Structural mechanisms of the degenerate sequence recognition by Bse634I restriction endonuclease

Elena Manakova, Saulius Gražulis, Mindaugas Zaremba, Giedre Tamulaitiene, Dmitrij Golovenko and Virginijus Siksnys*

Department of Protein–DNA Interactions, Institute of Biotechnology Vilnius University, Graiciuno 8, LT-02241 Vilnius, Lithuania

ABSTRACT

Restriction endonuclease Bse634I recognizes and cleaves the degenerate DNA sequence 5′-R/CCGGY-3′ (R stands for A or G; Y for T or C, '/' indicates a cleavage position). Here, we report the crystal structures of the Bse634I R226A mutant complexed with cognate oligoduplexes containing ACCGGT and GCCGGC sites, respectively. In the crystal, all potential H-bond donor and acceptor atoms on the base edges of the conserved CCGG core are engaged in the interactions with Bse634I amino acid residues located on the α6 helix. In contrast, direct contacts between the protein and outer base pairs are limited to van der Waals contact between the purine nucleobase and Pro203 residue in the major groove and a single H-bond between the O2 atom of the outer pyrimidine and the side chain of the Asn73 residue in the minor groove. Structural data coupled with biochemical experiments suggest that both van der Waals interactions and indirect readout contribute to the discrimination of the degenerate base pair by Bse634I. Structure comparison between related enzymes Bse634I (R/CCGGY), NgoMIV (G/CCGGC) and SgrAI (CR/CCGGYG) reveals how different specificities are achieved within a conserved structural core.

INTRODUCTION

Type II restriction endonucleases (REases) recognize specific DNA sequences (usually 4–8 bp in length) and cleave them at specific positions within or close to their recognition site. The majority of restriction enzymes faithfully recognize a unique target site in DNA (e.g. EcoRI binds to GAATTC) and discriminate it against other sequences with an extreme specificity [discrimination factor is up to 10^{10} (1,2)]. However, a significant number of REases is able to interact with degenerate recognition sequences meaning that more than one base is permitted at a particular position of the target site. For example, the thermophilic Bse634I restriction enzyme recognizes sequence R/CCGGY, where Y (pYrimidine) stands for 'C or T', R (puRine) stands for 'A or G', and '/' indicates the cleavage position (3). In fact, among 314 recognition sites of Type II REases categorized in the REBASE (4), 83 represent degenerate sequences, including 53 containing an R:Y degenerate base pair at various positions. Molecular mechanisms that enable REase cleavage at several alternative target sites, but prevent cutting at non-cognate DNA sequences, are not yet fully understood. Deciphering of these mechanisms may reveal novel strategies for the engineering of highly specific nucleases which target unique sites in genomes but are refractory for the off-side cleavage (5).

In order to understand molecular mechanisms of the degenerate sequence recognition, we have focused on the REases that contain an R:Y base pair in their target sites. Currently, there are 32 structures of TypeII REases in the DNA-bound form and 5 of them, namely BsoBI (C/YCGRG) (6), EcoO109I (RG/GNCCY) (7), HincII (GTY/RAC) (8), SgrAI (CR/CCGGYG) (9,10) and BstYI (R/GATCY) (11) represent enzymes which recognize R:Y degenerate target sites. In most cases, the crystal structures are solved with one version of the degenerate target site and only for HincII, the structures were obtained with two cognate oligonucleotide variants providing us with the first structural mechanism of the alternative sequence recognition. Crystallographic studies supported by biochemical experiments (12,13) demonstrate that HincII specificity for the degenerate GTYRAC sequence arises from the indirect readout of the conformational preferences at the central pyrimidine–purine step (8). However, it remains to be established whether the same or different mechanisms are responsible for the degenerate sequence recognition by other restriction enzymes.

Bse634I restriction enzyme is specific for the degenerate hexanucleotide sequence R/CCGGY and belongs to a large family of evolutionarily related REases that interact with

---

closely related DNA sequences containing the conserved CCGG or CCNGG core and cut their target sites before the first C (14,15) (Table 1). The crystal structure of Bse634I in the DNA free form revealed a tetramer that had two DNA-binding sites (16). Structural comparison of Bse634I and NgoMIV–DNA (17) complex suggested that both enzymes presumably share a conserved mechanism for the recognition of the conserved CCGG core (18), however, it did not provide an answer how Bse634I discriminates the degenerate outer base pairs.

Here, we present the crystal structures of the Bse634I (R226A mutant) in the complex with the oligoduplexes representing both palindromic variants (GCCGGC and ACCGGT) of the target site (Table 2). The structures of Bse634I with the AT-1 and AT-2 oligoduplexes containing the ACCGGT site were solved at 2.3 and 2.7 Å resolution, respectively. The crystal structure of Bse634I with the GC-1 oligoduplex, containing the GCCGGC site, was solved at 2.3 Å resolution. The crystal structures reveal structural principles for an alternative target site recognition by Bse634I. The structural data coupled with the biochemical experiments suggest that both van der Waals interactions and indirect readout contribute to the discrimination of the degenerate base pair by Bse634I. Furthermore, a direct structural comparison of Bse634I (R/CCGGY), NgoMIV (G/CCGGC) and SgrAI (CR/CCGGYG) in the DNA-bound form demonstrates how different specificities are achieved within the group of related enzymes.

MATERIALS AND METHODS

Oligonucleotides

All oligonucleotides used in this study were synthesized and HPLC purified by Metabion (Martinsried, Germany). For the DNA-binding and kinetic experiments, the upper strand of each DNA duplex was labeled at the 5′-end with $[\gamma\text{-}^{33}P]ATP$ (Hartmann Analytic, Braunschweig, Germany) using T4 DNA polynucleotide kinase (ThermoFisher, Vilnius, Lithuania). Oligoduplexes were assembled by slow annealing from 95°C to room temperature in a buffer [8 mM Tris-OAc (pH 7.9 at 25°C), 16.5 mM KOAc] or in water (for crystallization).

Protein purification and crystallization

Mutants P203G and P203S were constructed by two-step megaprimer method (24). The protein coding region of the resulting plasmids was sequenced in order to ensure that the only desired mutations were introduced. The wt Bse634I and R226A (25), P203G and P203S (this study) mutants were expressed and purified as described (18). The identity of the purified P203G and P203S mutant proteins was checked by mass spectrometry. The concentration of the protein monomers was estimated spectrophotometrically using extinction coefficient $34\,280\ M^{-1}\,cm^{-1}$. R226A–DNA complex was crystallized by the sitting drop vapor diffusion technique. The protein in a buffer containing 5 mM $CaCl_2$, 20 mM Tris–HCl, 50 mM NaCl, pH 7.5 was mixed with an equimolar amount of DNA and an equal volume of a crystallization solution was added. Final concentration of the protein was in the range from 4 to 8 mg/ml. The crystals grew in a week, and after several months degraded severely. Solutions used as reservoirs for crystallization of the R226A mutant complexes with GC-1 and AT-1 oligonucleotides contained 100 mM NaOAc pH 4.25–5.5, 10 mM $CaCl_2$ or $Ca(OAc)_2$, 4–8% (w/v) of

Table 2. DNA oligoduplexes used in this study

Name	Sequence[a]	Comment
NS	5′-AGCGTAGCACTGGGCTGCTAGTC-3′ 3′-TCGCATCGTGACCCGACGATCAG-5′	Non-cognate, unmodified
AT	5′-CGCACGATCA<u>CCGG</u>TGATGCACGC-3′ 3′-GCGTGCTAGT<u>GGCC</u>ACTACGTGCG-5′	Cognate, unmodified
GC	5′-CGCACGATCG<u>CCGG</u>CGATGCACGC-3′ 3′-GCGTGCTAGC<u>GGCC</u>GCTACGTGCG-5′	Cognate, unmodified
TA	5′-CGCACGATCT<u>CCGG</u>AGATGCACGC-3′ 3′-GCGTGCTAGA<u>GGCC</u>TCTACGTGCG-5′	Mis-cognate, unmodified
CG	5′-CGCACGATCC<u>CCGG</u>GGATGCACGC-3′ 3′-GCGTGCTAGG<u>GGCC</u>CCTACGTGCG-5′	Mis-cognate, unmodified
AU	5′-CGCACGATCA<u>CCGG**U**</u>GATGCACGC-3′ 3′-GCGTGCTAG**U**<u>GGCC</u>ACTACGTGCG-5′	Cognate, modified
IC	5′-CGCACGATC**I**<u>CCGG</u>CGATGCACGC-3′ 3′-GCGTGCTAGC<u>GGCC</u>**I**CTACGTGCG-5′	Cognate, modified
UA	5′-CGCACGATC**U**<u>CCGG</u>AGATGCACGC-3′ 3′-GCGTGCTAGA<u>GGCC</u>**U**CTACGTGCG-5′	Mis-cognate, modified
CI	5′-CGCACGATCC<u>CCGG</u>**I**GATGCACGC-3′ 3′-GCGTGCTAG**I**<u>GGCC</u>CCTACGTGCG-5′	Mis-cognate, modified
GC-1	5′-TCGCG<u>CCGG</u>CGCG-3′ 3′-GCGC<u>GGCC</u>GCGCT-5′	Crystallization
AT-1	5′-TCGCG<u>ACCGG</u>TGCG-3′ 3′-GCGT<u>GGCC</u>ACGCT-5′	Crystallization
AT-2	5′-TTCGA<u>CCGG</u>TCGA-3′ 3′-AGCT<u>GGCC</u>AGCTT-5′	Crystallization

[a]The central tetranucleotide CCGG of the Bse634I recognition sequence is underlined. Modified bases are in bold face.

Table 1. Structurally characterized REases that interact with /CCGG or /CCNGG target sequences[a]

REase	Subtype	Recognition sequence	Oligomeric state	PDB ID	References
Bse634I	IIF	R/CCGGY	Tetramer	1KNV	(18)
Cfr10I	IIF	R/CCGGY	Tetramer	1CFR	(19)
SgrAI	IIF	CR/CCGGYG	dimer/tetramer	3DVO, 3MQ6, 3DPG, 3DW9	(9,10)
NgoMIV	IIF	G/CCGGC	Tetramer	1FIU, 4ABT	(17)
Ecl18kI	IIF	/CCNGG	dimer/tetramer	2FQZ	(20)
PspGI	IIP	/CCWGG	dimer	3BM3	(21)
EcoRII	IIE	/CCWGG	dimer	1NA6, 3HQG	(22,23)

[a]'/' indicates the cleavage position

PEG8000. The crystals of the R226A complex with the AT-2 oligoduplex were obtained with the reservoir solution which contained 0.1 M Bis–Tris pH 5.5, 0.5% polyvinylpyrrolidone and 16% of PEG400.

Data collection and structure determination

Crystallographic data for R226A-GC-1 were collected at EMBL Hamburg outstation DESY X12 beamline, and data sets from crystals with AT-1 and AT-2 oligoduplexes were collected at X11 DESY beamline. MOSFLM (26,27), SCALA (28) and TRUNCATE (29) were used for data processing. Initial phases were obtained by molecular re-placement using MOLREP (30) and PDB entry 1KNV as an initial model. Coordinates for B-form DNA oligonucleo-tides were generated using NAB software http://structure.usc.edu/make-na/server.html. The model was rebuilt and refined using COOT (31), CNS (32) and REFMAC (33) programs. The crystal structure of the Bse634I complex with AT-2 was refined using non-crystallographic symmetry (NCS) restraints between all eight subunits present in the asymmetric unit. Data collection and refine-ment statistics are shown in Supplementary Table S1. All molecular scale representations were prepared using MOLSCRIPT (34) and RASTER3D (35) software.

DNA-binding experiments

Binding of the modified and unmodified DNA oligoduplexes by wt and mutant Bse634I was studied by the gel mobility shift assay. The radiolabeled oligoduplexes (0.1 nM) were mixed in 20 µl samples with different concentrations of Bse634I (varying from 0.1 nM to 5 µM in terms of monomer). The reaction was carried out in a buffer containing 40 mM Tris-OAc (pH 8.3 at 25°C), 5 mM Ca(OAc)$_2$, 0.1 mg/ml BSA and 10% (v/v) glycerol. Before loading on the gel the samples were incubated for 10 min at room temperature. The protein–DNA complexes were resolved by electrophoresis in 6% non-denaturing PAGE [acrylamide/N,N′-methylenebisacrylamide 29:1 (w/w)] in 40 mM Tris–OAc running buffer containing 5 mM Ca(OAc)$_2$ at 6 V/cm for 3 h. The gels were visualized using a Cyclone Phosphor-Imager (Perkin-Elmer, Wellesley, MA, USA).

DNA-cleavage experiments

The DNA cleavage reactions were performed by mixing the radiolabeled oligoduplexes (100 nM) with Bse634I (100 nM in terms of tetramer) in the reaction buffer [10 mM Tris-HCl (pH 8.5 at 37°C), 100 mM KCl, 10 mM MgCl$_2$ and 0.1 mg/ml BSA] at 25 and 50°C (reac-tions which required incubation time >10 min were carried out under mineral oil). The samples were collected at timed intervals and quenched with a loading dye solution [95% (v/v) formamide, 25 mM EDTA, 0.01% (w/v) bromphenol blue]. Separation of the DNA hydroly-sis products was performed by denaturing PAGE: the 20% polyacrylamide gel [acrylamide/N,N′-methylenebisa-crylamide 19:1 (w/w)] in Tris–borate buffer containing 8.5 M of urea was run at 30 V/cm. The radiolabeled DNA was detected as described above. The rate constants of the DNA cleavage were determined by fitting single exponentials to the time-courses of the substrate depletion in the experiments with Gnuplot software.

RESULTS

Crystals of Bse634I R226A mutant

Our multiple attempts to obtain crystals of the wild type Bse634I–DNA complex were unsuccessful, therefore, we switched to the crystallization trials of the mutational variants. We succeeded in the co-crystallization of the Bse634I R226A mutant with three different oligoduplexes GC-1, AT-1 and AT-2 (Table 2). Bse634I restriction enzyme is arranged as a tetramer comprised of two primary dimers and requires two DNA copies for its optimal activity (18). The Arg226 residue is located at the tetramerization interface of Bse634I and makes an in-tricate network of interactions with the amino acid residues of two neighboring subunits. Its replacement by alanine disrupts the Bse634I tetramer into monomers in solution and impairs the Bse634I cleavage at low protein concentrations [(25) and M.Z., unpublished results].

Two different crystal forms of the R226A mutant in complex with DNA were obtained. Data collection and refinement statistics is presented in the Supplementary Table S1. The crystals in the P2$_1$2$_1$2 space group were produced with the cognate oligoduplexes GC-1 and AT-1 (Table 2) representing both palindromic variants of the Bse634I recognition sequence GCCGGC and ACCGGT, respectively. The crystallographic asymmetric unit in both crystals contains two protein monomers and two DNA chains. The primary dimer is generated by rotation of the protein and DNA monomer around the crystallographic 2-fold axis.

The second crystal form in the P2$_1$ space group was obtained with the AT-2 oligoduplex, which differs from the AT-1 oligoduplex by the flanking sequence (Table 2). The asymmetric unit in the crystal consists of eight protein subunits that build up two tetramers (Supplementary Figure S1A). Two oligoduplexes are bound within the DNA-binding clefts of each Bse634I tetramer. Interestingly, three additional 'out-of-site' duplexes fill up the space between the protein tetramers and form the continuous quasi-infinite DNA strands, each built by the repetition of three crystallographically independent oligoduplexes. The unpaired 5′-terminal thymine of the AT-2 oligoduplex makes H-bonds to the last A:T pair of the next DNA molecule, forming a nearly perfect Hoogsteen pair (Supple-mentary Figure S1B).

Overall complex architecture

In both crystal forms, two primary dimers are stacked back-to-back to each other and form a tetramer (Figure 1A and Supplementary Figure S1A). Each primary dimer binds one DNA duplex, therefore the tetramer binds two DNA molecules simultaneously. In all three R226A–DNA complex structures, the tetramer arrangement is very similar to that of the Bse634I in the DNA-free form (18). Thus, although R226A substitution disrupts the Bse634I tetramer into monomers in solution at low protein concentrations, in the crystal, the mutant
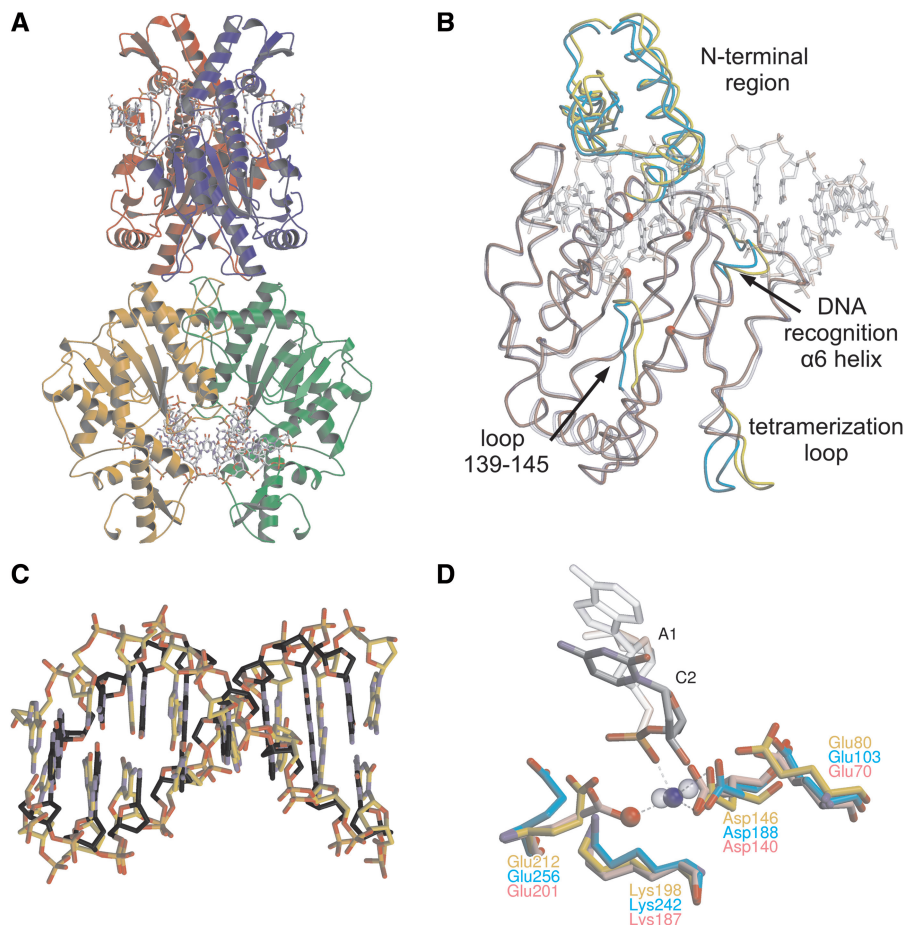
Figure 1. The crystal structure of Bse634I (R226A)–DNA complex. (**A**) General architecture of the tetramer bound to the GC-1 oligoduplex. Individual monomers are shown in different colors. Two DNA molecules bound by Bse634I are in a ball-and-stick representation. (**B**) Superposition of DNA-free Bse634I (PDB ID 1KNV) monomer A (transparent blue) with the monomer A (red) in the Bse634I–GC-1 complex. Protein chains are traced by Cα, DNA is shown in a transparent ball-and-stick representation and CPK colors. Protein chain fragments that differ in both structures are cyan and yellow for apo- and DNA bound monomers, respectively. Cα atoms of the catalytic motif (Glu80, Asp146, Lys198 and Glu212) are shown as red spheres. (**C**) Superposition of the eight central base pairs (recognition site and adjacent flanking base pairs) of the GC-1 oligoduplex complexed with Bse634I (gold) and B-form DNA of the same sequence generated by NAB software (grey). (**D**) Superposition of the catalytic center residues of the Bse634I and AT-1 complex (gold) and corresponding residues in NgoMIV–DNA (PDB ID 4ABT, pink) and SgrAI–DNA (PDB ID 3DVO, cyan) complexes. Ca^{2+} is shown as a blue sphere, a water molecule as a red sphere. Ca^{2+} in the active sites of NgoMIV and SgrAI are shown as blue transparent spheres.

protein is assembled into the tetramer similar to the wt enzyme.

Bse634I undergoes relatively small changes upon DNA binding (Figure 1B). First, the N-terminal part (residues 23–89) of the Bse634I monomer which is flexible in the apo-structure (18), moves closer to the DNA and becomes fixed by the DNA backbone contacts. The DNA binding also induces a slight shift of the protein segment (residues 139–145) in the vicinity of the active site residue Asp146 and upstream the α6 'recognition' helix (residues 202–205) which bears the residues involved in the central CCGG core recognition. The tetramerization loop (residues 258–264), which makes major contacts between the primary dimers within the tetramer, also shows slightly different conformations in the DNA bound- and apo-structures.

The DNA duplex bound to Bse634I in all three structures is in the B-form conformation (Figure 1C). Unpaired thymines at the 5′-end of the oligoduplex are highly flexible

and poorly resolved. According to CURVES (36) analysis, DNA in the complex with the R226A mutant shows a slight 11–18° bending towards the minor groove (the bending angle value varies depending on the individual structure). The most profound kink occurs at the central dinucleotide step of the site RC<u>CG</u>GY coinciding with the dyad axis of oligonucleotide.

Active center organization

The active center of Bse634I (Figure 1D) is formed by Glu80, Asp146, Lys198 and Glu212 residues which (except of Glu80) follow a sequence motif $PDX_{46-53}KX_{13}E$ characteristic for the REase family presented in Table 1. The Bse634I active site is structurally similar to those of NgoMIV (17), and SgrAI enzymes (9). Indeed, Cα atoms of the catalytic residues corresponding to Glu212 and Asp146 of Bse634I nearly coincide between all three

enzymes. However, the side chain residues of Glu212 are not well-resolved in the individual Bse634I subunits. This fact, probably, reflects an incompleteness of the active center assembly, whether due to the mutation, or due to a relatively low pH of the crystallization mixture.

Calcium ion was included in all crystallization mixtures, because it is a necessary pre-requisite for the DNA binding by Bse634I (16). Nevertheless, only in the crystal structure of Bse634I with AT-1 oligoduplex, an electron density that could be asserted to calcium was found in a close vicinity of the scissile phosphate. The octahedral coordination sphere of the metal ion is incomplete and only four ligands could be identified. The long axis of octahedron is built up by a phosphate oxygen of C2 nucleotide and a carbonyl oxygen of Leu197. Other two ligands are carboxyl oxygens of Asp146 and a water molecule (Figure 1D).

Contacts to the conserved CCGG core

Two symmetry-related α6 helices, dubbed recognition helices (18), protrude into the major groove of DNA, as it was predicted analyzing the apo-structure. The side chain atoms of the amino acid residues Arg202, Asp204 and Arg205 located at the N-terminus of the α6 helix and on the upstream loop are engaged in the sequence-specific hydrogen bond interactions with the donor and acceptor atoms of the CCGG bases. One protein subunit makes contacts in the major groove with one half-site of the symmetric CCGG sequence (Figure 2A). More specifically, the Nε and Nη2 atoms of the Arg202 and Arg205 residues make bidentate hydrogen bonds to the O6 and N7 atoms of G5 and G4 bases, respectively, while carboxyl oxygens of Asp204 bridge N4 atoms of the neighboring cytosines C2 and C3 (Figure 2B and C). The Nη1 atom of Arg205 also makes a direct hydrogen bond to the main chain oxygen atom of Ser200 and through the bridging water molecule interacts with the catalytic Lys198 residue (not shown). In the minor groove, the main chain oxygen atom of Gly69 is within a hydrogen bond distance (2.9–3.1 Å depending on the particular subunit) to the N2 amino group of the G5 base (Figure 2B).

Interactions with an outer R1:Y6 base pair

The recognition pattern of the conserved CCGG sequence by Bse634I is nearly identical to that of NgoMIV (17). However, in contrast to NgoMIV which strictly prefers the G:C base pair outside the CCGG core, Bse634I tolerates both G:C and A:T within its target site. The crystal structures of Bse634I in the complex with the oligoduplexes containing either GCCGGC (GC-1; Table 2) or ACCGGT sites (AT-1 and AT-2; Table 2) reveal the protein interaction networks with two alternative target sequences. In both structures, A1:T6 and G1:C6 base pairs are contacted by the Bse634I dimer from both minor and major groove sides (Figure 2D and E) and both monomers contribute the amino acid residues for the outer base pair recognition (Figure 2A). In the major groove, Pro203 is in the van der Waals contact distance both to A1 and G1 bases (Figure 2D and E). In the minor groove, the Nδ atom of Asn73 residue

coming from the neighboring monomer makes a hydrogen bond to the O2 atom of T6 and C6 pyrimidines.

The outer base pair R1:Y6 is distorted in the Bse634I–DNA complex. The local base pair step parameters as calculated by x3dna (37,38) are represented in the Supplementary Figure S2. The Y6 pyrimidine base is displaced from the minor groove by Asn73 and shifted towards the major groove relatively to the first flanking base pair (Supplementary Figure S2C and S2D). This results in the distortion of stacking between Y6 base and the first base outside of the recognition sequence.

Thus, Bse634I seems to combine both direct and indirect mechanisms to discriminate the RCCGGY sites from mis-cognate YCCGGR sequences. The direct interactions are limited by van der Waals contact to the purine base made by Pro203 in the major groove and a H-bond made to the complementary pyrimidine by the Asn73 in the minor groove. *In silico* mutagenesis suggests that the R1 purine replacement by the T base should introduce a steric clash of the C5 methyl group with the Pro203 residue and interfere with the Bse634I interaction with the TCCGGA site.

Bse634I binding and cleavage studies

In order to test this hypothesis, we analyzed the DNA binding and cleavage by Bse634I using a set of the oligoduplexes containing cognate (ACCGGT or GCCGGC), mis-cognate (TCCGGA or CCCGGG) and modified cognate and mis-cognate sites (Table 2). In the modified substrates, the outer T is replaced by U (uracil), which lacks the 5-methyl group in the pyrimidine ring, and G is replaced by I (inosine), which lacks 2-amino group in the purine ring. Flanking sequences in all oligoduplexes used in these experiments were identical (Table 2).

The electrophoretic mobility shift assay revealed that Bse634I binds the cognate oligoduplexes AT and GC (Figure 3A and C) at the low protein concentrations. Further increase of the Bse634I concentration gives rise to the second slowly migrating protein–DNA complex. In the case of the non-cognate oligoduplex NS (Figure 3I), only this slowly migrating complex is observed at the increased protein concentrations. Based on the DNA-binding patterns presented here and in the previous studies (16), we conclude that the complexes with the cognate DNA observed at low protein concentrations represent the specific complexes where Bse634I is bound to the recognition site while lower mobility complexes at increased protein concentrations correspond to non-specific complexes. The cognate oligoduplexes containing U instead of T (AU) and I instead of G (IC) showed the DNA-binding pattern similar to that for the specific oligoduplexes GC and AT indicating that the elimination of the 5-methyl group in thymine or 2-amino group in guanine in the outer base pair has no effect on the Bse634I binding (Figure 3E and G). The crystal structures are consistent with this observation since there are no direct or indirect contacts with methyl or amino groups in the Bse634I complexes with cognate DNA.

On the other hand, the oligoduplexes TA (Figure 3B) and CG (Figure 3D) containing mis-cognate sites are
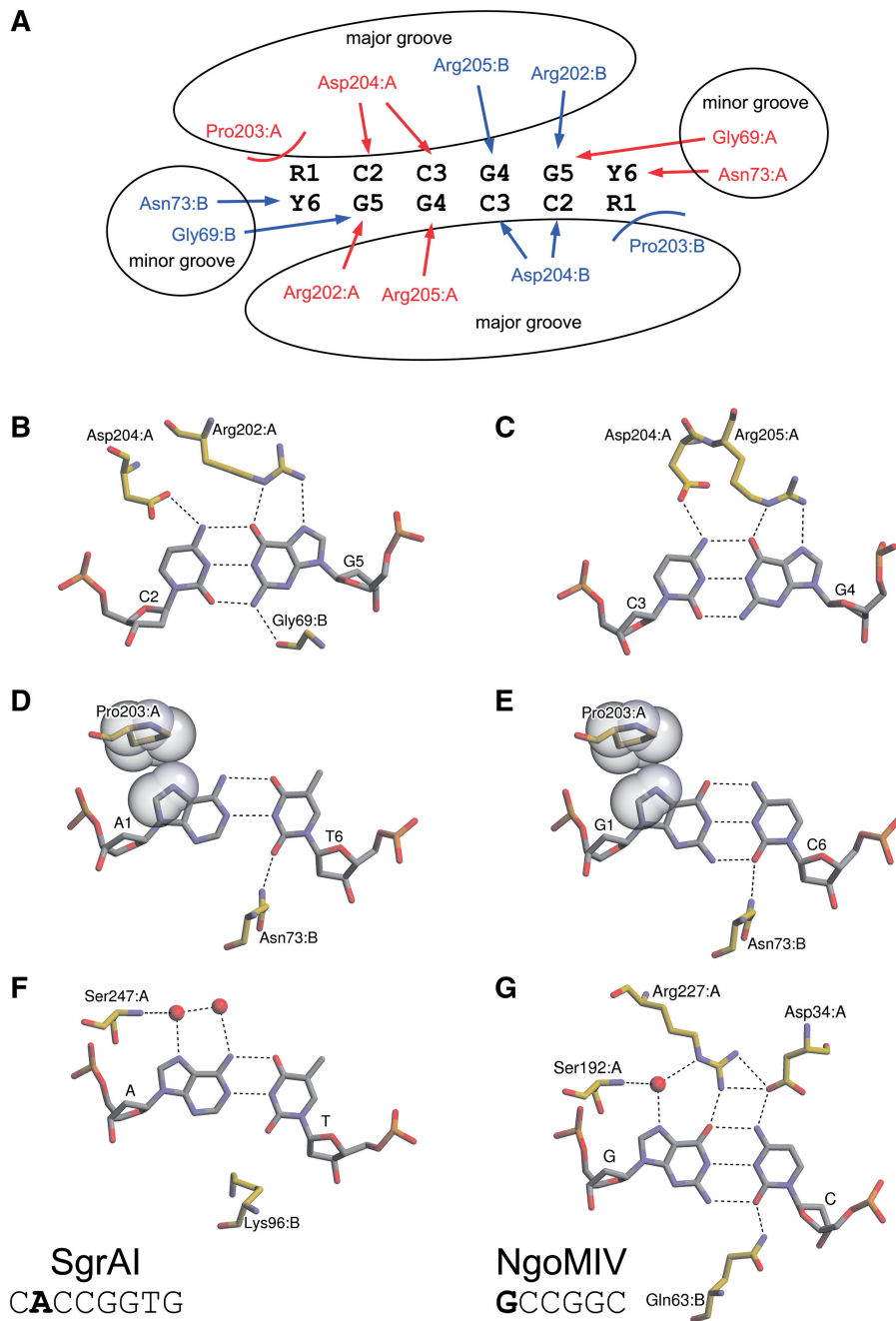
Figure 2. Specific interactions of Bse634I with the recognition sequence. (**A**) Schematic representation of the degenerate target site and interactions between the Bse634I and its target site. Residues belonging to monomers A and B are shown in red and blue, respectively. H-bonds are indicated by arrows, van der Waals interactions to purines–by curved lines. (**B**) Bse634I interactions with the C2:G5 base pair. (**C**) Bse634I interactions with the C3:G4 base pair. (**D**) Bse634I contacts to the outer A1:T6 base pair in the AT-1 complex. (**E**) Bse634I contacts to the outer G1:C6 base pair in the GC-1 complex. C8 and N7 atoms of R1 base and side chain atoms of Pro203 residue are shown in CPK representation. (**F**) SgrAI interactions with the A:T base pair in complex with the oligoduplex containing the cognate 8 bp site CACCGGTG (PDB ID 3DVO). (**G**) NgoMIV interactions with the G:C base pair in the 6 bp target site GCCGGC (PDB ID 4ABT).

bound by the protein similarly to the non-cognate DNA (Figure 3I). However, T to U replacement in the mis-cognate oligoduplex UA results in the different binding pattern (compare panels F and B in Figure 3). Indeed, the band corresponding to a specific protein–DNA complex appears in the gel, albeit at the higher protein concentrations in comparison to the cognate AT site.

The modified CI oligoduplex forms only a non-specific protein–DNA complex, similarly to the mis-cognate CG duplex (Figure 3H and D).

On the next step, we analyzed Bse634I cleavage of cognate, mis-cognate and modified oligoduplexes. The cleavage reaction was carried out under single turnover conditions using an optimal enzyme concentration that
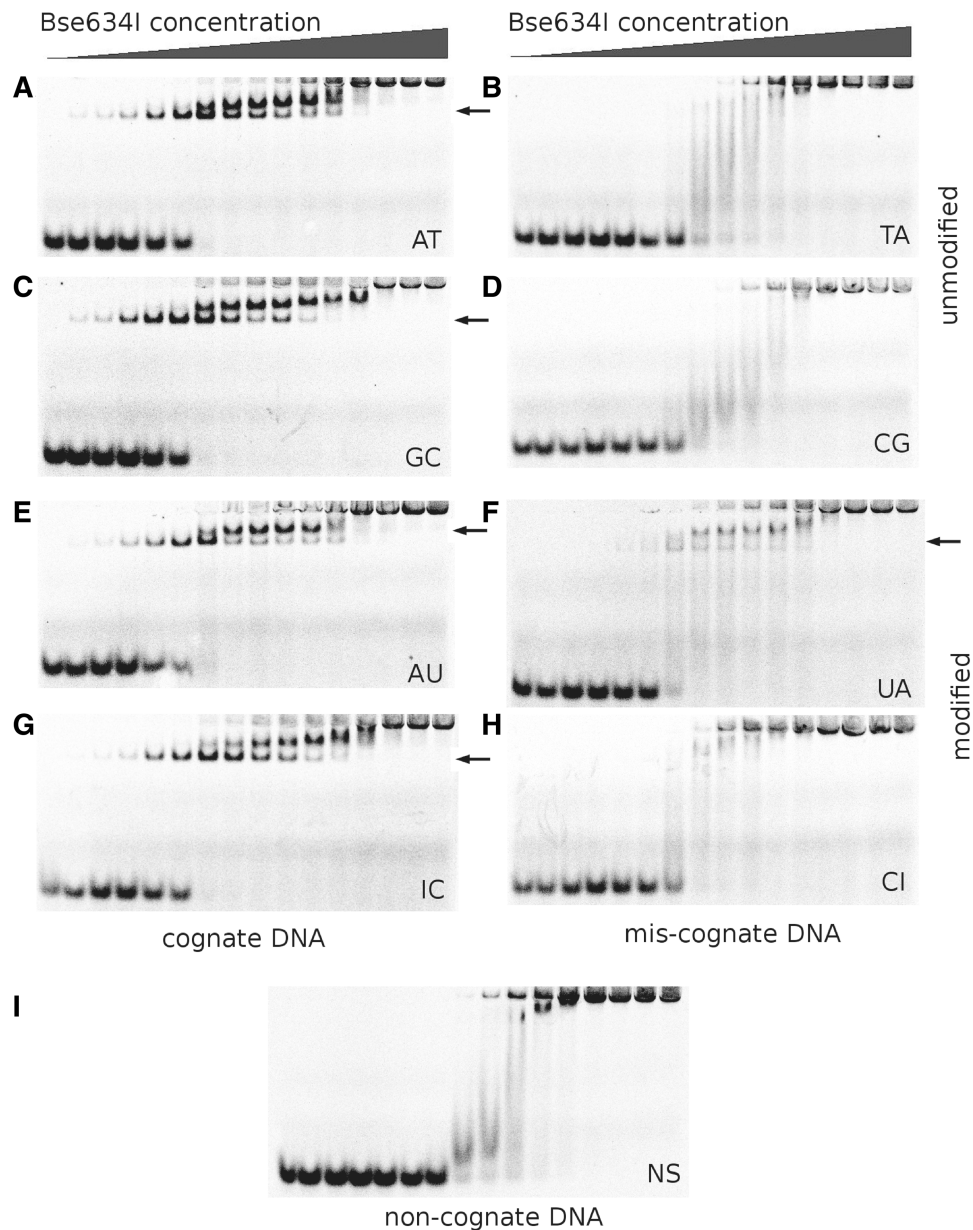
Figure 3. Binding of specific (**A**, **C**, **E** and **G**), mis-cognate (**B**, **D**, **F** and **H**), and non-cognate (**I**) DNA oligoduplexes by the wt Bse634I. Labels shown on the picture indicate oligoduplex (Table 2) used in the gel-shift experiment. I stands for Inosine and U for Uridine. Concentration of Bse634I in samples is increasing from 0.1 nM to 5 μM (in terms of monomer; see 'Materials and Methods' section). The first lane in all gels contains no protein. Specific or specific-like Bse634I–DNA complexes are designated by arrows.

yields a maximal cleavage rate. Reaction rate constants were determined as described in the 'Materials and Methods' section.

The cleavage rates for the cognate and the modified cognate oligoduplexes AT, AU, GC and IC are within the range 4–13 min^{-1} at 25°C (Table 3) indicating that modifications in the cognate AU and IC duplexes have no significant effect on the cleavage activity. However, in the case of the unmodified and the modified mis-cognate duplexes (TA, CG, UA and CI), the cleavage rates at 25°C were too slow to be reliably measured. Therefore, the Bse634I cleavage of TA, CG, UA and CI oligoduplexes was studied at 50°C taking an

advantage of the fact that Bse634I is a thermophilic enzyme (3). Under these conditions, the cleavage rate of the cognate oligoduplex was too high to be measured by the conventional techniques. On the other hand, the mis-cognate TA and CG duplexes even at 50°C were hydrolyzed very slowly (~20% of the oligoduplex was hydrolyzed after 48 h incubation). However, under these conditions, the Bse634I cleavage rate of the modified mis-cognate UA duplex was increased nearly by three orders of magnitude compared to the TA oligoduplex. The cleavage rate of the CI duplex was very close to that of the unmodified CG and TA oligonucleotides (Table 3).

Table 3. Rate constants of oligoduplex cleavage by Bse634I

Oligoduplex	k (min^{-1})					
	wt		P203S		P203G	
	25°C	50°C	25°C	50°C	25°C	50°C
AT	10 ± 1.3	–	15 ± 1.8	–	2.9 ± 0.2	–
GC	4.1 ± 0.9	–	7.8 ± 1.2	–	12 ± 4.0	–
AU	5.4 ± 1.3	–	11 ± 2.4	–	1.5 ± 0.4	–
IC	13 ± 2.9	–	17 ± 1.5	–	14 ± 2.1	–
TA	nc[a]	$5.7 ± 0.7 \times 10^{-5}$	nc	$5.7 ± 1.6 \times 10^{-4}$	nc	$1.2 ± 0.3 \times 10^{-4}$
CG	nc	$8.2 ± 0.8 \times 10^{-5}$	$1.3 ± 0.1 \times 10^{-4}$	$9.4 ± 1 \times 10^{-2}$	nc	$3.9 ± 0.5 \times 10^{-4}$
UA	nc	$6.9 ± 0.8 \times 10^{-2}$	$1.1 ± 0.1 \times 10^{-2}$	5.8 ± 1	$1.7 ± 2.6 \times 10^{-3}$	$3.8 ± 0.3 \times 10^{-1}$
CI	nc	$3.8 ± 0.9 \times 10^{-5}$	$1.4 ± 0.1 \times 10^{-4}$	$2.4 ± 0.8 \times 10^{-1}$	nc	$1.5 ± 0.9 \times 10^{-3}$
NS	nc	nc	nc	$5.4 ± 1.6 \times 10^{-5}$	nc	$3.7 ± 2 \times 10^{-5}$

[a]No cleavage products were observed after 48 h of incubation.

To further probe the importance of the van der Waals interactions in the discrimination of the degenerate R:Y base pair, we engineered P203G and P203S mutants replacing the proline with glycine (no side chain) or serine (equivalent to Pro in SgrAI) and performed DNA binding and cleavage experiments (Table 3, Supplementary Figures S3 and S4). Functional analysis revealed that the Pro203 replacement resulted in the ~10-fold increase in the oligoduplex TA cleavage rate in comparison to the wt Bse634I (Table 3). This finding would be consistent with our hypothesis that van der Waals interactions between Pro and T methyl group are important for the discrimination of the degenerate base pair by Bse634I. However, the Pro replacement by Gly or Ser seems to be not fully equivalent to the T substitution by U in DNA, since the P203G and P203S mutants display increased cleavage rate for all tested mis-cognate substrates indicating that the Pro mutations compromise Bse634I specificity.

DISCUSSION

Mechanisms of degenerate sequence recognition by Bse634I

The Bse634I restriction enzyme identified in the thermophilic *Bacillus stearothermophilus* strain cleaves two alternative target sites ACCGGT and GCCGGC (conserved sequences underlined) with nearly equal rates (Table 3). In order to understand structural mechanisms of the degenerate sequence recognition by Bse634I, we have solved the crystal structures of Bse634I in complex with three oligoduplexes representing two alternative sites.

Structural analysis revealed that the recognition of the conserved CCGG sequence is achieved by a direct readout of the unique pattern of the donor and acceptor atoms on the base edges of the CCGG tetranucleotide core (Figure 2A–C). The degenerate base pair outside the conserved core introduces an ambiguity for the base pair discrimination by a direct readout, since A:T and G:C base pairs expose a different set of donor and acceptor atoms in the major and minor grooves. Not surprisingly, there are no direct H-bonds with the G:C and A:T bases in the major groove and only the Pro203 residue makes van der Waals contact to the purine ring of the external base pair. In the minor groove, contacts to the degenerate base pair are limited to a single H-bond between the Nδ atom of Asn73 residue and O2 atom of the pyrimidine base. This raises a question how these interactions can discriminate RCCGGY sites against YCCGGR sites.

In silico experiments indicate that in the case of TCCGGA sequence a steric clash would occur between the C5 methyl group of T base and Pro203 residue. The importance of this van der Waals interaction for the R versus T discrimination is supported by biochemical data. Indeed, the modified UA oligoduplex (Table 2) which contains U instead of T and lacks 5-methyl group forms the specific Bse634I–DNA complex (Figure 3F) and is cleaved nearly three orders of magnitude faster in comparison to the unmodified mis-cognate oligoduplexes TA and CG (Table 3). Taken together, these data indicate that the steric clash between the C5 methyl group of T and Pro203 residue plays an important role in the R versus T discrimination by Bse634I. However, the Bse634I preference for R against T in the first position of the target site cannot be fully explained by a disruption of the favored van der Waals contact between Pro203 and the purine base. Cleavage rate of the UA oligoduplex containing the UCC GGA site is still much lower than of the cognate duplex.

Moreover, the steric expulsion mechanism does not provide an answer why C1:G6 base pair is not accepted by Bse634I. In the crystal structures with ACCGGT and GCCGGC oligoduplexes, the O2 atom of the outer pyrimidine Y6 makes a H-bond to the side chain Nδ2 of Asn73. Interestingly, the Gln63 residue in NgoMIV that is structurally equivalent to the Asn73 of Bse634I also makes a H-bond with O2 of the outer C6 base (Figure 2G). However, it seems unlikely that this single minor groove contact can discriminate between R and Y in the first position of the target site. Theoretically, the N3 nitrogen of A6 base should be able to make a H-bond with Asn73 similarly to Y6 pyrimidine.

Our present hypothesis is that the Bse634I selection against CCCGGG (and to some extent against the TCCGGA site) is achieved by an indirect readout mechanism due to a certain combination of distortions in DNA conformation. In the Bse634I structures with both cognate

oligoduplexes, the outer pyrimidine base is displaced towards the major groove resulting in the distortion of the stacking interaction between Y6 base and the first base outside of the recognition sequence (Supplementary Figure S2). Interestingly, SgrAI which recognizes CRCCGGYG sequence (9) displays a similar DNA distortion.

Other REases which interact with the recognition sites containing the R:Y base pair employ different mechanisms for the degenerate base pair recognition. There are five other REases with the known structure that accept the R:Y base pair in their palindromic recognition sites (summarized in the Supplementary Table S2). Three of them are solved in complex with the oligoduplex containing only one variant of the degenerate target site, while the HincII crystal structure was solved with two alternative variants of the recognition sequence. [EcoO109I crystal structure with oligoduplex containing the non-palindromic recognition sequence contains DNA in both overlapping orientations (7).] Crystal structure analysis indicates that the REases achieve specificity for their degenerate target sites by decreasing the number of direct contacts which often become limited to the N7 position of purines or due to the ambivalence of the water-mediated hydrogen bonding (Supplementary Table S2).

How different specificities are generated within a group of related restriction enzymes

It has been reported before that apo-Bse634I (RCCGGY) shows close structural resemblance to the 6 bp cutter NgoMIV (GCCGGC) and 8 bp cutter SgrAI (CRCCGGYG) (conserved CCGG sequence is underlined) (14,15). All three enzymes cut their target sites before the CCGG core and share similar architecture of the active site (Figure 1D). The Ca^{2+} ion present in the Bse634I–AT-1 complex occupies exactly the same position as the first Ca^{2+} ion in the SgrAI structure with cognate DNA (9). In Bse634I, Ca^{2+} ion is coordinated by the oxygen atoms of the scissile phosphate, the side chain oxygen atoms of Asp146 and the main chain carbonyl of the Leu197 preceding the catalytic Lys198. The water molecule which serves as a fourth Ca^{2+} ion ligand in Bse634I overlaps well with the water molecule coordinated by Ca^{2+} ion in SgrAI and NgoMIV structures. Furthermore, Glu80 of Bse634I coincides with Glu103 of SgrAI, which coordinates the second Ca^{2+} ion (Figure 1D). The electron density which could be assigned to the second metal ion, however, is absent in the Bse634I structure.

Structural comparison of the enzymes in the DNA-bound forms confirmed that Bse634I, SgrAI and NgoMIV share a conserved mechanism for the recognition of the CCGG core within their target sites. The key contacts to the CCGG bases (Figure 2B and C) are made by the conserved stretch of RXDR residues located at the N-terminal tip of the recognition helices and are identical between Bse634I, NgoMIV and SgrAI.

In contrast to NgoMIV, which is strictly specific for the GCCGGC site, Bse634I recognizes two alternative palindromic sequences ACCGGT and GCCGGC. Outside the CCGG core, NgoMIV makes an intricate network of H-bonds to the outer G:C base pair (Figure 2G) both in the major and minor grooves. These interactions unambiguously specify the G:C base pair. Unlike NgoMIV, Bse634I makes only two contacts to the base pair outside the CCGG core, namely the hydrogen bond by Asn73 to the O2 of pyrimidine in the minor groove and the van der Waals contact by Pro203 to the purine from the major groove (Figure 2A, D and E). Thus, the direct contacts made by Bse634I to the outer base pair are restrained in comparison to NgoMIV presumably due to an ambiguity introduced by the degenerate base pair.

The target site of Bse634I (RCCGGY) overlaps with the central part of SgrAI site CRCCGGYG (underlined). SgrAI makes no direct contacts with the degenerate R:Y base pair (9) except of the putative water-mediated H-bond between Ser247 and N7 atom of purine (Figure 2F). The Cα atom of Lys96 occupies a structurally equivalent position of Asn73 in Bse634I, however, H-bond to the O2 oxygen of thymine in the minor groove is missing. Thus, the direct contacts to the degenerate base made by SgrAI are restrained similarly to Bse634I. Moreover, in the SgrAI–DNA complex, the DNA is distorted similarly to Bse634I (Supplementary Figure S2). The side chain of Lys96 contacts the outer G (underlined) of the SgrAI recognition sequence CRCCGGYG from the minor groove and unstacks this base from the Y that just precedes it (9). Thus, it seems that both in SgrAI and Bse634I an indirect mechanism plays an important role in the discrimination of the degenerate base pair, however, it remains to be determined to what extent the specific structural features of the RY helical step (39–42) contribute to the discrimination of the degenerate base pair.

Structural comparison suggests that the N-terminal domains of Bse634I and SgrAI exhibit the plasticity which may contribute to the recognition of the 8 bp sequence by SgrAI. The Arg31 residue of SgrAI which makes bidentate H-bonds to the external G residue (underlined) in the CACCGGTG sequence is located in the loop (residues 19–39) at the N-terminal domain. The N-terminal loop (residues 10–26) of Bse634I is structurally equivalent to the loop (residues 19–39) in SgrAI. However, the N-terminal loop in all Bse634I structures is rather flexible and shows a different orientation or is disordered in the individual monomers. The structurally equivalent loop in SgrAI is much more rigid presumably due to the larger number of contacts to the DNA backbone and the protein body. Furthermore, the structural equivalent of the Arg31 residue is missing in the similar loop of Bse634I which is too short to reach the major groove of the DNA substrate outside of the hexanucleotide recognition sequence. Further structural studies of Kpn2I (T/CCGGA), AgeI (A/CCGGT) and BsaWI (W/CCGGW, where W stands for A or T) restriction enzymes ongoing in the lab should contribute to our understanding how Nature engineered different specificities within the conserved structural fold.

ACCESSION NUMBERS

Coordinates and structure factors of Bse634I (R226A) complexes with DNA are deposited under PDB ID 3V1Z (GC-1), 3V20 (AT-1) and 3V21 (AT-2).

REFERENCES

1. Engler,L.E., Welch,K.K. and Jen-Jacobson,L. (1997) Specific binding by EcoRV endonuclease to its DNA recognition site GA TATC. *J. Mol. Biol.*, **269**, 82–101.
2. Jen-Jacobson,L., Engler,L.E. and Jacobson,L.A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure*, **8**, 1015–1023.
3. Repin,V.E., Lebedev,L.R., Puchkova,L., Serov,G.D., Tereschenko,T., Chizikov,V.E. and Andreeva,I. (1995) New restriction endonucleases from thermophilic soil bacteria. *Gene*, **157**, 321–322.
4. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
5. Silva,G., Poirot,L., Galetto,R., Smith,J., Montoya,G., Duchateau,P. and Pâques,F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.*, **11**, 11–27.
6. van der Woerd,M.J., Pelletier,J.J., Xu,S. and Friedman,A.M. (2001) Restriction enzyme BsoBI-DNA complex: a tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure*, **9**, 133–144.
7. Hashimoto,H., Shimizu,T., Imasaki,T., Kato,M., Shichijo,N., Kita,K. and Sato,M. (2005) Crystal structures of type II restriction endonuclease EcoO109I and its complex with cognate DNA. *J. Biol. Chem.*, **280**, 5605–5610.
8. Horton,N.C., Dorner,L.F. and Perona,J.J. (2002) Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. *Nat. Struct. Biol.*, **9**, 42–47.
9. Dunten,P.W., Little,E.J., Gregory,M.T., Manohar,V.M., Dalton,M., Hough,D., Bitinaite,J. and Horton,N.C. (2008) The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.*, **36**, 5405–5416.
10. Park,C.K., Joshi,H.K., Agrawal,A., Ghare,M.I., Little,E.J., Dunten,P.W., Bitinaite,J. and Horton,N.C. (2010) Domain swapping in allosteric modulation of DNA specificity. *PLoS Biol.*, **8**, e1000554.
11. Townson,S.A., Samuelson,J.C., Xu,S.Y. and Aggarwal,A.K. (2005) Implications for switching restriction enzyme specificities from the structure of BstYI bound to a BglII DNA sequence. *Structure*, **13**, 791–801.
12. Joshi,H.K., Etzkorn,C., Chatwell,L., Bitinaite,J. and Horton,N.C. (2006) Alteration of sequence specificity of the type II restriction endonuclease HincII through an indirect readout mechanism. *J. Biol. Chem.*, **281**, 23852–23869.
13. Babic,A.C., Little,E.J., Manohar,V.M., Bitinaite,J. and Horton,N.C. (2008) DNA distortion and specificity in a sequence-specific endonuclease. *J. Mol. Biol.*, **383**, 186–204.
14. Tamulaitis,G., Solonin,A.S. and Siksnys,V. (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett.*, **518**, 17–22.
15. Pingoud,V., Conzelmann,C., Kinzebach,S., Sudina,A., Metelev,V., Kubareva,E., Bujnicki,J.M., Lurz,R., Lüder,G., Xu,S.-Y. *et al.* (2003) PspGI, a type II restriction endonuclease from the extreme thermophile Pyrococcus sp.: structural and functional studies to investigate an evolutionary relationship with several mesophilic restriction enzymes. *J. Mol. Biol.*, **329**, 913–929.
16. Zaremba,M., Sasnauskas,G., Urbanke,C. and Siksnys,V. (2005) Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: oligomeric structure-stability-function correlations. *J. Mol. Biol.*, **348**, 459–478.
17. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
18. Grazulis,S., Deibert,M., Rimseliene,R., Skirgaila,R., Sasnauskas,G., Lagunavicius,A., Repin,V., Urbanke,C., Huber,R. and Siksnys,V. (2002) Crystal structure of the Bse634I restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res.*, **30**, 876–885.
19. Bozic,D., Grazulis,S., Siksnys,V. and Huber,R. (1996) Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 A resolution. *J. Mol. Biol.*, **255**, 176–186.
20. 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.
21. Szczepanowski,R.H., Carpenter,M.A., Czapinska,H., Zaremba,M., Tamulaitis,G., Siksnys,V., Bhagwat,A.S. and Bochtler,M. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res.*, **36**, 6109–6117.
22. Golovenko,D., Manakova,E., Tamulaitiene,G., Grazulis,S. and Siksnys,V. (2009) Structural mechanisms for the 5'-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII. *Nucleic Acids Res.*, **37**, 6613–6624.
23. Zhou,X.E., Wang,Y., Reuter,M., Mücke,M., Krüger,D.H., Meehan,E.J. and Chen,L. (2004) Crystal structure of type IIE restriction endonuclease EcoRII reveals an autoinhibition mechanism by a novel effector-binding fold. *J. Mol. Biol.*, **335**, 307–319.
24. Barik,S. (1995) Site-directed mutagenesis by double polymerase chain reaction. *Mol. Biotechnol.*, **3**, 1–7.
25. Zaremba,M., Sasnauskas,G., Urbanke,C. and Siksnys,V. (2006) Allosteric communication network in the tetrameric restriction endonuclease Bse634I. *J. Mol. Biol.*, **363**, 800–812.
26. Leslie,A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
27. Leslie,A.G.W. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 48–57.
28. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
29. CCP4 (Collaborative Computational Project, N. 4). (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.*, **50**, 760–763.
30. Vagin,A.A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**, 1022–1025.

31. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

32. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54(Pt 5)**, 905–921.

33. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

34. Kraulis,P. (1991) Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.

35. Merritt,E.A. and Murphy,M.E. (1994) Raster3D Version 2.0 A program for photorealistic molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 869–873.

36. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.

37. Lu,X.-J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.

38. Lu,X.-J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.

39. Ulyanov,N.B. and Zhurkin,V.B. (1984) Sequence-dependent anisotropic flexibility of B-DNA. A conformational study. *J. Biomol. Struct. Dyn.*, **2**, 361–385.

40. Suzuki,M., Yagi,N. and Finch,J.T. (1996) Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.*, **379**, 148–152.

41. Olson,W.K., Gorin,A.A., Lu,X.J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.

42. Fujii,S., Kono,H., Takenaka,S., Go,N. and Sarai,A. (2007) Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.*, **35**, 6063–6074.