VILNIUS UNIVERSITY

Dmitrij Golovenko

STRUCTURAL AND FUNCTIONAL STUDIES OF RESTRICTION ENDONUCLEASES EcoRII, BfiI AND Bse634I

Doctoral dissertation
Physical science, biochemistry (04 P)

Vilnius, 2012

The work presented in this doctoral dissertation has been carried out at the Institute of Biotechnology, Vilnius University during 2004-2012.

Scientific consultants:

Dr. **Saulius Gražulis** (Vilnius University, physical sciences, chemistry - 03 P)

Prof. dr. **Virginijus Šikšnys** (Vilnius University, physical sciences, biochemistry - 04 P)

VILNIAUS UNIVERSITETAS

Dmitrij Golovenko

RESTRIKCIJOS ENDONUKLEAZIŲ EcoRII, BfiI IR Bse634I STRUKTŪRINIAI IR FUNKCINIAI TYRIMAI

Daktaro disertacija
Fiziniai mokslai, biochemija (04 P)

Vilnius 2012

Disertacinis darbas atliktas 2004-2012 metais Vilniaus Universiteto Biotechnologijos institute.

Moksliniai konsultantai:

Dr. **Saulius Gražulis**
(Vilniaus Universitetas, fiziniai mokslai, chemija - 03 P)

Prof. dr. **Virginijus Šikšnys**
(Vilniaus Universitetas, fiziniai mokslai, biochemija - 04 P)

CONTENTS

LIST OF ABBREVIATIONS.....7
INTRODUCTION.....9
1. LITERATURE OVERVIEW.....15
Restriction endonucleases 15
1.1. REases are the part of bacterial R-M systems.....15
1.2. Nomenclature of the Type II restriction enzymes.....15
1.3. Structural classification of the Type II restriction enzymes.....16
1.3.1. Nuclease domain PD-(D/E)XK.....17
1.3.2. Nuclease domain ββα-Me.....19
1.3.3. Nuclease domain GYI-IYG.....20
1.3.4. Phospholipase D nuclease domain.....21
1.3.5. “Half-pipe” nuclease domain.....22
1.4 Multidomain Type II restriction endonucleases.....24
1.4.1. Duplicated PD-(D/E)XK domains: Sau3AI.....25
1.4.2. Fused HTH and PD-(D/E)XK domains: FokI, SdaI, NaeI.....26
1.4.3. Fused B3-like and PD-(D/E)XK domains: EcoRII.....31
1.4.4. Fused B3-like and PLD nuclease domains: BfiI.....32
1.4.5. Fused MTase and PD-(D/E)XK domains: BpuSI.....34
1.4.6. Fused metal binding and PD-(D/E)XK domains: NotI.....35
1.5. Family of REases recognizing common DNA motif CCGG.....36
Plants DBDs interacting with DNA through β-sheet.....38
2.1. AP2 and B3 DNA binding domains.....39
2.1.1. Structural analysis of B3 domain and model of DNA binding.....42
2.1.2. Nucleic acid-binding B3 fold in context of cradle-loop metafold..45
3. MATERIALS AND METHODS.....47
3.1. Enzymes, sorbents and reagents.....47
3.2. E. coli Strains.....47
3.3. Plasmids.....47
3.4. Oligoduplexes.....48
3.5. Buffers.....50
3.5.1. Buffers related to EcoRII-N.....50
3.5.2. Buffers related to EcoRII-C.....51
3.5.3. Buffers related to BfiI, BfiI-C and BfiI-N-K107A.....51
3.5.4. Buffers related to Bse634I.....52
3.6. Expression and purification of EcoRII-Y41A and EcoRII-N.....52
3.7. Preparation and crystallization of EcoRII-N–DNA complex.....52
3.8. Preparation and crystallization of EcoRII-C–DNA complex.....53
3.9. Data collection from EcoRII-N–DNA and EcoRII-C–DNA crystals....54
3.10. EcoRII-N–DNA and EcoRII-C–DNA structure determination.....55
3.11. Purification and crystallization of BfiI domains with DNA.....56
3.12. Data collection and structure determination of BfiI domains.....57
3.13. Purification and crystallization of Bse634I with DNA.....58
3.14. Data collection and structure solution of Bse634I-DNA.....59

3.15. DNA-protein binding assay....60
3.16. DNA cleavage assay....61
3.17. Software used in research....62
3.17.1. Protein sequence analysis....62
3.17.2. Crystallographic and structural data analysis....62
4. RESULTS AND DISCUSSION....63
4.1. Analysis of EcoRII-C domain interactions with DNA....63
4.1.1. DNA oligoduplex is bound in two overlapping orientations....63
4.1.2. Overall structure of the EcoRII-C complex with DNA....64
4.1.3. EcoRII-C flips out the central T:A base pair within the CC(T/A)GG site....65
4.1.4. Putative mechanism of T:A recognition by EcoRII-C....66
4.1.5. EcoRII-C interactions with the symmetrical CC:GG half-sites....68
4.1.6. Active site of EcoRII-C....69
4.1.7. EcoRII-C sensitivity to DNA methylation....70
4.2. Analysis of EcoRII-N domain interactions with DNA....72
4.2.1. Overall structure of the EcoRII-N–DNA complex....72
4.2.2. Sequence recognition by EcoRII-N....73
4.2.3. Probing specificity determinants of the central T:A pair....75
4.2.4. Probing sensitivity to methylation and indirect recognition....76
4.2.5. Similarities of EcoRII-N to the TF factors of B3 family....78
4.3 Mechanism of the EcoRII action....80
4.3.1. Structural transitions occurring upon DNA binding to the N-terminal domains of EcoRII....80
4.3.2. Structural transitions occurring upon DNA binding to the C-terminal domains of EcoRII....82
4.4. EcoRII family proteins....84
4.5. Analysis of BfiI domain interactions with DNA....87
4.5.1. Asymmetric units of BfiI-C–DNA and BfiI-N-K107A crystals....87
4.5.2. DNA recognition by the BfiI-C domain....88
4.5.3. Non-specific contacts in the BfiI-C–DNA complex....91
4.5.4. B3-like domains share conservative DNA interaction mechanism 92
4.6. A model of BfiI activation....92
4.7. BfiI family proteins....94
4.8. Analysis of Bse634I interactions with DNA....96
4.8.1 Crystallization and crystals of Bse634I-DNA complexes....96
4.8.2 Overall architecture of Bse634I-DNA complexes 98
4.8.3 Bse634I active center organization....99
4.8.4 Bse634I contacts to the conserved CCGG core....100
4.8.5 Bse634I interactions with an outer R1:Y6 base pair....102
4.8.6 Bse634I binding and cleavage studies....104
CONCLUSIONS....108
LIST OF PUBLICATIONS....109
ACKNOWLEDGEMENT....110
REFERENCES....112

LIST OF ABBREVIATIONS

Specifying DNA binding site “R” stands for purine, “W” – for A or T, “Y” – for pyrimidine, “N” – for one from A, G, T or C.

5mC	5-methylcytosine
AFM	atomic force microscopy
bp	base pair
BSA	bovine serum albumin
B3-like	DNA binding domain of pseudobarrel fold
CPK	CPK (Corey-Pauling-Koltun) is a color scheme for atoms (C=gray; O=red; N=blue; S=yellow; P=orange)
DBD	DNA binding domain
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
^{5m}iC	5-methyl isocytosine
iG	isoguanine
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth
MTase	DNA methyltransferase
NCS	non crystallographic symmetry
PAGE	polyacrylamide gel electrophoresis
PDB ID	protein data bank identification code
PEG	polyethyleneglycol
PLD	phospholipase D
PMSF	phenylmethylsulfonyl fluoride
REase	restriction endonuclease
R-M	restriction-modification
RMSD	root-mean-square deviation
SDS	sodium dodecyl sulfate
TEMED	Tetramethylethylenediamine
TF	transcription factor
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
wt	wild type

INTRODUCTION

Restriction endonucleases (REases) are present in both bacterial and archaeal kingdoms, as a part of restriction-modification systems [1]. *In vivo* these systems defend prokaryotic cells against invading foreign (e.g., bacteriophage) DNA. Type II REases recognize short DNA sequences, usually 4-8 bp in length, and cleave both DNA chains within or close to their recognition sites generating sticky or blunt ends [2]. Nearly 4000 different Type II restriction endonucleases are characterized and totally cover ~300 unique specificities [3]. In this respect, REases constitute one of the largest families of functionally related proteins known to date. Crystallographic and bioinformatics studies indicate that REases can be subdivided into five evolutionarily distinct folds depending on the nuclease core involved in the phosphodiester bond cleavage. Type II REases are grouped into i) PD-(D/E)XK superfamily, ii) phospholipase D (PLD) superfamily, iii) “half-pipe” superfamily, iv) GIY-YIG superfamily and v) ββα-Me (HNH) superfamily [4]. Delineation of the repertoire of protein folds, providing three-dimensional portraits of the REases, should reveal how different folds are tailored to function as restriction enzymes and enable homology modeling of other family members using template structures.

Most of our knowledge on the structure and function of REases comes from the PD-(D/E)XK superfamily. Indeed, 34 crystal structures of PD-(D/E)XK superfamily REases are currently present in the Protein Data Bank [5]. Structure comparison of the family members reveals a common structural core, which serves as a scaffold for the active site comprised of two or three acidic residues (aspartate or glutamate) and one lysine residue. The conserved structural core is decorated by different structural elements, which often involve various DBD and often determine DNA sequence specificity [4]. The structural data for other REase families are scarce. For example, crystal structures of PLD superfamily enzymes in the DNA bound form were unknown when these PhD studies were initiated.

The major goal of this work was to explore the specificity-structure relationships within PD-(D/E)XK and phospholipase D superfamily enzymes using a combination of the crystallographic and biochemical methods. More specifically, we have focused on the Bse634I and EcoRII restriction enzymes belonging to the Cfr10I/NgoMIV/Bse634I branch of PD-(D/E)XK superfamily, and BfiI REase of PLD superfamily.

Cfr10I/NgoMIV/Bse634I subfamily of the PD-(D/E)XK REases. Restriction enzymes belonging to this subfamily recognize related nucleotide sequences containing conserved CCGG or CCNGG core that is flanked by different nucleotides, and cut phosphodiester bond in the conserved position before the first cytosine of the conserved nucleotide core (Table 1).

Table 1. Evolutionary conserved subfamily of Cfr10I/NgoMIV/Bse634I REases

Enzyme	Subtype	Recognition site	Oligomeric state	Functional requirements	References
HpyF100III	IIP	↓CCGG	n.d.*		[6]
AgeI	IIP	A↓CCGGT	n.d.		[7]
Kpn2I	IIP	T↓CCGGA	n.d.		[8]
NgoMIV	IIF	G↓CCGGC	tetramer	two cleavage sites	[9]
BsaWI	IIP	W↓CCGGW	n.d.		[3]
Cfr10I /Bse634I	IIF	R↓CCGGY	tetramer	two cleavage sites	[10][11]
SgrAI	IIF	CR↓CCGGYG	transient tetramer	two cleavage sites	[12]
PspGI	IIP	↓CCWGG	dimer		[13]
EcoRII	IIE	↓CCWGG	dimer	two allosteric sites	[14] [15]
Ecl18kI	IIF	↓CCNGG	transient tetramer	two cleavage sites	[16][6]
PfoI	IIP	T↓CCNGGA	n.d.		[17]

↓ indicates cleavage position; * not determined.

The Cfr10I/NgoMIV/Bse634I subfamily currently includes 12 enzymes specific for the 5'-CCGG, 5'-CCNGG, 5'-CCWGG, 5'-NCCGGN, 5'-TCCNGGA, 5'-CRCCGGYG sequences spanning from 4 nt to 8 nt.

Collectively, the enzymes belonging to this family represent ten different specificities (Table 1). The family members presumably have a conserved structural core of the monomer, however, they show variability in sequence specificity and quaternary structure. Structural studies of enzymes belonging to this family should allow us to understand how different specificities emerge within a common structural core. In parallel, understanding the oligomeric structure-function relationships within the family of the homologous restriction enzymes can provide clues to the biological and functional significance of different quaternary associations.

Tetrameric Type IIF restriction endonuclease Bse634I is one of the founding members of the family. It recognizes a degenerate 5'-RCCGGY sequence (where R stands for G or A, and Y stands for C or T). Its crystal structure in the DNA free form has been solved in 2002 [11], however the lack of the structure in the DNA-bound form hindered our understanding how recognition of the degenerate DNA sequence is achieved by Bse634I.

The EcoRII REase recognizes a pentanucleotide sequence 5'-CCWGG (W stands for A or T) where the perfect target site symmetry is broken in the center. It belongs to the Type IIE REases which require an additional DNA site for the cleavage to occur at the first site. Thus, one site is the actual target of cleavage, the other is an allosteric effector [18]. While the interaction with two DNA sites is a prerequisite for EcoRII cleavage, simultaneous binding of three recognition sites is necessary for a concerted cleavage of both DNA strands at a single site [15]. The crystal structure of the EcoRII solved in the DNA-free form shows a homodimer comprised of two monomers [14]. EcoRII monomer consists of two structurally distinct domains which are able to bind the 5'-CCWGG sequence [15] raising a question how two structurally different domains recognize the same target sequence. Furthermore, it was unclear how EcoRII interacts with the interrupted conserved 5'-CCWGG core (underlined) of the Cfr10I/NgoMIV/Bse634I subfamily (Table 1).

PLD nuclease superfamily. BfiI restriction enzyme recognizes the nucleotide sequence 5'-ACTGGG [19] and is unique in cleaving DNA in the absence of metal ions [20]. The crystal structure [21] and biochemical data [22] revealed that BfiI represents a different evolutionary lineage of restriction enzymes and belongs to the phospholipase D superfamily. The protein consists of two structural domains. The N-terminal catalytic domain is similar to Nuc [23], an EDTA-resistant nuclease from the phospholipase D superfamily. The C-terminal DNA-binding domain of BfiI exhibits a β-barrel-like structure very similar to the effector DNA-binding domain of the Mg^{2+}-dependent restriction enzyme EcoRII [14] and to the B3 DNA-binding domain of plant transcription factors [24]. BfiI presumably evolved through a domain fusion of a DNA-recognition element to a nonspecific nuclease akin to Nuc and elaborated a mechanism to limit DNA cleavage to a single double-strand break near the specific recognition sequence [25]. However, it remained to be determined how BfiI recognizes its target site and couples target recognition to catalysis. Furthermore, structural similarities between the DNA-binding domain of EcoRII and the C-terminal domain of BfiI raised a question how these domains interact with different target sites.

The specific aims of this study were:

1. To determine the mechanism of the sequence recognition by Bse634I using a combination of crystallographic and biochemical methods.

2. To determine the mechanism of the sequence recognition by EcoRII restriction enzyme using a combination of crystallographic and biochemical methods.

3. To determine the mechanism of the sequence recognition by BfiI using a combination of crystallographic and biochemical methods.

Scientific novelty.

We have solved the crystal structure of the Bse634I REase mutant (R226A) complexed with two alternative target sites 5'-ACCGGT and 5'-GCCGGC, respectively. The analysis of the crystal structures revealed for the first time that the degenerate base pairs recognition by Bse634I is achieved through the combination of direct and indirect readout mechanisms. The crystal structures of the N- and C-terminal domains of EcoRII solved in the DNA bound form revealed different structural mechanisms used for the recognition of the same target sequence 5'-CCWGG. We provided the first structural evidence that the C-terminal domain of EcoRII flips out the central nucleotide while interacting with its target site, enabling the EcoRII C-terminal domain to use conserved structural elements for the recognition of the CCGG core. Furthermore, the crystal structure of the EcoRII-N provided the first glimpse into the B3 family domain in the DNA bound form. Finally, the crystal structure of the C-terminal DNA binding domain of BfiI bound to the target site enabled the first structural comparison of two B3-family domains (EcoRII-N and BfiI-C) in the DNA-bound form.

Practical value of the research.

Due to their unique specificity, restriction endonucleases have gained widespread application as indispensable tools for *in vitro* manipulation and cloning of DNA. Delineation of the repertoire of the protein folds, providing three-dimensional portraits of the REases, should reveal how different folds are tailored to function as restriction enzymes and provide novel templates for homology modeling of REases.

Highly specific programmable DNA endonucleases (meganucleases) are desirable for the gene replacement technology and genome editing, prerequisites for gene therapy. Determination of crystal structures and deciphering molecular mechanisms of DNA recognition may contribute to the

development of meganucleases for genome manipulation. Indeed, determination of the crystal structure of the R226A mutant of Bse634I contributed to the engineering of the Bse634I-triple helix forming oligonucleotide conjugate [26]. Moreover, the EcoRII-N and BfiI-C structures in the DNA bound form provide templates for modeling B3 superfamily transcription factors and complexes which may contribute to plant engineering [27].

The major findings presented for defense in this thesis:

- The Bse634I restriction enzyme mutant R226A bound to the two symmetric variants of 5'-RCCGGY site employs conserved structural elements for the recognition of the CCGG core and uses both direct and indirect readout for discrimination of the degenerated R:Y base pair.
- The catalytic C-terminal domain of EcoRII flips out the central nucleotide while interacting with its 5'-CCWGG target site and employs conserved structural elements for the recognition of the CCGG core.
- The N-terminal allosteric domain of EcoRII employs a distinct mechanism for the recognition of the 5'-CCWGG site.
- The structural mechanism of the sequence recognition employed by the EcoRII-N and BfiI-C domains is shared by plant transcription factors of the B3 family.

1. LITERATURE OVERVIEW

Restriction endonucleases

1.1. REases are the part of bacterial R-M systems

The main biological function of a REase is to protect the host cell from bacteriophages by cleaving invading foreign DNA [28]. The host's own DNA is protected from the REase cleavage by an accompanying DNA methyltransferase which introduces the methyl group at the particular position within the recognition site and renders DNA resistant to REase cleavage [2]. Hemimethylated DNA is usually cut by Type II endonucleases extremely slowly [29], providing time for cognate DNA methyltransferase to modify DNA strand that is newly synthesized upon replication and therefore to protect the hemimetylated chromosomes [30]. R-M systems themselves are thought to be rapidly evolving genes because of their tendency for the frequent horizontal gene transfer [31]. They can be carried on mobile elements such as plasmids, phages, conjugative elements, genomic islands, transposons and integrones [32] providing stable maintenance through the mechanism of the postsegregational killing [33].

1.2. Nomenclature of the Type II restriction enzymes

Based on cofactor requirements and subunit composition REases are classified into four types [34]. Vast majority of biochemically and genetically characterized REases belongs to the Type II [3]. Type II REases require only Mg^{2+} ions as a cofactor and cleave DNA generating 3'-hydroxyl and 5'-phosphate termini, which can be either 'blunt' or 'sticky' with 3'- or 5'-overhangs. Type II REases are further subdivided into the 11 subtypes in respect to their structural organization, recognition sequence and number of copies required, etc. [34] (Table 2). A typical Type IIP REase is a symmetric homodimer which recognizes palindromic sequence and cleaves it at the fixed position within or near of the recognition sequence.

Table 2. Nomenclature of Type II restriction enzymes (based on [34])

	Defining feature	Example enzyme, recognition and cleavage site
A	Asymmetric recognition sequence	AciI, CCGC(-3/-1)
B	Target sequence is cleaved out	BcgI, (10/12)CGANNNNNNTGC(12/10)
C	R and M functions in one polypeptide	GsuI, CTGGAG(16/14)
E	Two or more targets; one cleaved, one an effector	EcoRII, ↓CCWGG
F	Two targets, both cleaved coordinately	SfiI, GGCCNNNN↓NGGCC
G	Cleavage affected by S-adenosyl-L-methionine	Eco57I, CTGAAG(16/14)
H	Similar to Type I gene structure	BcgI, (10/12)CGANNNNNNTGC(12/10)
M	Methylated target	DpnI, Gm6A↓TC
P	Symmetric target and cleavage sites	EcoRI, G↓AATTC
S	Asymmetric target and cleavage sites	BfiI, ACTGGG(5/4)
T	Heterodimeric organization	BslI, CCNNNNN↓NNGG

Enzymes of Type IIE subtype, like EcoRII and NaeI, bind two or more recognition sites, whereas only one of them is cleaved, while the other sites serve as allosteric effectors. Enzymes belonging to the Type IIF subtype, exemplified by NgoMIV and Bse634I, are homotetramers that bind and cleave two copies of their recognition site simultaneously. Enzymes of the Type IIG and the Type IIC have both cleavage and modification domains within a single polypeptide but differ by the requirement of S-adenosyl-L-methionine as an additional cofactor. Enzymes of the Type IIS subtype, for instance FokI and BfiI, cleave DNA outside of the asymmetric recognition sequence.

1.3. Structural classification of the Type II restriction enzymes

Initial attempts to classify Type II REases were based on the comparison of their recognition sequences and their modes of action. Lack of detectable sequence similarities between REases, as well as between REases and other protein families for a long time hindered their bioinformatics analysis [35]. Typically, the sequence similarity between REases (except isoschizomers) is so low that most of the relationships between known members of the family were identified only after the crystal structures were solved. The first crystal structures of the REases EcoRI [36] and EcoRV [37] revealed the conserved

catalytic core which was dubbed the PD-(D/E)XK fold due to the conserved active site residues proline, aspartate, glutamate and lysine. This fold was detected in all REase structures solved before 2005, until the structure of REase BfiI revealed a phospholipase D fold [21]. Later the crystal structure of PabI revealed a different 'half-pipe' fold [38] and REase Hpy99I was confirmed to belong to ββα-Me nuclease superfamily [39] establishing a link to homing nucleases. Recently the structure of Eco29kI revealed that it belongs to the GIY-YIG nucleases superfamily [40].

Currently there are 38 crystal structures of full length Type II REases deposited in PDB [5]. REases are classified according to their nuclease domains: 34 of them have catalytic domains of PD-(D/E)XK fold, ββα-Me and GIY-YIG folds are represented by two structures each and single structures of PLD and "half-pipe" folds are available (Table 3).

Table 3. Distribution of REases exhibiting different folds (based on [4])

Nuclease domain	Structures (complexes)	Annotated subfamilies	Fusions to other DBDs	References
PD-(D/E)XK	34(28)	98	HTH, PD-(D/E)XK, B3-like	[41]
ββα-Me	2(2)	14	HTH, ββα-Me	[39]
GIY-YIG	2(2)	2	-	[42]
PLD	1(0)	2	B3-like	[21]
half-pipe	1(0)	1	-	[38]

1.3.1. Nuclease domain PD-(D/E)XK

The conserved structural core of the PD-(D/E)XK nuclease superfamily was initially postulated after analyzing structures of the REases EcoRI and EcoRV [43]. The most recent definition of the consensus structural core is a four-stranded mixed β-sheet flanked by an α-helix on each side ordered as αββββαβ [44] (Figure 1A). The catalytic motif that gave name to the fold consists of the invariant aspartate followed by a proline, an a.a. sequence of varying length, an aspartate or glutamate residue, a hydrophobic residue denoted as X and a lysine residue [2]. While most REases share a conserved $PDX_{(10\text{-}20)}$(D/E)XK motif, a certain plasticity within motif is tolerated: for

example, a proline residue is missing in BsoBI [45] or MspJI [46], the lysine moiety appears to be replaced by the hydrophobic residue valine in SdaI [47], or by glutamine in BglII [48] or even by negatively charged glutamate in BamHI [49]. Alternatively, the structure of the Cfr10I REase showed that the second carboxylate residue (underlined) can come from a different part of the sequence $P^{133}DX_{55}KX_{13}\underline{E}$ [10]. Importance of the spatial rather than sequence conservation of the active center residues was proved by the 'swap' mutation S188E/E204S that has transformed the active site of Cfr10I to the canonical motif $P^{133}DX_{53}\underline{E}VK$ [50].

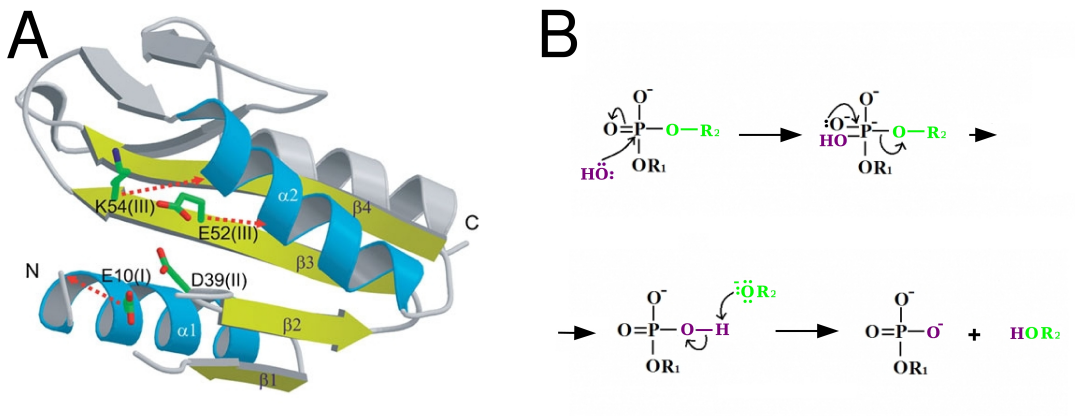


Figure 1. (A) Conserved structural core of PD-(D/E)XK nucleases (illustrated by the archaeal Holliday junction endonuclease [PDB ID 1OB8] containing the typical example of the motif) comprised of a mixed β-sheet (yellow) and two α-helices (blue). Positions of the active site residues (sticks) are depicted. Red arrows indicate possible 'migration' routes of the active site residue into alternative positions. The figure was adapted from [44]. (B) Schematic representation of a phosphodiester bond hydrolysis by the S_N2-type mechanism.

DNA phosphodiester bond is hydrolyzed by PD-(D/E)XK superfamily nucleases by the S_N2-type mechanism [51]. The conserved active site provides the key amino acid residues: a base for deprotonation of the water molecule attacking scissile phosphodiester bond; a Lewis acid that stabilizes the negatively charged pentavalent transition state; and an acid that protonates the leaving 3'-hydroxyl group (Figure 1B). Cleavage is dependent on divalent metal ions as a cofactor and optimal cleavage is observed with Mg^{2+}, which in some cases can be replaced Mn^{2+} and Co^{2+}. The metal ion stabilizes the pentavalent transition state phosphate by coordinating the two oxygen atoms of

the phosphate group; the metal is also supposed to participate in the activation of the attacking water molecule [51].

1.3.2. Nuclease domain ββα-Me

All ββα-Me superfamily enzymes have the structural motif that consists of two antiparallel β-strands (ββ) followed by an α-helix (α) with the catalytic metal ion (Me) sandwiched between them [52] (Figure 2A). The divalent metal ion at the active site is coordinated by one (e.g. in I-PpoI [53]), two (e.g. in T4 endonuclease VII [54]) or three (e.g. in colicin E7 [55]) amino acid ligands. The most commonly encountered metal ion is Mg^{2+}, although it can be substituted by Mn^{2+}, Co^{2+}, Cu^{2+} and Ni^{2+} [52]. Enzyme protomer usually has one or two ββα-Me repeats (e.g. Hpy99I), but only one β-hairpin harbors the active site histidine [39]. The superfamily is alternatively called "HNH nucleases" pointing out to conservative nonconsecutive active site residues [52]. HNH motif is approx. 34 amino acids long and contains invariant histidine and asparagine residues each located at the corresponding secondary structure of ββα-Me (Figure 2A). HNN variations of the HNH motif are also known [52].

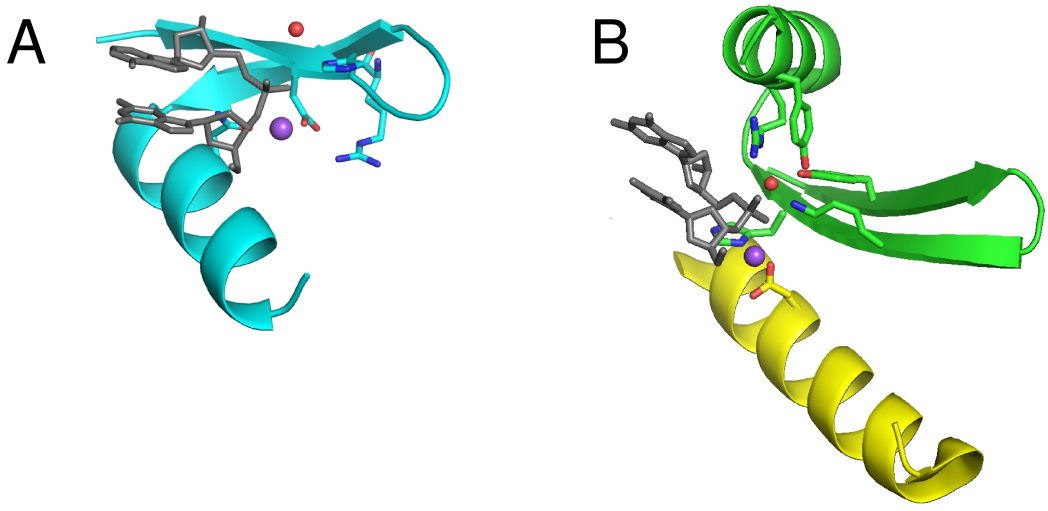


Figure 2. Secondary structure elements of ββα-Me (A) and GIY-YIG (B) REase active sites superfamilies. Figures are adapted from [42] using PDB IDs 3FC3 and 3OQC for substrate bound REases Hpy99I and Hpy188I, correspondingly. Metal ions (magenta) and water molecules (red) that attack scissile phosphate are shown as spheres, active site residues are shown as sticks.

Analysis of DNA bound structures of homing endonucleases and colicins [56] allowed to propose that a general cleavage mechanism for HNH nucleases is of S_N2-type (Figure 1B). Crystallographic snapshots of the active site of the DNA bound REase Hpy99I [42] confirms that the structurally conservative histidine residue from the end of the first strand of ββα motif acts as a general base that activates a water molecule for the nucleophilic attack of the scissile phosphodiester bond (Figure 2A), while the conservative asparagine (H<u>N</u>H) coordinates divalent metal ion that stabilizes the phosphoanion transition state. Meanwhile the structure of the DNA-bound REase PacI displays yet another variation of the HNH active site where histidine (<u>H</u>NH) has been replaced by arginine interacting with the 3'-hydroxyl leaving group [57].

1.3.3. Nuclease domain GYI-IYG

Conservative core of GYI-IYG superfamily has α/β-fold where three-stranded antiparallel β-sheet is flanked by two or three α-helices [58]. The canonical domain has ββαβα topology where the first two β-strands harbor the conserved GIY and YIG motifs, respectively, followed by conserved Arg and Glu residues. A minimal core region can be as short as 70 residues [59]. Phylogenetic analysis showed that both motifs are variable with only a few a.a. being ultimately conserved. The first crystal structure of a GYI-IYG nuclease in complex of DNA was that of the REase Hpy188I [42]. The structure revealed unexpected variant of structural motif with one of the conserved helices swapped between two domains (swapped helix shown yellow in Figure 2B) and showed modified variant of active site GYI-IYG motif, e.g. LVY-KIG (green in Figure 2B), as well as conserved Arg and Glu residues participate in the active site formation (Figure 2B). It was proposed on the basis of crystallographic studies of Hpy188I-DNA and Eco29kI-DNA complexes [40] that GIY-YIG nucleases use S_N2-type mechanism (Figure 1B) to catalyze the direct attack of a water molecule on the scissile phosphate atom in a single step nucleophilic substitution with the simultaneous stabilization of the phosphoanion transition state and the 3'-hydroxyl leaving group by a

divalent metal ion. According to the proposed mechanism the GIY invariant tyrosine serves as a general base for the attacking water molecule and the YIG tyrosine (replaced by a lysine in Hpy188I) is required as a hydrogen bond donor to the scissile phosphate oxygen atom [42]. The conserved YIG glycine makes space for the GIY tyrosine and the 5'-phosphate in the protein-product complex. The invariance of the Arg residue located on the α-helix is explained by its interactions with the attacking water molecule. The Glu residue located on the swapped α-helix is needed to anchor the active site Ca^{2+} ion [42] (Figure 2B).

Along with the adaptation for the recognition of the target site by restriction enzymes, GIY-YIG scaffold has encountered also in homing endonucleases where its function is to endure the sequence specific cleavage [60] or to enable structure-specific cleavage by the DNA repair enzymes such as UvrC [61].

1.3.4. Phospholipase D nuclease domain

The phospholipase D (PLD) superfamily is defined by the catalytic core of an α/β class that was observed for the first time in the secreted nonspecific endonuclease Nuc [23]. The crystal structure of the endonuclease Nuc revealed a dimer with 2400 $Å^2$ buried at the dimer interface. Each subunit is made of an eight-stranded mixed β-sheet flanked by α-helices from each side (Figure 3A). The Nuc dimer cleaves single and double stranded DNA [62] which is accommodated in a wide binding cleft [23]. The superfamily consensus sequence motif $HXKX_4DX_6GSXN$ proposed earlier by the bioinformatic analysis [63] in the Nuc monomer is located on loops between the secondary structure elements β4-β5 and β6-α4 [23]. Thus, the two motifs in the dimer, related by 2-fold symmetry, come together in the middle of the binding cleft to constitute a single active site. In the crystal structure of Nuc the competitive inhibitor tungstate is bound in the active center (Figure 3B).

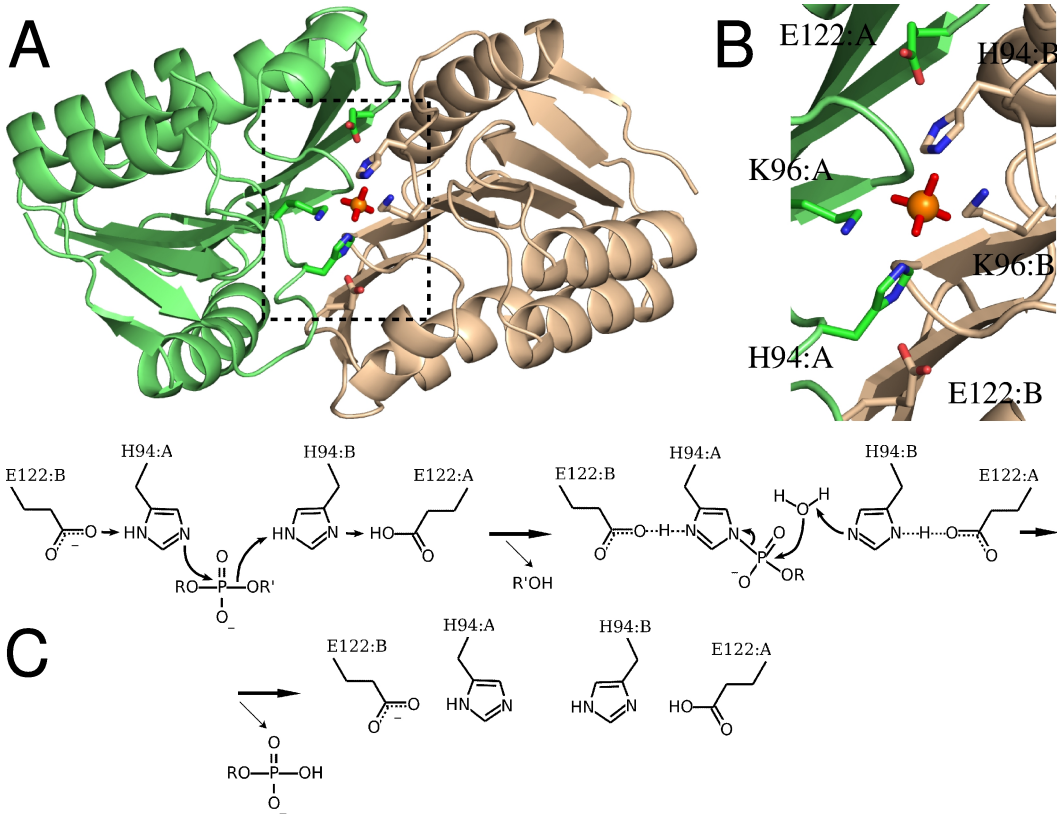


Figure 3. Phospholipase D nuclease domain represented by (A) dimeric (green and wheat) nonspecific nuclease Nuc [PDB ID 1BYS] and (B) close view of the active site in the presence of competitive inhibitor – tungstate ion WO_4^{2-}. (C) A proposed DNA cleavage mechanism by PLD nucleases.

During the first step of the cleavage reaction the conserved histidine from one monomer acts as the nucleophile attacking the scissile phosphate to generate a covalent phosphohistidine intermediate; its negative charge is stabilized by two lysines (Figure 3C). The histidine from the second monomer acts as a general acid, protonating the leaving 3'-hydroxyl group (Figure 3C). Biochemical experiments, computational methods and mutational studies show that this two-step phosphodiester hydrolysis scheme is also valid for other PLD nucleases, including the dimeric REase BfiI [64] and a pseudodimeric lysosomal endodeoxyribonuclease DNase IIα [65].

1.3.5. “Half-pipe” nuclease domain

Genome comparison studies and a search of putative restriction genes [66] allowed to identify a new REase PabI that catalyzes the cleavage of GTA↓C sequence [67].

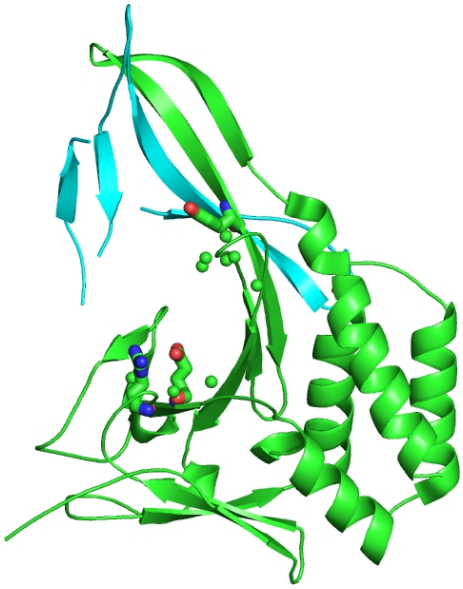


Figure 4. Crystal structure of the PabI [PDB ID 2DVY] that represents 'half-pipe' catalytic fold with one monomer in green and β-sheet of another monomer, that completes 'half-pipe' element, colored in cyan. Residues predicted by bioinformatic methods to be functionally important are shown by their Cα atoms, residues shown by mutagenesis to be crucial for DNA cleavage are represented by sticks.

The crystal structure of the apo enzyme [38] revealed a novel nuclease fold that consists of ten β-strands, five α-helices and two 3_{10} helices folded into the α/β structure with a topology of $\beta\beta\beta\beta3_{10}\beta\alpha\alpha\alpha\beta\beta\beta\beta3_{10}\beta\alpha\alpha$ (Figure 4). Two anti-parallel β-sheets made of three and six β-strands each are arranged in a perpendicular fashion with a large α-helical bundle supporting inner faces of both β-sheets. A highly curved anti-parallel β-sheet is formed when two monomers dimerize through two long β7 strands forming highly curved β-sheet made of twelve β-strands (Figure 4). In this aspect PabI is similar to the lipocalin fold, however it possesses only half of the barrel, therefore it was named "half-pipe"[38].

DNA-binding mode of this fold was modeled by computational docking of 24 bp long ideal B-DNA, assuming that the DNA should be bound in the positively charged cleft of the homodimer [38]. Based on this model thirteen putative residues involved in sequence recognition or catalysis of the phosphodiester bond cleavage were predicted (Figure 4). Substitution of these residues with alanine has indicated that Arg32, Glu63, and Tyr134 are crucial for the catalytic activity of R.PabI and might be involved in the active site formation (Figure 4) [38]. It was also experimentally demonstrated that the putative active site of 'half-pipe' nuclease is Mg^{2+} independent [38].

1.4 Multidomain Type II restriction endonucleases

In theory, one can expect many combinations of the discussed nucleolytic domains with sequence-recognizing machinery in functional restriction endonucleases. Indeed, crystallographic and bioinformatic analysis of REases identified a plethora of non-nuclease (or inactive nuclease) domains fused to the nucleolytic core [4] (Figure 5). Interestingly, only three of the five described domains – PD-(D/E)XK, ββα-Me and PLD – have been found in fusions with foreign domains. Most often the helix-turn-helix (HTH) or "winged" helix-turn-helix (wHTH) domains are attached to the PD-(D/E)XK and ββα-Me domains. In the context of REases HTH(wHTH) motif recognizes DNA of various lengths and composition. In the Type II REases SdaI [47] and FokI [68] this domain is located at the N-terminus.

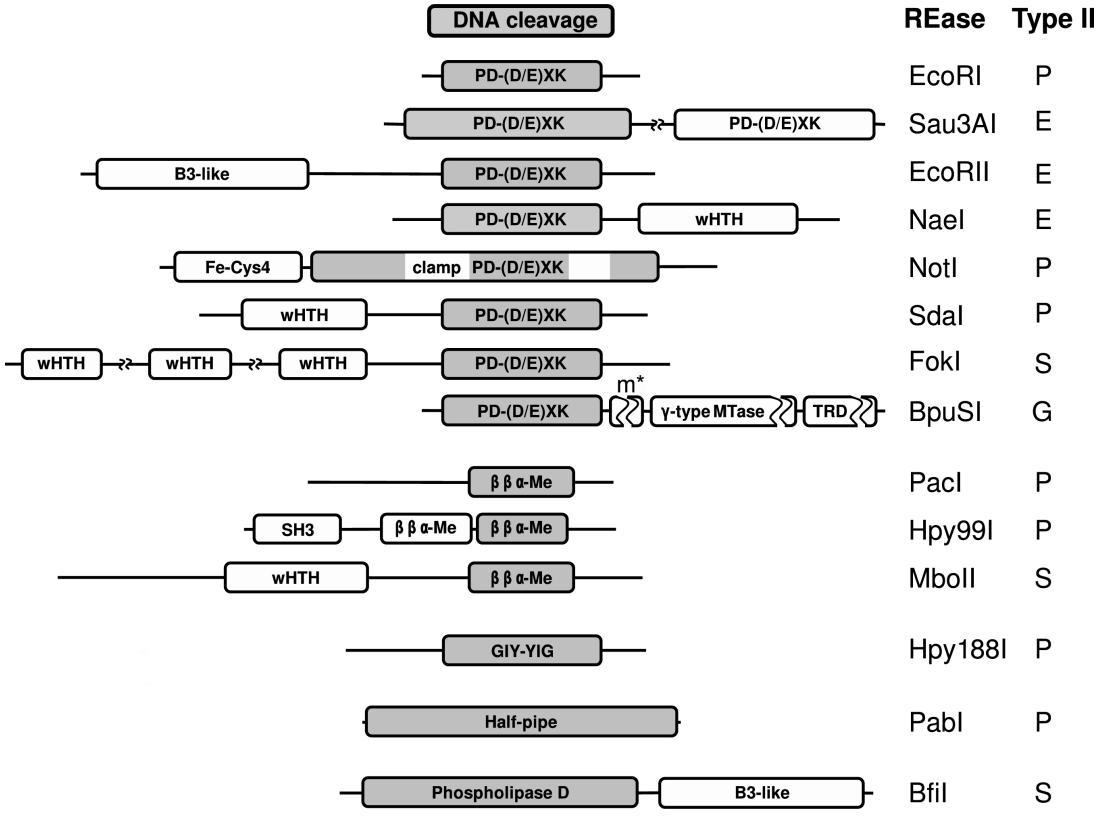


Figure 5. Variety of structural architectures among crystallographically (except MboII) characterized Type II restriction endonucleases is depicted on the level of primary sequences. Active cleavage domains are depicted in gray and aligned, for comparison the length of PD-(D/E)XK domain in EcoRI is 110 a.a.

In SdaI it probably dimerizes and binds symmetric 8 bp DNA site, meanwhile in FokI it recognizes asymmetric 5 bp DNA site as a monomer. In Type IIE

REase NaeI the HTH domain is located at the C-terminus of the protein and recognizes a symmetric 6 bp long DNA sequence as a dimer [69]. Other REases, such as NotI, have Fe-Cys4 metal binding domain attachment to as well as subdomain insertions within the PD-(D/E)XK nuclease domain [70], participating in the binding/recognition of the symmetric 8 bp target. B3-like β-barrel domain in the Type IIE REase EcoRII [14] binds pseudosymmetric 5 bp DNA [71] and in the BfiI REase [21] it is involved in the recognition of an asymmetric 6 bp DNA target [25]. Another β-barrel domain of an SH3 fold is found attached to the N-terminus of the ββα-Me domain in the Hpy99I structure [39]. However, so far the DNA binding function has not been demonstrated for this SH3 domain. A wHTH domain is observed in the gene sequence of MboII next to the catalytic ββα-Me, but currently there is no crystal structure of this domain pair. Up to now GIY-YIG and half-pipe catalytic domains were not found attached to the other DBD within REases [4].

Combining of REase domains is a promising method for creating artificial nucleases with designed specificity. REase derived nuclease domains of PD-(D/E)XK [72][73] and PLD [74][75] families have been already fused to various DBDs resulting in artificial nucleases. Also, the domain of the human methyl-CpG binding protein 2 (MeCP2) that is specific to methylated DNA [76] was fused to the cleavage domains of the REases FokI or BmrI to generate chimeric endonucleases that can be used in studies of CpG island modification in tumor suppressor genes [77].

1.4.1. Duplicated PD-(D/E)XK domains: Sau3AI

The Sau3AI REase recognizes the palindromic sequence 5'-↓GATC [3]. Bioinformatic protein sequence analysis identified that Sau3AI has two fragments of high sequence homology and probably consists of two similar N- and C-terminal domains of PD-(D/E)XK fold [78] (Figure 5). However at the a.a. sequence level the full catalytic $PD^{57}X_6E^{64}VK$ motif was identified only at N-terminal domain (a.a. 1-219) [78].

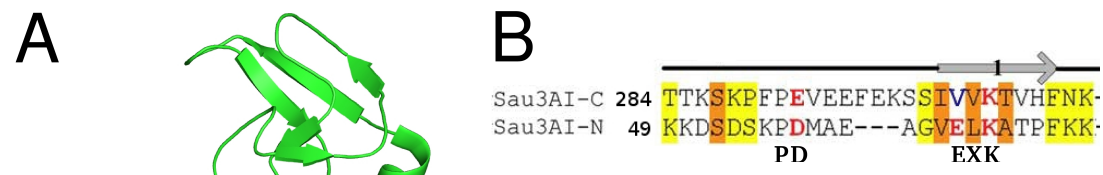


Figure 6. (A) Structure of the C-terminal effector Sau3AI domain [PDB ID 2REU]. (B) Fragment of the a.a. sequence alignment between N- and C-terminal domains of Sau3AI adapted from [79]. In (A) Val substituted active site Glu is colored in blue, conserved Lys – in red.

Computational model of Sau3AI suggested a pseudo-dimeric enzyme [78]. This model was later supported by biochemical studies that showed strong Sau3AI preference for DNA substrates with two GATC sites over those with one site [80]. Furthermore, electron microscopy [80] and tethered particle motion studies [81] have confirmed that Sau3AI interacts with two DNA sites. Recently, it was shown that an isolated C-terminal domain of Sau3AI is able to bind specifically to a single DNA site as a monomer but has no cleavage activity confirming it's role in effector DNA binding [79]. On the other hand the apo crystal structure of the C-terminal domain (a.a. 233-489) [79] indeed revealed the PD-(D/E)XK fold (Figure 6A) but the catalytically important Glu64 residue present in the N-terminal domain is replaced by a valine moiety (Figure 6B). This substitution is thought to be the major reason inactivating this domain.

1.4.2. Fused HTH and PD-(D/E)XK domains: FokI, SdaI, NaeI

The prototypical Type IIS REase FokI is a monomeric enzyme [82] that recognizes the asymmetric DNA sequence GGATG(9/13)[1] [83]. Limited proteolysis experiments revealed two domains (Figure 5): the N-terminal DNA binding domain and the C-terminal domain which is active as a nonspecific nuclease [84]. Crystal structure of the liganded FokI [68] revealed that the DNA bound N-terminal domain consists of three compactly clustered wHTH

1 Cleavage position downstream the recognition site is indicated in the parentheses.

motifs connected by a linker to the C-terminal domain of a typical PD-(D/E)XK fold (Figure 7A). Interestingly, this structure represents an inactive FokI complex, because the monomeric nuclease domain is positioned away from the cleavage site on the DNA (Figure 7A). Mutations introduced at the interface between the DNA binding and the catalytic domains relax the specificity of FokI, enabling cleavage at hemimethylated sites [85]. Thus, after binding to the recognition site the interface between the wHTH and the PD-(D/E)XK domains has to change allowing the catalytic domain to relocate the active site to the scissile phosphate [86].

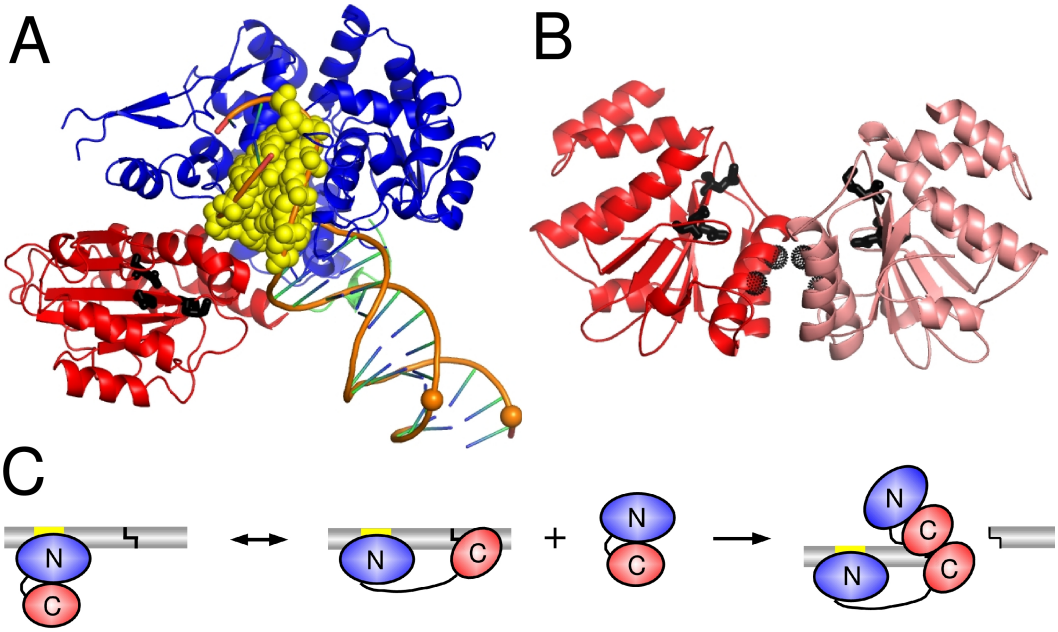


Figure 7. Crystal structures of (A) DNA bound FokI monomer [PDB ID 1FOK] and (B) dimeric catalytic domain of FokI [prepared from PDB ID 2FOK]. wHTH and PD-(D/E)XK domains are depicted in blue and red correspondingly; DNA recognition site is shown as yellow spacefill and scissile phosphate atoms as orange spheres, active site is shown as black sticks. In (B) different monomers are colored in red and light red, black dots marks positions of dimer-disrupting mutations. (C) Cartoons illustrate reaction pathway of FokI based on [86], color coding as in (A), cleavage site is defined by black stripes.

The crystal structure of the apo FokI protein [87] revealed that in order to form a catalytic unit with two active sites the catalytic domains of FokI form a dimer, resembling an orthodox dimeric REase, however, the interdomain interface in FokI is only 800 $Å^2$ (Figure 7B). The alteration of the dimerisation interface at the C-terminal domain with two point mutations (Figure 7B)

abolished the FokI activity [88]. Such relatively small interface explains why FokI is a monomer in solution [82].

Recently it was demonstrated that while FokI monomer is bound to the recognition site by its N-terminal domain, its active site, which is located on the C-terminal domain, cuts the scissile bond of the bottom DNA strand 13 nt downstream the site [89]. To cut the top strand 9 nt away of the site a second monomer of FokI is required or the original FokI monomer has to relocate its active site to the top strand. It was shown that the second monomer can be provided *in trans* [88][82] or *in cis* [90]. Recently fluorescence resonance energy transfer studies of the FokI reaction pathway on a DNA with a single recognition site demonstrated the motion of the DNA-bound effector and catalytic domains as well as the enzyme dimer assembly on DNA (Figure 7C) [86].

The SdaI restriction enzyme recognizing the 8 bp sequence 5'-CCTGCA↓GG [3] is an example of a multidomain Type IIP enzyme (Figure 5) that is specific to the palindromic DNA site [47]. Crystal structure of apo SdaI [47] revealed an N-terminal wHTH DNA recognition motif fused to a C-terminal PD-(D/E)XK nuclease domain (Figure 8A).

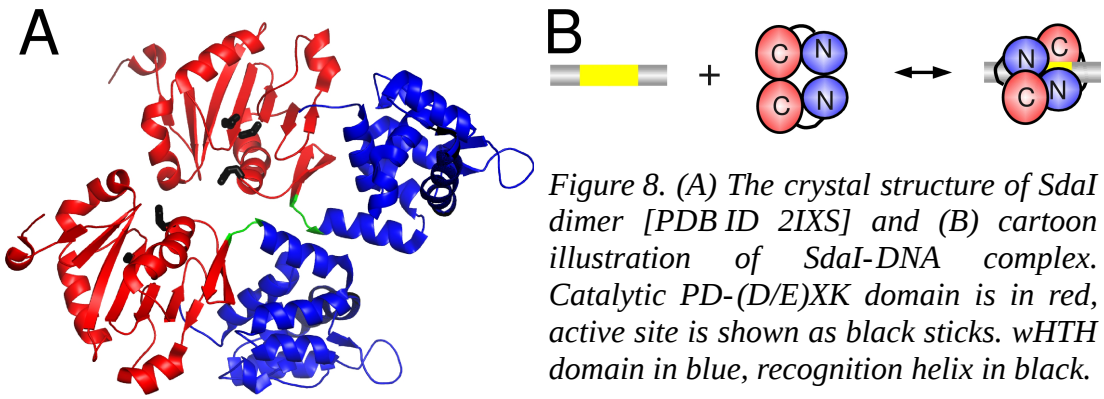


Figure 8. (A) The crystal structure of SdaI dimer [PDB ID 2IXS] and (B) cartoon illustration of SdaI-DNA complex. Catalytic PD-(D/E)XK domain is in red, active site is shown as black sticks. wHTH domain in blue, recognition helix in black.

The domains of SdaI were separated by limited proteolysis and their oligomeric state was analyzed by size-exclusion chromatography. Biochemical analysis of the SdaI domains obtained by limited proteolysis revealed that the DNA-binding domain is a monomer in solution while the catalytic domain is a dimer [47]. Contrary to FokI, an isolated catalytic domain of SdaI showed no

DNA cleavage. It was suggested that the SdaI dimer seen in the crystal structure is not biologically relevant, because the distance between the active sites (Figure 8A) is two times larger than the distance between scissile phosphates in B-form DNA [47]. However, an isolated N-terminal wHTH domain retained its function and bound cognate DNA albeit with much lower affinity than full length SdaI. Therefore, it was proposed that the functionally active SdaI dimer could bind the symmetric DNA recognition sequence with two wHTH motifs, each recognizing one 4 bp half-site only [47]. Thus, the N- and C-terminal domains have to rearrange to generate a biologically active SdaI dimer (Figure 8B) allowing N-terminal domains to bind DNA and C-terminal domains to position their catalytic residues at the scissile phosphate [47] to perform double strand DNA cleavage [91].

The Type IIE REase NaeI recognizes the 6 bp sequence 5'-GCC↓GGC [3]. Unlike the previously described REases FokI and SdaI it requires binding of two specific DNA sites to cleave only one of them, the other one being an allosteric effector [92]. Proteolytic cleavage identified two protease-resistant modules (Figure 5) [93], and the crystal structure of NaeI in apo-form [94] revealed dimeric organization. Each monomer of NaeI is composed of an N-terminal catalytic domain of the PD-(D/E)XK fold and a C-terminal domain with the HTH DNA-recognition motif (Figure 9A).

A crystal structure of DNA-bound NaeI [69] (Figure 9B) revealed two different ways of the symmetric site recognition by the catalytic and effector domains. The catalytic domain secured most of the base-specific interactions by placing β-turns from each subunit into the DNA major groove, while the active sites are positioned to the scissile phosphodiester bonds from the DNA minor groove side (Figure 9B). The effector domain symmetrically placed two DNA recognition helices, each against one half-site of the palindromic recognition sequence (Figure 9B) [69]. Schematically the apo structure of NaeI can be compared with closed “spring-type clothespeg” in which prongs are protein monomers (Figure 9C).

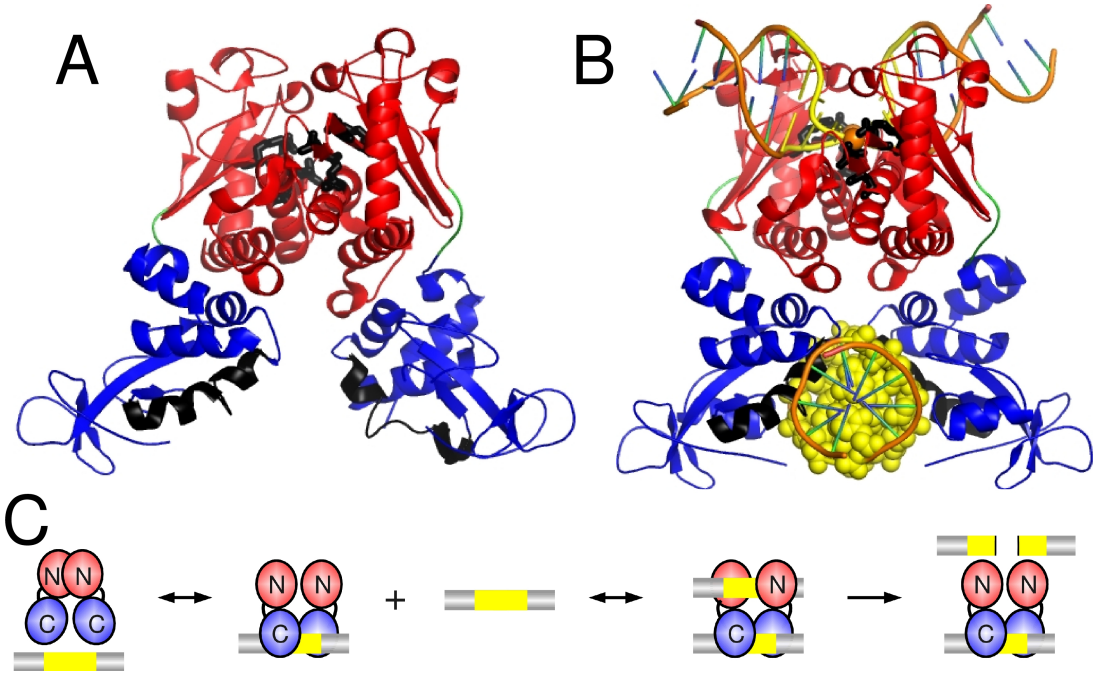


Figure 9. The crystal structures of NaeI in apo (A) [PDB ID 1EV7] and DNA-bound form (B) [PDB ID 1IAW]. The N-terminal catalytic PD-(D/E)XK domain is shown in red and HTH DBD is shown in blue. DNA is represented as cartoon, with recognition site in effector DNA and scissile phosphate atoms in substrate DNA highlighted as yellow and orange spheres, respectively. Recognition helix of HTH domain and active site residues in PD-(D/E)XK domain are shown in black. (C) Cartoons illustrate possible reaction pathway of NaeI, color coding for DNA recognition site and protein domains as in panel (B).

Thus, in the apo-form the dimeric catalytic domain is in the closed conformation and the space between PD-(D/E)XK units, the imaginary grip of the peg, is too small to accommodate the DNA double helix, while the DNA binding cleft of the effector domain, considered to be the top of the peg, is too wide [94] to form specific complex with effector DNA (Figure 9A). Comparison of the structures of apo and liganded NaeI shows a notable rearrangement of the dimer upon DNA binding. The DNA recognition cleft of the effector domain becomes narrower by 16 Å in the complex with DNA (Figure 9A and B). The catalytic cleft, in contrast, widens by 8 Å and completely encircles the second DNA duplex [69]. Mutations at the interdomain interface were proved to disrupt communication between the effector and the catalytic domains [95].

1.4.3. Fused B3-like and PD-(D/E)XK domains: EcoRII

The Type IIE REase EcoRII recognizes the 5'-↓CCWGG sequence and cleaves the phosphodiester bond before the first cytosine [96]. Experiments showed that EcoRII cleavage of the T3 and T7 phage DNA requires stimulation by an additional DNA molecule that works as an allosteric activator [18]. It was shown by limited proteolysis that EcoRII has two functional domains [97]. DNA binding by both isolated domains was proven by gel mobility shift analysis and size-exclusion chromatography [97][71]. The N-terminal domain EcoRII-N is a monomer in solution and binds DNA with the single recognition site 5'-CCWGG as a monomer. The C-terminal domain EcoRII-C cuts the single DNA site in the absence of effector DNA [97]. Furthermore EcoRII-C cleaves substrates with multiple recognition sites much faster than the full-length enzyme [97]. Interestingly, it was shown that EcoRII-C retains specific methylation sensitivity for internal cytosines within the 5'-CCWGG sequence [98]. Recently is was demonstrated that the isolated EcoRII-N domain in equimolar amounts is able to inhibit the catalytic activity of the EcoRII-C REase [99].

The first crystallization trials of EcoRII were reported in 1998 [100]. In 1999 crystals of apo enzyme were reported that diffracted X-rays to 4.0 Å resolution [101]. Low quality crystals of an EcoRII complex with 11 bp DNA were reported a year later [102]. Further work resulted in apo EcoRII crystals that diffracted to 2.8 Å resolution but still were unsuitable for structure solution [103]. The major breakthrough was made in 2003 when a crystal of the apo EcoRII R88A mutant diffracting to 2.1 Å was obtained [104] leading to a structure solution [14]. Different approaches were used to solve EcoRII structure with DNA including crystallization of inactive EcoRII mutants that bind DNA tightly [105].

Crystal structure of apo EcoRII-R88A mutant [PDB ID 1NA6] revealed a compactly packed dimer with each protomer consisting of the N-terminal effector domain with a novel DNA-binding pseudobarrel fold and the

C-terminal catalytic domain with a typical PD-(D/E)XK fold (Figure 10A and B). EcoRII has three putative interfaces for the specific interaction with DNA. In apo EcoRII two N-terminal domains sterically occlude the third DNA binding interface, the dimeric C-terminal domain (Figure 10A and B). In this autoinhibited form EcoRII could scan the substrate DNA for the effector CCWGG target site [106]. DNA cleavage experiments showed that the enzyme requires simultaneous binding of three rather than two recognition sites to achieve concerted DNA cleavage at a single site [15] and the three-site DNA interaction model for EcoRII has been proposed (Figure 10D). On the other hand EcoRII-C does not require interactions with the effector DNA for optimal cleavage [15] (Figure 10C).

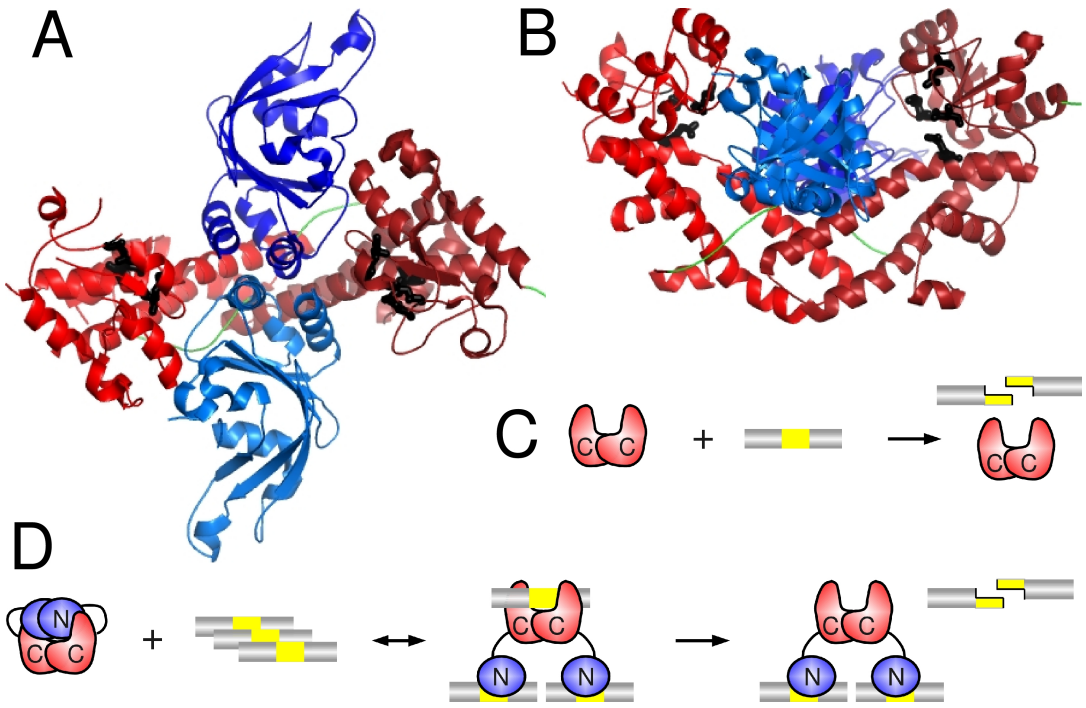


Figure 10. (A, B) The crystal structure of the apo EcoRII [PDB ID 1NA6] [14] in two orientations. Effector B3-like N-terminal domains are in dark and light blue, C-terminal dimeric PD-(D/E)XK domain is in dark and light red, active site residues are represented as black sticks. Cartoons illustrating reaction pathways for (C) REase EcoRII-C and (D) full length EcoRII as proposed by [15], color coding of domains as in (A).

1.4.4. Fused B3-like and PLD nuclease domains: BfiI

The Type IIS restriction endonuclease BfiI recognizes an asymmetric sequence 5'-ACTGGG(5/4) [19]. It was the first discovered Me^{2+}-independent

REase [20]. Analysis of the protein sequence and mutational data indicated that BfiI is a member of the PLD superfamily [22]. Apo and DNA-bound BfiI is a dimer in solution and it is able to cleave DNA when bound to a single recognition site, however it displays optimal activity after binding two sites [22]. BfiI hydrolyzes two strands of DNA outside of the recognition site in two consecutive steps: first the bottom strand is cleaved and only then the top strand is processed [107]. Limited proteolysis experiments have shown that BfiI consists of a dimeric N-terminal domain with a nonspecific nuclease activity and a C-terminal domain which binds to the recognition sequence as a monomer [25]. Random gene dissection experiments have reconfirmed the position of an interdomain region within BfiI [108].

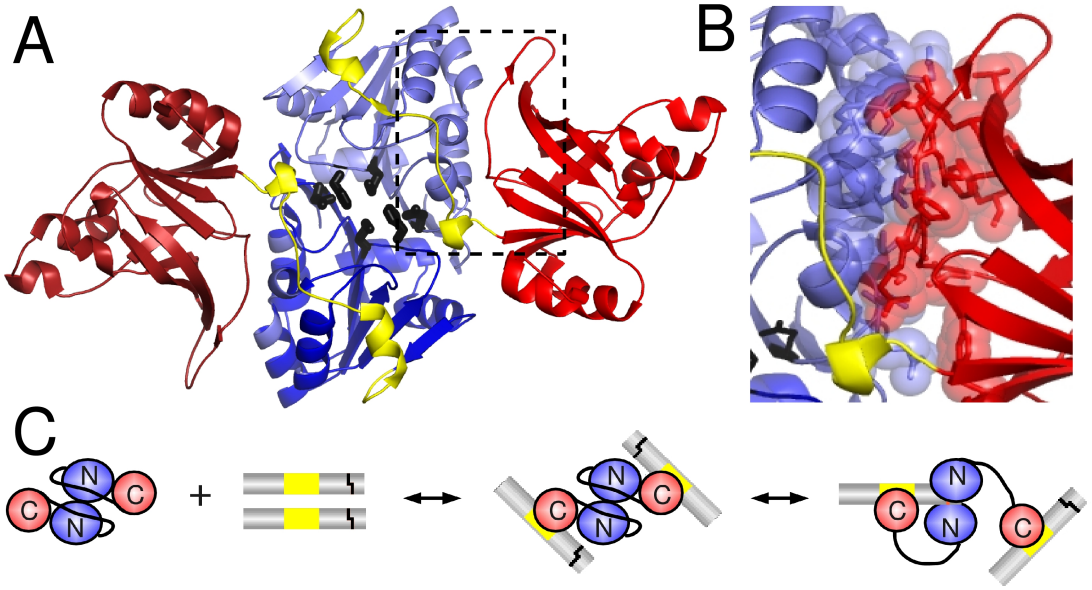


Figure 11. (A) The crystal structure of the apo BfiI [PDB ID 2C1L] and close view of interdomain interface within one of the monomers is shown in panel (B). B3-like C-terminal domains are in dark and light red, dimeric N-terminal PLD nuclease domain is in dark and light blue, interdomain linkers are colored yellow, active site residues are shown as black sticks. (C) Cartoons illustrating BfiI catalytic activity control as proposed by [21], color coding of domains as in (A), in DNA specific binding site is in yellow, cleavage site is defined by black stripe.

The crystal structure of apo BfiI [PDB ID 2C1L] solved at 1.9 Å resolution showed an unexpected dimer organization [21] (Figure 11A). The N-terminal domain belongs to the PLD superfamily and consists of two α/β-globules that dimerize through β-sheets and the dimer interface buries 3500 $Å^2$. Two

pseudobarrel DNA binding B3-like C-terminal domains are attached to the opposite sides of the BfiI-N, with ~1700 Å^2 (Figure 11B) buried at the interdomain interface that is made mainly by hydrophobic contacts. Cross contacts between the C- and the N-domains from different monomers contribute additional 500 Å^2. In the apo structure interdomain linkers of both monomers occlude active site of BfiI (Figure 11A), preventing DNA cleavage. Thus it was proposed that a large conformational rearrangement of the enzyme is triggered by binding of specific DNA by C-terminal domains, inducing proper positioning of the nuclease active site in the vicinity of target phosphodiester bond (Figure 11C) [21].

1.4.5. Fused MTase and PD-(D/E)XK domains: BpuSI

The RM.BpuSI protein (hereafter termed ‘BpuSI’) is a typical Type IIG REase with Restriction, Modification and Specificity modules (R-M-S) fused into a single polypeptide chain, which recognizes the 5 bp site GGGAC(10/14) [3]. It has been shown that BpuSI cleaves DNA substrates containing only a single target as efficiently as substrates with two target sites. A DNA-free structure of BpuSI solved recently has revealed an elongated monomer that consist of five individually folded domains: a PD-(D/E)XK endonuclease domain, located at the at N-terminus, is connected by a helical ‘m*’ domain to a γ-MTase domain, that is followed by a target recognition domain at the C-terminus (TRD), which in turn is interrupted by an additional domain that is suggested to be involved in DNA binding/recognition (Figure 12A) [109]. Within monomer interdomain contact surfaces are mostly hydrophilic and bury 740-1000 Å^2 of surface area. The REase domain has the overall fold that is closely related to FokI and an easily identifiable catalytic PD-(D/E)XK motif, which is partly covered by the MTase domain (Figure 12A) [109]. It has been proposed that the DNA cleavage and methylation activities of BpuSI's are coupled to large movements of the REase, MTase and TRD domains. Both activities are dependent upon the methylation status of the target site [109].

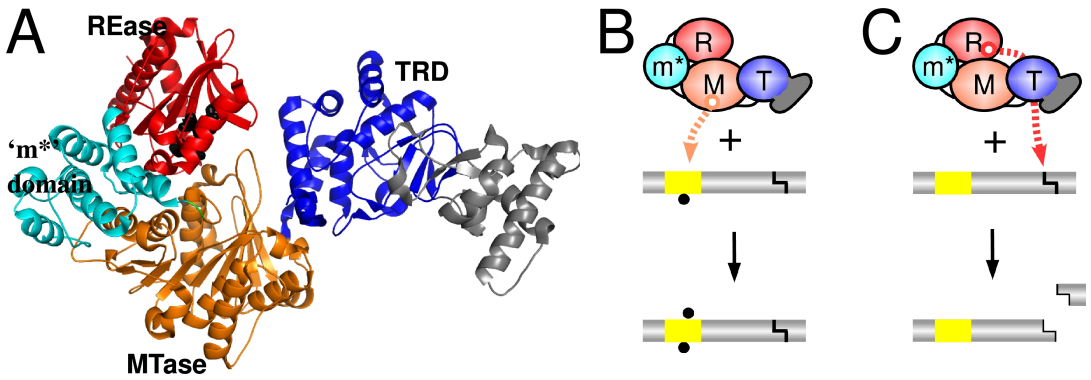


Figure 12. (A) The crystal structure of BpuSI in apo-form [PDB ID 3S1S] [109]. Color coding: the N-terminal PD-(D/E)XK nuclease domain in red and active site residues in black, MTase domain in orange, nuclease-MTase connector domain (m) in cyan, target recognition domain in blue and α/β domain with putative function in gray. (B, C) Cartoons illustrating BpuSI catalytic activity control by methylation status of target site as proposed by [109]. Color coding for domains is the same as in (A), DNA recognition site is in yellow, single and double black dots notes hemi- and dimethylated DNA site, correspondingly. Cleavage site is defined by a black stripe.*

When a hemimethylated target site is bound by the MTase and the TRD domain it invokes a conformational switch with the formation of a catalytically active MTase-DNA complex, resulting in the fully methylated DNA site (Figure 12B). On the other hand, recognition and binding of an unmethylated target triggers a different conformational change leading to the formation of a catalytically active REase-DNA complex and the DNA cleavage (Figure 12C) [109].

1.4.6. Fused metal binding and PD-(D/E)XK domains: NotI

The rare cutting Type IIP homodimeric restriction endonuclease NotI recognizes the 8 bp site 5'-GC↓GGCCGC [110]. X-ray analysis has revealed that the NotI protomer, unlike another rare cutter SdaI [111], consists of several large structural elements folded into three distinct domains. The N-terminal part of the NotI protein forms a metal binding domain Fe-Cys4, which is followed by a typical PD-E(D/E)XK catalytic domain, which in turn contains a protrusion folded into the third helical clamp domain. The catalytic domain hosts a DNA recognition helix and is terminated by extended domain-swapped helices [70].

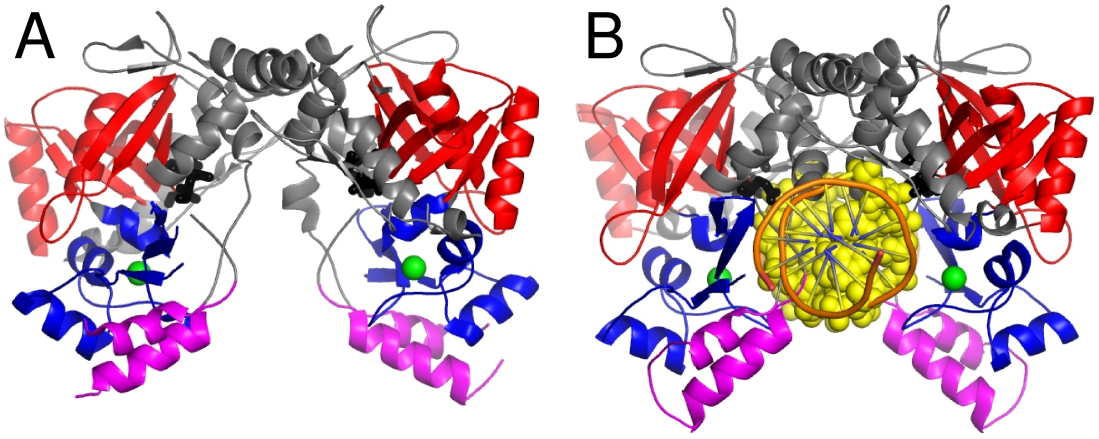


Figure 13. The crystal structures of NotI in (A) apo-form [PDB ID 3BVQ] and DNA-bound form (B) [PDB ID 3C25]. The catalytic PD-(D/E)XK domain is shown in red and active site residues are colored black, metal-binding domain is highlighted in blue, bound irons are shown as green spheres, clamp domain is colored in magenta. In (B) DNA is represented as cartoon, with recognition site shown as yellow spheres.

The compactly folded N-terminal Fe-Cys4 domain, made of a three-stranded antiparallel β-sheet and three α-helices, coordinates an iron ion by a tetrahedral Cys4 motif. This motif has not been observed in Type II REases before (Figure 13A) [111]. The DNA-bound structure of NotI shows that despite being tightly associated with the nuclease core by the clamp region, the metal-binding domain is not coupled to the enzyme's active site and makes only nonspecific contacts to the DNA backbone. A superposition of the apo structure onto the DNA-bound counterpart did not reveal any large rearrangements of the nuclease core. However, upon DNA binding significant motion of the clamp and the metal-binding domains is observed in both monomers. Hence, when NotI encloses the DNA, the distance between the iron ions decreases by ~6 Å [70] and the clamp regions are placed into the DNA major groove (Figure 13B).

1.5. Family of REases recognizing common DNA motif CCGG

Structural comparison of NgoMIV, Bse634I and Cfr10I allowed to propose existence of a distinct subfamily of structurally related REases all of which recognise sites containing the central 5'-CCGG tetranucleotide and cleave before the first cytosine residue [11]. Indeed, active site residues as well as

recognition determinants of the CCGG site are structurally conserved (Table 4) and were confirmed by mutagenesis [112]. Bioinformatic analysis revealed that such a.a. motifs can be identified in plethora of other REases that share the common tetranucleotide in the recognition sequence and also the cleavage site [113][112].

Table 4. Conservative CC:GG recognition mechanism observed for related dimeric and tetrameric Type II REases

REase	Subtype	Recognition site	C_1C_2:G_1G_2 half site			
			C1	C2	G1	G2
			the major groove interactions			
NgoMIV	IIF	G↓CCGGC	D193	D193	R191	R193
SgrAI	IIF	CR↓CCGGYG	D248	D248	R246	R249
Ecl18kI	IIF	↓CCNGG	D187	D187	R186	R188
PspGI	IIP	↓CCWGG	E165	E165	R164	R166
			the minor groove interactions			
NgoMIV	IIF	G↓CCGGC				
SgrAI	IIF	CR↓CCGGYG				
Ecl18kI	IIF	↓CCNGG			Q114	Q114
PspGI	IIP	↓CCWGG			Q94	Q94

Despite the fact that recognition sites of some REases such as Ecl18kI, PspGI and EcoRII are interrupted (Table 4), the functional importance of the common DNA recognition motifs was demonstrated by mutational analysis in Ecl18kI [112], EcoRII [71] and confirmed by the crystal structure of apo-EcoRII [14]. The mechanism of interaction with an interrupted recognition sequence was revealed when the crystal structure of liganded Ecl18kI became available [16]. Ecl18kI, contrary to the expected paradigm [43], does not change the EcoRI-like dimerisation mode, but rather deforms DNA by flipping the nucleotides of the central base pair out of the duplex (Figure 17A and B). The rest of the DNA double helix contracts; as a result the distance between scissile phosphates in liganded Ecl18kI and their positions with respect to the active site residues become very close to the values observed in the NgoMIV-DNA complex (Figure 17C). Similar combination of DNA distortion and a

conservative recognition motif is used by the Type IIP REase PspGI (5'-↓CCWGG) [13] (Table 5).

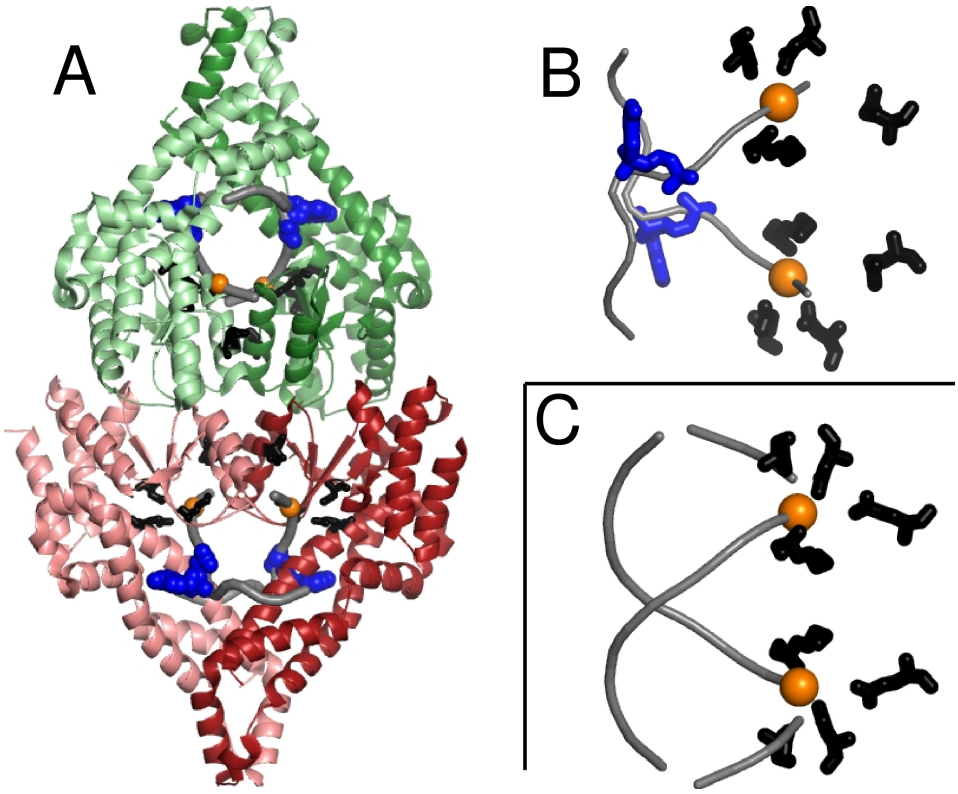


Figure 17. Crystal structure of tetrameric Type IIF REase Ecl18kI in DNA-bound form (A) [PDB ID 2FQZ]. Monomers of primary dimers are colored in dark/light red and green, active site residues are shown as black sticks, DNA is presented as smoothed gray ribbons, flipped bases shown as blue spheres. (B, C) Different DNA conformation in complex with (B) Ecl18kI and (C) NgoMIV [PDB ID 4ABT], phosphate atoms in 5'-phosphate groups in (C) or scissile phosphates in (A, B) are shown as orange spheres.

Plants DBDs interacting with DNA through β-sheet

Reichmann *et al.* identified more than 10 families of plant-specific transcription factors, classified according to the DNA binding domain they posses [114]. Recent structural and bioinformatic studies revealed that many plant DBDs are widespread in other kingdoms of life [115]. However, in contrast to prokaryotes or eukaryotes which usually employ α-helical DBD motifs, like HTH domains [116], plant DBDs often bind DNA through β-stranded regions. Such population of distinct DBDs and recognition

sequences may be important in preventing interference with gene expression by symbiotic or pathogenic organisms [115].

2.1. AP2 and B3 DNA binding domains

Cloning and sequencing of the transcription factor RAV1 (related to ABI3/VP1) from *Arabidopsis thaliana* showed two distinct a.a. sequence domains found at that time only in higher plant species [117]. The N-terminal DBD is homologous to the AP2 domain of the *Arabidopsis* transcription factors APETALA2 [118] (Figure 18). The conserved C-terminal domain is similar to the B3 domain, first identified as a third basic domain in the protein coded by VIVIPAROUS1 gene (VP1) [119] (Figure 18). TFs with such combination of domains were called the RAV subfamily [114].

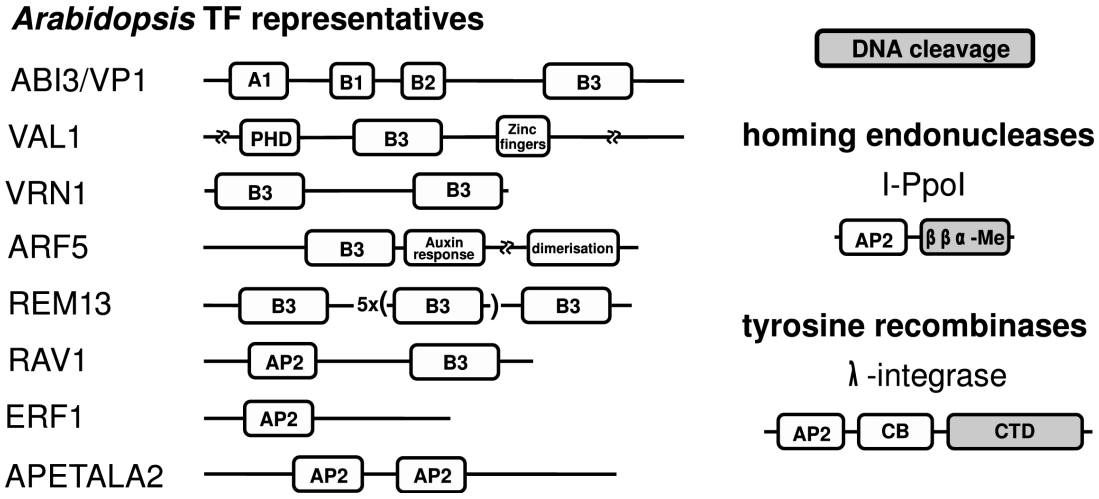


Figure 18. Schematic representation of TF carrying AP2 or/and B3 domains and of DNA modifying enzymes carrying AP2 domain. For comparison length of the B3 domain is 100 a.a.. Abbreviation used: A1 stands for an acidic domain, PHD – for a plant homeodomain, CB – for a domain involved in binding of core-type sites, CTD – for a C-terminal catalytic domain.

Binding site selection assays showed that RAV1 binds specifically to bipartite recognition sequences composed of 5'-CAACA and 5'-CACCTG motifs, which can be separated by 2-8 bp spacings as well as differently oriented [117]. Further analysis revealed that the AP2 and B3 domains can autonomously bind to the 5'-CAACA and 5'-CACCTG sites, correspondingly.

The AP2 domain consists of three β-strands followed by an α-helix (Figure 19A) and is widespread among many eukaryotic transcription factor families which differ by protein organization and have diverse recognition sites. Plant TF AP2 subfamily named ERF (Ethylene Responsive Factors) [120] is specific for 6 bp sequence 5'-RCCGAC. Bioinformatics analysis identified AP2 domain in cyanobacterium, ciliate and viruses [121]. Experimental studies revealed that AP2 domain is transferred by homing endonucleases encoded in plant germ line-limited mobile elements [122] that are present in malarial, cattle and human parasites [123]. Proteins from the Apicomplexan AP2 (ApiAP2) subfamily have been shown to contribute to the regulation of gene expression at various stage of *Plasmodium* development [124]. Homing endonucleases make DNA double strand break at the target site triggering the movement of mobile self-splicing introns [125].

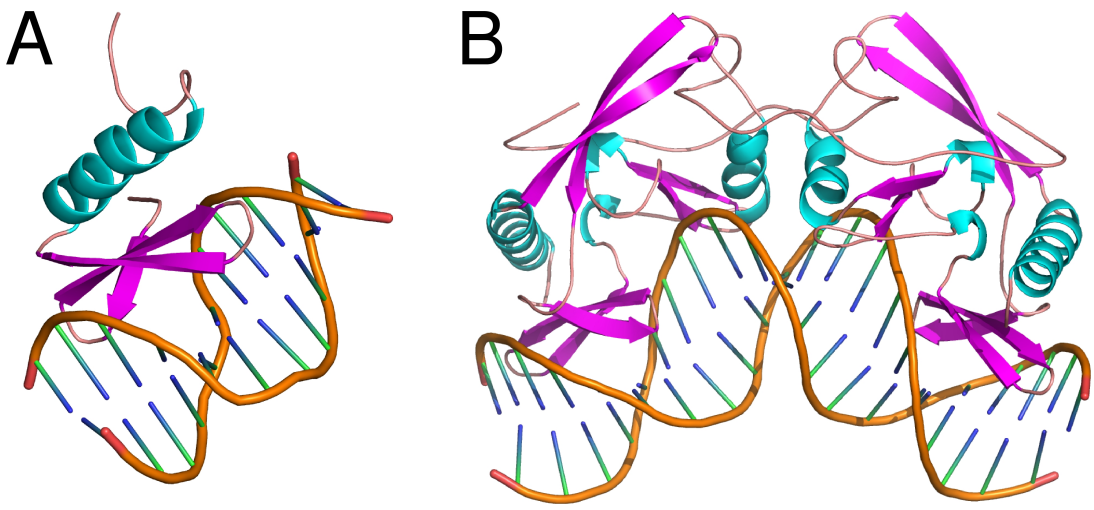


Figure 19. (A) NMR structure of AP2 domain from AtERF1 in complex with specific DNA [PDB ID 1GCC]. (B) Crystal structure of the DNA-bound dimeric I-PpoI homing endonuclease [PDB ID 1A73] with two AP2 domains placed at the DNA major groove.

In the prototype enzyme I-PpoI specific for the homing site 5'-CTCTCTTAA↓GGTAGC [126] the AP2 domain is fused to the N-terminal ββα-Me nuclease domain (Figures 18 and 19B) [127]. The tolerance of I-PpoI for an extended set of recognition sites may be evolutionary advantageous for lateral transfer of mobile intron *P. polycephalum* LSU3 [128].

Bioinformatic analysis revealed that the **B3 domain** is widespread among the plant TFs called the B3 family and a.a. sequence similarities of such TFs allowed to subdivide them into distinct B3 subfamilies. Biochemical studies showed that domains from each B3 subfamily have different specificities towards different DNA sequences: they bind the 5'-TGTCTC site in the ARF (auxin response factors) subfamily [129], the 5'-CATGCA site in the ABI3 (ABA-insensitive factor) subfamily [130] or have no determined sequence specificity in the REM (expressed in reproductive meristems) subfamily [131]. Also, in the REases EcoRII and BfiI the B3-like domain recognizes the 5'-CCTGG and 5'-ACTGGG sites, respectively, that illustrates adaptability of this domain to recognize DNA sites of different length and composition. Noteworthy, binding studies have shown that the isolated B3 domain of RAV1 binds the target site with K_d~10^{-7} M^{-1} while a non-specific sequence is bound with K_d~10^{-6} M^{-1} [24]. On the other hand, the isolated B3-like domains of REases EcoRII and BfiI bind their targets sites with high affinity (K_d~10^{-9} M^{-1}) and show virtually no nonspecific DNA binding up to micromolar protein concentrations [25][71].

Systematic analysis of B3 domain containing proteins in 11 sequenced plant species identified numerous combinations of B3 and additional domains (Figure 18) [132]. For example, B3 containing the TFs of the ARF subfamily have an increased affinity to specific DNA that is achieved by homodimerisation of the C-terminal domains of two ARF proteins (Figure 18) [132]. Alternatively, domains (Aux/IAA) that are responsible for heterodimerization of the TFs of the ARF subfamily with TFs of other families enable auxin signaling transduction cascade [133]. Also, the TF VRN1 (response to vernalisation) from the REM subfamily (Figure 18) is comprised of two B3 binding domains, and it is required in *Arabidopsis thaliana* and in other dicotyledonous plants for stable suppression of floral receptor FLOWERING LOCUS C (FLC) which encodes a repressor of flowering [134]. VRN1 displays strong nonspecific DNA binding *in vitro* analyzed by gel-shift

and surface plasmon resonance methods [135], *in vivo* studies proposed its contribution to meiotically unstable chromatin silencing [136].

Interestingly, an analysis of phylogenetic relationships of B3 proteins coupled with an analysis of structural residue conservation suggest that the N-terminal domain of EcoRII belongs to the VAL (VP1/ABI3-like) subfamily proteins that are thought to comprise the root of all B3 family [132]. The B3 domain TFs play an important role in the accumulation of seed storage reserves during seed development [137]. Regulation of the gene expression pathway underlining this physiological process is actively explored [138] and an atomic model of the specific B3 domain interaction with DNA may contribute to our understanding of this important pathway [27][139].

2.1.1. Structural analysis of B3 domain and model of DNA binding

The first NMR structure of the RAV1 B3 domain was reported by Yamasaki *et al.* [24] followed by the At1g16640 structure [140] from the REM subfamily. Both structures share the topology and the minimal DNA-binding pseudobarrel fold with EcoRII-N: seven β-strands are connected in exactly the same way to form a β-barrel that is opened between the strands β1 and β2 (Figure 20A and B). Two short conserved α-helices are inserted between the strands β2/3 and β5/6 (Figure 20A and B). The structural core of the β-barrel consists of hydrophobic side-chains.

NMR titration experiments were performed with labeled RAV1-B3 protein by adding increasing amount of unlabeled DNA [24] and chemical shift perturbations have been observed in heteronuclear single-quantum coherence spectra [141]. Chemical shift changes were completed when protein-DNA molar ratio reached 1:1 and allowed to identify a RAV1-B3–DNA interface: the strands β2, β4 and β5 were the most affected by binding and also some changes were observed for residues from the strand β1 and the helix α1 (Figure 20A).

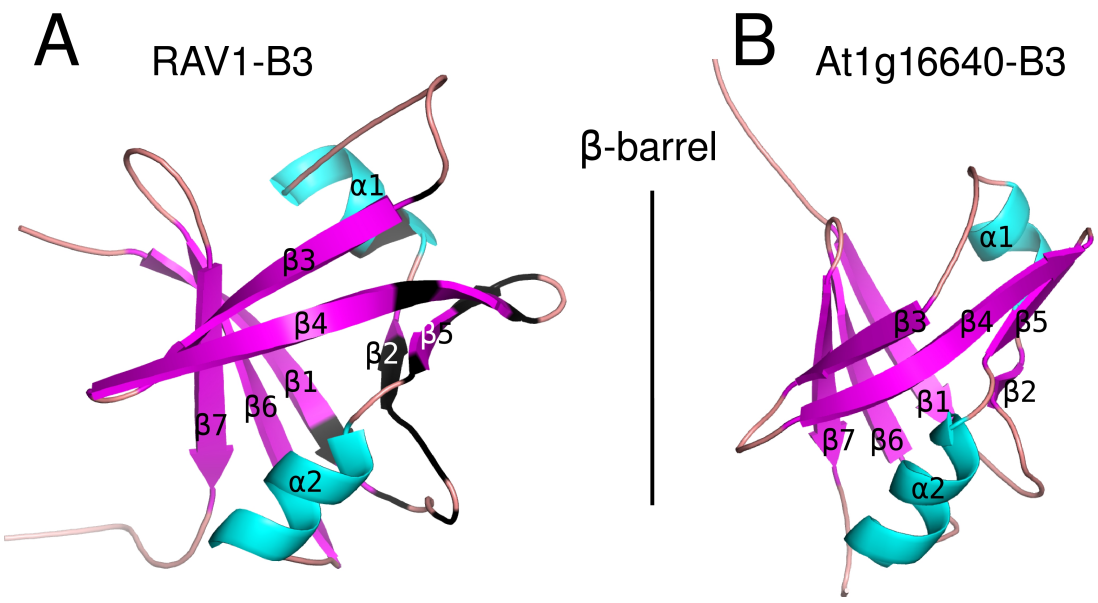


Figure 20. Structures of an open β-barrel of B3 domain from (A) transcription factor RAV1 [PDB ID 1WID] and (B) At1g16640 protein [PDB ID 1YEL]. Helices are colored in cyan, β-strands – magenta, loops – salmon. Black coloring in (A) highlights parts of the RAV1-B3 the most affected by DNA binding [24].

A structural model for the complex of RAV1-B3 with a canonical B-form DNA duplex was generated by docking (method described in [142]) the protein into the DNA major groove (coordinates of the modeled complex are not available). However, the results of the energy minimization calculations in such model should be taken with caution, because they were performed only for a limited number of complexes [24] and did not cover the whole range of absolute protein DNA orientations within the theoretically possible range 0 to 90 degrees.

Interestingly, homology models of ARF1-B3 and ABI3-B3 domains also display large positively charged surfaces containing amino acid residues highly conserved in other members of the family. These residues probably maintain a conserved DNA-binding mode through non-specific interactions [24]. On the other hand, it was proposed that the DNA binding cleft also contains base-contacting residues which contribute to the sequence specificity [24] and those residues should be conserved withing the family. For instance, B3 domains of each subfamily have the characteristic a.a. composition of putative DNA binding loops. For example, the RAV1 β-strands β4 and β5 and the loop

between them, which are largely affected by DNA binding, include the fragment W^{245}(N/R/K)SSQS (Figure 21A), whereas in the ARF subfamily, the corresponding fragment contains R^{68}G(Q/T)P(K/R)R (Figure 21B) and in the ABI3 subfamily protein it contains the W^{67}PNNKS sequence (Figure 21C). In contrast, the equivalent loop of the non-specific DNA binder At1g16640 is much shorter (Figure 20B) [140].

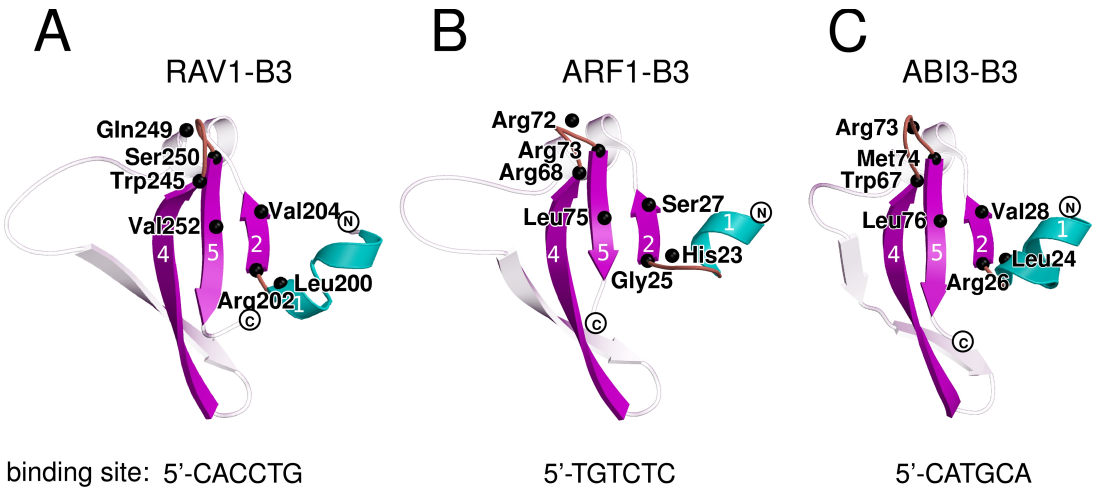


Figure 21. Putative DNA binding clefts of domains from proteins of B3 subfamilies. Helices are colored in cyan, β-strands in magenta, loops in salmon. For clarity protein secondary structure elements connecting β2 and β4 are shown in gray. Cα atoms of putative DNA contacting residues in B3 domains (A) of RAV1 [PDB ID 1WID], (B) of ARF1 and (C) of ABI3 are shown as black spheres [PDB files of ARF1-B3 and ABI3-B3 homology models retrieved from [24]]. Recognition sequences of corresponding domains are provided under binding clefts.

Recently another REM subfamily B3 domain from the VRN1 TF was crystallized and an X-ray diffraction dataset to 2.1 Å resolution was collected [143]. Interestingly, all attempts to use the RAV1 and At1g16640 atomic coordinates [PDB ID 1WID and 1YEL, correspondingly] to solve this structure by molecular replacement were unsuccessful [143]. Later, an NMR study of the same protein fragment provided backbone and side chain resonance assignments for 114 of 129 residues and revealed that its secondary structure elements, seven β-strands and three α-helices, have topology common for B3 domains [144]. This fact reflects a large conformational diversity of the B3 family proteins.

Close structural similarity of B3 domains to B3-like N-terminal domain of EcoRII [24], despite the low sequence identity (19% of identical a.a), allowed to speculate that B3 domains might have arisen evolutionarily in a common ancestor of plants and modern bacteria [27]. However, the possibility of horizontal transfer of EcoRII-N gene into higher plants from symbiotic or pathogenic bacteria species like *Mesorhizobium* and *Agrobacterium* is not excluded [115][139]. Indeed, the gene of the EcoRII isoschizomer Pca17AI has been found in chromosomal DNA isolated from the plant pathogenic bacteria *Pectobacterium carotovorum* and the protein was shown to be an active REase *in vitro* [145].

2.1.2. Nucleic acid-binding B3 fold in context of cradle-loop metafold

Structural classification by natural descent of β-barrel proteins with different topologies allowed to introduce the concept of the cradle-loop barrel metafold [146]. Cradle-loop barrels are named after two symmetric β-hairpins knotted into a six-stranded double-psi barrel [PDB ID 1CZ4] (Figure 22). The metafold also includes β-barrels that are topologically different from the double-psi barrel. Thus, the bacterial TFs MraZ [PDB ID 1N0G] [147] and AbrB-N [PDB ID 1YFB] [148] belong to the same metafold and have pseudodimeric and homodimeric swapped-hairpin barrel structures, correspondingly (Figure 22). Noteworthy, in the AbrB-N–DNA NMR model the cradle-loops enter the DNA major groove [149].

Possible evolutionary relationships for cradle-loop barrels were suggested when the NMR structure of the PhS018 protein from *Pyrococcus horikoshii* was solved [PDB ID 2GLW] [150] (Figure 22). The structure revealed a six-stranded barrel named RIFT (after the occurrence of this topology in the ribosomal protein L3, F1 ATPase and in translation factors) and allowed to propose the existence of an ancestral RIFT barrel that is homodimeric (Figure 22). It was suggested that the double-psi β-barrel fold may originate from the monomeric RIFT by the strand swapping mechanism (Figure 22). On the other hand, insertion of a strand into the hairpin of the homodimeric RIFT

may give rise to a homodimeric swapped-hairpin β-barrel, which further yielded a single chain barrel by duplication and fusion [146] (Figure 22). Finally, the B3 barrel [PDB ID 1WID] may have arisen through the fusion of β-hairpins from two homo-dimeric barrels: an ancestral RIFT and a swapped-hairpin (Figure 22) [151].

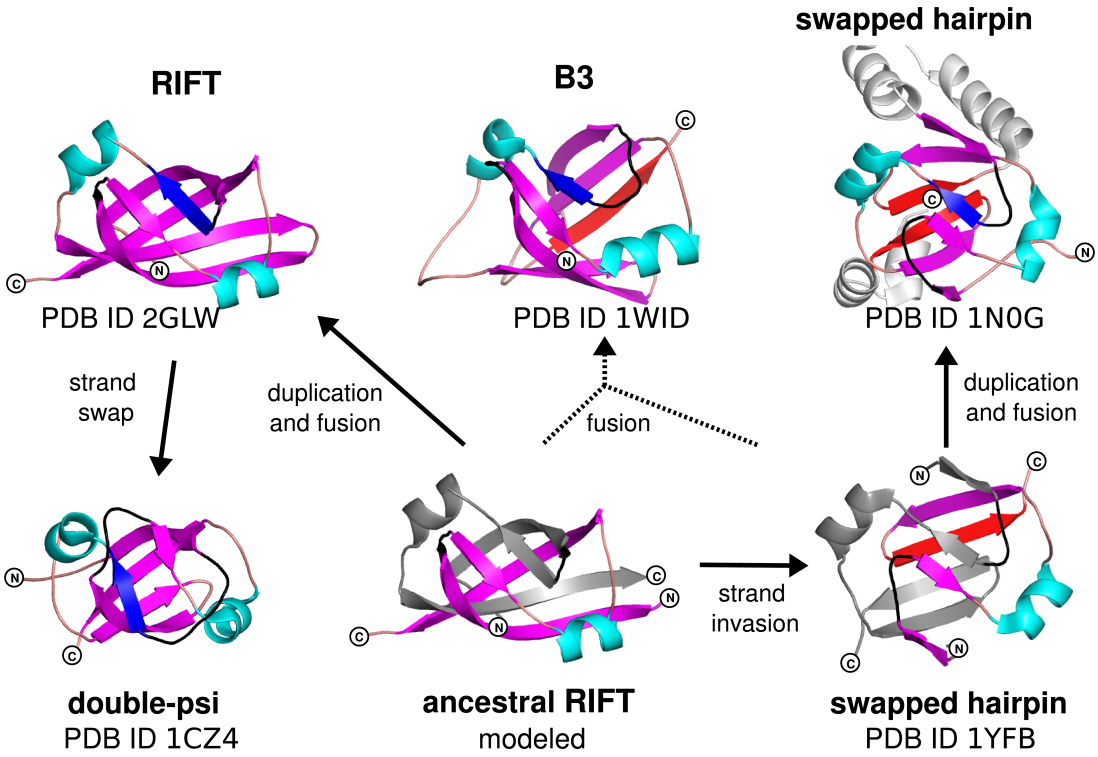


Figure 22. Topological evolution map of the cradle-loop barrel metafold starting from β-hairpin. Minimal β-hairpin in ancestral homodimeric β-barrel RIFT is colored in magenta (strands) and cyan (helix). In other proteins the invading strand is shown in red, the swapped strand in blue, cradle-loops are colored in black. Figure was prepared according to [151].

3. MATERIALS AND METHODS

3.1. Enzymes, sorbents and reagents

T4 polynucleotide kinase was from “Thermo Fisher Scientific” (Vilnius, Lithuania). Trypsin and thermolysin were from “Sigma” (Taufkirchen, Germany). For protein chromatography were used pre-packed HiTrap Heparin, Superdex 75 HR, Superdex 200 HR, Mono Q 5/50 GL columns from “GE Healthcare” (Uppsala, Sweden). Bradford reagent was from “BioRad”. Ultrafiltration devices with various MW cut-off used for the concentrating of proteins were from “Millipore” (Billerica, USA). Crystallization related materials such as plates, sealing tapes, crystallization screens, Izit Crystal Dye were from “Hampton Research” (Aliso Viejo, USA).

3.2. E. coli Strains

E. coli JM109:

F’ traD36 $proA^{+}B^{+}$ $lacI^{q}$ Δ(lacZ)M15 / $e14^{-}$ ($McrA^{-}$) Δ(lac-proAB) endA1 gyrA96 (Nal^{r}) thi-1 hsdR17 ($r_K^{-}m_K^{+}$) glnV44 relA1 recA1.

E. coli ER2267**:**

Δ(argF-lac)U169recA1/F’proAB $lacI^{q}$Δ(lacZ)M15zzf::mini-Tn 10(Kan^{r}).

E. coli ER2566:

F^{-} λ^{-} fhuA2 [lon] ompT lacZ::T7 geneI gal sulA11 Δ (mcrC-mrr)114::IS10 R (mcr-73::miniTn10)2 R(zgb-210::Tn10)1 (Tet^{s}) endA1 [dcm] λDE3.

Bacteria were transformed and cultures were grown according to the standard protocols [152].

3.3. Plasmids

EcoRII expression plasmids pDK1_wt_M.EcoRII (Cm^{r}), pQE30_R.EcoRII-N-H6 (Ap^{r}), pQE30_Y41A_R.EcoRII-H6 (Ap^{r}) were donated by M. Reuter (Institute of Virology, Berlin, Germany). Plasmids for

expression of the BfiI catalytic mutant pET21b-pBfiIR5.1-K107A (Ap^r) and pACYC-BfiIM9.1 (Cm^r) were donated by V. Morin and A. Lagunavičius, respectively.

3.4. Oligoduplexes

All oligonucleotides used in this study were purified by HPLC. All non-modified oligonucleotides, as well as modified ones (containing uridine and inosine) for the Bse634I studies were purchased from "Metabion" (Martinsried, Germany). Modified oligonucleotides containing uridine, 2-aminopurine, 5-methyl isocytosine or isoguanine for EcoRII studies were obtained from "Integrated DNA Technologies" (Coralville, USA). For the DNA binding and kinetic experiments one strand of each duplex was labeled at the 5'-end with [γ-^{33}P]ATP ("Hartmann Analytic", Braunschweig, Germany) using DNA labeling kit ("Thermo Fisher Scientific", Vilnius, Lithuania). Oligoduplexes were assembled by slow annealing from 95°C or 85°C to the room temperature in the buffer (8 mM Tris-OAc (pH 7.9 at 25ºC), 16.5 mM KOAc) or in water (for crystallization). Sequences and short descriptions of the oligoduplexes are provided in the Table 5.

Table 5. Oligoduplexes used in this study

Oligo-duplex	Sequence	Specification
Oligonucleotides[a] used for co-crystallization with EcoRII domains		
SP9	5'-GCCCTGGCG-3' 3'-CGGGACCGC-5'	9 bp cognate
SP12	5'-TAGCCTGGTCGA-3' 3'-ATCGGACCAGCT-5'	12 bp cognate
Oligonucleotides[ab] used for probing of EcoRII-N interactions with central base pair		
T/A	5'-CGCACGACTTCCTGGAAGAGCACGC-3' 3'-GCGTGCTGAAGGACCTTCTCGTGCG-5'	25 bp cognate
U/A	5'-CGCACGACTTCC**U**GGAAGAGCACGC-3' 3'-GCGTGCTGAAGGACCTTCTCGTGCG-5'	25 bp with modified central bp at recognition site
T/2	5'-CGCACGACTTCCTGGAAGAGCACGC-3' 3'-GCGTGCTGAAGG**2**CCTTCTCGTGCG-5'	25 bp with modified central bp at recognition site

Oligo-duplex	Sequence	Specification
U:A	5'-CGCACGACTTC**U**TGGAAGAGCACGC-3' 3'-GCGTGCTGAAGAACCTTCTCGTGCG-5'	25 bp with modified second bp at recognition site
A:U	5'-CGCACGACTTCCTAGAAGAGCACGC-3' 3'-GCGTGCTGAAGGA**U**CTTCTCGTGCG-5'	25 bp with modified fourth bp at recognition site
^{5m}isoC/ /isoG	5'-CGCACGACTTCC^{5m}**iC**GGAAGAGCACGC-3' 3'-GCGTGCTGAAGG **iG**CCTTCTCGTGCG-5'	25 bp with modified central bp
NS33	5'-AATGGGCTCGCACGGCGGCTATTATCGATTGTA-3' 3'-TTACCCGAGCGTGCCGCCGATAATAGCTAACAT-5'	33 bp non-cognate

Oligonucleotides[ac] used for probing of N- and C- EcoRII domains interactions with hemimethylated DNA

NM	5'-CGTAGCCTGGTCGATC-3' 3'-GCATCGGACCAGCTAG-5'	16 bp cognate
HM(A)	5'-CGTAGCCTGGTCGATC-3' 3'-GCATCGGA**M**CAGCTAG-5'	16 bp with 5-methylcytosine in the A-strand
HM(T)	5'-CGTAGC**M**TGGTCGATC-3' 3'-GCATCGGACCAGCTAG-5'	16 bp with 5-methylcytosine in the T-strand
DM	5'-CGTAGC**M**TGGTCGATC-3' 3'-GCATCGGA**M**CAGCTAG-5'	16 bp with 5-methylcytosine in both strands

Oligonucleotides[a] used for co-crystallization with BfiI-C

S12	5'-AGCACTGGGTCG-3' 3'-TCGTGACCCAGC-5'	12 bp cognate without BfiI cleavage site

Oligonucleotides used for co-crystallization with BfiI-NK107A

NS8	5'-TACCGGTA-3' 3'-ATGGCCAT-5'	8 bp non-cognate DNA

Oligonucleotides[a] used for co-crystallization with Bse634I-R226A

GC-1	5'-TCGCGCCGGCGCG-3' 3'-GCGCGGCCGCGCT-5'	13 bp cognate GC-1
AT-1	5'-TCGCACCGGTGCG-3' 3'-GCGTGGCCACGCT-5'	13 bp cognate AT-1
AT-2	5'-TTCGACCGGTCGA-3' 3'-AGCTGGCCAGCTT-5'	13 bp cognate AT-2

Oligonucleotides[bd] used for biochemical analysis of Bse634I

NS	5'-AGCGTAGCACTGGGCTGCTAGTC-3' 3'-TCGCATCGTGACCCGACGATCAG-5'	23 bp non-cognate
AT	5'-CGCACGATCACCGGTGATGCACGC-3' 3'-GCGTGCTAGTGGCCACTACGTGCG-5'	23 bp cognate unmodified
GC	5'-CGCACGATCGCCGGCGATGCACGC-3' 3'-GCGTGCTAGCGGCCGCTACGTGCG-5'	23 bp cognate unmodified
TA	5'-CGCACGATCTCCGGAGATGCACGC-3' 3'-GCGTGCTAGAGGCCTCTACGTGCG-5'	23 bp mis-cognate unmodified

Oligo-duplex	Sequence	Specification
CG	5'-CGCACGATCCCCGGGGATGCACGC-3' 3'-GCGTGCTAGGGGCCCCTACGTGCG-5'	23 bp mis-cognate unmodified
AU	5'-CGCACGATCACCGGUGATGCACGC-3' 3'-GCGTGCTAGUGGCCACTACGTGCG-5'	23 bp mis-cognate modified
IC	5'-CGCACGATCICCGGCGATGCACGC-3' 3'-GCGTGCTAGCGGCCICTACGTGCG-5'	23 bp mis-cognate modified
UA	5'-CGCACGATCUCCGGAGATGCACGC-3' 3'-GCGTGCTAGAGGCCUCTACGTGCG-5'	23 bp mis-cognate modified
CI	5'-CGCACGATCCCCGGIGATGCACGC-3' 3'-GCGTGCTAGIGGCCCCTACGTGCG-5'	23 bp mis-cognate modified

[a] recognition site of REase is underlined;

[b] **2** denotes 2-aminopurine, **I** denotes inosine, **U** denotes uracil, ^{5m}**iC** denotes 5-methyl isocytosine, **iG** denotes isoguanine;

[c] **M** stands for 5-methylcytosine;

[d] central tetranucleotide CCGG of the Bse634I recognition sequence is underlined.

3.5. Buffers

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

Electrophoresis buffer II: 100 mM Tris-borate (pH 8.2 at 25°C), 2 mM EDTA.

Loading dye solution: 95 % formamide, 25 mM EDTA (pH 9.0 at 25°C), 0.01 % bromphenol blue.

3.5.1. Buffers related to EcoRII-N

Storage buffer I: 20 mM Tris-HCl (pH 8.0 at 25°C), 200 mM KCl, 1.0 mM EDTA, 50% glycerol.

Proteolysis buffer: 20 mM Tris-HCl (pH 8.0 at 25°C), 200 mM KCl, 1 mM EDTA, 10% glycerol.

Concentration buffer: 20 mM Tris-HCl (pH 8.0 at 25°C), 0.2 M KCl, 10% glycerol, 0.1 mM EDTA.

Elution buffer I: 20 mM Tris-HCl (pH 8.0 at 25°C), 100 mM KCl, 1 mM EDTA, 10% glycerol.

Reservoir solution: 1 M NaOAc (pH 4.5, 25°C), 0.2 M LiOAc, 10% glycerol.

3.5.2. Buffers related to EcoRII-C

Storage buffer I: 10 mM Tris-HCl (pH 7.4 at 25°C), 200 mM KCl, 1.0 mM EDTA, 1.0 mM DTT, 50% glycerol.

Proteolysis buffer: 10 mM Tris-HCl (pH 7.5 at 25°C), 200 mM KCl, 1 mM DTT, 2 mM $Ca(OAc)_2$, 10% glycerol.

Storage buffer II: 10 mM Tris-HCl (pH 7.4 at 25˚C), 200 mM KCl, 1.0 mM EDTA, 1.0 mM DTT, 50% glycerol.

Concentration buffer: 20 mM Tris-HCl (pH 8.0 at 25°C), 0.4 M KCl, 10% glycerol, 0.1 mM EDTA, 5 mM of $Ca(OAc)_2$.

Reservoir solution: 26% PEG1500, 25% glycerol.

Storage buffer III: 10 mM Tris-HCl (pH 7.4 at 25˚C), 200 mM KCl, 1.0 mM EDTA, 50% glycerol.

Binding buffer: 10 mM Tris-HCl (pH 7.5 at 25˚C), 200 mM KCl, 2 mM DTT, 10% glycerol, 0.1 mg/ml BSA, 5 mM of $Ca(OAc)_2$.

Reaction buffer: 33 mM Tris-OAc (pH 7.9 at 25°C), 66 mM KOAc, 10 mM $Mg(OAc)_2$, 2 mM DTT, 0.1 mg/ml BSA

3.5.3. Buffers related to BfiI, BfiI-C and BfiI-N-K107A

Storage buffer: 10 mM Tris-HCl (pH 7.5 at 25°C), 150 mM KCl, 1 mM EDTA, 1 mM DTT, 50% glycerol.

Proteolysis buffer: 20 mM Tris-HCl (pH 7.5 at 25°C), 200 mM KCl, 1 mM DTT, 4 mM $Ca(OAc)_2$, 10% glycerol.

Concentration buffer: 10 mM Tris-HCl (pH 7.5 at 25°C), 100 mM KCl, 1 mM DTT, 1 mM EDTA, 10% glycerol.

Reservoir solution I: 0.49 M NaH_2PO_4, 0.91 M K_2HPO_4 (pH 6.9 at 25°C).

Reservoir solution II: 0.01 M $MgCl_2$, $Na(CH_3)_2AsO_2$ (pH 7.5 at 25°C), 1.0 M Li_2SO_4

3.5.4. Buffers related to Bse634I

Crystallization buffer: 20 mM Tris-HCl (pH 7.5 at 25°C), 50 mM NaCl, 5 mM $CaCl_2$.

Reservoir solution I: 100 mM NaOAc (pH 4.25-5.5 at 25°C), 10 mM $CaCl_2$ or $Ca(OAc)_2$, 4-8 % (w/v) of PEG8000.

Reservoir solution II: 0.1 M Bis-Tris (pH 5.5 at 25°C), 0.5 % polyvinyl-pyrrolidone, 16 % of PEG400.

3.6. Expression and purification of EcoRII-Y41A and EcoRII-N

Strain JM109 containing plasmid pDK1_wt_M.EcoRII (Cm^r) bearing *ecoRIIM* gene was used as a host for transformation with EcoRII coding plasmids: pQE30_R.EcoRII-Y41A-H_6 (Ap^r) harboring a gene of an inactive full-length EcoRII-Y41A mutant or pQE30_R.EcoRII-N-H_6 (Ap^r), which contains a gene of the EcoRII N-terminal domain. Cells were grown in LB medium and protein synthesis was induced with IPTG as described in [153]. The cells were resuspended in the lysis buffer and disrupted by sonication. N-terminal $(His)_6$-tag fused proteins were purified as described in [153], except that the wash and gradient buffers were without the addition of Triton X-100. Fractions containing EcoRII-N or EcoRII-Y41A proteins were validated by SDS PAGE, pooled, dialyzed against the corresponding Storage buffer I (see, sections 3.5.1. or 3.5.2.) and stored at -20°C. Protein concentrations were determined by measuring the absorbance at 280 nm, using the extinction coefficient of 45630 $M^{-1}cm^{-1}$ for a EcoRII-Y41A and 22943 $M^{-1}cm^{-1}$ for EcoRII-N.

3.7. Preparation and crystallization of EcoRII-N–DNA complex

All attempts to obtain X-ray diffracting crystals using $(His)_6$-tagged version of EcoRII-N bound to cognate oligoduplexes were unsuccessful. In addition to the $(His)_6$-tag at the N-terminus the expressed EcoRII-N fragment has seven residues of the flanking sequence at the C-terminus because of a cloning artifact [97]. We hypothesized that the existed N- and C-terminal tags could

hamper crystal packing. Therefore, we decided to generate a minimal EcoRII-N–DNA complex by limited proteolysis. Hence, 5 mg of $(His)_6$-tagged EcoRII-N were mixed with oligoduplex SP9 (Table 5) at the protein:DNA molar ratio 1:1.1 and digested for 4 hours with trypsin at the EcoRII:trypsin w/w ratio 75:1 in the Proteolysis buffer I (see, section 3.5.1.) at 37°C. Completeness of cleavage and homogeneity of protein were evaluated by SDS PAGE. The reaction was stopped by adding PMSF to the final concentration 0.1 mM. The proteolysis mixture containing the minimal EcoRII-N-DNA complex was diluted two-fold to adjust the concentration of KCl to 0.1 M and loaded onto a Heparin Sepharose column. The EcoRII-N–DNA complex was collected as a flow-through fraction in the Elution buffer (see, section 3.5.1.), then loaded onto Mono Q 5/50 GL column and eluted in a linear gradient at ~0.3÷4 M KCl. EcoRII-N–DNA complex was concentrated in the Concentration buffer (see, section 3.5.1.) to 18.5 mg/ml by ultrafiltration. Protein concentration was determined with the Bradford reagent. Crystals were grown using the sitting drop vapor diffusion technique at 18-20°C: 2.0 µl of the EcoRII-N–DNA solution were mixed with 0.5 µl of the Reservoir solution (see, section 3.5.1.). The EcoRII-N–DNA crystal appeared after 3 days from a clear drop.

3.8. Preparation and crystallization of EcoRII-C–DNA complex

The EcoRII-C terminal domain was obtained from the EcoRII-Y41A protein using limited proteolysis with thermolysin in the Proteolysis buffer (see, section 3.5.2.) according to the purification protocol described in [154]. Homogeneity of the enzyme preparation was evaluated by an SDS PAGE gel. For crystallization the protein was dialyzed against the Storage buffer II (see, section 3.5.2.) and an aliquot for biochemical experiments – against the Storage buffer III (see, section 3.5.2.), and stored at -20°C. Equimolar amounts of the EcoRII-C dimer and the 12 bp oligoduplex SP12 (Table 5) were mixed in the Concentration buffer (see, section 3.5.2.). The EcoRII-C–DNA complex was then concentrated to 4.6 mg/ml by ultrafiltration. Protein concentration

was determined with the Bradford reagent. Crystals of the complex were grown by the sitting drop vapor diffusion technique: 0.5 µl of the EcoRII-C–DNA solution was mixed with 0.2 µl of the Reservoir solution (see, section 3.5.2.). EcoRII-C–DNA crystals appeared after ~1 month from a clear drop at 18-20°C.

3.9. Data collection from EcoRII-N–DNA and EcoRII-C–DNA crystals

Crystals were transported to the DESY synchrotron (Hamburg, Germany) in crystallization plates and measured at EMBL outstation X13 beamline. The crystals of EcoRII-C–DNA were flash-cryocooled without any extra cryoprotection. Crystal belonged to the spacegroup $P2_12_12$ and diffracted to 2.6 Å. Statistics of the data collection are provided in Table 6.

Table 6. Data collection and refinement statistics of EcoRII domains

Dataset:	**EcoRII-N/SP9**	**EcoRII-C/SP12**
PDB ID	3HQF	3HQG
Temperature	100 K	100 K
Spacegroup	$P4_32_12$	$P2_12_12$
Parameters of the unit cell, Å	a=43.2, b=43.2, c=253.6, α=β=γ=90.0°	a=77.1, b=58.0, c=61.0, α=β=γ=90.0°
Resolution, Å (final shell)	63.5-2.5 (2.64-2.50)	47.84-2.6 (2.74-2.60)
Reflections unique (total)	9055 (62633)	8835 (119326)
Completeness (%) overall (final shell)	99.1 (94.3)	99.8 (99.8)
I/σ_I overall (final shell)	6.9 (7.0)	13.0 (2.0)
R_{merge}[2] overall (final shell)	0.07 (0.08)	0.05 (0.33)
B(iso) from Wilson (Å^2)	38	71
Refinement:		
Number of protein atoms	1334	1812
Number of DNA atoms	363	486

2 $R_{merge} = \Sigma_{\boldsymbol{h}} \Sigma_{i=1}^{n_h} \left| \langle I_{\boldsymbol{h}} \rangle - I_{\boldsymbol{h}i} \right| / \Sigma_{\boldsymbol{h}} \Sigma_{i=1}^{n_h} \left| I_{\boldsymbol{h}i} \right|$ where $I_{\boldsymbol{h}i}$ is an intensity value of *i*th measurement of reflection $\boldsymbol{h}$, $\boldsymbol{h}=(h,k,l)$, sum $\Sigma_{\boldsymbol{h}}$ runs over all measured reflections, and $\langle I_{\boldsymbol{h}} \rangle$ is an average measured intensity of the reflection $\boldsymbol{h}$. Number $n_{\boldsymbol{h}}$ is a number of measurements of reflection $\boldsymbol{h}$.

Dataset:	EcoRII-N/SP9	EcoRII-C/SP12
Number of solvent molecules	153	67
R_{cryst} (R_{free}), test set size 10%	0.187(0.229)	0.236 (0.293)
RMS bonds/angles	0.016/1.916	0.006/1.012
Average B-factors (Å^2)	15.0	45.7
main chain	14.3	46.3
side chains	16.9	48.0
DNA	13.3	39.6
solvent	12.6	39.7

Prior to flash-cryocooling EcoRII-N–DNA crystal was transferred into the 18.5:3.5 mixture of the Reservoir solution (see, section 3.5.1.) and PEG400. The crystal belonged to the spacegroup $P4_32_12$ and images with X-ray diffraction data were collected to 1.8 Å, 2.5 Å and 3.5 Å resolution limits, correspondingly. Unfortunately, due to the overlapping of spots, diffraction data from images at 1.8 Å were not processed. All data were processed with the programs MOSFLM [155], SCALA [156] and TRUNCATE [157]. The working datasets for EcoRII-N and EcoRII-C were to 2.5 Å and to 2.6 Å resolution correspondingly (Table 6).

3.10. EcoRII-N–DNA and EcoRII-C–DNA structure determination

The structures of the EcoRII-N–DNA and EcoRII-C–DNA complexes were solved by the molecular replacement method using the apo-EcoRII model PDB ID 1NA6 residues 4-170 and 284-402 of chain B, respectively. Molecular replacement was done with programs AMORE [158] or MOLREP [159]. Initial phases and electron density maps were calculated with SIGMAA [160] program using molecular replacement solutions followed by manual fitting using the program COOT [161]. Model refinement was carried out with CNS [162] and REFMAC [163]. Refinement proceeded in several cycles in combination with manual rebuilding. In the EcoRII-N complex 173 a.a. of the protein and DNA 9-mer were fully modeled into the electron density of the asymmetric unit.

In the asymmetric unit of the EcoRII-C–DNA complex 222 a.a. of a protein chain were traced in the density along with the 12-mer oligonucleotide (Table 6), which was modeled as two alternative strands with 50% occupancy each. During refinement with CNS [162] constrains were applied on DNA to hold together positions of overlapping backbone atoms from alternative chains.

3.11. Purification and crystallization of BfiI domains with DNA

The catalytically inactive BfiI-K107A mutant was expressed and purified as described [22]. Fractions containing BfiI-K107A were validated by SDS PAGE, pooled, dialyzed against the corresponding Storage buffer (see, section 3.5.3.) and stored at -20°C. Limited proteolysis in the presence of the cognate oligoduplex was used for the preparation of DNA-bound BfiI-C. Since BfiI binds two recognition sites, 10 mg of the BfiI-K107A mutant protein (250 nmoles) was mixed with oligoduplex S12 (Table 5) at the molar protein:DNA ratio 1:2.2 for a typical reaction. DNA complexed BfiI-K107A was digested with thermolysin at ratio 10:1 (w/w) in the Proteolysis buffer (see, section 3.5.3.) at 37°C for 20 hours. The reaction was stopped with PMSF as described [25]. The proteolysis mixture was diluted two-fold with the proteolysis buffer without KCl and loaded onto a Heparin Sepharose column. The BfiI-N-K107A domain binds to the column and was eluted in linear gradient of KCl at ~0.2 M. The BfiI-C–DNA complex (with the excess of free DNA) was collected as a flow through fraction and was loaded onto a Mono Q 5/50 GL column. The BfiI-C–DNA complex was separated from an excess DNA and eluted in a linear KCl gradient at ~0.5 M. All purification steps were performed at room temperature. The collected fractions of the BfiI-N-K107A and BfiI-C–DNA complexes were dialyzed against the Concentration buffer (see, section 3.5.3.).

The BfiI-C–DNA complex was concentrated to 1.7 mg/ml by ultrafiltration in the Concentration buffer (see, section 3.5.3.). Concentration of the complex was determined with the Bradford reagent. Crystals were grown using the sitting drop technique at 18-20°C by mixing 0.5 µl of the complex solution with 0.5 µl of the Reservoir solution I (see, section 3.5.3.). Measurable crystals

of BfiI-C–DNA were found after ~3 years since preparation and belonged to the spacegroup $P6_5$. Prior to flash-cryocooling, crystals were transferred into the 3:1 mixture of the Reservoir solution I and glycerol.

The BfiI-N-K107A domain was concentrated to 5 mg/ml by ultrafiltration in the Concentration buffer (see, section 3.5.3.) and then mixed with the DNA 8 bp duplex NS8. Crystals were grown using the sitting drop technique at 18-20°C by mixing 0.5 µl of the complex solution with 0.5 µl of the Reservoir solution II (see, section 3.5.3.). Crystals of BfiI-N-K107A appeared after ~1 week and belonged to the spacegroup I4. Prior to flash-cryocooling, crystals were transferred into the 3:1 mixture of the Reservoir solution II and glicerol.

3.12. Data collection and structure determination of BfiI domains

An X-ray diffraction dataset from the BfiI-C–DNA crystal was collected to 3.1 Å on the Rigaku RU-H3R rotating anode generator at the *Vilnius University Institute of Biotechnology, Vilnius, Lithuania*. A dataset from the BfiI-N-K107A crystal was collected to the 2.4 Å resolution at the analogous machine at the *International Institute of Molecular and Cell Biology, Warsaw, Poland, in the laboratory of dr. Matthias Bochtler*. MOSFLM [155], SCALA [156] and TRUNCATE [157] were used for data processing. Data collection statistics are presented in Table 7.

Table 7. Data collection and refinement statistics of BfiI domains

Dataset:	**BfiI-C/S12**	**BfiI-N-K107A/NS8**
Temperature	100 K	100 K
Spacegroup	$P6_5$	I4
Parameters of the unit cell, Å	a=175.18, b=175.18, c=35.79, α=γ=90.0°, β=120°	a=163.44, b=163.49, c=68.9, α=β=γ=90.0°
Resolution, Å (final shell)	50.57-3.2 (3.37-3.2)	40.96-2.44 (2.57-2.44)
Reflections unique (total)	10809 (75356)	32775 (87976)
Completeness (%) overall (final shell)	100 (100)	95.7 (83.8)

Dataset:	BfiI-C/S12	BfiI-N-K107A/NS8
I/σ_I overall (final shell)	13.3 (7.2)	7.9 (3.4)
R_{merge} overall (final shell)	0.147 (0.240)	0.117 (0.254)
B(iso) from Wilson, (Å^2)	20.9	17.9
Refinement:		
Number of protein atoms	2642	5854
Number of DNA atoms	972	-
R_{cryst} (R_{free}), test set size 10%	0.18 (0.22)	0.20 (0.24)
RMS bonds/angles	0.008/0.999	0.002/0.548
Average B factors (Å^2), total	32.0	17.5
main chain:	30.1	16.5
side chains:	33.1	18.5
DNA:	33.1	-

Initial phases the for BfiI-C–DNA and BfiI-N-K107A structures were obtained by the molecular replacement method using the apo-BfiI model PDB ID 2C1L. In the case of BfiI-C the residues 200-347 of the chain A were supplied for a multicopy search in MOLREP; in the case of BfiI-N-K107A a multicopy search was done using the residues 1-180 of chain A. Both structures were refined using REFMAC [163] with NCS restrains. After this, the complex was additionally refined with PHENIX [164] applying TLS restrains [165], and COOT [161] was used for model inspection and rebuilding. The refinement statistics are presented in Table 7. The refined coordinates and the structure factors are to be deposited to the RCSB database.

3.13. Purification and crystallization of Bse634I with DNA

The Bse634I wt protein as well as the R226A mutant [166] were expressed and purified as described [11]. Concentration of the protein monomers was estimated spectrophotometrically using the extinction coefficient 34280 $M^{-1}cm^{-1}$. The Bse634I-R226A–DNA complex was crystallized using the sitting drop vapor diffusion technique. The protein in the Crystallization buffer (see, section 3.5.4.) was mixed with an equimolar amount of the GC-1 or AT-1 oligoduplexes (Table 5) and an equal volume of the Reservoir solution I (see,

section 3.5.4.) was added. The Bse634I-R226A complex with the AT-2 DNA (Table 5) was crystallized by adding the Reservoir solution II (see, section 3.5.3.). The final concentration of the protein was in the range 4 to 8 mg/ml. Crystals grew in a week, and after several months degraded severely.

3.14. Data collection and structure solution of Bse634I-DNA

Crystallographic data for R226A-GC-1 were collected at the EMBL Hamburg outstation DESY X12 beamline, and datasets from crystals with AT-1 and AT-2 oligoduplexes were collected at the beamline X11. MOSFLM [155], SCALA [156] and TRUNCATE [157] were used for data processing. Initial phases were obtained by molecular replacement using MOLREP [159] and the Bse634I apo-form residues 95-293 from chain A [PDB ID 1KNV] as an initial model. The model was rebuilt and refined using COOT [161], CNS [162] and REFMAC [163] programs. The crystal structure of the Bse634I complex with AT-2 was refined using NCS restraints between all eight subunits present in the asymmetric unit. Data collection and refinement statistics are shown in Table 8.

Table 8. Data collection and refinement statistics of DNA-bound Bse634I-R226A

Dataset:	**Bse634I-R226A/ GC-1**	**Bse634I-R226A/ AT-1**	**Bse634I-R226A/ AT-2**
PDB ID	3V1Z	3V20	3V21
spacegroup	$P2_12_12$	$P2_12_12$	$P2_1$
Parameters of the unit cell, Å	a=72.01, b=83.63, c=123.34, α=β=γ=90°	a=71.94, b=87.39, c=125.13, α=β=γ=90°	a=106.88, b=115.31, c=130.24, α=γ=90°, β=112.50°
Resolution, Å	2.3	2.34	2.7
Reflections unique (total)	33372 (125217)	30943 (143192)	80299 (366610)
Completeness overall (outer shell), %	99.1 (95.2)	90.6 (64.5)	100 (100)
I/σ overall (outer shell)	9.1 (2.5)	22.8 (3.0)	20.3 (3.1)
R_{merge} overall (outer shell)	0.067 (0.307)	0.054 (0.332)	0.060 (0.388)

Dataset:	Bse634I-R226A/ GC-1	Bse634I-R226A/ AT-1	Bse634I-R226A/ AT-2
B(iso) from Wilson, (Å^2)	39.4	47.2	67.0
Refinement:			
Number of atoms*	4696	5344	22664
Number of solvent molecules	218	106	359
R_{cryst} (R_{free}), test set size 10%	0.23 (0.28)	0.24 (0.29)	0.23 (0.27)
RMS bonds/angles	0.007 (1.009)	0.009 (1.186)	0.01 (3.820)
Average B-factors all, Å^2	43.0	42.6	40.6
solvent	30.8	38.4	40.9
DNA	40.4	38.6	78.6

* alternative positions of atoms are reported as a single atom.

3.15. DNA-protein binding assay

Binding of the modified and unmodified DNA oligoduplexes by REases or REase domains was studied by the gel mobility shift assay. Samples containing increasing amounts of protein were mixed with ^{33}P-radiolabeled DNA in the reaction buffer with one or more supplements (Table 9), left for 20-30 min at room temperature and loaded onto the non-denaturing polyacrylamide gel. The gels were made of 6-8% acrylamide/N,N'-methylenebisacrylamide (29:1 (w/w)) dissolved in the electrophoresis buffer I (3.5) with 0.1 mM of EDTA or 5 mM of $Ca(OAc)_2$, polymerization was initiated by adding into mixture TEMED and ammonium persulphate. Electrophoresis was run at the room temperature at ~6 V/cm. Gels were dried and signal from radiolabeled DNA fragments was detected by exposing to BAS-MS image plates ("FujiFilm", Kanagava, Japan) or Cyclone Phosphor-Imager screens ("Perkin-Elmer", Wellesley, MA, USA).

Table 9. Conditions of performed DNA-binding experiments

Protein	Concentration of DNA, nM	Reaction buffer with 0.1 mg/ml of BSA	Supplement, mM			run time
			EDTA	Ca^{2+}	DTT	
EcoRII-N	2	40 mM Tris-HCl (pH 8.3 at 25ºC), 10% (v/v) glycerol	0.1	-	-	~2 hrs
EcoRII-C	0.1		-	5	2	
Bse634I	0.1		-	5	-	~3 hrs

3.16. DNA cleavage assay

DNA cleavage reactions were conducted at 25ºC or 50ºC (in the case of wt Bse634I) by mixing the radiolabeled oligoduplexes with the enzyme in the reaction buffer (Table 10). The reactions performed at 50ºC, which required incubation time >10 min, were carried out under mineral oil. Aliquots were removed at timed intervals and quenched by mixing with the Loading dye solution (see, section 3.5.). DNA hydrolysis products were separated by the denaturing PAGE: the 20 % polyacrylamide gel (acrylamide/N,N'–methylenebisacrylamide 19:1 (w/w)) in the Electrophoresis buffer II (see, section 3.5.) containing 8.0 M of urea was run at 30 V/cm for 2-3 hours [167]. Gels were dried and analyzed as described in section 3.15. Quantitative analysis of the DNA fragment distribution was performed with the OptiQuant 3.0 software "Perkin Elmer". When analyzing cleavage of the methylated oligoduplexes by EcoRII-C products of the oligoduplex NM cleavage (Table 5) with the REase MvaI were used as length markers.

Table 10. Conditions of performed DNA cleavage experiments

Enzyme	Concentration:		Reaction buffer with 0.1 mg/ml BSA	Supplement	
	of protein, nM	of DNA, nM		Mg^{2+}	DTT
EcoRII-C	1000 (dimer)	200	33 mM Tris-OAc (pH 7.9 at 25°C), 66 mM KOAc	10 mM	2 mM
Bse634I	100 (tetramer)	100	10 mM Tris-HCl (pH 8.5 at 37ºC), 100 mM KCl	10 mM	-

The cleavage of each DNA strand in all four oligoduplexes was analyzed using a non-linear regression. The rate constants of the DNA cleavage were determined by fitting the time-courses of the depletion of each substrate strand to a single exponent model with Gnuplot [168] or KyPlot [169].

3.17. Software used in research

3.17.1. Protein sequence analysis

Theoretical parameters such as pI, molecular weight and extinction coefficient for EcoRII, BfiI, Bse634I and their domains were calculated by the ProtParam [170]. Sequences of EcoRII or BfiI homologs were retrieved from public databases with BLAST [171] or obtained from REBASE database [3] picking either confirmed REases or putative ORFs that are accompanied by methyltransferase genes; alignments were produced by CLUSTALW [172] available on-line http://embnet.vital-it.ch/software/ClustalW.html using scoring matrix PAM; aligned sequences were rendered with ESPript [173].

3.17.2. Crystallographic and structural data analysis

Collected X-ray diffraction datasets were analyzed at crystal twinnig server at UCLA [174]. The searching of characteristic DNA signal in all datasets was performed by DIBER [175]. The quality of the atomic models was verified with the PROCHECK [176], WHAT IF [177] or MolProbity [178] programs. The DNA parameters values were calculated with CURVES [179]. Manually assigned protein–DNA contacts were compared to contacts calculated by NUCPLOT [180]. The solvent accessible surface areas were calculated with NACCESS [181]. NQ-Flipper [182] was used for checking rotamers assignment of both asparagine and glutamine. Quality of modeled loops was validated with ProSA [183]. Searching for protein structural similarities was performed by DALI [184]. The fragments of DNA or proteins were superimposed with local scripts (Saulius Gražulis, to be published). Intermediate structures between two protein or DNA conformations were calculated with Morph server [185]. Atomic coordinates of B-form DNA were generated by 3D-DART [186] or NAB [187]. Homologous structural models were constructed with an automated comparative protein modeling server SWISS-MODEL [188]. Graphical representations of atomic coordinates were performed either with PyMol [189] or MOLSCRIPT [190] and Raster3D [191].

4. RESULTS AND DISCUSSION

4.1. Analysis of EcoRII-C domain interactions with DNA

4.1.1. DNA oligoduplex is bound in two overlapping orientations

The EcoRII-C domain obtained by limited proteolysis was co-crystallized (see, section 3.8.) with the 12 bp non-palindromic oligoduplex SP12 (Table 5) containing the EcoRII recognition sequence 5'-CCWGG. The EcoRII-C structure was solved by molecular replacement using the C-terminal domain of EcoRII as a model (see, section 3.10.) and subsequently refined to 2.6 Å resolution against the dataset processed assuming $P2_12_12$ symmetry (Table 5). The asymmetric unit contains one protein monomer and one DNA strand. In the $P2_12_12$ spacegroup the protein chain is mapped by a twofold axis onto the other protein chain of the EcoRII-C dimer. The same twofold axis maps the DNA strand onto the complementary DNA strand of the duplex bound by the protein dimer. Since the DNA oligonucleotide used in the crystallization experiment was non-palindromic, however, the bound DNA duplex does not posses a strict twofold symmetry and therefore should not occupy the special position on the dyad axis. The apparent contradiction can be explained by assuming that the DNA duplex is disordered around the twofold axis of the protein dimer and is present in the crystal in two opposite orientations with 50:50% occupancy. Thus, DNA was modeled as two alternative DNA strands both having the same backbone position in the model. Such model predicts that the *average* electron density of the DNA can be described by a twofold symmetry present in the $P2_12_12$ spacegroup and in this way explain the observed symmetry of the crystal.

Another alternative explanation of the crystal packing would be that the crystal in fact has only the $P2_1$ symmetry, and the DNA duplex has a preferred orientation in the crystal. Such model would predict that the model with one DNA orientation has better fit to experimental data than the model with the opposite DNA orientation. When the data were processed and the structure

refined in the $P2_1$ symmetry, however, both DNA duplex orientations with respect to the polar $P2_1$ crystal axes yielded comparable R-factors and electron densities, confirming the equivalence of both DNA strand orientations as predicted by the $P2_12_12$ symmetry. Therefore, there was not reason to prefer the lower symmetry group, and all further work was performed with the higher symmetry $P2_12_12$ model.

4.1.2. Overall structure of the EcoRII-C complex with DNA

The secondary structure elements are identical in the apo-EcoRII [14] and in the liganded EcoRII-C. Two EcoRII-C monomers are arranged into a clamp-like structure and completely encircle the bound DNA (Figure 23). The subunit interface of the EcoRII-C dimer is bipartite and buries 2400 $Å^2$ per monomer.

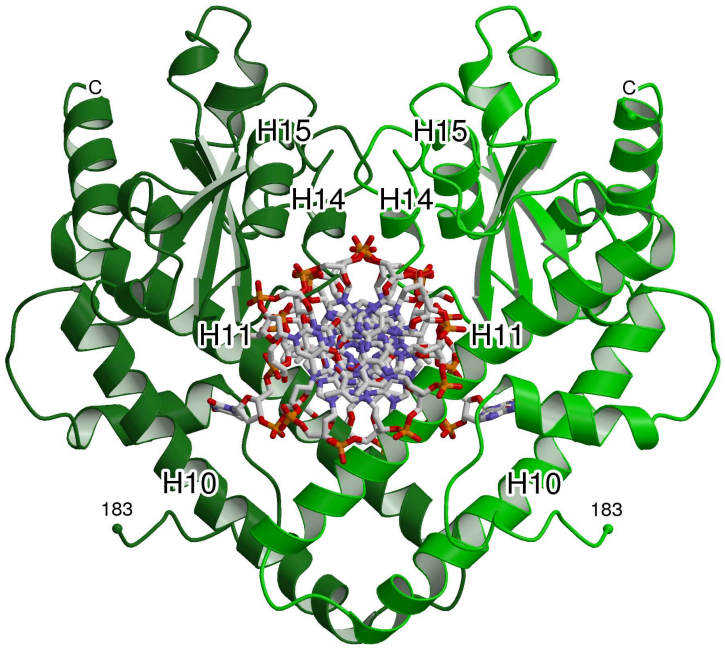


Figure 23. Crystal structure of EcoRII-C in complex with cognate 12 bp oligoduplex; view in the direction of long DNA axis. DNA is shown as a stick model in CPK. Different protein monomers colored in dark and light green, respectively.

In the N-terminal part of the dimer interface two long α-helices H10 and H11 interact (Figure 23). The N-terminal part of the interface is predominantly hydrophobic and buries 1600 $Å^2$ of the accessible surface area. The C-terminal part of the dimer interface is made by α-helices H14 and H15 located at the conserved catalytic core (Figure 23). The C-terminal interdomain contacts are dominated by hydrogen bonds and bury 840 $Å^2$ at the interface. Such dimerisation mode is similar to the dimerisation mode observed in the Type IIP

REase PspGI [13] and the Type IIF REase Ecl18kI [16]. The intersubunit contacts create a large bipartite DNA binding surface which occludes the DNA and buries nearly 4900 Å^2 per dimer at the EcoRII-C–DNA interface.

4.1.3. EcoRII-C flips out the central T:A base pair within the CC(T/A)GG site

The DNA in the complex with EcoRII-C has an unusual conformation. Bases of the central T:A base are extruded from the DNA double helix (Figure 23) and the phosphodiester backbone is compressed so that the helical 5 bp recognition site is transformed into the 4 bp helix (Figure 24). Similar DNA distortion is observed in the related REases Ecl18kI [16] and PspGI [13]. The conserved CC:GG dinucleotides within the 5'-CCNGG (Ecl18kI) and 5'-CCWGG (EcoRII-C/PspGI) recognition sites superimpose well (Figure 24). In contrary extrahelical nucleotides in EcoRII-C–DNA complex are overwound towards the 5'-end of the corresponding DNA chain in comparison to Ecl18kI–DNA complex but nearly coincide with those of PspGI–DNA (Figure 24). It could be suggested that these differences in the extrahelical base conformation are probably due to the differences in the central base pair specificity: EcoRII-C and PspGI accept only A or T nucleotides while Ecl18kI shows no specificity for the central nucleotide.

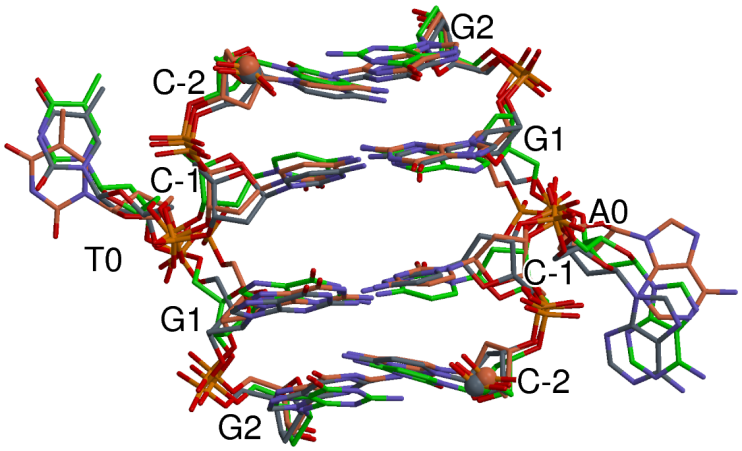


Figure 24. Stick representation of the target 5'-CCTGG sequence in crystal structures with EcoRII-C (green), Ecl18kI [PDB ID 2FQZ, coral] and PspGI [PDB ID 3BM3, gray] complexes. Scissile phosphates atoms are shown as spheres.

The nucleotide flips seen in EcoRII-C are in an excellent agreement with the solution studies, which show the ~12-fold fluorescence enhancement of the 2-aminopurine placed at the center of the recognition sequence [192]. The

compressed DNA conformation is presumably stabilized by interactions of the extruded backbone phosphates with the Arg262 residue which is absolutely conserved within the EcoRII family (Figure 39). Indeed, the replacement of the structurally equivalent Arg116 (Ecl18kI) and Arg96 (PspGI) residues by alanine completely abolished DNA binding and cleavage [193].

4.1.4. Putative mechanism of T:A recognition by EcoRII-C

Unstacked bases are accommodated into the EcoRII-C pocket that is formed by the residues Arg222, Glu225, Tyr226, Phe229 located on the α-helix H10 and the Arg265, Ala266 residues (Figure 25A) located on the α-helix H11 of the same monomer (Figure 23). Structural comparison indicates that in apo-EcoRII the binding pocket for the flipped base is already preformed and binding of the extrahelical base is accompanied by only minor changes in the conformations of the pocket residues. The EcoRII-C, PspGI and Ecl18kI pockets for the flipped nucleotides are strikingly similar (Figure 25) raising a question how EcoRII-C discriminates the 5'-CC(T/A)GG targets from the 5'-CC(C/G)GG sites. In EcoRII-C the extruded A0 or T0 bases are sandwiched between the aromatic ring of the Tyr226 residue and the side chain of the Arg222 residue (Figure 25). The C5 methyl group of the T0 base is positioned at the van der Waals contact distance (3.7 Å) to the aromatic ring of Phe229 within the binding pocket (Figure 25C). The N6 atom of the flipped out base A0 comes close (3.1 Å) to the Oε1 oxygen atom of Glu225, however, the geometry for the hydrogen bond is far from the optimal (Figure 25A). In theory, these interactions could allow to discriminate T:A base pair vs C:G in the EcoRII-C binding pockets. Indeed, replacement of cytosine for thymine should eliminate the favorable van der Waals interactions between the thymine's methyl group and the Phe229 aromatic ring while replacement of guanine for adenine should perturb hydrogen-bonding interactions with the side chain carboxylate oxygen of Glu225 (Figure 25A). Also, the guanine base would fall into a steric clash with the aromatic ring of Phe229 (atom N2 distance to atom Cζ 2.9 Å).

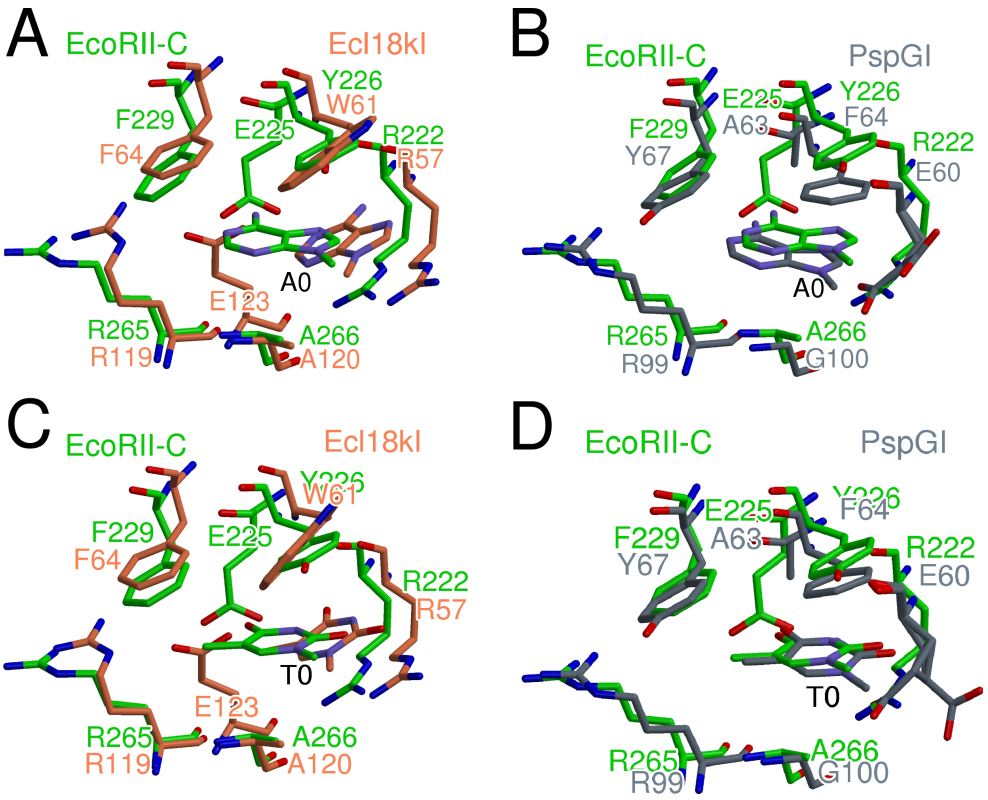


Figure 25. Superposition of binding pockets for the flipped A/T nucleotides (AC) between EcoRII-C (green) and Ecl18kI [PDB ID 2FQZ, coral] or (BD) between EcoRII-C (green) and PspGI [PDB ID 3BM3, gray].

This observation is in line with the results of the DNA binding and cleavage experiments [167] that showed that EcoRII-C cuts duplexes with symmetrical mismatches A/A or T/T pair with the same rate as the cognate A/T base pair, whereas the G/G or C/C bases at the central positions caused > 1000-fold decrease in the cleavage rates. According to the "double-check" mechanism for substrate cleavage by both EcoRII-C and PspGI [167] the strength of the Watson-Crick base pair interaction between bases in the central base pair as well as the direct check of the identity of the flipped bases in the pocket are critical for the stable binding of the protein with cognate DNA.

The functional importance of Glu225 and Phe229 residues is supported by nearly absolute conservation among the 13 protein sequences of EcoRII homologues (Figure 39). Surprisingly, in Ecl18kI, which shows no base preference in the flipped base binding pockets, the structural equivalents of conserved Glu225 and Phe229 residues are Glu123 and Phe64, respectively (Figure 25A). Moreover, in PspGI, which like EcoRII-C is specific for the

5'-CC(T/A)GG sequence, Tyr67 is structurally equivalent to Phe229 but Ala63 replaces the Glu225 (Figure 25B). Further crystallographic and mutational studies are required to determine the mechanism of T:A vs C:G base pair discrimination in the binding pockets of EcoRII-C and PspGI.

4.1.5. EcoRII-C interactions with the symmetrical CC:GG half-sites

EcoRII-C achieves the recognition of the symmetric CC:GG half-sites through the direct contacts of the consecutive Lys328, Asp329 and Arg330 residues located at the N-terminus of the symmetry-related α-helices H14 (Figure 23) which are dubbed "recognition helices" in analogy with other CC:GG recognizing REases [112]. Each monomer makes an identical set of contacts to the CC:GG half of the 5'-CC(T/A)GG sequence. Therefore, only one set of contacts to the CC:GG subsite will be discussed. The Oδ1 and Oδ2 carboxylate oxygens of Asp329 accept hydrogen bonds from N4 atoms of two neighboring cytosines (Figure 26). The amino group of the Lys328 donates a hydrogen bond to the O6 atom of the outer guanine (Figure 26A), while Arg330 makes bidentate hydrogen bond interactions with the O6 and N7 atoms of the inner guanine (Figure 26B). The main chain oxygen and the Oδ1 oxygen atoms of the Asn260 residue located on the α-helix H11 of the neighboring monomer are within hydrogen bond distance to the N2 atoms of the inner and the outer guanine residues in the minor groove, completing the interaction network with the CC:GG half-site (Figure 26).

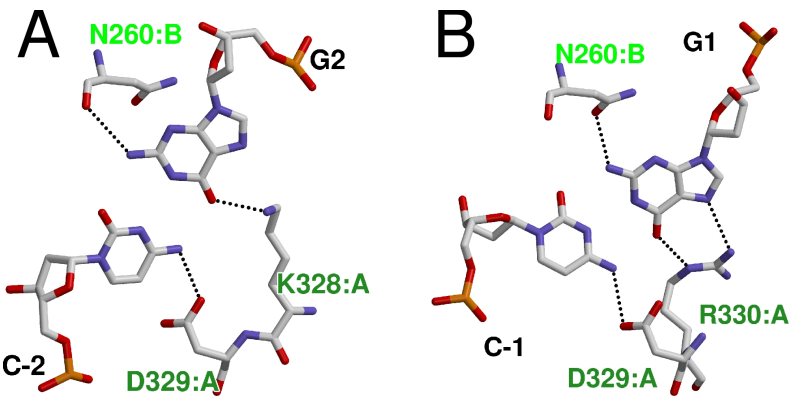


Figure 26. Interactions of EcoRII-C with the CC:GG half-site of the 5'-CC(A/T)GG sequence. DNA and protein residues are colored in CPK.

The structural model of sequence recognition is in agreement with the mutational studies of the Lys328, Asp329 and Arg330 residues [71]. Not

surprisingly, these residues are absolutely conserved in the EcoRII family (Figure 39). Thus, the structural study confirms that EcoRII-C belongs to the restriction enzyme family (Table 1) that recognizes CC:GG half-sites in different contexts through the common structurally conserved amino acids motif [13][16].

4.1.6. Active site of EcoRII-C

The active site of EcoRII-C–DNA belongs to the PD-(D/E)XK superfamily. The two active centers are formed by the Glu271, Asp299, Lys324 and Glu337 residues from each monomer and are located in a close proximity to the scissile phosphates of both DNA strands (Figure 27). The mutational and biochemical analysis is consistent with the active site function of these residues [71]. Moreover, the EcoRII-C active site residues are absolutely conserved across the EcoRII family, with only one exception in a protein from *Maricaulis maris* (Figure 39). Furthermore, a superposition of EcoRII-C and Ecl18kI in the DNA bound form shows that the active sites residues (Figure 27) as well as the scissile phosphate bonds of the two complexes (Figures 24 and 27) almost coincide.

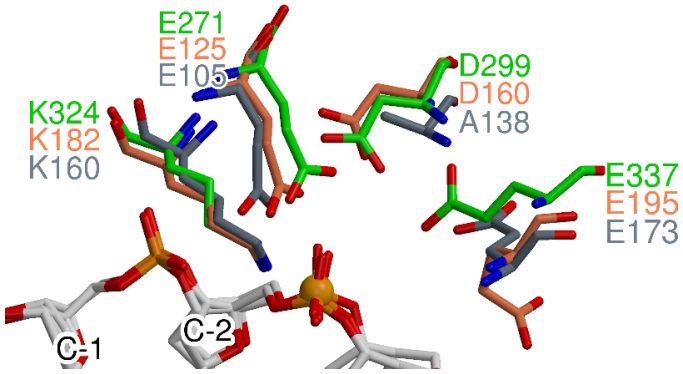


Figure 27. Superposition of DNA bound EcoRII-C active site (green) and active sites of Ecl18kI [PDB ID 2FQZ, coral] and PspGI-D138A [PDB ID 3BM3, gray]. DNA is colored in CPK, phosphorous atom of scissile phosphate is shown as orange sphere.

Despite the presence of Ca^{2+} in the crystallization buffer there is no electron density that could be interpreted as a Ca^{2+} ion in the vicinity of the active site of EcoRII-C. In the PspGI, the side chain conformation of the Glu173 differs from that of the structurally equivalent residues in EcoRII-C and Ecl18kI (Figure 27). The active site environment of PspGI, however, may

be perturbed due to the E138A mutation which was introduced to obtain diffracting crystals [13].

4.1.7. EcoRII-C sensitivity to DNA methylation

Previous studies [98] revealed that C5-methylation of the cytosine C-1 within the 5'-CCWGG sequence in both strands (dimethylated site) makes the M13 phage DNA resistant to the EcoRII-C cleavage. According to another study, wt EcoRII is still able to cleave target DNA that has C5-methylation in one strand (hemimethylated site) but only when activated by non-methylated duplexes [194]. Therefore, we decided to study EcoRII-C interaction with methylated DNA by gel shift method (see, section 3.15.). Our data show that EcoRII-C does not bind to the dimethylated (DM) oligoduplex containing 5-methylcytosine (5mC) in both strands, but still shows residual affinity to the oligoduplexes with hemimethylated 5'-CCWGG sites (HM(A) and HM(T)) (Figure 28). In the case of HM substrate shifted DNA bands in the gel are smeared, suggesting decreased complex stability.

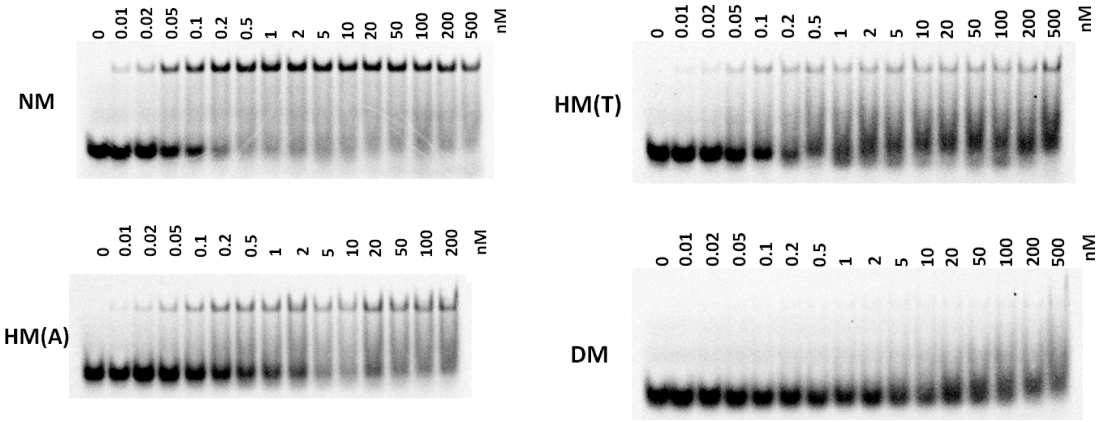


Figure 28. Gel mobility shift analysis of DNA binding by the EcoRII-C. The binding reactions (see, section 3.15.) contained either the non-methylated (NM), hemimethylated (HM) or dimethylated (DM) oligoduplex (Table 5). Protein concentration is indicated above each line. Oligoduplexes containing 5mC in A- or T-chain of the target site are designated as HM(A) and HM(T), respectively.

Oligoduplex cleavage studies (see, section 3.16.) show that methylation of both cytosines in the target site decreases EcoRII-C cleavage rate nearly by five orders of magnitude (Figure 29A). Hemimethylation of the 5'-CCWGG

site, however, affects cleavage rates of methylated and non-methylated strands differently. While the cleavage rate of methylated strand goes down by nearly four log orders, the non-methylated strand is cleaved only 100-200-fold slower than the unmethylated oligoduplex. *In silico* analysis shows that the introduction of 5mC should disrupt the recognition network at EcoRII-C interface between the key recognition residue Asp329 and cytosine C-2 (Figure 29B).

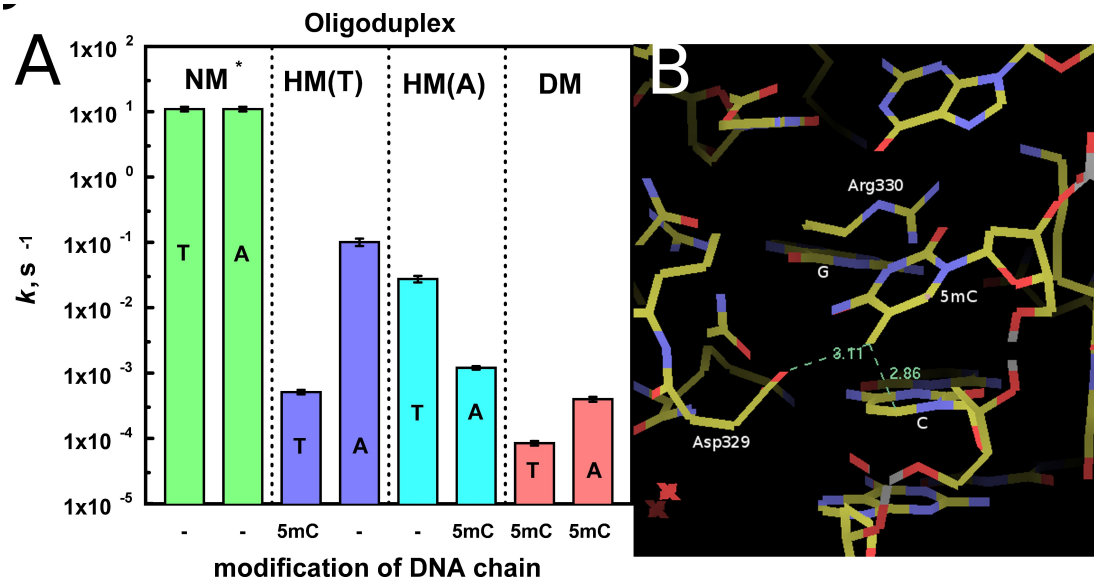


Figure 29. Functional and structural effects of hemimethylation on cleavage and computer simulation of methyl group position. (A) EcoRII-C cleavage of methylated and non-methylated DNA. Cleavage rates of oligoduplexes containing 5mC modification at the second C of 5'-*CC(A/T)GG by EcoRII-C were obtained under the single turnover conditions. Abbreviations are the same as in Figure 28. Rate constants are provided for the cleavage of each strand in the oligoduplex. Rate constants for the cleavage of the non-methylated duplex (NM)* are from [167]. (B) In silico modeling of the methylated C at the EcoRII-C target site. Both cytosines from one half-site are shown. Cytosine C-1 was in silico replaced by 5mC [161]. Dashed lines indicate distances between the C5 carbon atom of methyl group and neighboring atoms.*

In conclusion, the EcoRII-C cleavage rate differences between methylated and non-methylated strands suggest that in spite of the symmetric arrangement of EcoRII-C dimer, individual monomers act independently on separate DNA strands. Similar uncoupling of the monomer interactions due to the hemimethylation of the target site has been reported previously for the restriction enzyme EcoRI [195].

4.2. Analysis of EcoRII-N domain interactions with DNA

4.2.1. Overall structure of the EcoRII-N–DNA complex

EcoRII-N was crystallized (see, section 3.7.) with the 9 bp oligonucleotide duplex SP9 (Figure 30A) containing the 5'-CCWGG target sequence. The asymmetric unit contained one protein monomer bound to a single DNA copy (Figure 30B). The structure was solved by molecular replacement using the N-terminal domain of the EcoRII structure [PDB ID 1NA6] as a model (see, section 3.10.). The secondary structure of the DNA-bound EcoRII-N domain is identical to that of the B chain (residues 4-172) in the apo-EcoRII structure [14] and follows the same labeling. In the EcoRII-N–DNA complex structure ~2200 Å^2 of the protein surface is buried at this interface. The DNA in the EcoRII-N complex is not significantly distorted (Figure 30B) except that the major groove becomes ~3.0 Å wider in comparison to the canonical B-form.

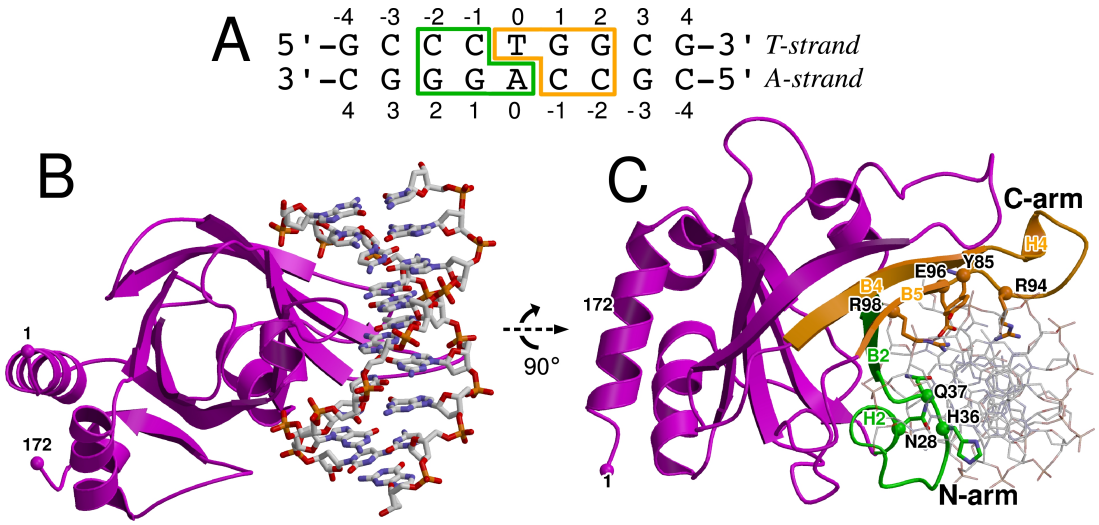


Figure 30. (A) Oligonucleotide used for crystallization of EcoRII-N. Pseudosymmetric parts of the target site are boxed in green (5'-end) and orange (3'-end) letters, respectively. (B, C) Two different views of EcoRII-N–DNA complex structure. In (C) the secondary structure elements of the N-arm that predominantly interact with CC:GG base pairs of the effector DNA are shown in green; those belonging to the C-arm and predominantly interacting with GG:CC are shown in orange. Residues and their Cα atoms are colored as corresponding arm color.

4.2.2. Sequence recognition by EcoRII-N

The EcoRII-N monomer interacts with the recognition sequence in a single defined orientation. The concave DNA binding cleft of the N-terminal domain

resembles a wrench with a skewed U-shaped opening that grips DNA from the major groove side (Figure 30B). The N-arm (residues 27-42) of the wrench is formed by an α-helix H2, a β-strand B2 and a connecting loop, while two antiparallel β-strands B4, B5 and a connecting loop comprise the C-arm (residues 78-101) (Figure 30C).

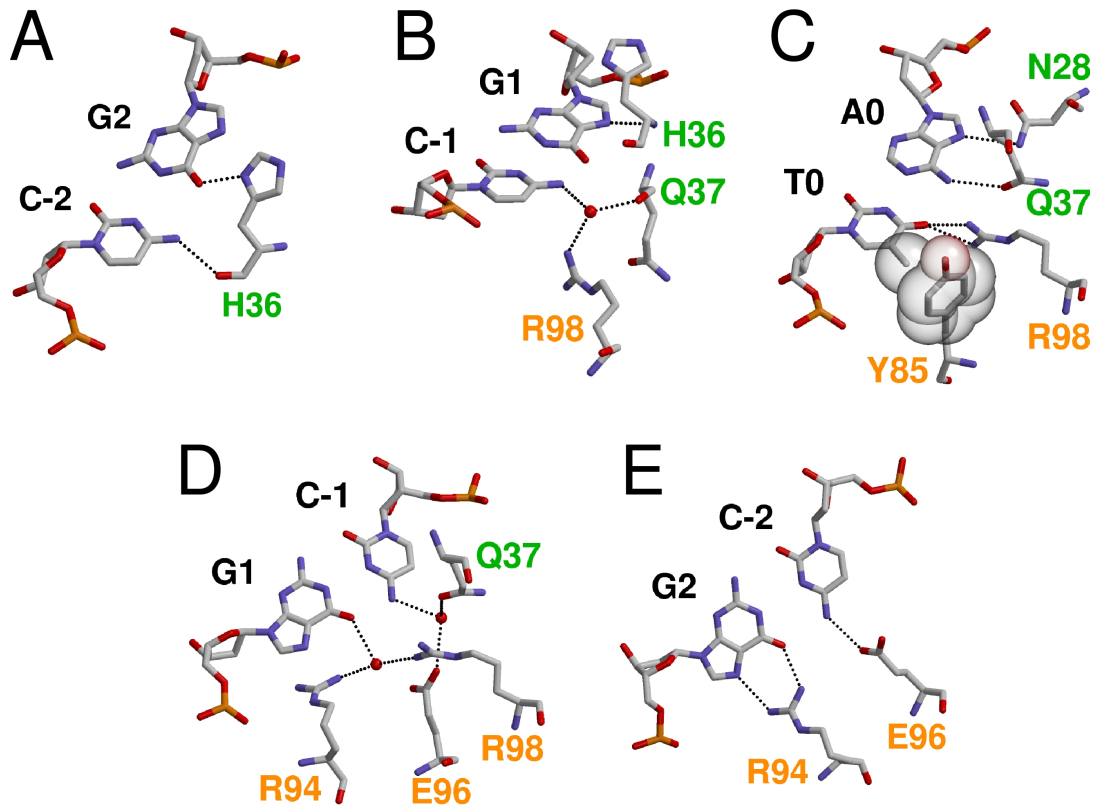


Figure 31. Recognition of the 5'-CCTGG sequence by EcoRII-N. (A-E) Panels for individual base pairs are arranged following the T-strand in 5'→3' direction. Residues belonging to the N- and C-arms are labeled in green and orange, respectively.

Both arms penetrate into the DNA major groove and the amino acid residues located on the arms interact with the base edges of the recognition sequence. The residues located on the N-arm interact predominantly with the "left" 5'-CC:GG half-site, while the residues in the C-terminal arm make contacts to the "right" 5'-CC:GG half-site (Figure 30C). Both arms contribute amino acid residues to the recognition of the central T:A base pair. The detailed interactions of the EcoRII-N residues in the DNA major groove are depicted in the Figure 31A-E as follows:

•*with the base pair C-2:G2*: The single His36 residue makes two direct hydrogen bonds to the donor and acceptor atoms of the C-2:G2 base pair. The main chain carbonyl oxygen atom of His36 accepts a hydrogen bond from the N4 atom of the cytosine, and the Nδ atom of its imidazole ring donates a hydrogen bond to the O6 atom of the guanine (Figure 31A).

•*C-1:G1*: A water molecule sandwiched by the guanidino group of Arg98 and the main chain carbonyl oxygen of Gln37 accepts a hydrogen bond from the N4 atom of the cytosine. The main chain nitrogen atom of His36 donates a hydrogen bond to the N7 atom of the guanine (Figure 31B).

•*T0:A0*: The N6 atom of the adenine donates a hydrogen bond to the Oε atom of Gln37 while N7 atom accepts a hydrogen bond from the Nδ atom of Asn28 (Figure 31C). The Nη1 and Nη2 atoms of Arg98 make hydrogen bonds to the thymine O4 atom. The methyl group of the thymine T0 makes a van der Waals contact with the aromatic ring of the Tyr85 residue.

•*G1:C-1*: The guanidino groups of Arg94 and Arg98 sandwich the water molecule that donates a hydrogen bond to the O6 atom of guanine. The carboxylate of Glu96 and the Oε atom of Gln37 bind the water molecule that accepts a hydrogen bond from the N4 atom of cytosine (Figure 31D)

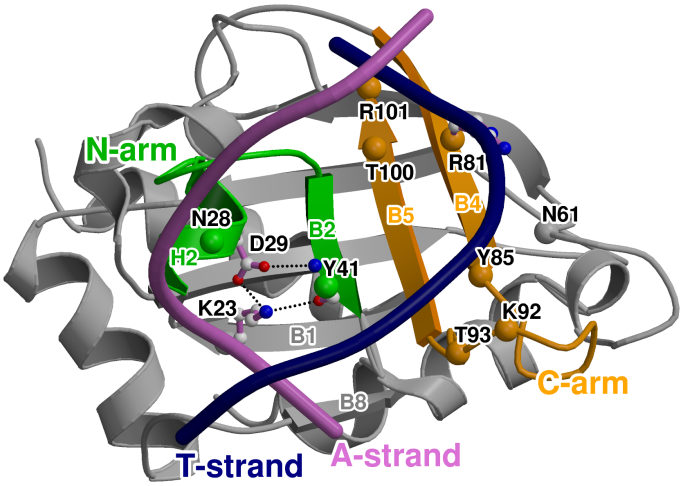


Figure 32. EcoRII-N contacts to the DNA phosphodiester backbone. Protein Cα atoms of residues that interact with the phosphodiester backbone of DNA are depicted as spheres. DNA backbone is depicted as helix.

•*G2:C-2*: The Arg94 nitrogen atoms Nη1 and Nη2 make hydrogen bonds to the N7 and O6 atoms of guanine. The carboxylate of Glu96 accepts a hydrogen bond from the cytosine N4 atom (Figure 31E).

Each arm of the wrench makes a set of contacts with the phosphodiester backbone of one particular DNA strand (Figure 32). Asn28 and Tyr41 from the N-arm interact with the DNA phosphates of the A-strand, while Arg81, Tyr85, Lys92, Thr93, Thr100 and Arg101 from the C-arm interact with the DNA phosphates of the T-strand.

4.2.3. Probing specificity determinants of the central T:A pair

Theoretically, the N- and C-arm interactions with the symmetrical CC:GG half-sites are compatible with both alternative binding orientations of the bound DNA, however, the central T:A base pair recognition interface is not symmetrical (Figure 31C). The van der Waals contact between the methyl group of the central T0 base and the Tyr85 residue and the hydrogen bonding interaction between the side chain oxygen of Gln37 and the 6-amino group of A0 implies the asymmetric EcoRII-N binding in one preferable orientation (Figure 33A). In the alternative binding mode, both the van der Waals contact and the hydrogen bond interactions will be disrupted. To determine the importance of these interactions for the EcoRII-N specificity we analyzed EcoRII-N binding to the modified oligoduplexes containing either uridine instead of T0 or 2-aminopurine instead of A0. Gel mobility shift analysis demonstrates that elimination of the methyl group by T to U (uracil) replacement or elimination of the N6 amino group in the case of A to 2-aminopurine replacement completely abolishes DNA binding by EcoRII-N (Figure 33C and D). The key role of Tyr85 in the EcoRII-N binding specificity is also supported by its absolute conservation among the EcoRII homologues (Figure 39). However, it is possible to reconstitute the set of specific interactions with the central base pair using modified bases 5-methyl isocytosine and isoguanine (compare panels A and B in Figure 33).

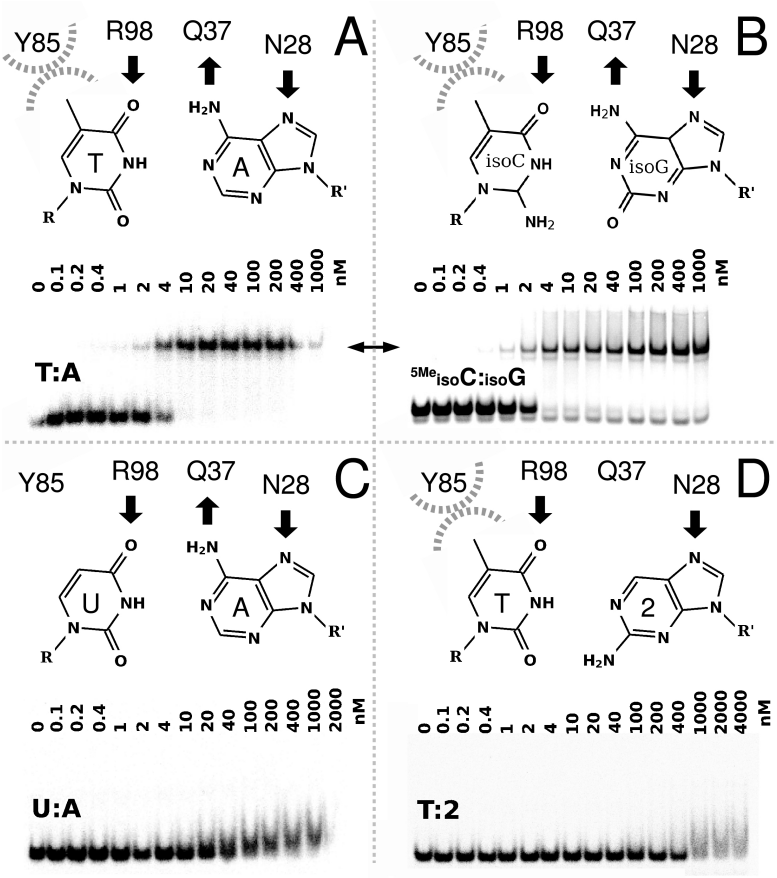


Figure 33. EcoRII-N interactions with oligoduplexes containing modified central base pair. Van der Waals contacts are shown as gray dashes, hydrogen bonds as black arrows pointed to acceptor. Binding reactions (see, section 3.15.) were performed with modified central base pair (Table 5). Horizontal two-direction arrow marks position of the specific complex.

Asymmetric binding pattern explains the results of crosslinking experiments where Tyr41 of EcoRII was found to be cross-linked with the 45% yield to the C-2 of the A-strand of the recognition sequence and no cross-link was observed for the T-strand [196].

4.2.4. Probing sensitivity to methylation and indirect recognition

In silico analysis shows that the conversion of the second cytosine to 5mC in the A-strand should displace the water molecule which is involved in specific hydrogen bond interactions with Gln37 and Glu96 (Figure 31D) and create a steric conflict of the methyl group with the preceding C base (Figure 34D, HM(A)). Introduction of 5mC instead of C-1 in the T-strand should produce a steric clash of methyl group with the Oη atom of Tyr85 involved in specific (Figure 31C) and non-specific (Figure 34, HM(T)) contacts with DNA. C5 methylation of the C-1 residue either in the A- or the T-strand completely abolishes EcoRII-N binding of hemimethylated DNA (Figure 34). Thus, EcoRII-N seems to be more sensitive to the target site

methylation in one of the strands than isolated EcoRII-C. This finding explains previous data on methylation sensitivity and cleavage activation of wt EcoRII [194].

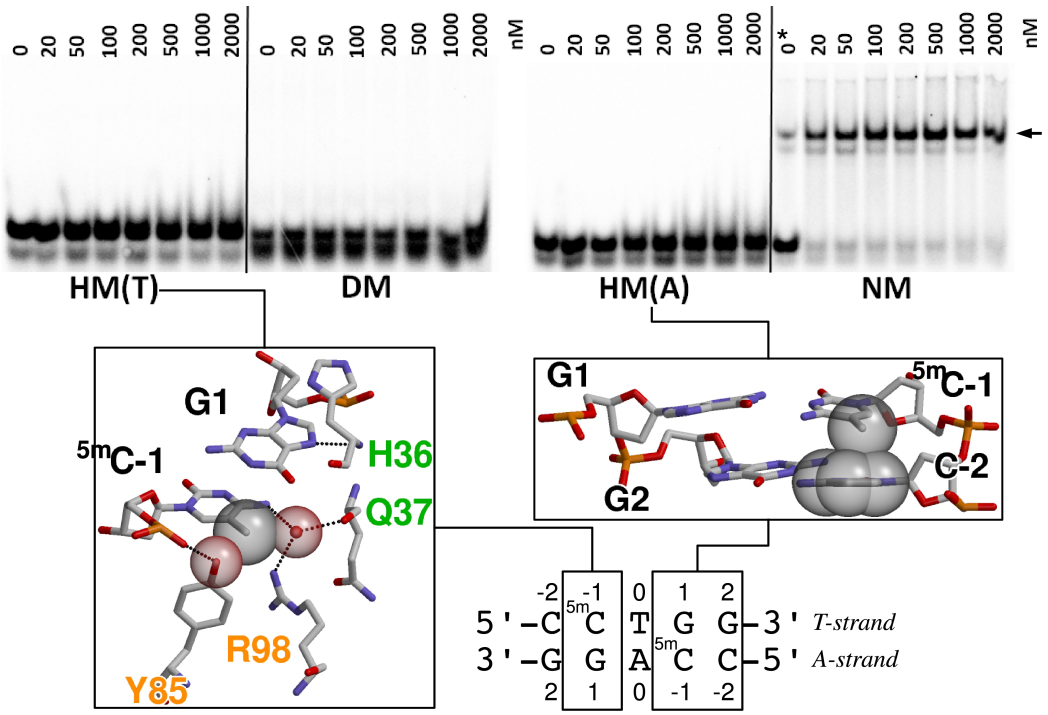


*Figure 34. Functional and structural effects of hemimethylation on cleavage and computer simulation of methyl group position. Gel mobility shift analysis of DNA binding by the EcoRII-N is presented in upper panels. The binding reactions (see, section 3.15.) contained either the non-methylated (NM), hemimethylated (HM) or dimethylated (DM) oligoduplexes (Table 5) and the protein at concentrations as indicated by each line. Arrow marks position of the specific complex. * - carryover of sample from the next line. Oligoduplexes containing 5mC in A- or T-chain of the target site are designated as HM(A) and HM(T), respectively. Below gels are shown in silico modeling of the methylated C at the EcoRII-N target site. Cytosine C-1 was in silico replaced by 5mC [161].*

Crystal structure suggests that water molecules at the EcoRII-N–DNA complex interface are important for the indirect readout of the second and the fourth base pairs (Figure 31B and D) as well as for the methylation sensitivity of EcoRII-N. The pattern of water-mediated specific hydrogen bonds should be disturbed by replacing the C-1:G1 base pair with U:A or G1:C-1 base pair with A:U, correspondingly (Figure 31D). Not surprisingly, EcoRII-N shows no binding to oligoduplexes containing the A:U base pair (designated as A:U and U:A in the Table 5) at either “left” or “right” half-sites (data not shown). This

observation indirectly supports the indirect readout mechanism by the water molecules modeled at the protein-DNA interface.

4.2.5. Similarities of EcoRII-N to the TF factors of B3 family

The crystal structure of apo-EcoRII revealed that EcoRII-N represents a novel fold [14]. NMR structures of B3 DNA binding domains of RAV1 [24] and At1g16640 [140] plant-specific transcription factors appeared to be strikingly similar to EcoRII-N. Indeed, a pairwise comparison of the EcoRII-N structure with the B3 domains of RAV1 and At1g16640 according to DALI [197] gives relatively high similarity Z-score values of 6.2 and 7.1, respectively. Phylogenetic analysis also confirms that EcoRII-N is closely related to the B3 family proteins [24][198]. Structural superposition of EcoRII-N–DNA and the RAV1-B3 domain (Figure 35A) suggests that the plant-specific transcription factors of the B3 family may bind DNA in a similar way. Furthermore, the N- and C-recognition arm of EcoRII-N have direct equivalents in RAV1, such as the α-helix H2, the β-strand B2 for the N-arm and the β-strands B4 and B5 for the C-arm, correspondingly (Figure 35A).

Moreover, putative recognition elements of RAV1-B3 are placed within the major groove of DNA with only minor steric clashes that can be easily relieved by assuming alternative side chain conformations or slightly different loop conformations. Side chains of the Arg202, Trp245 and Asn246 residues of the RAV1-B3 domain are in the vicinity of DNA bases (not shown), while Lys190, Arg241, Thr254 (Figure 35B) and Thr193, Ser195 are close to DNA phosphates.

An NMR titration experiment allowed Yamasaki *et al.* to predict the DNA binding residues of RAV1-B3 and to build a docking model of the RAV1-B3–DNA complex [24]. Many putative DNA binding residues of RAV1-B3 identified by the NMR titration experiment coincide with the DNA binding residues predicted by a structural superposition with EcoRII-N (Figure 35C). Since the coordinates for the RAV1-B3–DNA complex model obtained by

Yamasaki, *et al.* are not available, we were unable to provide the more detailed comparisons. However, a visual inspection of both models with similarly oriented protein barrels shows that DNA in the docking model [24] is rotated around the protein-DNA dyad axis by nearly 90 degrees with respect to the DNA in the model obtained by structural superposition with EcoRII-N.

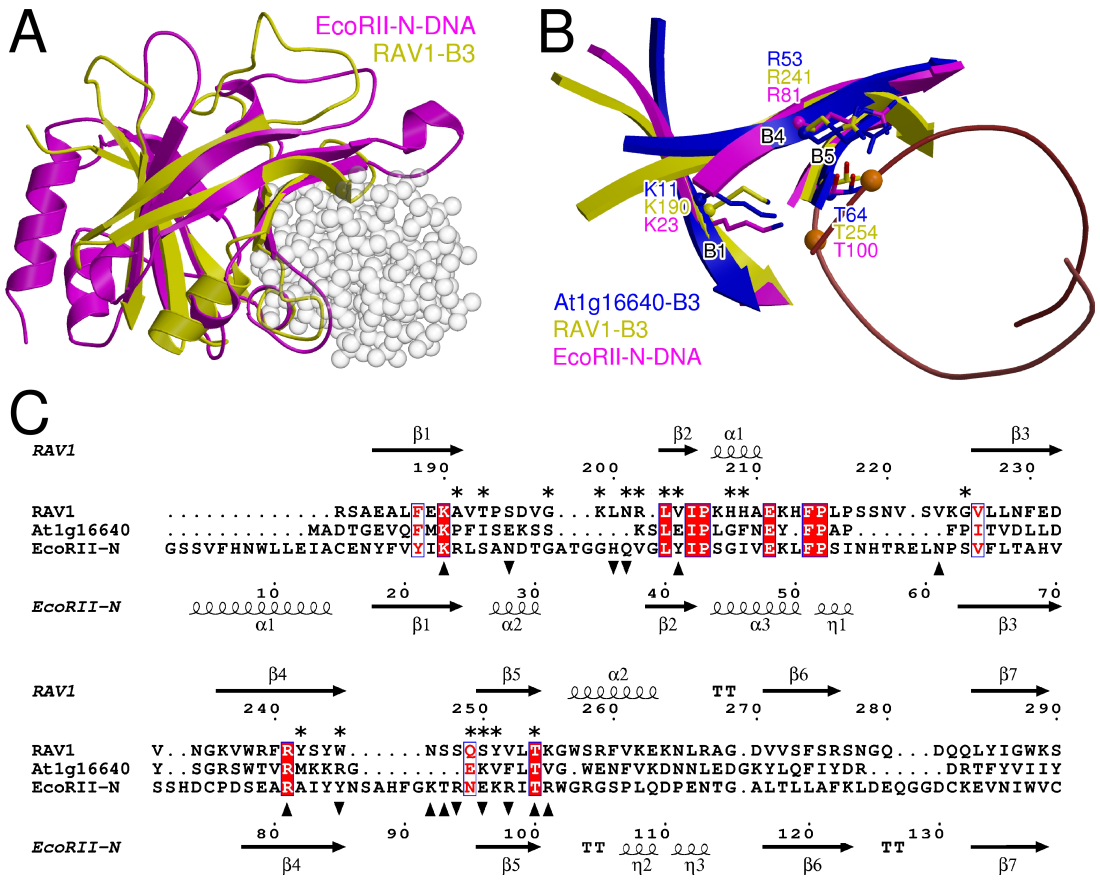


Figure 35. Similarities between EcoRII-N and plant transcription factors of B3 family. (A) Superposition of RAV1-B3 (yellow) [PDB ID 1WID] on EcoRII-N–DNA complex structure (magenta). DNA is depicted in the spacefill representation and shown in transparent gray. (B) Structurally conserved β-strands of EcoRII (magenta), RAV1-B3 (yellow), and At1g16640 (blue) [PDB ID 1YEL] bearing conserved amino acid residues involved in contacts with DNA phosphates. DNA backbone is shown in brown, phosphates as orange spheres. (C) Structural alignment of the B3 domains of RAV1 and At1g16640 with EcoRII-N represented as primary sequence level. Secondary structure elements of RAV1 and EcoRII-N are indicated above and below the alignment, correspondingly. EcoRII-N amino acid residues involved in specific and non-specific contacts are denoted by apex down or up triangles, correspondingly. The asterisks mark position of the RAV1 amino acid residues most affected by DNA binding as demonstrated by NMR spectra.

Structural conservation of the EcoRII-N residues Lys190, Arg241 and Thr254 (absolutely conserved in the EcoRII family (Figure 39)) involved in binding of the DNA backbone phosphates in RAV1-B3 and At1g16640-B3 (Figure 35B) and also within the B3 family (with one exception in case of the ABI-B3 model) (Table 11) [24] argues for the same DNA orientation in EcoRII-N and RAV1-B3 complexes.

Table 11. Sequence conserved residues of EcoRII interacting with the backbone of effectoric DNA and their structural equivalents in the B3 domain proteins

Protein:	EcoRII-N	RAV1-B3	ARF1-B3*	ABI3-B3*	At1g16440-B3
Recognition site	5'-CCTGG	5'-CACCTG	5'-TGTCTC	5'-CATGCA	non specific
Structurally conserved residues	K23	K190	K160	K574	K11
	R81	R241	R211	R623	R53
	T100	T254	T224	E639	T64

* homology models from [24].

4.3 Mechanism of the EcoRII action

4.3.1. Structural transitions occurring upon DNA binding to the N-terminal domains of EcoRII

In the absence of DNA the N-terminal domains of the wt EcoRII REase are positioned in the DNA binding cleft at the C-terminal domain dimer interface creating a steric block which interferes with DNA binding and cleavage. Therefore, it was suggested that in the absence of DNA EcoRII is in the autoinhibited state [14]. The EcoRII-N and EcoRII-C structures in the DNA bound form suggest possible structural changes that may occur during the transition of EcoRII from the autoinhibited apo-form to the functionally active DNA-bound form. Structural comparison of the N-terminal domains in the DNA-free and DNA-bound forms demonstrate that conformational changes of EcoRII-N occurring upon DNA binding are limited to the loop regions. The loop in the C-arm, which harbors the Arg94 residue involved in sequence recognition, moves by 3 Å but does not change its backbone conformation (Figure 36A). The Cα atoms of the loop residues 86-95 in the DNA-free and

bound forms superimpose with an RMSD of 0.25 Å. In contrast, the loop in the N-arm changes its conformation significantly upon DNA binding resulting in the RMSD value of 1.7 Å between the DNA-bound and free forms for the residues 31-38 (Figure 36A). These perturbations in the N-arm loop displace Val38 (Figure 36A) and position the Gln37 residue in a close proximity to the A0 base edge (Figure 36B). The Oε atom of the Gln37 residue in the DNA bound form moves ~8 Å with respect to its position in the DNA-free form and makes a hydrogen bond with the N6 atom of the central A0 nucleotide (Figure 36B). Nearly absolute conservation of Gln37 among the EcoRII homologues (Figure 39) is consistent with its key role in the cognate DNA recognition.

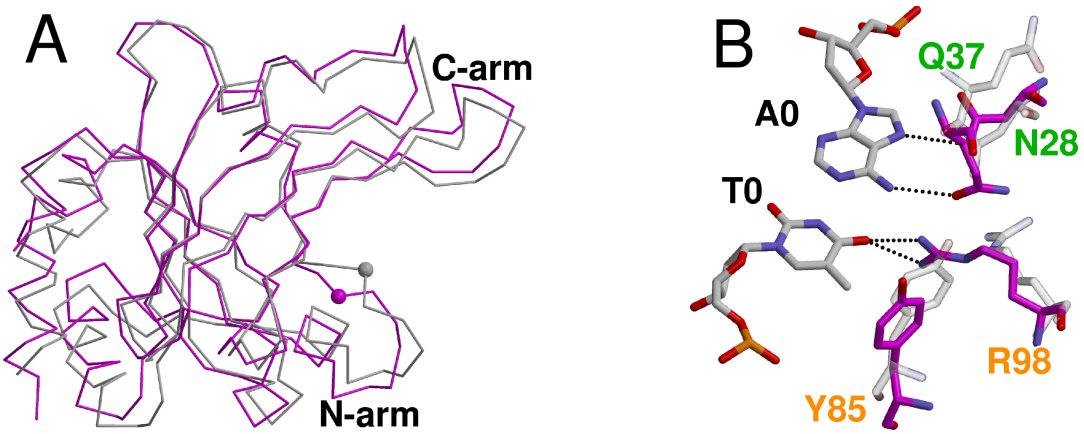


Figure 36. EcoRII-N conformational changes upon DNA binding. (A) Superposition of apo- (transparent) and DNA-bound (magenta) EcoRII-N domains. Cα atom of Val38 is shown as sphere (B) View illustrating conformation of residues involved in the central T:A base pair recognition in apo- (gray) and DNA-bound (magenta) EcoRII-N. Residues from the recognition N-arm labeled in green, from C-arm – in orange (as in Figure 30C).

A rigid body superposition of the EcoRII-N–DNA complexes on the apo-EcoRII (Figure 37A) results in a steric clash between the effector DNA and the C-terminal domain of apo-EcoRII. Thus, in such superposition the phosphates from the effectoric DNA A-strand at the nucleotide positions C-2 and G1 would face negatively charged Glu351 and Glu380 of the C-terminal domain (Figure 37B). Furthermore, the mainchain atoms of the nucleotides at the position C-1 and A0 would have a strong steric interference with the residues in the fragment 377-379 (Figure 37B). The above-mentioned

unfavorable interactions between the nuclease domain and the copy of the effector DNA may destabilize the interface between the N- and C-terminal domains in apo-EcoRII.

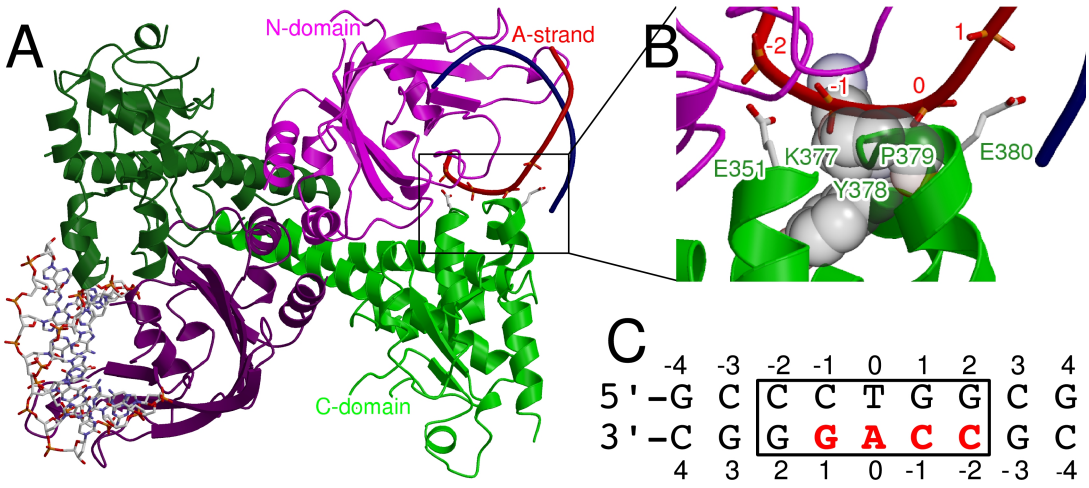


Figure 37. Model of EcoRII bound to the two copies of effector DNA. Based of superimposed complexes EcoRII-N–DNA on structure of autoinhibited EcoRII [PDB ID 1NA6]. Proteins chains of liganded EcoRII-N are omitted for clarity. (A) Structure of EcoRII where A protomer (dark magenta and dark green) is bound to DNA that is shown as sticks, meanwhile B protomer (light magenta and green) is bound to duplex represented by backbone helix. (B) Zoomed in unfavored electrostatic interactions and steric clashes between autoinhibited EcoRII-C (shown as sticks and spacefill, respectively) and backbone of DNA A-strand (red). (C) Nucleotides from EcoRII-N–DNA complex that would fall in clash in apo-EcoRII structure are shown in red within DNA duplex, recognition site is boxed.

In the apo-EcoRII structure the two N-terminal domains contact each other and their DNA binding clefts are exposed into solution. It is possible that binding of two effector DNA copies is required to trigger the conformational change that opens DNA substrate access to both EcoRII-C active sites. Indeed, biochemical studies suggest that only simultaneous binding of two rather than one effector DNA copies results in a concerted cleavage of the third DNA copy at the catalytic EcoRII-C domain [15].

4.3.2. Structural transitions occurring upon DNA binding to the C-terminal domains of EcoRII

In the apo-EcoRII (Figure 38A), the DNA binding interface at the EcoRII-C is not fully assembled compared to the DNA bound form

(Figure 38C). Thus, the EcoRII-C dimer has to rearrange in order to create DNA binding surface. The changes occur both at the N-terminal and C-terminal parts of EcoRII-C but to a different extent. The moderate rearrangements at the hydrophobic N-terminal interface of the EcoRII-C dimer include α-helices H10 and H11, and result in the increase of the buried accessible surface area from 1400 $Å^2$ in the apo-form (Figure 38A) to the 1600 $Å^2$ in the DNA-bound form (Figure 38C). However, rearrangement of dimer is not sufficient to fully encircle the bound DNA (Figure 38B) and conformational changes of monomers are required. The most significant structural change occurring within the monomer is ~25 degree bending of the long α-helix H11 at the Gly267 (compare panels A and C in Figure 38). Interestingly, this residue is absolutely conserved within the EcoRII family (Figure 39). It is also structurally conserved in the other two base flipping REases Ecl18kI and PspGI, suggesting that this residue may act as a hinge required for the rearrangement of dimer interface.

In apo-EcoRII recognition α-helices H14 are located at ~30 Å distance away from each other (Figure 38A). Hence, above-mentioned rearrangements of EcoRII-C provide that in active form this recognition helices comes together and dimerize through residues Gln333 (absolutely conserved in EcoRII family (Figure 39)). This and other newly developed dimerisation contacts at the C-terminal interface between PD-(D/E)XK domains bury 840 $Å^2$ of the accessible surface area closing the clamp around the DNA (Figure 38C). Much smaller interdomain surface and predominantly hydrophilic intermonomer interface suggests that the C-terminal part of EcoRII-C may act as an entrance gate which closes after encircling DNA.

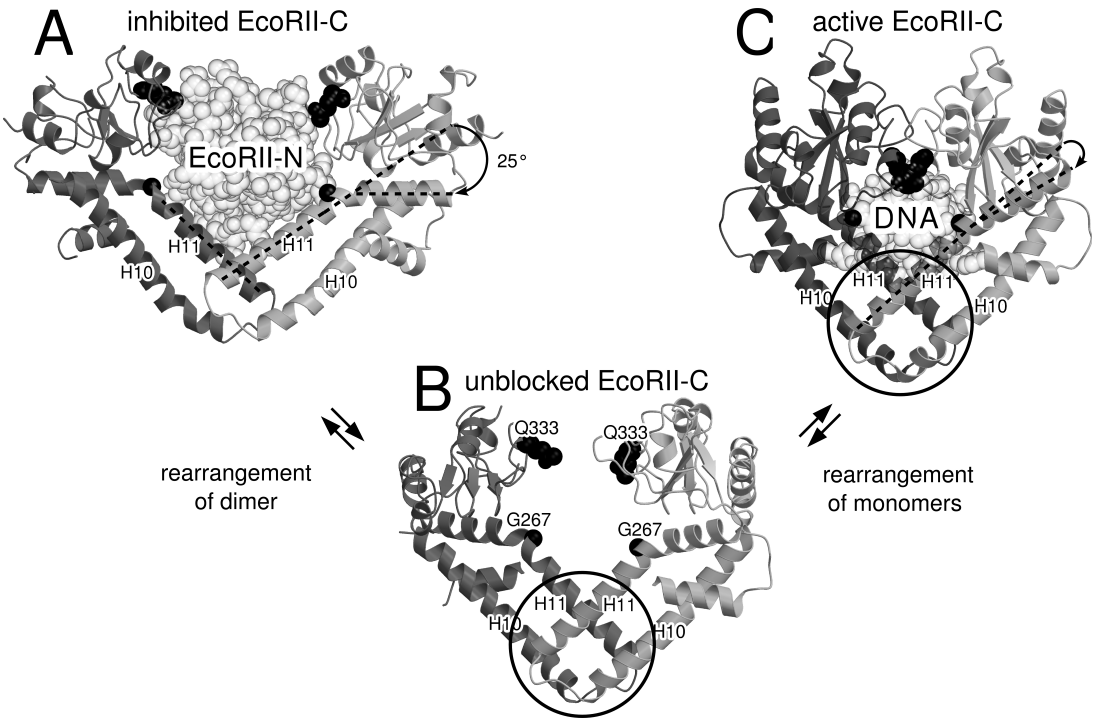
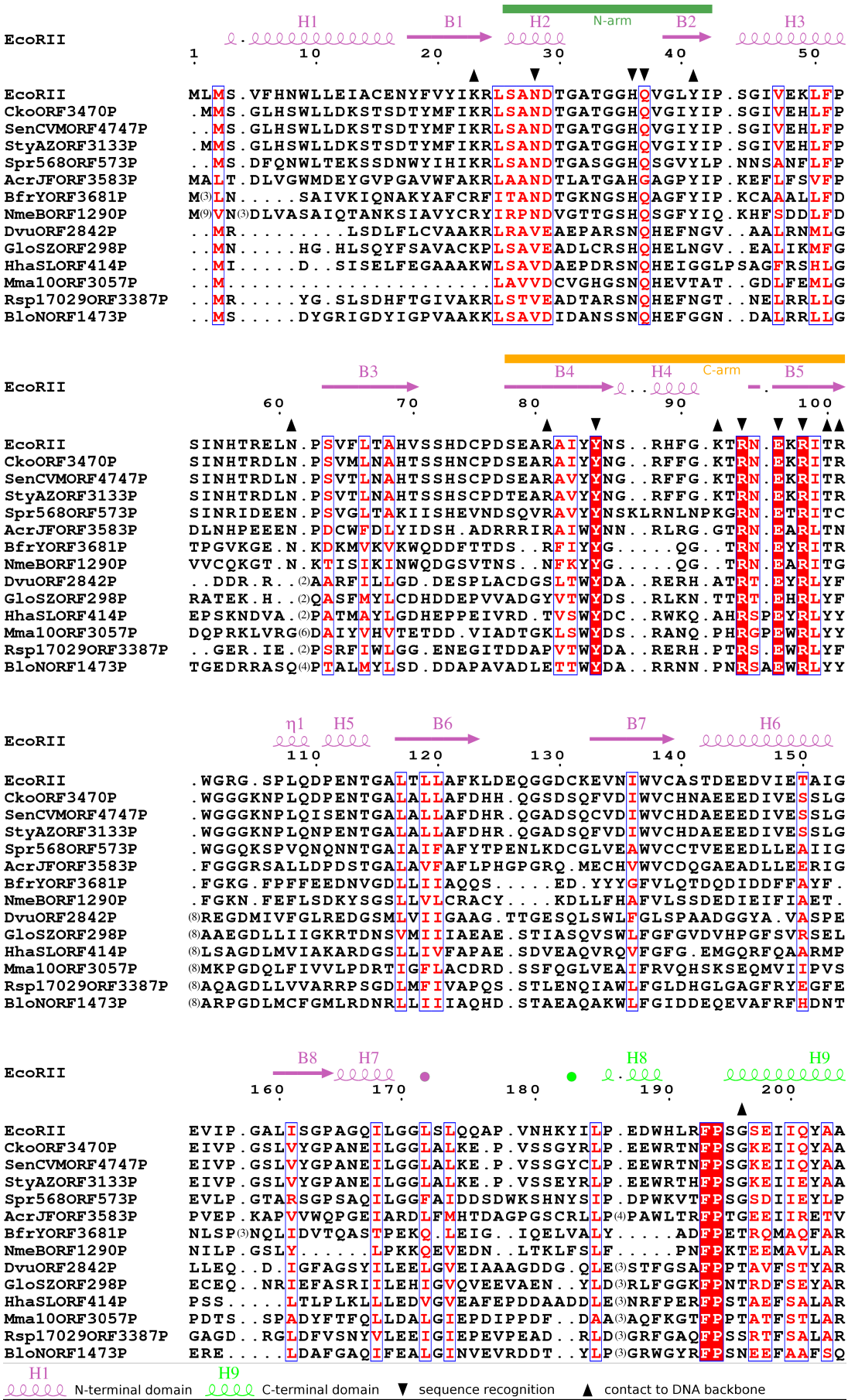


Figure 38. Possible structural changes of EcoRII-C during transition from the autoinhibited form in apo-EcoRII (A) towards activated DNA-bound form (C). (B) EcoRII dimer (N-domains are omitted) with fragments of α-helices H10 and H11 (encircled) oriented exactly as in DNA-bound EcoRII-C (C). Within both monomers Gln333 from recognition a-helix 14 is depicted in black spacefill representation, black sphere marks positions of the Cα atom of Gly267 (B). N-terminal domain in EcoRII (A) and DNA in EcoRII-C–DNA (C) are shown are shown as gray spacefills.

4.4. EcoRII family proteins

According to REBASE database [3] the sequence 5'-CCWGG represents one of the most common recognition sequences for restriction-modification system. Indeed, out of the 3392 RM systems in the REBASE, 225 recognize 5'-CCWGG. REases specific for the CCWGG sequence fall into two groups which differ in the cleavage position (CC↓WGG or ↓CCWGG). EcoRII is a prototype of REases which cut before the first C in CCWGG site. Bioinformatics analysis of REases indicates that EcoRII family includes 45 members [4]. Alignment of the non-redundant protein sequences of EcoRII family showed sequence conservation of structurally and functionally important residues: all active site residues and sequence recognition determinants are almost absolutely conserved (Figure 39).



(continued on the next page)

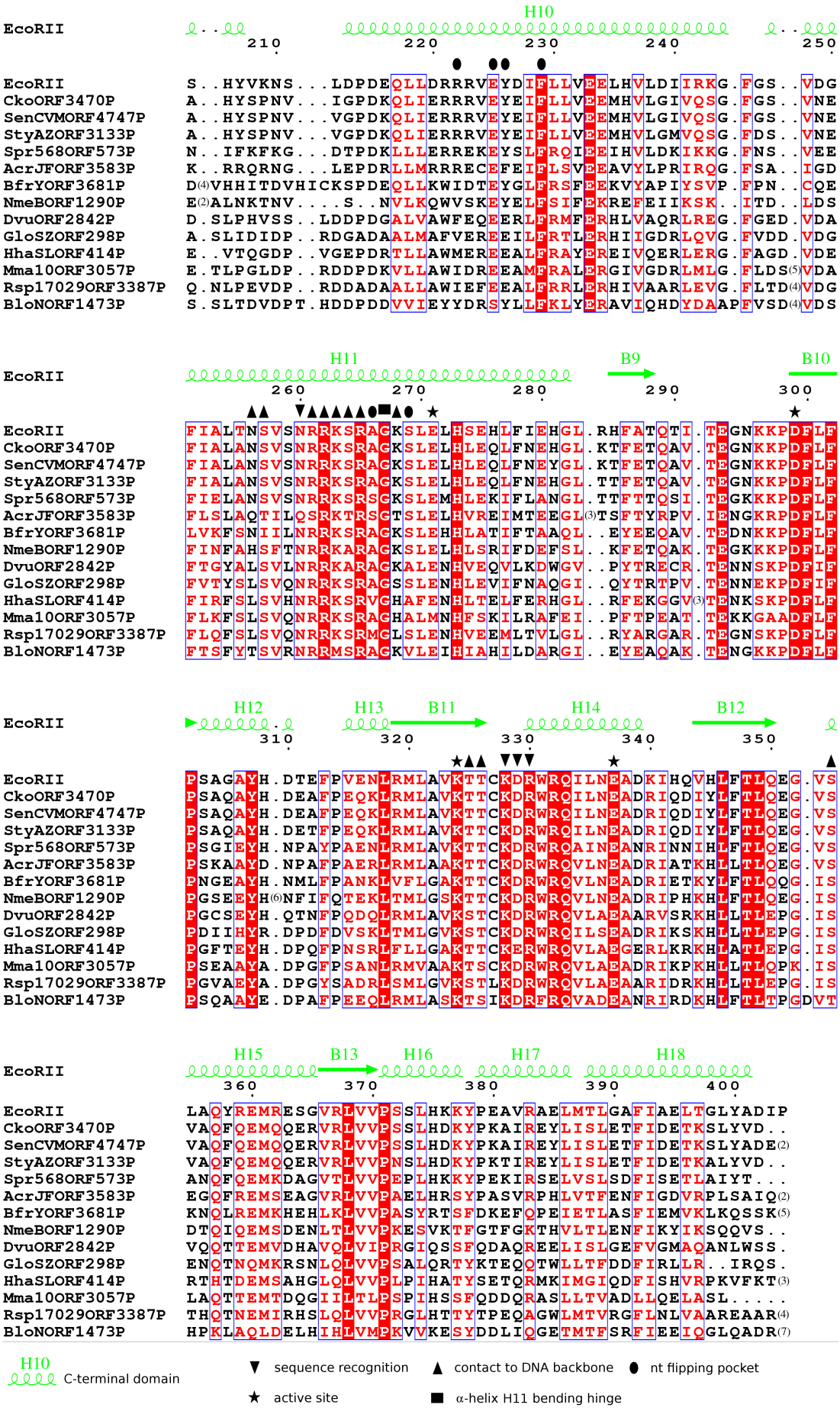

H10
EcoRII
210
220
230
240
250
EcoRII S..HYVKNS...LDPDEQLLDRRRVEYDIFLLVEELHVLDIIRKG.FGS..VDG
CkoORF3470P A..HYSPNV...IGPDKQLIERRRVEYEIFLLVEEMHVLGIVQSG.FGS..VNE
SenCVMORF4747P A..HYSPNV...VGPDKQLIERRRVEYEIFLLVEEMHVLGIVQSG.FGS..VNE
StyAZORF3133P A..HYSPNV...VGPDKQLIERRRVEYEIFLLVEEMHVLGMVQSG.FDS..VNE
Spr568ORF573P N..IFKFKG...DTPDKLLLERREKEYSLFRQIEEIHVLDKIKKG.FNS..VEE
AcrJFORF3583P K..RRQRNG...LEPDRLLMRRRECEFEIFLSVEEAVYLPRIRQG.FSA..IGD
BfrYORF3681P D(4)VHHITDVHICKSPDEQLLKWIDTEYGLFRSFEEKVYAPIYSVP.FPN..CQE
NmeBORF1290P E(2)ALNKTNV....S..NVLKQWVSKEYELFSIFEKREFEIIKSK..ITD..LDS
DvuORF2842P D.SLPHVSS..LDDPDGALVAWFEQEERLFRMFERHLVAQRLREG.FGED..VDA
GloSZORF298P A.SLIDIDP..RDGADAALMAFVEREEILFRTLERHIIGDRLQVG.FVDD..VDG
HhaSLORF414P E..VTQGDP..VGEPDRTLLAWMEREEALFRAYEREIVQERLERG.FAGD..VDE
Mma10ORF3057P E.TLPGLDP..RDDPDKVLLAWIDREEAMFRALERGIVGDRLMLG.FLDS(5)VDA
Rsp17029ORF3387P Q.NLPEVDP..RDDADAALLAWIEFEEALFRRLERHIVAARLEVG.FLTD(4)VDG
BloNORF1473P S.SLTDVDPT.HDDPDDVVIEYYDRSYLLFKLYERAVIQHDYDAAPFVSD(4)VDS
H11
B9
B10
EcoRII
260
270
280
290
300
EcoRII FIALTNSVSNRRKSRAGKSLELHSEHLFIEHGL.RHFATQTI.TEGNKKPDFLF
CkoORF3470P FIALANSVSNRRKSRAGKSLELHLEQLFNEHGL.KTFETQAV.TEGNKKPDFLF
SenCVMORF4747P FIALANSVSNRRKSRAGKSLELHLEQLFNEYGL.KTFETQAV.TEGNKKPDFLF
StyAZORF3133P FIALANSVSNRRKSRAGKSLELHLEQLFNEHGL.TTFETQAV.TEGNKKPDFLF
Spr568ORF573P FIELANSVSNRRKSRSGKSLEMHLEKIFLANGL.TTFTTQSI.TEGKKKPDFLF
AcrJFORF3583P FLSLAQTILQSRKTRSGTSLELHVREIMTEEGL(3)TSFTYRPV.IENGKRPDFLF
BfrYORF3681P LVKFSNIILNRRKSRAGKSLEHHLATIFTAAQL..EYEEQAV.TEDNKKPDFLF
NmeBORF1290P FINFAHSFTNRRKARAGKSLELHLSRIFDEFSL..KFETQAK.TEGKKKPDFLF
DvuORF2842P FTGYALSVLNRRKARAGKALENHVEQVLKDWGV..PYTRECR.TENNSKPDFIF
GloSZORF298P FVTYSLSVQNRRKSRAGSSLENHLEVIFNAQGI..QYTRTPV.TENNEKPDFIF
HhaSLORF414P FIRFSLSVHNRRKSRVGHAFENHLTELFERHGL..RFEKGGV(3)TENKSKPDFLF
Mma10ORF3057P FLKFSLSVQNRRKSRAGHALMNHFSKILRAFEI..PFTPEAT.TEKKGAADFLF
Rsp17029ORF3387P FLQFSLSVQNRRKSRMGLSLENHVEEMLTVLGL..RYARGAR.TEGNSKPDFLF
BloNORF1473P FTSFYTSVRNRRMSRAGKVLEIHIAHILDARGI..EYEAQAK.TENGKKPDFLF
H12
H13
B11
H14
B12
EcoRII
310
320
330
340
350
EcoRII PSAGAYH.DTEFPVENLRMLAVKTTCKDRWRQILNEADKIHQVHLFTLQEG.VS
CkoORF3470P PSAQAYH.DEAFPEQKLRMLAVKTTCKDRWRQILNEADRIQDIYLFTLQEG.VS
SenCVMORF4747P PSAQAYH.DEAFPEQKLRMLAVKTTCKDRWRQILNEADRIQDIYLFTLQEG.VS
StyAZORF3133P PSAQAYH.DETFPEQKLRMLAVKTTCKDRWRQILNEADRIQDIYLFTLQEG.VS
Spr568ORF573P PSGIEYH.NPAYPAENLRMLAVKTTCKDRWRQAINEANRINNIHLFTLQEG.VS
AcrJFORF3583P PSKAAYD.NPAFPAERLRMLAAKTTCKDRWRQVLNEADRIATKHLLTLQEG.VS
BfrYORF3681P PNGEAYH.NMLFPANKLVFLGAKTTCKDRWRQVLNEADRIETKYLFTLQQG.IS
NmeBORF1290P PGSEEYH(6)NFIFQTEKLTMLGSKTTCKDRWRQVLNEADRIPHKHLFTLQEG.IS
DvuORF2842P PGCSEYH.QTNFPQDQLRMLAVKSTCKDRWRQVLAEAARVSRKHLLTLEPG.IS
GloSZORF298P PDIIHYR.DPDFDVSKLTMLGVKSTCKDRWRQILSEADRIKSKHLLTLEPG.IS
HhaSLORF414P PGFTEYH.DPQFPNSRLFLLGAKTTCKERWRQVLAEGERLKRKHLATLEPG.IS
Mma10ORF3057P PSEAAYA.DPGFPSANLRMVAAKTSCKDRWRQVLAEADRIKPKHLLTLQPK.IS
Rsp17029ORF3387P PGVAEYA.DPGYSADRLSMLGVKSTLKDRWRQVLAEAARIDRKHLLTLEPG.IS
BloNORF1473P PSQAAYE.DPAFPEEQLRMLASKTSIKDRFRQVADEANRIRDKHLFTLTPGDVT
H15
B13
H16
H17
H18
EcoRII
360
370
380
390
400
EcoRII LAQYREMRESGVRLVVPSSLHKKYPEAVRAELMTLGAFIAELTGLYADIP
CkoORF3470P VAQFQEMQQERVRLVVPSSLHDKYPKAIREYLISLETFIDETKSLYVD..
SenCVMORF4747P VAQFQEMQQERVRLVVPSSLHDKYPKAIREYLISLETFIDETKSLYADE(2)
StyAZORF3133P VAQFQEMQQERVRLVVPNSLHDKYPKTIREYLISLETFIDETKALYVD..
Spr568ORF573P ANQFQEMKDAGVTLVVPEPLHKKYPEKIRSELVSLSDFISETLAIYT...
AcrJFORF3583P EGQFREMSEAGVRLVVPAELHRSYPASVRPHLVTFENFIGDVRPLSAIQ(2)
BfrYORF3681P KNQLREMKHEHLKLVVPASYRTSFDKEFQPEIETLASFIEMVKLKQSSK(5)
NmeBORF1290P DTQIQEMSDENLTLVVPKESVKTFGTFGKTHVLTLENFIKYIKSQQVS..
DvuORF2842P VQQTTEMVDHAVQLVIPRGIQSSFQDAQREELISLGEFVGMAQANLWSS.
GloSZORF298P ENQTNQMKRSNLQLVVPSALQRTYKTEQQTWLLTFDDFIRLLR..IRQS.
HhaSLORF414P RTHTDEMSAHGLQLVVPLPIHATYSETQRMKIMGIQDFISHVRPKVFKT(3)
Mma10ORF3057P LAQTTEMTDQGIILTLPSPIHSSFQDDQRASLLTVADLLQELASL.....
Rsp17029ORF3387P THQTNEMIRHSLQLVVPRGLHTTYTPEQAGWLMTVRGFLNLVAAREAAR(4)
BloNORF1473P HPKLAQLDELHIHLVMPKVVKESYDDLIQGETMTFSRFIEEIQGLQADR(7)
H10 C-terminal domain
▼ sequence recognition
▲ contact to DNA backbone
● nt flipping pocket
★ active site
■ α-helix H11 bending hinge

Figure 39. Protein sequence alignment of EcoRII homologues. Thirteen sequences of full-length non-redundant EcoRII homologues from putative restriction modification systems were obtained from REBASE and their entry codes are shown at the beginning of each sequence. Numbers above alignment correspond to the EcoRII sequence. EcoRII secondary structure elements [PDB ID 1NA6] are shown above the alignment and colored in magenta for EcoRII-N and in green for EcoRII-C. Symbols under the alignment designate the residue function in EcoRII REase. Abbreviations are provided in the Figure. C-termini of EcoRII-N and N-termini of EcoRII-C separately solved crystal structures are denoted by magenta and green cycles, correspondingly.

4.5. Analysis of BfiI domain interactions with DNA

4.5.1. Asymmetric units of BfiI-C–DNA and BfiI-N-K107A crystals

BfiI-C–DNA complex crystallized (see, section 3.11.) in a spacegroup $P6_5$ and the structure was determined by the molecular replacement method (see, section 3.12.). Structure refinement (Table 7) showed that the asymmetric unit contains two copies of BfiI-C–DNA complex (Figure 40A), that contact each other burring ~1000 $Å^2$ at protein interface.

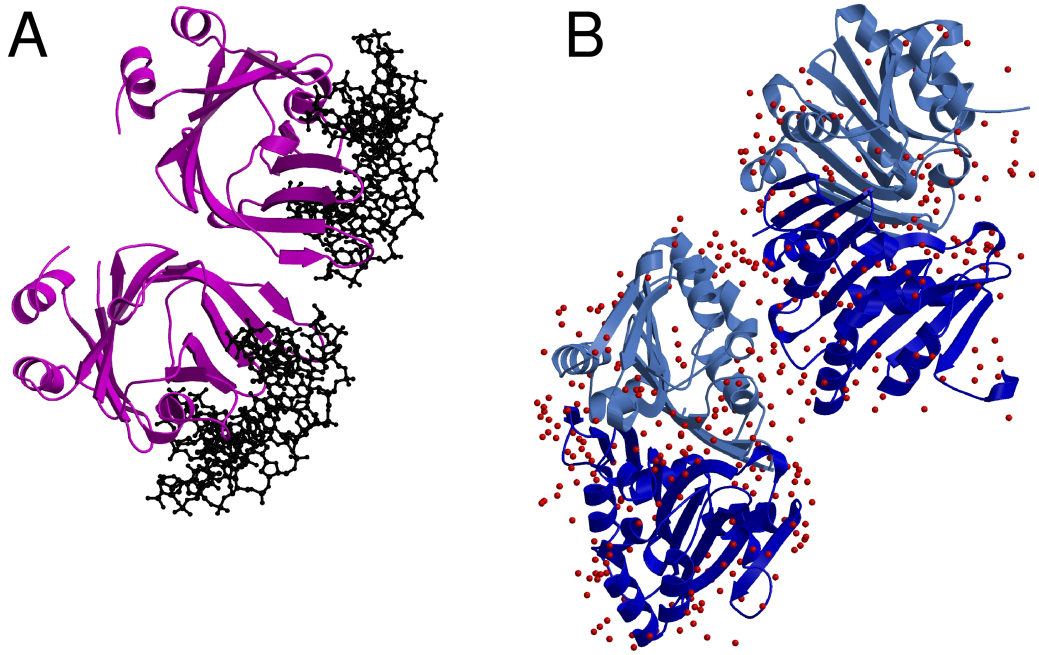


Figure 40. Asymmetric unit of (A) BfiI-C–DNA and (B) BfiI-N-K107A crystals. DNA bound BfiI-C is colored magenta, DNA is represented as black ball-and-sticks. Monomers of BfiI-N-K107A domain dimer are colored in light and dark blue, water molecules are in red.

Residues of both C-terminal domains within the asymmetric unit of the crystal structure (residues 199-358) were modeled in the electron density maps along with the part of the linker (residues 193-198). The secondary structure of the DNA-bound BfiI-C domain is identical to that of the A chain in the apo-BfiI structure (residues 193-358) except of a few 3_{10} helices (discussed in Chapter 4.6.) and follows the same labeling.

The N-terminal domain (residues 1-193) of the active site mutant K107A crystallized (see, section 3.11.) in a spacegroup I4 and the structure was determined by the molecular replacement method (see, section 3.12.). Structure refinement (Table 7) showed asymmetric unit comprised of two protein dimers and ~400 water molecules (Figure 40B). Full catalytic domain (residues 1-170) and part of the linker (residues 171-189) were modeled. Both BfiI-N-K107A dimers (residues 1-170) could be superimposed to apo-BfiI by Cα atoms with RMSD value of 0.35.

4.5.2. DNA recognition by the BfiI-C domain

DNA in the complex with the BfiI-C domain represents a canonical B-form conformation (Figure 41A and B). The DNA binding cleft of BfiI-C is very similar to that of EcoRII-N [21]. It is also shaped to grip DNA from the major groove side (Figure 41A and B), but has two additional loops which buttress DNA at the minor groove (Figure 42). The “N-arm” of the binding cleft is formed by an α-helix H6, β-strand B11 and the loop between them (residues 210-231). Two anti-parallel β-strands B14, B15 and the loop connecting them comprise the “C-arm” (residues 266-287) of the binding cleft (Figure 41A and B). Both arms penetrate into the DNA major groove and amino acid residues interact with the base edges of the recognition sequence. Bases at the 5'-end of the recognition site are bound exclusively by the N-arm, meanwhile 3'-end is contacted by the C-arm (Figure 41C). Most of recognition contacts to the base pairs T0:A0 and G1:C-1 in the middle of the target site are made predominantly by the C-arm with only few contacts to the bases in the T-strand (Figure 41C) coming from the N-arm.

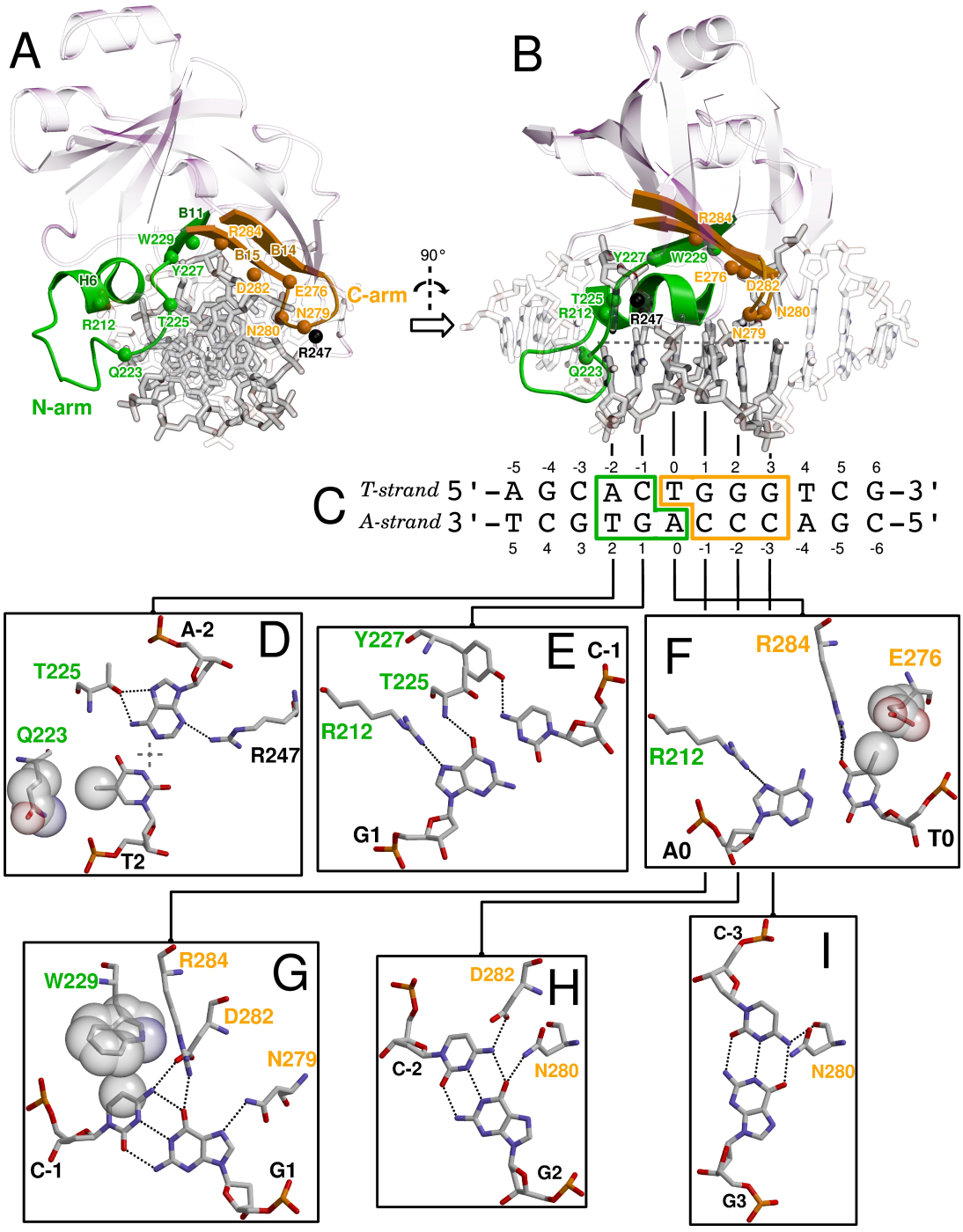


Figure 41. Mechanism of DNA recognition by BfiI C-terminal domain through hydrogen-bonding and van der Waals interactions. (A, B) View of the BfiI-C–DNA complex along the long DNA axis (A) and side view (B). Secondary structure elements of BfiI-C recognition cleft which are closer to the N-terminus: helix H6, strand B11 (N- arm) are colored in green; and those are closer to the C-terminus: strands B13, B14 (C- arm) – in orange. The same coloring scheme is applied for residues which are involved in the specific DNA recognition, their Cα atoms are represented as spheres. DNA bases of binding site are colored in dark gray in (A, B) and are boxed in (C) according what recognition arm they have hydrogen-bonding with. (D-I) Panels for individual base pairs are arranged following the T-strand in 5'→3' direction, complex is oriented as in panel A.

Detailed interactions of the BfiI-C residues with each base pair within site are described (Figure 41D-I) are depicted as follows:.

•*with the base pair A-2:T2:* In the DNA major groove the Oγ atom of Thr225 donates a hydrogen bond to the N7 atom and accepts one from N6 atom of the adenine (Figure 41D). The methyl group of the thymine could make putative van der Waals contact with side chain atoms Cβ and Cγ of Gln223. In the minor groove the N3 atom of the adenine can accept hydrogen bond from one of the Nη atoms of Arg247.

•*C-1:G1*: The Oη atom of the Tyr227 accepts hydrogen bond from the cytosine N4 atom. The N atom of Thr225 and guanidino group of Arg212 donate hydrogen bonds to O6 and N7 atoms of the guanine, respectively (Figure 41E).

•*T0:A0*: The Nη1 and Nη2 atoms of Arg284 make hydrogen bonds to the thymine O4 atom. The side chain atoms of Glu276 make van der Waals contact (<3.5 Å) with the methyl group of the thymine (Figure 41F). The guanidino group of Arg212 is in the hydrogen bond distance to the N7 atom of adenine, however geometry is not optimal.

•*G1:C-1*: The N7 atom of guanine accepts a hydrogen bond from the Nδ atom of Asn279. The Nη2 atom of Arg284 makes hydrogen bond to the guanine O6 atom. The Nδ atom of Asn280 is in the hydrogen bond distance to the O6 atom of guanine. The Oδ1 atom of Glu282 accepts a hydrogen bond from the cytosine N4 atom. Aromatic ring of Trp229 is in proper distance to make van der Waals interactions (3.6 Å) to C5 atom of cytosine (Figure 41G).

•*G2:C-2*: The O6 atom of guanine accepts a hydrogen bond from the Nδ atom of Asn280. The Oδ2 atom of Glu282 accepts a hydrogen bond from the cytosine N4 atom (Figure 41H).

•G3:C-3: The main chain O atom of Asn280 accepts a hydrogen bond from the N4 atom of cytosine (Figure 41I).

4.5.3. Non-specific contacts in the BfiI-C–DNA complex

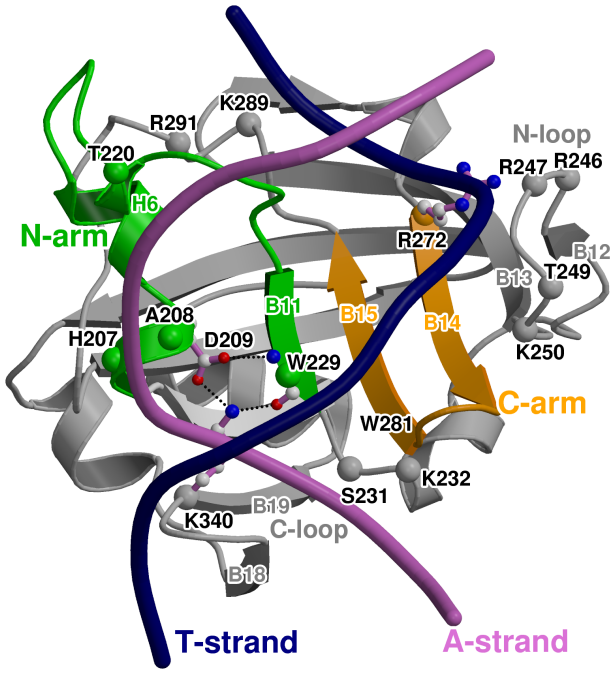


Figure 42. Nonspecific interactions between BfiI-C and 12 bp oligoduplex. Protein Cα atoms of residues that interact with the phosphodiester backbone of DNA are depicted as spheres and colored as protein secondary structure. Recognition N- and C-arms are colored in green and orange, correspondingly, other structural elements are in gray.

The backbone phosphates groups of the DNA T-strand are in contact with C-arm of binding cleft and loop between strands β12 and β13 (N-loop) (Figure 42). Positively charged residues K289, Arg291 and Arg272 from the C-arm make electrostatic interactions with phosphates of C-3 and A-2 nucleotides, respectively. N-loop through main chain nitrogen atoms of Arg246, Arg247 and sidechain of Thr249 contacts phosphates of C-1 and T0 nucleotides.

Phosphodiester backbone of DNA A-strand is contacted mainly by residues from the N-arm, loop next to the N-arm and the loop between strands β18 and β19 (C-loop) (Figure 42). The phosphate of C-1 is contacted by the backbone nitrogen atom of Ala208. Phosphates of C-1 and A0 are at the hydrogen bond distances with the atom Nε2 of His207 and the Oγ atom of Thr220, respectively. Side chain nitrogen atom of Lys232 makes salt bridge with the phosphate of A-4 nucleotide, meanwhile the backbone nitrogen atom contacts phosphate of C-3 nucleotide, which is additionally fixed by the side chain oxygen of Ser231. Nζ atom of Lys340 from C-loop participates in multiple interactions with phosphate of C-2, side chain oxygen atom of Asp209 and with the main chain oxygen atom of Tpr229. These interactions stick together secondary structure elements of N-arm to the C-2 phosphate of A-strand.

4.5.4. B3-like domains share conservative DNA interaction mechanism

Crystal structures of the BfiI-C–DNA and EcoRII-N–DNA complexes revealed that the orientation of DNA duplex in respect to the β-barrel in both cases is identical (compare Figures 32 and 42). Furthermore, structural elements of DNA recognition clefts (N- and C-arms) of both domains are anchored at DNA major groove by nonspecific interactions conveyed by conservative lysine and arginine residues. Thus, C-arms of both domains interact with the T-strand of the DNA duplexes by structurally coinciding arginine residues. C-arm of EcoRII-N is bound through Arg81 to the phosphate of C-2 nucleotide, while C-arm of BfiI-C interacts with the phosphate of A-2 nucleotide through Arg272 (compare Figures 32 and 42). In both domains the N-arm makes non-specific contacts with the A-strand through well conserved lysine residues. However, these residues are located in EcoRII and BfiI on different secondary structure elements. In EcoRII Lys23 located on β-strand B1 makes hydrogen bond to β-strand B2 and interacts electrostatically with α-helix H2 and phosphate of C-2 nucleotide (Figure 32). In BfiI Lys340 located on β-strand B19 makes hydrogen bond to β-strand B11 and interacts electrostatically with α-helix H6 and phosphate of C-2 nucleotide (Figure 42).

Crystal structure of BfiI-C–DNA supports our hypothesis (see, chapter 4.2.5.) that proteins of B3 fold share conservative DNA binding mechanism. Comparison of two DNA bound structures of B3-like domains allows us to propose the spatial conservation of positively charged residues involved in contacts to DNA backbone.

4.6. A model of BfiI activation

Superposition of DNA-bound BfiI-C with the C-terminal domains of apo-BfiI [PDB ID 2C1L] results in RMSD of 1.2 Å and 1.4 Å for the apo-BfiI chains A and B, respectively. The structural alignment of the β-barrel using Cα atoms of the 118 residue fragment (“β-sheet 1” as described in the Supplementary data in [21]), yields RMSD of only ~0.5 Å, confirming that

most of the β-barrel remains unchanged upon DNA binding but positions of the N- and C-arms and the N-loop differ significantly (Figure 43). The α-helix H6 and the β-strand B11 stay rigid upon DNA binding. The part of the loop (the residues 216-219) moves ~4.0 Å towards the DNA (Figure 43) bringing Thr225 close to the with first two base pairs of the recognition site (Figure 41D and E). The side chains of Arg212, Tyr227 and Trp229 have conformations similar to those observed in the apo-BfiI crystal structure.

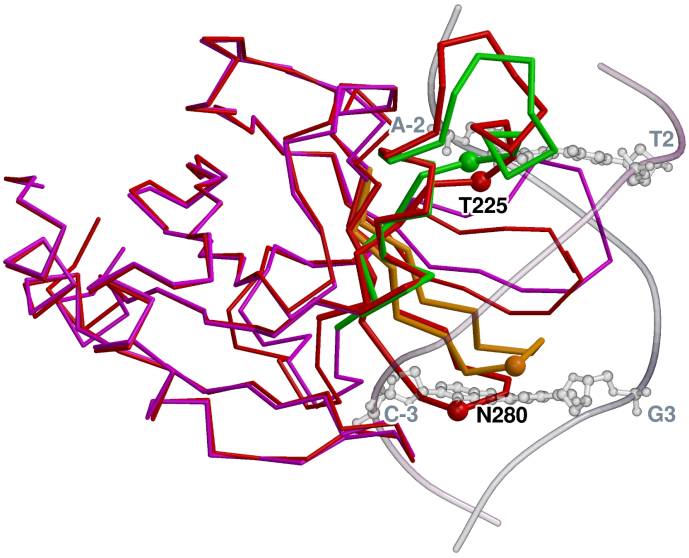


Figure 43. Superposition of C-terminal BfiI domain (red) to DNA-bound BfiI-C domain (magenta), recognition N- and C-arms are colored in green and orange, respectively. Residues making specific hydrogen bonds through main chain atoms are denoted by Cα. DNA is shown as transparent helix, first and last base pairs of the site are shown in a ball-and-stick representation.

Upon DNA binding the position of the C-arm also changes considerably with respect to the β-barrel core (Figure 43). The distance between the strands B15 and B18 increases. The movement of the C-arm results in repositioning of the Cα atom of Asn280 by approx. 4 Å(Figure 43). The DNA binding is also accompanied by folding of the residues 232-237 into a 3_{10}-helix, which results in the favorite dipole interaction of the backbone nitrogen atom of Lys232 and the phosphate moiety in the C-3 base of the A-strand (Figure 42).

A structural superposition of the BfiI C-terminal domain bound to DNA onto the full length apo protein revealed a significant reorganization of the C- and N-terminal domain interface. Indeed, the movement of the recognition C-arm into the DNA major groove (Figure 43) induces concerted movement of the Phe281, Phe275, Ala277, Tyr253 residues (Figure 44B). In addition, DNA backbone induced coil-to-3_{10}-helix compression puts Tyr235 out of the

complementarity with the N-terminal domain surface (compare panels A and B in Figure 44). Noteworthy, described hydrophobic interface is absolutely conserved in the homologues of BfiI, with the a single observed difference at the position 281, the change of Phe to Val (Figure 45). Thus, the observed movements may be of functional importance.

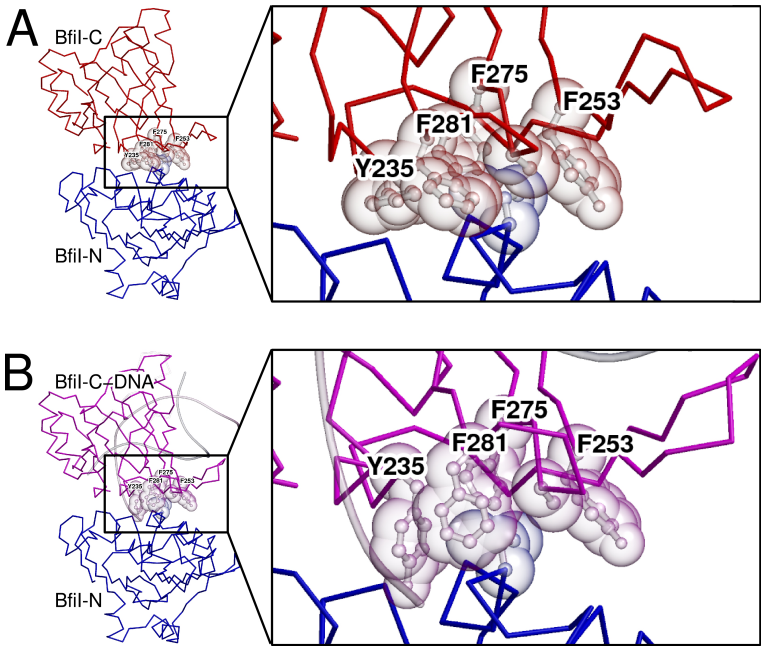


Figure 44. (A) Inter-domain interface in the apo-BfiI monomer, which is shown as ribbon. N- and C-terminal domains are colored in blue and red, residues from hydro-phobic cluster at the BfiI are labeled and shown as ball-and-sticks and spheres. (B) DNA (transparent helix) induced changes of interdomain interface at BfiI-C (magenta).

4.7. BfiI family proteins

Restriction enzyme R.BmrI [199] is an BfiI has isoshizomer of BfiI and shares the same 358 a.a. length and ~80% of sequence identity (Figure 45). Mutational analysis to alanine confirmed functional importance of predicted active site residues of BmrI [74]. Sequence alignment of BfiI close homologs found in REBASE (Figure 45) shows conserved active site residues and DNA binding residues at the C-terminal domain. Of note is that BmrI has Pro instead of Asp175 and no negatively charged residue in linker in vicinity of 175 position (±5 a.a.). Other negatively charged residues Asp185, Asp194 and Asp/Glu at positions 186, 195 are absolutely conserved within the family supporting possible function of the interdomain linker as DNA mimic [21].

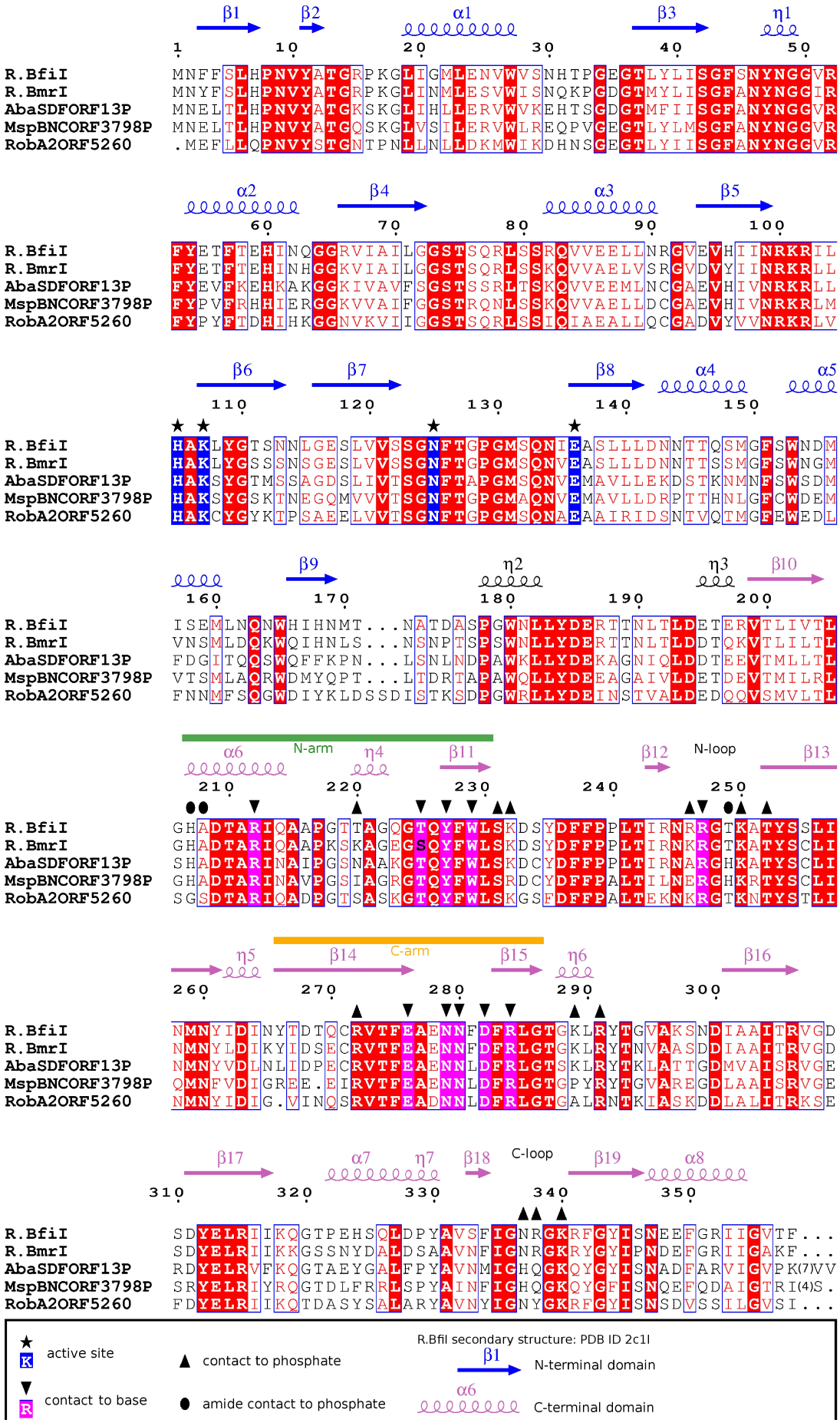

β1 β2 α1 β3 η1
1 10 20 30 40 50
R.BfiI MNFFSLHPNVYATGRPKGLIGMLENVWVSNHTPGEGTLYLISGFSNYNGGVR
R.BmrI MNYFSLHPNVYATGRPKGLINMLESVWISNQKPGDGTMYLISGFANYNGGIR
AbaSDFORF13P MNELTLHPNVYATGKSKGLIHLLERVWVKEHTSGDGTMFIISGFANYNGGVR
MspBNCORF3798P MNELTLHPNVYATGQSKGLVSILERVWLREQPVGEGTLYLMSGFANYNGGVR
RobA2ORF5260 .MEFLLQPNVYSTGNTPNLLNLLDKMWIKDHNSGEGTLYIISGFANYNGGVR
α2 β4 α3 β5
60 70 80 90 100
R.BfiI FYETFTEHINQGGRVIAILGGSTSQRLSSRQVVEELLNRGVEVHIINRKRIL
R.BmrI FYETFTEHINHGGKVIAILGGSTSQRLSSKQVVAELVSRGVDVYIINRKRLL
AbaSDFORF13P FYEVFKEHKAKGGKIVAVFSGSTSSRLTSKQVVEEMLNCGAEVHIVNRKRLL
MspBNCORF3798P FYPVFRHHIERGGKVVAIFGGSTRQNLSSKQVVAELLDCGAEVHIVNRKRLM
RobA2ORF5260 FYPYFTDHIHKGGNVKVIIGGSTSQRLSSIQIAEALLQCGADVYVVNRKRLV
β6 β7 β8 α4 α5
110 120 130 140 150
R.BfiI HAKLYGTSNNLGESLVVSSGNFTGPGMSQNIEASLLLDNNTTQSMGFSWNDM
R.BmrI HAKLYGSSSNSGESLVVSSGNFTGPGMSQNVEASLLLDNNTTSSMGFSWNGM
AbaSDFORF13P HAKSYGTMSSAGDSLIVTSGNFTAPGMSQNVEMAVLLEKDSTKNMNFSWSDM
MspBNCORF3798P HAKSYGSKTNEGQMVVVTSGNFTGPGMAQNVEMAVLLDRPTTHNLGFCWDEM
RobA2ORF5260 HAKCYGYKTPSAEELVVTSGNFTGPGMSQNAEAAIRIDSNTVQTMGFEWEDL
β9 η2 η3 β10
160 170 180 190 200
R.BfiI ISEMLNQNWHIHNMT...NATDASPGWNLLYDERTTNLTLDETERVTLIVTL
R.BmrI VNSMLDQKWQIHNLS...NSNPTSPSWNLLYDERTTNLTLDDTQKVTLILTL
AbaSDFORF13P FDGITQQSWQFFKPN...LSNLNDPAWKLLYDEKAGNIQLDDTEEVTMLLTL
MspBNCORF3798P VTSMLAQRWDMYQPT...LTDRTAPAWQLLYDEEAGGAIVLDETDEVTMILRL
RobA2ORF5260 FNNMFSQGWDIYKLDSSDISTKSDPGWRLLYDEINSTVALDEDQQVSMVLTL
α6 N-arm η4 β11 β12 N-loop β13
210 220 230 240 250
R.BfiI GHADTARIQAAPGTTAGQGTQYFWLSKDSYDFFPPLTIRNRRGTKATYSSLI
R.BmrI GHADTARIQAAPKSKAGEGSQYFWLSKDSYDFFPPLTIRNKRGTKATYSCLI
AbaSDFORF13P SHADTARINAIPGSNAAKGTQYFWLSKDCYDFFPPLTIRNARGHKATYSCLI
MspBNCORF3798P GHADTARINAVPGSIAGRGTQYFWLSRDCYDFFPALTILNERGHKRTYSCLI
RobA2ORF5260 SGSDTARIQADPGTSASKGTQYFWLSKGSFDFFPALTEKNKRGTKNTYSTLI
η5 β14 C-arm β15 η6 β16
260 270 280 290 300
R.BfiI NMNYIDINYTDTQCRVTFEAENNFDFRLGTGKLRYTGVAKSNDIAAITRVGD
R.BmrI NMNYLDIKYIDSECRVTFEAENNFDFRLGTGKLRYTNVAASDDIAAITRVGD
AbaSDFORF13P NMNYVDLNLIDPECRVTFEAENNLDFRLGTSKLRYTKLATTGDMVAISRVGE
MspBNCORF3798P QMNFVDIGREE.EIRVTFEAENNLDFRLGTGPYRYTGVAREGDLAAISRVGE
RobA2ORF5260 NMNYIDIG.VINQSRVTFEADNNLDFRLGTGALRNTKIASKDDLALITRKSE
β17 α7 η7 β18 C-loop β19 α8
310 320 330 340 350
R.BfiI SDYELRIIKQGTPEHSQLDPYAVSFICNRGKRFGYISNEEFGRIIGVTF...
R.BmrI SDYELRIIKKGSSNYDALDSAAVNFICNRGKRYGYIPNDEFGRIIGAKF...
AbaSDFORF13P RDYELRVFKQGTAEYGALFPYAVNMICHQGKQYGYISNADFARVIGVPK(7)VV
MspBNCORF3798P SRYELRIYRQGTDLFRRLSPYAINFICHQGKQYGFISNQEFQDAIGTRI(4)S.
RobA2ORF5260 FDYELRIIKQTDASYSALARYAVNYICNYGKRFGYISNSDVSSILGVSI...
active site
contact to base
contact to phosphate
amide contact to phosphate
R.BfiI secondary structure: PDB ID 2c1l
β1 N-terminal domain
α6 C-terminal domain

Figure 45. Protein sequence alignment of REase BfiI family. Three sequences of full-length BfiI homologues from putative restriction modification systems were obtained from REBASE and their entry codes are shown at the beginning of each sequence. Numbers above alignment correspond to the BfiI sequence and BfiI secondary structure elements [PDB ID 2C1L] are shown above the alignment, BfiI-N colored in blue, BfiI-C in magenta. Additionally, above alignment as green and orange stripes were marked recognition N- and C-arms of BfiI-C. Symbols above the alignment designate the residue function in EcoRII restriction enzyme. Stars denotes active site residues; downward black triangles stands for residues that contact DNA bases; ovals and upward triangles denotes residues that makes nonspecific contacts to DNA main chain or side chain, correspondingly.

4.8. Analysis of Bse634I interactions with DNA

4.8.1 Crystallization and crystals of Bse634I-DNA complexes

Multiple attempts to obtain crystals of the wt Bse634I-DNA complex were unsuccessful therefore crystallization trials of the mutational Bse634I variants were carried out. Co-crystallization of the Bse634I R226A mutant with three different oligoduplexes GC-1, AT-1 and AT-2 (Table 5) were successful. Bse634I restriction enzyme is arranged as a tetramer comprised of two primary dimers and requires two DNA copies for its optimal activity [10]. The Arg226 residue is located at the tetramerization interface of Bse634I and makes an intricate network of interactions with the amino acid residues of two neighboring subunits. Its replacement by alanine disrupts the Bse634I tetramer into monomers in solution and impairs the Bse634I cleavage at low protein concentrations [166].

Two different crystal forms of the R226A mutant in complex with DNA were obtained. Data collection and refinement statistics is presented in the (Table 8). The crystals in the $P2_12_12$ spacegroup were produced with the cognate oligoduplexes GC-1 and AT-1 (Table 5) representing both palindromic variants of the Bse634I recognition sequence GCCGGC and ACCGGT, respectively. The crystallographic asymmetric unit in both crystals contains two protein monomers and two DNA chains. The primary dimer is generated

by rotation of the protein and DNA monomer around the crystallographic two-fold axis.

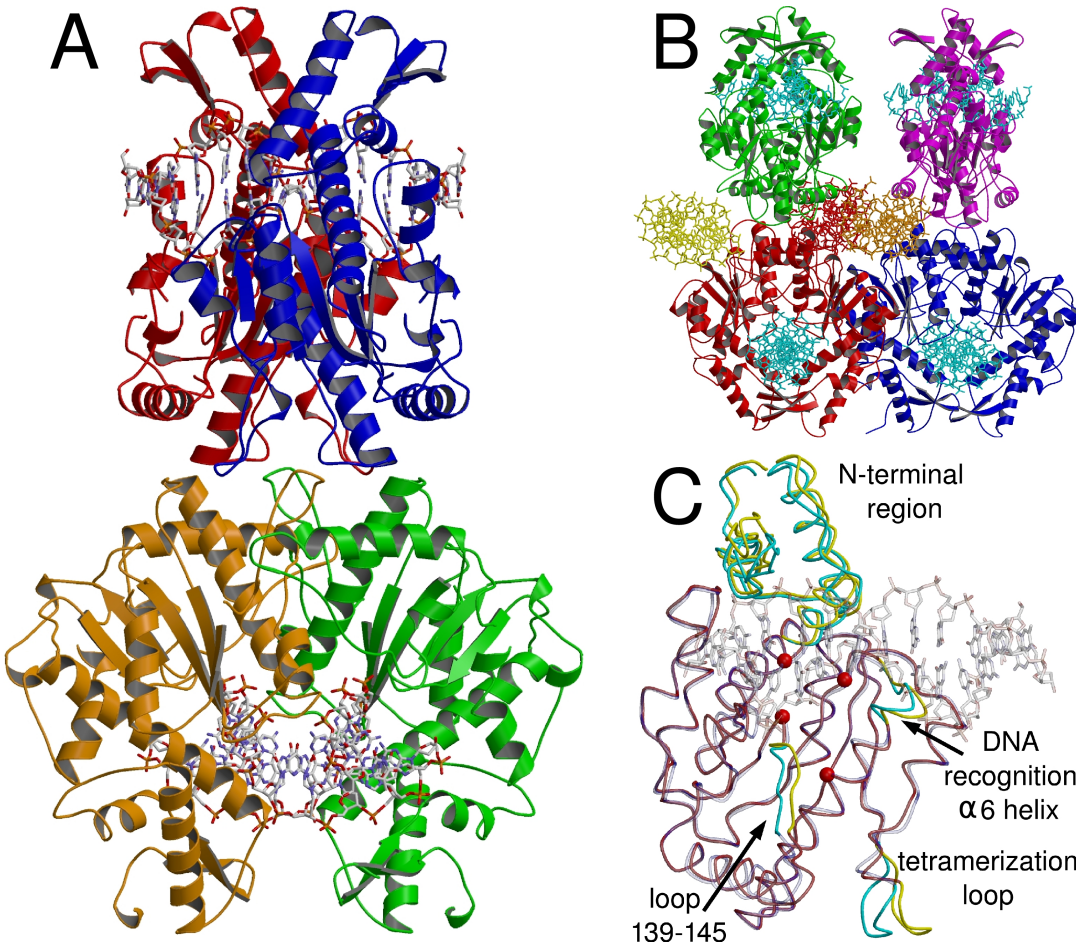


Figure 46. The crystal structure of Bse634I(R226A)-DNA complex. (A) General architecture of the Bse634I-R226A tetramer (cartoon) bound to the GC-1 oligoduplex and (B) the asymmetric unit of $P2_1$ crystals obtained with AT-2. In different colors are shown monomers in (A) and primary dimers in (B). DNA molecules bound by Bse634I are in a ball-and-stick representation, in (B)"out-of-site" DNA duplexes are in yellow, red and orange. (C) Superposition of DNA-free Bse634I (PDB ID 1KNV) monomer A (transparent blue) with the monomer A (red) in the Bse634I-GC-1 complex. Protein chains are traced by Cα. Protein chain fragments (ribbons) that differ in both structures are cyan and yellow for apo- and DNA bound- monomers, respectively. Cα atoms of the catalytic motif (Glu80, Asp146, Lys198 and Glu212) are shown as red spheres.

The second crystal form in the $P2_1$ spacegroup was obtained with the AT-2 oligoduplex, which differs from the AT-1 oligoduplex by the flanking sequence (Table 5). The asymmetric unit in the crystal consists of eight protein subunits that build up two tetramers (Figure 46B). Two oligoduplexes are bound within the DNA-binding clefts of each Bse634I tetramer. Interestingly, three

additional “out-of-site” duplexes fill up the space between the protein tetramers and form the continuous quasi-infinite DNA strands, each built by the repetition of three crystallographically independent oligoduplexes.

4.8.2 Overall architecture of Bse634I-DNA complexes

In both crystal forms two primary dimers are stacked back-to-back to each other and form a tetramer (Figure 46A). Each primary dimer binds one DNA duplex, therefore the tetramer binds two DNA molecules simultaneously. In all three R226A-DNA complex structures the tetramer arrangement is very similar to that of the Bse634I in the DNA-free form [10]. Thus, although R226A substitution disrupts the Bse634I tetramer into monomers in solution at low protein concentrations, in the crystal the mutant protein is assembled into the tetramer similar to the wt enzyme.

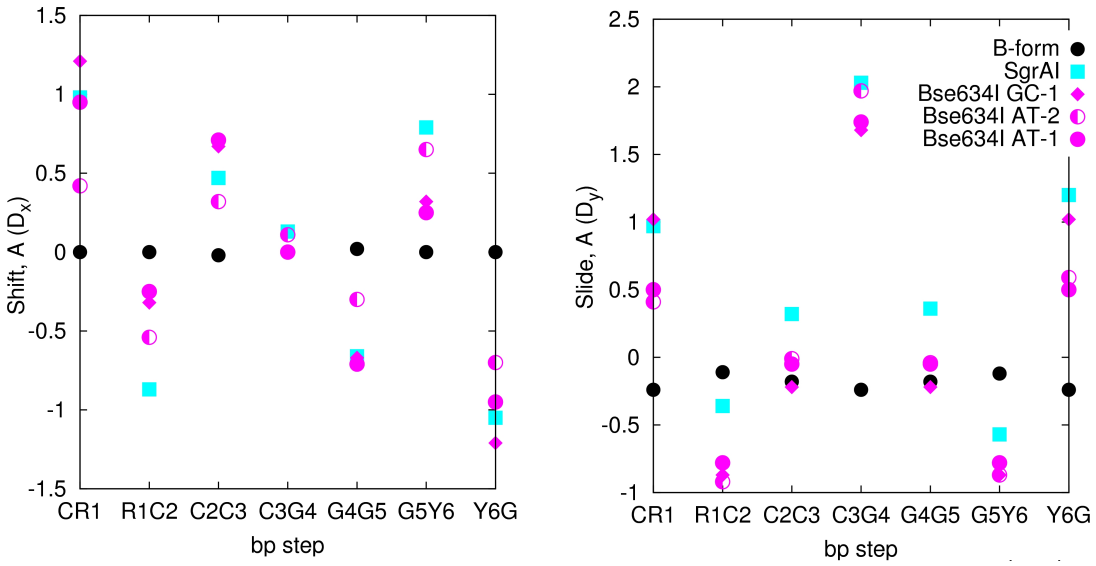


Figure 47. Plots of local base pair step parameters of DNA in Bse634I-R226A and SgrAI [PDB ID 3DVO] [12] complexes (calculated by x3dna [200]). Corresponding labels are displayed within graphs.

Bse634I undergoes relatively small changes upon DNA binding (Figure 46C). First, the N-terminal part (residues 23-89) of the Bse634I monomer which is flexible in the apo-structure [10], moves closer to the DNA and becomes fixed by the DNA backbone contacts. The DNA binding also induces a slight shift of the protein segment (residues 139-145) in the vicinity

of the active site residue Asp146 and upstream the α6 “recognition” helix (residues 202-205) which bears the residues involved in the central CCGG core recognition. The tetramerization loop (residues 258-264), which makes major contacts between the primary dimers within the tetramer, also shows slightly different conformations in the DNA bound- and apo-structures.

The DNA duplex bound to Bse634I in all three structures is in the B-form conformation (Figure 46A). Unpaired thymines at the 5'-end of the oligoduplex are highly flexible and poorly resolved. According to CURVES [179] analysis DNA in the complex with the R226A mutant shows a slight 11-18° bending towards the minor groove (the bending angle value varies depending on the individual structure). The most profound kink occurs at the central dinucleotide step of the site RCCGGY (underlined) coinciding with the dyad axis of oligonucleotide. The outer base pair R1:Y6 is distorted in the Bse634I-DNA complex: the Y6 pyrimidine base is displaced from the minor groove by Asn73 and shifted towards the major groove relatively to the first flanking base pair (Figure 47). This results in the distortion of stacking between Y6 base and the first base outside of the recognition sequence.

4.8.3 Bse634I active center organization

The active center of Bse634I is formed by Glu80, Asp146, Lys198 and Glu212 residues (Figure 48) which (except of Glu80) follow a sequence motif $PDX_{46\text{-}53}KX_{13}E$ characteristic for the REase superfamily. The Bse634I active site is structurally similar to those of NgoMIV [9], and SgrAI enzymes [12] (Figure 48). Indeed, Cα atoms of the catalytic residues nearly coincide between all three enzymes. However, the side chain residues of Glu212 are not well resolved in the individual Bse634I subunits. This fact, probably, reflects an incompleteness of the active center assembly, whether due to the mutation, or due to a relatively low pH of the crystallization mixture.

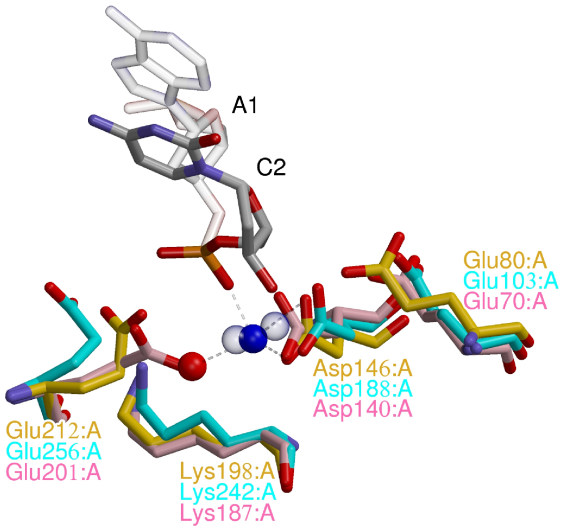


Figure 48. Superposition of the catalytic center residues of the in the Bse634I-R226A-AT-1 complex (gold) and corresponding residues in NgoMIV-DNA [PDB ID 4ABT, pink] and SgrAI-DNA [PDB ID 3DVO, cyan] complexes. Ca^{2+} is shown as a blue sphere, a water molecule as a red sphere. Ca^{2+} in the active sites of NgoMIV and SgrAI are shown as blue transparent spheres. Nucleotides of scissile phosphodiester bond are shown as gray sticks.

Calcium ion was included in all crystallization mixtures, because it is a necessary prerequisite for the DNA binding by Bse634I [201]. Nevertheless, only in the crystal structure of Bse634I with AT-1 oligoduplex an electron density that could be asserted to calcium was found in a close vicinity of the scissile phosphate. The octahedral coordination sphere of the metal ion is incomplete and only four ligands could be identified. The long axis of octahedron is built up by a phosphate oxygen of C2 nucleotide and a carbonyl oxygen of Leu197. Other two ligands are carboxyl oxygens of Asp146 and a water molecule (Figure 48).

4.8.4 Bse634I contacts to the conserved CCGG core

Two symmetry-related α6 helices, dubbed recognition helices, protrude into the major groove of DNA, as it was predicted analyzing the apo-structure [11]. The side chain atoms of the amino acid residues Arg202, Asp204 and Arg205 located at the N-terminus of the α6 helix and at the upstream loop are engaged in the sequence-specific hydrogen bond interactions with the donor and acceptor atoms of the CCGG bases. One protein subunit makes contacts in the major groove with one half-site of the symmetric CCGG sequence (Figure 49A).

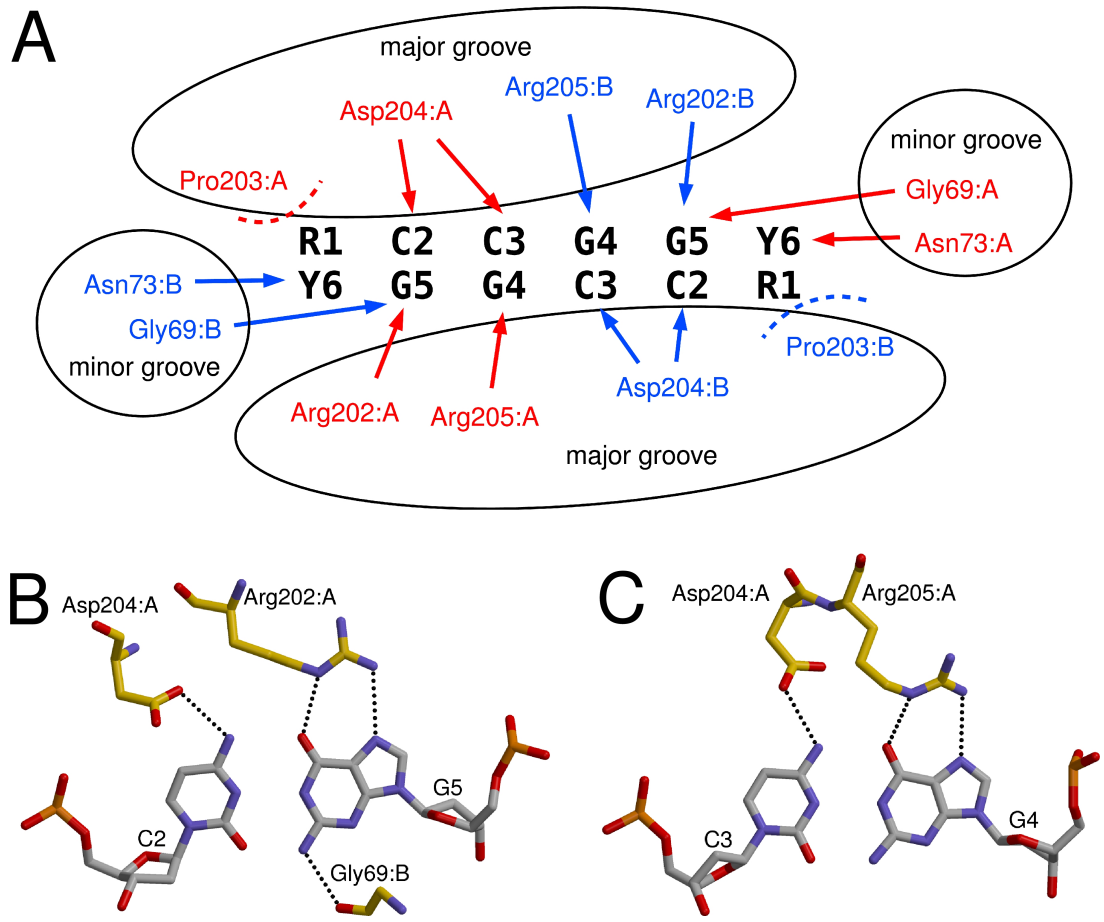


Figure 49. Specific interactions of Bse634I with the recognition sequence. (A) Schematic representation of the degenerate target site and interactions between the Bse634I and its target site. Residues belonging to monomers A and B are shown in red and blue, respectively. Hydrogen bonds are indicated by arrows, van der Waals interactions to purines as dashed lines. (B) Bse634I interactions with the C2:G5 base pair. (C) Bse634I interactions with the C3:G4 base pair.

More specifically, the Nε and Nη2 atoms of the Arg202 and Arg205 residues make bidentate hydrogen bonds to the O6 and N7 atoms of G5 and G4 bases, respectively, while carboxyl oxygens of Asp204 bridge N4 atoms of the neighboring cytosines C2 and C3 (Figure 49B and C). The Nη1 atom of Arg205 also makes a direct hydrogen bond to the main chain oxygen atom of Ser200 and through the bridging water molecule interacts with the catalytic Lys198 residue (not shown). In the minor groove the main chain oxygen atom of Gly69 is within a hydrogen bond distance (2.9-3.1 Å depending on the particular subunit) to the N2 amino group of the G5 base (Figure 49B).

Structural comparison of the enzymes in the DNA-bound forms confirmed that Bse634I, SgrAI [12] and NgoMIV [9] share a conserved mechanism for the recognition of the CCGG core within their target sites. The key contacts to the CCGG bases (Figure 49B and C) are made by the conserved stretch of RXDR residues located at the N-terminal tip of the recognition helices and are identical between Bse634I, NgoMIV and SgrAI.

4.8.5 Bse634I interactions with an outer R1:Y6 base pair

The recognition pattern of the conserved CCGG sequence by Bse634I is nearly identical to that of NgoMIV [9]. However, in contrast to NgoMIV which strictly prefers the G:C base pair outside the CCGG core, Bse634I tolerates both G:C and A:T within its target site. On the other hand the target site of Bse634I (RCCGGY) overlaps with the central part of SgrAI site CRCCGGYG (underlined). The crystal structures of Bse634I in the complex with the oligoduplexes containing either GCCGGC (GC-1, Table 5) or ACCGGT sites (AT-1 and AT-2, Table 5) reveal the protein interaction networks with two alternative target sequences. In both structures A1:T6 and G1:C6 base pairs are contacted by the Bse634I dimer from both minor and major groove sides (Figure 50A and B) and both monomers contribute the amino acid residues for the outer base pair recognition (Figure 49A). In the major groove Pro203 is in the van der Waals contact distance both to A1 and G1 bases (Figure 50A and B). In the minor groove the Nδ atom of Asn73 residue coming from the neighboring monomer makes a hydrogen bond to the O2 atom of T6 and C6 pyrimidines.

In contrast to Bse634I, NgoMIV makes an intricate network of hydrogen-bond to the outer G:C base pair (Figure 50D) both in the major and minor grooves. These interactions unambiguously specify the G:C base pair. On the other hand, SgrAI makes no direct contacts with the degenerate R:Y base pair [12] except of the putative water mediated hydrogen-bond between Ser247 and N7 atom of purine (Figure 50C). The Cα atom of Lys96 occupies a

structurally equivalent position of Asn73 in Bse634I, however hydrogen-bond to the O2 oxygen of thymine in the minor groove is missing.

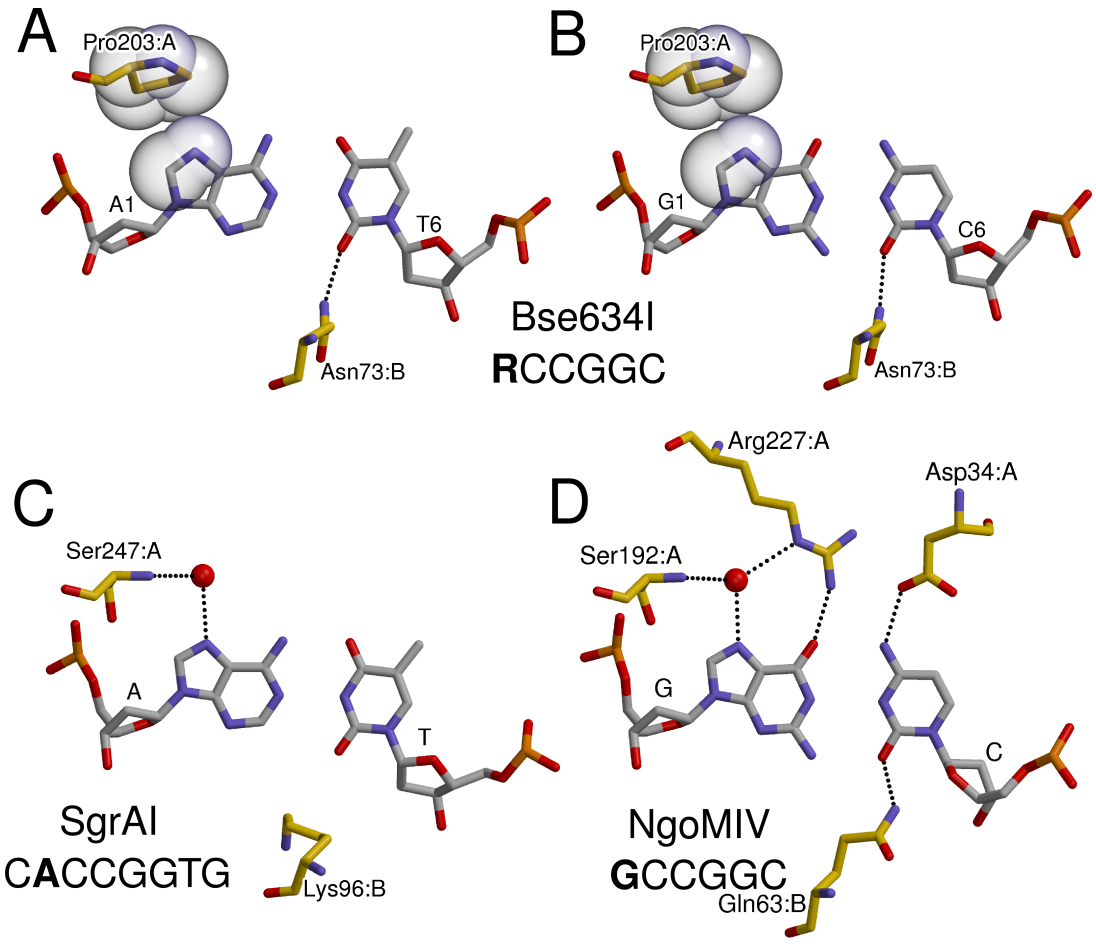


Figure 50. (A, B) Contacts of Bse634I to the outer (A) A1:T6 base pair in the AT-1 complex and (B) to the outer G1:C6 base pair in the GC-1 complex. C8 and N7 atoms of purine R1 base within RCCGGY site and side chain atoms of Pro203 residue are shown in CPK representation. For comparison are presented (C) SgrAI interactions with the A:T base in complex with the oligoduplex containing the cognate 8 bp site CACCGGTG [PDB ID 3DVO] and (D) NgoMIV interactions with the G:C base pair in the 6 bp target site GCCGGC [PDB ID 4ABT].

Thus, the direct contacts to the degenerate base made by SgrAI are restrained similarly to Bse634I. Moreover, in the SgrAI-DNA complex the DNA is distorted similarly to Bse634I. The side chain of Lys96 contacts the outer G (underlined) of the SgrAI recognition sequence CRCCGGYG from the minor groove and unstacks this base from the Y that just precedes it [12].

Thus, Bse634I seems to combine both direct and indirect mechanisms to discriminate the RCCGGY sites from mis-cognate YCCGGR sequences.

However it remains to be determined to what extent the specific structural features of the RY helical step [202][203][204] contribute to the discrimination of the degenerate base pair. The direct interactions are limited by van der Waals contact to the purine base made by Pro203 in the major groove and a hydrogen-bond made to the complementary pyrimidine by the Asn73 in the minor groove. *In silico* mutagenesis suggests that the R1 purine replacement by the T base should introduce a steric clash of the C5 methyl group with the Pro203 residue and interfere with the Bse634I interaction with the TCCGGA site.

4.8.6 Bse634I binding and cleavage studies

To probe the importance of the van der Waals interactions in the discrimination of the degenerate R:Y base pair were performed DNA binding and cleavage by Bse634I studies using a set of the oligoduplexes containing cognate (ACCGGT or GCCGGC), mis-cognate (TCCGGA or CCCGGG) and modified cognate and mis-cognate sites (Table 5). In the modified substrates the outer T is replaced by U (uracil), which lacks the 5-methyl group in the pyrimidine ring, and G is replaced by I (inosine), which lacks 2-amino group in the purine ring. Flanking sequences in all oligoduplexes used in these experiments were identical (Table 5). The electrophoretic mobility shift assay (see, section 3.15.) revealed that Bse634I binds the cognate oligoduplexes AT and GC (Figure 51A and C) at the low protein concentrations. Further increase of the Bse634I concentration gives rise to the second slowly migrating protein-DNA complex. In the case of the non-cognate oligoduplex NS (Figure 51I) only this slowly migrating complex is observed at the increased protein concentrations.

Based on the DNA binding patterns presented here and in the previous studies [201] we conclude that the complexes with the cognate DNA observed at low protein concentrations represent the specific complexes where Bse634I is bound to the recognition site while lower mobility complexes at increased protein concentrations correspond to non-specific complexes.

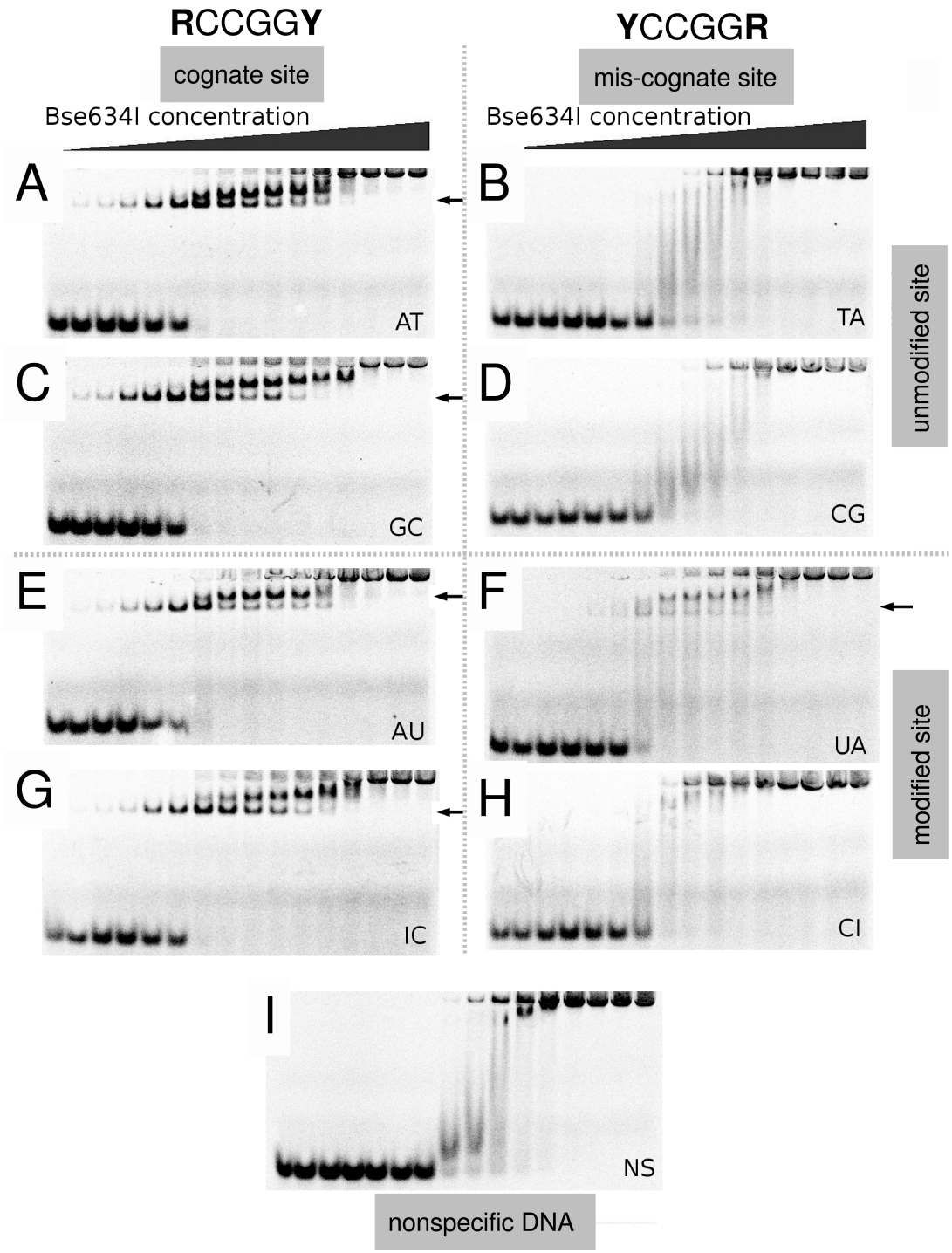


Figure 51. Binding of specific and mis-cognate DNA oligoduplexes by the wt Bse634I. Labels shown on the picture indicate oligoduplex (see, Table 5) used in the gel-shift experiment. I stands for Inosine; U – for Uridine; NS – duplex without CCGG site. Concentration of Bse634I in samples is increasing from 0.1 nM to 5 μM (in terms of monomer) (see, section 3.15.). The first lane in all gels contains no protein. Specific or specific-like Bse634I-DNA complexes are designated by arrows.

The cognate oligoduplexes containing U instead of T (AU) and I instead of G (IC) showed the DNA binding pattern similar to that for the specific

oligoduplexes GC and AT indicating that the elimination of the 5-methyl group in thymine or 2-amino group in guanine in the outer base pair has no effect on the Bse634I binding (compare panels E and G in Figure 51). The crystal structures are consistent with this observation since there are no direct or indirect contacts with methyl or amino groups in the Bse634I complexes with cognate DNA. On the other hand, the oligoduplexes TA (Figure 51B) and CG (Figure 51D) containing mis-cognate sites are bound by the protein similarly to the non-cognate DNA (Figure 51C). However, T to U replacement in the mis-cognate oligoduplex UA results in the different binding pattern (compare panels F and B in Figure 51). Indeed, the band corresponding to a specific protein-DNA complex appears in the gel, albeit at the higher protein concentrations in comparison to the cognate AT site. The modified CI oligoduplex forms only a non-specific protein DNA complex, similarly to the mis-cognate CG duplex (Figure 51H and D).

On the next step we analyzed Bse634I cleavage of cognate, mis-cognate and modified oligoduplexes. The cleavage reaction (see, section 3.16.) was carried out under single turnover conditions using an optimal enzyme concentration that yields a maximal cleavage rate.

Table 12. Rate constants of oligoduplex cleavage by Bse634I

Oligoduplex	**k, (min^{-1})**	
	at 25°C	**at 50°C**
AT	10±1.3	-
GC	4.1±0.9	-
AU	5.4±1.3	-
IC	13±2.9	-
TA	*no cleavage**	$5.7\pm0.7\times10^{-5}$
CG	*no cleavage*	$8.2\pm0.8\times10^{-5}$
UA	*no cleavage*	$6.9\pm0.8\times10^{-2}$
CI	*no cleavage*	$3.8\pm0.9\times10^{-5}$
NS	*no cleavage*	*no cleavage*

* no cleavage products were observed after 48 hours of incubation.

The cleavage rates for the cognate and the modified cognate oligoduplexes AT, AU, GC and IC are within the range 4-13 min^{-1} at 25°C (Table 12) indicating that modifications in the cognate AU and IC duplexes have no significant effect on the cleavage activity. However, in the case of the unmodified and the modified mis-cognate duplexes (TA, CG, UA and CI) the cleavage rates at 25°C were too slow to be reliably measured. Therefore, the Bse634I cleavage of TA, CG, UA and CI oligoduplexes was studied at 50°C taking an advantage of the fact that Bse634I is a thermophilic enzyme [205]. Under these conditions the cleavage rate of the cognate oligoduplex was too high to be measured by the conventional techniques. On the other hand, the mis-cognate TA and CG duplexes even at 50°C were hydrolyzed very slowly (~20% of the oligoduplex was hydrolyzed after 48 hours of incubation). However, under these conditions the Bse634I cleavage rate of the modified mis-cognate UA duplex was increased nearly by three orders of magnitude compared to the TA oligoduplex. The cleavage rate of the CI duplex was very close to that of the unmodified CG and TA oligonucleotides (Table 12).

Taken together, these data indicate that the steric clash between the C5 methyl group of T and Pro203 residue plays an important role in the R vs T discrimination by Bse634I. However, the Bse634I preference for R against T in the first position of the target site cannot be fully explained by a disruption of the favored van der Waals contact between Pro203 and the purine base. Cleavage rate of the UA oligoduplex containing the UCCGGA site is still much lower than of the cognate duplex. Our present hypothesis is that the Bse634I selection against CCCGGG (and to some extent against the TCCGGA site) is achieved by an indirect readout mechanism due to a certain combination of distortions in DNA conformation.

CONCLUSIONS

1. The C-terminal catalytic domain dimer of EcoRII flips out the central nucleotide pair in the pseudosymmetric 5'-CCWGG site transforming it into the symmetric 5'-CCGG sequence. Extrusion of the central base pair enables EcoRII-C to use a conserved mechanism for the recognition of CC:GG dinucleotides.

2. The N-terminal allosteric domain of EcoRII binds to the 5'-CCWGG site as a monomer in a single orientation determined by the specific interactions with the central T:A pair in the major groove.

3. The structural mechanism of the sequence recognition employed by the EcoRII-N and BfiI-C domains is shared by plant transcription factors of the B3 family.

4. Bse634I restriction enzyme (R226A) mutant employs conserved structural elements for the recognition of the CCGG core in the 5'-RCCGGY site and uses both direct and indirect readout mechanisms for discrimination of the degenerated R:Y base pair.

LIST OF PUBLICATIONS

The thesis is based on the following original publications:

1. **Golovenko D**., Manakova E., Tamulaitienė G., Gražulis S. and Šikšnys V. (2009) Structural mechanisms for the 5'-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII. *Nucleic Acids Res* 37:6613-6624
2. Manakova E., Gražulis S., Zaremba M., Tamulaitienė G., **Golovenko D**. and Šikšnys V. (2012) Structural mechanisms of the degenerate sequence recognition by Bse634I restriction endonuclease. *Nucleic Acids Res* 40:6741-6751

The material of the thesis was presented in these conferences:

1. **Golovenko D.**, Manakova E., Gražulis S. and Šikšnys V. (2008) Crystallographic studies of EcoRII-DNA complex: one enzyme – two DNA recognition mechanisms. The 10th International EMBL PhD Symposium, Heidelberg, Germany; October 23-25.
2. **Golovenko D.**, Manakova E., Gražulis S. and Šikšnys V. (2010) Structural mechanisms for the DNA recognition by restriction endonuclease EcoRII. The 6th New England BioLabs Meeting on DNA Restriction and Modification, Bremen, Germany; August 1-6.

Other publications:

1. Čapkauskaitė E., Baranauskienė L., **Golovenko D.**, Manakova E., Gražulis S., Tumkevičius S. and Matulis D. (2010) Indapamide-like benzenesulfonamides as inhibitors of carbonic anhydrases I, II, VII, and XIII. *Bioorg Med Chem* 18:7357-7364
2. Sūdžius J., Baranauskienė L., **Golovenko D.**, Matulienė J., Michailovienė V., Torresan J., Jachno J., Sukackaitė R., Manakova E., Gražulis S., Tumkevičius S. and Matulis D. (2010) 4-[N-(substituted 4-pyrimidinyl)amino]benzenesulfonamides as inhibitors of carbonic anhydrase isozymes I, II, VII, and XIII. *Bioorg Med Chem* 18:7413-7421
3. Baranauskienė L., Hilvo M., Matulienė J., **Golovenko D.**, Manakova E., Dudutienė V., Michailovienė V., Torresan J., Jachno J., Parkkila S., Maresca A., Supuran C.T., Gražulis S. and Matulis D. (2010) Inhibition and binding studies of carbonic anhydrase isozymes I, II and IX with benzimidazo[1,2-c][1,2,3]thiadiazole-7-sulphonamides. *J Enzyme Inhib Med Chem* 25:863-870

ACKNOWLEDGEMENT

The financial support from the Max Planck Society, Research Council of Lithuania and MoBiLi project is gratefully acknowledged.

I thank dr. Saulius Gražulis for encouraging me to start my studies in macromolecular crystallography, for his extensive supervision operating X-ray machines/facilities and open source software and sincere open communication, also for his invaluable suggestions preparing this doctoral thesis and summary of the thesis.

I would like to express my appreciation to prof. Virginijus Šikšnys for providing excellent research objects, invaluable discussions and suggestions, help with preparation of publications and this doctoral thesis.

Special thanks to dr. Elena Manakova for making available the crystals of DNA complexed EcoRII-C, for crystallizing Bse634I-R226A complexes with DNA as well as for gel-shift and kinetics experiments on Bse634I. Her critical comments on this doctoral thesis are very appreciated.

I thank drs. Giedrė Tamulaitienė and Matthias Bochtler (International Institute of Molecular and Cell Biology, Warsaw, Poland) for their assistance during beam-time. I am grateful to drs. Giedrius Sasnauskas and Gintautas Tamulaitis who guided me in the kinetic and gel-shift studies. I would like to thank dr. Mindaugas Zaremba for suggestions performing limited proteolysis experiments and all my colleagues in Vilnius for discussions, comments and suggestions.

I am grateful to prof. Monika Reuter (Institute of Virology at Humboldt University, Berlin, Germany) for providing expression vectors of EcoRII-N and EcoRII-Y41A constructs. I am grateful to drs. Arūnas Lagunavičius (“Thermo Fisher Scientific”, Vilnius) and Violeta Morin (University of Konstanz, Baden-Württemberg, Germany) for constructing and providing plasmid carrying BfiI methylases and expression vector of BfiI-K107A.

I thank drs. Matthew Groves, Gleb Bourenkov and Manfred S. Weiss at EMBL beamlines X11, X12 and X13 at the DORIS storage ring, Hamburg for the invaluable help with beamline operation.

I would like to acknowledge suggestions and help with structure refinement software provided by dr. Garib Murshudov (MRC Laboratory of Molecular Biology, Cambridge, UK) during the workshop at International School of Crystallography (Erice, Italy) as well as provided by dr. Haim Rozenberg (Weizmann Institute of Science, Rehovot, Israel) during the visit at Weizmann Institute of Science (Rehovot, Israel).

My greatest gratitude belongs to my parents for their concern, immense support and understanding.

REFERENCES

1. Pingoud A, Fuxreiter M, Pingoud V, Wende W (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, **62**:685-707
2. Pingoud A JA (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res*, **29**:3705-3727
3. Roberts RJ, Vincze T, Posfai J, Macelis D (2010) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*, **38**:D234-6
4. Orlowski J BJ (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res*, **36**:3552-3569
5. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol*, **10**:980
6. Zaremba M, Owsicka A, Tamulaitis G, Sasnauskas G, Shlyakhtenko LS et al. (2010) DNA synapsis through transient tetramerization triggers cleavage by Ecl18kI restriction enzyme. *Nucleic Acids Res*, **38**:7142-7154
7. Mizuno H, Suzuki T, Akagawa M, Yamasato K, Yamada Y (1990) Purification, properties and determinations of recognition sequence and cleavage site of restriction endonuclease from "Agrobacterium gelatinovorum" IAM 12617, a marine bacterium (AgeI). *Agric Biol Chem*, **54**:1797-1802
8. Lubys A, Jurenaite S, Janulaitis A (1999) Structural organization and regulation of the plasmid-borne type II restriction-modification system Kpn2I from Klebsiella pneumoniae RFL2. *Nucleic Acids Res*, **27**:4228-4234
9. Deibert M, Grazulis S, Sasnauskas G, Siksnys V, Huber R (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat Struct Biol*, **7**:792-799
10. Bozic D, Grazulis S, Siksnys V, Huber R (1996) Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 A resolution. *J Mol Biol*, **255**:176-186
11. Grazulis S, Deibert M, Rimseliene R, Skirgaila R, Sasnauskas G et al. (2002) Crystal structure of the Bse634I restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res*, **30**:876-885
12. Dunten PW, Little EJ, Gregory MT, Manohar VM, Dalton M et al. (2008) The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res*, **36**:5405-5416
13. Szczepanowski RH, Carpenter MA, Czapinska H, Zaremba M, Tamulaitis G et al. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res*, **36**:6109-6117
14. Zhou XE, Wang Y, Reuter M, Mucke M, Kruger DH et al. (2004) Crystal structure of type IIE restriction endonuclease EcoRII reveals an autoinhibition

mechanism by a novel effector-binding fold. *J Mol Biol*, **335**:307-319
15. Tamulaitis G, Sasnauskas G, Mucke M, Siksnys V (2006) Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease. *J Mol Biol*, **358**:406-419
16. Bochtler M, Szczepanowski RH, Tamulaitis G, Grazulis S, Czapinska H et al. (2006) Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J*, **25**:2219-2229
17. Gaigalas M, Maneliene Z, Kazlauskiene R, Petrusyte M, Janulaitis A (2002) PfoI, a unique type II restriction endonuclease that recognises the sequence 5'-T downward arrow CCNGGA-3'. *Nucleic Acids Res*, **30**:e98
18. Pein CD, Reuter M, Meisel A, Cech D, Kruger DH (1991) Activation of restriction endonuclease EcoRII does not depend on the cleavage of stimulator DNA. *Nucleic Acids Res*, **19**:5139-5142
19. Vitkute J, Maneliene Z, Petrusyte M, Janulaitis A (1998) BfiI, a restriction endonuclease from Bacillus firmus S8120, which recognizes the novel non-palindromic sequence 5'-ACTGGG(N)5/4-3'. *Nucleic Acids Res*, **26**:3348-3349
20. Sapranauskas R, Sasnauskas G, Lagunavicius A, Vilkaitis G, Lubys A et al. (2000) Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of Salmonella typhimurium. *J Biol Chem*, **275**:30878-30885
21. Grazulis S, Manakova E, Roessle M, Bochtler M, Tamulaitiene G et al. (2005) Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc Natl Acad Sci U S A*, **102**:15797-15802
22. Lagunavicius A, Sasnauskas G, Halford SE, Siksnys V (2003) The metal-independent type IIs restriction enzyme BfiI is a dimer that binds two DNA sites but has only one catalytic centre. *J Mol Biol*, **326**:1051-1064
23. Stuckey JA DJ (1999) Crystal structure of a phospholipase D family member. *Nat Struct Biol*, **6**:278-284
24. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T et al. (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell*, **16**:3448-3459
25. Zaremba M, Urbanke C, Halford SE, Siksnys V (2004) Generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. *J Mol Biol*, **336**:81-92
26. Silanskas A, Zaremba M, Sasnauskas G, Siksnys V (2012) Catalytic activity control of restriction endonuclease--triplex forming oligonucleotide conjugates. *Bioconjug Chem*, **23**:203-211
27. Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. *Trends Plant Sci*, **13**:647-655
28. Tock MR DD (2005) The biology of restriction and anti-restriction. *Curr*

Opin Microbiol, **8**:466-472
29. Nelson, M., Raschke, E. and McClelland, M. (1993) Effect of site-specific methylation on restriction endonucleases and dna modification methyltransferases. *Nucleic Acids Res.*, **21**:3139-3154
30. Murray NE (2002) 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**:3-20
31. Jeltsch A PA (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J Mol Evol*, **42**:91-96
32. Furuta K (2011).. In: Restriction-Modification Systems As Mobile Epigenetic Elements, Volume , Edition, ed. Roberts M, e., ed. (Landes bioscience), .
33. Handa,N. and Kobayashi,I. (1999) Post-segregational killing by restriction modification gene complexes: observations of individual cell deaths. *Biochimie*, **81**:931-938
34. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*, **31**:1805-1812
35. Kosinski J, Feder M, Bujnicki JM (2005) The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics*, **6**:172
36. Kim YC, Grable JC, Love R, Greene PJ, Rosenberg JM (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**:1307-1309
37. Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS et al. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J*, **12**:1781-1795
38. Miyazono K, Watanabe M, Kosinski J, Ishikawa K, Kamo M et al. (2007) Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res*, **35**:1908-1918
39. Sokolowska M, Czapinska H, Bochtler M (2009) Crystal structure of the beta beta alpha-Me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res*, **37**:3799-3810
40. Mak AN, Lambert AR, Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure*, **18**:1321-1331
41. Niv MY, Ripoll DR, Vila JA, Liwo A, Vanamee ES et al. (2007) Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res*, **35**:2227-2237
42. Sokolowska M, Czapinska H, Bochtler M (2011) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res*, **39**:1554-1564

43. Venclovas C, Timinskas A, Siksnys V (1994) Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins*, **20**:279-282
44. Laganeckas M, Margelevicius M, Venclovas C (2011) Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. *Nucleic Acids Res*, **39**:1187-1196
45. van der Woerd MJ, Pelletier JJ, Xu S, Friedman AM (2001) Restriction enzyme BsoBI-DNA complex: a tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure*, **9**:133-144
46. Horton JR, Mabuchi MY, Cohen-Karni D, Zhang X, Griggs RM et al. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res*, :
47. Tamulaitiene G, Jakubauskas A, Urbanke C, Huber R, Grazulis S et al. (2006) The crystal structure of the rare-cutting restriction enzyme SdaI reveals unexpected domain architecture. *Structure*, **14**:1389-1400
48. Lukacs CM, Kucera R, Schildkraut I, Aggarwal AK (2000) Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 A resolution. *Nat Struct Biol*, **7**:134-140
49. Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**:656-663
50. Skirgaila R, Grazulis S, Bozic D, Huber R, Siksnys V (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J Mol Biol*, **279**:473-481
51. Horton NC PJ (2004) DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, **43**:6841-6857
52. Belfort M, Stoddard B, Wood D, Derbyshire V (2008).Homing endonucleases and inteins, Volume , Edition (Springer).
53. Galburt EA, Chevalier B, Tang W, Jurica MS, Flick KE et al. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat Struct Biol*, **6**:1096-1099
54. Biertümpfel C, Yang W, Suck D (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature*, **449**:616-620
55. Ko TP, Liao CC, Ku WY, Chak KF, Yuan HS (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure*, **7**:91-102
56. Pommer AJ, Cal S, Keeble AH, Walker D, Evans SJ et al. (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J Mol Biol*, **314**:735-749
57. Shen BW, Heiter DF, Chan S, Wang H, Xu S et al. (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease PacI.

Structure, **18**:734-743
58. Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT et al. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res*, **27**:2115-2125
59. Dunin-Horkawicz S, Feder M, Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**:98
60. Taylor GK SB (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res*, :
61. Truglio JJ, Rhau B, Croteau DL, Wang L, Skorvaga M et al. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J*, **24**:885-894
62. Pohlman RF, Liu F, Wang L, Moré MI, Winans SC (1993) Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101. *Nucleic Acids Res*, **21**:4867-4872
63. Koonin EV (1996) A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem Sci*, **21**:242-243
64. Sasnauskas G, Zakrys L, Zaremba M, Cosstick R, Gaynor JW et al. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3'-5' and 5'-3' strands by rotating a single active site. *Nucleic Acids Res*, **38**:2399-2410
65. Cymerman IA, Meiss G, Bujnicki JM (2005) DNase II is a member of the phospholipase D superfamily. *Bioinformatics*, **21**:3959-3962
66. Chinen A, Naito Y, Handa N, Kobayashi I (2000) Evolution of sequence recognition by restriction-modification enzymes: selective pressure for specificity decrease. *Mol Biol Evol*, **17**:1610-1619
67. Ishikawa K, Watanabe M, Kuroita T, Uchiyama I, Bujnicki JM et al. (2005) Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon Pyrococcus abyssi. *Nucleic Acids Res*, **33**:e112
68. Wah DA, Hirsch JA, Dorner LF, Schildkraut I, Aggarwal AK (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**:97-100
69. Huai Q, Colandene JD, Topal MD, Ke H (2001) Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat Struct Biol*, **8**:665-669
70. Lambert AR, Sussman D, Shen B, Maunus R, Nix J et al. (2008) Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure*, **16**:558-569
71. Tamulaitis G, Mucke M, Siksnys V (2006) Biochemical and mutational

analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett*, **580**:1665-1671
72. Ramalingam, S., Kandavelou, K., Rajenderan, R. et al. (2011) Creating designed zinc-finger nucleases with minimal cytotoxicity. *J. Mol. Biol.*, **405**:630-641
73. Fonfara, I., Curth, U., Pingoud, A. et al. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**:847-860
74. Chan S, Bao Y, Ciszak E, Laget S, Xu S (2007) Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases with novel substrate specificities. *Nucleic Acids Res*, **35**:6238-6248
75. Zhang, P., Bao, Y., Higgins, L. et al. (2007) Rational design of a chimeric endonuclease targeted to noti recognition site. *Protein Eng. Des. Sel.*, **20**:497-504
76. Wakefield RI, Smith BO, Nan X, Free A, Soteriou A et al. (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol*, **291**:1055-1065
77. Fomenkov, A., Too, PH., Chan, S. et al. (2008) Targeting dna 5mcpg sites with chimeric endonucleases. *Anal. Biochem.*, **381**:135-141
78. Bujnicki JM (2001) A model of structure and action of Sau3AI restriction endonuclease that comprises two MutH-like endonuclease domains within a single polypeptide. *Acta Microbiol Pol*, **50**:219-231
79. Xu C, Yu F, Xu S, Ding Y, Sun L et al. (2009) Crystal structure and function of C-terminal Sau3AI domain. *Biochim Biophys Acta*, **1794**:118-123
80. Friedhoff P, Lurz R, Lüder G, Pingoud A (2001) Sau3AI, a monomeric type II restriction endonuclease that dimerizes on the DNA and thereby induces DNA loops. *J Biol Chem*, **276**:23581-23588
81. Gemmen GJ, Millin R, Smith DE (2006) Tension-dependent DNA cleavage by restriction endonucleases: two-site enzymes are "switched off" at low force. *Proc Natl Acad Sci U S A*, **103**:11555-11560
82. Vanamee ES, Santagata S, Aggarwal AK (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol*, **309**:69-78
83. Sugisaki H KS (1981) New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene*, **16**:73-78
84. Li L, Wu LP, Chandrasegaran S (1992) Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A*, **89**:4275-4279
85. Waugh DS SR (1994) A novel class of FokI restriction endonuclease mutants that cleave hemi-methylated substrates. *J Biol Chem*, **269**:12298-12303
86. Pernstich C HS (2012) Illuminating the reaction pathway of the FokI restriction endonuclease by fluorescence resonance energy transfer. *Nucleic Acids Res*, **40**:1203-1213

87. Wah DA, Bitinaite J, Schildkraut I, Aggarwal AK (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**:10564-10569
88. Bitinaite J, Wah DA, Aggarwal AK, Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**:10570-10575
89. Sanders KL, Catto LE, Bellamy SRW, Halford SE (2009) Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res*, **37**:2105-2115
90. Rusling DA, Laurens N, Pernstich C, Wuite GJL, Halford SE (2012) DNA looping by FokI: the impact of synapse geometry on loop topology at varied site orientations. *Nucleic Acids Res*, **40**:4977-4987
91. Bilcock DT, Daniels LE, Bath AJ, Halford SE (1999) Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J Biol Chem*, **274**:36379-36386
92. Conrad M TM (1992) Modified DNA fragments activate NaeI cleavage of refractory DNA sites. *Nucleic Acids Res*, **20**:5127-5130
93. Colandene JD TM (1998) The domain organization of NaeI endonuclease: separation of binding and catalysis. *Proc Natl Acad Sci U S A*, **95**:3531-3536
94. Huai Q, Colandene JD, Chen Y, Luo F, Zhao Y et al. (2000) Crystal structure of NaeI-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J*, **19**:3110-3118
95. Colandene JD TM (2000) Evidence for mutations that break communication between the Endo and Topo domains in NaeI endonuclease/topoisomerase. *Biochemistry*, **39**:13703-13707
96. Boyer HW, Chow LT, Dugaiczyk A, Hedgpeth J, Goodman HM (1973) DNA substrate site for the EcoRII restriction endonuclease and modification methylase. *Nat New Biol*, **244**:40-43
97. Mucke M, Grelle G, Behlke J, Kraft R, Kruger DH et al. (2002) EcoRII: a restriction enzyme evolving recombination functions?. *EMBO J*, **21**:5262-5268
98. Kruger DH RM (2005) Reliable detection of DNA cytosine methylation at CpNpG sites using the engineered restriction enzyme EcoRII-C. *Biotechniques*, **38**:855-856
99. Szczepek M, Mackeldanz P, Möncke-Buchner E, Alves J, Krüger DH et al. (2009) Molecular analysis of restriction endonuclease EcoRII from Escherichia coli reveals precise regulation of its enzymatic activity by autoinhibition. *Mol Microbiol*, **72**:1011-1021
100. Karpova, E. and Pusey, M (1998) Peculiarity of crystallization of restriction endonuclease EcoRII. , **7th ICCBM May 3-8**:p. 96
101. Karpova EA, Meehan E, Pusey ML, Chen L (1999) Crystallization and preliminary X-ray diffraction analysis of restriction endonuclease EcoRII. *Acta Crystallogr D Biol Crystallogr*, **55**:1604-1605

102. Karpova, E., Chen, L, Meehan, E and Pusey, M (2000) Purification of restriction endonuclease EcoRII and its co-crystallization with DNA-substrate. , **Abstracts**:p. 163
103. Zhou EX, Reuter M, Meehan EJ, Chen L (2002) A new crystal form of restriction endonuclease EcoRII that diffracts to 2.8 A resolution. *Acta Crystallogr D Biol Crystallogr*, **58**:1343-1345
104. Zhou XE, Wang Y, Reuter M, Mackeldanz P, Kruger DH et al. (2003) A single mutation of restriction endonuclease EcoRII led to a new crystal form that diffracts to 2.1 A resolution. *Acta Crystallogr D Biol Crystallogr*, **59**:910-912
105. Chen, L., Qiu, L, Szczepek, M, Reuter, M and Meehan, E (2008) Structural Studies of Type IIE Restriction Endonuclease EcoRII-DNA complexes. , **Alabama EPSCoR Annual Conference Abstracts**:p. 56
106. Gilmore JL, Suzuki Y, Tamulaitis G, Siksnys V, Takeyasu K et al. (2009) Single-molecule dynamics of the DNA-EcoRII protein complexes revealed with high-speed atomic force microscopy. *Biochemistry*, **48**:10492-10498
107. Sasnauskas G, Halford SE, Siksnys V (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc Natl Acad Sci U S A*, **100**:6410-6415
108. Sapranauskas, R. and Lubys, A. (2005) Random gene dissection: a tool for the investigation of protein structural organization. *BioTechniques*, **39**:395-402
109. Shen BW, Xu D, Chan S, Zheng Y, Zhu Z et al. (2011) Characterization and crystal structure of the type IIG restriction endonuclease RM.BpuSI. *Nucleic Acids Res*, **39**:8223-8236
110. Qiang BQ SI (1987) NotI and SfiI: restriction endonucleases with octanucleotide recognition sequences. *Methods Enzymol*, **155**:15-21
111. Tamulaitiene G SV (2008) NotI is not boring. *Structure*, **16**:497-498
112. Tamulaitis G, Solonin AS, Siksnys V (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett*, **518**:17-22
113. Pingoud V, Kubareva E, Stengel G, Friedhoff P, Bujnicki JM et al. (2002) Evolutionary relationship between different subgroups of restriction endonucleases. *J Biol Chem*, **277**:14306-14314
114. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**:2105-2110
115. Yamasaki K, Kigawa T, Inoue M, Watanabe S, Tateno M et al. (2008) Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol Biochem*, **46**:394-401
116. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*, **29**:231-262

117. Kagaya Y, Ohmiya K, Hattori T (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res*, **27**:470-478
118. Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proc Natl Acad Sci U S A*, **94**:7076-7081
119. McCarty DR, Hattori T, Carson CB, Vasil V, Lazar M et al. (1991) The Viviparous-1 developmental gene of maize encodes a novel transcriptional activator. *Cell*, **66**:895-905
120. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun*, **290**:998-1009
121. Magnani, E., Sjölander, K. and Hake, S. (2004) From endonucleases to transcription factors: evolution of the ap2 dna binding domain in plants. *Plant Cell*, **16**:2265-2277
122. Wuitschick JD, Lindstrom PR, Meyer AE, Karrer KM (2004) Homing endonucleases encoded by germ line-limited genes in Tetrahymena thermophila have APETELA2 DNA binding domains. *Eukaryot Cell*, **3**:685-694
123. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res*, **33**:3994-4006
124. Painter, HJ., Campbell, TL. and Llinás, M. (2011) The apicomplexan ap2 family: integral factors regulating plasmodium development. *Mol. Biochem. Parasitol.*, **176**:1-7
125. Ellison EL VV (1993) Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol Cell Biol*, **13**:7531-7539
126. Wittmayer PK, McKenzie JL, Raines RT (1998) Degenerate DNA recognition by I-PpoI endonuclease. *Gene*, **206**:11-21
127. Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**:49-95
128. Argast GM, Stephens KM, Emond MJ, Monnat RJJ (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol*, **280**:345-353
129. Ulmasov T, Hagen G, Guilfoyle TJ (1997) ARF1, a transcription factor that binds to auxin response elements. *Science*, **276**:1865-1868
130. Suzuki M, Kao CY, McCarty DR (1997) The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell*, **9**:799-807
131. Franco-Zorrilla JM, Cubas P, Jarillo JA, Fernández-Calvín B, Salinas J et al. (2002) AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. *Plant Physiol*, **128**:418-

427
132. Wang Y, Deng D, Zhang R, Wang S, Bian Y et al. (2012) Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol Biol Rep*, **39**:6267-6282
133. Wang Y, Deng D, Shi Y, Miao N, Bian Y et al. (2012) Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights gained from analyzing maize ARF genes. *Mol Biol Rep*, **39**:2401-2415
134. Michaels SD AR (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, **11**:949-956
135. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science*, **297**:243-246
136. Mylne JS, Barrett L, Tessadori F, Mesnage S, Johnson L et al. (2006) LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. *Proc Natl Acad Sci U S A*, **103**:5012-5017
137. Santos-Mendoza, M., Dubreucq, B., Baud, S. et al. (2008) Deciphering gene regulatory networks that control seed development and maturation in arabidopsis. *Plant J.*, **54**:608-620
138. Peng FY WR (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC Genomics*, **12**:286
139. Romanel EAC, Schrago CG, Couñago RM, Russo CAM, Alves-Ferreira M (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One*, **4**:e5791
140. Waltner JK, Peterson FC, Lytle BL, Volkman BF (2005) Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci*, **14**:2478-2483
141. Campagne S, Gervais V, Milon A (2011) Nuclear magnetic resonance analysis of protein-DNA interactions. *J R Soc Interface*, **8**:1065-1078
142. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T et al. (2004) A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. *J Mol Biol*, **337**:49-63
143. King G, Hill JM, Martin JL, Mylne JS (2009) Expression, purification and preliminary X-ray diffraction studies of VERNALIZATION1(208-341) from Arabidopsis thaliana. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **65**:291-294
144. Mylne JS, Mas C, Hill JM (2012) NMR assignment and secondary structure of the C-terminal DNA binding domain of Arabidopsis thaliana VERNALIZATION1. *Biomol NMR Assign*, **6**:5-8
145. Waleron K, Waleron M, Osipiuk J, Podhajska AJ, Lojkowska E (2006)

Identification of a DNA restriction-modification system in Pectobacterium carotovorum strains isolated from Poland. *J Appl Microbiol*, **100**:343-351
146. Alva V, Koretke KK, Coles M, Lupas AN (2008) Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. *Curr Opin Struct Biol*, **18**:358-365
147. Chen S, Jancrick J, Yokota H, Kim R, Kim S (2004) Crystal structure of a protein associated with cell division from Mycoplasma pneumoniae (GI: 13508053): a novel fold with a conserved sequence motif. *Proteins*, **55**:785-791
148. Coles M, Djuranovic S, Söding J, Frickey T, Koretke K et al. (2005) AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. *Structure*, **13**:919-928
149. Sullivan DM, Bobay BG, Kojetin DJ, Thompson RJ, Rance M et al. (2008) Insights into the nature of DNA binding of AbrB-like transcription factors. *Structure*, **16**:1702-1713
150. Coles M, Hulko M, Djuranovic S, Truffault V, Koretke K et al. (2006) Common evolutionary origin of swapped-hairpin and double-psi beta barrels. *Structure*, **14**:1489-1498
151. Schwede T PM (2008).Computational structural biology: methods and applications, Volume , Edition (World Scientific).
152. Sambrook J, Fritsch E, Maniatis T (1989).Molecular cloning. a laboratory manual., Volume , Edition (Cold Spring Harbor Laboratory Press).
153. Reuter M, Kupper D, Meisel A, Schroeder C, Kruger DH (1998) Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites. *J Biol Chem*, **273**:8294-8300
154. Shlyakhtenko LS, Gilmore J, Portillo A, Tamulaitis G, Siksnys V et al. (2007) Direct visualization of the EcoRII-DNA triple synaptic complex by atomic force microscopy. *Biochemistry*, **46**:11128-11136
155. Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62**:48-57
156. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, **62**:72-82
157. French G WK (1978) On the treatment of negative intensity observations. *Acta Crystallogr*, **A34**:517-525
158. Navaza J (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr*, **57**:1367-1372
159. Vagin A TA (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, **56**:1622-1624
160. Read RJ (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr*, **A42**:140-149
161. Emsley P CK (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**:2126-2132
162. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P et al. (1998) Crystallography & NMR system: A new software suite for macromolecular

structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**:905-921
163. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol*, **374**:300-321
164. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**:213-221
165. Painter J ME (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr*, **62**:439-450
166. Zaremba M, Sasnauskas G, Urbanke C, Siksnys V (2006) Allosteric communication network in the tetrameric restriction endonuclease Bse634I. *J Mol Biol*, **363**:800-812
167. Tamulaitis G, Zaremba M, Szczepanowski RH, Bochtler M, Siksnys V (2008) How PspGI, catalytic domain of EcoRII and Ecl18kI acquire specificities for different DNA targets. *Nucleic Acids Res*, **36**:6101-6108
168. Williams T KC (2007). Gnuplot 4.2: an interactive plotting program. (, ed.), vol. , ed. pp. ,
169. Yoshioka K (2002) KyPlot—a user-oriented tool for statistical data analysis and visualization. *CompStat.*, **17**:425–437
170. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, Bairoch A (2005).The proteomics protocols handbook, Volume , Edition (Humana Press).
171. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**:3389-3402
172. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**:4673-4680
173. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*, **31**:3320-3323
174. Padilla JE YT (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr D Biol Crystallogr*, **59**:1124-1130
175. Chojnowski G BM (2010) DIBER: protein, DNA or both?. *Acta Crystallogr D Biol Crystallogr*, **66**:643-653
176. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol*, **231**:1049-1067
177. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, **8**:52-6, 29
178. Chen VB, Arendall WB3, Headd JJ, Keedy DA, Immormino RM et al.

(2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**:12-21
179. Lavery R SH (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn*, **6**:63-91
180. Luscombe NM, Laskowski RA, Thornton JM (1997) NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res*, **25**:4940-4945
181. Hubbard S TJ (1993).'naccess', computer program, department of biochemistry and molecular biology, university college london., Volume , Edition ().
182. Weichenberger CX SM (2007) NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. *Nucleic Acids Res*, **35**:W403-6
183. Wiederstein M SM (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, **35**:W407-10
184. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics.*, **24**:2780-2781
185. Krebs WG GM (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res*, **28**:1665-1675
186. van Dijk M BA (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res*, **37**:W235-9
187. Macke T, C.D. (1998).Modeling unusual nucleic acid structures. In: Molecular Modeling Of Nucleic Acids, Volume , Edition, ed. N.B. Leontes and J. SantaLucia J, e., ed. (Washington, DC: American Chemical Society), p. pp. 379-393.
188. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, **31**:3381-3385
189. Schrodinger L (2010) The PyMOL Molecular Graphics System, Version 1.0r2. , :
190. Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography*, **24**:946-950
191. Merritt EA MM (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **50**:869-873
192. Tamulaitis G, Zaremba M, Szczepanowski R, Bochtler M, Siksnys V (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res*, **35**:4792–4799
193. Pingoud V, Conzelmann C, Kinzebach S, Sudina A, Metelev V et al. (2003) PspGI, a type II restriction endonuclease from the extreme thermophile

Pyrococcus sp.: structural and functional studies to investigate an evolutionary relationship with several mesophilic restriction enzymes. *J Mol Biol*, **329**:913-929
194. Petrauskene OV, Gromova ES, Romanova EA, Volkov EM, Oretskaya TS et al. (1995) DNA duplexes containing methylated bases or non-nucleotide inserts in the recognition site are cleaved by restriction endonuclease R.EcoRII in presence of canonical substrate. *Gene*, **157**:173-176
195. Jen-Jacobson L, Engler LE, Lesser DR, Kurpiewski MR, Yee C et al. (1996) Structural adaptations in the interaction of EcoRI endonuclease with methylated GAATTC sites. *EMBO J*, **15**:2870-2882
196. Mucke M, Pingoud V, Grelle G, Kraft R, Kruger DH et al. (2002) Asymmetric photocross-linking pattern of restriction endonuclease EcoRII to the DNA recognition sequence. *J Biol Chem*, **277**:14288-14293
197. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**:2780-2781
198. Wang Y, Deng D, Zhang R, Wang S, Bian Y et al. (2012) Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol Biol Rep*, :
199. Bao Y, Higgins L, Zhang P, Chan S, Laget S et al. (2008) Expression and purification of BmrI restriction endonuclease and its N-terminal cleavage domain variants. *Protein Expr Purif*, **58**:42-52
200. Lu X OW (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc*, **3**:1213-1227
201. Zaremba M, Sasnauskas G, Urbanke C, Siksnys V (2005) Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: oligomeric structure-stability-function correlations. *J Mol Biol*, **348**:459-478
202. Ulyanov NB ZV (1984) Sequence-dependent anisotropic flexibility of B-DNA. A conformational study. *J Biomol Struct Dyn*, **2**:361-385
203. Suzuki M, Yagi N, Finch JT (1996) Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett*, **379**:148-152
204. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A*, **95**:11163-11168
205. Repin VE, Lebedev LR, Puchkova L, Serov GD, Tereschenko T et al. (1995) New restriction endonucleases from thermophilic soil bacteria. *Gene*, **157**:321-322