VILNIUS UNIVERSITY
INSTITUTE OF BIOTECHNOLOGY

Gintautas Tamulaitis

STRUCTURAL AND FUNCTIONAL STUDIES OF THE Ecl18kI AND EcoRII RESTRICTION ENZYMES SPECIFIC FOR INTERRUPTED PALINDROMIC SITES

Doctoral dissertation

Physical sciences, biochemistry (04 P), proteins, enzymology (P310)

Vilnius, 2008
The work presented in this doctoral dissertation has been carried out at the
Institute of Biotechnology from 2000 to 2008.

Supervisor:

prof. dr. Virginijus Siksnys (Institute of Biotechnology)

(physical sciences, biochemistry (04 P), proteins, enzymology (P310))
CONTENTS

LIST OF ABBREVIATIONS 4
INTRODUCTION 5
LITERATURE OVERVIEW 9
I. The diversity of restriction endonucleases 9
II. The similarity of restriction endonucleases 12
III. Similarities between restriction enzymes recognizing similar DNA sequences 16
  1. Bse634I, Cfr10I and NgoMIV family 16
  2. EcoRI and MunI family 22
  3. BglI and SfiI family 27
  4. BcnI, Mval and MutH family 32
  5. BamHI, BglII and BstYI family 37
  6. HpaII and MspI family 42
  7. Ecl18kI, EcoRII and PspGI family 46
  8. Concluding remarks: structural and functional similarities between restriction enzymes 46
IV. Restriction endonuclease EcoRII 46
  1. The resistance of rarely occurring EcoRII specific recognition sites against the cleavage by EcoRII 47
  2. EcoRII can be activated to cleave refractory DNA recognition sites 47
  3. Preliminary EcoRII mechanistic models 48
    3.1. Kruger’s model 48
    3.2. Petrauskene’s model 49
    3.3. Karpova’s model 50
  4. EcoRII-DNA interactions: two sites versus one site 51
  5. EcoRII reactions on modified oligoduplexes 52
  6. Putative DNA binding sites of EcoRII 54
  7. Domain organization of EcoRII 55
  8. EcoRII dimer interface disruption by a single amino acid replacement 56
  9. Concluding remarks: EcoRII – archetypal Type IIE restriction enzyme 57
MATERIALS AND METHODS 58
V. Materials 58
  1. Chemicals 58
  2. Enzymes 58
  3. E. coli strains 58
  4. DNA 59
  5. Oligonucleotides 59
  6. Buffers 61
VI. Methods 62
  1. Electrophoresis 62
    1.1. Denaturing (SDS) polyacrylamide gel electrophoresis of proteins 62
    1.2. Non-denaturing electrophoresis through agarose 63
    1.3. Non-denaturing polyacrylamide gel electrophoresis 63
    1.4. Denaturing polyacrylamide gel electrophoresis 64
2. Expression and purification of Ecl18kI
   2.1. Ecl18kI plasmids and strains
   2.2. Site-directed mutagenesis of Ecl18kI
   2.3. Purification of Ecl18kI
3. Expression and purification of EcoRII
   3.1. EcoRII plasmids and strains
   3.2. Site-directed mutagenesis of EcoRII
   3.3. Purification of EcoRII
4. Expression and purification of PspGI and MvaI
   4.1. Expression and purification of PspGI
   4.2. Expression and purification of MvaI
5. Characterization of R.Ecl18kI protein
6. Ecl18kI mass-spectrometry analysis
7. Protein sequence analysis
8. Circular Dichroism spectroscopy
9. Analytical ultracentrifugation
10. EcoRII limited proteolysis
11. Gel mobility shift assay
   11.1. DNA binding of wt Ecl18kI and mutant proteins
   11.2. DNA binding of wt EcoRII and mutant proteins
   11.3. DNA binding of wt, W61Y and W61A Ecl18kI, EcoRII-C, PspGI and MvaI
   11.4. Gel mobility shift assay
   11.5. Dissociation constant $K_d$ calculation
12. DNA cleavage assay
   12.1. DNA cleavage assay of Ecl18kI mutant proteins
   12.2. DNA cleavage assay of EcoRII mutant proteins
   12.3. DNA cleavage assay of wt, W61Y and W61A Ecl18kI, EcoRII-C and PspGI
13. Structure solution of Ecl18kI-DNA complex
14. Fluorescence spectroscopy of Ecl18kI, EcoRII-C and PspGI
15. Gel filtration of EcoRII
16. EcoRII, EcoRII-C and PspGI reactions with supercoiled plasmid substrates
   16.1. Construction and purification of plasmid substrates
   16.2. DNA cleavage reactions
   16.3. Data analysis
17. EcoRII-DNA AFM analysis
   17.1. Construction and purification of DNA fragments for AFM
   17.2. EcoRII-DNA AFM analysis
18. Structure analysis

RESULTS AND DISCUSSION

VII. Protein sequence similarities within the restriction endonuclease subfamily specific for CCGG/CCNGG sites
VIII. Active site/C:G binding residues in the restriction enzymes Ecl18kI and EcoRII
   1. Identification of the putative catalytic/C:G binding site of Ecl18kI by biochemical analysis
      1.1. Mutational analysis of the possible catalytic/metal binding site of Ecl18kI
1.2. Mutational analysis of possible DNA binding site of Ecl18kI
2. Confirmation of the putative Ecl18kI catalytic/DNA binding site by Ecl18kI-DNA X-ray structure
3. Identification of the putative catalytic/DNA binding site of EcoRII by biochemical analysis
   3.1. Mutational analysis of possible catalytic site at EcoRII
   3.2. Mutational analysis of the possible DNA binding site of EcoRII
4. Confirmation of the EcoRII putative catalytic/DNA binding site by EcoRII X-ray apo-structure
5. Family of restriction enzymes bearing CCGG tetranucleotide in their recognition sequence

IX. How Ecl18kI, EcoRII-C and PspGI achieve recognition of its target site?
1. Restriction endonuclease Ecl18kI flips the central base pair within its recognition site
2. 2-aminopurine steady-state fluorescence analysis of Ecl18kI, EcoRII-C and PspGI enzymes
   2.1. Probes for Ecl18kI, EcoRII-C and PspGI triggered nucleotide flipping
   2.2. Fluorescence of Ecl18kI-DNA in the presence of Ca$^{2+}$ ions
   2.3. Fluorescence of Ecl18kI-DNA in the absence of Ca$^{2+}$ ions
   2.4. Fluorescence of the complex of Ecl18kI pocket mutant-DNA
   2.5. Fluorescence of the EcoRII-C and PspGI complexes with DNA
3. Nucleotide flipping phenomenon

X. Detailed mechanism of EcoRII interaction with DNA
1. Stoichiometry of the EcoRII-DNA complex: number of possible DNA binding interfaces in EcoRII
   1.1. Gel shift analysis of EcoRII-N-DNA complex
   1.2. Gel filtration analysis of the EcoRII-N-DNA complex
   1.3. Gel filtration analysis of the EcoRII-C-DNA complex
2. EcoRII kinetics studies to verify three DNA binding model
   2.1. Construction of plasmid DNA substrates for kinetics studies
   2.2. EcoRII cleavage of plasmid DNA at [E]≥[S]
   2.3. EcoRII cleavage reactions in trans
   2.4. EcoRII cleavage of plasmid DNA at [E]<[S]
   2.5. Plasmid DNA cleavage by EcoRII-C
   2.6. Plasmid DNA cleavage by PspGI
   2.7. Summary of EcoRII reactions on the 1-, 2- and 3-site plasmids
   2.8. EcoRII and NaeI: Type IIE restriction enzymes that follow different mechanisms
3. Direct visualization of the EcoRII-DNA triple synaptic complex by AFM
   3.1. Construction of DNA fragments for EcoRII AFM studies
   3.2. Visualization and characterization of EcoRII-PCR3 complexes
   3.3. Characterization of EcoRII-PCR1 complexes
   3.4. EcoRII monomer in the complexes
   3.5. Three-site synaptic complexes formed by other proteins

XI. Future prospects
CONCLUSIONS
ACKNOWLEDGEMENT
LIST OF PUBLICATIONS
REFERENCES
LIST OF ABBREVIATIONS

2-AP 2-aminopurine  
aa amino acid(s)  
AdoMet S-adenosyl-L-methionine  
AFM atomic force microscopy  
ATP adenosine-5’-triphosphate  
a. u. arbitrary unit(s)  
bp base pair(s)  
BSA bovine serum albumin  
CPK CPK coloring conventionally depicts hydrogen as white, carbon as black or grey, nitrogen as blue, oxygen as red, phosphorus as orange, sulfur as yellow, magnesium as green.  
ds double stranded (DNA)  
DTT 1,4-dithiothreitol  
EDTA ethylenediaminetetraacetic acid  
FLL linear DNA form with one double strand brake  
HPLC high performance liquid chromatography  
IPTG isopropyl-β-D-thiogalactopyranoside  
kb kilo base pairs  
L2 plasmid DNA cleavaged into two fragments  
L3 plasmid DNA cleavaged into three fragments  
λexc excitation wavelength  
λem emission wavelength  
M. methyltransferase  
N A, T, G or C; adenine, thymine, guanine or cytosine  
nt nucleotid(s)  
PAA polyacrylamide  
PDB ID Protein Data Bank identification code  
OC open-circular  
PAAG polyacrylamide gel  
PAGE polyacrylamide gel electrophoresis  
PCR polymerase chain reaction  
R purine A or G, adenine or guanine  
R. or REase restriction endonuclease  
R-M restriction modification  
RT room temperature  
S C or G, cytosine or guanine  
SC supercoiled  
SDS sodium dodecyl sulfate  
SeMet selenomethionine  
TEMED N,N,N’,N’-tetramethylethlenediamine  
Tris 2-amino-2-hydroxymethyl-1,3-propanediol  
TSC triple synaptic complex  
Y pyrimidine T or C, thymine or cytosine  
W A or T, adenine or thymine  
wt wild type
INTRODUCTION

Type II restriction endonucleases (REases) recognize short DNA sequences, typically 4-8 bp in length, and cleave both phosphodiester bonds at fixed position within or close to their recognition sites leaving 5'-phosphate and 3'-hydroxyl groups (reviewed by). Almost 3800 Type II REases specific for more than 260 different nucleotide sequences have been identified by screening various bacterial species (http://rebase.neb.com). Therefore, Type II restriction enzymes probably represent the largest family of functionally related enzymes; however the structural and mechanismical divergence within the family is still poorly understood.

Type II REases are currently subdivided into four families which show evolutionary relationships to different nuclease lineages: PD-(E/D)XK, PLD, HNH and GIY-YIG. PabI restriction enzyme that shows a novel nuclease fold according to the recently published crystal structure may become a founding member of yet another nuclease family.

Type II restriction enzymes of PD-(E/D)XK family are best characterized both structurally and in respect to the mechanism of DNA cleavage. For example, among 26 Type II REase structures present in PDB, 24 belong to the PD-(E/D)XK family. According to the bioinformatic analysis of restriction enzyme sequences available in REBASE, nearly ~70% of them could be assigned to the PD-(E/D)XK family. Enzymes belonging to the latter family are Mg$^{2+}$-dependent endonucleases that use acidic residues from the conserved PD-(E/D)XK motif for metal ion coordination/catalysis. The active site residues are anchored to the conserved structural core, composed of a five stranded β-sheet and two flanking α-helices. The PD-(E/D)XK family restriction enzymes show different specificities and cleavage patterns raising an intriguing question how the conserved structural fold is adapted to interact with different nucleotide sequences.
In general, we would like to understand how different specificities and cleavage patterns evolved within the family of restriction enzymes containing a conserved CC:GG motif in their target sites (see Table 1). In this study, we focused on the structural and biochemical mechanisms by which restriction enzymes Ecl18kI and EcoRII, specific for the \( \downarrow \text{CCNGG} \) and \( \downarrow \text{CCWGG} \) sequences (\( \downarrow \) shows the cleavage position, N stands for the any nucleotide, W – for the A/T), respectively, achieve their function.

Table 1. Family of restriction enzymes containing the conserved CC:GG motif in their recognition sequence

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Subtype</th>
<th>Recognition sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HpyF100III</td>
<td>IIP</td>
<td>( \downarrow \text{CCGG} )</td>
</tr>
<tr>
<td>AgeI</td>
<td>IIP</td>
<td>( \text{A} \downarrow \text{CCGGT} )</td>
</tr>
<tr>
<td>Kpn2I</td>
<td>IIP</td>
<td>( \text{T} \downarrow \text{CCGGA} )</td>
</tr>
<tr>
<td>NgoMIV</td>
<td>IIF</td>
<td>( \text{G} \downarrow \text{CCGGC} )</td>
</tr>
<tr>
<td>BsaWI</td>
<td>IIP</td>
<td>( \text{W} \downarrow \text{CCGGW} )</td>
</tr>
<tr>
<td>Bse634I/Cfr10I</td>
<td>IIF</td>
<td>( \text{R} \downarrow \text{CCGGY} )</td>
</tr>
<tr>
<td>SgrAI</td>
<td>IIF</td>
<td>( \text{C} \downarrow \text{CCGGYG} )</td>
</tr>
<tr>
<td>PspGI</td>
<td>IIP</td>
<td>( \downarrow \text{CCWGG} )</td>
</tr>
<tr>
<td>EcoRII</td>
<td>IIE</td>
<td>( \downarrow \text{CCWGG} )</td>
</tr>
<tr>
<td>Ecl18kI</td>
<td>IIP</td>
<td>( \downarrow \text{CCNGG} )</td>
</tr>
<tr>
<td>PfoI</td>
<td>IIP</td>
<td>( \text{T} \downarrow \text{CCNGGA} )</td>
</tr>
</tbody>
</table>

Ecl18kI REase has been identified in the Enterobacter cloaceae 18k strain\(^\text{13}\). The protein sequence is \( \sim 99\% \) identical to the isoschizomeric restriction enzymes SsoII\(^\text{14}\), SenPI\(^\text{15}\) and StyD4I\(^\text{16}\), which share weak sequence similarity to EcoRII and PspGI endonucleases specific for CCWGG sequence\(^\text{17}\). The structural and mechanical details of Ecl18kI were practically unknown before this study.

EcoRII is an archetypal Type IIE enzyme, which in contrast to the orthodox Type IIP REases, shows no cleavage activity on DNA containing a single recognition site\(^\text{18}\). To activate cleavage it requires binding of two copies of the CCWGG sequence, however only one cognate DNA copy is cleaved while the second acts as an allosteric activator\(^\text{18-20}\). Despite of the numerous publications, the EcoRII mechanism remained controversial and therefore required further experimental studies.

The specific aims of this study

1. To identify the active site and determinants of sequence recognition of REases Ecl18kI and EcoRII.
2. To elucidate how restriction enzymes Ecl18kI and EcoRII interacting with the interrupted palindromic sites CCNGG and CCWGG, respectively, accommodate degenerate central nucleotides within their target sites.

3. To determine DNA cleavage mechanism of restriction endonuclease EcoRII.

**Scientific novelty**

We demonstrate that restriction enzymes Ecl18kI and EcoRII interacting with interrupted target sites CCNGG and CCWGG, respectively, share similar structural fold, active site structure and mechanism of the CC:GG dinucleotide recognition with Bse634I/Cfr10I and NgoMIV endonucleases interacting with palindromic RCCGGY and GCCGGC sites, respectively. We show for the first time that Ecl18kI restriction enzyme flips central base pair within CCNGG target to achieve sequence specificity and DNA cleavage. Base flipping has been reported before for the enzymes (DNA methyltransferases, repair enzymes, etc.) that perform chemistry on the extruded base. Here we provide first experimental evidence that base flipping is employed to achieve specificity of the restriction enzyme in solution. Moreover, we demonstrate that PspGI REase and the catalytic domain of EcoRII (EcoRII-C) restriction enzyme, both specific for the CCWGG sequence, also flip central nucleotides. We also show for the first time that Type IIE REase EcoRII interacts with three rather than two target sites to achieve concerted DNA cleavage at the target site.

**Scientific value**

Restriction enzymes are widely used in biotechnology as molecular scissors that allow cutting DNA at fixed sites and are indispensable in gene cloning and other DNA manipulations. Engineering restriction enzymes with new tailor made specificities could provide new molecular tools for biotechnology. However, the success in the protein engineering of the restriction enzymes is limited. The strategies of the rational design of novel enzymes require a detailed knowledge of the fundamental mechanisms
employed by the restriction enzymes to achieve their specificities. This study demonstrates how evolutionary related REases achieve their specificity and provide novel clues for protein engineers.

**The new findings presented for defense in this thesis**

1. REases Ecl18kI (CCNGG), EcoRII (CCWGG), Bse634I/Cfr10I (RCCGGY) and NgoMIV (GCCGGC) interacting with different but related target sites share conserved active site architecture and structural mechanism of the CC:GG recognition.

2. REase Ecl18kI flips both central nucleotides within the CCNGG recognition sequence and buries the extruded bases in the pockets within the protein.

3. Binding of DNA by the restriction enzymes Ecl18kI, EcoRII-C and PspGI enhances the fluorescence of 2-aminopurines placed at the centers of their recognition sequences providing evidence that these enzymes flip central nucleotide(s) within their target sites.

4. EcoRII REase requires simultaneous binding of three rather than two recognition sites to achieve concerted DNA cleavage at a single site.
LITERATURE OVERVIEW

Type II restriction endonucleases (REases) represent one of the largest families of functionally related enzymes; however protein sequence comparison in general reveals no or very weak similarities between the restriction enzymes raising the question of whether REases evolved independently. With the exception of the catalytic motif, little sequence similarity has been observed between the more than 400 Type II restriction enzymes that have been sequenced to date (http://rebase.neb.com/cgi-bin/statlist). In the first part of the literature review I have analysed available data on the structure-function relationships between restriction enzymes recognizing similar DNA sequences. In the second part of the review I have summarized the available literature data on the archetypal Type IIE REase EcoRII\textsuperscript{21,22} to provide a background for the critical evaluation of the experimental data obtained in this work.

I. The diversity of restriction endonucleases

REases are components of restriction modification (R-M) systems that occur ubiquitously among bacteria, archaia and in viruses of certain unicellular algae\textsuperscript{1}. Their main function is to defend their host against foreign DNA. This is achieved by cleaving incoming DNA that is recognized as foreign by the absence of a characteristic modification (N4 or C5 methylation at cytosine or N6 methylation at adenine) at defined sites within the recognition sequence. The host DNA is resistant to the cleavage as these sites are modified. Additional functions have been attributed to restriction enzymes, including maintenance of species identity among bacteria and generation of genetic variation\textsuperscript{1}. REases of \textit{Chlorella} viruses may have a nutritive function by helping degrade host DNA or preventing infection of a cell by another virus\textsuperscript{23}. Certain types of R-M systems can also be considered as selfish DNA
elements. In general, bacteria and archaea harbor numerous R-M systems. For example, in *Helicobacter pylori* more than 20 putative R-M systems, comprising greater than 4% of the total genome, have been identified in two completely sequenced *H. pylori* strains. Several types of REases exist that differ in subunit composition and cofactor requirement. Commonly, four types are distinguished.

As of 22 September 2008, REBASE ([http://rebase.neb.com/cgi-bin/statlist](http://rebase.neb.com/cgi-bin/statlist)) lists 3871 characterized restriction enzymes: 91 of Type I, 3764 of Type II, 11 of Type III and 5 of Type IV. The predominance of Type II enzymes certainly is biased by their usefulness for recombinant DNA work. The analysis of published genome sequences suggests a somewhat more even distribution among putative R-M systems: approximately 29% Type I, 45% Type II, 8% Type III and 18% Type IV.

Type II REases differ from the Type I, III and IV enzymes by a more simplified subunit organization. They are usually homodimeric enzymes that interact with palindromic DNA sequences and contain two active sites each responsible for the catalyzing cleavage in one DNA strand to make a double strand break within or close to their recognition site. For the cleavage Type II restriction enzymes do not require ATP or GTP. Recently it was demonstrated that several restriction enzymes act as homotetramers or monomers. All Type II REases except BfiI, BmrI and PabI require divalent metal ions (usually Mg$^{2+}$) as a cofactor for DNA cleavage.

According to the structural organization, recognition sequence, cofactor requirements and cleavage mechanism Type II REases are divided into few subtypes (Fig. 1). Orthodox Type IIP enzymes are homodimers that recognize palindromic sequences of 4–8 bp in length, and cleave DNA within this sequence in both strands, producing 3’-hydroxyl and 5’-phosphate ends. Some of them like BglII (Fig. 1) recognize discontinuous palindromes, interrupted by a segment of specified length but unspecified sequence. The DNA fragments produced have ‘blunt’ or ‘sticky’ ends with 3’- or
5’-overhangs. Type IIA enzymes recognize asymmetric sequences and cleave DNA away from the sequence or within the sequence. Type IIB enzymes like AloI (Fig. 1) cleave DNA at both sides of the recognition sequence. Type IIC enzymes have both cleavage and modification domains within one polypeptide. Type IIE enzymes exemplified by EcoRII and NaeI (Fig. 1) need to interact with two copies of their recognition sequence for efficient cleavage, one copy being the actual target of cleavage, the other serving as an allosteric effector.

**Figure 1. Different subtypes of Type II restriction endonucleases.** Recognition sequences are shown by red boxes. Gray arrows indicate cleavage sites, green arrow indicate the allosteric effect of Type IIE enzymes. Orange triangles indicate methylated target. Subtypes illustrated according 26.

Type IIF enzymes like SfiI or NgoMIV (Fig. 1) are typically homotetrameric REases that also interact with two copies of their recognition site, but cleave both of them in a more or less concerted manner. Type IIG enzymes, similar to a subgroup of Type IIC enzymes, have both cleavage and modification domains within single polypeptide and that may be stimulated or inhibited by AdoMet. Type IIH enzymes behave like Type II enzymes, but their genetic organization resembles Type I R-M systems. Type IIM enzymes like DpnI (Fig. 1) recognize a specific methylated sequence and cleave the
DNA at a fixed site. Type IIS enzymes like FokI and BfiI (Fig. 1) cleave target DNA outside of the asymmetric recognition sequence. Type IIT enzymes are heterodimeric enzymes. Some Type II restriction enzymes only nick DNA.

Many Type II restriction enzymes have not yet been characterized in detail and quite a number of enzymes was not assigned to the particular subtype\textsuperscript{1,26}.

II. The similarity of restriction endonucleases

Type II REases comprise a very diverse group of enzymes. Most of them belong to the so called PD-(E/D)XK superfamily of endonucleases\textsuperscript{1}. Restriction enzymes with novel nuclease folds recently were discovered and assigned to the PLD, HNH, GIY-YIG, PabI families\textsuperscript{4-10,38,39}.

With the exception of the catalytic motif, little, if any, sequence similarity has been observed between the more than 400 Type II restriction enzymes that have been sequenced to date. The few exceptions are isoschizomers, that cleave the same sequence at the same position, e.g. EcoRI and Rsrl (\textsuperscript{G}↓AATTC)\textsuperscript{40}, MthTI and NgoQII (GG↓CC)\textsuperscript{41}, Cfr9I and XmaI (C↓CCGGG)\textsuperscript{42}, Bse634I and Cfr10I (R↓CCGGY)\textsuperscript{31}, SbfI and SdaI (CCTGCA↓GG), BsuBI and PstI (CTGCA↓G)\textsuperscript{43}. Most isoschizomers, however, do not share significant sequence similarity. Limited sequence similarity has also been observed in some cases among restriction enzymes that recognize related sequences, e.g. EcoRI (G↓AATTTC) and MunI (C↓AATTG)\textsuperscript{44}; SsoII, StyD4I (↓CCNGG) and EcoRII, PspGI (↓CCWGG)\textsuperscript{17}; Alw26I (GTCTC(N\textsubscript{1}/N\textsubscript{5})↓) and Eco31I (GGTCTC(N\textsubscript{1}/N\textsubscript{5})↓) and Esp3I (CGTCTC(N\textsubscript{1}/N\textsubscript{5})↓)\textsuperscript{45}, BsuBI, PstI (CTGCA↓G) and SbfI, SdaI (CCTGCA↓GG)\textsuperscript{43}.

Therefore, it was thought for a long time that restriction enzymes are not evolutionarily related\textsuperscript{46}. This view began to change as crystal structures of Type II REases became available, demonstrating that restriction enzymes do have a similar structural core that harbors the active site with the characteristic
PD-(E/D)XK motif\textsuperscript{47} (Table 2). Furthermore, a statistical analysis revealed a significant correlation between the amino acid sequences of restriction enzymes and their recognition sequences and mode of cleavage; these findings were interpreted as evidence for an evolutionary relationship among Type II REases\textsuperscript{1,48}.

With the determination of more crystal structures it became clear that REases of PD-(E/D)XK family (Table 2) have a similar structural core. This core consists of a five-stranded mixed β-sheet and two flanked by α-helices, as first recognized by a comparison of the structures of EcoRI and EcoRV\textsuperscript{12,47,49}. Intriguingly, this core is also present in other proteins with a nuclease function (Table 2). The conserved core harbors the catalytic center: it brings into spatial proximity two or three carboxylates, typically one aspartate and one glutamate or aspartate residue, and one lysine residue\textsuperscript{12} (Table 2). Within the common core composed of five-stranded mixed β-sheet only four β-strands are absolutely conserved; two of these strands, the second and third, serve as a scaffold for the catalytic residues of the PD-(D/E)XK motif, the remaining β-strands are critical for the formation of the β-sheet and the hydrophobic core. The fifth β-strand can be parallel (as in EcoRI REase) or antiparallel (as in EcoRV REase) to the fourth strand\textsuperscript{1}. Based on structural differences, in particular the topology of the secondary structure elements (orientation of the fifth β-strand) and the arrangement of the subunits (which depends on the cleavage pattern), the Type II REases of PD-(D/E)XK superfamily are divided into an EcoRI and EcoRV classes or branches\textsuperscript{1,47,50}.

Restriction enzymes that belong to EcoRI-like class (BamHI, BglIII, Bse634I, BsoBI, BstYI, Cfr10I, EcoO109I, EcoRI, FokI, MunI, NgoMIV) usually approach the DNA and positions the central β-sheet in the major groove, recognize the DNA mainly via an α-helix and a loop (therefore sometimes EcoRI-like class is referred as α-class\textsuperscript{51}) and in general produce 5’ staggered ends\textsuperscript{47,50}.
Enzymes of the EcoRV-like class (BcnI, BglI, EcoRV, HincII, HinP1I, MspI, MvaI, NaeI, PvuII, SfiI) usually approach the DNA and positions the central β-sheet in the minor groove, use a β-strand and a β-like turn for DNA recognition (β-class according\textsuperscript{51}) and in general produce blunt or 3’ staggered ends\textsuperscript{47,50}.

However, the division into the structural classes is not absolute. For example, SdaI restriction enzyme according to its topology belongs to the EcoRI-like class\textsuperscript{43} but after cleavage it produces 3’-staggered ends and according to the suggested model positions the central β-sheet in the minor groove as the EcoRV-like class enzyme. Monomeric endonucleases BcnI, HinP1I, MspI and MvaI produce 5’-staggered ends but position the central β-sheet in the minor groove and have the same topology as EcoRV-like class restriction enzymes\textsuperscript{33,35,36,65}.

The formation of a highly specific hydrogen bond network is a characteristic feature of the specific protein-DNA complex of REases. This hydrogen bond network comprises contacts to the bases (‘direct read-out’) as well as to the sugar-phosphate backbone (‘indirect readout’).\textsuperscript{1} For the sequence-specific recognition all REases form mostly direct and also water-mediated hydrogen bonds with the bases in the major groove as all enzymes specifically interacting with DNA\textsuperscript{81}. In some cases, hydrogen bonds to the donor and acceptor atoms exposed in the DNA minor groove also contribute to the sequence recognition. In addition, van der Waals contacts and hydrophobic interactions are formed to the bases of the recognition sequence. Most of the specific contacts are implemented between one subunit and one half-site of the palindromic recognition sequence; a few are directed to the other half-site. A characteristic feature of the recognition process of restriction enzymes is its high redundancy.
Table 2. Crystal structures of Type II restriction endonucleases of the PD-(D/E)XK family and related enzymes*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Subtype</th>
<th>Recognition site</th>
<th>Catalytic residues**</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EcoRI-family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BamHI</td>
<td>II</td>
<td>G(^{+})GATCC</td>
<td>D77</td>
<td>E111</td>
</tr>
<tr>
<td>BglII</td>
<td>II</td>
<td>A(^{+})GATCT</td>
<td>N69</td>
<td>E286</td>
</tr>
<tr>
<td>Bse634I</td>
<td>II</td>
<td>R(^{+})CCGGY</td>
<td>E80</td>
<td>D146</td>
</tr>
<tr>
<td>BsoBI</td>
<td>II</td>
<td>C(^{+})YCGRG</td>
<td>E252</td>
<td>D212</td>
</tr>
<tr>
<td>BstYI</td>
<td>II</td>
<td>R(^{+})GATCY</td>
<td>E75</td>
<td>D119</td>
</tr>
<tr>
<td>Cfr10I</td>
<td>II</td>
<td>R(^{+})CCGGY</td>
<td>E71</td>
<td>D134</td>
</tr>
<tr>
<td>EcoO109I</td>
<td>II</td>
<td>RG(^{+})GNCCY</td>
<td>D77</td>
<td>D110</td>
</tr>
<tr>
<td>EcoRI</td>
<td>II</td>
<td>G(^{+})AATTC</td>
<td>D59</td>
<td>E91</td>
</tr>
<tr>
<td>FokI</td>
<td>II</td>
<td>GATG(N(<em>9)/N(</em>{13}))</td>
<td>D450</td>
<td>D467</td>
</tr>
<tr>
<td>MunI</td>
<td>II</td>
<td>C(^{+})AATTG</td>
<td>D38</td>
<td>E98</td>
</tr>
<tr>
<td>NgoMIV</td>
<td>II</td>
<td>G(^{+})CCGGC</td>
<td>E70</td>
<td>D140</td>
</tr>
<tr>
<td>SdaI</td>
<td>II</td>
<td>CCTGCA(^{+})GG</td>
<td>Q178</td>
<td>D233</td>
</tr>
<tr>
<td><strong>Related to EcoRI-family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TnsA</td>
<td>specific</td>
<td>E63</td>
<td>D114</td>
<td>E149</td>
</tr>
<tr>
<td><strong>EcoRV-family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BcnI</td>
<td>II</td>
<td>CC(^{+})GG</td>
<td>E41</td>
<td>D55</td>
</tr>
<tr>
<td>BglII</td>
<td>II</td>
<td>GCCNNNN(^{+})NGGC</td>
<td>E87</td>
<td>D116</td>
</tr>
<tr>
<td>EcoRV</td>
<td>II</td>
<td>GAT(^{+})ATC</td>
<td>E45</td>
<td>D74</td>
</tr>
<tr>
<td>HineI</td>
<td>II</td>
<td>GTY(^{+})RAC</td>
<td>E38</td>
<td>D114</td>
</tr>
<tr>
<td>HinP1I</td>
<td>II</td>
<td>G(^{+})CGC</td>
<td>E18</td>
<td>D62</td>
</tr>
<tr>
<td>MspI</td>
<td>II</td>
<td>C(^{+})CGG</td>
<td>E35</td>
<td>D99</td>
</tr>
<tr>
<td>Mval</td>
<td>II</td>
<td>CC(^{+})WG</td>
<td>E36</td>
<td>D50</td>
</tr>
<tr>
<td>NaeI</td>
<td>II</td>
<td>GCC(^{+})GGC</td>
<td>E70</td>
<td>D86</td>
</tr>
<tr>
<td>PvuII</td>
<td>II</td>
<td>CAG(^{+})CTG</td>
<td>E55</td>
<td>D58</td>
</tr>
<tr>
<td>SfiI</td>
<td>II</td>
<td>GGCCNNNN(^{+})NGGCC</td>
<td>E55</td>
<td>D79</td>
</tr>
<tr>
<td><strong>Related to EcoRI-family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-)-exonuclease</td>
<td>non-specific</td>
<td>E85</td>
<td>D119</td>
<td>E129</td>
</tr>
<tr>
<td>RecB endonuclease</td>
<td>non-specific</td>
<td>E1020</td>
<td>D1067</td>
<td>D1080</td>
</tr>
<tr>
<td>RecU endonuclease</td>
<td>structure-specific</td>
<td>E36</td>
<td>D88</td>
<td>E101</td>
</tr>
<tr>
<td>Hef nuclease domain</td>
<td>structure-specific</td>
<td>E558</td>
<td>D583</td>
<td>E593</td>
</tr>
<tr>
<td>Hjc resolvase P. furiosus</td>
<td>structure-specific</td>
<td>E9</td>
<td>D33</td>
<td>E46</td>
</tr>
<tr>
<td>Hjc resolvase S. solfataricus</td>
<td>structure-specific</td>
<td>E12</td>
<td>D42</td>
<td>E55</td>
</tr>
<tr>
<td>Hje resolvase S. solfataricus</td>
<td>structure-specific</td>
<td>E10</td>
<td>D39</td>
<td>E52</td>
</tr>
<tr>
<td>MutH</td>
<td>(^{+})GATC</td>
<td>E56</td>
<td>D70</td>
<td>E77</td>
</tr>
<tr>
<td>T7 endonuclease I</td>
<td>structure-specific</td>
<td>E20</td>
<td>D55</td>
<td>E65</td>
</tr>
<tr>
<td>Vsr endonuclease</td>
<td>C(^{+})TWGG</td>
<td>E25</td>
<td>D51</td>
<td></td>
</tr>
</tbody>
</table>

* Structures of EcoRIII\(^{+}\) and Ecl18kI\(^{+}\) are not included in this table since they will be discussed below. Table adapted from1.

** The additional catalytical residue to the residues from PD-(D/E)XK motif was suggested but the functional importance was not proved biochemically for all the enzymes.
III. Similarities between restriction enzymes recognizing similar DNA sequences

Further I will focus on the available structures of REases recognizing similar DNA sequences and yielding similar cleavage pattern. I will examine the individual similarities in their active sites and mechanisms of sequence recognition.

1. Bse634I, Cfr10I and NgoMIV family

The first large group of related and structurally characterized enzymes comprises REases Bse634I/Cfr10I (R↓CCGGY) and NgoMIV (G↓CCGGC), which recognition sequences contain common CCGG motif. These enzymes cut DNA before the first C of the conserved CCGG sequence. The structures of Cfr10I\(^{56}\) and Bse634I\(^{31}\) were determined in the absence of DNA, whereas the structure of NgoMIV\(^{29}\) was solved in the DNA bound form. Unlike orthodox Type II REases that function as dimers, Bse634I/Cfr10I and NgoMIV are active as tetramers and therefore belong to the Type IIF subtype\(^{26}\).

Overall structure. The polypeptide chains of the Bse634I, Cfr10I and NgoMIV are folded in compact α/β structures. They share a common PD-(D/E)XK family core comprised of five β-strands and two α-helices core; however other structural elements are different (Fig. 2). The pair of isoschizomers Bse634I and Cfr10I share 30% sequence identity and 50% similarity that suggest similar folds. Indeed the crystal structures of both proteins are very similar and could be treated as nearly the same structure\(^{31}\) (Fig. 2). The primary dimers of Bse634I, Cfr10I and NgoMIV are structurally similar to that of dimeric restriction enzymes that produce 4 bp 5’-staggered ends after DNA cleavage. The structural elements of these enzymes used for the dimer formation are similar to those of the orthodox REases such as EcoRI, BamHI or MunI\(^{52,58,60}\).
Dimerization interface of NgoMIV and Bse634I/Cfr10I is mostly conserved and is mediated through the contacts between helices α7, α6 and α7 which cross over their counterpart of the related subunit in a V-like fashion. Noteworthy, the N-termini of the latter α-helices are directed towards the major groove of DNA and carry aspartate and arginine residues involved in the C:G base pair recognition (see below). Hence, the dimerization and DNA recognition interfaces of these enzymes are intertwined. Some additional protein segments also contribute to the dimer interface formation.

Figure 2. Structures of tetrameric restriction endonucleases Bse634I, Cfr10I and NgoMIV (PDB ID 1KNV31, 1CFR56 and 1FIU29). Two primary dimers positioned back-to-back to each other are depicted in colored-grey and dark-grey. In one monomer of the dimer the central β-sheet is colored orange, dimerization and tetramerization α-helices light blue, the NgoMIV arm participating in tetramerization cyan and the rest structural elements light green. DNA molecules bound to the NgoMIV are shown in brown.

Tetramers of restriction enzymes Bse634I, Cfr10I and NgoMIV are arranged as dimers of primary dimers positioned back-to-back to each other with DNA binding clefts facing opposite directions (Fig. 2). The structural elements participating in the tetramer formation in NgoMIV and Bse634I/Cfr10I are somewhat different. In the Bse634I/Cfr10I tetramers, structural elements involved in the dimerization (α6 and α7 helices) also contribute to the interface between the two primary dimers. In NgoMIV, beside
the dimerization helices $\alpha 7$, the most extensive contacts are made by the tetramerization loops that spans across the neighboring monomer (Fig. 2). Interestingly, despite of the extensive interactions across of the dimer-dimer interface the single alanine replacement of the W220 residue of Cfr10I (W228 in Bse634I) located at the C-terminus of the $\alpha 7$helices ($\alpha 6$ in Bse634I), disrupts the tetramerization interface generating a protein dimer$^{28,32}$.

Active site architecture. Structural studies revealed conserved features of the divalent metal ion binding sites at the catalytic centers of REases of PD-(E/D)XK family$^{1}$. Three charged residues (two acidic and a lysine) are located at the ends of the conserved $\beta$-sheet in the vicinity of the scissile phosphate. The structure of Cfr10I$^{56}$ supported by the biochemical studies$^{84}$ revealed new variation of the PD-(E/D)XK active site. In Cfr10I a serine is present instead of the second acidic residue in the PD-(E/D)XK motif. Structural data indicate that despite its different location in the linear sequence, E204 of Cfr10I spatially substitutes for the second acidic residue in the PD-(E/D)XK motif. Thus, in Cfr10I the sequence motif$^{133}$PDX$_{55}$KX$_{13}$E$^{204}$ corresponds to the active site, which is structurally similar to those of the other restriction enzymes. Another one glutamate residue E71 structurally coincides with the E45 residue of EcoRV, which is involved in the binding of the second Mg$^{2+}$ ion$^{63}$. The role of E71, D134 and E204 in Cfr10I catalytic activity was confirmed by mutational studies$^{84}$. The double Cfr10I mutant with redesigned active site structure

![Figure 3. Catalytic/metal binding site of the Bse634I, Cfr10I and NgoMIV restriction endonucleases (PDB ID 1KNV$^{31}$, 1CFR$^{56}$ and 1FIU$^{29}$). Superposition of the active site residues of Bse634I (yellow), Cfr10I (crimson) and NgoMIV (blue). Mg$^{2+}$ ions from NgoMIV–DNA structure are shown as green spheres and labeled Me$_A$ and Me$_B$ according to$^{29}$. The scissile phosphate is shown as an orange sphere. Figure adapted from$^{83}$.](image-url)
S188E/E204S exhibited 10% of the wt enzyme cleavage activity. This experiment suggests that spatial rather than sequence conservation plays the dominant role in the formation of the restriction enzyme active site.

NgoMIV structure was solved in the presence of DNA and Mg\(^{2+}\), therefore DNA was cleaved\(^{29}\). Two Mg\(^{2+}\) ions denoted as A and B are present in the NgoMIV active center (Fig. 3). Structural comparisons reveal the conserved spatial location of the metal chelating residues of Bse634I, Cfr10I and NgoMIV, suggesting similar architecture of the catalytic/metal binding sites\(^{29,31,56}\) (Fig. 3). Therefore, it is very likely that Bse634I and Cfr10I also coordinate two metal ions at their active sites. The metal ion B would be chelated by acidic residues of the conserved sequence motif PDX\(_{46-55}\)KX\(_{13}\)E. Aspartate residues belonging to the latter sequence motif and conserved glutamates E80/E71/E70 located on the \(\alpha\)-helix would contribute to the binding site of the metal ion A (Fig. 3).

Thus, both NgoMIV and Bse634I/Cfr10I show a structurally similar organization of their catalytic/metal binding sites, which differs from those of other restriction enzymes in the linear sequence position of the active site residues\(^{82}\).

**DNA recognition.** NgoMIV structure is solved in the presence of DNA and provides the view how the DNA recognition is achieved. As expected for the enzymes producing 4 nt 5'-overhangs, NgoMIV binds the DNA from the major groove side\(^1\). Direct readout takes place primarily in the DNA major groove with one interaction occurring from the minor groove side.

NgoMIV-DNA recognition is symmetrical as it is common for REases: a single subunit of NgoMIV makes hydrogen bonds to the bases of one half site GCC in the major groove, and a single minor groove contact to the C base of the outer GC pair comes from the neighboring subunit\(^{29}\) (Fig. 4A). The two subunits of the protein bind DNA in a classical “crossover” manner\(^{47}\), where the “left” subunit cleaves and makes most of its phosphate contacts to the left DNA half-site, but makes almost all of its base pair contacts to the right
subunit, and *vice versa* for the “right” subunit. The central CCGG part of the NgoMIV recognition sequence is contacted by the residues located on a short contiguous stretch of amino acids with the sequence $^{191}\text{RSDR}^{194}$ just upstream at the N-terminus of the $\alpha_7$ helix (Fig. 4B).

**Figure 4. Structural elements that are involved in the recognition of NgoMIV (PDB ID 1FIU).** (A) Principle scheme of NgoMIV-DNA recognition. NgoMIV monomers are depicted in the different colours. Protein dimer and the scissile phosphates are shown as lagger and smaller spheres, respectively. Protein monomers are depicted in different colours. (B) NgoMIV structural elements that are involved in the recognition of CGGCCG sequence. Structural elements are colored for the different monomers like in (A). (C) Structural elements of NgoMIV involved in the building of the catalytic/metal binding site and discrimination of the central CG nucleotides. Mg$^{2+}$ ions are shown as green spheres. Figure (C) was adapted from $^{83}$.

Arginines R191 and R194 donate bidentate hydrogen bonds to the guanines while carboxylate oxygens of D193 bridge exocyclic amino groups of the neighboring cytosines (Fig. 5A). Thus, three residues R191, D193 and R194 of a single subunit unambiguously specify inner and middle C:G nucleotides within one half-site of the NgoMIV recognition site 5’-GCC. Equivalent residues located on the symmetry related $\alpha_7$ helix make the same set of contacts to the central C:G base pairs of the other GGC-3’ half-site of the recognition sequence. Besides their key role in the inner C:G recognition, D193 and R194 residues also contribute to the dimerization interface making two salt-bridges with R194 and D193 residues of neighboring subunit, respectively. The recognition helix $\alpha_7$ also contributes to the tetramerization
interfaces and might be involved in the cross-talking interactions between individual monomers in the NgoMIV tetramer (Fig. 2).

Figure 5. NgoMIV-DNA recognition (PDB ID 1FIU\textsuperscript{29}). (A) Recognition of the inner C:G base pairs. (B) Recognition of the outer G:C base pair. Only contacts to one half-site of the recognition sequence GCCGGC are shown. (a) and (b) indicate for amino acids which correspond to the different monomers of primary dimer of NgoMIV. Water molecules are shown as red spheres. Hydrogen bonds are shown as dotted lines. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

The sequence motif RSDR present in NgoMIV has the structurally equivalent R(P/S)DR residues in Bse634I\textsuperscript{31} and Cfr10I\textsuperscript{56}, suggesting that these enzymes use the same residues and structural mechanisms to interact with the central bases\textsuperscript{82}. Structural elements of NgoMIV involved in catalysis and CCGG sequence discrimination are intertwined\textsuperscript{29}. Indeed, residues R191, D193 and R194, involved in the recognition of the central CCGG nucleotides are interspaced between the catalytic/metal binding site residues K187 and E201 (Fig. 4C).

The recognition of the outer G:C base pair by NgoMIV is achieved by a recognition module comprised of three non-contiguous structural elements: loop preceding helix α2, helix α8 and helix α3 of the neighboring subunit (Fig. 4B). The major groove contacts to the outer G:C base pair arise from the
D34 and R277 residues located on the loops upstream of helix α2 and helix α8, respectively (Fig. 4B, 5B). The amino group of Q63, located at the N-terminus of the helix α3 of the neighboring subunit, donates a hydrogen bond to the O2 atom of the cytosine in the minor groove. Thus both subunits of a primary dimer contribute to the recognition of the outer G:C pair of one half-site (Fig. 5B). Since the crystal structures of both Cfr10I\(^\text{56}\) and Bse634I\(^\text{31}\) were solved in the absence of the DNA, the recognition mechanism for discriminating the outer base pairs of recognition sequences of Bse634I/Cfr10I is still unknown. The recognition sequence RCCGGY of Bse634I/Cfr10I is less stringent in comparison to that of NgoMIV (GCGGCG). Bse634I/Cfr10I can tolerate two alternative purine nucleotides (A or G) at the first position of its recognition site while eliminating pyrimidines. It was suggested that the interactions with the N7 nitrogen atom of the outer purine base of the Bse634I/Cfr10I sequence could discriminate G:A bases against C:T\(^\text{82}\).

Thereby, structural comparisons between NgoMIV and Bse634I/Cfr10I suggest that recognition mechanisms of the common CCGG tetranucleotide are similar, while the mechanism of outer base pair discrimination is probably different.

2. EcoRI and MunI family

A similar interaction pattern at the protein-DNA recognition interface was shown also for the EcoRI and MunI restriction enzymes, specific for G↓AATTC and C↓AATTG sequences, respectively\(^\text{44}\). Comparison of the protein sequences of MunI and EcoRI revealed local similarities that mapped to the structural elements of EcoRI involved in catalysis and recognition of the common AATT tetranucleotide\(^\text{44}\). The structural comparisons of the crystal structures in the presence of specific DNA\(^\text{58,60,85}\) revealed that the recognition of the AATT tetranucleotide, common to both EcoRI and MunI recognition sites, is achieved by a similar structural mechanism. The mechanism of the external base pair discrimination of EcoRI and MunI, however, differs.
**Overall structure.** Although EcoRI and MunI have the same topology and share a similar α/β core, the central β-sheet in MunI is significantly more twisted than in EcoRI (Fig. 6). Noteworthy, all α-helices located on the convex side of the β-sheet are involved in either DNA recognition or dimerization.

![EcoRI and MunI](image)

*Figure 6. Structures of dimeric restriction endonucleases EcoRI and MunI bound to DNA (PDB IDs 1ERI and 1D02).* In one monomer of the dimer the central β-sheet is colored orange, dimerization α-helices light blue, and the rest structural elements light green. DNA molecules are shown in brown.

EcoRI and MunI produces fragments with 5’-overhangs after DNA cleavage and their dimerization mode is strikingly similar. Three different regions are involved in contacts between the monomers. α4 helixes of EcoRI and 3_{10} helices of MunI coming from the different subunits cross over one another in an X-like fashion (Fig. 6). The loop connecting strands β1 and β2 of both proteins protrudes away from the core of the protein and spans across the neighboring subunit. Additionally, in MunI α5 helix forms a head-to-tail contact to helix α4 of the other monomer and in EcoRI α5 helix interacts with α4 helix from other subunit forming four-helix bundle.

The DNA conformation is remarkably similar in the MunI and EcoRI DNA complexes; the DNA is centrally kinked and has large roll angles at the inner (AATT) base steps.

**Active-site architecture.** Mapping of the EcoRI and MunI structures, active site residues E98 and K100 of MunI spatially overlap with E111 and K113 of EcoRI (Fig. 7). The second conserved acidic residue is missing in the MunI structure since D83A mutant was used instead of the wt for structure
solution. The importance of D83 residue in MunI catalysis, however, was demonstrated by the mutational studies\textsuperscript{86}. The metal ion is absent in MunI structure since D83A was used and replacement of D83 by alanine probably decreases the metal-binding affinity. The putative Mg\textsuperscript{2+} ion position in MunI is occupied by tightly bound water molecule. However, in MunI, only a single metal-binding site was identified that structurally coincides with that of EcoRI, but for EcoRI additional possible D59 was suggested as possible metal-binding residue\textsuperscript{31}.

**DNA recognition.** Both EcoRI and MunI structures were solved in the presence of DNA and allows comparing directly DNA recognition mechanisms for both enzymes. Both MunI and EcoRI face the DNA major groove and protein makes direct contacts to the bases exclusively from the major groove side. Although the recognition sequences of EcoRI and MunI differ only in the external base pairs, different structural elements are involved in the base-specific contacts. In the EcoRI-DNA complex, two long loops (‘inner arm’ formed by residues 115-141 and ‘outer arm’ formed by residues 171-200) extend from the core of the protein and surround the DNA. Residues 137-140 form so called ‘extended chain motif’ located at the C-termini of the ‘inner arm’ (Fig. 8).

The extended chain motif makes a structural platform for localization of a few residues involved in the sequence recognition. The residues located at the
C-terminal ends of these loops are involved in the direct contacts with DNA bases in the recognition sequences of EcoRI. In the case of MunI-DNA complex, the ‘outer arm’ is missing while the ‘inner arm’ (formed by residues 102-118) is considerably shorter and has a different conformation (Fig. 8).

![EcoRI and MunI](image)

**Figure 8. Recognition loops of EcoRI and MunI (PDB ID 1ERI and 1D02).** Protein and DNA are shown in backbone. DNA is colored in orange. ‘Outer’ and ‘inner’ arms of EcoRI and ‘inner’ arm of MunI are colored in violet and red, respectively. Outer base pairs of targets and amino acids recognizing them are depicted in CPK and blue colours, respectively.

While EcoRI uses amino acid residues located on the extended chain motif, helices $\alpha_4$ and $\alpha_5$ to achieve the specificity for its recognition sequences, MunI combines all residues involved in the base-specific contacts within one short segment (residues 115-121) located between the helices $3_{10}3$ and $3_{10}4$ and the N-termini of the $3_{10}4$ helix of MunI.

The recognition of the middle and inner A:T base pairs in EcoRI and MunI is similar and is accomplished mainly by the residues from the conserved sequence motif GNAXER$^{44}$ located on the structurally equivalent helices $\alpha_4$ (EcoRI) and $3_{10}4$ (MunI) and by a few residues just upstream of these helices. The structurally equivalent set of amino acid residues of MunI (G116, N117, A118, R121 and Q102) and EcoRI (G140, D141, A142, R145 and Q115) is involved in the recognition of the central AATT sequence (Fig. 9A).

The discrimination of the external base pair by EcoRI and MunI is achieved by different mechanisms. The single R115 residue of MunI
recognizes the outer C:G base pair. Amino acid residues involved in the recognition of the external G:C base pair by EcoRI are located on the two separate structural elements: R200 and R203 come from the helix α5 and A138 is located on the extended chain motif (Fig. 9B)

Figure 9. DNA recognition by EcoRI and MunI (PDB ID 1ERI and 1D02).

Superposition of recognition residues of EcoRI (crimson) and MunI (blue) of the inner A:T base pairs (A) and the outer base pair (B). Only contacts to one half-site of the recognition sequence are shown. DNA is depicted in CPK using light grey and grey for carbon atoms from EcoRI-DNA and MunI-DNA complexes, respectively. In (A) only DNA from MunI-DNA complex is shown. (a) and (b) indicate for amino acids which correspond to the different monomers of primary dimer. Water molecule from MunI-DNA complex is shown as a red sphere. Hydrogen bonds are shown as dotted lines. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

EcoRI and MunI make an extensive set of contacts with the sugar-phosphate backbone of DNA. In the EcoRI-DNA interface, contacts to six phosphates appear to act as ‘clamp’ to position base recognition elements and stabilize distorted DNA conformation. It was suggested that these ‘clamp’ phosphates are positioned so that the protein can exert a torsional strain to ‘kink’ the DNA. Thus, contacts between EcoRI and these phosphates presumably contribute to the indirect DNA-binding specificity of EcoRI along to the direct contacts to DNA bases. Comparison of the protein-DNA backbone contacts between EcoRI and MunI DNA complexes reveals that contacts with the ‘clamp’ phosphates come from the same parts of the proteins and are highly conserved (Fig. 10). Since the DNA in MunI and EcoRI complexes is
distorted in the same way, the conserved ‘clamp’ contacts are involved in the stabilization of the ‘kinked’ DNA conformation.\(^{60}\)

![Figure 10. EcoRI and MunI interactions with DNA phosphates (PDB ID 1ERI\(^{58}\) and 1D02\(^{60}\)). Only one monomer of a protein and one DNA strand are depicted. The C\(\alpha\) of the amino acids of EcoRI (A) and MunI (B) which make direct or water mediated hydrogen bonds or nonbonded contact (<3.35\(\text{Å}\)) with DNA phosphates are shown as red (identical contacts in EcoRI and MunI) or orange (different) spheres. DNA phosphorus atoms are colored violet and the scissile phosphate is marked as larger spheres.]

3. BglI and SfiI family

Another structurally characterized pair of the enzymes recognizing overlapping DNA sequences and possessing the same cleavage pattern is BglI (GCCNNNN\(^{\downarrow}\)NGGC) and SfiI (GGCCNNNN\(^{\downarrow}\)NGGC).\(^{62,68}\) Both restriction enzymes recognize interrupted palindromic DNA sequences and cleave target to produce 3 nt 3'-overhangs. BglI functions as a dimer; however, SfiI is a tetramer. Despite that there is little sequence similarity between the two enzymes, their modes of DNA recognition are unusually similar. Over 80% of the SfiI residues can be structurally overlapped with those in BglI (RMSD of \(~2.4\ \text{Å}\)), even though the sequence identity between the two enzymes is very low (<15\%). The way in which the BglI dimer binds DNA is very similar to the mode in which the SfiI A/B and C/D dimers dock with DNA.\(^{68}\)
**Overall structure.** The SfiI and BglI monomers have the familiar for restriction enzymes α/β core comprised of a central twisted β-sheet surrounded by α helices (Fig. 11). The α/β core is embellished by a β-substructure consisting of several small β strands (β4, β5, β6, β10, and β11) and a long loop (loop E) that grazes the DNA major groove in the native complex. The residues responsible for the specific recognition of DNA are arranged on this β-substructure both for SfiI and BglI (Fig. 11).

![Figure 11. Structures of dimeric restriction endonuclease BglI (PDB ID 1DMU) and tetrameric restriction endonuclease SfiI (PDB ID 2F03) bound to DNA.](image)

Figure 11. Structures of dimeric restriction endonuclease BglI (PDB ID 1DMU) and tetrameric restriction endonuclease SfiI (PDB ID 2F03) bound to DNA. In one monomer of the dimer the central β-sheet is colored orange, dimerization α-helices light blue, the recognition β-substructure red and the rest structural elements light green. DNA molecules are shown in brown. In SfiI-DNA complex two primary dimers positioned back-to-back to each other are depicted in colored-grey and dark grey-dark grey.

The major difference between the two enzymes is that dimeric interface in SfiI is much smaller. In BglI, a long N-terminal helix (αl) from one subunit wraps around the other subunit, providing additional surface contacts to stabilize the dimer. There is no equivalent of this helix in SfiI, resulting in far fewer dimer contacts. The BglI dimer interface is extremely large and is formed by residues comprising one side of the BglI subunit. The dimer
interface is formed by several segments of the polypeptide chain: the N-terminus, the C-terminal region of helix \( \alpha_3 \), helix \( \alpha_4 \), the loop between \( \beta_1 \) and \( \beta_2 \), just before strand \( \beta_5 \), and the loops between \( \beta_7 \) and \( \beta_8 \), and between \( \beta_8 \) and \( \beta_9 \). The SfiI dimer interface is formed primarily by strands \( \beta_1 \) and \( \beta_2 \) and the connecting loop B from one subunit interacting with helices \( \alpha_2 \) and \( \alpha_3 \) from the other subunit, and vice versa\(^{68}\).

There is a large DNA binding cleft formed by monomers, which is long enough to accommodate DNA duplex with 11 bp and 13 bp recognition sequences for BglII and SfiI, respectively (Fig. 11). The DNA is bound with the minor groove facing the protein, as in the EcoRV and PvuII protein–DNA complexes\(^{63,67}\). The DNA curves away from the SfiI \( \alpha/\beta \) core in a relatively smooth fashion, with a total bend of \( \sim 25^\circ \). Overall, the DNA has a B-form conformation with average helical twist and rise of 34.35° and 3.35 Å, respectively. The BglII DNA is also B-form and curves away from the \( \alpha/\beta \) core, though to a lesser extent (20°). There are no kinks in the SfiI and BglII DNA axes of the type seen in EcoRV and EcoRI DNAs, or any major unwinding at the central base-pair step as seen in EcoRV\(^{63}\), EcoRI\(^{58}\), and BglII\(^{53}\) DNAs. The direction of curvature is the same as in most restriction enzyme complexes, namely away from the \( \alpha/\beta \) core.

The SfiI tetramer has a box-like shape, wherein the four subunits are arranged into a dimer of dimers (Fig. 11). In the native complex, two DNA molecules are bound to opposite sides of the SfiI tetramer, in a back-to-back arrangement, similarly to Bse634I, Cfr10I and NgoMIV\(^{29,31,56}\).

**Active site architecture.** BglII and SfiI possess similar active site architecture, which corresponds to the canonical PD-(D/E)XK motif: the SfiI catalytic residues, D79, D100, and K102, follow the general consensus and overlap with the catalytic residues in BglII, namely D116, D142, and K144 (Fig. 12). Intriguingly, in contrast to other Type IIF restriction enzymes with known structures (Bse634I, Cfr10I and NgoMIV), in SfiI the second catalytic residue comes from the \( \beta \)-meander and not from a distant helix\(^{68}\). Thus, even
though SfiI is tetrameric and belongs to the Type IIF family of restriction enzymes, its subunit structure and DNA-binding arrangement is much more akin to the dimeric Type IIP enzyme BglI.

The BglI structure reveals two bound metals in the active site\(^6\) and in the structure of SfiI-DNA the density for only a single metal is seen\(^6\). The BglI–DNA structure reveals that E87 superimposes with the second Me\(^{2+}\) binding E45 of EcoRV. In the BglI active site, however, E87 does not participate in Ca\(^{2+}\) binding directly, but instead forms a hydrogen bond to the Ca\(^{2+}\)-bound water molecule. SfiI contains E55 at the equivalent position (Fig. 12).

**Figure 12.** Catalytic/metal binding site of BglI and SfiI restriction endonucleases (PDB ID 1DMU\(^6\) and 2F03\(^6\)). Superposition of the active site residues of BglI (crimson) and SfiI (blue). DNA is depicted in CPK using light grey and grey for carbon atoms (Ca\(^{2+}\) ions as pale green and green spheres) from BglI-DNA and SfiI-DNA complexes, respectively. The scissile phosphates are shown as orange spheres.

DNA recognition. Individual subunits of the BglI and SfiI dimers sit entirely on one half-site of the recognition sequence and there is no cross-over mode of binding seen in NgoMIV, EcoRI/MunI (see above), BamHI/BglII/BstYI, where a protein subunit forms contacts to both DNA half-sites\(^6\). DNA recognition occurs in the major groove via residues located primarily on the strand \(\beta_{11}\) of the \(\beta\) substructure and in the loop E which partially enters the major groove. Three of the four G:C base pairs in the SfiI recognition half-site (GGCC) are in common with the BglI half-site (GCC), and the recognition of these is remarkably similar between the two enzymes\(^6,6\) (Fig. 13).
Figure 13. Comparison of base-specific DNA interactions in BglII and SfiI complexes (PDB ID 2F03 and 1DMU). Superposition of recognition residues of BglII (crimson) and SfiI (blue). Only contacts to one half-site of the recognition sequence are presented for clarity. DNA is depicted in CPK using light grey and grey for carbon atoms from BglII-DNA and SfiI-DNA complexes, respectively. Water molecules are shown as red colour spheres. Hydrogen bonds are shown as dotted lines. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

Specifically, R218 and R220 in SfiI and R277 and R279 in BglII donate bidentate hydrogen bonds to the guanines in first and third positions ($G_3C_2C_1$), respectively, while K208 in SfiI and K266 in BglII donate a single hydrogen bond to O6 of the second guanine ($G_3C_2C_1$). Also, the main-chain carbonyl of this same lysine accepts a hydrogen bond from the N4 of the second cytosine.
(G₃C₂C₁). In both enzymes, these arginines and lysines are located on the strand β₁₁ and loop E that bind tangentially to the major groove. Also, S210 in SfiI and its structural analog D268 in BglII make hydrogen bond through water molecule with cytosine in first position (G₃C₂C₁). E106 in SfiI and D150 in BglII that are located in the loop between strands β₃ and β₄, make hydrogen bond with third cytosine (G₃C₂C₁).

Difference between the enzymes recognition is that whereas SfiI does not partake in any minor groove contacts in the recognition of its sequence, BglII makes a series of direct and water-mediated hydrogen bonds via K73 located on the loop A between the helices α₃ and α₄. Interestingly, the equivalent loop in SfiI is much shorter and it does not enter the minor groove nor carry the equivalent of K73. Naturally, SfiI is unique in recognizing the fourth G:C base-pair (G₄G₃C₂C₁). This outer G:C base pair is recognized by R109 that is located at the tip of the strand β₄. R109 makes bidentate hydrogen bonds with the outer guanine. There are no direct base contacts with non-cognate bases in both enzymes.

4. BcnI, MvaI and MutH family

REase superfamily is extremely diverse and includes along with restriction enzymes, λ-exonuclease, Vsr endonuclease, TnsA endonuclease, RecB endonuclease, RecU endonuclease, Hjc and Hje Holliday-junction resolvase, Hef nuclease domain, T7 endonuclease I and the DNA repair protein MutH¹ (Table 2). Structural comparisons of crystal structures of MvaI and BcnI in the presence of specific DNA revealed that these REases share many common features with MutH protein but not with the other restriction enzymes³⁵,³⁶. MutH is a part of DNA mismatch repair protein complex and shows no catalytic activity alone, at least under the physiological conditions. In complex with its partner proteins, MutH is a nickase that cleaves only the unmethylated strand in hemimethylated DNA, its physiological substrate⁹¹. Biochemical and crystallographic studies indicate that MutH interacts with its
target DNA as a monomer. Although the MutH target sequence GATC is palindromic, N6-methylation of the adenine in one strand breaks this symmetry, so that MutH binds DNA only in one orientation and cleaves only one strand. Biochemical and crystallographic characterization of related REases BcnI (CC↓SGG) and MvaI (CC↓WGG) revealed that they are active as a monomers and share many structural features with MutH, including a two subdomain architecture, mobility of the hinge region that connects the subdomains, active-sites that take their productive conformations only in the presence of metal and DNA, and very similar DNA-binding modes, despite unrelated target sequences.

Figure 14. Structures of monomeric restriction endonucleases BcnI, MvaI and DNA repair protein MutH (PDB ID 2ODI, 2OAA, 2AOR) bound to DNA. In all proteins the central β-sheet is colored orange, the target recognition elements red and the rest structural elements light green. DNA molecules are shown in brown, in MutH complex DNA is hemimethylated.

**Overall structure.** BcnI, MvaI and MutH are all monomeric and share a characteristic two subdomain structure (Fig. 14). In the case of MutH, the two subdomains have been termed the ‘N-arm’ and the ‘C-arm’ of the protein and for MvaI and BcnI – ‘catalytic lobes’ and ‘recognition lobes, respectively. For all three enzymes, there is strong evidence that the hinge is highly mobile. In the absence of DNA, the angles between the catalytic and the recognition subdomains are very different for the different enzymes. In the presence of DNA, all enzymes take very similar conformations, which appear to be dictated by the interactions with DNA (Fig. 14). Therefore, presumably, all three enzymes can act as ‘clamps’ for DNA.
Figure 15. Superposition of structures of catalytic and recognition subdomains of BcnI, MvaI and MutH (PDB ID 2ODI\textsuperscript{36}, 2OAA\textsuperscript{35}, 2AOR\textsuperscript{26}). Superposition of structures of catalytic (A) and recognition (B) subdomains of BcnI (crimson), MvaI (blue) and MutH (yellow). DNA is depicted light grey, grey and black from BcnI-DNA, MvaI and MutH-DNA complexes, respectively.

BcnI and MvaI are structurally very similar throughout, but the similarity to MutH is mostly concentrated in the catalytic subdomain, which can be described as the core nuclease fold with a few extra decorations (Fig. 15A). The recognition subdomains of BcnI and MvaI are very similar and can be approximately described as two β-sheets that are stacked against each other to almost form a barrel (Fig. 15B). The MutH recognition subdomain is far more irregular and quite different, despite the presence of some common elements, especially in the region that contacts DNA.

Active site architecture. Active sites of BcnI/MvaI/MutH

Figure 16. Catalytic/metal binding site of BcnI, MvaI and MutH (PDB ID 2ODI\textsuperscript{36}, 2OAA\textsuperscript{35}, 2AOR\textsuperscript{26}). Superposition of the active site residues of MvaI (crimson), BcnI (blue) and MutH (yellow). DNA is depicted in light grey, grey and dark grey (Ca\textsuperscript{2+} ions as pale green, green and olive green spheres) from BcnI-DNA, MvaI-DNA and MutH-DNA complexes, respectively. The scissile phosphates are shown as orange spheres.
proteins are situated in the N-terminal part or catalytic subdomain. Structural superposition of the DNA-bound forms of the catalytic subdomains shows that the most conserved regions anchor the active-site residues (Fig. 16). Interestingly, the locations of the active-site aspartates (BcnI D55, MvaI D50 and MutH D70) superimpose only in the DNA-bound conformations, but vary drastically in the apo-structures. In their productive conformations, the active-site loops of BcnI, MvaI and MutH bind two divalent metal ions.

**DNA binding.** The DNA binds analogously to the BcnI, MvaI and MutH monomers. In all cases, the catalytic subdomains approach the DNA from the minor groove sides, whereas the recognition subdomains interact with the DNA from the major groove side. Only one DNA strand comes close to the single active-site in the monomer.

As BcnI catalyzes double strand breaks, the binding modes with central ‘C’ in the proximal position and with the central ‘G’ in the proximal position must both occur in solution. Similarly, MvaI must be able to accommodate either the strand with central ‘T’ (T-strand) or the strand with the central ‘A’ (A-strand) proximally. Surprisingly, only the binding modes with proximal pyrimidines have been observed in the co-crystals of BcnI and MvaI with cognate oligoduplexes.

As BcnI (CC↓SGG) and MvaI (CC↓WGG) recognize related sequences, one might have expected that they recognize equivalent bases analogously. In indeed, there are a lot of interactions that are conserved (see Fig. 17 for details). Actually, in the major groove side only the C:G base pairs in the (-2) position are recognized identically by BcnI and MvaI. In the case of C:G at (-1) position, R216 in BcnI overtakes the role of H225 in MvaI. However, there some differences in the interaction of BcnI and MvaI with the conserved base pairs in their target sites. In the major groove in the (+2) position BcnI misses an equivalent for R209 in MvaI and in the (+2) position MvaI misses equivalent for N72 in BcnI (Fig. 17).
Figure 17. Specific interactions of BcnI and MvaI with the common C:G base pairs of their recognition sequences (PDB ID 2ODI$^{36}$, 2OAA$^{35}$). Superposition of recognition residues of BcnI (crimson) and MvaI (blue). DNA is depicted in CPK using light grey and grey for carbon atoms from BcnI-DNA and MvaI-DNA complexes, respectively. Water molecules are shown in red spheres. Hydrogen bonds are shown as dotted lines. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

On the minor groove side, most residues form indirect, water-mediated hydrogen bonds with the bases that are not conserved between BcnI and MvaI. Also, MvaI is engaged in two direct hydrogen bonds with the base pair outside
the recognition site, while BcnI does not form direct hydrogen-bonding interactions with DNA outside its target sequence \(^92\).

BcnI and MvaI have different specificities for the central base pair (G:C in the case of BcnI, A:T in the case of MvaI). Therefore, it does not come as a surprise that interactions with this base pair (position (0)) are not conserved (Fig. 18). Note, that BcnI and MvaI must also be able to accommodate the alternative binding mode, which switches the roles of purine and pyrimidine.

The specificities of the MvaI, BcnI REases and of MutH are unrelated. Therefore the architecture of the recognition subdomain of MutH differ from BcnI and MvaI and MutH enzyme uses entirely different machinery for the sequence discrimination \(^92\).

5. BamHI, BglII and BstYI family

One more group of structurally characterized restriction enzymes which interact with overlapping recognition sequences includes BamHI (G↑GATCC), BglII (A↑GATCT) and BstYI (R↑GATCY). BstYI restriction enzyme shares 22% and 30% sequence identities with BamHI and BglIII, respectively. For all
these enzymes structures of apo-form and complexes with specific DNA are solved\textsuperscript{52,53,55,94-98}. Structural comparison of these proteins reveals similarities in proteins fold and active sites, and surprisingly shows different mechanism of recognition of the common GATC tetranucleotide.

**Overall structure.** The $\alpha/\beta$ core of BamHI, BglII and BstYI is composed of six-stranded $\beta$-sheet surrounded by seven, five or seven $\alpha$-helices, respectively, two of which are involved in dimerization in all three proteins (Fig. 19). The loops preceding the dimerization helices fit into the DNA major groove and carry the residues for specific sequence recognition.

![Figure 19. Structures of dimeric restriction endonucleases BamHI, BglII and BstYI bound to DNA (PDB ID 2BAM, 1D2I, 1VRR).](image)

In one monomer of the dimer the central $\beta$-sheet is colored orange, dimerization $\alpha$-helices light blue, recognition substructures red and the rest structural elements light green. DNA molecules are shown in brown.

BamHI and BglII contain a substructure at the base of the $\alpha/\beta$ core that wraps around the backside of the DNA. In BstYI this substructure is much more minimal. In BamHI, this substructure consists of a small two-stranded $\beta$ sheet and an $\alpha$ helix that unfolds on specific DNA binding\textsuperscript{52,97}. In BglII, this substructure is elaborated into a full five-stranded $\beta$ sandwich domain that orients one loop into the minor groove where participates in recognition of bases and another into the major groove where makes numerous contacts with DNA phosphates (totally in all the protein, 8 direct and 20 mediated via water molecules) (Fig. 19). The fitting of these loops into grooves on the backside of the DNA leads to the complete encirclement of the DNA in the BglII complex\textsuperscript{53,96}. The BstYI complex reveals an analogous substructure that
positions a loop into the DNA major groove, making contacts with the sugar-phosphate backbone\textsuperscript{55,98}.

BstYI contains additional extended “arm” subdomain which participates in dimerization and could be responsible for the thermostability of BstYI\textsuperscript{98} (Fig. 19). The $\beta_2$ and $\beta_3$ strands of this subdomain extend across the top of the DNA and several residues along $\beta_3$ make contacts with the phosphate backbone. Additionally, it provides contacts with the DNA backbone and Y115 from the “arm” makes a direct contact with the base outside the recognition sequence\textsuperscript{55}. This interaction, which occurs in both DNA half sites, acts to fix both ends of the DNA duplex at either side of the binding cleft. Thus, the extended “arm” subdomain in BstYI enables the enzyme to make additional DNA contacts that are not present in either the BamHI or BglII complexes.

All three enzymes' dimers approach DNA from the major groove side and wrap around the “back” side of the DNA. Upon DNA binding the dimers undergo conformational changes, although distinct in all three cases. The BamHI dimer binds DNA via a “clamp-like” motion, in which each subunit closes onto the DNA by a 10° rotation around the DNA axis\textsuperscript{52}. BglII accommodates DNA by a “scissor-like” motion in which each subunit swings open by as much as 50° in a direction parallel to the bound DNA axes\textsuperscript{96}. BstYI incorporates both clamping and scissor-like motions, each subunit rotates by $\sim$5° around the DNA axis (clamping) and $\sim$8° perpendicular to the DNA axis (scissor-like)\textsuperscript{55}. The most striking conformational change occurs in the C-terminal part (“arm”) of BamHI (residues 194-213). In this region an ordered $\alpha$ helix ($\alpha_7$) in apo-enzyme unfolds in specific complex and becomes extended and partially disordered\textsuperscript{52,99}. One “arm” from one subunit fits into the minor groove and another one from the second subunit follows the DNA sugar-phosphate backbone.

In BamHI specific complex DNA is relatively straight, while in BglII and BstYI specific complexes DNA is distorted by mild bending away from the proteins $\alpha/\beta$ core and by local unwinding and overwinding\textsuperscript{55,99}. 
Active site architecture.

BamHI, BglII and BstYI active sites are deviate from the classic PD-(D/E)XK consensus motif and contain glutamate (E113 in BamHI) or glutamine (Q95 in BglII or Q130 in BstYI) instead of lysine. Superposition of BamHI/BglII/BstYI active sites is shown in Fig. 20. Dorner and Schildkraut experimentally demonstrated that changing E113 in BamHI to lysine or glutamine inactivates the enzyme.

In some restriction enzymes an extra acidic residue coordinates the second metal ion at the active site. Crystal structures of the BamHI pre-reactive and post-reactive complexes have shown that E77 together with D94 coordinate a second metal in the active site. In contrast, in the BglII co-crystal structure, only a single octahedrally coordinated Mg ion was seen at the active site. At the corresponding position BglII has a less acidic asparagine residue N69. No divalent cations were used in the crystallization of the BstYI-DNA complex, but E75 was suggested as putative analog of E77 and N69 in BamHI and BglII, respectively. Since, in the BstYI-DNA structure, E75 is too distant to coordinate a metal ion in the active site, it in the presence of metal ions this residue should move closer to the other active site residues (Fig. 20).

DNA recognition. As the BstYI recognition sequence (RGATCY) overlaps with that of BamHI (GGATCC) and BglII (AGATCT) one could
expect that the common inner G:C and middle A:T base pairs could be recognized in a similar way. Surprisingly, all three enzymes exhibit different patterns of specific contacts with bases (Fig. 21).

Figure 21. Specific interaction of BamHI, BglII and BstYI with common base pairs of their recognition sequences (PDB ID 2BAM94, 1D2I53, 1VRR55). Superposition of recognition residues of BamHI (yellow), BglII (red) and BstYI (blue) of G:C and A:T base pairs. Only contacts to one half-site of the recognition sequence are shown. DNA is depicted in CPK using light grey, grey and dark grey for carbon atoms from BamHI-DNA, BglII-DNA and BstYI-DNA complexes, respectively. (a) and (b) indicate for amino acids which correspond to the different monomers of primary dimer. Water molecules are shown in red spheres. Hydrogen bonds are shown as dotted lines and van der Waals contacts as dotted arrow. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

The G base within the common GATC subsite is recognized by through the different set of contacts made by structurally equivalent amino acid residues N116 (BamHI), N98 (BglII) and K133 (BstYI) (Fig. 21). However the last C base is recognized similarly by all three enzymes: N4 of cytosine is hydrogen-bonded to the main chain carbonyl of D154 (BamHI), N140 (BglII) or S172 (BstYI) which are located at equivalent positions. Additionally,
guanine (-2) is recognized by BamHI R122 in major groove and BglII R192 in minor groove.

Curiously, there are no direct or water-mediated hydrogen bonds to the inner A:T base pair (GATC) in the BstYI-DNA complex at all (Fig. 21). Nonpolar portion of K133 makes van der Waals contacts with the methyl group of the thymine of the inner A:T base pair. It was suggested that part of the specificity for an inner A:T base pair in the BstYI recognition sequence could be derived from “indirect readout,” given the undertwisting at the central base pair step (RGATCY)\textsuperscript{55}. Like K133, the nonpolar portions of N116 in BamHI and N98 in BglII also mediate similar van der Waals contacts with the inner thymine (Fig. 21). But in contrast to BstYI, in the BamHI-DNA complex, the N116 make hydrogen bonds with the inner thymine and inner adenine through water molecule\textsuperscript{52}. In the BglII-DNA complex, the inner adenine is hydrogen bonded to S97 from another subunit; Y190 from both subunits make hydrogen bond network through water molecule\textsuperscript{53}.

6. HinP1I and MspI family

The co-crystal structures of restriction enzymes HinP1I (G↓CGC)\textsuperscript{34,65} and MspI (C↓CGG)\textsuperscript{33,101} further illustrate remarkable structural diversity of REases. HinP1I and MspI like BcnI and MvaI (see above) interact with DNA as monomers. The enzymes make specific contacts with all 4 base pairs in the recognition sequence. Without significant sequence homology, MspI and HinP1I display striking structural similarity but differ from MvaI and BcnI. Moreover, almost all the structural elements of HinP1I can be matched in MspI, including both the DNA recognition and catalytic elements. Despite both enzymes have structural equivalents for DNA recognition, the recognition mode of the common central CG dinucleotide is different\textsuperscript{33,34,65}. 
Figure 22. Structures of monomeric restriction endonucleases HinP1I and MspI (PDB ID 1YNM and 1SA3) bound to DNA. In both proteins the central β-sheet is colored orange, the recognition β-sheet red and the rest structural elements light green. DNA molecules are shown in brown.

**Overall structure.** HinP1I/MspI proteins fold into an α/β architecture similar to PD-(E/D)XK family enzymes. Apart of the common central five-stranded mixed β–sheet sandwiched on both sides by α-helices, HinP1I/MspI, similar to BglII/SfiI, contain additional smaller β–sheet (β5, β8, β9, β10, β11 for HinP1I and β4, β8, and β7 for MspI) that makes specific contacts with the DNA bases in the major groove (Fig. 22). The HinP1I and MspI structures are very similar, except an extra C-terminal helix in HinP1I and an extra N-terminal helix in MspI. The DNA duplexes in the complexes are primarily in B-form with no major bends or kinks.

Structure of HinP1I with DNA revealed unexpected dimerization mode in the crystal lattice (Fig. 23). Two monomers of HinP1I are orientated in a back-to-back fashion, positioning the catalytic and the DNA binding sites on the opposite surfaces of the dimer facing away from one another, rather than towards one another as it...
occurs in all other structurally characterized REase homodimers. HinP1I dimer contacts are similar to those observed in the tetrameric restriction enzymes Bse634I, Cfr10I and NgoMIV. Some biochemical data suggest that HinP1I can form a dimer at high protein concentrations.

**Active site architecture.** HinP1I and MspI has PD-(D/E)XK-like catalytic site. The catalytic residues of EcoRV, D74, D90 and K92 align spatially with HinP1I residues D62, Q81 and K83, in which glutamine occurs in place of the second acidic residue (Fig. 24A).

![Figure 24. Superposition HinP1I and MspI (PDB ID 2FKH and 1SA3).](image)

Superposition of the poten tially active site residues (A) and recognition β-sheets (B) of HinP1I (crimson) and MspI (blue) are shown. DNA is depicted in CPK using light grey and grey for carbon atoms, for HinP1I-DNA and MspI-DNA complexes respectively. Ca$^{2+}$ ions from Hinp1I-DNA complex are shown as green spheres. The scissile phosphates are shown as orange spheres. Only residues which make hydrogen bonds with DNA bases in the major groove are shown.

The catalytic motif of HinP1I appears to be SDX$_{18}$QXK, an unusual unique feature that is also represented in MspI as TDX$_{17}$NXK, in which asparagine takes the place of the second acidic residue. In addition, E18 in HinP1I and E35 in MspI are superimposable on E45 in EcoRV, suggesting their possible role in metal ion binding/catalysis.

**DNA recognition.** HinP1I and MspI have common CG:CG dinucleotide in their recognition sequences ($G^4$CGC and $C^4$CGG, respectively) and cleave their target sites in the same position. In the MspI-DNA and HinP1I-DNA complexes, the enzymes approach and recognize DNA mainly from the major
groove. In addition to the phosphate interactions, all eight bases of the tetranucleotide recognition sequence have direct hydrogen bond interactions with one HinP1I/MspI molecule. All the amino acids involved in the specific contacts with DNA within the major groove are located on, or near to, the recognition β-sheet (Fig. 24B).

**Figure 25.** Specific interaction of HinP1I and MspI with the central base pairs CG:GC of their recognition sequences (PDB ID 1YNM and 1SA3). Superposition of recognition residues of HinP1I (crimson) and MspI (blue). DNA is depicted in CPK using light grey and grey for carbon atoms from HinP1I-DNA and MspI-DNA complexes, respectively. Water molecules are shown in red spheres. Hydrogen bonds are shown as dotted lines. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.

The central dinucleotide is recognized by HinP1 in DNA major groove by Q93, K223 and Q236. MspI recognition network is more complex. MspI makes direct and water mediated hydrogen bonds to the central CG:CG in the major groove by S127, Y249, S251 and Q259 and E130. It makes also a water mediated hydrogen bond to the cytosine (-1) base form the minor groove by S27 (Fig. 25). Thus, both enzymes recognize the common CG:CG differently; the only structural equivalent Q236 in HinPI coincide with Q259 in MspI that make contact with guanine in (+1) position (GCGC and CCGG, respectively).
7. EclI8kI, EcoRII and PspGI family

In 1998 paper Morgan et al.\textsuperscript{17} identified a conserved segment of 87 amino acid residues (34% similarity and 20% identity) in the amino acid sequences of four restriction enzymes EcoRII/PspGI and SsoII/StyD4I (isoschizomers of EclI8kI) that recognize CCWGG and CCNGG sequences, respectively, and cleave before the first C. They proposed that this segment could be a part of a common DNA recognition domain for the common CC\_GG sequence; however they were unable to identify amino acid residues involved in DNA recognition.

8. Concluding remarks: structural and functional similarities between restriction enzymes

Literature analysis provided in the section III. shows that often restriction enzymes specific for the overlapping nucleotide sequences and yielding identical cleavage pattern use conserved structural mechanisms for the interaction with common nucleotide subsets in their target sites. Indeed, similar recognition mechanism for conserved nucleotide subsites was demonstrated for the Bse634I/Cfr10I/NgoMIV, EcoRI/MunI, BglII/SfiI and BcnI/MvaI REase families. This conclusion, however, is not absolute, since BamHI/BglII/BstYI family members display considerably different protein-DNA contacts to the common GATC sequence within their recognition sites. The same is true for the HinP1I/MspI subfamily.

IV. Restriction endonuclease EcoRII

EcoRII REase recognizes the DNA sequence CCWGG and cleaves the phosphodiester bond at the 5’ end of the first C of the unmethylated sequence\textsuperscript{21,22}. EcoRII is Type IIE restriction enzyme which interacts with two copies of their recognition sequences\textsuperscript{26,102-104}. One of them serves as an allosteric activator which is not cleaved but stimulates the cleavage of the second DNA copy. REase EcoRII was subjected for biochemical studies over
more than 30 years. In this chapter, I will summarize basic facts available in the literature on EcoRII restriction enzyme.

1. The resistance of rarely occurring EcoRII specific recognition sites against the cleavage by EcoRII

EcoRII was among the very first R-M enzymes for which the DNA recognition sequence was determined\textsuperscript{21,22}. As REases are a part of R-M systems and the substrate of the restriction enzyme is foreign DNA, the infecting and growing of virulent bacteriophages is restricted in the bacteria strains were the genes of methyltransferase and REase are expressed. Unusually and differently from other R-M systems it was noticed that the single-stranded DNA phages fd and M13 are not restricted in the cells carrying EcoRII-encoding plasmids\textsuperscript{105} but \textit{in vivo} they are substrates for EcoRII methylation\textsuperscript{106}. Vovis\textsuperscript{107} noted that complete digestion of phage f1 replicative form DNA could not be obtained by addition of EcoRII, and Hattman\textsuperscript{108} reported that \textphi\textsubscript{X}174 replicative form DNA could be completely digested only in the presence of unmodified heterologous DNA. The similar phenotype of EcoRII was described by Kruger investigating the ability to control growth of the virulent bacteriophages T7 and T\textsubscript{3}\textsuperscript{109}. The common feature of the resistant for EcoRII cleavage phage DNAs was the rare existence of the specific EcoRII sites; for example there are only two and three EcoRII sites, respectively, in T7 and T\textsubscript{3} DNA of \textasciitilde 40 kbp size.

2. EcoRII can be activated to cleave refractory DNA recognition sites

Inability of EcoRII enzyme to restrict the infecting phage was explained by Kruger et al.\textsuperscript{18,109}. It was suggested that the resistance of DNA molecules for EcoRII cleavage could be related to the low frequency of EcoRII sites in the phage DNA. First, they demonstrated that DNA of T7 and T\textsubscript{3} phage genomes become susceptible to EcoRII cleavage in the presence of pBR322 \textit{dcm}\textsuperscript{−} plasmid or phage \textlambda\textit{dcm}\textsuperscript{−} DNA which contain 6 and 70 EcoRII sites,
respectively. Second, they showed that complete cleavage of phage DNA occurs down to a ratio of 2 pBR322 EcoRII sites to 1 T3 DNA site. When the ratio was lowered further, the cleavage of T3 DNA started to decline\textsuperscript{18}. Authors concluded that EcoRII is a prototype of REases, which require simultaneously binding at least 2 recognition sites for their activity. The cleavage resistance of refractory EcoRII sites was explained by the low site frequency in DNA molecule. They speculated that the optimal distance between two sites or the orientation of the asymmetric recognition sites might play an important role in the EcoRII function.

Later Pein et al. demonstrated that synthetic oligonucleotide duplexes of 14 bp with a single EcoRII site were sufficient to activate the cleavage of T3 DNA as well as the other DNA molecules\textsuperscript{19}. To understand the activation mechanism, Pein et al. argued whether the activator DNA necessarily should be a cleavable. He showed that the EcoRII derived pBR322 cleavage products stimulated T3 DNA cleavage\textsuperscript{20}. Moreover, even oligonucleotides mimicking EcoRII cleavage products were efficient activators as long as they provided the 5’-phosphate group. On the other hand, oligonucleotide duplexes displaying intact recognition sequences with modified base (6-methyl-2’-deoxyadenosine, 2’-deoxyinosine, N4-methyl-2’-deoxycytidine) were still capable of activating the enzyme without being cleaved themselves\textsuperscript{20}. So, these data showed that activator molecule stimulated the cleavage of DNA substrate but was not cleaved by EcoRII enzyme.

3. Preliminary EcoRII mechanistic models

Several models of EcoRII action were proposed by different groups studying EcoRII.

3.1. Kruger’s model

Based on the T7 and T3 phage DNA cleavage studies Kruger suggested EcoRII mechanism. According his model, EcoRII first binds reversibly to a single site on a DNA molecule. Cleavage can occur only if the second specific
DNA binding site on the enzyme is occupied before the enzyme dissociates from the first DNA molecule (Fig. 26)\textsuperscript{110}.

Gabbara and Bhagwat\textsuperscript{111} studied EcoRII mechanism using short 14 bp oligoduplexes. They demonstrated that EcoRII cleavage of oligoduplex with one recognition site follows a sigmoidal concentration dependence. At constant protein and low activator concentrations, DNA was poorly cleaved by the enzyme. Furthermore, increase of DNA concentration resulted in the sigmodal dependence of the cleavage reaction rate on substrate concentration which implied substrate cooperativity. Moreover, at high concentrations the activator oligoduplex molecules acted both as activators and substrates suggesting that EcoRII could not discriminate between the activator and substrate molecules.

\textbf{3.2. Petrauskene’s model}

Petrauskene et al. further studied the two site binding model suggested by Krugger using synthetic oligonucleotide duplexes. To determine the stoichiometry of the active EcoRII-substrate complex, Petrauskene et al.\textsuperscript{112} carried out EcoRII cleavage experiments using increasing enzyme concentrations at the constant DNA concentration. The cleavage reaction rates increased with

\textbf{Figure 26. Kruger’s model of DNA cleavage by EcoRII.} Solid lines represent DNA molecules, open boxes represent EcoRII sites, ovals represent EcoRII enzyme molecules. Enzyme first binds reversibly to a single site on a DNA molecule. Cleavage can occur only if a second specific DNA binding site on the enzyme is occupied before the enzyme dissociates from the first DNA molecule. Figure adapted from\textsuperscript{110}.

\textbf{Figure 27. Petrauskene’s model of DNA binding by EcoRII.} Solid lines represent DNA molecules, circle represent EcoRII enzyme monomer. Inactive complex with one recognition site is formed in a first binding step and in the second step the active complex of two coordinated recognition sites is formed. Figure adapted from\textsuperscript{112,113}.
the increasing enzyme concentration until it became comparable to DNA concentration. The excess of EcoRII over DNA inhibited the enzyme activity. These data were interpreted as simultaneous interaction of EcoRII enzyme with two DNA recognition sites\textsuperscript{112}.

Testing the EcoRII activation with non-hydrolyzable oligoduplexes containing pyrophosphate bond(s), Petrauskene et al. concluded that enzyme cuts one of two coordinated recognition sites in one catalytic event\textsuperscript{114}.

They also speculated, that the activator molecule is not an allosteric effector but acts as a second substrate\textsuperscript{113}. Based on gel filtration data which showed that EcoRII existed as a dimer in solution\textsuperscript{115} and oligoduplex cleavage data Petrauskene et al. suggested a slightly modified EcoRII mechanism\textsuperscript{112,113}. According the proposed model, EcoRII-substrate interaction involves two steps: first, enzyme forms inactive complex with one recognition site forming complex A. Then with increasing DNA concentration, as the concentrations of DNA sites and protein become more nearly equal, the second recognition site interacts with the dimer and the complex A* forms. Complex A* is unstable in the absence of Mg\textsuperscript{2+} because of negative repulsion of the DNA phosphate groups. Dissociation of complex A* produces two B complexes. Complex B is active and DNAs are hydrolyzed in complex\textsuperscript{116}.

In contrast to the model where EcoRII dimer interacted with two DNA molecules, Karpova et al.\textsuperscript{116} claimed that the active complex is formed by a monomeric EcoRII subunit and one DNA recognition site (Fig. 28).

3.3. Karpova’s model

Figure 28. Karpova’s hypothetical model of interaction of EcoRII with DNA recognition sites. Ovals represent the dimer of EcoRII, unbroken parallel lines indicate DNA substrate and broken parallel lines represent the product of the endonuclease reaction. Two EcoRII subunits bind one cognate recognition site forming complex A. Then with increasing DNA concentration, as the concentrations of DNA sites and protein become more nearly equal, the second recognition site interacts with the dimer and the complex A* forms. Complex A* is unstable in the absence of Mg\textsuperscript{2+} because of negative repulsion of the DNA phosphate groups. Dissociation of complex A* produces two B complexes. Complex B is active and DNAs are hydrolyzed in complex\textsuperscript{116}.
Authors came to this conclusion by studying the binding of EcoRII to various synthetic DNA duplexes, which contained either the modified or the canonical EcoRII recognition sequence. This model implies that after occupation of both DNA-binding sites on the EcoRII dimer, the complex dissociates into two monomer-DNA complexes that are both catalytically active in the presence of Mg\(^2+\). After product release the monomeric EcoRII subunits are supposed to form dimeric enzyme molecules again.

4. EcoRII-DNA interactions: two sites versus one site

Cooperative binding properties of EcoRII were studied by different methods. To demonstrate the stoichiometry of the active DNA-enzyme complex Reuter et al. titrated the EcoRII enzyme against a constant substrate concentration\(^{117}\). Plasmid pBR322 \(dcm^-\) DNA (6 EcoRII sites) was incubated with varying amounts of EcoRII and the completeness of plasmid DNA cleavage in various protein concentrations was monitored. Authors concluded that 0.25-0.5 EcoRII dimers per recognition site are necessary for the formation of an active EcoRII complex\(^{117}\). Higher and lower ratios of enzyme dimers to recognition sites result formation of inactive DNA-enzyme complexes. To characterize the native EcoRII-DNA complexes, gel filtration experiments were carried out. Only single species of EcoRII-DNA complex eluted from a gel filtration column with molecular weight corresponding to a dimeric enzyme bound to two DNA substrate molecules\(^{117}\).

Trying to discriminate between the looping and sliding mechanisms for target site search by EcoRII restriction enzyme, Reuter et al. used the Lac repressor as “molecular barrier”\(^{117}\). The linear 445 bp DNA fragment containing two EcoRII sites at a distance of 191 bp and a \(lac\) operator located in between was chosen as a substrate. The Lac repressor masks about 24 nucleotides on the DNA and is released by the inducer IPTG. Trying to ensure only \(in\ \text{cis}\) EcoRII reactions on DNA fragment substrate concentration was held low at 2.5 nM. The Lac repressor placed between two EcoRII recognition
sites did not inhibited REase activity, indicating that cooperativity between EcoRII sites is achieved by bending or looping rather than “tracking” of the intervening DNA stretch.

Cleavage of linear DNA substrates with differently spaced interacting EcoRII sites revealed an inverse correlation between the cleavage rate and the site distance. Cleavage rate of DNA fragments changed starting from more than 80% for a DNA fragment with a distance of 5 bp between two EcoRII sites to 6% for DNA fragment with interspaced two sites with 947 bp. It was suggested that probability of finding the second DNA site by EcoRII bound to the first site decreases with increasing distance between the sites and ~1000 bp distance becomes critical for loop formation by EcoRII. The cleavage of two 71-mers with a distance of 5 bp between two EcoRII sites with different orientation revealed that EcoRII cleavage is independent on the recognition site orientation.

Transmission electron microscopy provided direct evidence that EcoRII induces loop formation after binding of two EcoRII recognition sites 191 bp apart on linear DNA molecules. Single molecule technique confirmed that EcoRII enzyme is able to form stable loops in the absence and presence of divalent metal ions. Recently Takahashi et al. examined binding and cleavage kinetics of EcoRII by following the frequency (mass) changes of a DNA-immobilized quartz-crystal microbalance (QCM). They observed that 2-site DNA-immobilized QCM was cleaved by EcoRII in the presence of Mg$^{2+}$ ions but for 1-site DNA-immobilized QCM only the binding behavior was observed. Authors concluded that EcoRII requires two recognition sites for cleavage.

5. EcoRII reactions on modified oligoduplexes

Yolov reported that EcoRII enzyme is able to cleave the synthetic DNA polymer with repeated EcoRII recognition sequences were the central A:T base pair in the EcoRII site is replaced by mismatch A:A or T:T. Using the
same the synthetic DNA polymer substrates authors demonstrated that EcoRII endonuclease catalyzes single-strand cleavages both in dA- and dT-containing strands of the recognition site if the cleavage of the other strand has been blocked by chemical modification (phosphoamide or pyrophosphate internucleotide bonds) of scissile bond or if the other strand has been cleaved.

Petrauskene et al. used oligoduplexes containing 1-(β-D-2'-deoxy-threo-pento-furanosyl) cytosine (dCx) and/or 1-(β-D-2'-deoxy-threo-pentofuranosyl) thymine (dTx) in place of dC and dT residues in the recognition site as a substrates for EcoRII cleavage. Hydrolysis of dCx-containing DNA duplexes by EcoRII endonuclease was blocked and binding affinity was strongly depended on the location of an altered sugar. Contrary, a DNA duplex containing a dTx residue was cleaved by the enzyme.

DNA duplexes containing the natural methylated bases N6-methyladenine (m^6Ade), N4-methylcytosine (m^4Cyt) or C5-methylcytosine (m^5Cyt) in one strand of the recognition sequence were resistant to EcoRII cleavage. Hydrolysis of these modified duplexes was observed in the presence of the canonical substrate. Incorporation of m^4Cyt or m^5Cyt into both strands of the recognition sequence precluded such activation by a canonical substrate. EcoRII also failed to cleave substrate analogs in which one of the nucleosides in the recognition site was replaced by the 1,2-dideoxyribose or by 1,3-propanediol (DNA with an abasic site). The hydrolysis of DNA duplexes with non-nucleotide inserts was also activated in the presence of canonical substrate. EcoRII also did not cleave oligoduplexes containing 2-aminopurine placed no in the centre of the recognition site, nor in place of inner and outer G, or both. But the hydrolysis of the duplex containing 2-aminopurine in place of adenine was activated in the presence of the canonical substrate.
6. Putative DNA binding sites of EcoRII

To identify DNA binding regions of EcoRII Reuter et al.\textsuperscript{127} used matrix-bound peptide scans that represented the complete EcoRII amino acid sequence as overlapping peptides. Dodecapeptides overlapping by nine amino acids were prepared by spot synthesis. After the screening two separate DNA-binding regions emerged: amino acids 88-102 (binding site I, RHFG\textsuperscript{K}TRNEKRITRW) and amino acids 256-273 (binding site II, NSVSNRR\textsuperscript{K}SRAGKSLELH), which share the consensus motif KXXRXK.

Peptide substitution analogs of both potential DNA binding sites of EcoRII protein were synthesized in which every amino acid of the original sequence was replaced by all others. Results obtained with these substitution analogs demonstrated that the replacement of the basic amino acids within the consensus strongly affected DNA binding of the respective peptides. Consequently, EcoRII mutant proteins were generated by replacing the consensus lysine residues by alanine or glutamic acid in one or both DNA-binding sites. Mutant proteins showed different effects – mutations in DNA-binding site I attenuated DNA binding, whereas corresponding mutations in DNA-binding site II inhibited DNA cleavage without changing the DNA binding affinity in comparison to the wt EcoRII\textsuperscript{127}. Authors hypothesized that both DNA binding sites I and II contribute to the DNA binding interface of EcoRII.

Reuter et al. also made an attempt to identify the possible active site residues in EcoRII protein. Since most Type II restriction enzymes possesses PD-(E/D)XK like active site Reuter et al. identified several putative catalytic motives and tested their role in EcoRII function by mutational analysis. She generated E96A, D130A and E234A mutants and studied catalytic properties of mutant proteins\textsuperscript{127}. Substitution of glutamic acid at positions 96 by alanine (possible catalytic motif \textsuperscript{76}PDX\textsubscript{13}EKR\textsuperscript{96}) eliminated DNA cleavage and binding of EcoRII, whereas when alanine was introduced into equivalent positions of the two other putative catalytic motifs, \textsuperscript{111}PEX\textsubscript{17}DCK\textsuperscript{132} and \textsuperscript{214}PDX\textsubscript{18}ELH\textsuperscript{236},
DNA cleavage activity was not changed. Authors concluded that E96 is part of a potential catalytic motif located in DNA binding site I\(^{127}\).

Reuter et al. performed EcoRII sequence homology studies detected local homologies of DNA binding site II with SsoII REase (\(^4\)CCNGG), isoschizomer of Ecl18kI, and weak homologies with enzymes with recognition sequences containing C:G or G:C pairs in their recognition sequences (CviAII, NaeI, DpnII, LlaII, CglII, MthTI, NgoPII, ScrFI). Therefore Reuter et al. proposed that EcoRII binding site II is responsible for DNA recognition\(^{127}\).

Authors concluded that the organization of EcoRII functional regions resembles that proposed for NaeI, another Type IIE enzyme\(^{128,129}\). Reuter et al. also speculated that EcoRII monomer may have even two active site, similar like PI-SceI homing endonuclease\(^{130}\). As it emerged that E271 is essential for EcoRII DNA hydrolysis, authors suggested that both glutamates, E96 and E271 (possible catalytic motifs \(^{76}\)PDX\(_{18}\)EKR\(^{98}\) and \(^{215}\)PDX\(_{55}\)ELH\(^{273}\), respectively), may be considered to be part of putative catalytic motifs\(^{117,118}\).

7. Domain organization of EcoRII

In order to analyze the possible domain structure of EcoRII, Mucke et al.\(^{131}\) subjected EcoRII to limited proteolysis. The trypsin and chymotrypsin digestion patterns were determined in the presence and in the absence of specific DNA. EcoRII proteolysis revealed two-protease-resistant domains involving approximately the N- and C-terminal halves of the protein. Analytical ultracentrifugation revealed that EcoRII-N domain is monomer in solution. DNA binding protected the N-terminal domain of EcoRII against proteolysis. Isolated N-terminal domain of EcoRII (EcoRII-N) bound cognate DNA with an affinity ~ 20 fold less than wt EcoRII but did not cleave it.

In contrast, EcoRII C-terminal domain cleaved DNA specifically and did not require anymore the presence of the second DNA copy. Analytical ultracentrifugation experiments revealed that EcoRII-C domain is a dimer in solution. EcoRII-C cleaved DNA substrates independent of the number and the
distance of the DNA recognition sites in vitro and in vivo. DNA of phage T3 that resists cleavage by wt EcoRII was completely cleaved by EcoRII-C\(^{131}\). In contrast to wt EcoRII, EcoRII-C restricted growth of phage T3 by 1.5 orders of magnitude\(^{104}\). Therefore, it was concluded that EcoRII-C represents a standard Type IIP REase which cleave specifically at a single DNA site.

In order to identify the amino acids of EcoRII that interact specifically with the recognition sequence Mucke et al. used photocross-linking technique\(^{132}\). In EcoRII recognition sequence, either each C was substituted with 5-iododeoxycytidine or each A, G, T was substituted with 5-iododeoxyuridine. Iodopyrimidine bases were excited using a UV laser to result in covalent cross-linking products. EcoRII protein was photocross-linked to the 5′C of the CCAGG strand of the recognition sequence. Interestingly, photocross-link of EcoRII to the 5′C of the CCTGG strand was not obtained suggesting that the contact of EcoRII to the bases of the recognition sequence is asymmetric. Tryptic digestion of free and of cross-linked EcoRII, followed HPLC separation of the individual peptides and Edman degradation identified amino acids 25–49 of EcoRII as the cross-linking peptide.

Mutational analysis of the electron-rich amino acids H36 and Y41 of this peptide indicated that Y41 is the amino acid involved in the cross-link and that it therefore contributes to specific DNA recognition by EcoRII. Other protein mutants K92E/K97E, R94A, E96A, R98A and R101A showed also severely impaired DNA binding\(^ {104,127}\). Therefore, it was concluded that Y41 and amino acids from binding site I is part of the DNA binding site of the effector N-domain which is responsible for the DNA recognition\(^ {127,132}\).

8. EcoRII dimer interface disruption by a single amino acid replacement

Interestingly, a single amino acid replacement of V258 by aspargine resulted in EcoRII dimer disruption into monomers\(^ {118}\). Analysis of the oligomerization state of mutant V258N showed a high percentage of protein monomers in solution. V258N showed 16% of wt cleavage activity, but bound
specific DNA with the same affinity as wt enzyme. It was suggested that V258 is located in the EcoRII dimer subunit interface.

9. Concluding remarks: EcoRII – archetypal Type IIE restriction enzyme

Summarizing numerous biochemical studies Reuter et al.\textsuperscript{103,104} suggested 2-site binding mechanism of EcoRII action which is very similar to another Type IIE restriction enzyme NaeI\textsuperscript{30,66,103,104,133-135} (Fig. 29).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure29.png}
\caption{Model of Type IIE restriction endonuclease EcoRII interaction with DNA. The two domains of a monomer are symbolized by a circle and a rectangle. Two monomer of homodimer are presented in white and black grid. Circle represents DNA binding domain; rectangle endonuclease domain. The lines between the rectangles symbolize dimerization contacts. Black lines with boxes represent DNA with specific DNA recognition sites. DNA first binds to the DNA binding domains of the dimer. In a second event, the DNA binding site in the endonuclease domains becomes occupied. The enzyme forms a DNA loop with DNA containing two recognition sites. The DNA recognition site bound to the endonuclease domains will be cleaved. Figure adapted from\textsuperscript{104}.

According this mechanism EcoRII DNA binding domain initially binds to specific DNA. The binding of the activator site changes the protein conformation and promotes the binding of the substrate site by the endonuclease-like domain. Thus, the complex of EcoRII dimer with two DNA recognition sites is the active form of the enzyme. In the active complex, the DNA recognition site, which is bound to the endonuclease-like domains of the dimer, is the substrate site that will be cleaved.}
MATERIALS AND METHODS

V. Materials

1. Chemicals

All chemicals used in this study were of the highest quality available.

2. Enzymes

T4 polynucleotide kinase, DNA polymerases TaqI and Pfu, calf intestine alkaline phosphatase, T4 DNA ligase, bovine serum albumin and all REases used for DNA manipulations were obtained from Fermentas UAB (Vilnius, Lithuania). BSA for Bradford assay was obtained from Pierce (Rockford, USA) Thermolysin was purchased from Serva (Heidelberg, Germany) or Sigma (Taufkirchen, Germany). All these products were used according to the manufacturer's instructions. Recombinant preparation of restriction enzyme EcoRII-C was provided by M. Reuter (Institute of Virology, Berlin, Germany).

3. *E. coli* strains

*E. coli* JM109:

\[ F^+ \text{traD36 proA}^+ B^+ \text{lacI}^+ \Delta(lacZ)M15 \ / \ e14^- (McrA^-) \Delta(lac-proAB) \text{endA1 gyrA96 (Nal}^r) \text{thi-1 hsdR17 (r}^K_m^+ \text{glnV44 relA1 recA1} \]

*E. coli* BL21 (DE3):

\[ F^- \text{ompT gal [dcm] [lon] hsdR}^B (r^B_m^B) \text{; an E. coli B strain) \lambda DE3} \]

*E. coli* HMS174 (DE3):

\[ F^- \text{recA1 hsdR(r}_{K12}^+ m_{K12}^-) \text{ Rif}^\lambda \text{DE3} \]

*E. coli* ER2566 (DE3):

\[ F^- \lambda^- \text{fhuA2 [lon] ompT lacZ::T7 geneI gal sulA11 \Delta (mcrC-mrr)114::IS10 R (mcr-73::miniTn10)2 R(zgb-210::Tn10)1 (Tet}^r) \text{endA1 [dcm] \lambda DE3} \]
E. coli GM31:

\( \text{F}^{-} \text{dcm}^{-6} \text{thr}^{-1} \text{araC14 leuB6 fhuA31 lacY1 tsx}^{-78} \text{glnV44 galK2 galT22 hisG4 rpsL136 (Str}'^{\text{r}} \text{) xylA5 mtl}^{-1} \text{thi}^{-1} \)

4. DNA

\( \lambda \text{dam} \text{dcm}^{-} \) DNA, pUC19 dam dcm, pUC19 and pBluescript IIKS (+) plasmids were obtained from Fermentas UAB (Vilnius, Lithuania). pET21b (+) plasmid was obtained from Novagen (Madison, USA). pKpn2I-ori plasmid was a kind gift from S. Jurenaite-Urbanaviciene (Institute of Biotechnology, Vilnius, Lithuania) and A. Lubys (Fermentas UAB, Vilnius, Lithuania). pHSG415\text{ts}_{\text{wt}}\_M.Ecl18kI (Cm\text{'}) , pUC129\_wt\_R.Ecl18kI (Ap\text{'}) and pQE30\_wt\_R.Ecl18kI-H\text{6} (Ap\text{'}) Ecl18kI expression plasmids were provided by A.S. Solonin (Institute of Biochemistry and Physiology of Microorganisms, Pushchino, Russia). pDK1\_wt\_M.EcoRII (Cm\text{'}) , pQE30\_wt\_R.EcoRII-H\text{6} (Ap\text{'}) , pQE30\_R.EcoRII-N-H\text{6} (Ap\text{'}) , pQE30\_Y41A\_R.EcoRII-H\text{6} (Ap\text{'}) , pQE30\_E271A\_R.EcoRII-H\text{6} (Ap\text{'}) , pQE30\_D299A\_R.EcoRII-H\text{6} (Ap\text{'}) , pDK11\_wt\_R.M.EcoRII (Ap\text{'}) EcoRII expression plasmids and pVH1 (Kn\text{'}) (bearing \text{lac}I\text{q} gene) plasmid were donated by M. Reuter (Institute of Virology, Berlin, Germany). pACYC184\_wt\_M.PspGI (Cm\text{'}) and pET21a (+)_wt \_R.PspGI (Ap\text{'}) PspGI expression plasmids were donated by New England Biolabs (Ipswich, USA).

5. Oligonucleotides

All non-modified oligodeoxynucleotides were purchased from Metabion (Martinsried, Germany) and 2-aminopurine containing oligodeoxynucleotides were obtained from Integrated DNA Technologies (HPLC grade, Coralville, USA).

All double-strand oligonucleotides used in this study and their short descriptions are given in Table 3. Oligoduplexes were made by annealing two oligodeoxyribonucleotides with complementary sequences.
Table 3. Oligoduplexes used in this study*

<table>
<thead>
<tr>
<th>Oligoduplex</th>
<th>Sequence</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP16a</td>
<td>5’-CGTAGCCTGGGGCTAG-3’</td>
<td>16 bp cognate oligoduplex for Ecl18kI gel shift analysis</td>
</tr>
<tr>
<td></td>
<td>3’-GCATCGGACCACCAGTC-5’</td>
<td></td>
</tr>
<tr>
<td>NSP16</td>
<td>5’-AGCGTAGCCTGGGCTCCT-3’</td>
<td>16 bp non-cognate oligoduplex for Ecl18kI gel shift analysis</td>
</tr>
<tr>
<td></td>
<td>3’-TCGCATCGGTACCCGGA-5’</td>
<td></td>
</tr>
<tr>
<td>SP16b</td>
<td>5’-CGTAGCCTGGTCGATC-3’</td>
<td>16 bp cognate oligoduplex for EcoRII-N gel filtration analysis</td>
</tr>
<tr>
<td></td>
<td>3’-GCATCGGACCACGCT-5’</td>
<td></td>
</tr>
<tr>
<td>SP12</td>
<td>5’-TAGCCTGGTCAGA-3’</td>
<td>12 bp cognate oligoduplex for EcoRII-C gel filtration analysis</td>
</tr>
<tr>
<td></td>
<td>3’-ATCGGACCAGCT-5’</td>
<td></td>
</tr>
<tr>
<td>SP33</td>
<td>5’-AATGGGCTCGACCGCTCTTTGTA-3’</td>
<td>33 bp cognate oligoduplex for EcoRII gel shift analysis and plasmid cleavage stimulation in trans</td>
</tr>
<tr>
<td></td>
<td>3’-TTACCCGAGCGTGGGACCATAATAGCT AACAAT-5’</td>
<td></td>
</tr>
<tr>
<td>NSP33</td>
<td>5’-AATGGGCTCGACCGCTCTTTGTA-3’</td>
<td>33 bp non-cognate oligoduplex for EcoRII gel shift analysis and plasmid cleavage stimulation in trans</td>
</tr>
<tr>
<td></td>
<td>3’-TTACCCGAGCGTGGGACCATAATAGCT AACAAT-5’</td>
<td></td>
</tr>
<tr>
<td>NSP25/28</td>
<td>5’-CGCACGCCTTGACCAAGAGCAGCT-3’</td>
<td>25/28 bp non-cognate oligoduplex for MvaI, PspGI EcoRII-C, Ecl18kI W61A, W61Y, wt gel shift analysis</td>
</tr>
<tr>
<td></td>
<td>3’-GCGTGCTGAACAGTTCGTGGTG-5’</td>
<td></td>
</tr>
<tr>
<td>T/A</td>
<td>5’-CGCACGCCCTTCCTGGAAGACGACACTA-3’</td>
<td>25 bp cognate oligoduplex without 2-AP for 2-AP fluorescence studies</td>
</tr>
<tr>
<td></td>
<td>3’-GCGTGCCGGAGGACCTCGTGGTGAT-5’</td>
<td></td>
</tr>
<tr>
<td>T/2</td>
<td>5’-CGCACGCCCTTCTTGGGAAGCAGACACTA-3’</td>
<td>25 bp cognate oligoduplex with 2-AP for 2-AP fluorescence studies</td>
</tr>
<tr>
<td></td>
<td>3’-GCGTGCGGAGG2CTCGTGGTGAT-5’</td>
<td></td>
</tr>
<tr>
<td>2FS</td>
<td>5’-CGCACGCCCTTTCTTGGAAGACGACACTA-3’</td>
<td>25 bp 2-AP-position-non-cognate oligoduplex for 2-AP fluorescence studies</td>
</tr>
<tr>
<td></td>
<td>3’-GCGTGCGGAGG2CGACCTCGTGGTGAT-5’</td>
<td></td>
</tr>
</tbody>
</table>

* Ecl18kI, EcoRII, PspGI and MvaI recognition site is in boldface; 2 – 2-aminopurine is in boldface and underlined. Oligoduplex 2FS contains 2-aminopurine introduced immediately adjacent to the target site.

Oligonucleotides used in DNA binding studies were labeled on 5’-end. Radioactive labels were introduced into the 5’-ends of individual DNA strands prior to the annealing with unlabelled strands. 5’-labelling of the top strand was performed with \([^{33}\text{P}]\) ATP (Hartmann Analytic, Braunschweig, Germany) and DNA labeling kit (Fermentas UAB, Vilnius, Lithuania). Labeling reactions
were performed according to the manufacturer's instructions except that DNA was kept in ~2-4 fold excess over the radioactive nucleoside triphosphates.

For the annealing oligoduplexes, appropriate oligodeoxynucleotides (Table 3) were mixed with a 1.05-1.1-fold molecular excess of complementary strands in the Reaction buffer III, heated to 95°C (85°C for modified oligonucleotides) and allowed to cool slowly over several hours to room temperature.

6. Buffers

Lysis buffer I: 20 mM Tris-HCl (pH 7.9 at 25°C), 5 mM imidazole, 500 mM NaCl.

Equilibration buffer I: 10 mM K-phosphate (pH 7.4 at 25°C), 1 mM EDTA, 7 mM 2-mercaptoethanol.

Equilibration buffer II: 20 mM Tris-HCl (pH 8.0 at 25°C), 100 mM KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 10% glycerol.

Storage buffer I: 10 mM K-phosphate (pH 7.4 at 25°C), 100 mM KCl, 2 mM DTT, 0.1 mM EDTA, 50% glycerol.

Storage buffer II: 10 mM Tris-HCl (pH 7.4 at 25°C), 200 mM KCl, 1 mM EDTA, 1 mM DTT, 50% glycerol.

Storage buffer III: 10 mM Tris–HCl (pH 7.4 at 25°C), 200 mM KCl, 1 mM DTT, 1 mM EDTA, 50% glycerol, 0.1% triton X-100.

CD buffer I: 30 mM Tris-HCl (pH 7.0 at 25°C), 150 mM NaCl.

CD buffer II: 10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 10% glycerol, 0.01% triton X-100.

Proteolysis buffer: 10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 1 mM DTT.

Loading dye solution I: 75 mM EDTA (pH 9.0 at RT), 0.01% bromphenol blue in 50% (v/v) glycerol.

Loading dye solution II: 25 mM EDTA (pH 9.0 at RT), 0.01% (w/v) bromphenol blue in 95% (v/v) formamide.
**Reaction buffer I:** 33 mM Tris-acetate (pH 7.9 at 15°C), 66 mM K-acetate.
**Reaction buffer II:** 33 mM Tris-acetate (pH 7.9 at 37°C), 66 mM K-acetate.
**Reaction buffer III:** 33 mM Tris-acetate (pH 7.9 at 25°C), 66 mM K-acetate.

**Electrophoresis buffer:** 40 mM Tris-acetate (pH 8.3 at 25°C).

**Gel filtration buffer I:** 100 mM Na-phosphate (pH 6.7 at 25°C), 100 mM Na-sulphate.

**Gel filtration buffer II:** 20 mM Tris-HCl (pH 7.5 at 25°C), 200 mM KCl, 10% glycerol, 5 mM CaCl$_2$.

**DNA elution buffer I:** 10 mM Tris-HCl (pH 7.8 at 25°C).

**DNA elution buffer II:** 10 mM HEPES (pH 7.4 at 25°C).

**Binding buffer:** 40 mM HEPES (pH 8.4 at 25°C).

---

**VI. Methods**

1. **Electrophoresis**

1.1. *Denaturing (SDS) polyacrylamide gel electrophoresis of proteins*

Denaturing SDS-PAGE of proteins was employed to verify homogeneity of protein preparations (see sections VI.2-4.) and resolve proteolytic fragments of EcoRII (see section VI.10.). The electrophoretic buffer consisted of 25 mM Tris, 190 mM glycine (pH 8.3 at RT) and 0.1% SDS (w/v). The stacking and separating gels consisted of 4% (in 125 mM Tris-HCl (pH 6.8 at RT) and 0.1% SDS (w/v)) and 12% (in 375 mM Tris-HCl (pH 8.8 at RT) and 0.1% SDS (w/v)) acrylamide/N,N’-methylenebisacrylamide (37.5:1 (w/w)), respectively$^{136}$. Protein samples were mixed 1:1 (v/v) with sample buffer (100 mM Tris-HCl (pH 6.8 at RT), 4% SDS (w/v), 20% (v/v) glycerol, 200 mM DTT, trace of Bromphenol blue) and placed for 5 min into a bath with 95 °C water before loading. Electrophoresis was run at room temperature for 1-1.5 hours at 25 V/cm. Gels were stained with Coomasie Blue.

Digital images of the gels were taken with BioDocAnalyze (*Biometra, Goettingen, Germany*) gel documenting system. Amounts of protein were
quantified by densitometric analysis using 1-D Main software (*Advanced American Biotechnology, Fullerton, USA*).

1.2. Non-denaturing electrophoresis through agarose

Separation of DNA fragments (see section VI.12.) and separation of supercoiled (SC), open circular (OC) and linear forms of plasmid DNA (FLL, L2, L3) (see section VI.16.) were performed in 0.8-1.2% agarose gels in the electrophoretic buffer containing 100 mM H$_3$BO$_3$-NaOH, 15 mM sodium acetate, 2 mM EDTA (pH 8.2 at RT) and 0.5 µg/ml ethidium bromide. DNA samples were mixed with 1/3 volume of Loading dye solution I and electrophoresed at 3 V/cm until the bromphenol blue dye migrated approximately for 3 cm. Gels were dried, digital images of the gels were taken and amounts of DNA were quantified as described above (see section VI.1.1.)

DNA fragments required for genetic engineering procedures (see sections VI.2.2. and VI.3.2.), AFM (see section VI.17.) and SC form of the plasmid for EcoRII cleavage reactions (see section VI.16.1.) were separated in 0.8-2.0% agarose gels in the ethidium bromide-free electrophoretic buffer containing 40 mM Tris-acetate and 1 mM EDTA (pH 8.3 at RT). The gel slices containing required DNA were excised according to the ethidium bromide stained markers. DNA was recovered by phenol extraction: the gel slices were chopped up and an equal volume of phenol saturated with pH 8.0 buffer was added. After rigorous mixing, the samples were kept for 30 min at -70 °C and centrifuged at RT for 15 min at 13000 g. The DNA-containing aqueous phase was removed and extracted twice with CHCl$_3$. DNA was precipitated by adding 1/7.7 volume of 3 M Na-acetate (pH 7.0 at 25 °C) and equal volume of 2-propanol. Then DNA was washed using 70% (v/v) ethanol solution, dried incubating for 5-7 min at 37 °C and dissolved in water.

1.3. Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis was employed in gel-shift experiments (see section VI.11.). The electrophoretic buffer consisted
of 40 mM Tris-acetate (pH 8.3 at 25°C) and 5-10 mM Ca-acetate or 0.1 mM EDTA. The gels consisted of 6-8% acrylamide/N,N’-methylenebisacrylamide (29:1 (w/w)) in the electrophoretic buffer, polymerisation was initiated by adding TEMED and ammonium persulphate. The gels were 1 mm thick and ~20 cm in length. Prior to gel casting, one of the glass plates was processed with “bind silane” (3-methacryloxypropyltrimethoxysilane) and the other with “repeal silane” (5-7% (v/v) dichlorodimethylsilane in CHCl₃). Samples of radiolabeled DNA, binding buffer and protein were mixed in the wells of ELISA plate, left for 10 min at room temperature and then loaded on the gel. Electrophoresis was run at room temperature for 2-3 hours at ~6 V/cm.

After electrophoresis the glass plate with “repeal silane” was removed and the gel was dried on the glass plate with “bind silane” under a hot air flow. Radiolabeled DNA was detected in the dried gels using BAS-MS image plates (FujiFilm, Kanagawa, Japan) and Cyclone™ phosphorimager (Perkin Elmer, Massachusetts, USA). The amounts of various DNA fragments were quantified with OptiQuant 3.0 software (Perkin Elmer, Massachusetts, USA).

1.4. Denaturing polyacrylamide gel electrophoresis

Denaturing PAGE was employed in oligonucleotide DNA cleavage studies (see section VI.12.3.). The electrophoretic buffer consisted of 100 mM Tris-borate (pH 8.2 at RT) and 2 mM EDTA. The gels consisted of 20% acrylamide/ N,N’-methylenebisacrylamide (29:1 (w/w)) and 7 M urea in the electrophoretic buffer.

Samples of radiolabeled DNA were mixed 1:1 (v/v) with Loading dye solution II, placed for 5 min into a bath with 95 °C water and chilled on ice. Electrophoresis was run for 2 hours at 30 V/cm without the samples and for another 2-3 hours with the DNA samples. Gels were dried and analysed as described above (see section VI.1.3.).
2. Expression and purification of Ecl18kI

2.1. Ecl18kI plasmids and strains

For the characterization of wt Ecl18kI protein *E. coli* strain JM109 carrying the pHSG415ts_wt_M.Ecl18kI (Cm') and pUC129_wt_R.Ecl18kI (Ap')13 were used.

The *E. coli* strain JM109 and the plasmid pQE30_R.Ecl18kI-H6 (Ap') plasmid13 were used for construction and expression of (histidine)$_6$-tagged wt and E125A, D160A, K182A, R186A, E187A, R188A, E195A mutants Ecl18kI. A compatible plasmid, pHSG415ts_wt_M.Ecl18kI (Cm'), carried the gene for the Ecl18kI methyltransferase.

For the production of wt Ecl18kI for the crystallization and fluorescence experiments the ec18kIR gene was re-cloned from pUC129_wt_R.Ecl18kI (Ap') plasmid to the pET21b (+) vector and expressed in the BL21 (DE3) strain carrying the plasmid pVH1 [Knr] (with lacI$^{137}$) and plasmid pHSG415ts_wt_M.Ecl18kI (Cm') bearing the ec18kIM gene. Plasmid pET21b (+)_wt_R.Ecl18kI (Ap') was used for construction and expression W61Y and W61A mutants of Ecl18kI.

2.2. Site-directed mutagenesis of Ecl18kI

Site-directed mutagenesis of E125A, D160A, K182A, R186A, E187A, R188A and E195A Ecl18kI mutants was performed by the two-step ‘megaprimer’ method$^{138}$. The plasmid pQE30_wt_R.Ecl18kI-H$_6$ (Ap') was used as a template for PCR with a primer containing the appropriate mispair and a second primer upstream or downstream of the ec18kIR gene. The PCR products were purified and used in a second round of mutagenesis with a primer downstream or upstream of ec18kIR. The resultant product was cleaved by BamHI/HincII or BglII/HincII and ligated back into pQE30_wt_R.Ecl18kI-H$_6$ (Ap') that had been cleaved with the same enzymes. Competent host cells were prepared and transformed with heterologous DNA by the CaCl$_2$ method$^{136}$. Both strands of the entire region between
BamHI/HincII or BglII/HincII sites we re sequence d to confirm that the
designed amino acid change was the only change in the DNA sequence.

W61Y and W61A mutants of Ecl18kI without any tags were obtained by
modified QuickChange Mutagenesis protocol\textsuperscript{139}. pET21b (+)_wt _R.Ecl18kI (Ap\textsuperscript{r}) plasmid was amplified by PCR using Pfu polymerase 
\textit{(Fermentas UAB, Vilnius, Lithuania)} and two complementary (partially
overlapping) primers containing the desired mutation. After PCR methylated
parental non-mutated plasmid was digested with DpnI \textit{(Fermentas UAB, 
Vilnius, Lithuania)}. PCR product – circular, nicked ds DNA was transformed
into \textit{E. coli} BL21 (DE3), carrying the plasmid pVH1 [Kn\textsuperscript{i}] and plasmid
pHSG415\textsuperscript{158}_wt_M.Ecl18kI (Cm\textsuperscript{i}) bearing the \textit{ecl18kIM} gene, competent cells
by the CaCl\textsubscript{2} method\textsuperscript{136}. The plasmid DNA was isolated by the alkaline lysis
procedure and purified using phenol-chloroform extraction\textsuperscript{136}. Sequencing of
the entire genes of the mutants confirmed that only the designed mutations had
been introduced.

\textbf{2.3. Purification of Ecl18kI}

\textit{(Histidine)}\textsubscript{6}-tagged wt and mutant Ecl18kI variants were purified to
approximately 90\% purity on Ni-NTA \textit{(Novagen, Madison, USA)} and AH-
Sepharose \textit{(GE Healthcare, Uppsala, Sweden)} columns (all purification
procedures of all proteins were performed at 4\textdegree C). A 250 ml overnight culture
of \textit{E.coli} JM109 bearing the pQE30_R.Ecl18kI-H\textsubscript{6} (Ap\textsuperscript{i}) expression plasmid
was incubated at 37\textdegree C for 3.5 h after IPTG induction at final 0.4 mM
concentration. Cells were harvested by centrifugation at 2500 g for 15 min and
stored frozen. The frozen cell pellet was suspended in 10 ml of Lysis buffer
(20 mM Tris-HCl (pH 7.9 at 25\textdegree C), 5 mM imidazole, 500 mM NaCl),
sonicated 4 x 30 s and centrifuged for 30 min at 18 000 g to remove cell debris.
The supernatant was applied to a 2 ml Ni-NTA \textit{(GE Healthcare, Uppsala, 
Sweden)} column equilibrated with Lysis buffer and Ecl18kI endonuclease was
eluted using a linear 5-1000 mM imidazole gradient. The fractions containing
Ecl18kI enzyme were pooled, dialyzed against Equilibration buffer I (10 mM
K-phosphate (pH 7.4 at 25°C), 1 mM EDTA, 7 mM 2-mercaptoethanol) with 100 mM NaCl, and applied to a 2 ml AH-Sepharose (GE Healthcare, Uppsala, Sweden) column. The column was eluted using a linear 100-1000 mM NaCl gradient, fractions containing wt or mutant Ecl18kI were pooled and dialyzed against Storage buffer I (10 mM K-phosphate (pH 7.4 at 25°C), 100 mM KCl, 2 mM DTT, 0.1 mM EDTA, 50% glycerol) and stored at -20°C.

Ecl18kI for the wt protein characterization was expressed in *E. coli* JM109 cells containing the pUC129_wt_R.Ecl18kI (Ap⁵) plasmid in the same conditions as for (Histidine)₆-tagged Ecl18kI (see above). Protein was purified to homogeneity through subsequent chromatography on Heparin-Sepharose, Q-Sepharose, Blue-Sepharose (GE Healthcare, Uppsala, Sweden) and Hydroxylapatite (Calbiochem, Gibbstown, USA) columns and dialyzed against Storage buffer I. The protein was >95% homogeneous by sodium dodecyl sulphate-PAGE analysis.

Wt, W61Y and W61A Ecl18kI proteins for the crystallization and fluorescence experiments were expressed in *E. coli* BL21 (DE3) cells containing the pET21b (+)_R.Ecl18kI (Ap⁵) plasmid. Cells were cultured overnight at 30°C in LB medium supplemented with required antibiotics. The overnight culture was used to inoculate 400 ml of the fresh LB medium (1/20 (v/v)), bacteria were grown to the mid-log phase (absorbance at 600 nm 0.5 to 0.7), 1 mM IPTG was added and further cultured for another 4 h at 30°C. The cells were harvested and proteins were purified to homogeneity through subsequent chromatography on Phosphocellulose P11 (Whatman, Maidstone, England), Q-Sepharose, Bordo-Sepharose and AH-Sepharose (GE Healthcare, Uppsala, Sweden) columns. Ecl18kI was dialyzed against storage Storage buffer II (10 mM Tris-HCl (pH 7.4 at 25°C), 200 mM KCl, 1 mM EDTA, 1 mM DTT, 50% glycerol) and stored at -20°C. The proteins were >95% homogeneous by SDS-PAGE analysis. For the production of SeMet variant, the Ecl18kI was expressed in *E. coli* BL21 (DE3) cells containing the pET21b (+)_wt_R.Ecl18kI (Ap⁵) plasmid. Recombinant SeMet variant of
Ecl18kI was prepared by inhibiting the methionine synthesis pathway and expressing the protein in the presence of D,L-selenomethionine (Sigma, Taufkirchen, Germany) following the published procedure\textsuperscript{140}. SeMet variant of Ecl18kI was purified as wt and stored in Storage buffer II. Mass-spectrometry analysis of the purified SeMet variant of Ecl18kI showed that it contains all 12 methionines replaced with SeMet (G. Lukinavicius, data not shown).

The Ecl18kI proteins concentrations were determined by measuring the absorbance at 280 nm and using an extinction coefficients of 77,660 M\(^{-1}\)cm\(^{-1}\) for wt protein, 75,640 M\(^{-1}\)cm\(^{-1}\) for W61Y mutant and 72,660 M\(^{-1}\)cm\(^{-1}\) for W61A Ecl18kI. All extinction coefficients were calculated for Ecl18kI dimer.

3. Expression and purification of EcoRII

3.1. EcoRII plasmids and strains

*E. coli* strain JM109 (pDK1\_wt\_M.EcoRII (Cm\(^r\)), pQE30\_R.EcoRII-H\(_6\) (Ap\(^r\)))\textsuperscript{117} was used for cloning and expression of (histidine)_\(_6\)-tagged wt and mutants of EcoRII.

For the production of wt EcoRII, *E. coli* strain HM174 (DE3) bearing the plasmids pVH1 (Kn\(^r\)) and pDK11\_R.M.EcoRII (Ap\(^r\))\textsuperscript{137} were used.

3.2. Site-directed mutagenesis of EcoRII

The Y41A\textsuperscript{132}, E271A and D299A mutant plasmids were donated by M. Reuter (Institute of Virology, Berlin, Germany). Mutants K324A, K328A, D329A, R330A and E337A were obtained by a two-step ‘megaprimer’ method\textsuperscript{138} (see section VI.2.2.). Both strands of the *ecoRII* gene were sequenced to confirm that only the designed mutations had been introduced.

3.3. Purification of EcoRII

*E. coli* strain JM109 (pDK1\_wt\_M.EcoRII (Cm\(^r\)), pQE30\_R.EcoRII-H\(_6\) (Ap\(^r\))) was used for the protein expression\textsuperscript{117}. Cells carrying genes encoding EcoRII, EcoRII-N, E271A, D299A, K324A and E337A proteins were grown in Luria Broth medium as described\textsuperscript{117}. To increase protein yield M9 minimal
medium supplemented with 4 μg/ml thiamine, 100 μg/ml carbenicillin, 25 μg/ml kanamycin and 20 μg/ml of L-tryptophan, L-histidine and L-methionine was used for the growth of cells harboring K328A, D329A and R330A mutants.

All the (histidine)$_6$-tagged proteins were expressed and purified as described. The fractions containing EcoRII enzyme were pooled and dialyzed against the Storage buffer III (10 mM Tris–HCl (pH 7.4 at 25°C), 200 mM KCl, 1 mM DTT, 1 mM EDTA, 50% glycerol, 0.1% triton X-100) and stored at -20°C. All protein preparations were to >95% pure according to SDS–PAGE. Protein concentrations were measured by Bradford assay using bovine serum albumin (Pierce, Rockford, USA) for the calibration.

For expression of wt EcoRII protein, a single transformant containing the pDK11_R.M.EcoRII (Ap$^5$) plasmid in HMS174 (DE3, pVH1 (Kn$^r$)) cells was cultured overnight at 37°C in LB medium supplemented with carbenicillin (100 μg/ml). The overnight culture was used to inoculate 400 ml of the fresh LB medium (1/20 (v/v)), bacteria were grown to the mid-log phase (absorbance at 600 nm 0.5 to 0.7), 1 mM IPTG was added and further cultured for another 5 h. The cells were harvested, suspended in Equilibration buffer I supplement with 100 mM NaCl, 10% (v/v) glycerol and 0.1% (v/v) triton X-100, disrupted by sonication and cell debris removed by centrifugation. The supernatant was purified through subsequent chromatography on Phosphocellulose P11 (Whatman, Maidstone, England), Heparin-Sepharose and AH-Sepharose (GE Healthcare, Uppsala, Sweden) columns by elution with NaCl gradients. The fractions containing EcoRII enzyme were pooled and dialyzed against the storage Storage buffer III and stored at -20°C. The protein preparation was to >95% pure according to SDS-PAGE. The wt EcoRII concentration was measured by Bradford assay (see above).

For kinetic studies recombinant preparation of restriction enzyme EcoRII-C was provided by M. Reuter (Institute of Virology, Berlin, Germany).
For gel filtration and 2-AP fluorescence studies EcoRII-C protein was obtained by thermolysin digestion of the Y41A mutant of EcoRII which possesses low cleavage activity. One milligram of Y41A EcoRII was digested for 3 h at 25°C with 0.1 mg of thermolysin (Sigma, Taufkirchen, Germany) in a final volume of 2 ml (as described in section VI.10.). The solution was loaded onto a HiTrap Heparin column (GE Healthcare, Uppsala, Sweden) pre-equilibrated with Equilibration buffer II (20 mM Tris-HCl (pH 8.0 at 25°C), 100 mM KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 10% glycerol) and protein was eluted from the column with a linear KCl gradient. The fractions were analyzed with a λ*dam−dcm* DNA (Fermentas UAB, Vilnius, Lithuania) DNA cleavage assay that monitored the EcoRII-C activity which is unleashed during the proteolysis and SDS-PAGE. Fractions containing EcoRII-C were pooled, dialyzed against Storage buffer II and stored at -20°C. SDS-PAGE analysis revealed that the sample contained >95% of the C-terminal domain and showed no traces of the full-length protein. The concentration of EcoRII-C was determined by measuring the absorbance at 280 nm using an extinction coefficient of 43 240 M⁻¹cm⁻¹ calculated for the dimer.

4. Expression and purification of PspGI and MvaI

4.1. Expression and purification of PspGI

The wt PspGI protein was expressed in E. coli ER2566 (DE3, pACYC184_wt_M.PspGI (Cm⁵) and pET21a (+)_wt_R.PspGI (Ap⁴)) and was purified as described. Fractions containing PspGI were pooled, dialyzed against Storage buffer II and stored at -20°C. PspGI preparation was to >95% pure according to SDS–PAGE. The concentration of PspGI was determined by measuring the absorbance at 280 nm using an extinction coefficient of 50 920 M⁻¹cm⁻¹ calculated for the dimer.
4.2. Expression and purification of MvaI

MvaI REase was cloned and purified as described\textsuperscript{35}.

5. Characterization of R.Ecl18kI protein

Sequencing of plasmid pUC129_wt_R.Ecl18kI (Ap\textsuperscript{r}) bearing the \textit{ecl18kIR} gene revealed three nucleotide mutations in comparison to the \textit{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 CGG). PSI-BLAST search reveals that our Ecl18kI variant is 100\% identical to the StyD4I protein sequence except that StyD4I has an additional N-terminal extension of 13 aa\textsuperscript{16}. Sequence analysis of the \textit{ecl18kIR} gene reveals two putative starts of translation corresponding to M1 and M14 residues of the deduced protein sequence. Translation from the M1 codon would result in the R.Ecl18kI protein of the same length as StyD4I. N-terminal sequencing of the purified R.Ecl18kI protein (data not shown), however, indicated that translation is started from the codon of M14 residue yielding a protein of 305 aa residues in length. The mass-spectrometry analysis of purified R.Ecl18kI protein yielded a molecular weight value of 35 892Da (V. Luksa, data not shown) that is very close to the value of 35 897 Da calculated for the 305 aa protein.

6. Ecl18kI mass-spectrometry analysis

The mass-spectrometry analysis of purified wt and SeMet variant of Ecl18kI proteins to estimate the molecular weight was done by V. Luksa (\textit{Sicor Biotech UAB, Vilnius, Lithuania}) and G. Lukinavicius (\textit{Institute of Biotechnology}) using Hewlett-Packard 1100 series single quadruple system mass spectrometer with electrospray ionization.
7. Protein sequence analysis

The theoretical pI, molecular mass and extinction coefficients for all proteins were determined with the ProtParam tool at http://www.expasy.ch/tools/protparam.html.

FASTA pairwise similarity search between protein sequence of Ecl18kI and other restriction enzymes revealed weak local similarities to Cfr10I endonuclease. Using Multalin sequence alignment server (http://prodes.toulouse.inra.fr/multalin/multalin.html) we were able to extend the initial pairwise alignment between Ecl18kI and Cfr10I to a larger set of REases specific for CCNGG/CCWGG or RCCGGY/GCCGGC nucleotides. In the initial alignment generated automatically by Multalin structurally conserved PD residues at the active sites of Bse634I, Cfr10I and NgoMIV, however, were misaligned (probably due to the different number of amino acid in spacings between the conserved regions) and therefore were realigned manually. Analysis of highly diverged group of nucleases identified conserved VD residues in Ecl18kI, PD residues in EcoRII and PD residues in PspGI sequences, respectively, as homologues to the PD residues from the PD|55KX13E active site motif of Cfr10I. Therefore, we manually aligned those residues of Ecl18kI, EcoRII and PspGI with structurally conserved PD residues of Bse634I, Cfr10I and NgoMIV generating the final form of Fig. 30.

8. Circular Dichroism spectroscopy

To demonstrate that wt and mutant proteins of Ecl18kI and EcoRII has no differences in secondary structure CD spectra of the proteins were recorded. Circular dichroism spectra were recorded on Jasco J-710 spectropolarimeter (Tokyo, Japan) using a cylindrical cuvette with a 0.5 mm light-path length at 17°C. Three measurements between 180 to 260 nm were performed and data averaged. CD spectra of Ecl18kI mutant proteins were recorded in CD buffer I (30 mM Tris-HCl (pH 7.0 at 25°C), 150 mM NaCl) and EcoRII mutants were
recorded in CD buffer II (10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 10% glycerol, 0.01% triton X-100).

9. Analytical ultracentrifugation

For determination of molecular masses in solution, sedimentation equilibrium experiments in the analytical ultracentrifuge were exploited. Analytical ultracentrifugation of Ecl18kI was done by C. Urbanke in Hanover Medicinische Hochschule (Germany) in an An-60 rotor with a Coulter-Beckman Model XL-A analytical ultracentrifuge equipped with UV absorption detection using charcoal filled epon centerpieces. Molar masses was evaluated when the sedimentation equilibrium was reached as described\textsuperscript{28}. Analytical ultracentrifugation revealed that in solution wt R.Ecl18kI is a dimer (C. Urbanke, data not shown).

10. EcoRII limited proteolysis

The EcoRII-C and EcoRII-C mutants were obtained by thermolysin (Serva, Heidelberg, Germany or Sigma, Taufkirchen, Germany) digestion of 0.5 mg/ml (5.4 µM in terms of dimer) protein solution at 25°C in Proteolysis buffer (10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 1 mM DTT) containing 20–45% glycerol, 0.015–0.1% triton X-100 and 2 mM Ca-acetate. EcoRII: thermolysin ratio was 10:1 (w/w) for wt, E271A, D299A, K324A, E337A and 2.5:1 (w/w) for K328A, D329A, R330A. Reactions were terminated by adding EDTA to 9 mM and heating the samples for 3 min at 40°C.

The proteolysis mix contained >90% EcoRII-C according to SDS–PAGE analysis and was used in DNA binding and gel filtration experiments without further purification. The EcoRII-C concentration in the proteolysis mix was evaluated by densitometric analysis of wt EcoRII standard and EcoRII-C bands in the Coomassie stained SDS–PAGE gel. The N-terminus sequence analysis was performed in the ZMMK-Servicelabor (University of Cologne, Germany).
11. Gel mobility shift assay

11.1. DNA binding of wt Ecl18kI and mutant proteins

For DNA binding experiments with (histidine)$_6$-tagged wt Ecl18kI and E125A, D160A, K182A, R186A, E187A, R188A, E195A Ecl18kI mutants the specific SP16a and non-specific NSP16 oligoduplexes (Table 3) were used in final concentration of 1 nM. DNA was incubated with varying amounts of Ecl18kI for 10 min at room temperature in the Reaction buffer II (33 mM Tris-acetate (pH 7.9 at 37°C), 66 mM K-acetate) containing 0.1 mg/ml BSA, 10% glycerol and 0.1 mM EDTA or 10 mM Ca-acetate.

11.2. DNA binding of wt EcoRII and mutant proteins

For DNA binding experiments with wt EcoRII, EcoRII-N, EcoRII-C and EcoRII-C mutants the concentration of $^{33}$P-labelled specific SP33 and non-specific NSP33 (Table 3) oligoduplexes was kept at 1 nM.

For DNA binding analysis of the wt EcoRII and EcoRII-N, different amounts of the recombinant wt EcoRII or EcoRII-N were incubated for 10 min at room temperature with DNA in 20 µl of Electrophoresis buffer containing 0.1 mg/ml BSA, 10% glycerol and 0.1 mM EDTA or 5 mM of Ca-acetate.

For DNA binding analysis of EcoRII-C and EcoRII-C mutants, different amounts of proteolysis mixture (stopped with adding EDTA without heating and without any further purification, see section VI.10.) were incubated at room temperature for 5 min with DNA in 20 µl of Proteolysis buffer (10 mM Tris–HCl (pH 7.5), 200 mM KCl, 1 mM DTT) containing 10% glycerol, 0.1 mg/ml BSA with or without 10 mM of Ca-acetate.

11.3. DNA binding of wt, W61Y and W61A Ecl18kI, EcoRII-C, PspGI and MvaI

For DNA binding experiments with wt, W61Y and W61A Ecl18kI, EcoRII-C, PspGI and MvaI with oligoduplexes T/A, T/2, NSP and NSP25/28 (Table 3) $^{33}$P-labelled DNAs were used in final concentration of 0.1 nM.
Different amounts of the proteins were incubated for 10 min at room temperature with DNA in 20 µl of Electrophoresis buffer containing 0.1 mg/ml BSA, 10% glycerol and 0.1 mM EDTA or 5 mM of Ca-acetate.

11.4. Gel mobility shift assay

Free DNA and protein-DNA complexes were separated through non-denaturing 6% or 8% PAAG (29:1, acrylamide/N,N’-methylene-bisacrylamide) in Electrophoresis buffer containing 0.1 mM EDTA or 5-10 mM Ca-acetate. Electrophoresis was run at room temperature for 2-3 hours at ~6 V/cm.

11.5. Dissociation constant $K_d$ calculation

After electrophoresis gels were dried and analyzed on Cyclone (PerkinElmer, Downers Grove, USA) phosphorimager. The amounts of complexed and non-complexed DNA were quantified with OptiQuant 3.0 software (PerkinElmer, Downers Grove, USA). The $K_d$ values for cognate oligoduplex were calculated by fitting the quantified data to equation:

$$y = \{s_0 - x - K_d + \left[(s_0 + x + K_d)^2 - 4s_0x\right]^{0.5}\}/2 \quad (1)$$

where $y$ is the free DNA concentration (in terms of nM) at each protein concentration $x$, $s_0$ is the DNA concentration in binding mixture, and $K_d$ is the dissociation constant of protein–DNA complex. Data analysis used the KyPlot 2.0 software[^144].

12. DNA cleavage assay

12.1. DNA cleavage assay of Ecl18kI mutant proteins

Cleavage activity of Ecl18kI active site and DNA binding mutants was monitored under single turnover conditions using supercoiled pUC19 dam $\overline{dcm}$ plasmid (Fermentas UAB, Vilnius, Lithuania) containing 12 Ecl18kI recognition sites. Wt Ecl18kI or mutant proteins at 75 nM (in terms of dimer) concentration were mixed with 2.5 nM of pUC19 in the Reaction
buffer I (33 mM Tris-acetate (pH 7.9 at 15°C), 66 mM K-acetate) containing 0.1 mg/ml BSA and the reaction was initiated by adding Mg-acetate to give a final concentration of 10 mM. Aliquots were removed after fixed time intervals and mixed with Loading dye solution I (containing EDTA) with 0.3% SDS. Cleavage experiments were performed at 15°C to make the reaction rates slow enough to collect samples manually in the case of wt Ecl18kI. 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. 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.

12.2. DNA cleavage assay of EcoRII mutant proteins

The DNA cleavage activity of active site and DNA binding mutants of EcoRII (E125A and E187A mutants of Ecl18kI as well) was determined by incubating different amounts of protein in 50 µl Reaction buffer II (33 mM Tris–acetate (pH 7.9 at 37°C), 66 mM K-acetate) containing 10 mM Mg-acetate, 0.1 mg/ml BSA and 1 µg phage λ dam dcm DNA. The reactions were carried out for 1-16 hours at 37°C. The reactions were stopped by adding 25 µl of Loading dye solution I containing EDTA and analyzed by electrophoresis using a 1.0% (w/w) agarose gel. Gels were visualized under UV light and the completeness of the λ DNA cleavage evaluated visually. One unit of restriction enzyme activity is the amount of protein required to completely hydrolyze 1 µg of λ DNA in 60 minutes in a total reaction volume of 50 µl. Wt EcoRII had a specific activity of $2 \times 10^5$ units/mg.
12.3. DNA cleavage assay of wt, W61Y and W61A Ecl18kI, EcoRII-C and PspGI

The wt Ecl18kI, W61Y and W61A Ecl18kI, EcoRII-C and PspGI ability to cleave 2-AP containing oligoduplex substrates a 25 bp oligoduplexes T/A, T/2 and 2FS containing a $^{33}$P-label either in the top or the bottom DNA strand were used (Table 3). Cleavage rates of both strands were evaluated separately. Ecl18kI cleavage reactions were conducted at 20°C in the Reaction buffer III (33 mM Tris-acetate, pH 7.9 at 25°C, 66 mM K-acetate) containing 10 mM Mg-acetate and 0.1 mg/ml BSA using 200 nM of oligoduplex and 300 nM of protein. EcoRII-C and PspGI cleavage reactions were performed in the same Reaction buffer III at 25°C using 200 nM of oligoduplex and 1000 nM of protein. Aliquots were removed at timed intervals and quenched by mixing with Loading dye solution II before denaturing gel electrophoresis. The samples were analyzed and quantified as described in\textsuperscript{146}.

13. Structure solution of Ecl18kI-DNA complex

The Ecl18kI-DNA complex structure was solved by the multiple anomalous diffraction (MAD) technique using the selenomethionine variant of the protein and a KBr soak at 1.7 Å resolution by the M. Bochtler group (International Institute of Molecular and Cell Biology, Warsaw, Poland). Required oligoduplexes were purchased from Metabion (HPLC grade, Martinsried, Germany):

\begin{align*}
5'\text{--CGCCAGGGC--3'} & \quad 5'\text{--CGCCAGGGC--3'} \\
3'\text{--GCGGTCCCCG--5'} & \quad 3'\text{--GCGGXCCCG--5'},
\end{align*}

where X is 5-iododeoxyuridine.

Coordinates and structure factors have been submitted to the RCSB Protein Data Bank with Accession Codes 2FQZ (2.0Å resolution data) and 2GB7 (1.7Å resolution data).
14. Fluorescence spectroscopy of Ecl18kI, EcoRII-C and PspGI

All fluorescence measurements were acquired in photon counting mode on a Fluoromax-3 (Jobin Yvon, Stanmore, UK) spectrofluorometer equipped with Xe lamp. Sample temperatures were maintained at 25°C by a circulating water bath. Oligoduplexes T/2 or 2FS were used as 2-AP labeled DNA (Table 3). Emission spectra (340–420 nm) were recorded at an excitation wavelength $\lambda_{\text{ex}} = 320$ nm with excitation and emission bandwidths of 2 and 8 nm, respectively. At least two scans were averaged for each spectrum. Sample emission spectra were collected in Reaction Buffer III in the presence and absence of 10 mM Ca-acetate on 250 nM DNA alone or 250 nM DNA mixed with a 5-fold excess of the protein to ensure saturation of the fluorescence signal. Control spectra used for the background subtraction corrections were collected under identical conditions except that oligoduplex T/A was used instead of the fluorescent DNA (Table 3). The fluorescence emission value of the corrected spectrum was determined at the emission maximum (Table 4) for each sample. For the oligoduplex titration experiment, emission spectra of the 250 nM oligoduplex T/2 with protein in a 0-2000 nM range were collected.

<table>
<thead>
<tr>
<th>Oligoduplex</th>
<th>$\lambda_{\text{em}}^\text{max} , \text{nm}$ in the absence of Ca$^{2+}$</th>
<th>$\lambda_{\text{em}}^\text{max} , \text{nm}$ in the presence of Ca$^{2+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T/2</td>
<td>370</td>
<td>370</td>
</tr>
<tr>
<td>T/2 + Ecl18kI wt</td>
<td>367</td>
<td>367</td>
</tr>
<tr>
<td>T/2 + Ecl18kI W61Y</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>T/2 + Ecl18kI W61A</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>T/2 + EcoRII-C</td>
<td>370</td>
<td>360</td>
</tr>
<tr>
<td>T/2 + PspGI</td>
<td>370</td>
<td>360</td>
</tr>
<tr>
<td>T/2 + MvaI</td>
<td>370</td>
<td>370</td>
</tr>
</tbody>
</table>

* Fluorescence emission maxima of oligoduplex NSP and oligoduplex NSP+protein corrected spectra are 370 nm. Table was taken from$^{147}$. 


15. Gel filtration of EcoRII

Gel filtration experiments of wt EcoRII, EcoRII-N, EcoRII-N-DNA, EcoRII-C wt and mutants were carried out at room temperature on Waters Breeze (*Waters, Milford, USA*) HPLC system using a TSK-GEL Super SW2000 column (*Tosoh Bioscience, Stuttgart, Germany*) pre-equilibrated with Gel filtration buffer I (100 mM Na-phosphate (pH 6.7 at 25°C), 100 mM Na-sulphate). The samples of EcoRII-N (1.5 µM in terms of monomer), wt EcoRII (2 µM dimer) and DNA were prepared in 20 µl of the same buffer. The EcoRII-N-DNA complexes were prepared by mixing 30 pmol of EcoRII-N with different amounts of specific SP16b oligoduplex (Table 3) to give molar ratios of monomer:duplex at 2:1 and 1:2, respectively. For oligomerization state analysis of the wt EcoRII-C and EcoRII-C mutants 20 µl proteolysis mixture (stopped with EDTA without heating and without any further purification, see section VI.10.) obtained in the absence of specific oligoduplex was loaded on the column.

Gel filtrations of EcoRII-C-DNA was carried out at room temperature on an ÄKTA FPLC (*GE Healthcare, Uppsala, Sweden*) system using a Superdex 200 HR column (*GE Healthcare, Uppsala, Sweden*) pre-equilibrated with Gel filtration buffer II (20 mM Tris-HCl (pH 7.5 at 25°C), 200 mM KCl, 10% glycerol, 5 mM CaCl₂). The samples of EcoRII-C purified after limited proteolysis by thermolysin (stopped with adding EDTA and without further purification, see section VI.10.) (2.6 µM in terms of dimer) and various concentrations of SP12 oligoduplex (Table 3) were prepared in 100 µl of the same buffer.

Elution from the column was monitored by measuring absorbance at 215 and 260 nm. A calibration curve was generated by measuring the elution volumes of a series of standard proteins of known molecular masses (*Bio-Rad, Hercules, USA*). The molecular masses of EcoRII-DNA complexes were calculated by interpolating their elution volumes onto the calibration curve.
16. EcoRII, EcoRII-C and PspGI reactions with supercoiled plasmid substrates

16.1. Construction and purification of plasmid substrates

Plasmids containing a single, two and three EcoRII recognition sites pEcoRII-1, pEcoRII-2 and pEcoRII-3 were engineered on the basis of the plasmid pKpn2I-ori (a derivative of pKpn2RM3.7 plasmid) containing three EcoRII recognition sites with different flanking sequences. To engineer 2304 bp pEcoRII-1 plasmid containing a single CCWGG site in the p15A-like ori region, MvaI269I-Van91I flanked DNA fragment was replaced by pUC18 fragment bearing ampicillin resistance gene but lacking EcoRII sites.

Ligation of the oligonucleotide duplex:

\[
\begin{align*}
5' &\quad CGCACG\underline{CCTGG\ TATTAT} & 3' \\
3' &\quad CGAGCGTGC\underline{GGACC\ TATAATAGC} & 5'
\end{align*}
\]

containing the CCTGG sequence (underlined) and flanking sequences identical to those surrounding the first EcoRII sites into Eco24I-Bsu15I-pre-cleaved pEcoRII-1 yielded 2285 bp pEcoRII-2, with the sites separated by 362 bp.

Ligation of the oligonucleotide duplex:

\[
\begin{align*}
5' &\quad GATCCGTACG\underline{CCTGG\ TAT} & 3' \\
3' &\quad GCATGCG\underline{GGACC\ TATAATAGC} & 5'
\end{align*}
\]

containing the CCTGG sequence (underlined) and flanking sequences identical to those surrounding the both EcoRII sites, into BglII-pre-cleaved pEcoRII-2 plasmid yielded 2303 bp pEcoRII-3, with the sites separated by 362 bp and 568 bp. Constructed plasmids were sequenced to ensure that they contain desired number of EcoRII recognition sites with identical flanking sequences and identical site orientations.

The plasmids were isolated from *E. coli* GM31 dcm\(^{-}\) strain and purified twice by ultracentrifugation through a CsCl gradient in the presence of ethidium bromide, typically \(~ \sim 65\%\) were as supercoiled monomers. To eliminate the heavier forms in the plasmids preparations the freeze and squeeze
technique for DNA recovery from agarose was used (see section VI.1.2.). Approximately, 90% of the DNA in the plasmids preparations were supercoiled monomers and ~10% randomly nicked open-circles. The concentrations of pEcoRII-1, pEcoRII-2 and pEcoRII-3 plasmid DNA were evaluated by densitometric analysis of ethidium bromide stained agarose gels (see section VI.1.1.). Various amounts of purified plasmids were loaded on agarose together with known amounts of supercoiled pUC19 DNA (*Fermentas UAB, Vilnius, Lithuania*).

33 bp oligoduplexes SP33 and NSP33 (Table 3) were used for the stimulation reactions *in trans*. The flanking sequences surrounding EcoRII recognitions site in the SP33 were identical to those surrounding the EcoRII sites in pEcoRII-1, pEcoRII-2 and pEcoRII-3 plasmids.

**16.2. DNA cleavage reactions**

Reactions of wt EcoRII, EcoRII-C and PspGI with plasmid substrates were conducted at 25°C in a Reaction buffer III containing 10 mM Mg-acetate and 0.1 mg/ml of BSA. Reaction mixtures typically contained 1.5 nM plasmid substrates and 0.05-500 nM of wt EcoRII or 0.2-125 nM of PspGI or 0.1-20 nM of EcoRII-C. For the experiments of oligonucleotide activation of wt EcoRII 33 bp specific and non-specific oligonucleotide duplexes were added to a final concentration of 0.2 nM-16 µM. Reactions were initiated by adding enzyme premixed with Mg-acetate to a mixture of the other reaction components. Aliquots were removed at fixed time intervals and the reaction stopped by adding 1/3 volume of Loading dye solution I containing EDTA. The samples were separated by electrophoresis through agarose. The amount of supercoiled (SC), open-circular (OC), and linear DNA forms (FLL, L2 and L3) was evaluated by densitometric analysis of ethidium bromide-stained gels. In the case of the 2-site and 3-site plasmid substrates, the OC and FLL (for 2-site plasmid) forms and OC, FLL and L2 (for 3-site plasmid) forms of DNA may contain additional nick(s) at the second (or third) recognition site, respectively; however these forms could not be resolved in the gel.
16.3. Data analysis

Exponential function was fitted to the supercoiled plasmid depletion curves obtained under excess of enzyme and at the equimolar protein-DNA concentrations and apparent first-order reaction rate constants \((k_{\text{obs}})\) were determined. For the reactions obtained under the substrate excess conditions the reaction rate constants \(k_{\text{cat}}\) were determined from the linear parts of the reaction progress curves by linear regression. Data analysis used the KyPlot 2.0 software\(^{144}\).

17. EcoRII-DNA AFM analysis

17.1. Construction and purification of DNA fragments for AFM

The fragments for EcoRII AFM analysis were obtained by PCR amplification of selected regions from the pBluescript IIKS (+) plasmid. PCR primers were designed to amplify a defined region of interest to include the correct number of EcoRII restriction sites. The PCR1 fragment containing single EcoRII site 205 bp long and PCR3 fragment containing three EcoRII sites 810 bp long were constructed. The PCR products were cut and purified from the gel using the QIAquick gel extraction kit (Qiagen, Hilden, Germany) and eluted with DNA elution buffer I (10 mM Tris-HCl (pH 7.8 at 25°C)). The concentrations of DNA fragments were determined by measuring the absorbance at 260 nm. The correct positioning of the restriction sites was verified using the MvaI REase, which recognizes the same restriction site (CCWGG) yet does not require binding to multiple sites\(^{35}\), thus cleaving with a higher efficiency (data not shown).

17.2. EcoRII-DNA AFM analysis

For visualization of triple synaptic complex formed by EcoRII with DNA AFM analysis was exploited. The samples were prepared, images were acquired and data analysis was performed by prof. dr. Y. Lyubchenko group (University of Nebraska Medical Center, USA). Images were acquired in tapping mode in air using a Multimode SPM Nanoscope IV system.
(Veeco/Digital Instruments, Santa Barbara, USA). For each type of EcoRII-DNA complex formed, data were obtained for the length of the DNA fragments, the volume of the protein, and the yields of the various complexes using Femtoscan Online (Advanced Technologies Center, Moscow, Russia). EcoRII binding reaction mixtures (10 µl total) contained 10-100 nM PCR3 fragment, 20-200 nM EcoRII protein (in terms of dimer) in Binding buffer (40 mM HEPES (pH 8.4 at 25°C)) containing 5 mM CaCl₂. Binding reactions for the wt EcoRII protein with the PCR1 fragment were performed at a 1:1 protein:DNA ratio. The final reaction mixture contained 209 nM EcoRII and 205 nM PCR1 in Binding buffer. Reaction mixtures were incubated at room temperature for 15 min and then fixed with 0.5% glutaraldehyde for 10 min. The fixation reaction was stopped by adding 1-2 µl of 2 M Tris-HCl (pH 7.1). After completion of the fixation step, the complexes were purified using Montage PCR filter units (Millipore, Bedford, USA). To purify the samples, 300 µl of DNA elution buffer II (10 mM HEPES (pH 7.4 at 25°C)) was spin-filtered through them. The final samples were resuspended in same buffer. Appropriate dilutions were made before the deposition of the samples onto mica to provide an even spread of DNA on the surface for AFM imaging. AFM imaging and data analysis were performed as described in 150.

18. Structure analysis

The image of the protein structures in all the figures were produced with Molscript 151 and Raster3D 152. Superposition of the structures was performed with Top3D 153 and secondary structures were assigned with Stride 154.
RESULTS AND DISCUSSION

VII. Protein sequence similarities within the restriction endonuclease subfamily specific for CCGG/CCNGG sites

Pairwise FASTA search using REase Ecl18kI (\textsuperscript{1\downarrow}CCNGG, cleavage position indicated by the $\downarrow$) sequence as a query allowed us to detect weak local similarities to the Type IIF tetrameric Cfr10I (R\textsuperscript{1\downarrow}CCGGY) endonuclease (data not shown). Further analysis (see section VI.6.) revealed that the similarity region identified between Ecl18kI and Cfr10I could be further extended to a larger family of the restriction enzymes specific for \textsuperscript{1\downarrow}CCNGG/\textsuperscript{1\downarrow}CCWGG and R\textsuperscript{1\downarrow}CCGGY/G\textsuperscript{1\downarrow}CCGGC nucleotides (Fig. 30). In fact, only few residues are strictly conserved (Fig. 30) and this alignment by itself lacks stringent statistical significance to be trusted without further information. Fortunately, it has a strong structural support: crystal structures for Cfr10I\textsuperscript{56}, NgoMIV\textsuperscript{29} and Bse634I\textsuperscript{31} REases have been solved giving a possibility to map the conserved residues to the known three-dimensional structures. Since the NgoMIV structure is solved both in the presence of DNA and metal cofactors, it was used as a reference in the further analysis.

Figure 30. Conserved sequence region between Ecl18kI, EcoRII, PspGI, Bse634I, Cfr10I and NgoMIV (adapted from\textsuperscript{83}). Residues that are identical or similar are boxed and shown in red and yellow, respectively. Consensus residues are shown above the Ecl18kI sequence. Putative Ecl18kI, EcoRII and PspGI catalytic sequence discrimination residues are noted with an asterisk. Secondary structure assignment of NgoMIV, as determined by X-ray crystallography (PDB ID 1FIU\textsuperscript{29}), is shown below the sequences.
The most conserved region corresponding to the consensus K-\(\text{(K/R)-(D/E)}\)R------E spans across the \(\beta_3\) strand and \(\alpha_7\) helix of NgoMIV (Fig. 30). According to the crystal structure\(^2\) residues located on \(\beta_3\) strand and \(\alpha_7\) helix of NgoMIV form a part of the catalytic/metal binding site and provide an interface for the recognition of the central C:G nucleotides (see section III.1 and Fig. 4C).

Acidic residues D140 and E201, chelating first Mg\(^{2+}\) ion in NgoMIV, and lysine K187 presumably involved in catalysis (see section III.1 and Fig. 3) are conserved within Ecl18kI, EcoRII and PspGI subfamily (Fig. 30). According to the crystal structure, D140 from \(^{139}\text{PD}X_{46}\text{K}X_{13}E^{201}\) motif and E70 residue located on the \(\alpha_3\)-helix constitute a second metal ion binding site in NgoMIV (see section III.1 and Fig. 3). While the D140 residue of NgoMIV is conserved within Ecl18kI, EcoRII and PspGI subfamily (Fig. 30), the location of the second acidic residue is less obvious. It was suggested\(^3\) that E125 residue of Ecl18kI might be a structural equivalent of the E71 residue of Cfr10I. Thus, on the basis of the alignment and the structural mapping of the conserved residues (Fig. 30) we propose that sequence motif \(\text{E}X_{26-33}(P/V)\text{D}X_{21-24}\text{K}X_{12}E\) conserved within Ecl18kI, EcoRII and PspGI correspond to the active site that is structurally similar to that of Bse634I, Cfr10I and NgoMIV enzymes (Table 5).

Table 5. Functional amino acids within the subfamily of restriction enzymes specific for the CCGG/CCNGG sites*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Catalysis/metal ion chelation</th>
<th>C:G recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecl18kI</td>
<td>E125 D160 K182 E195</td>
<td>R186 E187 R188</td>
</tr>
<tr>
<td>EcoRII</td>
<td>E271 D299 K324 E337</td>
<td>K328 D329 R330</td>
</tr>
<tr>
<td>PspGI</td>
<td>E105 D138 K160 E173</td>
<td>R164 E165 R166</td>
</tr>
<tr>
<td>Bse634I</td>
<td>E80 D146 K198 E212</td>
<td>R202 D204 R205</td>
</tr>
<tr>
<td>Cfr10I</td>
<td>E71 D134 K190 E204</td>
<td>R194 D196 R197</td>
</tr>
<tr>
<td>NgoMIV</td>
<td>E70 D140 K187 E201</td>
<td>R191 D193 R194</td>
</tr>
</tbody>
</table>

* Catalytic/C:G recognition residues were taken from the alignment shown in Fig. 30 and from\(^3\).

NgoMIV residues involved in the recognition of the central CCGG tetranucleotide are located on a short stretch of amino acids \(^{191}\text{RSDR}^{194}\) interspaced between catalytic/metal binding site residues K187 and E201,
respectively\textsuperscript{29} (see section III.1 and Fig. 4). According to the alignment NgoMIV residues R191, D193 and R194 are conserved within the Ecl18kI, EcoRII and PspGI subfamily (Fig. 30). Thus, it is tempting to speculate that two positive amino acid residues (R/K) and one negative amino acid residue (D/E) of Ecl18kI, EcoRII and PspGI contribute to the discrimination of C:G nucleotides at the recognition site (Table 5).

Thus, on the basis of the sequence similarities and the mapping of the conserved residues to the crystal structure of NgoMIV, we hypothesize that two subfamilies of endonucleases Ecl18kI/EcoRII/PspGI and Bse634I/Cfr10I/NgoMIV, specific, respectively, for CCNGG/CCWGG and RCCGGY/GCCGGC sites, share conserved catalytic/metal binding site architecture and the mechanism of CC:GG dinucleotide recognition\textsuperscript{83}.

VIII. Active site/C:G binding residues in the restriction enzymes Ecl18kI and EcoRII

To demonstrate the functional importance of the predicted putative catalytic and C:G discriminating residues of Ecl18kI and EcoRII we employed biochemical and structural methods.

First, we generated a set of alanine mutants, expressed them in \textit{E. coli}, purified and studied their DNA binding and cleavage properties. To simplify and facilitate purification procedures both wt and mutant Ecl18kI and EcoRII proteins were expressed as N-terminal (histidine)\textsubscript{6}-tagged proteins. The following rationale was behind the mutational analysis. If the residues are the catalytic moieties of the enzyme, their replacement should generate mutant proteins which have virtually no cleavage activity but they may still be able to bind DNA like the wt enzyme. On the other hand, if amino acid residues are involved in DNA binding, their replacement should affect both DNA binding and cleavage ability.

In order to prevent DNA cleavage, binding of restriction enzymes to DNA is usually studied in the absence of the Mg\textsuperscript{2+} cofactor, or in the presence
of Ca$^{2+}$ ions that do not support catalysis$^{155-157}$. Endonucleolytic activity of the Ecl18kI and EcoRII mutant proteins was assessed using $\text{dam}^{-}\text{dcm}^{-}$ DNA since the recognition sites for these enzymes may overlap with sites for $\text{dcm}$ methylation.

1. Identification of the putative catalytic/C:G binding site of Ecl18kI by biochemical analysis

1.1. Mutational analysis of the possible catalytic/metal binding site of Ecl18kI

We generated E125A, D160A, K182A and E195A mutants of Ecl18kI (see section VI.2.2.) and studied their DNA cleavage (see section VI.12.1.) and binding properties (see section VI.11.1.).

Table 6. DNA cleavage and binding activities of wt and mutant Ecl18kI proteins*

<table>
<thead>
<tr>
<th></th>
<th>Specific activity, %</th>
<th>$K_d$, nM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>100</td>
<td>3.5±0.2</td>
</tr>
<tr>
<td>E125A</td>
<td>&lt;0.1***</td>
<td>&gt;200****</td>
</tr>
<tr>
<td>D160A</td>
<td>0</td>
<td>&gt;50**</td>
</tr>
<tr>
<td>K182A</td>
<td>0</td>
<td>0.3±0.1</td>
</tr>
<tr>
<td>E195A</td>
<td>4**</td>
<td>0.2±0.1</td>
</tr>
<tr>
<td>R186A</td>
<td>0.2**</td>
<td>&gt;100****</td>
</tr>
<tr>
<td>E187A</td>
<td>&lt;0.1****</td>
<td>7.0±0.2</td>
</tr>
<tr>
<td>R188A</td>
<td>0</td>
<td>&gt;100****</td>
</tr>
</tbody>
</table>

* Table was adapted from$^{83}$. ** 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 Ecl18kI equals 0.07±0.02 s$^{-1}$. *** Specific activity was estimated using phage $\lambda$. $\text{dam}^{-}\text{dcm}^{-}$ DNA cleavage assay (see Methods). **** We were unable to determine the $K_d$ value due to the weak smeared bands.

The pUC19 $\text{dam}^{-}\text{dcm}^{-}$ DNA cleavage experiments revealed that D160A and K182A mutant proteins completely lost both double-strand cleavage and nicking activities (Table 6). Two other mutants E125A and E195A exhibited significantly decreased specific DNA activity (Table 6) and impaired double-strand cleavage ability (complete DNA cleavage was not observed). The significantly compromised mutants’ catalytic activities suggest their active site
function. However, a similar phenotype with loss both of the DNA binding and cleavage activity would be characteristic for the mutants of the DNA binding residues.

![Gel shift analysis of DNA binding by Ecl18kI and mutant proteins.](image)

**Figure 31. Gel shift analysis of DNA binding by Ecl18kI and mutant proteins.** Increasing amounts of protein (indicated below the corresponding lanes in terms of dimer in nM) were incubated with 1 nM $^{33}$P labelled specific and non-specific 16 bp oligoduplexes and protein-DNA complexes were analyzed in polyacrylamide gel under non-denaturing conditions (see Methods). All gels except (A) were run in the presence of 10 mM Ca-acetate.

In order to determine whether the mutations affected the ability of Ecl18kI to bind the substrate, the DNA binding of wt and mutant proteins was analyzed using gel shift assay. Two synthetic 16 bp DNA duplexes, one containing the recognition site CCWGG and another lacking an Ecl18kI site...
were used in binding experiments as specific and non-specific DNA, respectively (Table 3). Binding experiments demonstrated that DNA binding by Ecl18kI without divalent metal ions is weak ($K_d = 236 \pm 33$ nM, see section VI.11.5. for $K_d$ calculation) (Fig. 31A). The inclusion of Ca$^{2+}$ ions into the reaction mixtures and gel running buffers, however, resulted in clear shifts (Fig. 31B) confirming previous reports that many REases, which show no DNA binding in the gel shift assay in the absence of metal ion, however, display DNA binding specificity in the presence of Ca$^{2+}$ ions that do not support DNA cleavage$^{84,155,156}$. While wt Ecl18kI binds specific DNA with high affinity ($K_d = 3.5 \pm 0.2$ nM), non-specific DNA binding remains weak.

Therefore, gel shift experiments with Ecl18kI mutants were performed in the presence of 10 mM of Ca-acetate both in the binding and the gel running buffers (calculated $K_d$ values are summarized in Table 6). The gel shift analysis of Ecl18kI mutants revealed that putative active site mutants fall into two classes with respect to DNA binding. The DNA binding affinity of K182A and E195A mutants of Ecl18kI increases relative to the wt enzyme (Table 5). In contrast, the DNA binding ability of E125A and D160A mutants is highly compromised (Table 6 and Fig. 31C). K182A and E195A mutants of Ecl18kI resulted in the loss or significant decrease of endonucleolytic activity while DNA binding increased in comparison to the wt enzyme (Table 6). Previous mutational studies of the active site residues of restriction enzymes, e.g. EcoRI$^{155}$, MunI$^{86}$ and BamHI$^{158}$, revealed that alanine replacements do not interfere with the 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 the proposed active site function of K182 and E195 residues in Ecl18kI. The gel shift analysis, however, indicates that alanine replacement of the putative metal chelating residue D160 decreases DNA binding affinity (Fig. 31C). 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. 31A). D160 of Ecl18kI is
homologous to the D146, D134 and D140 residues of Bse634I, Cfr10I and NgoMIV, respectively (Table 5). In Bse634I/Cfr10I/NgoMIV these carboxylates are involved in the bridging interactions between two metal ions at the active site (Fig. 3). The detailed role of both metal ions is not clear yet; however it is possible that the metal ion(s) contribute both to catalysis and DNA binding. Indeed, Ca$^{2+}$ ions are necessary for the specific DNA binding by Cfr10I$^{157}$. Noteworthy, a D134A mutation compromised the DNA binding affinity of Cfr10I in the presence of Ca$^{2+}$ ions (Zaremba, M., Skirgaila, R. and Siksnys, V., unpublished results). It is conceivable that D134A replacement in Cfr10I abolishes the coordination of Ca$^{2+}$ ion that is necessary for the stabilization of the enzyme complex with specific DNA. Similarly, if the D160 residue of Ecl18kI chelates Ca$^{2+}$ ion that is required for the specific DNA binding, it is not surprising that the D160A mutation compromises the DNA binding ability. Interestingly, alanine replacement of E125 residue presumably involved in the coordination of the metal ion Me$_A$ also diminishes cognate DNA binding. Hence, it is tempting to speculate that the binding of the Me$_A$ metal ion at the active site is important for the specific DNA binding by Ecl18kI.

Thus, the mutational analysis data are in principle consistent with the hypothesis that E125, D160, K182 and E195 residues of Ecl18kI constitute the catalytic/metal binding site similar to that of NgoMIV$^{83}$.

1.2. Mutational analysis of possible DNA binding site of Ecl18kI

In order to demonstrate the functional importance of the putative DNA binding residues R186, E187 and R188 of Ecl18kI we generated a set of alanine mutants and studied again DNA cleavage (see section VI.11.1.) and binding properties (see section VI.10.1.) of the mutant proteins. Alanine substitution of R186 and R188 residues of Ecl18kI resulted in the complete loss of the endonucleolytic activity accompanied by highly compromised DNA binding (Fig. 31D and Table 6). That confirms the importance of R186 and R188 residues in the specific DNA binding. Sequence-specific contacts of the
restriction enzymes, however, are often redundant and comprised by the amino acid residues located on the different structural elements\textsuperscript{12}. Therefore we cannot rule out the possibility that other residues located on the structural elements not conserved between Ecl18kI and NgoMIV also contribute to the recognition interface of Ecl18kI. Interestingly, while E187A substitution nearly abolished the catalytic activity, specific DNA binding affinity decreased only ~ 2-fold (Table 6) in comparison to the wt Ecl18kI.

Thus, our mutational and biochemical studies are consistent with the hypothesis that Ecl18kI, specific for \textsuperscript{↓}CCNGG sequence, shares a conserved structural mechanism of the C:G base pair recognition with restriction enzymes Bse634I, Cfr10I and NgoMIV, specific for the \textsuperscript{R↓}CCGGY/G↓CCGGC sequences\textsuperscript{83}.

Circular dichroism (see section VI.8.) and analytical ultracentrifugation (see section VI.9.) experiments (data not shown) revealed no changes in the secondary or quaternary structure of the alanine mutants relative to the wt enzyme. Therefore, it is unlikely that perturbation of the protein conformation or oligomerization state leads to the loss of function.

Of note is that the mutation of the homologous residues in the isoschizomeric SsoII restriction enzyme revealed similar phenotypes as in Ecl18kI\textsuperscript{159}.

2. Confirmation of the putative Ecl18kI catalytic/DNA binding site by Ecl18kI-DNA X-ray structure

The Ecl18kI-DNA complex structure was solved at 1.7 Å resolution\textsuperscript{80}. Structural comparison between Ecl18kI and NgoMIV restriction enzymes reveals that the Ecl18kI residues, identified by the mutational analysis, indeed correspond to the active site residues of NgoMIV (Fig. 32)\textsuperscript{80}. In NgoMIV, D140 bridges two Mg\textsuperscript{2+} ions in the active site, and we presumed that D160 in the Ecl18kI has the same role. However, Mg\textsuperscript{2+} ions are absent in the Ecl18kI crystals, which could only be grown in the presence of the metal-chelator
EDTA. Other catalytic/metal chelating residues of NgoMIV are also conserved in Ecl18KI (Fig. 32).

The Cα atom of E201 in NgoMIV coincides with the Cα atom of E195 in Ecl18KI, but the side chain is in a different conformation. It is possible that the conformation of E195 may change in the presence of metal ions.

Biochemical and structural data of Ecl18KI is consistent with the presence of E125, D160, K182 and E195 in the active site. Hence, Ecl18KI possesses an NgoMIV-like active site as predicted.

Figure 32. Superimposed active sites of Ecl18KI and NgoMIV (adapted from). Active site residues of Ecl18KI E125, D160, K182 and E195 superimpose with NgoMIV E70, D140, K187 and E201 residues (shown in gold and olive green, respectively). Two Mg²⁺ ions present in NgoMIV-DNA complex are shown as grey spheres.

Figure 33. Sequence recognition by Ecl18KI and NgoMIV (adapted from). Ecl18KI (A) and NgoMIV (B) interactions with one half-site of the recognition sequence. Only direct interactions with one half-site outer and inner C:G base pairs are shown. (a) and (b) indicate for amino acids which correspond to the different monomers of primary dimer. Recognition sequences are shown below. The depicted base pair is underlined in the recognition sequence.
Structural comparison between Ecl18kI and NgoMIV revealed that the recognition pattern of the C:G nucleotides in Ecl18kI is strikingly similar to that of NgoMIV confirming our prediction (Fig. 33)\(^80\). Ecl18kI contacts the bases of the recognition sequence primarily from the major groove side. Major groove contacts are exclusively mediated by R186, E187 and R188. Arginines 186 and 188 donate bidentate hydrogen bonds to the outer and inner guanines, respectively. The side chain oxygen atoms of E187 accept one hydrogen bond from each of the two neighboring cytosines of the recognition sequence (Fig. 33). Sequence-specific minor groove contacts are mediated exclusively by Q114. The main chain oxygen atom of this residue accepts a hydrogen bond from the outer guanine, and the side chain forms two hydrogen bonds with the inner guanine.

Thus, Ecl18kI residues R186, E187 and R188 coincide with NgoMIV residues R191, D193 and R194 which unambiguously specify inner and middle C:G nucleotides within the one half-site of the NgoMIV recognition site GCCGGC, as it was predicted\(^83\).

3. Identification of the putative catalytic/DNA binding site of EcoRII by biochemical analysis

Former attempts to locate the active site of EcoRII searching for the canonical PD-(D/E)XK motifs revealed few candidate residues\(^104,118,127\), however only E96 and E271 residues appeared to be important for catalysis (see section IV.7.). Thus, the sequence location of the catalytic/metal binding residues of EcoRII was not yet unambiguously predicted. We hypothesized that EcoRII C-terminal part sequence motifs \(^{271}EX_{27}PDX_{24}KK_{12}E^{337}\) specify a putative catalytic/metal binding site and K328, D329 and R330 residues correspond to the C:G discrimination determinants of the EcoRII recognition sequence (see section VII.). In order to demonstrate the functional importance of the putative active site (E271, D299, K324 and E337) and C:G binding residues (K328, D329 and R330) of EcoRII we generated a set of alanine
mutants (see section VI.3.2.) and studied DNA cleavage (see section VI.12.2.) and binding properties of the mutant proteins (see section VI.11.2.).

3.1. Mutational analysis of possible catalytic site at EcoRII

Phage λ dam⁻ dcm⁻ DNA cleavage analysis revealed that the catalytic activity of E271A, D299A, K324A and E337A mutants was completely abolished or significantly compromised suggesting their active site function (Table 7). However, a similar loss of function phenotype would be characteristic for the mutants of the DNA binding residues. So in order to test whether the alanine mutations of the putative catalytic residues at the C-terminal part affect specific DNA binding, we studied the DNA binding properties of the mutant proteins.

Table 7. DNA cleavage and binding of wt EcoRII and mutant proteins*

<table>
<thead>
<tr>
<th></th>
<th>EcoRII Specific activity, %**</th>
<th>K_d, nM****</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>100</td>
<td>11±1</td>
</tr>
<tr>
<td>E271A</td>
<td>0</td>
<td>&gt;2000</td>
</tr>
<tr>
<td>D299A</td>
<td>0</td>
<td>&gt;2000</td>
</tr>
<tr>
<td>K324A</td>
<td>0</td>
<td>8±1</td>
</tr>
<tr>
<td>E337A</td>
<td>&lt;0.01***</td>
<td>14±2</td>
</tr>
<tr>
<td>K328A</td>
<td>&lt;0.01***</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>D329A</td>
<td>&lt;0.01***</td>
<td>80±20</td>
</tr>
<tr>
<td>R330A</td>
<td>0</td>
<td>&gt;600</td>
</tr>
</tbody>
</table>

* Table was adapted from\textsuperscript{160}. ** The DNA cleavage activity of full-length EcoRII proteins was determined as described in Methods. *** Incomplete DNA hydrolysis after 16 h incubation. **** K_d value for the C-terminal domain binding to specific 33 bp oligoduplex (Table 3) is provided.

Previous limited proteolysis studies indicated that EcoRII consists of two physically separate domains that exhibit different functions\textsuperscript{131}. EcoRII-N binds the CCWGG sequence but lacks DNA cleavage activity, while EcoRII-C \textit{per se} is a functional orthodox REase that cleaves DNA containing specific CCWGG sequence before the first C\textsuperscript{131} (see section IV.8.). To avoid possible DNA binding interference by the N- and C-terminal domains in the full-length EcoRII, the mutant proteins were subjected to limited proteolysis by thermolysin (see section VI.10.) to yield EcoRII-C mutants (E271A-C, D299A-C, K324A-C and E337A-C).
Since previous attempts failed to demonstrate the EcoRII-C DNA binding ability, we sought to find out the EcoRII-C DNA binding conditions. Many restriction enzymes, which show no DNA binding in the gel shift assay in the absence of metal ion, however, display DNA binding specificity in the presence of Ca$^{2+}$ ions that do not support DNA cleavage by REases. Indeed, gel shift experiments in the absence of Ca$^{2+}$ revealed no binding of EcoRII-C to the 33 bp oligoduplex either containing or lacking the recognition site (Fig. 34A). We have demonstrated that Ca$^{2+}$ ions increased the Ecl18kI binding affinity to DNA (see section VIII.1.1). Therefore, gel shift analysis of the DNA binding by EcoRII-C was performed in the presence of 5 mM of Ca$^{2+}$ ions both in the binding and gel running buffers. Analysis of the DNA binding in the buffers supplemented with Ca$^{2+}$ ions revealed EcoRII-C binding to the cognate oligoduplex ($K_d = 11 \pm 1$ nM) but showed no binding to the non-specific DNA (Fig. 34B).

**Figure 34. Gel shift analysis of DNA binding by EcoRII-C**. The binding reactions contained either the specific or the non-specific 33P-labelled 33 bp oligoduplex (1 nM) and the EcoRII-C at concentrations (in terms of dimer) as indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions (see Methods). The last lane of each gel contained wt EcoRII instead of EcoRII-C. Gels were run in the presence of 0.1 mM EDTA (A) or 5 mM of Ca-acetate (B).
In the next step, DNA binding by the EcoRII-C mutants E271A-C, D299A-C, K324A-C and E337A-C was analysed in the presence of Ca\(^{2+}\). The gel shift analysis indicates that E337A-C and K324A-C mutants show the same DNA binding affinity as the EcoRII-C confirming catalytic/active site function of the latter residues (Fig. 35A and Table 7). DNA binding ability of D299A-C and E271A-C mutants, however, is significantly compromised (Table 7). D299 residue of EcoRII (equivalent to D160 of EcI18kI) is homologous to the D134 and D140 residues of Cfr10I and NgoMIV (Table 5), respectively, and therefore may be involved in the bridging interactions with metal ions (see section Fig. 3).

Figure 35. Gel shift analysis of DNA binding by EcoRII mutants E337A-C and R330-C\(^{160}\). The binding reactions contained either the specific or the non-specific 33P-labelled 33 bp oligoduplex (1 nM) and the protein at concentrations (in terms of dimer) as indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions (see Methods). The last lane of each gel contained wt EcoRII instead of E337A-C (A) or R330A-C (B). Gels were run in the presence of 5 mM of Ca-acetate.

DNA binding studies revealed that Ca\(^{2+}\) ions are absolutely required for specific DNA binding by EcoRII-C (Fig. 34). Replacement of negatively charged D299 residue by alanine may disturb Ca\(^{2+}\) ion coordination at the active site of EcoRII-C thus abrogating DNA binding. Similar phenotype of the equivalent D160 residue of EcI18kI was observed (Table 6). Interestingly, the alanine replacement of E271 residue, presumably involved in the coordination
of the metal ion Me₄, also diminishes cognate DNA binding. Structurally equivalent E125A of Ecl18kI exhibited analogous phenotype (Table 6).

Therefore, it is tempting to speculate that the binding of the Me₄ metal ion at the active site of EcoRII-C is important for the specific DNA binding similarly to Ecl18kI.

3.2. Mutational analysis of the possible DNA binding site of EcoRII

In order to test the assumption that the C-terminal K328, D329 and R330 residues contribute to the discrimination of C:G base pairs within the CCWGG site characteristic for EcoRII we have generated K328A, D329A and R330A mutants of EcoRII and tested their DNA cleavage and binding ability.

We have found that the alanine replacement of K328 and R330 residues abolished both the DNA cleavage and binding ability confirming the importance of these residues in the specific DNA binding (Table 7 and Fig. 35B). Interestingly, while the D329A substitution nearly abolished the catalytic activity of D329A-C, specific DNA binding affinity decreased only ~10-fold (Table 7) in comparison to the wt EcoRII-C.

Thus, our mutational and biochemical studies are consistent with the hypothesis that EcoRII, specific for \( \text{CCWGG} \) sequence, shares similar catalytic/metal binding site and conserved structural mechanism of the C:G base pair recognition with the restriction enzymes Bse634I, Cfr10I and NgoMIV, specific for the \( \text{RCGGY/GCCGGC} \) sequences\(^{160}\).

All mutant proteins show circular dichroism spectra (see section VI.8.) identical to the wt enzyme (data not shown), indicating that the mutations did not induce any gross conformational changes in the protein. Gel filtration studies (see sections VI.15. and X.1.3.) of the EcoRII-C mutants revealed that all proteins elute at a time that corresponds to an apparent \( M_w \) of ~ 54–57 kDa (data not shown) that is close to the value of 52 kDa calculated for EcoRII-C dimer. Hence, the EcoRII-C mutants remain in a dimeric form, similar to the wt EcoRII-C.
4. Confirmation of the EcoRII putative catalytic/DNA binding site by EcoRII X-ray apo-structure

Zhou et al. solved the crystal structure of an EcoRII R88A mutant at 2.1 Å resolution\textsuperscript{79}. The structure confirmed that EcoRII monomer has two domains linked through a hinge loop. The N-terminal effector-binding domain EcoRII-N has a novel DNA recognition fold with a prominent possible DNA binding cleft. The C-terminal catalytic domain EcoRII-C has a REase-like fold.

Figure 36. Conserved active sites and DNA recognition elements between EcoRII and tetrameric restriction enzymes Bse634I, Cfr10I and NgoMIV (adapted from\textsuperscript{160}). (A) Superposition of EcoRII (skyblue), Bse634I (yellow), Cfr10I (pink) and NgoMIV (olivegreen) proteins reveals putative active site residues of EcoRII. Two Mg\textsuperscript{2+} ions from NgoMIV–DNA structure are shown as grey spheres and labeled Me\textsubscript{A} and Me\textsubscript{B} according to\textsuperscript{29}. The scissile phosphate is shown as an orange sphere. (B) Residues involved in the recognition of C:G nucleotides in NgoMIV (olivegreen) and their structural equivalents in EcoRII (skyblue). Recognition sequence is shown below. The depicted base pair is underlined in the recognition sequence.

Structural comparison reveals that in EcoRII apo-structure C-terminal residues E271, D299, K324 and E337 superimpose with E70, D140, K187 and E201 residues at the active site of NgoMIV (Fig. 36A) as it was predicted from the sequence alignment\textsuperscript{83}. Conserved K324 residue of EcoRII is presumably involved in catalysis similarly to K187 NgoMIV\textsuperscript{29} and many other REases\textsuperscript{12}. Structural superposition suggests that EcoRII residues E271 and D299 may contribute to the coordination of the second metal ion Me\textsubscript{A} similarly to
Bse634I, Cfr10I and NgoMIV (Fig. 3). Therefore, it is likely that EcoRII like NgoMIV coordinates two metal ions at the active site.

Structural comparison reveals that K328, D329 and R330 of EcoRII are homologous to the R191, D193 and R194 residues of NgoMIV (Fig. 36B) and therefore suggest that these residues could contribute to the discrimination of C:G base pairs as it was predicted from sequence alignment (Fig. 30)\textsuperscript{83}.

5. Family of restriction enzymes bearing CCGG tetranucleotide in their recognition sequence

Type II restriction enzymes represents the large family of functionally related enzymes, however little sequence similarity has been observed between them\textsuperscript{1} (see reviewed in sections I.-III.). Amino acid sequence similarities appear to be restricted to the enzymes that recognize the same or similar DNA sequences and cleave DNA at the same position\textsuperscript{48}.

Our studies reveal the restriction enzymes family which is specific for the $\downarrow$CCNGG/$\downarrow$CCWGG and R$^+$CCGGY/G$^+$CCGGC sequences (see Table 5). In principle, the recognition sequence of these enzymes bear the common CCGG (intact or interrupted by N or W) tetranucleotide and they all cleave their recognition sites before the first C resulting in a different cleavage pattern (4 or 5 nt 5’ overhangs). Of note, other enzymes like HpyF100III ($^+$CCGG, Tamulaitiene, G., Sukackaite, R. and Siksnys, V., personal communication), Kpn2I (T$^+$CCGGA) and SgrAI (CR$^+$CCGGYG)$^{159,160}$ containing the CCGG tetranucleotide within their target sites could be incorporated into this group (see Table 1).

Interestingly, this group contains restriction enzymes which belong to the different Type II subtypes. PspGI is a homodimer and belongs to the orthodox Type IIP\textsuperscript{142}. Bse634I, Cfr10I, EcoRII and NgoMIV require simultaneous binding of two substrates (or one substrate with two recognition sites) for efficient cleavage (see reviewed in sections III. and IV.). EcoRII belongs to Type IIE enzymes\textsuperscript{103,104}, which are homodimeric proteins with two distinct
binding sites for their respective recognition sequences, one being the active site and the other the effector site. Bse634I, Cfr10I and NgoMIV are Type IIF tetrameric enzymes, which are composed of four identical subunits that possess four active sites which in a coordinate manner cleave two recognition sites (four DNA strands)\textsuperscript{29,31,56}.

**IX. How Ecl18kI, EcoRII-C and PspGI achieve recognition of its target site?**

Our mutational, biochemical and structural studies are consistent with the hypothesis that Ecl18kI and EcoRII, specific for the CCNGG and CCWGG, respectively, share a conserved structural mechanism of the C:G base pair recognition with restriction enzymes Bse634I, Cfr10I and NgoMIV specific for the RCCGGY/GCCGGC sequences. In principle, the recognition sequence of Ecl18kI (CCNGG) and EcoRII (CCWGG) might be considered as an interrupted version of CCGG tetranucleotide found at the recognition sites of Bse634I, Cfr10I and NgoMIV. How could this single nucleotide insertion at the recognition site be accommodated by Ecl18kI and EcoRII while maintaining the conserved catalytic/metal binding site and sequence recognition elements employed by the Bse634I, Cfr10I and NgoMIV enzymes?

BglI (GCCNNNN\textsuperscript{↓}NGGC) and SfiI (GGCCNNNN\textsuperscript{↓}NGGCC) restriction enzymes are very similar to the blunt end-cutter EcoRV (GAT\textsuperscript{↓}ATC) in terms of monomer structure, but generate 3 nt 3’ overhangs rather than blunt ends, because the enzymes dimerize in very different ways\textsuperscript{62,68}. One may assume that changes in the dimer interface of Ecl18kI and EcoRII could allow accommodation of the extra nucleotide in the target site while keeping the same structural elements for the sequence recognition/catalysis employed by Bse634I, Cfr10I and NgoMIV.
1. Restriction endonuclease Ecl18kI flips the central base pair within its recognition site

The superposition of Ecl18kI-DNA and NgoMIV-DNA structures revealed that Cα-atoms of the residues which are involved in cleavage and C:G recognition in Ecl18kI and NgoMIV were found to almost coincide (Fig. 37)\textsuperscript{80}. Therefore, the different cleavage patterns generated by Ecl18kI and NgoMIV, ↓CCNGG and G↓CCGGC, respectively, are not due to the differences of the dimerization modes. Instead, the Ecl18kI structure reveals a novel mechanism that provides an elegant explanation of the different specificities and cleavage patterns of Ecl18kI and NgoMIV.

The DNA-duplex in complex with Ecl18kI is dramatically deformed. The hydrogen bonds of the central base pair are broken, and the nucleotides are flipped out, which places their bases outside the sugar-phosphate backbone (Fig. 38A, 39A)\textsuperscript{80}. Each extruded base is accommodated into a pocket of Ecl18kI made by the side chain atoms of R57 on one face and the indole ring of W61 on the other face (Fig. 38B).

The DNA is strongly kinked at the side of the nucleotide flips (Fig. 39C). This deformation opens the minor groove and bends the DNA towards the major groove. Together, the register shift and the kink reduce the distance between the scissile DNA phosphates to 17.2 Å, which is within 0.1 Å identical to the distance between the scissile phosphates in the complex of NgoMIV with cleaved DNA (Fig. 39)\textsuperscript{80}.
Thus, Ecl18kI generates longer overhangs than NgoMIV simply because the enzyme makes the central 5 bp duplex in its recognition sequence mimic a 4 bp duplex.

Therefore, the two enzymes can use a conserved DNA recognition module, yet recognize different sequences, and form superimposable dimers, yet generate different cleavage patterns. **Hence, Ecl18kI is the first example of**
a restriction endonuclease that flips nucleotides to achieve specificity for its recognition site$^{80}$.

The evolutionary link between Ecl18kI, EcoRII-C and PspGI prompts the question whether nucleotide flipping occurs in EcoRII-C and PspGI restriction enzymes that recognize not CCNGG but CCWGG sequence. Available biochemical and structural data$^{79,80,83,142,160}$ argue that DNA recognition appears to be similar in Ecl18kI, EcoRII-C and PspGI, which would argue for a conserved nucleotide flipping mechanism.

Moreover, R57 and W61, which sandwich the flipped bases in the Ecl18kI–DNA co-crystal structure, spatially coincide with the R222 and Y226 of EcoRII-C and R161 and F64 of PspGI. The structure of the PspGI enzyme is not yet known. PspGI modeling studies suggest significant similarities to Ecl18kI$^{161}$. According to PspGI model R161 and F64 residues are counterparts of R57 and W61 in Ecl18kI. Moreover, genetic studies support the PspGI model and provide indirect evidence that PspGI may flip central nucleotides within the sequence that matches its target site except for the presence of a G:C pair instead of the A:T pair at the center$^{161}$. However, in contrast to Ecl18kI, which shows no specificity for the central base pair, EcoRII-C and PspGI prefer A:T at the central position of their recognition sequences.
2. 2-aminopurine steady-state fluorescence analysis of Ecl18kI, EcoRII-C and PspGI enzymes

Nucleotide flipping by Ecl18kI was demonstrated in the crystal structure. To demonstrate that flipping occurs also in solution 2-aminopurine (2-AP) steady-state fluorescence was employed. At neutral pH 2-AP makes a Watson–Crick base pair with T, which is only slightly weaker than the natural A:T base pair\(^{162}\). The 2-AP fluorescence is highly quenched in polynucleotides due to the stacking interactions with neighboring bases\(^{163}\) and therefore increases strongly when the base is flipped out of the DNA helix\(^{164,165}\). We have also used 2-AP steady-state fluorescence assay to examine if EcoRII-C and PspGI base flip nucleotides while interacting with their target sites.

2.1. Probes for Ecl18kI, EcoRII-C and PspGI triggered nucleotide flipping

A number of 25 nt oligoduplexes that contain the fluorescent base analog 2-AP at different positions were designed (Table 3). In the oligoduplex T/2, 2-AP was incorporated within the recognition sequence instead of A in the central base position. In the oligoduplex 2FS (‘2-AP flanking specific’), 2-AP was introduced immediately adjacent to the target site.

Table 8. 2-AP containing DNA binding by Ecl18kI, EcoRII-C, PspGI and MvaI

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>(K_d), nM* in the absence of Ca(^{2+})</th>
<th>(K_d), nM* in the presence of Ca(^{2+})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecl18kI wt</td>
<td>170±70</td>
<td>0.11±0.03</td>
</tr>
<tr>
<td>Ecl18kI W61Y</td>
<td>190±100</td>
<td>0.64±0.07</td>
</tr>
<tr>
<td>Ecl18kI W61A</td>
<td>170±40</td>
<td>1.2±0.4</td>
</tr>
<tr>
<td>EcoRII-C</td>
<td>&gt;1000</td>
<td>0.4±0.1</td>
</tr>
<tr>
<td>PspGI</td>
<td>&gt;500</td>
<td>0.5±0.1</td>
</tr>
<tr>
<td>MvaI</td>
<td>0.08±0.03</td>
<td>0.08±0.03</td>
</tr>
</tbody>
</table>

* DNA binding \(K_d\) is the same for oligoduplexes T/A, T/2 and 2FS. All enzymes with oligoduplex NSP lacking the specific target (Table 3) yielded complexes in the protein concentrations higher than with specific DNA (see Fig. 41). Table adapted from\(^{147}\).

It was shown previously that Ecl18kI specifically binds DNA in absence and presence of Ca\(^{2+}\) though with much higher affinity in the presence of Ca\(^{2+}\) (Fig. 31). Also it was shown that EcoRII-C (Fig. 34) and PspGI\(^{142}\) absolutely require Ca\(^{2+}\) ions for DNA binding. Therefore, the 2-AP influence on DNA
binding ability was tested for Ecl18kI, EcoRII-C and PspGI in the presence and absence of Ca\(^{2+}\) ions (Table 8 and Fig. 41).

Figure 41. Gel mobility shift analysis of the interactions between Ecl18kI, EcoRII-C, PspGI and oligoduplexes. Ecl18kI binding in the absence (A) and presence (B) of Ca\(^{2+}\) ions and EcoRII-C (C) and PspGI (D) binding in the presence of Ca\(^{2+}\) ions are illustrated. Oligoduplex T/A containing sequence CC(T/A)GG, oligoduplex T/2, containing sequence CC(T/2)GG or oligoduplex NSP lacking specific sequence (see Table 3) were used gel shift reactions. The binding reactions contained \(^{32}\)P-labeled oligoduplex (0.1 nM) and the protein at concentrations as indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions (see Methods). Gels were run in the presence of 0.1 mM of EDTA or 5 mM of Ca\(^{2+}\).

Gel shift experiments revealed that 2-AP incorporation into the target sequence had no effect on the affinity of Ecl18kI for cognate DNA (Table 8 and Fig. 41). To test if 2-AP containing oligoduplexes are substrates for the enzymes the cleavage ability of Ecl18kI, EcoRII-C and PspGI was tested. In the buffer supplemented with Mg\(^{2+}\) ions, Ecl18kI, EcoRII-C and PspGI cleaved
2-AP containing and lacking oligoduplexes with similar rates (Table 9 and data not shown). Thus, our data show that a fluorescent base 2-AP does not interfere with DNA binding and cleavage of EcL18kI, EcoRII-C and PspGI enzymes and can be used as a spectroscopic probe in the fluorescence experiments.

2.2. Fluorescence of EcL18kI-DNA in the presence of Ca$^{2+}$ ions

We titrated the 2-AP containing oligonucleotides with EcL18kI in the buffer supplemented with Ca$^{2+}$ ions and monitored the change of the 2-AP fluorescence intensity at 367 nm (Fig. 42A). The free oligoduplex containing 2-AP at the central position showed low signal because the fluorescence was quenched due to the base stacking interactions (Fig. 42B). When EcL18kI bound oligoduplex T/2, which contains 2-AP in the central position, at saturating concentrations the fluorescence intensity increased 6.5-fold. In contrast, only small changes were observed with oligoduplex 2FS, which carries 2-AP outside of the target site (Fig. 42B,C). The change of fluorescence intensity for the oligoduplex T/2 suggests that the 2-AP stacking with DNA bases is disrupted. It is compatible with the nucleotide flipping, which has been shown to enhance 2-AP fluorescence to varying extents in different systems$^{164-172}$.

2.3. Fluorescence of EcL18kI-DNA in the absence of Ca$^{2+}$ ions

Gel shift analysis showed highly decreased binding of EcL18kI to cognate DNA in the absence of Ca$^{2+}$ ions (Fig. 41A). However, we found that at much higher enzyme and DNA concentrations used in the fluorescence titration experiments, EcL18kI formed a binary complex with cognate DNA in the absence of Ca$^{2+}$ ions (Fig. 42D). The K$_d$ value obtained from the titration data was 52 ± 12 nM. In the Ca$^{2+}$-free buffer, EcL18kI binding to the oligoduplex T/2 increased the fluorescence intensity 28-fold at saturating protein concentrations, while only small changes were observed with oligoduplex 2FS (Fig. 42E,F). Therefore, the 2-AP signal in the buffer without Ca$^{2+}$ ions was ~4 times higher than the signal in the buffer supplemented with
Ca\textsuperscript{2+}. These results suggest that the structure of the complex formed in the presence of Ca\textsuperscript{2+} ions may differ from that formed without Ca\textsuperscript{2+}. It is possible, that in the presence of Ca\textsuperscript{2+} ions the flipped bases are more firmly trapped in the binding pockets resulting in the more pronounced fluorescence quenching while in the absence of Ca\textsuperscript{2+} ions they are more dynamic and therefore show higher fluorescence.

Figure 42. Fluorescence study of base flipping by Ecl18kI in solution\textsuperscript{147}. Titration of 250 nM 2-AP containing oligoduplex T/2 with increasing amounts of Ecl18kI in the presence (A) and absence (D) of Ca\textsuperscript{2+} ions, respectively. Corrected 2-AP emission spectra of Ecl18kI-DNA complexes (1250 nM Ecl18kI and 250 nM of oligoduplexes T/2 or 2FS) in the presence (B) and absence (E) of Ca\textsuperscript{2+} ions (see Methods). Diagrams in (C) and (F) illustrate the maximum fluorescence intensity values of the corrected fluorescence emission spectra presented in (B) and (E), respectively.

2.4. Fluorescence of the complex of Ecl18kI pocket mutant-DNA

2-AP fluorescence is often quenched in the hydrophobic environment of a protein\textsuperscript{171-174}. In the structure of Ecl18kI-DNA complex, the flipped nucleotides are accommodated in the pockets that are lined by tryptophan W61. In order to test whether W61 quenches 2-AP fluorescence, we replaced this residue with tyrosine and alanine. The Ecl18kI W61Y retained 0.07-0.2% specific activity compare to wt and W61A variant did not cleave cognate
oligoduplex T/A or the 2-AP containing oligoduplex T/2 radioactively labeled at either strand, but both retained the ability to bind both oligoduplexes albeit at 6-10-fold decreased affinity according to the gel shift assay (Table 8 and 9).

### Table 9. Ecl18kI mutants W61Y and W61A cleavage rates

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_1$, s$^{-1}$*</th>
<th>oligoduplex T/A</th>
<th>oligoduplex T/2</th>
<th>oligoduplex 2FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecl18kI wt</td>
<td>$0.22\pm0.05$</td>
<td>$0.33\pm0.04$</td>
<td>$0.29\pm0.05$</td>
<td></td>
</tr>
<tr>
<td>Ecl18kI W61Y</td>
<td>$4.5\pm0.03\times10^{-4}$</td>
<td>$2.7\pm0.1\times10^{-4}$</td>
<td>$2.0\pm0.4\times10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>Ecl18kI W61A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

* For clarity only the cleavage rates constants of the first DNA strand are presented. Both wt and W61Y Ecl18kI cleaved both strands with the same cleavage rate.

Binding of Ecl18kI W61Y to oligoduplex T/2 in the presence of Ca$^{2+}$ ions increased the 2-AP fluorescence intensity ~80-fold and W61A ~125-fold (Fig. 43) suggesting that both mutants are able to flip out the central nucleotide(s). 2-AP fluorescence in the ternary W61Y-DNA-Ca$^{2+}$ complex was 12-times higher and W61A-DNA-Ca$^{2+}$ even ~20 -times higher than in the wt Ecl18kI-DNA-Ca$^{2+}$ complex.

Thus, the W61Y and W61A mutants’ 2-AP fluorescence data support the assumption that low 2-AP fluorescence in the ternary complex with the wt protein is due to the quenching of the extruded base by stacking interactions with the W61 residue. However, one cannot exclude that increased space in the binding pocket of the mutants may allow a different orientation of the extrahelical 2-AP and affect the fluorescence intensity. Wt Ecl18kI-DNA complex a 2-AP fluorescence does
not necessarily indicate nucleotide flipping, since it could also be attributed to a less drastic DNA unstacking deformation.

However, the much higher 2-AP fluorescence increase in the W61 mutant-DNA-Ca\(^{2+}\) ternary complex compared to the wt-DNA-Ca\(^{2+}\) ternary complex strongly suggests that in the latter complex the flipped fluorophore from DNA duplex comes close to the indole ring of W61 for efficient quenching.

Moreover, the lack of activity of the Ecl18kI W61A mutant in the presence of Mg\(^{2+}\) ions and the nearly 10-fold reduced affinity to DNA are also consistent with a loss of interactions between the flipped nucleotide and the indole ring of W61. Otherwise the decreased binding and cleavage activity of W61 mutants suggest that although the pocket mutants manage to flip nucleotides, but W61 is important in enzyme binding and cleavage function.

2.5. Fluorescence of the EcoRII-C and PspGI complexes with DNA

To answer the intriguing question whether EcoRII-C and PspGI also flip the central W nucleotide from the recognition sequence CCWGG we performed 2-AP fluorescence assay for EcoRII-C and PspGI too. For 2-AP fluorescence studies purified EcoRII-C obtained by limited proteolysis was used (see section VI.10.).

EcoRII-C turned out oligoduplex T/2 binding was accompanied by ~12-fold increase of fluorescence (Fig. 44), suggesting that the base is extruded from the double
helix. PspGI titration of 2-AP containing oligonucleotide T/A showed an increase of 2-AP fluorescence and resulted in a 64-fold increase of the signal at saturation in comparison to the free oligoduplex (Fig. 44). Thus, according to the 2-AP assay, EcoRII-C and PspGI flip central nucleotide(s) while interacting with their recognition site like Ecl18kI.

Table 10. 2-AP fluorescence analysis of Ecl18kI, PspGI and EcoRII-C*

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>F(enzyme-T/2)/F(T/2)</th>
<th>F(enzyme-2FS)/F(2FS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecl18kI wt</td>
<td>6.5 (28.1**)</td>
<td>2.9 (1.1**)</td>
</tr>
<tr>
<td>Ecl18kI W61Y</td>
<td>76.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Ecl18kI W61A</td>
<td>123.6</td>
<td>2.2</td>
</tr>
<tr>
<td>EcoRII-C</td>
<td>62.9</td>
<td>1.6</td>
</tr>
<tr>
<td>PspGI</td>
<td>11.8</td>
<td>1.2</td>
</tr>
<tr>
<td>MvaI</td>
<td>0.9</td>
<td>0.8</td>
</tr>
</tbody>
</table>

* F stands for fluorescence intensity. All experiments (except, for wt Ecl18kI) were carried out in buffer supplemented with Ca$^{2+}$. F(enzyme-T/2)/F(T/2) and F(enzyme-2FS)/F(2FS) is a ratio of fluorescence of enzyme-DNA complex and free DNA for T/2 and 2FS oligoduplexes, respectively. The fluorescence intensity F values are of the corrected fluorescence emission spectra at fluorescence maximum (see Table 4). ** For wt Ecl18kI the fluorescence intensity values shown in the brackets were measured in the buffer supplemented without Ca$^{2+}$.

The 2-AP fluorescence increase upon addition of EcoRII-C and PspGI supports the idea that these enzymes also flip the central nucleotides of their target sequences. The 2-AP fluorescence intensity differences for Ecl18kI, EcoRII-C and PspGI (Table 10) likely reflect the nature of the enzyme pockets that accommodate the flipped bases. A structure based alignment indicates that these pockets are lined by tryptophan W61 in Ecl18kI, tyrosine Y226 in EcoRII-C and phenylalanine F64 in PspGI (Fig. 40). In the absence of the crystal structures of EcoRII-C and PspGI complexes with DNA, it remains unclear whether the differences in 2-AP fluorescence in the enzyme-DNA complexes are purely due to different hydrophobicities, or whether changes in the orientation of the aromatic side chains or other alterations around the flipped nucleotides contribute to the observed effects.

MvaI restriction enzyme recognizes the CC↓WGG sequence identical to that recognized by EcoRII-C and PspGI but cleaves it before the W nucleotide↑. In the MvaI-DNA complex structure the DNA conformation
does not deviate essentially from the canonical B-form and there is no evidence for base flipping\(^{35}\). Binding studies in solution revealed that MvaI binds 2-AP containing oligonucleotide T/2 (Table 8); however, this did not lead to an increase of 2-AP fluorescence (Fig. 44).

3. Nucleotide flipping phenomenon

The results of the 2-AP fluorescence assay provided direct evidence that Ecl18kI, EcoRII-C and PspGI un-stack bases at the center of their recognition sequences and flip them into the enzyme pockets\(^{147}\). Moreover, the 2-AP assay shows that nucleotide flipping observed in Ecl18kI co-crystal structure is also relevant in solution. Our data compliment genetic experiments on PspGI, which had provided indirect evidence that this restriction enzyme flips the central cytosine in the sequence CCCGG, which is related to the PspGI target sequence, but is not cleaved by PspGI\(^{161}\). Recently, Daujotyte et al.\(^{176}\) using chloracetaldehyde assay confirmed that the central cytosine in the CCSGG sequence is unstacked or flipped out by REases Ecl18kI and PspGI.

Base or nucleotide flipping has been shown to occur in many systems. First observed by X-ray crystallography for the bacterial C5-cytosine methyltransferases HhaI\(^{177}\) and HaeIII\(^{178}\), nucleotide flipping (base extrusion) has been documented later for other methyltransferases\(^{179-181}\), glycosylases\(^{182-185}\), glycosyltransferases\(^{186,187}\) and various DNA repair enzymes\(^{188-193}\).

We have provided first experimental evidence that REases that do not perform chemistry on DNA bases also flip nucleotides in their target sites. In contrast to the enzymes (methyltransferases, glycosylases, etc.) which flip bases to get an access to the DNA base to perform certain modification on the extruded base, Ecl18kI, EcoRII-C and PspGI extrude bases to achieve the specificity for the interrupted target site and adjust the cleavage pattern. It is tempting to suggest, that other enzymes or DNA binding proteins may employ this mechanism in their interactions with DNA.
X. Detailed mechanism of EcoRII interaction with DNA

EcoRII is Type IIE restriction enzyme which interacts with two copies of their recognition sequences\textsuperscript{26,102-104}. One of them serves as an allosteric activator which is not cleaved but stimulates cleavage of the second DNA copy. EcoRII was subjected for biochemical studies over more than 30 years. Several EcoRII mechanisms have been suggested, however, to our opinion, neither of them (see overviewed in section IV.) is able to explain the details of the allosteric activation mechanism of EcoRII. Therefore, we decided to re-investigate the mechanism of EcoRII REase.

1. Stoichiometry of the EcoRII-DNA complex: number of possible DNA binding interfaces in EcoRII

To elucidate the possible stoichiometry of the EcoRII-DNA complex we performed DNA binding experiments of EcoRII-N and EcoRII-C. To avoid possible DNA binding interference by the N- and C-terminal domains in full-length EcoRII, purified recombinant EcoRII-N and purified EcoRII-C obtained by limited proteolysis were used (see section VI.3.3.).

1.1. Gel shift analysis of EcoRII-N-DNA complex

Gel shift analysis of the specific 191 bp DNA fragment binding by the isolated EcoRII-N revealed two bands with different electrophoretic mobilities\textsuperscript{131}. It was suggested that the two bands of EcoRII-N-DNA complexes could be due to either non-homogeneity of the enzyme preparation or dimerization of EcoRII-N in the presence of specific DNA\textsuperscript{131}.

We have analyzed EcoRII-N binding to the 33 bp cognate oligoduplex using gel shift assay (Table 3 and see section VI.10.2.). We used the same recombinant version of the EcoRII-N containing an N-terminal (histidine)$_6$-tag which has been used in the binding experiments by Mucke et al\textsuperscript{131}. EcoRII-N bound readily to the 33 bp duplex with the recognition site ($K_d = 1.6 \pm 0.3$ nM), but not to the duplex lacking the cognate site (Fig. 45).
Noteworthy, EcoRII-N bound cognate DNA in the absence of divalent metals ions that were absolutely necessary for EcoRII-C binding (Fig. 34). Contrarily, EcoRII-N showed impaired DNA binding in the presence of divalent metal ions (data not shown). The apparent $K_d$ value for EcoRII-N binding to the 33 bp cognate duplex, however, was nearly $\sim$20-fold lower in comparison to the 191 bp cognate DNA fragment. The weak faster-moving band visible in the gel is presumably due to the truncated protein version that co-purifies with the EcoRII-N on the $\text{Ni}^{2+}$-column.

![Figure 45. Gel shift analysis of DNA binding by the EcoRII-N](image)

The binding reactions contained either the specific (A) or the non-specific (B) $^{33}$P-labelled 33 bp oligoduplex (1 nM), and the protein at concentrations (in terms of monomer) as indicated by each lane. Samples were analyzed by PAGE under non-denaturing conditions (see Methods). Gels were run in the presence of 0.1 mM EDTA.

Indeed, SDS–PAGE analysis reveals two bands present in the recombinant EcoRII-N preparation that differ slightly in the molecular weight (data not shown) but both react with the anti-EcoRII antibodies. In the absence of DNA EcoRII-N is susceptible to proteolytic degradation in vitro. Expression of EcoRII-N in *E. coli* may be accompanied by its partial proteolytic degradation at the C-terminus yielding a truncated protein version that co-purifies with the full-length EcoRII-N on a $\text{Ni}^{2+}$-column.

In the crystal structure the N-terminal domains of EcoRII make a dimer (Fig. 41). Analytical ultracentrifugation indicates that the isolated EcoRII-N in solution is a monomer which shows a tendency to dimerize at higher protein concentrations ($K_d$ value for protein dimerization $\sim1.85 \pm 0.19$ µM). Otherwise, the EcoRII-N shows DNA binding in the low nanomolar range of
protein concentrations (Fig. 38) far below the $K_d$ for the EcoRII-N dimer suggesting that a monomer binds a single DNA copy. If, however, EcoRII-N dimerizes upon DNA binding, then protein-DNA complex may include one or two DNA copies per dimer.

1.2. Gel filtration analysis of the EcoRII-N-DNA complex

In the next step to determine the stoichiometry of the EcoRII-N-DNA complex we employed gel filtration analysis (see section VI.14.). Apparent $M_w$ values were determined from the elution volumes by reference to a series of standard proteins of known $M_w$.

![Figure 46. Gel filtration of EcoRII-N on Super SW2000 column](image)

Figure 46. Gel filtration of EcoRII-N on Super SW2000 column\textsuperscript{160}. (A) Molecular masses of wt EcoRII, EcoRII-N and EcoRII-N-DNA complexes. The apparent $M_w$ of wt EcoRII (red circle), EcoRII-N (blue open inverted triangle), specific 16 bp DNA (green open square) and EcoRII-N–DNA complex (violet open triangle) were calculated by interpolation from the standard curve obtained using a set of proteins of known $M_w$ (black circles). (B) Gel filtration analysis of EcoRII-N–DNA complexes at various protein:DNA ratios. Size-exclusion chromatography was performed as described in Methods.

EcoRII-N in the absence of DNA eluted at a volume that corresponds to an apparent $M_w$ of 23 kDa (Fig. 46, blue line). The apparent $M_w$ of EcoRII-N determined by gel filtration is thus close to that expected for a monomer: the theoretical molecular mass of the EcoRII-N is 22.9 kDa. When the 16 bp duplex (Table 3) was applied to the column, it eluted as an apparent $M_w$ of ~27 kDa (Fig. 46, green line). The apparent $M_w$ of the duplex is 2.5-times higher than its actual molecular mass of 10.7 kDa. This is due to the duplex cylindrical shape and much higher frictional ratio than the spherically shaped
standard proteins. In the gel filtration experiment of the complex of EcoRII-N and 16 bp specific oligonucleotide at EcoRII-N (monomer):DNA (duplex) ratio 2:1 (Fig. 46, violet line) the mixture eluted in two peaks: one corresponding to an apparent $M_w$ of $\sim 38$ kDa (theoretical $M_w$ of EcoRII-N monomer-DNA is 33.6 kDa); another at the same time as the free EcoRII-N protein, again denoting an apparent $M_w$ of $\sim 22$ kDa. The two-fold excess of EcoRII-N over the DNA thus presumably leaves one fraction of the protein in its free state, the $\sim 22$ kDa species, and another fraction contains monomer bound to one duplex, the $\sim 38$ kDa species. Further gel filtration experiment used the same amount of EcoRII-N but with increased amounts 16 bp oligoduplex, so that the ratio of EcoRII-N (monomer):DNA (oligoduplex) was 1:2. This eluted in two peaks again: one corresponding to an apparent $M_w$ of $\sim 38$ kDa (the monomer bound to one duplex), another at the same time as the free DNA (an apparent $M_w$ of $\sim 27$ kDa) (Fig. 46, orange line).

Hence, **EcoRII-N is a monomer in solution and EcoRII-N binds DNA as a monomer with 1:1 stoichiometry**\(^1\).

### 1.3. Gel filtration analysis of the EcoRII-C-DNA complex

In the gel filtration experiment wt EcoRII elutes as 73 kDa (Fig. 46) $M_w$ that corresponds to EcoRII dimer (theoretical $M_w$ $\sim 90.5$ kDa) in agreement with previously published data\(^1,1^{17}\).

EcoRII-C has a typical restriction enzyme fold according to the crystal structure of apo-EcoRII\(^7\). Therefore, it is likely, that the C-terminal domain of EcoRII interacts with DNA as a dimer, similarly to other Type II restriction enzymes. To test this assumption we have analyzed stoichiometry of the EcoRII-C-DNA complex by gel filtration. For gel filtration studies purified EcoRII-C, obtained by limited proteolysis, was used (see sections VI.3.3. and VI.10.).

Apparent $M_w$ values were determined from the elution volumes by reference to a series of standard proteins of known $M_w$ similarly as described
for EcoRII-N (see above). EcoRII-C in the absence of DNA eluted at a volume that corresponds to an apparent $M_w$ of 68 kDa (Fig. 47, blue line). The apparent $M_w$ of EcoRII-C determined by gel filtration is similar to that expected for the dimer (theoretical $M_w$ 52 kDa). When the 12 bp duplex (Table 3) was applied to the column, it eluted as an apparent $M_w$ of 18 kDa (Fig. 47, green line). The difference between the experimental and theoretical (8 kDa) $M_w$ of the duplex could be explained with same duplex cylindrical shape (see above). In the gel filtration experiment of the complex of EcoRII-C and 12 bp specific oligonucleotide at EcoRII-C (dimer):DNA (duplex) ratio 0.3:1 (Fig. 47, violet line) the mixture eluted in two peaks: one corresponding to an apparent $M_w$ of $\sim$59 kDa (theoretical $M_w$ of EcoRII-C dimer-DNA is 60 kDa); another at the same time as the free oligoduplex, again denoting an apparent $M_w$ of $\sim$18 kDa. No additional peaks corresponding to the different complexes were observed in other EcoRII-C (dimer):DNA ratios (data not shown). The free EcoRII-C protein eluted after the EcoRII-C-DNA complex suggesting that EcoRII-C in the presence of DNA makes more compact globule.

**Figure 47.** Gel filtration of EcoRII-C on Superdex 200 HR 10/30 column. (A) Molecular masses of EcoRII-C and EcoRII-C-DNA complexes. The apparent $M_w$ of EcoRII-C (blue open inverted triangle), specific 12 bp DNA (green open square) and EcoRII-C-DNA complex (violet open triangle) were calculated by interpolation from the standard curve obtained using a set of proteins of known $M_w$ (black circles). (B) Gel filtration analysis of EcoRII-C-DNA complex. Size-exclusion chromatography was performed as described in Methods.

Hence, according to gel filtration EcoRII-C is a dimer in solution and binds one copy of oligoduplex. Since the EcoRII N-terminal domain is a
monomer that also binds a single copy of cognate DNA the full-length EcoRII contains three putative DNA binding interfaces: one at the C-terminal domain dimer and two at each of the N-terminal domains\textsuperscript{160}.

2. EcoRII kinetics studies to verify three DNA binding model

Structural and biochemical studies demonstrate a modular architecture of EcoRII (Fig. 48A)\textsuperscript{79,131,160}.

![Figure 48. Scheme illustrating possible DNA binding models of EcoRII\textsuperscript{195}.](image)

(A) EcoRII apo-structure (PDB ID 1NA6\textsuperscript{79}). N-terminal domains are depicted in red and C-terminal domains in blue. (B) Modular architecture of EcoRII is represented by different shapes and colors: the N-terminal allosteric DNA binding domains are shown as red ovals, the C-terminal catalytic domains are shown as blue kidney shape structures. DNA molecules are depicted as grey rectangular shapes. The ‘two-site binding’ model assumes that the N-terminal domain dimer binds a single DNA copy while the C-terminal domain dimer binds the second one. ‘Three-site binding’ model assumes that each N-terminal domain binds a separate DNA copy while the C-terminal domain dimer binds the third one.

Results of the gel filtration and gel shift analysis suggest that each N-terminal domain of EcoRII interacts with a single DNA copy, and the dimer of C-terminal domains binds one more DNA copy, hence wt EcoRII dimer possesses three potential DNA binding sites (Fig. 48B, left cartoon). This hypothesis is in a sharp contrast to the currently accepted two DNA binding model for EcoRII\textsuperscript{103,104,194} (Fig. 48B, right cartoon), which is similar to the model suggested for another Type IIE restriction enzyme NaeI\textsuperscript{66}.

In order to test the functional significance of two possible models we have studied the EcoRII cleavage of plasmid DNA substrates.
2.1. Construction of plasmid DNA substrates for kinetics studies

Supercoiled plasmids bearing one or two copies of the recognition sequence proved to be useful tools for testing whether a restriction enzyme requires two sites for its optimal activity: the orthodox homodimeric REases cleave both plasmid substrates with equivalent rates, while restriction enzymes requiring two recognition sequences strongly prefer the two-site substrates. In order to estimate the number of sites required for the optimal EcoRII catalysis we have engineered plasmids containing a single, two and three EcoRII target sites. Since the cleavage rate at the particular recognition site often depends on the sequences flanking that site, it is essential to ensure that target sites in the two and three-site plasmids are identical with each other and to the site in the one-site DNA, in terms of their sequence context. Multicopy plasmids such as pUC18 or pBR322 contain multiple recognition sites for EcoRII and therefore require elimination of several EcoRII sites to yield one, two and three-site plasmids, respectively. Unfortunately, both pUC18 and pBR322 contain at least two CCWGG sites in their ori region. Our attempts to eliminate the EcoRII sites in the ori region of pUC18 by site-directed mutagenesis were unsuccessful (data not shown).

![Figure 49](image-url)  
**Figure 49.** Plasmids containing a single, two and three EcoRII recognition sites used in the cleavage experiments. EcoRII sites are shown as black boxes. All sites have the same orientation and identical flanking sequences (see Methods for the details). Numbers in the brackets indicate relevant positions of the recognition sites in respect to the arbitrary EcoRII site chosen as zero.

Therefore, plasmids pEcoRII-1, pEcoRII-2 and pEcoRII-3 (Fig. 49) containing, respectively, one, two and three EcoRII target sites with identical flanking sequences and identical site orientation were engineered on the basis...
of the low-copy number pKpn2I-ori plasmid (a derivative of pKpn2RM3.7
plasmid\textsuperscript{148}), which contains a single EcoRII site in the plasmid p15A-like ori
region\textsuperscript{149}. Plasmids with two or three sites were engineered by inserting one or
two copies of synthetic oligodeoxynucleotide duplex bearing the EcoRII
recognition sequence into the single-site plasmid DNA as described in section
VI.16.1.

EcoRII cleavage of these plasmids was studied under different
experimental conditions (see section VI.16.2.). The individual DNA cleavage
steps during the catalytic cycle can be monitored in a single-turnover reaction
([E]>[S]). Contrary, in a multiple-turnover reaction ([E]<[S]) under the
substrate excess conditions only small fraction of enzyme is bound to the DNA
and observed reaction products are then the products liberated from the
enzyme at the end of the reaction rather than intermediates still bound to the
enzyme.

2.2. EcoRII cleavage of plasmid DNA at \([E]\geq[S]\)

For each substrate, rates were measured from the decline of substrate
concentration with time. A single exponential was fitted to the time-course of
supercoiled (SC) DNA decay to obtain a value for the apparent first order
reaction rate constant \(k_{obs}\). EcoRII reaction rates on 1-, 2- and 3-site plasmid
DNA substrates were measured at two different enzyme concentrations. At the
optimal enzyme concentration (2 nM of enzyme and 1.5 nM of 2- or 3-site
plasmid DNA) experimentally determined cleavage rates of SC plasmids
reached their highest value. At the enzyme excess over DNA (100 nM EcoRII
and 1.5 nM) cleavage rates for the 2- and 3-site plasmids became significantly
lower in comparison to the optimal conditions. Cleavage reactions were
initiated either by adding the Mg-acetate solution to the premixed
enzyme-DNA solution or by adding the enzyme premixed with Mg-acetate to
the DNA solution. No significant difference was observed between the two
mixing procedures (data not shown).
Figure 50. EcoRII restriction enzyme on 1-, 2- and 3-site plasmid substrates. The cartoons above the graphs illustrate various DNA forms that can exist during EcoRII reactions on plasmids with 1, 2 or 3 EcoRII sites. Plasmid DNA concentration was fixed at 1.5 nM while the enzyme concentration varied depending on the reaction conditions. Reactions at optimal conditions (A-C) contained 2 nM EcoRII, and reaction at enzyme excess conditions (D-F) contained 100 nM of EcoRII. Reactions in the presence of oligonucleotide on 1-site plasmid (G) contained 100 nM EcoRII and 3000 nM of cognate 33 bp oligonucleotide, on 2-site (H) and 3-site (I) plasmids 2 nM of EcoRII and 60 nM of cognate 33 bp oligonucleotide, respectively. The reactions were performed at 25 °C in 33 mM Tris-acetate (pH 7.8), 66 mM K-acetate, 10 mM Mg-acetate and 0.1 mg/ml BSA. Samples were removed from the reactions at timed intervals, quenched with loading dye solution containing EDTA and analyzed as
In the case of the 1-site plasmid cleavage at EcoRII concentration of 2 nM only ~ 10% of SC plasmid DNA is transformed into the OC DNA form (nicked DNA) after 1 h incubation (Fig. 50A). Further increase of EcoRII concentration (up to 100 nM) has no effect on the cleavage rate or reaction pattern (Fig. 50D). Hence, EcoRII cleavage of 1-site plasmid DNA even under enzyme excess over DNA is very slow confirming previous reports\(^{18}\).

In contrast to the 1-site plasmid, the EcoRII cleavage of the 2-site plasmid at the optimal enzyme concentration is much faster \((k_{\text{obs}} = 1.3\pm0.1 \text{ min}^{-1})\) and yields a set of different reaction products (Fig. 50B). The SC DNA is initially converted to the OC DNA nicked at one of the sites that is further transformed into the full-length linear (FLL) DNA form cleaved at both strands at a single site. A small fraction of the final reaction product L2 – linear DNA cleaved at both sites is also formed albeit at slow rate. Linear DNA cleaved at both sites may result due to the reaction \textit{in trans} (see, below). Alternatively, cleaved EcoRII site may activate cleavage at the second site. Indeed, oligonucleotide duplexes with simulated EcoRII ends are capable of enzyme activation\(^{20}\) similarly to SgrAI\(^{197,198}\) and Bse634I\(^{32}\) restriction enzymes.

EcoRII cleavage pattern of the 3-site plasmid DNA at the optimal enzyme concentration shows one important difference compared to the 2-site plasmid (Fig. 50C). \textbf{SC DNA form of the 3-site plasmid is directly converted into the FLL form by a concerted cleavage at both strands at a single site, while the 2-site plasmid is first converted into the nicked OC DNA intermediate.}

described in Methods to determine the amounts of the following forms of the plasmid DNA: supercoiled DNA (SC), open triangles (red); open-circular DNA (OC), filled circles (blue); linear DNA cut at one EcoRII site (FLL), open squares (green); linear DNA cut at two EcoRII sites (L2), filled inverted triangles (magenta); linear DNA cut at three EcoRII sites (L3), open diamonds (cyan). In the case of the 2-site and 3-site plasmid substrates, the OC and FLL (in case of 2-site plasmid) and OC, FLL and L2 (in case of 3-site plasmid) forms of DNA may contain additional nick(s) at the second (third) recognition site, however these forms are not resolved in the gel. The continuous line (in (B), (C), (E), (F), (G), (H) and (I)) represents the best fit of a single exponential to the time course of the supercoiled plasmid DNA cleavage.
Cleavage rates of SC forms of the 3- and 2-site plasmid DNA, however, show only a small difference ($k_{obs(3-site)} = 3.2\pm0.1 \text{ min}^{-1}$ and $k_{obs(2-site)} = 1.3\pm0.1 \text{ min}^{-1}$, respectively). Upon formation of the double strand break at the single site the remaining FLL containing two EcoRII recognition sites, is converted into the L2 form presumably following the reaction pathway described for the 2-site plasmid substrate.

In the case of the 1-site plasmid increasing concentrations of EcoRII do not change either the reaction rate or product distribution pattern indicating that DNA is saturated by the enzyme (Fig. 50A,D).

On the 2-site plasmid, an increase of EcoRII concentration from 2 nM to 100 nM results in a significant decrease of SC DNA cleavage rate and a dramatic decline in the yield of FLL form (Fig. 50B,E). Similar effect is observed for the 3-site plasmid. Indeed, the OC DNA form becomes a predominant reaction product and the amount of FLL is significantly decreased (Fig. 50F). The reaction pattern of the 3-site plasmid DNA cleavage at enzyme excess over DNA becomes similar to that of the 2-site plasmid DNA (Fig. 50E).

2.3. EcoRII cleavage reactions in trans

EcoRII cleavage of the 2- or 3-site plasmid DNA indicates that the presence of at least two recognition sites in cis stimulates DNA cleavage. It was previously reported that the cognate oligonucleotides bearing the EcoRII recognition site stimulate cleavage of the resistant plasmid DNA substrates with EcoRII sites occurring at a very low frequency\textsuperscript{19,20}. Therefore, we have studied the effect of oligonucleotides on the EcoRII cleavage of 1-, 2- or 3-site plasmids. Addition of the oligonucleotide to the reaction mixture containing a plasmid DNA and enzyme should result in EcoRII distribution between different DNA-bound forms: an enzyme bound to the plasmid, an enzyme bound simultaneously both to the plasmid and oligonucleotide and an enzyme bound solely to the oligonucleotide. The amount of the different enzyme bound forms will depend on the enzyme concentration, oligonucleotide/plasmid DNA
ratio and equilibrium binding constants for the enzyme-plasmid and enzyme-oligonucleotide complexes. Therefore, we have studied the effect of the oligonucleotides on plasmid DNA cleavage at the various reaction conditions.

**EcoRII cleavage of the 1-site plasmid DNA in the presence of cognate oligonucleotide**

EcoRII cleavage of the 1-site plasmid DNA under [E]>[S] conditions (1.5 nM 1-site plasmid and 100 nM of EcoRII) is very slow and yields only small amount of the OC DNA form nicked at one strand (Fig. 50D). Addition of 3000 nM of the cognate 33 bp oligonucleotide (Table 3) to the reaction mixture results in the dramatic increase of the cleavage rate and yields primarily a linear DNA cleaved at both strands (Fig. 50G). Decrease of the oligonucleotide concentration to 500 nM results in the simultaneous formation of both OC and FLL DNA forms (data not shown). Non-cognate oligonucleotide lacking the EcoRII-recognition site has no effect on the cleavage rate either at 500 nM or 3000 nM concentrations (data not shown).

**EcoRII cleavage of 2-site plasmid DNA in the presence of cognate oligonucleotide**

EcoRII cleavage of the 2-site plasmid containing two recognition sites *in cis* at equimolar amounts of enzyme (2.0 nM) and DNA (1.5 nM) yields initially the OC DNA that is further converted into a FLL cleaved at a single site (Fig. 50B). Addition of 60 nM of the cognate oligonucleotide results in the change of the reaction pattern (Fig. 50H). Indeed, in the presence of oligonucleotide, the SC form of the 2-site plasmid is simultaneously converted into nearly equimolar mixture of OC and FLL DNA forms, which are further cleaved to yield the final reaction product – linear DNA cleaved at both sites (Fig. 50H).

**EcoRII cleavage of the 3-site plasmid DNA in the presence of cognate oligonucleotide**
Cleavage pattern of the 3-site plasmid DNA in the presence of the cognate oligonucleotide (Fig. 50I) is nearly identical to that of the 3-site plasmid DNA in the absence of oligonucleotide. In both cases the SC plasmid DNA is converted into the FLL DNA form and virtually no OC DNA cleaved at one strand is formed. The cleavage rate of the FLL form containing two recognition sites is slightly increased in comparison to the cleavage in the absence of oligonucleotide. That is not surprising keeping in mind that the cognate oligonucleotide in trans stimulates cleavage of the 2-site DNA (Fig. 50H).

2.4. EcoRII cleavage of plasmid DNA at $[E]<[S]$

Under the substrate excess conditions (1.5 nM of plasmid and 0.1 nM of EcoRII) only a small fraction of enzyme is bound to the DNA and observed reaction products are then the products liberated from the enzyme at the end of the reaction rather than intermediates still bound to the enzyme. Under these conditions EcoRII shows virtually no cleavage of the 1-site plasmid DNA (Fig. 51A).

EcoRII cleavage of the 2-site plasmid under the multiple turnover conditions liberates mainly OC DNA cut just in one strand; only a small fraction of the DNA was cut at the both strands at a single site (Fig. 51B). Hence, nearly all SC DNA form of the 2-site plasmid is converted to the OC intermediate, which accumulates in solution. Therefore, to generate FLL DNA form cut at a single site, EcoRII has to re-bind the OC form and cleave the second phosphodiester bond at the same site in a separate catalytic cycle. The EcoRII cleavage pattern of the 3-site plasmid under the multiple turnover conditions (Fig. 51C) is similar to that of the 2-site plasmid with the only important difference that the FLL DNA form cleaved at both strands at a single site rather than nicked OC DNA is liberated into the solution. Under these conditions only a trace amount of L2 DNA and L3 DNA forms cut at two and three recognition sites are formed. Reaction product analysis by agarose gel electrophoresis does not allow discriminating between the linear and nicked
linear DNA forms. To generate the doubly cut L2 DNA from the linear FLL DNA form of the 3-site plasmid, EcoRII most likely follows the mechanism established for the 2-site plasmid (see above).

Figure 51. EcoRII and NaeI cleavage of plasmid DNA substrates under the multiple-turnover reaction conditions. The cartoons above graphs illustrate the various DNA forms that can exist during EcoRII and NaeI reactions on plasmids with 1, 2 or 3 cognate sites. In the case of the 2-site and 3-site plasmid substrates, the OC and FLL (for 2-site plasmid) forms and OC, FLL and L2 (for 3-site plasmid) forms of DNA may contain additional nick(s) at the second (or third) recognition site, respectively; however these forms are not resolved in the gel. Panels (A)-(C) display EcoRII cleavage of 1-, 2- or 3-site plasmids. Plasmid DNA concentration was fixed at 1.5 nM, EcoRII concentration was 0.1 nM. The reactions were performed at 25°C in 33 mM Tris-acetate (pH 7.8), 66 mM K-acetate, 10 mM Mg-acetate and 0.1 mg/ml BSA and analyzed as described in Figure 50 legend. Different DNA forms are color-coded the same as in Figure 50 legend. Panels (D) and (E) display NaeI cleavage of 1-site and 2-site plasmid and was adapted from (30) (with permission J. Mol. Biol.).

In this case EcoRII first nicks at one of the recognition sites of FLL DNA during a single binding event, dissociates from the DNA and has to re-associate to cleave next phosphodiester bond at one of the sites to yield either linear DNA L2 or FLL nicked at the second site. The doubly nicked FLL
form migrates in the gel with the same mobility as an intact or nicked at a single site FLL form and could not be resolved in the agarose gel under standard conditions.

2.5. Plasmid DNA cleavage by EcoRII-C

It was shown previously that the isolated C-terminal domain of EcoRII efficiently cleaves single-site DNA\textsuperscript{131}. Therefore, we have tested cleavage of plasmids containing 1-, 2- and 3-recognition sites by the isolated recombinant EcoRII-C.

Under the multiple turnover conditions EcoRII-C rapidly transforms 1-, 2- and 3-site plasmids into the final reactions products (Fig. 52A-C). In the case of 2- and 3-site plasmids DNA cut at a single site is released and accumulates in solution. The subsequent binding and cleavage result in the formation of final reaction products. Noteworthy, ~40% of FLL DNA accumulates during the EcoRII-C multiple turnover reactions on the 2- and 3-site substrates (Fig. 52B,C). This is exactly the amount of the reaction intermediate expected for an independent action of the restriction enzyme at each site of the multi-site plasmid substrate\textsuperscript{199-201}. The EcoRII-C reactions under the single turnover conditions were too fast to be analyzed by the conventional methods.

2.6. Plasmid DNA cleavage by PspGI

PspGI is a homodimer; however, it is not stimulated by simultaneous binding to two recognition sites\textsuperscript{142}. We have studied cleavage of plasmids containing 1, 2 and 3 recognition sites by PspGI under the enzyme excess conditions. We have found that PspGI cleaves effectively 1-, 2- and 3-site plasmids (Fig. 52D-F). The reactions proceed directly to FLL suggesting that both DNA strands at the recognition site are cleaved in a concerted fashion. In the case of 2- and 3-site plasmids, linear DNA forms containing single and two recognition sites, respectively, are formed as reaction intermediates.
Figure 52. Cleavage of supercoiled 1-site, 2-site and 3-site plasmids by EcoRII-C and PspGI restriction enzymes. The cartoons above the graphs illustrate the various DNA forms that can exist during EcoRII-C and PspGI reactions on plasmids with 1, 2 or 3 EcoRII sites. In the case of the 2-site and 3-site plasmid substrates, the OC and FLL (for 2-site plasmid) forms and OC, FLL and L2 (for 3-site plasmid) forms of DNA may contain an additional nick at the second (or third) recognition site, respectively; however these forms are not resolved in the gel. Plasmid DNA concentration was fixed at 1.5 nM. In (A), (B) and (C) the reaction mixture contained 0.2 nM of the C-terminal domain EcoRII (EcoRII-C). In (D), (E) and (F) the reaction mixture contained 20 nM of PspGI. The reactions were performed at 25°C in 33 mM Tris-acetate (pH 7.8), 66 mM K-acetate, 10 mM Mg-acetate and 0.1 mg/ml BSA and analyzed as described in Figure 50 legend. Different DNA forms are color-coded the same as in Figure 50 legend. The continuous line represents the best fit of a single exponential to the time course of the supercoiled plasmid DNA cleavage (in (D), (E) and (F)).

Under the multiple turnover conditions, however, the linear DNA cleaved at a single site slowly accumulated in solution suggesting that PspGI dissociates after the cleavage at a single site and the dissociation becomes a rate-limiting step under the multiple-turnover conditions (data not shown).
2.7. Summary of EcoRII reactions on the 1-, 2- and 3-site plasmids

Experimental data presented in Figs. 50 and 51 clearly demonstrate that EcoRII displays distinct reaction patterns on plasmids with 1-, 2- and 3- copies of the recognition sequence. Indeed, the 1-site plasmid is virtually refractory to EcoRII cleavage at all enzyme concentrations tested (Figs. 50A,D and Fig. 51A) and becomes susceptible to cleavage only upon addition of the cognate oligonucleotide (Fig. 50G).

Contrary to the 1-site plasmid, the 2-site plasmid substrate is readily cleaved by EcoRII under optimal ([E]≈[S]) reaction conditions (Fig. 50B), yielding first nicked OC DNA form that is further slowly transformed into FLL and L2 DNA forms containing double-strand breaks. Cleavage rate of the 2-site plasmid however becomes significantly slower under the excess of EcoRII in the reaction mixture (Fig. 50E). The dominant reaction product released after each turnover of EcoRII on the 2-site substrate in a steady-state reaction is OC DNA containing single-strand break(s) (Fig. 51B). Linear DNA form FLL, however becomes the major reaction product upon EcoRII cleavage of the 2-site plasmid under the optimal reaction conditions in the presence of the cognate oligonucleotide (Fig. 50H).

Similarly to the 2-site substrate, EcoRII also readily cleaves the 3-site plasmid (Fig. 50C and Fig. 51C). The steady-state reaction rates measured from the decline of supercoiled plasmid substrate concentration with time are similar both for 2-site and 3-site plasmids (\(k_{\text{cat(2-site)}} = 0.3\pm0.1 \text{ min}^{-1}, k_{\text{cat(3-site)}} = 0.6\pm0.1 \text{ min}^{-1}\)). Comparable reaction rates for EcoRII cleavage of 2-site and 3-site plasmids are observed also under the equimolar concentrations of the substrate and the enzyme (\(k_{\text{obs(2-site)}} = 1.3\pm0.1 \text{ min}^{-1}, k_{\text{obs(3-site)}} = 3.2\pm0.1 \text{ min}^{-1}\)). Of note is that similarly to the 2-site plasmid, EcoRII cleavage of the 3-site substrate is impeded under excess of the enzyme in the reaction mixture (Fig. 50F). The only important difference between EcoRII reactions on the 2- and 3-site plasmids is that on the 3-site plasmid the primary reaction product is FLL form cut at both strands while on the 2-site dominates OC DNA form.
containing single-strand breaks. Furthermore, FLL DNA is virtually the only reaction product that accumulates in solution during the EcoRII reaction on the 3-site substrate under steady-state reaction conditions (Fig. 51C), indicating that EcoRII makes one double-strand break per turnover before dissociating from the DNA. Indeed, the reactions at equimolar enzyme and 3-site plasmid concentrations demonstrate that large fraction of the supercoiled substrate is directly converted into the FLL product (Fig. 50C), suggesting that EcoRII cleaves both DNA strands at one of the sites in a concerted reaction.

The stimulation of EcoRII cleavage by the second recognition site introduced into the plasmid reported here is in a good agreement with the previous data\textsuperscript{19,20}. According to the currently prevailing ‘2-site binding’ model (Fig. 29 and Fig. 48B)\textsuperscript{103,104,194} binding of the single cognate DNA copy by the N-terminal domains of EcoRII stimulates binding and cleavage of the second DNA copy at the C-terminal domain dimer. However, this model is insufficient to account for the striking differences in the reaction products formed during the reaction course of EcoRII on the 2-site and the 3-site plasmid substrates (Fig. 50 and Fig. 51).

The differences in the reaction patterns, however, can be explained by the ‘3-site binding’ model (Fig. 48B). Zhou et al.\textsuperscript{79} proposed that EcoRII activity is regulated by autoinhibition. In the apo-form of EcoRII two N-terminal effector domains sterically block DNA access to the catalytic/binding site of the enzyme located at the cleft between the C-terminal domains (autoinhibition). Authors suggested that in the presence of DNA with EcoRII recognition sequence DNA binds to the effector clefts first. Effector binding induces domain re-orientation by moving two N-domains away from the catalytic half-sites and moving two C-domains closer to form an active dimer with one substrate DNA duplex (activation). However, according the ‘3-site binding’ model, binding of single DNA copy at the N-terminal domain of EcoRII reconfigures the protein conformation and opens an access for the cognate DNA enabling catalysis. But only \textit{synapsis of two DNA sites through the}
**N-terminal halves results in a concerted double-strand break at the C-terminal domain dimer**

In the case of 1-site plasmid, one of the N-terminal domains binds a solitary recognition site and presumably activates the catalytic domain for DNA binding/cleavage. However, no cleavage of 1-site plasmid DNA occurs since the C-terminal domain can bind and cleave the only EcoRII recognition site that is already occupied by the N-terminal domain (Fig. 53). The low residual EcoRII activity on the one-site substrate (Fig. 50A,D) may be due to the \textit{in trans} cleavage by the activated EcoRII molecule.

**Figure 53. Schematic illustration of EcoRII interactions with 1-, 2- and 3-site plasmids under different experimental conditions**

Possible EcoRII complexes formed on the 1-, 2- and 3-site plasmids under the \textit{optimal} ([E]=[S]) or \textit{enzyme excess} ([E]>>[S]) conditions are depicted. DNA cleavage products formed during the first catalytic cycle are indicated below the corresponding complexes. The N-terminal domain of EcoRII is shown in red and C-domain in blue, black rectangles in plasmids represent the EcoRII recognition sites.

EcoRII binding to the 2-site plasmid through one of the N-terminal domains opens DNA binding surface at the C-terminal domain, resulting in the competition for the second binding site between the N- and C-terminal domains (Fig. 53). Since the catalytic site of EcoRII is located at the
C-terminal domain, only synapsis of two recognition sites through one of the N-terminal domains and the C-terminal domain dimer yields the catalytically competent complex (Fig. 53). On a 2-site plasmid EcoRII makes only single-strand breaks during each enzyme turnover (Fig. 51B). It is likely, that due to the binding competition the C-terminal domain spends only part of the time bound to the cognate site and the lifetime of the catalytically productive complex becomes too short to allow the cleavage of both DNA strands during the single turnover (Fig. 53). Alternatively, cognate DNA binding by both N-terminal domains may be absolutely required for concerted cleavage even if the lifetime of the productive complex is long enough for double strand break to occur.

Contrary to the 2-site substrate, EcoRII binding to the 3-site substrate enables simultaneous binding of all three recognition sites by EcoRII dimer (Fig. 53). In this case the C-terminal catalytic domain dimer cleaves both DNA strands at a single site in a concerted manner (Fig. 50C) and after each enzyme turnover FLL DNA form cleaved at both strands is released into solution (Fig. 51C). The reaction intermediate is FLL DNA with two sites, which is further cleaved by EcoRII at one of the sites to yield the L2 product (Fig. 51C).

The ‘3-site binding’ model for EcoRII-DNA interaction also allows to rationalize differences in the reaction patterns and velocities of plasmid DNA cleavage at the elevated enzyme concentrations. Indeed, under the enzyme excess, more than one EcoRII dimer may interact with a single copy of the 2- or 3-site plasmid substrate via the N-terminal domains, thereby interfering with C-terminal domain binding and cleavage (Fig. 53). This accounts for the observed decline in hydrolysis rate and higher amount of OC products formed (Fig. 50E,F).

The changes in the reaction products and velocities of plasmid DNA cleavage by EcoRII observed in the presence of cognate oligonucleotide are also consistent with the ‘3-site binding’ model. Addition of the cognate oligonucleotide in trans into the reaction mixture enables formation of the
catalytically competent EcoRII molecules due to the oligonucleotide binding to one or two N-terminal domains. The C-terminal catalytic domain of such oligonucleotide-activated molecules can cleave \textit{in trans} plasmid substrates even with a single recognition site (Fig. 50G). On a 2-site plasmid addition of cognate oligonucleotide \textit{in trans} also stimulated cleavage and increased the yield of FLL products (Fig. 50B,H). Such changes in the reaction pattern and velocity suggest the presence of reactive EcoRII molecules bound to the three copies of cognate DNA in solution similarly to the EcoRII reaction on the 3-site plasmid. On the other hand, cognate oligonucleotide had little effect on the cleavage rate and distribution of the initial reaction products in case of the 3-site plasmid (Fig. 50C and I). This is not surprising if one assumes that the EcoRII complex with recognition sites located \textit{in cis} is more stable than that \textit{in trans}, that is common for many restriction enzymes interacting with multiple recognitions sites\textsuperscript{202}.

Contrary to wt EcoRII, the 1-, 2- and 3-site plasmids are cleaved by the isolated EcoRII-C at equal rates similarly to PspGI, supporting previous findings that EcoRII-C \textit{per se} is an orthodox Type IIP restriction enzyme and does not require interactions with multiple recognition sites for its optimal activity (Fig. 52)\textsuperscript{131}.

\textbf{2.8. EcoRII and NaeI: Type IIE restriction enzymes that follow different mechanisms}

NaeI is another member of Type IIE restriction enzymes and needs two copies of the recognition site for optimal activity\textsuperscript{134,135,203}. It recognizes the CCCGGG sequence and cleaves it yielding blunt-ends. The NaeI reactions on 1- and 2-site plasmids under the multiple turnover conditions reported previously\textsuperscript{30} (Fig. 51D,E) were compared with EcoRII reactions on 1- and 2-site plasmids. Under the multiple turnover conditions NaeI, like EcoRII, cleaves plasmid DNA containing a single recognition site very slowly (Fig. 51D). On the 2-site plasmid NaeI cleaves concertededly both DNA strands at a single site yielding the FLL form that is released into the solution Fig. 51E,
while in case of EcoRII only OC products are observed (Fig. 51B). However, the FLL form is released into the solution upon EcoRII cleavage of the 3-site plasmid (Fig. 44C). Thus, while NaeI needs to interact with two recognition sites for the concerted DNA cleavage at a single site, EcoRII requires three recognition sites.

We think that the observed differences in the reaction patterns of NaeI and EcoRII on a 2-site plasmid reflect the fundamental differences in their structural organization. The crystal structure of NaeI shows that its subunit interface has two clefts with the potential for binding DNA: the N-terminal cleft is similar to the active-site cleft of the orthodox restriction enzyme EcoRV, whereas the C-terminal cleft resembles the DNA binding site of a Type I topoisomerase. In the apo-structure of NaeI, the DNA binding and catalytic sites at the cleft between the N-terminal domains are not fully assembled. Binding of the single DNA copy at the cleft formed by the C-terminal domain dimer triggers conformational rearrangements that promote binding of the second copy at the N-terminal domain dimer resulting in the concerted DNA cleavage at both strands at a single site. The structure of the NaeI-DNA complex resembles to the schematically ‘2-site binding model’ which is depicted in Fig. 48B. Contrary, the crystal apo-structure and biochemical studies of EcoRII show a different architecture: EcoRII contains three putative DNA binding interfaces: one at the C-terminal domain dimer and two at each of the N-terminal domains.

The mode of EcoRII action is more like Type IIS REases rather than Type IIE enzyme NaeI. Type IIS enzyme FokI requires two copies of its recognition sequence and cleaves at only one recognition site similarly to Type IIE enzymes. Two FokI monomers bind two DNA copies through their N-terminal DNA recognition domains triggering dimerization of the C-terminal catalytic domains. The cleavage domain of FokI, however, is not involved in the specific DNA recognition, but cuts in its dimeric form both DNA strands at one cleavage site in one turnover. Metal-independent
Type IIIS restriction enzyme BfiI, which is a dimer already in the absence of DNA, has only one active site at the catalytic N-terminal domain dimer interface which is similar to unspecific Nuc, an EDTA-resistant nuclease from the phospholipase D superfamily. The C-terminal DNA recognition domains of BfiI, that are structurally similar to the N-terminal effector DNA binding domains of EcoRII, each bind a cognate DNA copy and then a single active site acts sequentially, first on the bottom and then the top strand to introduce a double strand break at one of the sites. Grafting determinants for sequence-specific DNA binding in the catalytic domains of FokI or BfiI would transform them into the Type IIE restriction enzymes mechanistically similar to EcoRII.

There are many restriction enzymes that become active only after interacting with several copies of the target sequence similarly to EcoRII or NaeI. For example, NarI is inactive at rare recognition sites GG↓CGCC but can be activated to cleave the resistant sites by adding a duplex with the recognition sequence. Structural organization of NarI is unknown, however, the NarI cleavage of plasmid DNA containing a single and two recognition sites has been reported recently. NarI cleaved plasmid DNA containing a single recognition site in just one strand converting the SC substrate to the OC form, like EcoRII, but more effectively. The SC plasmid with two sites was also converted first to OC DNA, cut in one strand at one or both sites and then the OC form was cleaved at the second strand to yield a FLL form cut at both strands at a single site. The FLL afterwards was very slowly converted to the final reaction product L2 cleaved at both sites, similarly to EcoRII.

_NarI reaction pattern on a plasmid with one and two recognitions sites implies that there could be more Type IIE enzymes following the same mechanism described here for EcoRII_. It will be interesting to see if NarI similarly to EcoRII requires three recognition sites for the concerted DNA cleavage of both phosphodiester bonds at a single site.
3. Direct visualization of the EcoRII-DNA triple synaptic complex by AFM

EcoRII kinetic reactions on the 1-, 2- and 3-site plasmids showed that simultaneous binding of three recognition sites is necessary for optimal activity of EcoRII. In order to visualize the EcoRII-DNA triple synaptic complex atomic force microscopy was applied. In order to test the three-site binding model of EcoRII, a DNA fragment containing three EcoRII recognition sites was employed. The formation of a triple synaptic complexes (TSC) mediated by the enzyme binding should lead to the formation of a two-loop molecule (Fig. 54).

**3.1. Construction of DNA fragments for EcoRII AFM studies**

Two different DNA fragment designs were examined in this study, the PCR1 fragment, which contains one recognition site, and the PCR3 fragment, which contains three recognition sites. The fragments were obtained by PCR amplification of selected regions from the pBluescript IIKS (+) plasmid (see section VI.17.1.). PCR primers were designed to amplify a defined region of interest to include the desired number of EcoRII restriction sites. The restriction sites are denoted A1, T, and A2 (Fig. 55). The letters indicate the central base A or T in the EcoRII recognition site, CCWGG, which is found on the leading strand. The PCR1 fragment is 205 bp long, with 5 bp recognition site A2 in the center and 100 bp flanks on either side (Fig. 55). The 810 bp PCR3 fragment was obtained by PCR to include all three restriction sites. The design of the PCR3 fragment consists of 100 bp arms, with segments between
each 5 bp recognition site, which are 312 and 283 bp long. A1 and A2 are two outer sites, with the T site between the two A sites (Fig. 55).

![Diagram](image)

**Figure 55. DNA fragments used in EcoRII AFM studies**. The measurements of linear fragments containing one-site (fragment PCR1) and three-site (fragment PCR3) of EcoRII recognition sites are indicated in base pairs and nanometers.

The distances between the sites A1 and T, as well as between A2 and T, were made approximately twice the DNA persistence length (312 and 283 bp, respectively) to ease potential problems with loop formation due to the relatively high rigidity of double-stranded DNA

3.2. Visualization and characterization of EcoRII-PCR3 complexes

The PCR3 fragment and wt EcoRII protein were mixed to form a complex as described in section VI.17.2. The Mg$^{2+}$ cations were replaced with Ca$^{2+}$ cations to prevent DNA cleavage and the complexes were cross-linked with glutaraldehyde.

![AFM images](image)

**Figure 56. AFM images of the EcoRII-DNA complexes**. (A) Large scan with various complexes. A zoomed image of the two-loop complex is shown as an inset. (B-E) Gallery of AFM images for two-loop DNA protein complexes, where the size of the bar (50 nm) applies to all images.
Fig. 56 shows an AFM image of the complex with a two-loop complex in the middle of the image. A zoomed image of this two-loop complex is shown in the inset. A collage of several two-loop complexes is shown in Fig. 56B-E. These are molecules of qualitatively similar morphologies: two loops of relatively similar sizes and short linear arms.

We measured the lengths of loops and of linear arms for the molecules to test whether the complexes are formed by specific binding of the enzyme to the cognate recognition sites on the DNA fragment. The results for the measurements of arms, loops, and entire contour lengths of the two-loop complexes are shown as histograms in Fig. 57A-C and in the Table 11.

![Histograms of DNA and protein parameters measurements](image)

**Figure 57. Histograms of protein and DNA parameters measurements in EcoRII-PCR3 two-loop structure complexes.** Histograms of measured DNA fragment length values: arm length (A), loop length (B), and contour length (C). The lengths correspond to expected values, which are 34 nm (A), 97.9 and 107.8 nm (B), and 278.8 nm (C). (D) Histogram for the EcoRII protein volume measurements. The expected size for the EcoRII dimer is 175.1 nm3.

Note that the histogram for the loop sizes contains the combined data for both loops, as their sizes were too similar to distinguish between them, so the data appear as a one-peak histogram. The expected sizes for corresponding distances for DNA contour length and the length between sites A1 and T, as well as between A2 and T (loops), and also the ends of the molecules, were calculated assuming a B-conformation of the DNA duplex. It is seen that the anticipated values (Fig. 55) are very close to the measured ones. Thus, *the AFM obtained data confirm that EcoRII enzyme is able specifically bind three DNA sites in the synaptic complex forming the two-loop structures*.
We also performed the measurements of the protein volume to evaluate the molecular mass of the protein and thus to determine the stoichiometry of the protein within the complex. The data for the volume measurements for two-loop complexes are shown as a histogram in Fig. 57D. The volume data were converted into the molecular mass and vice versa using the conversion coefficient as described in\textsuperscript{150}. From these conversions, the anticipated volume of an EcoRII dimer is 175.1 nm\textsuperscript{3}. This value is close to the mean value of 167 nm\textsuperscript{3} for the volume of the protein obtained from the histogram in Fig. 57D. Thus, the AFM data show that the two-loop complexes are formed by EcoRII and that dimerization of the protein is sufficient for that\textsuperscript{150}.

Figure 58. AFM data for different types of complexes of EcoRII with the PCR3 template\textsuperscript{150}. (A) Loop synthetic complexes. Inset images (i and ii) illustrate small-loop and large-loop synaptic complexes, respectively. (B) One-occupied-site complexes. (C) Two-occupied-site complexes. (D) Complexes with three-sites occupied by three separate EcoRII molecules. Representative AFM images for each type of complex are shown next to the corresponding histograms. Experimental protein volume measurements are indicated for each type of complexes.

In addition to two-loop complexes, the enzyme forms other types of complexes with the same DNA template. Additional EcoRII complexes with the PCR3 fragment were identified and divided into four major groups, as shown in the insets in Fig. 58. The first group comprises the synaptic
complexes, including the small loop and large loop structures. Insets in Fig. 58A (i) and (ii) show representative AFM images of these structures. The small loop can be formed by the protein binding at the central recognition site (T, Fig. 55) and one of the outer sites (A1 or A2, Fig. 55), and the big loop can be formed by the protein binding at two outer sites (A1 and A2, Fig. 55). Three other complexes (Fig. 58B-D) are observed on a linear DNA template with different numbers of protein blobs per molecule: one-site occupied fragments (B), two-site occupied fragments (C), and three-site occupied fragments (D).

These complexes were analyzed in a manner similar to that used for the two-loop complexes. The results of the analysis provided in the Table 11 confirm that looped complexes represent site-specific complexes.

Table 11. Measured length values for wt EcoRII-PCR3 complexes*

<table>
<thead>
<tr>
<th>Complex type</th>
<th>Contour length, nm</th>
<th>Arm length(s), nm</th>
<th>Loop length, nm</th>
<th>Segment length(s), nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>small loop (N** = 80)</td>
<td>279.2±0.5</td>
<td>33.0±0.3</td>
<td>135.6±0.4</td>
<td>106.7±0.3</td>
</tr>
<tr>
<td>big loop (N = 43)</td>
<td>279.2±1.1</td>
<td>32.6±0.3</td>
<td>207.0±1.2</td>
<td></td>
</tr>
<tr>
<td>double loop (N = 22)</td>
<td>276.1±0.5</td>
<td>33.0±0.2</td>
<td>101.5±0.5</td>
<td></td>
</tr>
<tr>
<td>1-site occupied (N = 52)</td>
<td>277.8±0.9</td>
<td>30.7±0.4</td>
<td>134.6±0.7</td>
<td></td>
</tr>
<tr>
<td>2-sites occupied (N = 36)</td>
<td>276.3±0.5</td>
<td>31.9±0.3</td>
<td>137.0±0.5</td>
<td>106.0±0.3</td>
</tr>
<tr>
<td>3-sites occupied (N = 18)</td>
<td>276.0±0.5</td>
<td>31.3±0.3</td>
<td></td>
<td>103.7±0.4</td>
</tr>
</tbody>
</table>

* All values corresponded well to the expected values. Multiple peaks on the histogram could not be distinguished for values that were close together, resulting in a number that fell between the two expected values. Standard errors of the mean from the Gaussian fit are indicated for each value. ** N indicates the number of complexes. Table was adapted from [50].

We also characterized the stoichiometry of the complexes by performing protein volume measurements. The volume measurement results for the synaptic looped complexes are shown in Fig. 58A. The volumes for all looped complexes are centered around the value of 180 nm$^3$ (or a molecular mass of 93.8 kDa) corresponding to the dimer state of the enzyme (theoretical M$\text{w}$...
90.5 kDa). However, when the EcoRII protein is found associated with one single site, the histogram for the protein volume shows a bimodal distribution (Fig. 58B-D), corresponding to the values for either a monomer or a dimer.

The yields for each type of EcoRII-PCR3 complex vary. The two-loop complexes were found to have a yield of 1.7%. The yield of the complex with a big loop is 2-times higher (3.2%), but the small loop complexes are the most abundant type of complex among all specific complexes (52.2%). Interactions with a single recognition site were also observed in substantial yields (33.5%). Other types of non-looped complexes (two-sites-occupied and three-sites-occupied) constituted 9.4%. The low yield of TSC complexes (1.7%) is not surprising as it is comparable with the low yield of the large one-loop complex (3.2%). A surprising finding is that the yield of small loop complexes occurs at a percentage of 52.2% and is the highest yield among all types of possible EcoRII-PCR3 complexes.

The probability of various sizes loop formation can contribute to the yield of synaptic complexes. Such probability is quantitatively characterized by the DNA cyclization J-factor. According the J-factor for the 300-600 bp the highest yield of looped complexes should be for the large loop with slightly lower yields for the small loops (data not shown). The data obtained for yields for large and small loops are just the opposite, suggesting that the probability of DNA cyclization does not explain the observed effect in the loop formation.

The difference in the sequences of the sites (Fig. 55) involved in the formation of large and small loops is another potential explanation for the observed non-even looping effect. Site orientation importance hypothesis is in line with the photo-cross-linking experiments, which demonstrate that the EcoRII N-terminal domain binding to the recognition site is asymmetric.

3.3. Characterization of EcoRII-PCR1 complexes

We analyzed complexes of EcoRII with a one-site DNA fragment PCR1. Both linear and synaptic complexes formed by two DNA molecules (trans X-
type complexes) were identified and characterized. The most abundant were the linear complexes with one protein bound to the DNA template (64.6%); the synaptic X-type complexes had a yield of 12.1% (data not shown). The length measurement for the shortest arm of the linear complex (30.1 nm) and the length measurement for all four arms of the synaptic complex (33.4 nm) corresponded to the expected values for specific binding (34 nm) (data not shown). The measured EcoRII volume revealed that synaptic complex volume could be approximated with one Gaussian with a maximum of 200.1 nm$^3$ (104.2 kDa) which is close to the expected value for a dimer (175.2 nm$^3$) (data not shown). The histogram of protein volume for single-site-occupied complexes linear complexes DNA fragment revealed two Gaussian maxima at 88.9 and 182 nm$^3$, corresponding to both monomeric and dimeric forms of the enzyme (data not shown). Control measurement of the free protein revealed that the protein appeared as spherical blobs with the mean volume value (185.6 ± 5.8 nm$^3$) corresponding to the protein dimer.

3.4. EcoRII monomer in the complexes

Additional insights into the EcoRII-DNA interactions come from the protein volume measurements of various complexes. The AFM data show that the dimeric form of the enzyme is required for the synaptic and three-site complex formation, but the monomeric form also provides stable one-site binding (Fig. 58). On the other hand, the protein dimers are found at individual binding sites as well. According to the gel filtration experiment (Fig. 47) and previously published data$^{115,117}$, EcoRII exists in solution as a dimer, so the appearance of the monomeric form in the complex could indicate protein-protein interactions that are weaker than protein-DNA interactions and reflect the dynamic nature of the EcoRII-DNA complex. Indeed, if the protein binds DNA as a dimer and then dissociates, one should expect to see some partially dissociated complexes as intermediates. Note that the intermediate states of the complex can be detected with AFM.
Fig. 59 shows a set of such intermediate complexes with a monomer bound to each individual site for different synaptic complexes: two-loop complexes (A), small-loop complexes (B and C), and big-loop complexes (D).

![Image of synaptic complexes](image)

**Figure 59. AFM images of partially dissociated synaptic complexes**. Complexes are shown for two-loop complex (A) small-loop complex (B, C) and large-loop complex (D). Protein monomeric and dimeric forms are indicated with M and D respectively.

Such synaptic complexes were not counted and considered to be not fully formed complexes. In the case of non-synaptic complexes (one, two, or three sites occupied by protein), it is impossible to distinguish between fully formed and partially dissociated complexes, and that could explain the co-existence of both monomer and dimer binding modes.

3.5. *Three-site synaptic complexes formed by other proteins*

The results obtained from AFM show that EcoRII is capable of bringing together three recognition sites to form a TSC, as well as forming typical synaptic complexes requiring the interaction of two DNA recognition sites. Importantly, the dimerization of the protein is sufficient for accomplishing all of these binding modes. These findings are fully in line with the “3-site model” proposed from EcoRII biochemical experiments. The visualization of the three-site complex by AFM is the direct justification of this model.

The protein-DNA synaptosomes are key intermediates for the various site-specific DNA interactions and the traditional model assumes the participation of two DNA sites in the formation of the complex. EcoRII AFM experiments revealed that three sites can be involved in the formation of a complex stabilized by one protein dimer. However, the interactions requiring three sites are not that very rare. Three sites can be brought together, but
typically, it requires the participation of at least two proteins. For example, a three-site synaptic complex is formed during bacteriophage Mu DNA transposition\textsuperscript{221,222}. In this insertion process, the transpositional enhancer is actively involved in direct interaction with the Mu left and right ends, leading to the formation of a three-site synaptic (LER) complex. A similar tripartite interaction including the involvement of two different site-specific DNA-binding proteins provides the Hin-mediated inversion reaction\textsuperscript{223-227}. In this system, the invertasome tripartite complex is formed by the Hin recombinase tetramers holding hixL and hixR sites and Fis protein dimer, which brings the enhancer to the complex. However, the three-site interaction can be formed transiently in the process of searching for specific sites via sliding of non-specifically formed synaptic complexes. This model was suggested by Cherepanov et al.\textsuperscript{228} on the basis of electron microscopy analysis of the integration of human immunodeficiency virus Type 1 (HIV-1) cDNA by the HIV-1 integrase. The loop sizes of the two-loop complexes varied over a broad range, leading the authors to hypothesize that two-loop or three-site synaptic complexes are transient states in the process of finding the specific site by protein sliding along DNA.

XI. Future prospects

The structural and biochemical studies carried out in this work revealed that Ecl18kI (\textsuperscript{↓}CCNGG) and EcoRII (\textsuperscript{↓}CCWGG) share similar active site architecture and DNA recognition elements with the tetrameric REases Bse634I (R\textsuperscript{↓}CCGGY), Cfr10I (R\textsuperscript{↑}CCGGY) and NgoMIV (G\textsuperscript{↓}CCGGC)\textsuperscript{80,83,160}. It was shown that these two REases flip the central base from their recognition sequence for the CC:GG half-sites recognition and for the adjusting the cleavage pattern\textsuperscript{80,147}. Therefore, the Ecl18kI/EcoRII and Bse634I/Cfr10I/NgoMIV enzymes can use a conserved DNA recognition module to recognize different sequences. Unlike Ecl18kI, which accepts both G:C and A:T base pairs at the center of its recognition sequence CCNGG,
EcoRII, particular the C-terminal domain EcoRII-C cleaves only target sequences with a central A:T pair, raising an intriguing question on the mechanisms of the central base pair discrimination in the CCNGG and CCWGG sites.

Base flipping by REases is a new mechanism employed for sequence-specific DNA recognition. 2-AP steady state fluorescence experiments carried out in this study paves the way for future stopped flow experiments to measure base flipping in a real time and for the future time resolved fluorescence studies to identify possible intermediate states on the base flipping pathway.

EcoRII studies carried out in this work provided the first example of REase that requires simultaneous binding of three recognition sites to achieve concerted DNA cleavage at a single site. AFM experiments confirmed the three site binding model for EcoRII and revealed that EcoRII-DNA interactions could be very dynamic. Further studies using single-molecule techniques, including time-lapse AFM, capable of detecting transient states of molecules and their complexes, could be an instrumental in unraveling the EcoRII site search mechanisms and the role of formation of the triple synaptic complex in particular.

Analytical ultracentrifugation of Ecl18kI indicates that the protein is a dimer in solution in the absence of DNA. The Ecl18kI-DNA crystal structure revealed in the asymmetric unit Ecl18kI tetramer bound with the two DNA molecules. Further studies are required to demonstrate if Ecl18kI tetramer observed in the crystal is functionally important in solution.
CONCLUSIONS

1. Our data demonstrate that Ecl18kI and EcoRII, specific for the CCNGG and CCWGGG sequences, respectively, share a conserved catalytic/metal binding site architecture and conserved structural mechanism of the C:G base pair recognition with restriction enzymes Bse634I/Cfr10I/NgoMIV specific for the RCCGGY/GCCGGC sequences.

2. We show that to achieve its specificity and cleavage pattern Ecl18kI flips both central nucleotides within the CCNGG sequence and buries the extruded bases in the pockets within the protein. Therefore, the Ecl18kI and Bse634I/Cfr10I/NgoMIV enzymes can use a conserved DNA recognition module to recognize different sequences, and form superimposable dimers, yet generate different cleavage patterns. Hence, Ecl18kI is the first restriction endonuclease that flips nucleotides to achieve specificity for its recognition site.

3. We demonstrate that Ecl18kI, EcoRII-C and PspGI binding enhance the fluorescence of 2-aminopurines placed at the centers of their recognition sequences suggesting that restriction endonucleases Ecl18kI, EcoRII-C and PspGI flip nucleotide(s) in solution.

4. We provide kinetic and biochemical evidence that Type IIE restriction endonuclease EcoRII requires simultaneous binding of three rather than two recognition sites to achieve concerted DNA cleavage at a single site. Atomic force microscopy allows direct visualization of synaptic EcoRII restriction enzyme-DNA complexes involving three DNA binding sites. Reaction mechanism suggested for EcoRII differs from that of another Type IIE enzyme NaeI.
I dedicate my thesis to my wife Giedrè.

I would like to express my special appreciation to my supervisor prof. dr. Virginijus Siksnys for his invaluable educational discussions and suggestions, sincere open communication, tolerance and for his enthusiasm in organization of various indoor and outdoor activities.

I am grateful to all my co-workers in the Laboratory of Protein-DNA Interactions for creating a good atmosphere in the lab. I thank dr. Giedrius Sasnauskas, dr. Mindaugas Zaremba and dr. Elena Manakova who contributed to the work presented in this dissertation. My special thanks to dr. Giedrius Sasnauskas who guided me in the kinetic studies and dr. Mindaugas Zaremba for their invaluable discussions and suggestions.

I am grateful to dr. Matthias Bochtler and Roman H. Szczepanowski (International Institute of Molecular and Cell Biology, Warsaw, Poland) for the crystal structure of Ecl18kI-DNA and to prof. dr. Yuri L. Lyubchenko and dr. Luda S. Shlyakhtenko (University of Nebraska Medical Center, USA) for the EcoRII AFM experiments.

I am grateful to prof. dr. Alfred Pingoud, dr. Wolfgang Wende (Justus-Liebig University, Giessen, Germany) and for allowing me the opportunity to carry out CD experiments and gain invaluable experience in their lab. I thank Dr. Sebastian Scholz (Justus-Liebig University, Giessen, Germany) for helping me with CD equipment. I thank prof. dr. Claus Urbanke (Medizinische Hochschule Hannover, Germany) for the analytical ultracentrifugation experiments.

I thank dr. Virginijus Lukša (Sicor Biotech UAB, Vilnius) and dr. Gražvydas Lukinavičius (Institute of Biotechnology) for their help with mass spectrometry analysis.

I thank Magdalena Kaus-Drobek and Monika Sokolowska (International Institute of Molecular and Cell Biology, Warsaw, Poland) for Mval cloning and protein purification.

I am grateful dr. Dalia Daujotytė (MRC Laboratory of Molecular Biology, Cambridge, Great Britain), prof. dr. Saulius Klimašauskas (Institute of Biotechnology, Vilnius), dr. Robert K. Neely, dr. Anita C. Jones and prof. David T.F. Dryden (University of Edinburgh, Great Britain) for their discussions on fluorescence experiments.

I would also like to thank to people who enabled detailed studies of Ecl18kI and EcoRII. I am grateful to dr. Alexander S. Solonin and dr. Marat M. Denjumukhametov (Institute of Biochemistry and Physiology of Microorganisms, Pushchino, Russia) for the cloning Ecl18kI restriction-
modification system. My thanks for dr. Monika Reuter (Institute of Virology, Berlin, Germany) for the clones of wt EcoRII and EcoRII-N, the Y41A, E271A and D299A EcoRII mutant plasmids. I also thank dr. Monika Reuter (Institute of Virology, Berlin, Germany) for her discussions and critical remarks. Thanks to New England Biolabs (Ipswich, USA) for PspGI clones.

I thank dr. Sonata Jurénaitē-Urbanavičienė (Institute of Biotechnology, Lithuania) and dr. Arvydas Lubys (Fermentas UAB, Vilnius) providing me the pKpn2I-ori plasmid.

I thank Fermentas UAB (Vilnius) for providing me pUC19 dam dcm− and pBluescript IIKS (+) plasmids, molecular biology tools and access to laboratory equipment.

I am grateful for the previous members of the lab dr. Remigijus Skirgaila and dr. Arūnas Lagunavičius for developing my experimental skills during my first years of experimental work. I am grateful to dr. Jolanta Vitkutė (Fermentas UAB, Vilnius) for her concern.

I thank assoc. prof. dr. Jurgis Kadiauskas (University of Vilnius, Department of Biochemistry and Biophysics) encouraging me to start my studies in biochemistry.

Many thanks for my secondary school teachers: my teacher of chemistry Janina Greblikienė, my teacher of physics Petras Jonušas, for their charismatic introduction in to the weird world of the laboratory and also to my teacher of mathematics Stasys Pikelis.

Many thanks for my uncle prof. dr. Sigitas Tamulevičius and aunt dr. Eglė Tamulevičienė for inspiring me to embark on a scientific career, for their constant interest in my work and their concern.

A lot of thanks to my friends for their honest and open communication. My thanks to my parents and parents-in-law for their concern and understanding. Particular thanks to my wife Giedrė for immense support both at work and at home.

The work presented in this thesis has been carried out at the Institute of Biotechnology, Vilnius from 2000 to 2008. The financial support from Howard Hughes Medical Institute, BIOCEL, International Quality Network and the Lithuanian State Science and Studies Foundation is gratefully acknowledged.
The thesis is based on the following original publications:


The material of the thesis was presented in these conferences:


Other publications:

REFERENCES


