BspRI restriction endonuclease: cloning, expression in *Escherichia coli* and sequential cleavage mechanism

Tamás Raskó, András Déró, Éva Klement, Krystyna Ślaska-Kiss, Eszter Pósfai, Katalin F. Medzihradszky, Daniel R. Marshak, Richard J. Roberts and Antal Kiss

1 Institute of Biochemistry, 2 Institute of Biophysics, 3 Proteomics Research Group, Biological Research Center of the Hungarian Academy of Sciences, Temesvári krt. 62, 6726 Szeged, Hungary, 4 Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158-2517, 5 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, 6 PerkinElmer Inc., 940 Winter Street, Waltham, MA 02451 and 7 New England Biolabs Inc., 240 County Road, Ipswich, MA 01938-2723, USA

Received September 16, 2009; Revised May 31, 2010; Accepted June 7, 2010

**ABSTRACT**

The GGCC-specific restriction endonuclease BspRI is one of the few Type IIP restriction endonucleases, which were suggested to be a monomer. Amino acid sequence information obtained by Edman sequencing and mass spectrometry analysis was used to clone the gene encoding BspRI. The *bspRIR* gene is located adjacent to the gene of the cognate modification methyltransferase and encodes a 304 aa protein. Expression of the *bspRIR* gene in *Escherichia coli* was dependent on the replacement of the native TTG initiation codon with an ATG codon, explaining previous failures in cloning the gene using functional selection. A plasmid containing a single BspRI recognition site was used to analyze kinetically nicking and second-strand cleavage under steady-state conditions. Cleavage of the supercoiled plasmid went through a relaxed intermediate indicating sequential hydrolysis of the two strands. Results of the kinetic analysis of the first- and second-strand cleavage are consistent with cutting the double-stranded substrate site in two independent binding events. A database search identified eight putative restriction-modification systems in which the predicted endonucleases as well as the methyltransferases share high sequence similarity with the corresponding protein of the BspRI system. BspRI and the related putative restriction endonucleases belong to the PD-(D/E)XK nuclease superfamily.

**INTRODUCTION**

Type IIP restriction endonucleases (REase) are characterized by recognition sequences displaying dyad axes of symmetry (palindromes), and constitute the most abundant class of characterized restriction enzymes (1). The first Type IIP REases, which were biochemically characterized, were shown to consist of two identical subunits: EcoRI (2), BclI (3), BstI (4) BamHI (5). Recognition of a symmetric recognition sequence by a homodimeric protein and cutting the two strands simultaneously using two active sites was an attractive model also because of the economy of the required protein synthesis, as first pointed out by Kelly and Smith (6). For a long time, the results of crystallographic studies supported the generalization that Type IIP REases are homodimers, e.g. (7–11) or tetramers (12).

To our knowledge, the first Type IIP REase, which was suggested to exist as a monomer was BspRI (13). BspRI of *Bacillus sphaericus* recognizes the sequence GGCC and cuts after the second G to produce blunt ends (14). The conclusion that the enzyme consists of a single subunit was derived from a comparison of molecular masses determined under native (gel filtration) and denaturing (SDS–polyacryamide gel electrophoresis) conditions. Later, based mostly on similar biochemical evidence as
for BspRI, a few other Type IIP REases were also reported to consist of a single polypeptide chain, such as BsuRI (GG/CC) (15), BcnI (CC/SGG) (16) DpnI (GmA/TC) (17), Sau96I G/GNCC (18), BshFI (GG/CC) (19). However, because of a lack of supporting structural data, the notion of monomeric Type IIP REases received little attention.

This changed when the X-ray structure of an MspI–DNA specific recognition complex was reported in 2004. MspI was shown to interact with its symmetric recognition sequence (C/GGG) as a monomer (20–21). Soon other articles describing structures of similar asymmetric complexes of three other Type IIP enzymes followed: HinPI (G/GCG) (22, 23), MvaI (CC/WGG) (24) and BcnI (CC/SGG) (25) establishing a new paradigm to think about this class of REases.

The gene of the BspRI methyltransferase (bspRIM), the cognate methyltransferase (MTase) of BspRI endonuclease, was cloned and expressed in Escherichia coli (26), but attempts to clone the BspRI REase gene (bspRIR) were not successful (A. Kiss, unpublished). The acceptance of the idea of monomeric Type IIP REases prompted us to revisit the project and to try to clone the bspRIR gene by an approach, which was not dependent on the expression of R.BspRI in E. coli.

Here we report the results of experiments, in which we used amino acid sequence information to identify the silent BspRI endonuclease gene on the previously cloned fragment, in the vicinity of the MTase gene. Replacing the native TTG start codon with ATG, and cloning the gene in an expression vector providing a strong promoter and Shine–Dalgarno sequence resulted in high-level expression of BspRI REase in E. coli. A database search identified eight putative restriction-modification systems, in which the predicted REases as well as the predicted MTases share high sequence similarity with the corresponding protein of the BspRI system. Secondary-structure prediction was used to determine whether R.BspRI can be assigned to any family of characterized metal-dependent REases. The DNA cleavage mechanism of BspRI was studied using a plasmid containing a single BspRI recognition site. This substrate allowed us to analyze kinetically both nicking and double-strand cleavage at the unique target site.

MATERIALS AND METHODS

Strains and growth conditions

Bacillus sphaericus R, originally isolated as a culture contaminant, is the native host of the BspRI R-M system (14). Bacillus sphaericus has recently been reclassified as Lysinibacillus sphaericus (27). Escherichia coli ER1821 F− glnV44, e14− (McrA−) endA1 thi-1 Δ(mcrC-mrr)114::IS10 obtained from New England Biolabs was used as cloning host. ER1821(DE3) was made by lysogenizing ER1821 with λDE3 using the λDE3 lysogenization kit of Novagen. ER1821(DE3) expresses T7 RNA polymerase upon induction with isopropyl-β-D-thiogalactopyranoside (IPTG). Bacteria were grown in LB medium (28) at 30°C (B. sphaericus) or at 37°C (E. coli). Ampicillin (Ap) and chloramphenicol (Cm) were used at 100 and 25 μg/ml, respectively. For BspRI overproduction ER1821(DE3 + pLysS + pET3H-BspRI) was grown to OD600 ~ 0.5, then BspRI production was induced by adding 0.4 mM IPTG to the culture and growth was continued for 4–5 h at 30°C.

DNA preparations

Bacillus sphaericus R genomic DNA was prepared from a 50 ml culture. Cells were sedimented by centrifugation, washed with 10 ml 20 mM Tris–HCl pH 8.0, then resuspended and lysed in a solution containing 50 mM Tris–HCl pH 7.5, 50 mM EDTA, 0.2% SDS and 200 μg/ml proteinase-K. After incubation at 37°C overnight, the DNA solution was extracted three times with phenol/ chloroform and precipitated with ethanol. The precipitated DNA was collected by a glass rod, dried and dissolved in TE buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA).

Plasmid pES1 contains the gene of the BspRI MTase on a ~9 kb BamHI fragment of B. sphaericus DNA cloned in pBR322 (26) (Figure 1A). pTZ-Bsp3 carries the segment of the bspRIR gene corresponding to the Q11–K97 peptide. It was constructed by PCR amplification using B. sphaericus genomic DNA as template and AK106/AK108 as primers (Figure 1B), and subsequent cloning of the PCR product in the commercial plasmid vector pTZ57R/T (Fermentas). pTZ-Bsp3 encodes the N-terminal M1–K97 peptide of R.BspRI. It was constructed by PCR-synthesis using pES1 as template and AK113/AK108 as primers (Figure 1B), and cloning the PCR product in pTZ57R/T. Plasmid pTZ-Bsp5, which contains the complete BspRI system, was made by inserting the 2920 bp EcoRV fragment of pES1 carrying part of the bspRIR gene and the intact bspRIM gene (Figure 1A) into the unique EcoRV site of pTZ-Bsp3. The orientation of the BspRI genes in pTZ-Bsp5 is opposite to the lac transcription on the plasmid.

To construct a plasmid overexpressing R.BspRI, first the EcoRV fragment of pES1 (Figure 1) was cloned into the SmaI site of pBAD24 (29) to yield pBAD-Bsp2. pBAD-Bsp2 lacks the beginning of the bspRIR gene. To reconstruct the complete BspRI system, the SalI-NdeI fragment of pTZ-Bsp5 was inserted between the Acc65I and SalI sites of pBAD24. The resulting plasmid (pBAD-Bsp3) encodes a BspRI variant, which carries a four amino acid extension at the N-terminus (MVLD). To facilitate purification of R.BspRI, first the EcoRV fragment of pES1 was cloned into the EcoRV fragment of pBAD24 (29) to yield pBAD-Bsp2. pBAD-Bsp2 lacks the beginning of the bspRIR gene. To reconstruct the complete BspRI system, the SalI-NdeI fragment of pTZ-Bsp5 was inserted between the Acc65I and SalI sites of pBAD24. The resulting plasmid (pBAD-Bsp3) encodes a BspRI variant, which carries a four amino acid extension at the N-terminus.
Plasmid 

extension (underlined): MHHHHHLDMAQ...

Plasmid pC194 (32) containing single sites for BspRI/BsuRI and HinP1I (33) was used as substrate to analyze single- and double-strand cleavage by these enzymes.

Plasmid DNA was prepared from *E. coli* cells by standard methods (28) or using commercial kits. For preparation of pC194, *B. subtilis* BD364(pC194) cells were treated with 5 mg/ml lysozyme before starting purification with the GenElute HP Plasmid Midiprep kit (Sigma).

**Oligonucleotides**

Oligonucleotides were purchased from Integrated DNA Technologies (IDT) or were synthesized in the BRC, Szeged. For approximate positions of the PCR primers see Figure 1B. AK106 and AK108 are pools of degenerate oligonucleotides, where R = A, G; Y = C, T; I = inosine. 

AK106 (5'-CAR AAR GTI GCI AAY ATI TTY ATI AAY) corresponds to the Q11KVANIFIN19 peptide of BspRI endonuclease (sense strand).

AK108 (5'-YTT YTG CCA RTT) corresponds to the N94WQK97 peptide of BspRI endonuclease (anti-sense strand).

AK111 (5'-GAT GGG TCT AAG ATA CTA TT) corresponds to the N29SILDPS97 peptide of BspRI endonuclease (anti-sense strand).

AK112 (5'-GAA ATG ATT TAT ATG ATG TG) hybridizes down-stream of the bspRM gene stop codon (sense strand).

AK113 (5'-GTC GAC ATG GCG CAG AGA AAA TAT GGT GCA) corresponds to the A2QRKYGA8 peptide of BspRI endonuclease, and carries an ATG start codon and a SalI site (underlined) as 5'-extension (sense strand).

**Other DNA techniques**

Restriction digestion, polymerase chain reaction, agarose gel electrophoresis and cloning in *E. coli* plasmid vectors were carried out using standard procedures (28). Restriction endonucleases, DNA polymerase large (Klenow) fragment, Taq DNA polymerase and T4 DNA ligase were purchased from Fermentas or New England Biolabs. DNA sequence was determined by an automated sequencer (ABI).

**Purification of BspRI endonuclease**

The following method was used to purify BspRI from *B. sphaericus* R as well as from arabinose-induced *E. coli* ER1821(pBAD-Bsp3) cells. Cells (30 g) were suspended in 50 ml buffer A (10 mM potassium phosphate pH 7.4, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 10% glycerol and 100 mM sodium chloride) and disrupted...
by sonication. After removing cell debris by centrifugation (18 000 r.p.m., 20 min), the cell extract was loaded onto a 150 ml phosphocellulose (Whatman P11) column equilibrated with the same buffer. Proteins were eluted from a 0.1–1.0 M NaCl gradient. Peak fractions were pooled, diluted with 10 mM Tris–HCl pH 7.4, and loaded onto a 30 ml heparin–agarose column equilibrated with buffer A. After elution with a 0.1-1.0 M NaCl gradient peak fractions were pooled and loaded directly onto a 30 ml hydroxyapatite column equilibrated with buffer A. BspRI was eluted with a 10–300 mM potassium-phosphate gradient. Peak fractions were pooled and dialysed against a buffer containing 10 mM Tris–HCl pH 7.5, 75 mM NaCl, 5% glycerol and loaded onto a 6 ml Resource-S column (Pharmacia). BspRI was eluted with a gradient 0–1 M NaCl in 20 mM potassium-phosphate (pH 7.5) buffer.

For purification of N-terminally His-tagged BspRI, ER1821(DE3 + pLysS + pET3H-BspRI) cells obtained from 11 IPTG-induced culture were resuspended in 50 ml buffer E (50 mM potassium phosphate, pH 7.4, 0.15 M NaCl, 5% glycerol, 10 mM 2-mercaptoethanol) containing 10 mM imidazole and disrupted by sonication. Cell debris was removed by centrifugation, and the supernatant was applied onto a 5 ml Ni-agarose column (His-Select C18, 75 mm column, flow rate 1000 l/min, gradient: 5–50% B in 30 min, solvent A: 0.1 % TFA/5 % acetonitrile in water; solvent B: 0.085% TFA in 95 % acetonitrile). Fractions were collected directly on the MALDI target. Post source decay (PSD) data were acquired in 10–12 segments, lowering the reflectron voltage by 25% in each step, then stitching the data together.

**Mass spectrometry analysis**

**In-gel digestion.** Gel pieces containing the R.BspRI band were cut out from SDS-polyacrylamide gels and soaked in 50% acetonitrile containing 25 mM NH₄HCO₃ to remove the Commassie stain and salts. Dihydrol and the free sulfhydryl groups were alkylated with iodoacetamide. After additional washing steps, the protein was digested in-gel with side-chain protected porcine trypsin for 4.5 h at 37°C. The resulting peptides were extracted with 2% formic acid in 50% acetonitrile.

**Peptide derivatization.** A portion of the digest was derivatized as described earlier (37). Briefly, to 5 µl of the digest 30 µl SPITC reagent (4-sulfophenyl isothiocyanate, 20 µg/µl in 25 mM NH₄HCO₃) was added and the pH of the reaction mixture was adjusted to pH 9.0 by the addition of NH₄OH. After 30 min at 55°C, the reaction was terminated with formic acid, then the peptides were purified on a C18 ZipTip according to the manufacturer's instructions.

**Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) MS.** The tryptic digest was analyzed unfraccionated prior to and after the derivatization using 2,5-dihydroxybenzoic acid as the matrix. Both mixtures were also fractionated by reversed phase HPLC (C18, 180 µm × 150 mm column, flow rate 1 µl/min, gradient: 5–40% B in 35 min, then up to 80% B in 10 min. Solvent A: 0.1 % TFA/5 % acetonitrile in water; solvent B: 0.085% TFA in 95 % acetonitrile). Fractions were collected directly on the MALDI target. Post source decay (PSD) data were acquired in 10–12 segments, lowering the reflectron voltage by 25% in each step, then stitching the data together.

**LC-MS/MS.** The underderivatized digest was also subjected to on-line LC-MS/MS analysis on an ABI QSTAR ESI-QqTOF mass spectrometer in information dependent acquisition (IDA) mode: 1 s MS acquisitions were followed by 5 s collision-induced dissociation (CID) analyses on computer-selected multiply charged ions. Nano-HPLC: C18, 75µm × 150 mm column, flow rate 300 nl/min, gradient: 5–50% B in 30 min, solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile.

Database searches were performed against the NCBI non-redundant protein database using the Protein Prospector software package (http://prospector.ucsf.edu). De novo sequencing was performed manually.

**DNA cleavage analysis**

For kinetic analysis, supercoiled pC194 plasmid DNA (2.84 nM) was incubated in 33 mM Tris–acetate pH 7.9, 10 mM Mg–acetate, 66 mM K–acetate, 0.1 mg/ml BSA (Fermentas Tango buffer) with His-tagged BspRI endonuclease (0.0054 pM) at 37°C. Aliquots were withdrawn at timed intervals and added to excess EDTA to stop digestion. Digestion of pC194 with HinP1I (New England Biolabs) and BsuRI (Fermentas) was tested using buffers recommended by the manufacturers. HinP1I and BsuRI
were used at 0.016 and 0.05 U/μl concentrations, respectively. BsuRI digestions were performed at room temperature. Plasmid isofoms were separated by electrophoresis in 1% agarose gel at low voltage (1.25 V/cm) and stained with ethidium bromide after the run. Amounts of DNA in the individual bands were determined by densitometry of the gel photograph using the GeneTools software (version 4.01, Synoptics). The kinetics of the cleavage reactions were analyzed in the framework of the reaction scheme shown in Figure 4B using the MATLAB Program Package (MathWorks Inc.). The differential equation system was solved numerically, using the Newton method (iteration step 0.001 s).

Bioinformatic tools

DNA and protein sequence similarity searches were performed using the BLAST (38,39) or the ClustalW2 (40) programs. Secondary structure predictions were performed with the Jpred 3 program (41). In all cases default settings were used.

RESULTS

Partial amino acid sequence of the BspRI REase

The gene of the BspRI MTase was originally cloned on a ~9 kb BamHI fragment. *E. coli* cells carrying pES1 (Figure 1A) expressed BspRI MTase, but did not show phage restriction and no BspRI endonuclease activity was detectable in the cell extracts (26). Cloning of longer overlapping fragments or screening a plasmid gene library for clones restricting non-modified phage failed to yield a clone expressing BspRI endonuclease.

To use an approach that is not dependent on the expression of R.BspRI in *E. coli*, the enzyme was purified from *B. sphaericus*, and a short N-terminal amino acid sequence (AQKRYGALQKVANIFINQVFTHFKG) was determined by Edman sequencing.

To obtain additional sequence information, purified BspRI was digested with trypsin and the peptides were subjected to mass spectrometry (MS) analysis as described in ‘Materials and Methods’ section. The tryptic digest was extensively analyzed by MALDI and electrospray mass spectrometry following off- or on-line HPLC fractionation. No proteins could be identified from the PSD and CID spectra by database search. To aid *de novo* sequencing, a portion of the digest was sulfonated on the N-termini of the peptides. Such derivatization usually leads to almost exclusive y-ion formation in PSD analysis. Peptide sequences determined manually from the MS/MS (PSD and/or CID) data are summarized in Table 1.

The bspRIR gene

The peptide sequences obtained by Edman degradation and MS analysis were used to design primers for PCR-amplification of a section of the bspRIR gene. Two primers (AK106 and AK108) were synthesized. Primer AK106 corresponded to amino acids Q_{10}KVANIFIN_{18}, whereas AK108 corresponded to the tetrapeptide NWQK (Table 1, Figure 1B). To reduce complexity of the AK106 pool, the neutral base inosine, which can form stable base pairs with all four bases (42), was used at positions with greater ambiguity. PCR amplification, using *B. sphaericus* DNA as template, produced an ~250 bp fragment, which was cloned in pTZ25R/T to yield pTZ-Bsp1. Sequencing of the insert revealed that the cloned fragment encodes several peptides previously detected by MS indicating that pTZ-Bsp1 carries a portion of the bspRIR gene. Unexpectedly, PCR synthesis using the same primers but pES1 plasmid DNA as template produced a similar fragment, suggesting that at least a part of the bspRIR gene was present on pES1.

To determine the approximate distance and relative orientation of the bspRIR and bspRIM genes, four PCR reactions were performed using *B. sphaericus* DNA as template and the following combinations of primers: AK108 + AK112, AK106 + AK112, AK106 + AK111 and AK108 + AK111 (Figure 1B). AK111 and AK112 were designed on basis of the previously determined sequence flanking the bspRIM gene (43). Only the AK106 + AK111 combination yielded a PCR product (~850 bp). The same result was obtained when pES1 plasmid DNA was used as template. These results showed that the genes of the BspRI R-M system are closely located in tandem arrangement with the REase gene being upstream (Figure 1).

Another conclusion following from this observation was that the entire bspRIR gene must be on the BamHI fragment cloned in pES1. This was surprising because of the lack of endonuclease activity in the clone. It seemed possible that the methods used (endonuclease assay in crude extracts and phage restriction) were not sensitive enough to detect low BspRI activity. This question was addressed by testing how inactivation of the MTase would affect viability of the clone. To obtain an m^r+ plasmid, the small SalI fragment carrying the 5'-terminal half of the bspRIM gene in pES1 (Figure 1A) was deleted. *Escherichia coli* cells carrying the resulting plasmid were perfectly viable. Taking into account the large number of BspRI sites in the *E. coli* genome and that restriction cuts producing blunt ends are likely to be highly damaging due to the absence of DNA–ligase-mediated repair (44), the viability of the m^r+ clone convincingly showed that BspRI expression from pES1 in *E. coli* was undetectable.

The nucleotide sequence of a 1028 bp segment preceding and that of a 535 bp segment following the published sequence (43) was determined (accession number: X15758). Comparison of the deduced amino acid sequence with the peptide sequences determined by Edman sequencing and MS analysis (Figures 1 and 2) unequivocally identified the ORF encoding R.BspRI. A great majority of the peptides detected in the unfractionated tryptic digest (29/34) fit to the amino acid sequence derived from the DNA sequence (Figure 2). This ORF starts with TTG at 168 and ends with TAG at 1082 defining a 304 amino acid protein. TTG is not an unusual start codon in *B. sphaericus*. Approximately 8% of the genes of another *B. sphaericus* strain (C3–41), whose sequence has recently been published (45), have TTG initiation codon (Xiaomin Hu, Nucleic Acids Research, 2010, Vol. 38, No. 20 7159
personal communication). The stop codon of the REase and the ATG start codon of the MTase are separated by 77 bp. Re-sequencing part of the MTase gene identified an error in the published sequence: the correct amino acid at position 394 is Ala rather than Thr.

The G + C content of the sequenced region (36.6%) corresponds well to the G + C content of the genome of B. sphaericus C3–41 (37.29%) (45). The region encompassing the BspRI R-M system (3178 bp) is devoid of BspRI recognition sites.

The nucleotide sequence offered an explanation for the lack of expression of R.BspRI in E. coli. First, TTG is an inefficient translational initiator in E. coli (46). Second, there is only a weak Shine–Dalgarno sequence preceding the start codon. Third, there is no typical E. coli promoter upstream of the initiation codon: a TATAAT sequence is present, but the −35 sequence is missing (Figure 1B).

To test whether the lack of expression of R.BspRI in E. coli was due to the lack of proper transcriptional and translational signals, the plasmid pBAD-Bsp3 was constructed as described in ‘Materials and Methods’ section. In pBAD-Bsp3 the bspRIR gene has an ATG start codon and the vector provides a strong Shine–Dalgarno sequence as well as an inducible E. coli promoter (araBAD). Arabinose induction led to high expression of R.BspRI indicating that the lack of expression of the native gene in E. coli was caused by the absence of proper transcriptional and translational signals.

**Sequence analysis of BspRI endonuclease**

The DNA sequence defines a protein with a calculated $M_r$ of 34 278 Da and a theoretical isoelectric point of 8.76. The results of Edman-sequencing showed that the mature protein does not contain the N-terminal formyl-Met.

A comparison with protein sequences in the GenBank database identified nine proteins (all annotated as hypothetical proteins) displaying relatively high sequence similarity to R.BspRI (Table 2). Seven of the predicted proteins are very similar in size to R.BspRI, the number of amino acids falling between 293 and 302. The genes of eight proteins are located adjacent to genes, whose predicted translation products show the signature motifs of C5-MTases, and share strong sequence similarity with M.BspRI (Table 2). The significant amino acid sequence similarity, the similar size and the co-localization with a C5-MTase gene strongly suggests that all eight proteins are restriction endonucleases (Table 2). As a result of evolutionary self-defence, R-M genes are typically characterized by the absence or paucity of the recognition site specific for the system (1). A search of the DNA regions encompassing the ORFs of the predicted REases and the counterpart C5-MTases revealed a striking scarcity of BspRI (GGCC) sites (Table 2), suggesting that these R-M systems might be functional and might have GGCC specificity, or if they are inactive, they probably have lost activity only recently. The ninth protein sharing amino acid sequence similarity with R.BspRI is a predicted protein of a strain of Streptococcus thermophilus. The protein is much shorter than R.BspRI, and the similarity extends for only the N-terminal half of the enzyme. In this case, the BLAST search did not find a gene encoding a MTase-like protein, instead it identified a gene whose translational product shows similarity with proteins playing a role in regulation of REase in some R-M systems. No significant similarity was found between R.BspRI and its isoschizomer R.BsuRI, which contrasts sharply with the very high sequence identity (67%) found between M.BspRI and its closest homolog in the database, M.BsuRI. Interestingly, the genes of the putative BspRI-like R-M system in the Bacteroides sp. 3_1_33FAA genome (Table 2) are located next to the genes of another putative R-M system annotated as ‘DNA cytosine MTase’ (ZP_06087300) and ‘type II restriction enzyme HaeIII’ (ZP_06087301), suggesting that this bacterium might contain two R-M systems of identical specificity. Whereas the amino acid sequences of the two predicted C5-MTases are highly similar, the REases of the two systems do not show significant similarity (not shown).

Based on crystallographic data and on structure predictions, most Type II REases are classified into five superfamilies (47). R.BspRI as well as the other R.BspRI-like proteins identified in the BLAST search (Table 2) appear to belong to the largest superfamily characterized by the PD-(D/E)XK motif forming the active site. This assignment is supported by results of secondary-structure predictions, which identified the βββββ core typical for PD-(D/E)XK nucleases (Supplementary Figure S1), and, in all but two proteins, the essential charged residues at characteristic positions (D in the βIb chain and EXK in the βIII chain) (48) (Figure 3). In two of the R.BspRI-like proteins (those of Lysinibacillus sphaericus C3–41 and Providencia alcalifaciens), parts of the predicted active site motif are
missing (Figure 3), suggesting that these proteins are inactive.

**BspRI cleaves the two DNA strands sequentially**

Previous results showed that R.BspRI is a monomer in free state (13). This raised questions about the mode of substrate recognition and cleavage. If BspRI is a monomer with one active site, how does it cleave the two strands of the symmetrical recognition sequence? The most likely mechanism appeared to be cutting the DNA in two steps: first making a nick, then, in a second binding event, cleaving the other strand. To test this model, a 2910 bp plasmid (pC194) having a single BspRI site was digested under steady state conditions with His-tagged BspRI, and conversion of the supercoiled plasmid DNA into the nicked and linear forms was followed as described in ‘Materials and Methods’ section. During the digestion, the nicked DNA accumulated before being converted to the linear form (Figure 4A), showing that the enzyme first cleaved just one strand. The amount of DNA in the different forms was quantitated by densitometry and the kinetics of the cleavage reactions were analyzed in the framework of the proposed reaction scheme shown in Figure 4B. The fit of the derived curves to the experimental data (Figure 4C) shows that the proposed reaction scheme is consistent with the results. Although the available experimental data did not allow determination of the full set of elementary rate constants shown in the reaction scheme, several of their pairwise respective ratios proved to be stable in the framework of the present model. For example we could establish that the rate constants of the first and second cleavage do not differ within the error of the data (~25%), suggesting independency of these processes. Accumulation of a significant amount of nicked DNA during the reaction assumes dissociation of the majority of the DNA•E complex before the final cutting process takes place. This implies that BspRI cuts one strand at a time, and the second cut occurs after binding of the enzyme to the other strand. Introducing an alternative pathway characterized by flipping of the
### Table 2.

Putative R-M systems showing the highest sequence similarity to R.BspRI

<table>
<thead>
<tr>
<th>Organism</th>
<th>Putative REase</th>
<th>Putative MTase</th>
<th>Frequency of GGCC sites&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinetobacter haemolyticus</td>
<td>M.BspRI (ZP_06729281)</td>
<td>ZP_06729282 (M.AhaBGORF3490P)</td>
<td>302/51</td>
</tr>
<tr>
<td>Gardnerella vaginalis</td>
<td>M.BspRI (ZP_03936874)</td>
<td>ZP_03936875 (M.GenETORF3500P)</td>
<td>298/44</td>
</tr>
<tr>
<td>Roseburia intestinalis</td>
<td>M.BspRI (ZP_04741871)</td>
<td>ZP_04741872 (M.RinLORF5004P)</td>
<td>293/61</td>
</tr>
<tr>
<td>Bacteroides sp. 3_1_33FAA</td>
<td>M.BspRI (ZP_06087299)</td>
<td>ZP_06087297 (M.BspFAAORF965P)</td>
<td>297/36</td>
</tr>
<tr>
<td>SD CMC 3f</td>
<td>M.BspRI (ZP_06618188)</td>
<td>ZP_06618187 (M.SDCM3FORF992P)</td>
<td>295/41</td>
</tr>
<tr>
<td>C3-41</td>
<td>M.BspRI (YP_001696252)</td>
<td>YP_001696251 (M.LspCORF497P)</td>
<td>249/32</td>
</tr>
<tr>
<td>Uncultured marine crenarchaeote</td>
<td>M.BspRI (ZP_06618188)</td>
<td>ZP_06618187 (M.SDCM3FORF992P)</td>
<td>297/36</td>
</tr>
<tr>
<td>Lysinibacillus sphaericus</td>
<td>M.BspRI (YP_001696251)</td>
<td>YP_001696252 (M.LspCORF497P)</td>
<td>295/41</td>
</tr>
<tr>
<td>Providencia alcalifaciens</td>
<td>M.BspRI (ZP_03318551)</td>
<td>ZP_03318550 (M.PalDORF1485P)</td>
<td>290/36</td>
</tr>
<tr>
<td>Uncultured marine crenarchaeote</td>
<td>M.BspRI (ZP_06618188)</td>
<td>ZP_06618187 (M.SDCM3FORF992P)</td>
<td>297/36</td>
</tr>
<tr>
<td>CNRZ1066</td>
<td>M.BspRI (YP_141100)</td>
<td>YP_141100 (M.BspRI)</td>
<td>188/9.15</td>
</tr>
</tbody>
</table>

<sup>a</sup>Identified by BLAST (blastp) search of the GenBank non-redundant protein database.

<sup>b</sup>Number of BspRI recognition sites in the DNA regions encompassing the ORFs of the predicted REases and the counterpart C5-MTases.

<sup>c</sup>One of the flanking genes encodes a putative Vsr DNA mismatch endonuclease.

<sup>d</sup>One of the flanking genes encodes a putative regulatory protein of an R-M system.

**DISCUSSION**

Amino acid sequence information obtained by sequencing the purified enzyme was used to identify the gene of the BspRI endonuclease, which was not expressed in the cloning host E. coli. The TTG start codon was silent in E. coli. The TTG start codon, the poor Shine-Dalgarno sequence and the lack of an E. coli promoter may explain why the bspRIR gene, in its native form, was silent in E. coli. Of these factors, the effect of the suboptimal initiation codon appears to be the most important. This can be concluded from the phenotype of the plasmid pTZ-Bsp5, which was an intermediate in the process of constructing pBAD-Bsp3. In pTZ-Bsp5 the original TTG initiation codon is already replaced with ATG, but even the rudimentary AGG Shine-Dalgarno sequence present in the native bspRIR gene is missing, and the orientation of the bspRIRM genes is opposite to the lac transcription starting from the pTZ57R vector. Nevertheless, E. coli cells harboring pTZ-Bsp5 produce BspRI endonuclease, suggesting that the increased expression was predominantly due to the replacement of the TTG initiation codon. Although TTG is much less efficient as start codon in E. coli than

enzyme from the cut to the uncut strand (Figure 4B, step 5) improved the fit only slightly, suggesting that double cutting may occur without formal dissociation of the monomer, but with a significantly lower probability (~20%). More complicated models, including enzyme dimerization on the DNA, did not further improve the fit, whereas the existence of sequential cutting steps in the reaction scheme remained obligatory.

Cleavage by untagged BspRI was analyzed less thoroughly, but it displayed similar kinetics as the His-tagged enzyme (not shown). Accumulation of nicked intermediate was not dependent on specific reaction conditions, it was observed in all buffers tested (Supplementary Figure S2).

The isoschizomeric BsuRI (GG/CC), another REase reported to be a monomer (15) produced a similar accumulation of the open circular form before reaching complete cleavage (Figure 4A). Thus R.BsuRI, which does not share sequence similarity with R.BspRI and is a considerably larger protein (576 versus 304 aa) (49), also appears to cut the double-stranded substrate in two consecutive reactions.
ATG (46), the observed dramatic difference in BspRI expression is surprising, and may indicate interactions between the start codon and downstream sequences of the mRNA that might be formed differently in B. sphaericus and E. coli (50).

The different start codons (TTG for R.BspRI and ATG for M.BspRI) suggest more efficient translation initiation for the MTase than for the REase, which can be important for ensuring safe protection of the host DNA against the cognate REase, especially when R-M genes enter a new cell. A regulatory mechanism based on the different efficiencies of translation signals was suggested to operate also in some other R-M systems (18,49). Such a mechanism could be an alternative to transcriptional control coordinating REase and MTase expression (51).

Putative R-M systems are typically identified in new genome sequences on basis of the signature sequence motifs characterizing DNA MTases. In most cases, due to their great variety, REase genes cannot be recognized directly (1). In this work, sequence similarity to a REase led to the discovery of putative R-M systems in the database. The sequence conservation characterizing both enzymes suggests common evolutionary origin for these R-M systems, which is puzzling for such diverse group of organisms (Table 2). Sequence analysis suggests that R.BspRI and its nine putative homologs are typical PD-(D/E)XK nuclease. Sequence conservation between the enzymes is highest in the region encompassing the predicted active site (Figure 3 and Supplementary Figure S1).

One of the goals of this work was to determine whether BspRI, which was a monomer in free state (13), acts on its target sequence as a monomer or dimerizes before the cleavage reaction takes place. For example, SalI and Eco29kI were also shown to exist predominantly as monomers (52,53), but assemble on the target sequence to act as dimers (54,55). In spite of using a wide range of conditions, no specific complex of BspRI with cognate DNA could be detected by gel electrophoretic mobility shift assay (our unpublished results). As an alternative approach, we chose to characterize the stoichiometry of the interaction using a plasmid containing a single BspRI site. Since its introduction (56) the use of plasmid DNA substrates containing a single target site has been a very productive method in studying the cleavage mechanism of REases because it allows separate kinetic analysis of single and double strand cleavage (57). This technique has been applied under steady-state as well as single-turnover conditions, e.g. (58–62). However, with few exceptions (61), the requirement for a single-site plasmid tended to restrict such studies to enzymes with longer recognition sequences. Finding a plasmid containing just a single GGCC site allowed us to analyze kinetically the first- and second-strand cleavage reactions by BspRI. The results of this analysis were consistent with the model of BspRI cleaving the two strands in two consecutive binding reactions. Qualitatively similar cleavage

\[
\begin{array}{cccccc}
\text{X} & \text{A} & \text{B} & \text{II} & \text{III} & \text{IV} \\
\text{a} & \text{b} & \text{c} & \text{d} & \text{e} & \text{f}
\end{array}
\]
kinetics were observed in sub-optimal buffers (Supplementary Figure S2) indicating that sequential cleavage of the two strands is an inherent property of the enzyme. The simplest interpretation of these data is that BspRI acts as a monomer. This assignment is supported by the similar time-course of cleavage detected for HinP1I, an enzyme shown by structural evidence to act as a monomer (Figure 4A).

Interestingly, all Type IIP REases that have been shown, by biochemical or structural evidence, to exist as monomers, recognize short sequences (4–5 bp; ‘Introduction’ section). Further work will determine whether this is just a coincidence or reflects an inherent property of the recognition mechanism.

**ACCESSION NUMBER**

X15758.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.
ACKNOWLEDGEMENTS
We thank David Dubnau and Lise Raleigh for the strains BD364(pC194) and ER1821, respectively, and Ibolya Anton for the technical assistance.

FUNDING
Hungarian Scientific Research Fund (T038343 to A.K.); National Science Foundation (DMB 8217553.000 to R.J.R.); Exxon Corporation (grant to R.J.R.); National Institutes of Health (NCRR RR015804 and P41RR01614 to the UCSF Mass Spectrometry Facility, Director A.L. Burlingame, to K.F.M., in part). Funding for open access charge: New England Biolabs.

Conflict of interest statement. None declared.

REFERENCES


