BpuJI initially cut the labelled bottom DNA strand 4 nt from the 3'-terminus and afterwards processed it further to shorter fragments. The cleavage pattern produced by the isolated CTD of BpuJI on the non-specific 179 bp PCR fragment is very similar to that observed for the full-length BpuJI in the presence of specific DNA, suggesting that the end-directed nuclease activity of the intact protein results from the CTD.

To characterize the end-directed activity in more detail, a set of non-cognate oligoduplexes having different DNA ends was employed. The 37 and 25 nt oligonucleotides were annealed to form either blunt-ended 37/37 and 25/25 duplexes, or 37/25 duplex with 3'-overhang and 25/37 duplex with 5'-overhang, respectively. The duplexes were either 3'- or 5'-labelled in the top strand and subjected to cleavage by isolated BpuJI CTD (Figure 6B). The CTD cut 3'-labelled strand in the blunt ended 37/37 oligoduplex to yield a single ~3 nt product indicating that initial cleavage occurs from 3'-end. Cleavage analysis of the

37/37 duplex labelled at 5'-terminus reveals further digestion of the top strand resulting in the accumulation of smaller DNA fragments.

The top strand in the 37/25 oligoduplex with a protruding 3'-end was also cleaved by CTD, however more slowly than the blunt-ended 37/37 substrate. The cut site was located on the top strand within the duplex region of the oligonucleotide. The top strand in the 25/37 oligoduplex with a recessed 3'-end was the poorest substrate for the CTD. The same strand in the blunt-ended 25/25 duplex was cleaved ~4 nt from the 3'-terminus at respective rate and further processed to smaller products. Thus, the CTD of BpuJI shows preference for the blunt-ended oligoduplex and first cuts DNA 3 nt from the 3'-end.

Taken together, these results indicate that BpuJI has a modular structure with two functional domains. The NTD binds to the recognition site as a monomer, while the CTD is an active stand-alone nuclease, which shows preference to cleave ~3 nt from the 3'-terminus at blunt DNA ends. The catalytic domain in the intact BpuJI protein has a very low activity and can be activated by either proteolytic cleavage or specific DNA binding.

The C-terminal domain is related to archaeal Holliday junction resolvases of PD-(D/E)XK superfamily

To explore whether the BpuJI catalytic domain is related to any known nuclease structures, the fold recognition analysis was performed. Fold recognition results revealed that the best-matching structures are those of AHJR enzymes Hjc/Hje (35–37). These archaeal resolvases are members of the PD-(D/E)XK nuclease superfamily (38,39) and represent some of the closest approximation of the minimal core fold (Supplementary Figure 5A). Moreover, the catalytic triads of Holliday junction resolvases perfectly aligned with D³⁴⁸, E³⁶⁷ and K³⁶⁹ of the BpuJI CTD suggesting that residues of the YD³⁴⁸...E³⁶⁷VK³⁶⁹ motif form the active site of BpuJI (Supplementary Figure 5B).

To test this prediction, D348, E367 and K369 were targeted to alanine replacement mutagenesis. Single amino acid substitutions D348A, E367A and K369A strongly impaired the catalytic activity of the enzyme. The mutants showed only trace of the sequence-directed nuclease activity and no end-directed nuclease activity on specific PCR fragment (Supplementary Figure 6A), both in the absence and presence of the cognate oligonucleotide duplex. In gel shift assay D348A, E367A and K369A mutants bound to the 16/16(SP) duplex with BpuJI recognition site and not to the 16/16 duplex, as wt protein (Supplementary Figure 6B). Thus D348A, E367A and K369A proteins are defective in catalysis and not in DNA recognition, as expected for active site mutants.

BpuJI CTD-like domains are present in diverse sequence contexts and are distributed throughout all three kingdoms of life

Taken together, fold recognition and experimental data confidently assigns the BpuJI CTD to the diverse PD-(D/E)XK superfamily with distinct affinity to AHJRs.

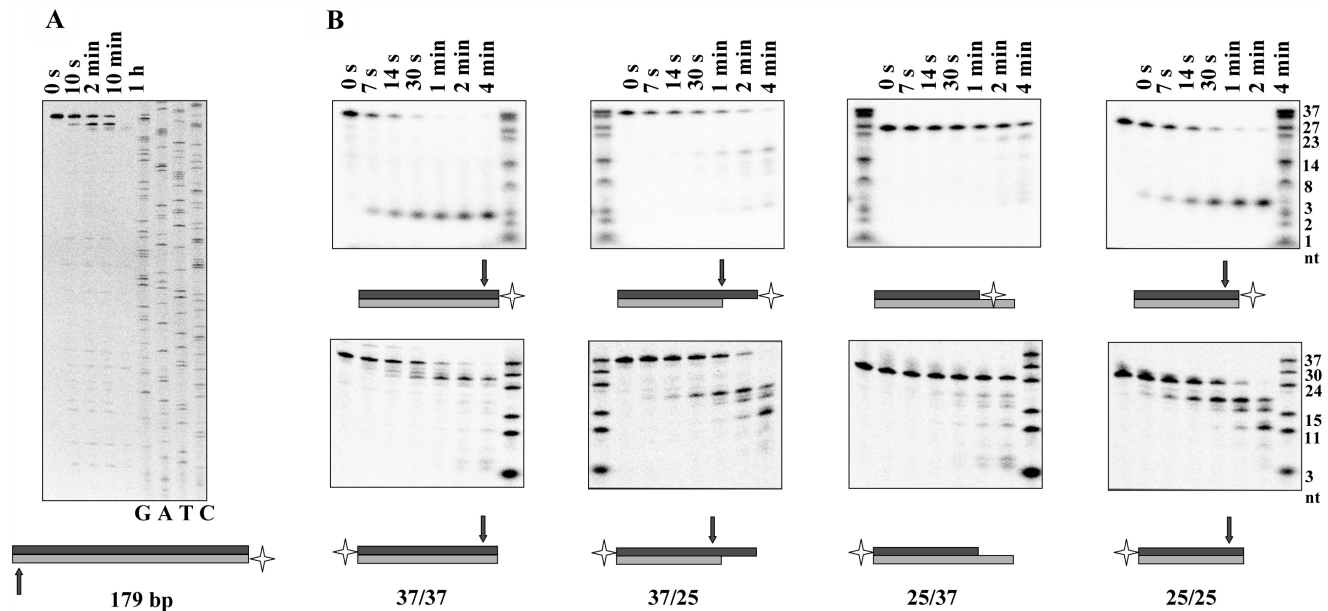


Figure 6. The 3'-end directed endonucleolytic activity of the isolated C-terminal domain of BpuJI. (A) Cleavage of the PCR fragment by the C-terminal domain. Total of 1 nM of the PCR fragment, 5'-labelled in bottom strand, was incubated with 3 μ M of the C-terminal domain at 4°C. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through the standard sequencing gel. Lanes G, A, T, C are sequence ladders. (B) Cleavage of oligonucleotide duplexes by the C-terminal domain. Total of 2 nM of oligoduplex, 3'-labelled (upper gels) or 5'-labelled (lower gels) in top strand, was incubated with 100 nM of the C-terminal domain at 25°C. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through denaturing 20% polyacrylamide. In cartoons below the gels the position of the label is indicated by a star, the arrows show the position of initial cleavage.

Despite of the fold similarity, the saturated PSI-BLAST search (20) starting with the BpuJI CTD sequence did not retrieve AHJR sequences, however it identified BpuJI CTD-like domains in quite a diverse pool of proteins. This suggests that homologs, detected with PSI-BLAST, are more closely related to the BpuJI CTD than AHJRs. Several protein families could be identified on the basis of common domain architecture (Figure 7A).

Family I includes uncharacterized bacterial and archaeal proteins (~900–1200 a.a.) that represent SWI2/SNF2-like superfamily II helicase fused to the BpuJI CTD-like domain (Figure 7A). This type of fusion has been noted before (38), however no putative protein function was suggested. We have noticed that the domain organization in Family I proteins is the same as in the Res subunits of characterized Type III R-M enzymes (Figure 7B) (16,40). Moreover, in absolute majority of cases we have detected an adenine-specific DNA MTase gene in close proximity. Therefore, both the domain organization and the genome context suggest that most if not all Family I proteins are Res subunits of Type III R-M enzymes. Yet, there are two major differences that distinguish Family I proteins from the currently characterized Type III R-M systems. First, the helicase region in the Family I proteins is more closely related to Rad54, while the corresponding region in Res subunits of conventional Type IIIA and IIIB enzymes is closer to Rad25 (Figure 7B). Second, putative Mod subunits colocalized with Family I sequences and those of the characterized Type IIIA and IIIB systems belong to

different orthologous clusters (COG1743 and COG2189, respectively).

Family II proteins in addition to the BpuJI CTD-like domain feature a common region that is most closely related to the N-terminal nucleotide-binding domain of HSP90 family of molecular chaperones (COG0326) (Figure 7A). This group includes long (~1100–2600 a.a.), mostly eukaryotic, proteins. To our knowledge, this is the first time the PD-(D/E)XK nuclease domain is found fused to the HSP90-like domain. What is the biological function of these proteins remains to be determined.

The third group (Family III) is comprised of archaeal and bacterial proteins (~400 a.a.) sharing a conserved N-terminal region. Unexpectedly, we have found that this region shows similarities to the NTD of McrB, the GTPase subunit of the 5'-methylcytosine-specific restriction enzyme McrBC (Figure 7). In McrB of *E. coli* the NTD is responsible for the recognition of the methylated target DNA sequence (41). By inference, results of our sequence analysis suggest that Family III proteins could be REases acting on 5'-methylcytosine DNA targets. However, unlike the two-component restriction system McrBC, these putative REases have both recognition and nuclease domains encoded within a single polypeptide chain.

Family IV consists of bacterial proteins (~250–400 a.a.) that have a conserved N-terminal region fused with the BpuJI CTD-like nuclease domain (Figure 7A). We have found that this conserved N-terminal region in some other proteins is fused to the H–N–H nuclease domain instead

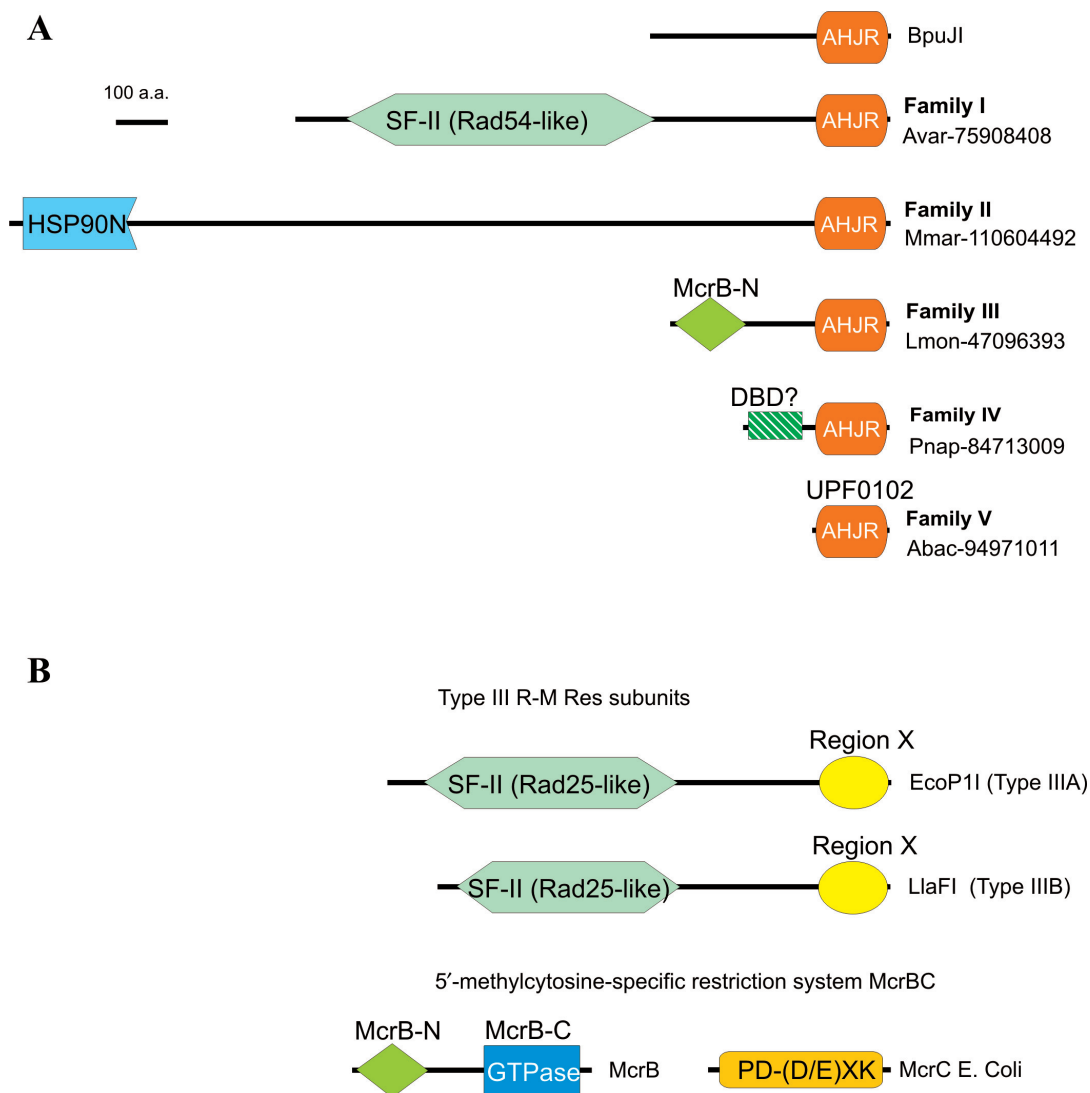


Figure 7. Domain architectures of (A) protein families sharing homology with the BpuJI nuclease domain identified in this study and (B) previously characterized Type III Res subunits and the McrBC system. AHJR, the archaeal Holliday junction resolvase domain similar to the C-terminal catalytic domain of BpuJI; for domains of superfamily II helicases, SF-II, the closest homolog of known structure is indicated; HSP90N, the N-terminal domain of HSP90; McrB-N and McrB-C, respectively, the N-terminal DNA binding and the C-terminal GTPase domains of the McrB subunit of 5'-methylcytosine-specific restriction enzyme; DBD?, unclassified putative DNA binding domain; UPF0102, uncharacterized protein family with the predicted single-domain nuclease fold related to AHJRs; Region X and PD-(D/E)XK, endonuclease domains of Type III and McrC restriction enzymes, respectively. Each architecture in (A) is illustrated with a single representative and corresponds to the grouping in the alignment (Supplementary Figure 5B).

of the BpuJI CTD-like domain. Although this conserved region does not show any apparent sequence similarity to characterized domains, its presence together with either nuclease domain suggests its likely involvement in DNA binding/recognition. However, in most cases we failed to find proximal MTase genes in genomic DNA sequences suggesting that this group of proteins do not function as restriction enzymes. Alternatively, they might act on methylated DNA targets, similarly to McrA.

We could also detect a statistically significant relationship ($E < 0.005$) between the BpuJI C-terminal nuclease domain and a large family of short (~130 a.a.) uncharacterized bacterial proteins (Pfam : UPF0102) (Family V, Figure 7A). The members of this family earlier have been

computationally identified as bacterial counterparts of AHJR enzymes (38), however their biological function has not been demonstrated.

We have detected BpuJI CTD homologs in bacteria, archaea and eukaryotes, yet the phylogenetic distribution pattern in eukaryotes is most intriguing. We found corresponding homologs in phylogenetically diverse branches of eukaryotic evolutionary tree including fungi/metazoa group, plants and green algae, apicomplexans, euglenozoa, entamoebas and oomycetes. With the exception of ascomycete fungi, all eukaryotic BpuJI CTD homologs form a closely related group with pairwise sequence identities in most cases exceeding 30%. For example, a BLAST search with a plant homolog can

retrieve fish and protists homologs with a significant *E*-value ($<10^{-3}$) and *vice versa*. Therefore, it came as a surprise that in metazoa group we could detect the relatives of the BpuJI CTD only in fish, molluscs and cnidarians. The inability to detect BpuJI CTD homologs outside of this group of marine animals is obviously not the result of insufficient sensitivity, since BLAST could easily find BpuJI CTD homologs in much more evolutionary distant eukaryotic organisms. It appears that more detailed studies are required to identify evolutionary mechanisms that led to this puzzling distribution pattern of BpuJI CTD homologs in eukaryotes.

DISCUSSION

The BpuJI REase identified in *Bacillus pumilus* recognizes the 5'-CCCGT sequence and generates a defined fragmentation pattern corresponding to the cleavage at the target site (Figure 1). However, in contrast to the Type II REases, which cut phosphodiester bonds at fixed location in respect to the recognition sequence, BpuJI makes multiple cuts at both DNA strands at distances varying from 2 to 23 nt (Figure 2). The cleavage positions are predominantly located downstream of the recognition site, however few cuts occur upstream of the target sequence. BpuJI is a homodimer in solution (Supplementary Figure 2) and requires Mg^{2+} -ion as the only cofactor for DNA cleavage. Type IIB subtype REases, like BcgI (42) cleave double-stranded DNA at specified positions on both sides of their recognition sequences to excise a short fragment containing the recognition site. BcgI and other Type IIB enzymes, however, are multi-functional proteins with both MTase and REase activities, and require both Mg^{2+} and AdoMet for DNA cleavage. The organization of the BpuJI R-M system and cofactor requirements make it distinct from the Type IIB REases. BpuJI also differs from the Type I, III and IV restriction enzymes which cleave at multiple sites but are multi-functional proteins which use translocation mechanisms and require ATP or GTP to cleave DNA at sites distant from their binding sites (40).

The asymmetric recognition sequence and the organization of the BpuJI R-M system (single REase and two separate MTases for different strands) suggest that it could be categorized as Type IIS system. However, according to the current nomenclature the overriding criterion for inclusion as a Type II enzyme would be that it yields a defined fragmentation pattern and cleaves either within or close to its recognition sequence at a fixed site or with known and limited variability (1). Therefore, the current definition still allows classification of HphI (11) and BfiI (12) REases, which show some defined variability in the cleavage positions, as Type IIS enzymes but creates an uncertainty in the case for BpuJI due to the unspecified cuts.

BpuJI like FokI (7) and other Type IIS enzymes (43) has a modular architecture. Proteolysis revealed that BpuJI is composed of two major domains: an NTD which lacks catalytic activity but binds specifically to the recognition sequence as a monomer, and a CTD, which

forms a dimer possessing nuclease activity (Figure 5). The two functional domains in BpuJI are presumably connected by a flexible linker. BpuJI therefore consists of two physically separate domains, with catalytic and dimerization functions in the C-terminus and DNA recognition functions in the N-terminus. Full length BpuJI is a dimer in solution and therefore has two surfaces for specific DNA binding provided by the NTDs and forms synaptic complexes by binding concurrently to two separate segments of specific DNA (Figure 3B). Binding of two DNA copies by BpuJI promotes cleavage. Indeed, BpuJI cleavage of the PCR fragment is highly stimulated by addition of the oligoduplex, which carries the recognition sequence (Figure 2). Thus, BpuJI similarly to other Type IIS enzymes (10,26,27,33), displays optimal catalytic activity when bound to two sites.

Thus, the only principal difference between the BpuJI and orthodox Type IIS enzymes is the variability of the cleavage position. The data provided here demonstrate that the unique cleavage pattern of BpuJI is determined by the properties of the catalytic CTD. In the full-length BpuJI the catalytic domain has a very low activity and can be activated either by proteolytic cleavage or specific DNA binding. The activated catalytic CTD of BpuJI becomes a nuclease that has no defined sequence specificity but shows a 3'-end directed endonucleolytic activity and preferentially cuts ~ 3 nt away from the 3'-terminus of the blunt-ended DNA. The isolated catalytic domain of BpuJI is a dimer and each monomer contains a single active site similar to the PD-(D/E)XK superfamily enzymes. The preferential cut of one of the strands at the 3'-terminus by the activated CTD suggests that only one active site per dimer is used for cleavage. In the absence of the target sequence, the endonucleolytic activity of the catalytic domain in BpuJI is repressed (Figure 8), since BpuJI shows only trace activity on the non-cognate DNA. Cognate DNA binding by the NTDs of BpuJI, presumably triggers a conformational change that relieves the autoinhibition and activates the CTD (Figure 8). If the BpuJI recognition site is embedded within a long DNA fragment, the activated catalytic CTD may become capable of cleaving phosphodiester bond(s) at random position (determined by the linker length or/and sequence context) in the vicinity of the recognition site. Independent cleavage of the second strand nearby of the first cut however will be required to generate the double strand break, if only one active site per dimer is used for the cleavage. A linear DNA resulting from BpuJI cleavage at the recognition site may be further subjected to 3'-end directed endonucleolytic cleavage by the activated CTD. The BpuJI cleavage at the recognition site followed by the 3'-end directed endonucleolytic action results in the complicated cleavage pattern of the PGR fragment (Figure 2).

Despite of the common function, Type II restriction enzymes are phylogenetically diverse (44). Indeed, the catalytic domains of well characterized Type IIS REases were found to be recruited from at least three different nuclease superfamilies: PD-(D/E)XK (7), H-N-H (or $\beta\beta\alpha$ Me) (45), or PLD (43). The PD-(D/E)XK superfamily comprises not only the REases but also other

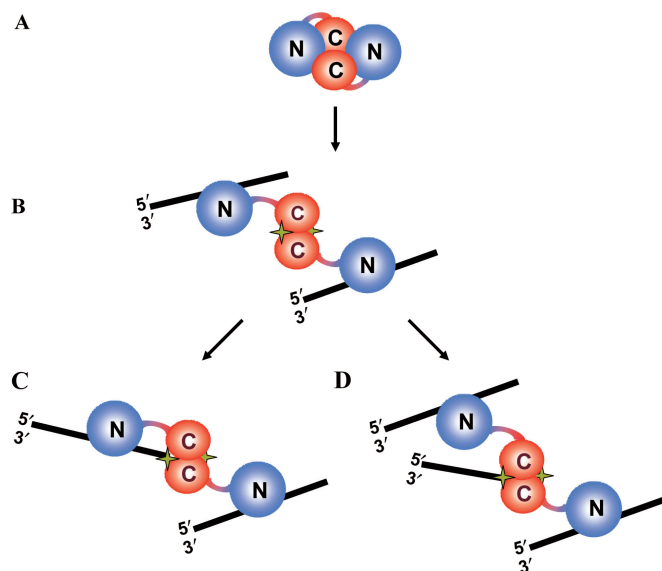


Figure 8. Cartoon representation of BpuJI action. BpuJI is a dimer and contains two domains: the N-terminal and the C-terminal (shown in blue and red, respectively). Each C-terminal domain possesses one active site, represented as a green star. In the absence of the target sequence nuclease activity of the C-terminal domain is repressed (A). Binding of cognate DNA by the N-terminal domains triggers a conformational change (B) and activates the catalytic domain. Activated C-terminal domain nicks DNA at random position in the vicinity of target sequence (C). In addition, activated C-terminal domain possesses the 3'-end directed nuclease activity and may cleave another DNA molecule (D).

nucleases implicated in DNA replication and repair. The catalytic domain of the archetypal Type IIS enzyme FokI belongs to the PD-(D/E)XK superfamily and shows structural resemblance to the phage lambda exonuclease (46). Fold recognition approach and mutational analysis reveal that the catalytic domain of BpuJI also belongs to the PD-(D/E)XK superfamily but is structurally more related to the AHJR. The closest structures are those of Holliday junction resolving enzymes Hjc/Hje (35–37). The Holliday junction resolving enzymes are structure-specific endonucleases that usually function as dimers. These enzymes bind to the four way DNA junctions and introduce paired nicks in the opposite strands to generate nicked duplex DNA species that are subsequently religated to produce recombinant products (39,47). It is likely that BpuJI restriction enzyme evolved through the fusion of the DNA binding domain to the archeal Holliday junction resolving nuclease and adapted the nucleolytic/nicking activity to generate a site-specific double strand break in DNA. Our data indicate that the catalytic domain of BpuJI acts as a stand-alone nuclease that possesses the 3'-end directed endonucleolytic activity on the double-stranded DNA. The presence of the single domain homologs that contain perfectly conserved active site suggests that at least some of them may be active as nucleases, albeit not in the context of R-M systems since the MTase gene in the close vicinity is missing. Bioinformatics analysis indicates that AHJR-like catalytic domain can also be found in different sequence contexts

and display wide phylogenetic distribution. One of the protein families containing the AHJR-like catalytic domain is likely to represent a new subfamily of the Type III R-M systems while another one may be related to the McrBC-like restriction enzymes (41). The function of the proteins belonging to the two other families that harbour the AHJR-like catalytic domain has yet to be identified.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Fermentas UAB for the BpuJI R-M system clone. We thank Donatas Vaitkevicius (Fermentas) for advice and help during the initial steps of BpuJI purification, Will Mawby (University of Bristol) for N-terminal sequencing, Mindaugas Zarembe and Giedrius Sasnauskas (Institute of Biotechnology) for discussions and comments on the manuscript. Research in the VS laboratory was supported by an intramural grant from the Institute of Biotechnology and “Center of Excellence - Biocell” Contract no: QLK2-CT-2002-30575. CV research was supported in part by the Howard Hughes Medical Institute International Research Scholar grant. The BpuJI sequence has been deposited in GenBank under the accession number EF409421.

Funding to pay the Open Access publication charges for this article was provided by “Center of Excellence - Biocell” Contract no: QLK2-CT-2002-30575.

REFERENCES

- Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2005) REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.*, **33**, D230–D232.
- Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
- 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.
- Sugisaki,H. and Kanazawa,S. (1981) New restriction endonucleases from *Flavobacterium okeanokoites* (FokI) and *Micrococcus luteus* (MluI). *Gene*, **16**, 73–78.
- Kaczorowski,T., Skowron,P. and Podhajska,A.J. (1989) Purification and characterization of the FokI restriction endonuclease. *Gene*, **80**, 209–216.
- Wah,D.A., Hirsch,J.A., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
- Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. USA*, **95**, 10570–10575.
- Catto,L.E., Ganguly,S., Milsom,S.E., Welsh,A.J. and Halford,S.E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res.*, **34**, 1711–1720.

10. Soundararajan, M., Chang, Z., Morgan, R.D., Heslop, P. and Connolly, B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J. Biol. Chem.*, **277**, 887–895.
11. Kleid, D., Humayun, Z., Jeffrey, A. and Ptashne, M. (1976) Novel properties of a restriction endonuclease isolated from *Haemophilus parahaemolyticus*. *Proc. Natl. Acad. Sci. USA*, **73**, 293–297.
12. Sasnauskas, G., Halford, S.E. and Siksnys, V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc. Natl. Acad. Sci. USA*, **100**, 6410–6415.
13. Kim, Y.G. and Chandrasegaran, S. (1994) Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci. USA*, **91**, 883–887.
14. Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA*, **93**, 1156–1160.
15. Murray, N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
16. McClelland, S.E. and Szelcun, M.D. (2004) The Type I and III restriction endonucleases: Structural Elements in Molecular Motors that Process DNA. In Pingoud, A. (ed). *Restriction Endonucleases*, Vol. 114. Springer-Verlag, Berlin Heidelberg, pp. 111–132.
17. Cesnaviciene, E., Petrusyte, M., Kazlauskienė, R., Maneliene, Z., Timinskas, A., Lubys, A. and Janulaitis, A. (2001) Characterization of AtoI, a restriction-modification system of a new type. *J. Mol. Biol.*, **314**, 205–216.
18. Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.
19. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
20. Li, W., Pio, F., Pawlowski, K. and Godzik, A. (2000) Saturated BLAST: an automated multiple intermediate sequence search used to detect distant homology. *Bioinformatics*, **16**, 1105–1110.
21. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
22. Pei, J., Sadreyev, R. and Grishin, N.V. (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics*, **19**, 427–428.
23. Do, C.B., Mahabhashyam, M.S., Brudno, M. and Batzoglou, S. (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.*, **15**, 330–340.
24. Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z. et al. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.*, **33**, D192–D196.
25. Barik, S. (1995) Site-directed mutagenesis by double polymerase chain reaction. *Mol. Biotechnol.*, **3**, 1–7.
26. Bath, A.J., Milsom, S.E., Gormley, N.A. and Halford, S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024–4033.
27. Lagunavicius, A., Sasnauskas, G., Halford, S.E. and Siksnys, V. (2003) The metal-independent type IIs restriction enzyme BfiI is a dimer that binds two DNA sites but has only one catalytic centre. *J. Mol. Biol.*, **326**, 1051–1064.
28. Oller, A.R., Vanden Broek, W., Conrad, M. and Topal, M.D. (1991) Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species. *Biochemistry*, **30**, 2543–2549.
29. Wentzell, L.M., Nobbs, T.J. and Halford, S.E. (1995) The SfiI restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J. Mol. Biol.*, **248**, 581–595.
30. Embleton, M.L., Williams, S.A., Watson, M.A. and Halford, S.E. (1999) Specificity from the synapsis of DNA elements by the Sfi I endonuclease. *J. Mol. Biol.*, **289**, 785–797.
31. Sektas, M., Kaczorowski, T. and Podhajska, A.J. (1992) Purification and properties of the MboII, a class-IIS restriction endonuclease. *Nucleic Acids Res.*, **20**, 433–438.
32. Tucholski, J., Skowron, P.M. and Podhajska, A.J. (1995) MmeI, a class-IIS restriction endonuclease: purification and characterization. *Gene*, **157**, 87–92.
33. Vanamee, E.S., Santagata, S. and Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69–78.
34. 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.
35. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. and Morikawa, K. (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure*, **9**, 197–204.
36. Bond, C.S., Kvaratskhelia, M., Richard, D., White, M.F. and Hunter, W.N. (2001) Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. USA*, **98**, 5509–5514.
37. Middleton, C.L., Parker, J.L., Richard, D.J., White, M.F. and Bond, C.S. (2004) Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic Acids Res.*, **32**, 5442–5451.
38. Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) Survey and summary: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
39. Lilley, D.M. and White, M.F. (2001) The junction-resolving enzymes. *Nat. Rev. Mol. Cell. Biol.*, **2**, 433–443.
40. Dryden, D.T., Murray, N.E. and Rao, D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
41. Pieper, U., Schweitzer, T., Groll, D.H. and Pingoud, A. (1999) Defining the location and function of domains of McrB by deletion mutagenesis. *Biol. Chem.*, **380**, 1225–1230.
42. Kong, H., Morgan, R.D., Maunus, R.E. and Schildkraut, I. (1993) A unique restriction endonuclease, BcgI, from *Bacillus coagulans*. *Nucleic Acids Res.*, **21**, 987–991.
43. 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.
44. Bujnicki, J.M. (2004) Molecular phylogenetics of restriction endonucleases. In Pingoud, A. (ed). *Restriction Endonucleases*. Springer-Verlag, Berlin Heidelberg, pp. 63–94.
45. Kriukiene, E., Lubiene, J., Lagunavicius, A. and Lubys, A. (2005) MnlII—the member of H-N-H subtype of Type IIS restriction endonucleases. *Biochim. Biophys. Acta.*, **1751**, 194–204.
46. Kovall, R.A. and Matthews, B.W. (1998) Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, **95**, 7893–7897.
47. Nishino, T., Ishino, Y. and Morikawa, K. (2006) Structure-specific DNA nucleases: structural basis for 3D-scissors. *Curr. Opin. Struct. Biol.*, **16**, 60–67.