Alternative arrangements of catalytic residues at the active sites of restriction enzymes

Gintautas Tamulaitis[a], Alexander S. Solonin[b], Virginijus Siksnys[a,*]

[a] *Institute of Biotechnology, Graiciuno 8, Vilnius 2028, Lithuania*
[b] *Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russia*

Received 26 February 2002; revised 20 March 2002; accepted 20 March 2002

First published online 3 April 2002

Edited by Gunnar von Heijne

Abstract A catalytic sequence motif PDX_{10-30}(E/D)XK is found in many restriction enzymes. On the basis of sequence similarities and mapping of the conserved residues to the crystal structure of *Ngo*MIV we suggest that residues D160, K182, R186, R188 and E195 contribute to the catalytic/DNA binding site of the *Ecl*18kI restriction endonuclease. Mutational analysis confirms the functional significance of the conserved residues of *Ecl*18kI. Therefore, we conclude that the active site motif $^{159}VDX_{21}KX_{12}E$ of *Ecl*18kI differs from the canonical PDX_{10-30}(E/D)XK motif characteristic for most of the restriction enzymes. Moreover, we propose that two subfamilies of endonucleases *Ecl*18kI/*Psp*GI/*Eco*RII and *Cfr*10I/*Bse*634I/*Ngo*-MIV, specific, respectively, for CCNGG/CCWGG and RCCGGY/GCCGGC sites, share conserved active site architecture and DNA binding elements.

Key words: Restriction endonuclease *Ecl*18kI; Active site; Mutagenesis

1. Introduction

Type II restriction endonucleases recognize specific DNA sequences, typically 4–8 bp in length, and cleave phosphodiester bonds in the presence of Mg^{2+}, within or close to these sequences [1]. Almost 3000 type II restriction endonucleases specific for more than 200 different nucleotide sequences have been identified by screening various bacterial species [2]. Therefore, restriction endonucleases probably represent the largest family of functionally related enzymes, however the structural and mechanistical divergence within the family is still poorly understood. Protein sequence comparisons in general reveal no or very weak similarities raising the question of whether restriction endonucleases evolved independently and elaborated individual ways to interact with their target nucleotide sequences.

Structural studies however demonstrate that despite of the lack of sequence homologies, type II restriction endonucleases possess a conserved structural core composed of a five-stranded β-sheet and two flanking α-helices, and similar architecture of the active site [3,4]. Pioneering structural studies of *Eco*RI [5] and *Eco*RV [6] revealed that sequence motifs PDX_{19}EAK and PDX_{15}DIK, respectively, correspond to the structurally invariant parts of their active sites. The PDX_{10-30}(D/E)XK motif was later identified in many sequences of type II restriction enzymes, suggesting active site similarities [1]. Structural studies of *Cfr*10I [7], *Ngo*MIV [8] and *Bse*634I [9] supported by mutational analysis [10], however, revealed that conserved spatial arrangement of the active residues might be achieved through a different linear sequence motif. Indeed, the sequence motif $PDX_{46-55}KX_{13}E$ corresponds to the conserved active sites of *Cfr*10I, *Bse*634I, and *Ngo*MIV but differs from the canonical PDX_{10-30}(E/D)XK motif characteristic for most of the restriction enzymes.

The *Ecl*18kI restriction endonuclease recognizes the pentanucleotide sequence CCNGG and cleaves it before the first C base [11]. The protein sequence of *Ecl*18kI [11] is ~99% identical to those of isoschizomeric restriction enzymes, *Sso*II [12], *Sen*PI [13] and *Sty*D4I [14]. Noteworthy, *Sty*D4I/*Sso*II showed significant sequence similarity to *Psp*GI/*Eco*RII endonucleases specific for overlapping CCWGG sequence [15]. The catalytic/DNA binding site residues for the *Ecl*18kI restriction endonuclease however have not yet been identified.

In this report we propose that residues D160, K187, R186, R188 and E195 contribute to the catalytic/DNA binding site of *Ecl*18kI and provide mutational and biochemical data supporting our hypothesis.

2. Materials and methods

2.1. Strains, plasmids and DNA

Recombinant wild-type (wt) and mutant genes of *Ecl*18kI were cloned and expressed in *Escherichia coli* JM109 strain (F′traD36-proA$^+$B$^+$lacI^{q}Δ(lacZ)M15/e14$^-$(McrA$^-$)Δ(lac-proAB)endA1gyrA96(-Nal^r)thi-1hsdR17(r_k^- m_k^+)glnV44relA1recA1). Strain JM109 containing pUC129 (Ap^r) harboring the *Ecl*18kI restriction–modification system was used for expression of wt *Ecl*18k*IR* protein. Strain JM109 containing plasmid pHSG415ts (Cm^r) bearing the *ecl18kIM* gene was used as a host for transformation of pQE30 plasmid (Ap^r) containing the *ecl18kIR* gene fused to the N-terminal $(His)_6$-tag encoding sequence.

Competent host cells were prepared and transformed with heterologous DNA using the $CaCl_2$ method. The plasmid DNA was isolated by the alkaline lysis procedure and purified using phenol:chloroform extraction [16].

pQE30 plasmid was used to engineer mutations by the two-step 'megaprimer' method [17]. The following mispaired primers were used: 5′-GGC ATA ACA AG**A** **GC**T ACA AGT TTA C-3′ (D160A); 5′-CAG CGT TGT **TGC** AGC GGA AAT C-3′ (K182A); 5′-GCC ATC GCT C**CG** **C**CA GCG TTG-3′ (R186A); 5′-CTT GCC A**TG** **C**CT CCC TCA GC-3′ (R188A); 5′-CTA TTA AC**C** **GC**T TCA GGT AC-3′ (E195A).

*Corresponding author. Fax: (370)-2-602116.
E-mail address: siksnys@ibt.lt (V. Siksnys).

Abbreviations: aa, amino acids; Ap and Ap^r, ampicillin and ampicillin resistance; BSA, bovine serum albumin; Cm and Cm^r, chloramphenicol and chloramphenicol resistance; EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane; wt, wild-type

0014-5793/02/$22.00
PII: S0014-5793(02)02621-2

Both strands of the entire region between *Bam*HI/*Hinc*II or *Bgl*II/*Hinc*II sites were sequenced to confirm that the designed amino acid (aa) change was the only change in the DNA sequence.

2.2. Protein purification

$(His)_6$-tagged wt and mutant *Ecl*18kI variants were purified to approximately 90% purity on Ni-NTA (Novagen) and AH-Sepharose (Amersham Biosciences) columns. A 250 ml overnight culture of *E. coli* JM109 bearing the pQE30 expression plasmid was incubated at 37°C for 3.5 h after isopropyl-1-thio-β-D-galactopyranoside induction at a final 0.4 mM concentration. Cells were harvested by centrifugation at $2500\times g$ for 15 min and stored frozen. The frozen cell pellet was suspended in 10 ml of binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM tris(hydroxymethyl)aminomethane (Tris)–HCl (pH 7.9 at 25°C)), sonicated 4×30 s and centrifuged for 30 min at $18\,000\times g$ to remove cell debris. The supernatant was applied to a 2 ml Ni-NTA column equilibrated with reconstitution buffer and *Ecl*18kI endonuclease was eluted using a linear 0.005–1 M imidazole gradient. The fractions containing *Ecl*18kI enzyme were pooled, dialyzed against buffer containing 10 mM potassium phosphate (pH 7.4 at 25°C), 1 mM ethylenediaminetetraacetic acid (EDTA), 7 mM 2-mercaptoethanol and 100 mM NaCl, and applied to a 2 ml AH-Sepharose column. The column was eluted using a linear 0.1–1.0 M NaCl gradient, fractions containing wt or mutant *Ecl*18kI were pooled and dialyzed against 10 mM potassium phosphate (pH 7.4 at 25°C), 100 mM KCl, 2 mM dithiothreitol, 0.1 mM EDTA, 50% glycerol and stored at −20°C.

Wt *Ecl*18k*IR* protein was expressed in *E. coli* JM109 cells containing the pUC129 plasmid and purified to homogeneity through subsequent chromatography on Heparin-Sepharose, Q-Sepharose, Blue-Sepharose (Amersham Biosciences) and hydroxylapatite (Calbiochem) columns. The protein was >95% homogeneous by sodium dodecylsulfate–PAGE analysis. The protein concentration was determined by measuring the OD_{280} and using an extinction coefficient of 38 830 M^{-1} cm^{-1} for a monomer.

2.3. DNA cleavage assay

Cleavage activity of *Ecl*18kI was monitored under single turnover conditions using supercoiled $dam^{-}dcm^{-}$ pUC19 plasmid (Fermentas) containing 12 *Ecl*18kI recognition sites. Wt *Ecl*18kI or mutant proteins at 75 nM (in terms of dimer) concentration were mixed with 2.5 nM of pUC19 in the assay buffer containing 33 mM Tris-acetate (pH 7.9 at 15°C), 66 mM potassium acetate, 0.1 mg/ml bovine serum albumin (BSA) and the reaction was initiated by adding $(CH_3COO)_2Mg$ to give a final concentration of 10 mM. Aliquots were removed after fixed time intervals (the shortest accessible reaction time was 5 s) and mixed with loading dye solution containing EDTA. Cleavage experiments were performed at 15°C to make the reaction rates slow enough to collect samples manually in the case of wt *Ecl*18kI. The DNA samples were separated in agarose gel and the amount of supercoiled form of plasmid DNA was determined by densitometric analysis of ethidium bromide-stained gels [18]. The decline in the concentration of supercoiled substrate with time was fitted to a single exponential to give an apparent first-order reaction rate constant k_1. The increase of protein concentration above 75 nM had no effect on the cleavage rate indicating that cleavage experiments were performed at saturating enzyme concentrations both in the case of wt and mutant proteins.

2.4. Gel mobility shift assay

The 16 bp double-stranded oligodeoxynucleotide

5′ – CGTAG**CCTGG**GGCTAG – 3′

3′ – GCATC**GGACC**CCGATC – 5′

containing the recognition sequence of *Ecl*18kI (shown in bold) and the 16 bp double-stranded oligodeoxynucleotide

5′ – AGCGTAGCACTGGGCT – 3′

3′ – TCGCATCGTGACCCGA – 5′

lacking the recognition sequence were used in binding experiments as specific and non-specific DNA, respectively. For each duplex, one strand was 5′-labeled with [γ-^{33}P]ATP using T4 polynucleotide kinase (Fermentas). DNA (final concentration of 1 nM) was incubated with varying amounts of *Ecl*18kI for 10 min at room temperature in the binding buffer (33 mM Tris-acetate (pH 7.9 at 37°C), 66 mM potassium acetate, 0.1 mg/ml BSA, 10% glycerol and 10 mM calcium acetate) and samples were applied to 6% polyacrylamide gel (29:1, acrylamide/bisacrylamide). Electrophoresis was run for 2 h at 5 V/cm in 40 mM Tris-acetate (pH 8.0 at 25°C) buffer containing 10 mM of calcium acetate. After electrophoresis gels were dried and analyzed on a Cyclone (Packard Instrument Co.) phosphoimager. The following equation was used to fit the quantitated data to obtain K_d values:

$$y = 100\{K_d + x + s_0 - [(K_d + x + s_0)^2 - 4xs_0]^{0.5}\}/(2s_0)$$

where, y – percentage of complexed DNA, x – initial protein concentration and s_0 – initial DNA concentration, K_d – dissociation constant.

2.5. Sequence alignment

FASTA pairwise similarity search between the protein sequence of *Ecl*18kI and other restriction enzymes revealed weak local similarities to *Cfr*10I endonuclease. Using *Multalin* [19] sequence alignment server (http://prodes.toulouse.inra.fr/multalin/multalin.html) we were able to extend the initial pairwise alignment between *Ecl*18kI and *Cfr*10I to a larger set of restriction endonucleases specific for CCNGG/CCWGG or RCCGGY nucleotides. In the initial alignment generated automatically by *Multalin* structurally conserved PD residues at the active sites of *Cfr*10I, *Bse*634I and *Ngo*MIV, however, were misaligned (probably due to the different number of aa in spacings between the conserved regions) and therefore were realigned manually. Analysis of a highly diverged group of nucleases [20] identified conserved ^{159}VD residues in *Ecl*18kI, ^{137}PD residues in *Psp*GI and ^{298}PD residues in *Eco*RII sequences, respectively, as homologues to the PD residues from the $PDX_{55}KX_{13}E$ active site motif of *Cfr*10I. Therefore, we manually aligned those residues of *Ecl*18kI/*Psp*GI/*Eco*RII with structurally conserved PD residues of *Cfr*10I, *Bse*634I and *Ngo*MIV generating the final form of Fig. 1a.

3. Results

3.1. Characterization of Ecl18kIR protein

Sequencing of plasmid pUC129 bearing the *ecl18kIR* gene revealed three nucleotide mutations in comparison to the *ecl18kIR* sequence Y16897 present in GenBank. Two mutations at positions corresponding to V228 (GTA vs GTT) and K272 (AAA vs AAG) were silent, however the third one resulted in R277Q replacement (CAG vs CCG). PSI-BLAST search reveals that our *Ecl*18kI variant is 100% identical to the *Sty*D4I protein sequence except that *Sty*D4I has an additional N-terminal extension of 13 aa [14]. Sequence analysis of the *ecl18kIR* gene reveals two putative starts of translation corresponding to Met1 and Met14 residues of the deduced protein sequence. Translation from the Met1 codon would result in the *Ecl*18k*IR* protein of the same length as *Sty*D4I. N-terminal sequencing of the purified *Ecl*18k*IR* protein (data not shown), however, indicated that translation is started from the codon of Met14 residue yielding a protein of 305 aa residues in length. The mass-spec analysis of purified *Ecl*18k*IR* protein yielded a molecular weight value of 35 892 Da (Luksa, V., unpublished results) that is very close to the value of 35 897 Da calculated for the 305 aa protein. Analytical ultracentrifugation revealed that in solution *Ecl*18k*IR* is a dimer (Urbanke, C., unpublished results).

3.2. Identification of the putative catalytic/DNA binding site of Ecl18kI

Pairwise FASTA search (data not shown) using *Ecl*18kI protein sequence as a query detects weak local similarities to the *Cfr*10I endonuclease (recognition site R/CCGGY). Further analysis revealed (Fig. 1a) that the similarity region identified between *Ecl*18kI and *Cfr*10I is conserved within a larger family of restriction endonucleases specific for CCNGG/

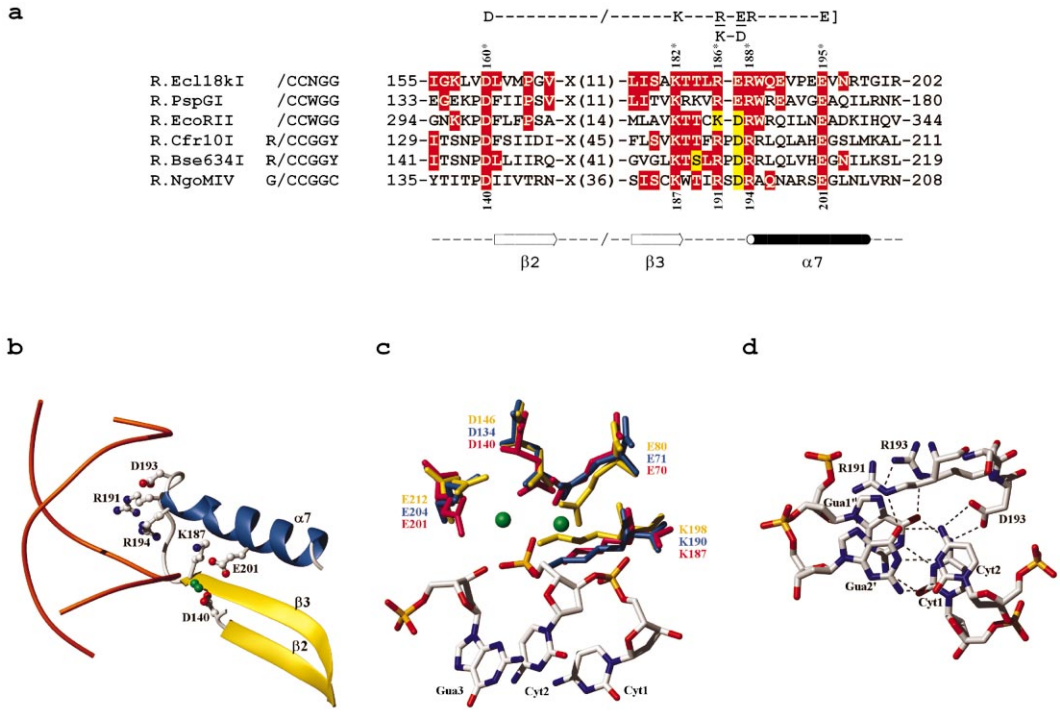

Fig. 1. Sequence conservation and mapping of the conserved residues to the crystal structure of *Ngo*MIV. a: Conserved sequence region between *Ecl*18kI, *Psp*GI, *Eco*RII, *Cfr*10I, *Bse*634I and *Ngo*MIV. Residues that are identical or similar to the corresponding *Ecl*18kI residues are boxed and shown in red and yellow, respectively. Consensus residues are shown above the *Ecl*18kI sequence. The mutated *Ecl*18kI residues are noted with an asterisk and numbered. Numbers below the *Ngo*MIV sequence indicate catalytic/DNA binding residues according to the crystal structure [8]. Secondary structure assignment of *Ngo*MIV, as determined by X-ray crystallography, is shown below the sequence; b: structural elements of *Ngo*MIV involved in the building of the catalytic/metal binding site and discrimination of the central CG nucleotides; c: superposition of the active site residues of *Ngo*MIV (yellow), *Cfr*10I (magenta) and *Bse*634I (red). Mg^{2+} ions are shown as green spheres; d: recognition of the central C:G nucleotides by *Ngo*MIV.

CCWGG and RCCGGY/GCCGGC nucleotides. In fact, only few residues are strictly conserved (Fig. 1a) and the alignment by itself lacks stringent statistical significance to be trusted without further information. Fortunately, it has a strong structural support: crystal structures for *Cfr*10I [7], *Ngo*MIV [8] and *Bse*634I [9] restriction endonucleases have been solved giving a possibility to map the conserved residues to the known three-dimensional structures. Since the *Ngo*MIV structure is solved both in the presence of DNA and metal cofactors, it was used as a reference in the further analysis.

The most conserved region corresponding to the consensus K- - - (R/K)-(D/E)R- - - - - -E (Fig. 1a) spans across the β3-strand and α7-helix of *Ngo*MIV. According to the crystal structure [8] residues located on the β3-strand and α7-helix of *Ngo*MIV form a part of the catalytic/metal binding site and provide an interface for the recognition of the central C:G nucleotides (Fig. 1b).

In the post-reactive *Ngo*MIV–DNA complex [8] two Mg^{2+} ions are located at the catalytic/metal binding site that is conserved within *Ngo*MIV, *Cfr*10I and *Bse*634I restriction endonucleases (Fig. 1c). Therefore, it is very likely that *Cfr*10I and *Bse*634I also coordinate two metal ions at their active sites. Structure-based sequence alignment between *Cfr*10I, *Bse*634I and *Ngo*MIV [9] reveals that the first metal ion is chelated by acidic residues from the sequence motif $PDX_{46-55}KX_{13}E$ that differs, however, from the canonical active site motif $PDX_{10-30}(E/D)XK$ characteristic for most of the restriction enzymes [1].

The acidic residues D140 and E201, chelating first Mg^{2+} ion in *Ngo*MIV, and lysine K187 presumably involved in catalysis are conserved within the *Ecl*18kI/*Psp*GI/*Eco*RII subfamily (Fig. 1a). According to the crystal structure [8] aspartate D140 from the $PDX_{46}KX_{13}E$ motif and the E70 residue located on the α3-helix constitute a second metal ion binding site in *Ngo*MIV (Fig. 1c). While the D140 residue of *Ngo*MIV is conserved within the *Ecl*18kI/*Psp*GI/*Eco*RII subfamily (Fig. 1a), the location of the second acidic residue is less obvious. It was suggested [20] that the E125 residue of *Ecl*18kI might be a structural equivalent of the E71 residue of *Cfr*10I, however it does not show up in our version of the local alignment (Fig. 1a). Thus, on the basis of alignment and structural mapping of conserved residues (Fig. 1) we propose that the sequence motif $(P/V)DX_{21-24}KX_{12}E$ conserved within *Ecl*18kI, *Psp*GI and *Eco*RII corresponds to the active site that is structurally similar to that of *Cfr*10I/*Bse*634I/*Ngo*MIV enzymes.

Structural elements of *Ngo*MIV involved in catalysis and sequence discrimination are intertwined [8]. Indeed, residues involved in the recognition of the central CCGG nucleotides are located on a short stretch of aa ^{191}RSDR interspaced between catalytic/metal binding site residues K187 and E201, respectively (Fig. 1b). Arginines R191 and R194 donate bidentate hydrogen bonds to the neighboring guanines while carboxylate oxygens of D193 bridge exocyclic amino groups of the neighboring cytosines (Fig. 1d). Thus, three residues R191, D193 and R194 of a single subunit unambiguously specify inner and middle C:G nucleotides within the first

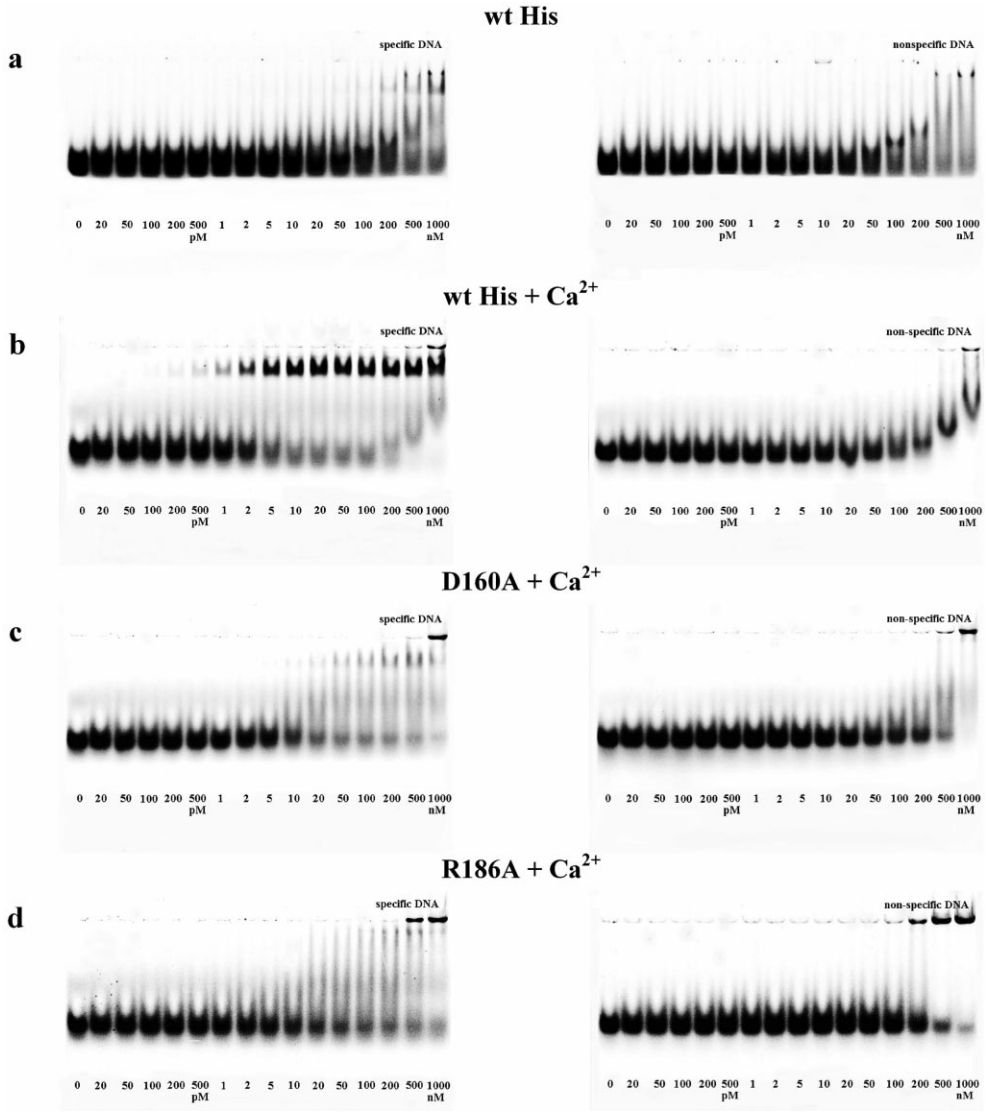

Fig. 2. Gel shift analysis of DNA binding by *Ecl*18kI and mutant proteins. Increasing amounts of protein (indicated below the corresponding lanes) are incubated with specific and non-specific 16 bp oligodeoxynucleotides and protein–DNA complexes analyzed in polyacrylamide gel under non-denaturing conditions. All gels except (a) were run in the presence of 10 mM $Ca(OAc)_2$.

half-site of the *Ngo*MIV recognition site GCCGGC. According to the alignment (Fig. 1a) residues R191 and R194 involved in the recognition of the C:G nucleotides by *Ngo*MIV are conserved within the *Ecl*18kI/*Psp*GI/*Eco*RII subfamily specific for CCNGG/CCWGG sequences. Thus, it is tempting to speculate that R186 and R188 residues of *Ecl*18kI contribute to the discrimination of C:G nucleotides at the recognition site.

3.3. DNA cleavage activity of wt Ecl18kI and mutant proteins in vitro

In order to demonstrate the functional importance of the conserved putative catalytic (D160, K182 and E195) and sequence discriminating (R186 and R188) residues of *Ecl*18kI, we generated a set of Ala mutants and studied DNA cleavage and binding properties of mutant proteins. To simplify and facilitate purification procedures both wt and mutant proteins were expressed as N-terminal $(His)_6$-tagged proteins and purified to approximately 90% purity. Endonucleolytic activity of *Ecl*18kI and mutant proteins was assessed using *dam*$^-$*dcm*$^-$ pUC19 plasmid DNA since the recognition sites for *Ecl*18kI may overlap with sites for *dcm* methylation. Using the plasmid cleavage assay the nicking activity of mutant proteins can be monitored alongside a double-strand cleavage activity. The

Table 1
DNA cleavage and binding activities of *Ecl*18kI and mutant proteins

*Ecl*18kI	Activity (%)	K_d (nM)
Wt	100	3.5 ± 0.2
D160A	0	>50[b]
K182A	0	0.3 ± 0.1
E195A	4[a]	0.2 ± 0.1
R186A	0.2[a]	>100[b]
R188A	0	>100[b]

[a]An apparent first-order rate constant for the pUC19 cleavage under the single turnover reaction conditions was used for calculation of the residual activity in %. The first-order rate constant for pUC19 cleavage by wt *Ecl*18kI equals 0.07 ± 0.02 s^{-1}.
[b]We were unable to determine the K_d value due to the weak smeared bands.

pUC19 cleavage experiments revealed that D160A, K182A and R188A mutants completely lost both double-strand cleavage and nicking activities (Table 1). Two other mutants (E195A and R186A) exhibited significantly decreased specific DNA nicking activity (Table 1) and impaired double-strand cleavage ability (complete DNA cleavage was not observed). Of note is that mutation of the homologous residues in the isoschizomeric *Sso*II restriction enzyme revealed similar phenotypes (Pingoud, A., unpublished results).

3.4. Effect of mutations on the DNA binding properties of Ecl18kI and mutant proteins

In order to determine whether the mutations affected the ability of *Ecl*18kI to bind substrate, the DNA binding of wt and mutant proteins was analyzed using the gel shift assay. Two synthetic 16 bp DNA duplexes, one containing the recognition site 5′-CCTGG and another lacking an *Ecl*18kI site, were used in binding experiments as specific and non-specific DNA, respectively. Preliminary binding experiments demonstrated that DNA binding by *Ecl*18kI is weak (Fig. 2a) and sensitive to the experimental conditions (pH and salt composition of binding and gel running buffers). The inclusion of Ca^{2+} ions into the reaction mixtures and gel running buffers, however, resulted in clear and reproducible shifts (Fig. 2b) confirming previous reports that Ca^{2+} ions that do not support DNA cleavage stimulate DNA binding specificity of restriction enzymes [21–23]. Therefore, gel shift experiments (except of those presented in Fig. 2a) were performed in the presence of 10 mM of $Ca(OAc)_2$ both in the binding and gel running buffer. Gel shift analysis demonstrates (Fig. 2b) that wt *Ecl*18kI binds specific DNA with high affinity ($K_d \sim 3$ nM) while non-specific DNA binding remains low. The *Ecl*18kI mutants with ablated DNA cleavage ability (Table 1) fall into two classes with respect to DNA binding. The DNA binding affinity of K182A and E195A mutants of *Ecl*18kI is increased relative to the wt enzyme (Table 1). In contrast, the DNA binding ability of D160A, R186A and R188A (Fig. 2c,d and Table 1) is highly compromised. The DNA binding pattern of R188A mutant is similar to that of R186A and therefore it is not shown. Circular dichroism and analytical ultracentrifugation experiments (Urbanke, C., unpublished results) revealed no changes of secondary or quaternary structure of alanine mutants relative to the wt enzyme. Therefore, it is unlikely that perturbation of protein conformation or oligomerization state leads to the loss of function.

4. Discussion

Protein sequence alignment (Fig. 1a) suggests that restriction endonucleases *Ecl*18kI/*Psp*GI/*Eco*RII and *Cfr*10I/*Bse*634I/*Ngo*MIV, specific, respectively, for CCNGG/CCWGG and RCCGGY/GCCGGC sites, share conserved active site/DNA binding elements.

Indeed, residues D140, K187 and E201 of *Ngo*MIV involved in metal ion binding/catalysis are conserved in *Ecl*18kI (Fig. 1a). Alanine substitution of homologous K182 and E195 residues of *Ecl*18kI resulted in the loss or significant decrease of endonucleolytic activity while DNA binding was increased in comparison to the wt enzyme (Table 1). Previous mutational studies of the active site residues of restriction enzymes, e.g. *Eco*RI [24], *Mun*I [25] and *Bam*HI [26], revealed that Ala replacements do not interfere with DNA binding or even enhance binding affinity in comparison to the wt enzymes. Thus, observed DNA binding and cleavage phenotypes of K182A and E195A mutants are in principle consistent with a proposed active site function of K182 and E195 residues in *Ecl*18kI.

Gel shift analysis, however, indicates that alanine replacement of putative metal chelating residue D160 decreases DNA binding affinity (Fig. 2c). The DNA binding pattern of D160A mutant exhibits characteristic smeared bands similar to that of wt enzyme in the absence of Ca^{2+} ions (Fig. 2a). Aspartate 160 of *Ecl*18kI is homologous to the D140 and D134 residues of *Ngo*MIV and *Cfr*10I, respectively. In *Ngo*MIV and *Cfr*10I carboxylates of D140 and D134 are involved in bridging interactions between two metal ions at the active site (Fig. 1c). The detailed role of both metal ions is not yet clear, however it is possible that metal ions contribute both to catalysis and DNA binding. Indeed, Ca^{2+} ions are necessary for specific DNA binding by *Cfr*10I [23] and other restriction enzymes [21,22,27]. Noteworthy, a D134A mutation compromised DNA binding affinity of *Cfr*10I in the presence of Ca^{2+} ions (Zaremba, M., Skirgaila, R. and V.S., unpublished results). It is conceivable that D134A replacement in *Cfr*10I abolishes coordination of Ca^{2+} ion that is necessary for the stabilization of enzyme-specific DNA complex. Similarly, if the D160 residue of *Ecl*18kI chelates Ca^{2+} ion that is required for the specific DNA binding, it is not surprising that the D160A mutation compromises DNA binding ability.

Thus, the mutational data presented here are in principle consistent with the hypothesis that D160, K182 and E195 residues of *Ecl*18kI constitute a catalytic/metal binding site similar to that of *Ngo*MIV. The active signature motif $^{159}VDX_{21}KX_{12}E$ of *Ecl*18kI is similar to the active site motif $PDX_{46-55}KX_{13}E$ characteristic for *Cfr*10I/*Bse*634I/*Ngo*MIV [9] but differs from the canonical $PDX_{10-30}(E/D)XK$ motif of most of the restriction enzymes. Conservation of homologous residues within the *Ecl*18kI/*Eco*RII/*Psp*GI subfamily (Fig. 1) suggests a similar catalytic/metal binding function of the conserved residues. Attempts to locate the active site of *Eco*RII searching for canonical $PDX_{10-30}(E/D)XK$ motifs, revealed few candidate residues [28], however only E96 residue located in the N-terminal part of *Eco*RII appeared to be important for catalysis. Thus, the sequence location of the catalytic/metal binding residues of *Eco*RII is not yet unambiguously predicted. We suggest that sequence motifs $^{298}PDX_{24}KX_{12}E$ and $^{137}PDX_{21}KX_{12}E$ specify a putative catalytic/metal binding site of *Eco*RII and *Psp*GI, respectively. Thus, the active site motif $PDX_{46-55}KX_{13}E$ characteristic for *Cfr*10I/*Bse*634I/*Ngo*MIV might be more common than it was initially thought.

Structural studies of the *Ngo*MIV endonuclease demonstrate that residues involved in catalysis and sequence discrimination are intertwined. Indeed, R191, D193 and R194 residues of *Ngo*MIV involved in major groove recognition of inner and middle C:G base pairs are located just upstream of α7-helix harboring the metal chelating residue E201 (Fig. 1b). According to Fig. 1a, R186 and R188 residues of *Ecl*18kI are homologous to the R191 and R194 residues of *Ngo*MIV. Alanine substitution of R186 and R188 residues of *Ecl*18kI resulted in the complete loss of endonucleolytic activity accompanied by highly compromised DNA binding (Table 1). Thus, residues R186 and R188 of *Ecl*18kI might be involved in sequence discriminating contacts. Sequence-specific contacts of restriction enzymes, however, are often redundant

and comprised by amino acid residues located on the different structural elements [1]. Therefore we cannot rule out the possibility that other residues located on the structural elements not conserved between *Ecl*18kI and *Ngo*MIV also contribute to the recognition interface of *Ecl*18kI.

Thus, our mutational studies and biochemical studies are consistent with the hypothesis that *Ecl*18kI specific for CCNGG sequence shares a conserved structural mechanism of the C:G base pair recognition with restriction enzymes *Cfr*10I/*Bse*634I/*Ngo*MIV specific for the RCCGGY/GCCGGC sequence. In principle, the recognition sequence of *Ecl*18kI / CCNGG ('/' indicates cleavage site) might be considered as an interrupted version of CCGG tetranucleotide found at the recognition sites of *Cfr*10I/*Bse*634I/*Ngo*MIV. How could this single nucleotide insertion at the recognition site be accommodated by *Ecl*18kI while maintaining the conserved catalytic/metal binding site and sequence recognition elements employed by the *Cfr*10I/*Bse*634I/*Ngo*MIV enzymes? The crystal structure of *Bgl*I [29] specific for the interrupted GCCNNN/NNGGC sequence demonstrates that it maintains recognition/catalytic functions similar to those of *Eco*RV restriction enzyme specific for the continuous GATATC site, however the interface between two *Bgl*I subunits is dramatically changed to accommodate a 5 bp insertion. It is plausible that a similar mechanism allows *Ecl*18kI to adopt an extra N base in the CCNGG sequence while keeping the same structural elements for the sequence recognition/catalysis employed by *Ngo*MIV.

5. Note added in proof

While this article was in press a paper by Pingoud et al. (JBC Online) was published suggesting that isoschizomeric *Sso*II restriction endonuclease exhibits a similar arrangement of the active site.

Acknowledgements: We thank A. Pingoud for communicating unpublished results, M.M. Den'mukhametov and M.V. Zakharova for providing plasmids, C. Urbanke for analytical ultracentrifugation, V. Luksa for the mass-spec analysis, G. Sasnauskas for the help with the figures and R. Skirgaila for discussions. This work was supported by HHMI International Research Scholarship Grant #55000336 to V.S. and Lithuanian Science Foundation doctoral fellowship to G.T.

References

[1] Pingoud, A. and Jeltsch, A. (2001) Nucleic Acids Res. 29, 3705–3727.
[2] Roberts, R.J. and Macelis, D. (2001) Nucleic Acids Res. 29, 268–269.
[3] Venclovas, C., Timinskas, A. and Siksnys, V. (1994) Proteins 20, 279–282.
[4] Aggarwal, A.K. (1995) Curr. Opin. Struct. Biol. 5, 11–19.
[5] Kim, Y., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M. (1990) Science 249, 1307–1309.
[6] Winkler, F.K. (1992) Curr. Opin. Struct. Biol. 2, 93–99.
[7] Bozic, D., Grazulis, S., Siksnys, V. and Huber, R. (1996) J. Mol. Biol. 255, 176–186.
[8] Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V. and Huber, R. (2000) Nat. Struct. Biol. 7, 792–799.
[9] Grazulis, S., Rimseliene, R., Skirgaila, R., Sasnauskas, G., Lagunavicius, A., Repin, V., Urbanke, C., Huber, R. and Siksnys, V. (2002) Nucleic Acids Res. 30, 876–885.
[10] Skirgaila, R., Grazulis, S., Bozic, D., Huber, R. and Siksnys, V. (1998) J. Mol. Biol. 279, 473–481.
[11] Den'mukhametov, M.M., Zakharova, M.V., Kravets, A.N., Pertsev, A.V., Sineva, E.V., Repik, A.V., Beletskaya, I.V., Gromova, E.S. and Solonin, A.S. (1997) Mol. Biol. (Mosk.) 31, 831–838.
[12] Karyagina, A.S., Lunin, V.G., Degtyarenko, K.N., Uvarov, V.Y. and Nikolskaya, I.I. (1993) Gene 124, 13–19.
[13] Ibanez, M., Alvarez, I., Rodriguez-Pena, J.M. and Rotger, R. (1997) Gene 196, 145–158.
[14] Miyahara, M., Ishiwata, N. and Yoshida, Y. (1997) Biol. Pharm. Bull. 20, 201–203.
[15] Morgan, R., Xiao, J.-P. and Xu, S.-Y. (1998) Appl. Environ. Microbiol. 64, 3669–3673.
[16] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[17] Barik, S. (1995) Mol. Biotechnol. 3, 1–7.
[18] Sasnauskas, G., Jeltsch, A., Pingoud, A. and Siksnys, V. (1999) Biochemistry 38, 4028–4036.
[19] Corpet, F. (1988) Nucleic Acids Res. 16, 10881–10890.
[20] Bujnicki, J.M. and Rychlewski, L. (2001) J. Mol. Microbiol. Biotechnol. 3, 69–72.
[21] Vipond, I.B. and Halford, S.E. (1995) Biochemistry 34, 1113–1119.
[22] Lagunavicius, A., Grazulis, S., Balciunaite, E., Vainius, D. and Siksnys, V. (1997) Biochemistry 36, 11093–11099.
[23] Skirgaila, R. and Siksnys, V. (1998) Biol. Chem. 379, 595–598.
[24] Grabowski, G., Jeltsch, A., Wolfes, H., Maass, G. and Alves, J. (1995) Gene 157, 113–118.
[25] Lagunavicius, A. and Siksnys, V. (1997) Biochemistry 36, 11086–11092.
[26] Engler, L.E., Sapienza, P., Dorner, L.F., Kucera, R., Schildkraut, I. and Jen-Jacobson, L. (2001) J. Mol. Biol. 307, 619–636.
[27] Nastri, H.G., Evans, P.D., Walker, I.H. and Riggs, P.D. (1997) J. Biol. Chem. 272, 25761–25767.
[28] Reuter, M., Schneider-Mergener, J., Kupper, D., Meisel, A., Mackeldanz, P., Krueger, D.H. and Schroeder, C. (1999) J. Biol. Chem. 274, 5213–5221.
[29] Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. and Phillips, S.E.V. (1998) EMBO J. 17, 5466–5476.