VILNIUS UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Rasa Sukackaitė

STRUCTURAL AND FUNCTIONAL STUDIES OF THE RESTRICTION ENDONUCLEASE BpuJI

Doctoral dissertation

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

Vilnius, 2009

The work presented in this doctoral thesis was carried out at the Institute of Biotechnology from 2004 to 2009.

Supervisor:

prof. dr. Virginijus Šikšnys (Institute of Biotechnology, physical sciences, biochemistry-04 P, proteins, enzymology-P310)

Scientific consultant:

dr. Saulius Gražulis (Institute of Biotechnology, physical sciences, biochemistry-04 P, proteins, enzymology-P310)

VILNIAUS UNIVERSITETAS

BIOTECHNOLOGIJOS INSTITUTAS

Rasa Sukackaitė

RESTRIKCIJOS ENDONUKLEAZĖS BpuJI STRUKTŪRINIAI IR FUNKCINIAI TYRIMAI

Daktaro disertacija

Fiziniai mokslai, biochemija (04 P), baltymai, enzimologija (P310)

Vilnius, 2009

Disertacija parengta Biotechnologijos institute 2004-2009 metais.

Mokslinis vadovas:

prof. dr. Virginijus Šikšnys (Biotechnologijos institutas, fiziniai mokslai, biochemija-04 P, baltymai, enzimologija-P310)

Mokslinis konsultantas:

dr. Saulius Gražulis (Biotechnologijos institutas, fiziniai mokslai, biochemija-04 P, baltymai, enzimologija-P310)

Table of Contents

List of abbreviations

AdoMet	S-adenosyl-L-methionine
AFM	atomic force microscopy
AHJR	archaeal Holliday junction resolvase
BSA	bovine serum albumin
COG	cluster of orthologous groups of proteins
CTD	C-terminal domain
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetraacetic acid
HhH	helix-hairpin-helix
HTH	helix-turn-helix
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth
MES	2-(N-morpholino)-ethanesulfonic acid
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethyleneglycol
REase	restriction endonuclease
R-M	restriction-modification
SDS	sodium dodecyl sulfate
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
wHTH	winged-helix
wt	wild type

Introduction

Type II restriction endonucleases (REases) recognize short DNA sequences (4-8 bp) and cleave the target within or close to the recognition sequence to generate 5'- or 3'-overhangs or blunt ends[1]. Due to a very high specificity and simple substrate requirements, Type II REases are common tools in biotechnology. Their practical value stimulated extensive searches for new enzymes. The REBASE database[2] currently lists more than 3600 biochemically or genetically characterized Type II REases.

Type II REases form a structurally diverse family. Structurally characterized Type II restriction enzymes present five unrelated three-dimensional folds, several different variants of active sites and catalytic mechanisms, and a plethora of modes for protein–protein and protein–DNA interactions[3].

The majority of restriction enzymes belong to the PD-(D/E)XK fold, which also includes other nucleases involved in DNA repair and recombination[3]. REases belonging to the PD-(D/E)XK fold show remarkable structural variety[4, 5]. Many orthodox Type II enzymes like EcoRI or BamHI are arranged as homodimers comprised of two subunits. Each subunit contains a single active site and acts independently catalyzing cleavage of phosphodiester bonds on 3′→5′ and 5′→3′ DNA strands, respectively[1, 5]. Individual subunit is often folded as a single-domain protein and interacts with one half of the palindromic recognition site. The dimerization mode determines the relative position of the active sites in respect to the scissile phopsphates and therefore governs the cleavage pattern, e.g. 5'-overhangs or blunt ends[6].

The Type IIS restriction endonuclease FokI contains a separate domain for specific DNA binding, tethered to a nonspecific nuclease of the PD-(D/E)XK fold[7, 8]. Such structural arrangement allows Type IIS restriction enzymes to cleave phopshodiester bonds at distant sites from the recognition sequence.

FokI recognizes the asymmetric sequence 5'-GGATG and cuts the top and bottom DNA strands 9 and 13 nt downstream of this site. The modular architecture was also shown for Type IIS restriction enzyme BfiI, which has a barrel-like DNA recognition domain fused to a catalytic module of the phospholipase D rather than the PD-(D/E)XK family[9, 10].

The vast majority of the Type II enzymes cut DNA at strictly fixed positions in respect to their target sites; however, some of the Type IIS enzymes were reported to cleave DNA with limited variability[2]. For example, BfiI usually cuts the top and bottom DNA strands 4 and 5 nt downstream of the 5'-ACTGGG target site, but additional cleavages of the top strand 6 and 7 nt away from the recognition site are also observed[11]. Limited variability of the cleavage site was also reported for several Type IIB enzymes, e.g. AloI (5'-(7/12-13) GAACN$_6$TCC (12-13/7))[12]. The molecular mechanisms for the cleavage site variability are unknown.

Restriction endonuclease BpuJI from the *Bacillus pumilus* RFL1458 strain is specific for the asymmetric sequence 5'-CCCGT (K. Stankevičius, personal communication). Preliminary BpuJI characterization revealed that, in contrast to other characterized Type IIS enzymes, BpuJI could have a variable cleavage pattern (Z. Maneliene, personal communication).

The overall goal of the PhD project was to elucidate the molecular and structural mechanisms that lead to the unusual cleavage pattern of BpuJI restriction enzyme.

The specific aims of this study were:

1. Perform detailed analysis of the BpuJI cleavage pattern and cleavage rates using different substrates.

2. Determine the oligomerization state and putative domain structure of BpuJI.

3. Perform crystallographic studies of BpuJI or its domains.

Scientific novelty. BpuJI is the first functionally and structurally characterized Type II restriction endonuclease that has an extremely variable DNA cleavage pattern. We have shown by limited proteolysis and subsequent biochemical experiments that BpuJI is arranged of separate DNA-binding and catalytic domains. We provided here the first experimental evidence that the unusual cleavage pattern of BpuJI results from the site-specific nuclease activity and the end-directed nuclease activity of the BpuJI catalytic domain, manifested upon BpuJI binding to the target sequence. We have demonstrated by bioinformatics analysis that the BpuJI catalytic domain is similar to the archaeal Holliday junction resolvases and is present in different uncharacterized proteins including putative nucleases that cut methylated DNA. We have solved the crystal structure of the BpuJI DNA binding domain in the DNA-bound form which shows that the target sequence recognition is achieved by amino acid residues located on two helix-turn-helix motifs.

Practical value. The variability of the cleavage site can limit the application of certain restriction endonucleases for cloning and other experiments. Elucidation of the structure-function relationships in BpuJI catalysis contributes to our understanding of the mechanisms leading to variable cleavage by Type IIS enzymes. Bioinformatics analysis revealed that catalytic domain of BpuJI is shared by a group of putative nucleases which may recognize methylated DNA sequences. Such nucleases may have practical applications for the analysis of genome methylation status. The crystal structure of the BpuJI recognition domain-DNA complex provides a mechanism for asymmetric target recognition which could also be applicable to other restriction endonucleases.

The major findings presented for defence in this thesis:

1. BpuJI is dimer in solution and has a modular architecture with separate domains for DNA recognition and catalysis.
2. The BpuJI catalytic domain shows an end-directed nuclease activity,

which leads to the complicated cleavage pattern near the target site.

3. The BpuJI nuclease domain has a PD-(D/E)XK fold, while the recognition domain folds into two winged-helix subdomains.
4. The specificity-determinants of the BpuJI recognition domain are located in the two recognition helices of the HTH motifs and an N-terminal arm.
5. The BpuJI recognition domain shows structural similarity to the nicking endonuclease Nt.BspD6I.

1. Literature overview

PD-(D/E)XK nucleases include a large number of enzymes involved in DNA repair, recombination and protection against foreign DNA invasion[13]. These nucleases possess a common structural core which consists of a mixed β-sheet flanked by α-helices[14]. The structural core brings into spatial proximity two carboxylates and one lysine residue that form the catalytic centre[1]. These residues make a bipartite catalytic motif PD-(D/E)XK[1] which give the name for the group. PD-(D/E)XK nucleases display various elaborations of the common fold in the form of large insertions and terminal extensions, which are often involved in DNA recognition[3]. Even entire DNA binding domains, such as helix-turn-helix (HTH), can be fused to the PD-(D/E)XK core structure[8, 15]. Various structural elaborations and oligomerization modes generate PD-(D/E)XK enzymes that are specific for DNA sequence (REases[1], MutH[16]), structure (Holliday junction resolvases, Mus81 family enzymes[17]) or particular DNA strand (λ exonuclease[18], RecBCD[19]).

This literature review starts with an overview of the PD-(D/E)XK and HTH domain architecture, followed by structural analysis of individual PD-(D/E)XK nucleases: λ exonuclease, REases, mismatch repair nucleases, structure-specific nucleases, TnsA, I-Ssp6803I and RecBCD. The major goal of this review is to demonstrate that incorporation of the conserved PD-(D/E)XK domain into various structural contexts results in different biological functions.

1.1 The PD-(D/E)XK domain

1.1.1 The common structural core

The common structural core of PD-(D/E)XK nucleases was first identified by comparison of the crystal structures of EcoRI and EcoRV complexes with cognate DNA[22]. This structural core consists of a mixed five-stranded β-sheet

and two α-helices packed against it (Figure 1A), and can be described as a sandwich in which α-helices and β-strands lie in two stacked layers. Subsequent crystallographic studies revealed the similar core structure in many REases and various other nucleases involved in DNA repair and recombination[1, 13]. The SCOP database[14] currently assigns to the PD-(D/E)XK fold 33 families: 17 families of REases, 10 families of other nucleases and 6 families of hypothetical proteins. tRNA-intron endonuclease catalytic domain, eukaryotic RPB5 N-terminal domain and TBP-interacting protein also are assigned to the PD-(D/E)XK fold, although they do not have a nuclease function.

The structural core of the PD-(D/E)XK fold in the SCOP database is defined as a three-layer α/β/α sandwich that contains a mixed beta-sheet of five strands, following order 12345 (strands 2 and, in some families, 5 are antiparallel to the rest)[14]. Recent analysis of the structures of 22 different REases shows, however, that the two helices that occur most frequently in the existing structures do not flank the β-sheet, but are positioned on the same face of the sheet. Thus, the αβα-core consists of two helices and five β-strands that do not constitute an αβα structural sandwich[24].

1.1.2 The catalytic motif

The conserved structural core harbours the catalytic site that includes two carboxylates, one Asp and the second Asp or Glu, and one Lys[1, 25]. These residues form the hallmark catalytic motif $PDY_{10\text{-}20}$(D/E)XK (where Y is any amino acid, and X is a hydrophobic residue), e.g. $PD^{116}...D^{142}IK$ in BglI[26], $PD^{119}...E^{129}LK$ in λ-exonuclease[18]. Despite the variable distance between the two acidic residues in the primary sequence, the first Asp appears at the N-terminus on one β-strand, adjacent to the Asp or Glu on the next strand (Figure 1A)[25].

The catalytic motif in some nucleases exhibit deviations from the

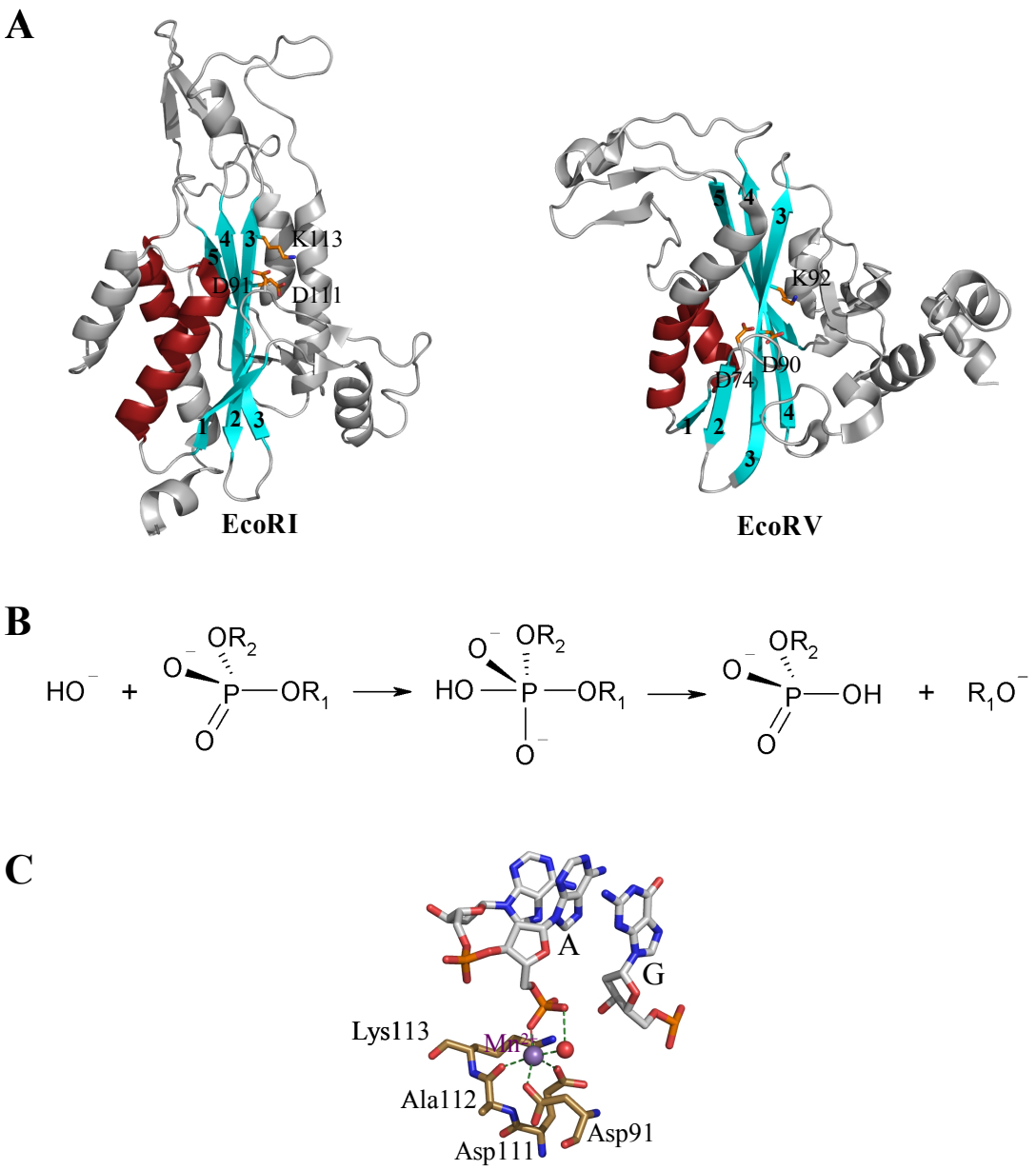

Figure 1. *(A) The common structural core[22] in the EcoRI (PDB:1eri)[20] and EcoRV (PDB:4rve)[21] REases. The structural core, which consists of a mixed β-sheet (cyan) and two α-helices (red), harbours the catalytic residues: two carboxylates and one lysine (shown in stick representation and labelled). The β-strands in the conserved sheet are numbered. The structural elements that do not belong to the core structure are in grey. These and other structure representations in the Literature overview were made with PyMol[23]. (B) A scheme for phosphodiester bond hydrolysis by S_N2-type mechanism[1, 25]. An activated nucleophile attacks the phosphorous atom to form a pentavalent transition state. (C) A metal binding site in the EcoRI-DNA complex (PDB:1qps). A manganese ion (purple sphere) is coordinated by the scissile phosphate, two carboxylates, a mainchain carbonyl and a water molecule (red sphere) which is also associated with the scissile phosphate. The substrate is cleaved in the crystal structure.*

consensus: the Pro is often missing (e.g. VD^{212}...E^{240}LK in BsoBI[27], QD^{70}...E^{77}LK in MutH[16]); the Lys is replaced by Glu in BamHI[28] (ID^{94}...E^{111}FE), Gln in BglII (ID^{84}...E^{93}VQ) and Val in SdaI (PD^{233}...E^{248}AV)[15]; the Mus81 family enzymes contain Arg instead of a hydrophobic residue (e.g. GD^{583}...E^{593}RK in Hef nuclease[29]); the second acidic residue is replaced by Asn in MspI[30] (TD^{99}...N^{117}IK) and Gln in similar enzyme HinP1I[31] (SD^{62}...Q^{81}VK^{83}). The second acidic residue is also missing in Cfr10I and related REases[32-36]. It was shown, however, that the counterpart of the second acidic residue comes from a different part of the Cfr10I sequence (PD^{134}...S^{188}VK...E^{204})[32, 37]. Thus, the only absolutely conserved residue of the catalytic motif is the first Asp. This variability makes it difficult to identify an active site in the PD-(D/E)XK nucleases solely based on sequence comparisons and usually requires analysis of the three-dimensional structures[38].

1.1.3 The mechanism of DNA cleavage

Phosphodiester bond hydrolysis by Type II restriction endonucleases follows an S_N2 type mechanism (Figure 1B)[1, 25]. To catalyse this reaction, three assisting groups are required: (i) a base to deprotonate the water molecule; (ii) a Lewis acid to stabilize the pentavalent transition state with two negative charges; and (iii) an acid to protonate the 3'-hydroxyl leaving group[1, 25].

PD-(D/E)XK nucleases need magnesium as a cofactor for DNA cleavage[25]. *In vitro* Mg^{2+} ions can be replaced by Mn^{2+} and some other dvivalent metal ions, but not by Ca^{2+}. One or two metal ions were observed at different sites in the crystal structures of the tertiary complexes of the enzyme, cognate DNA and metal ion(s). The comparison of these structures revealed a common binding site that is formed by the first and second acidic residues of the catalytic motif (P<u>DY</u>$_{10\text{-}20}$<u>(D/E)X</u>K) and a main chain carbonyl of the hydrophobic residue (PDY$_{10\text{-}20}$(D/E)<u>X</u>K)[1, 25]. The metal ion that occupies this binding site is also coordinated to a water molecule, which is associated with

the scissile phosphate (Figure 1C). The first and in some cases the second carboxylate of the catalytic motif can also form a second metal binding site[1, 25] .

Based on different structures of the tertiary complexes, one-, two- and even three- metal mechanisms were proposed[5, 25]. These mechanisms also differ in which element of the active site acts as a general base that activates the water molecule. The general base could be the metal ion, the Lys of the catalytic motif or the scissile phosphate itself (substrate assisted catalysis). For all these groups a considerable shift in the pK_a value would be required. The exact number of metal ions required for catalysis and the detailed mechanism still is a matter of debate even for best-characterized REases. Also, it is not clear if all PD-(D/E)XK nucleases follow the same mechanism.

1.2 The helix-turn-helix domain

PD-(D/E)XK nucleases employ various structural elaborations of the catalytic core, which can comprise even entire domains, to recognize DNA substrates[3]. One of DNA binding domains that are often found fused to the PD-(D/E)XK nuclease domain is the helix-turn-helix (HTH) domain[3, 8, 15]. HTH domains are critical determinants for DNA binding in the most prevalent prokariotic and some eukariotic transcription factors. These domains have been also recruited for various other functions such as DNA repair, replication, RNA metabolism and protein-protein interactions[39].

1.2.1 The tri-helical core

The structural core of the HTH domain is comprised of three α-helices[39]. The tri-helical core has an approximately triangular outline when it is displayed by placing the third helix in the front and in the horizontal orientation. The 2nd and the 3rd helices of the tri-helical bundle constitute a DNA binding HTH motif (Figure 2A). The sharp turn between these helices is a defining feature of the domain and usually do not tolerate insertions or

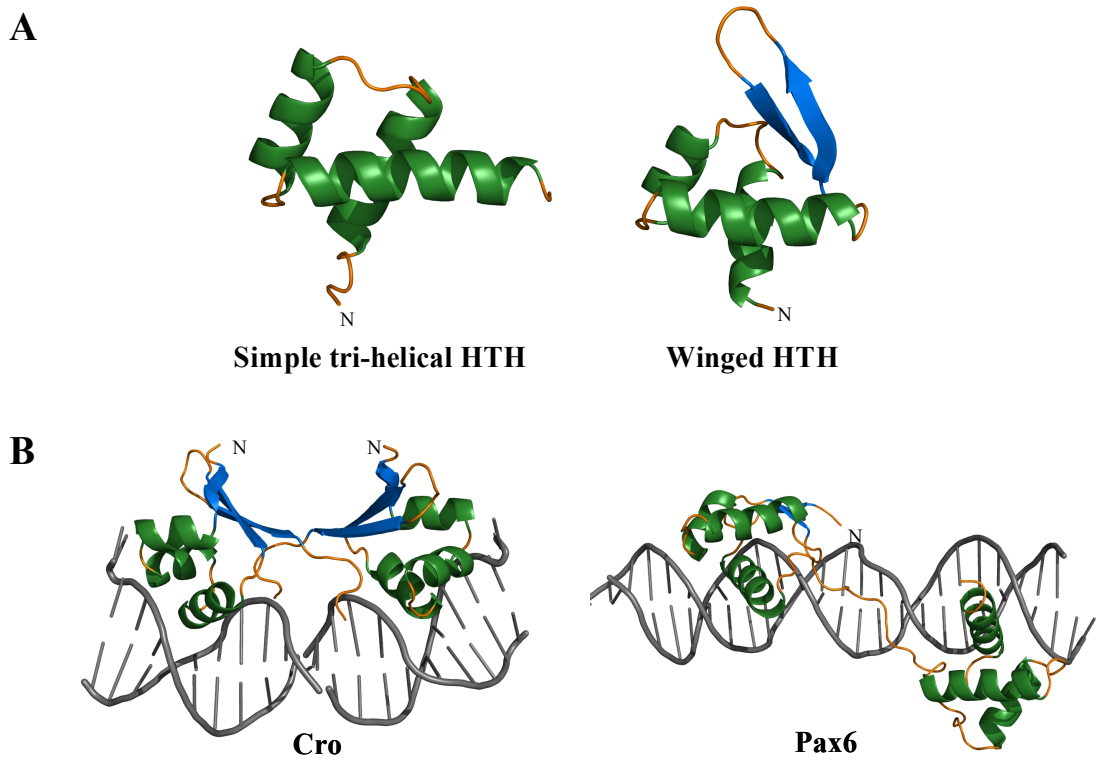

***Figure 2**. The HTH domain. (A) Representative structures of the simple tri-helical HTH (PDB:2hdd)[40] and the winged HTH (PDB:1smt)[41] domains. Helices and strands are colored in green and blue, respectively. (B) The phage lambda Cro repressor (PDB:6cro)[42] and the human Pax6 paired domain (PDB:6pax)[43] bound to DNA (grey). The recognition helices are inserted into the major groove.*

distortions, while the loop between the 1st and the 2nd helices shows more variability[39].

The basic tri-helical version of the HTH domain that is comprised entirely of the three core helices is widely seen across the three super-kingdoms of life[39]. Various N-terminal and C-terminal extensions to the core HTH fold also are common. A large group of such domains are the winged HTH (or winged helix) domains that harbour a C-terminal β-strand hairpin unit, a wing (Figure 2A)[39, 44].

1.2.2 DNA binding by HTH domains

The second helix of the HTH motif (or the 3rd helix of the tri-helical core) is known as a recognition helix[39]. The recognition helix is inserted into the

major DNA groove (Figure 2B) allowing protein side chains to make extensive base-specific contacts. Additional protein-DNA contacts can be mediated by the elaborations of the structural core such as an N-terminal arm or loops of the winged-helix domain[44, 45].

Due to DNA curvature, the recognition helix cannot access more than five bases in the major groove. Usually it binds to four base pairs[46]. Therefore, several HTH domains are combined to recognize longer DNA sequences (Figure 2B) either by dimerization (e.g. phage lambda Cro repressor[42]) or repeating them in tandem (e.g. Pax6 paired domain[43]).

1.3 Structure-function relationships in PD-(D/E)XK nucleases

1.3.1 λ exonuclease

Phage λ exonuclease is involved in homologous recombination[47]. It binds to a free end of double-stranded DNA and processively degrades one strand in the 5'→3' direction releasing mononucleotides. The 3'-overhangs generated by λ exonuclease are further processed by β protein which promotes annealing of complementary strands[47].

The crystal structure revealed that λ exonuclease consists of three subunits that share the structural core with restriction enzymes[18, 48]. The three protomers form a toroid (Figure 3A) with a funnel shaped central channel, tapering from an inner diameter of about 30 Å at the wider end to 15 Å at the narrow end. Although the crystal structure determined does not include DNA, it suggests that λ exonuclease encircles the substrate during hydrolysis (Figure 3B): the central channel is large enough to accommodate double-stranded DNA at the wide end but can accommodate only single-stranded DNA at the other end; and the active site is inside the channel[48, 49].

The toroidal structure explains the resistance of closed circular or nicked

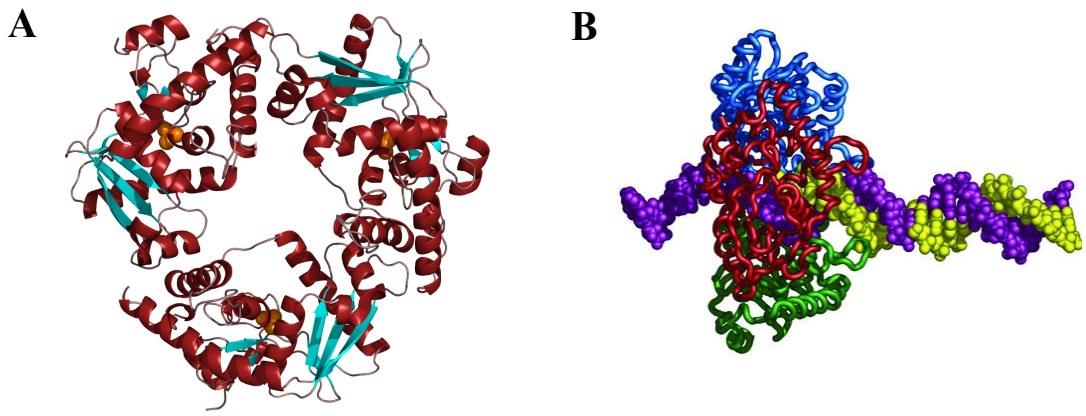

Figure 3. *λ exonuclease. (A) The toroidal structure of the λ exonuclease trimer (PDB:1avq)[48]. Helices and strands are colored in red and cyan, respectively. Three phosphate ions (orange spheres) are bound near the catalytic residues of each monomer. (B) The proposed model of the λ exonuclease-DNA complex. The three monomers are shown in red, blue, and green. The partially digested 5'-strand is shown in yellow, while the 3'-strand that threads through the enzyme is shown in purple. This image is from Subramanian et al[49].*

DNA substrates to hydrolysis by λ exonuclease, as the trimer can bind a new DNA strand only at a double-stranded break[48, 49]. Also, once bound, the toroid remains on the nondigested strand of DNA until the end of the DNA is reached or until the protein dissociates into monomers. Indeed, λ exonuclease is able to hydrolyse thousands of nucleotides per binding event. Microscopic observation of digestion of a single DNA molecule by λ exonuclease revealed that the processivity of the enzyme was >3000 bp and digestion rate was ~1000 bases/s[50]. Subsequent single-molecule approaches also revealed high processivity of λ exonuclease (18,000±8000 base pairs), but the digestion rate was ~32 nucleotides/s[51]. It was suggested that the melting of a base from the DNA is the rate-limiting step in the catalytic cycle of λ exonuclease[51].

Three residual phosphate ions (one per monomer) were found in the crystal structure of λ exonuclease (Figure 3A) coordinated by residues highly conserved in the λ exonuclease family of recombination nucleases[48]. The phosphate ion is in close proximity to the catalytic site and could mimic the phosphate group at the 5' end of double-stranded DNA. This hypothesis was confirmed by mutational analysis[49]. It was shown that binding of the

5'-phosphorylated end is important for the processivity of λ exonuclease[49].

1.3.2 Type II restriction endonucleases

REases are parts of restriction-modification (R-M) systems, which are common in bacteria. REases cleave foreign DNA in response to specific recognition sequences, while host DNA is protected by methylation at these sequences[52]. Their classical function is a defense against bacteriophage infection. In addition, R-M systems could modulate the frequency of recombination[53] and create a barrier against an unregulated uptake of foreign DNA, maintaining the species identity[54]. Also, it was proposed that R-M systems can act as selfish, mobile genetic elements[55].

Type II R-M systems have the simplest organization, with endonuclease and methyltransferase as separate enzymes[4]. Other R-M systems (Type I, Type III and Type IV) are complex, multisubunit molecular machines which will be discussed later. Type II REases recognize short DNA sequences (4-8 bp) and cleave DNA either within or close to the recognition sequence at a fixed site or with known and limited variability[4]. The DNA fragments produced have 3'-hydroxyl and 5'-phosphate ends that are either 'blunt' or 'sticky' with 3'- or 5'-overhangs. Because of the high specificity of DNA recognition and cleavage, Type II REases are common tools for recombinant DNA work[1].

The REBASE database[2] lists more than 3600 biochemically or genetically characterized Type II restriction enzymes, recognizing more than 260 distinct specificities. The crystal structures were determined for 31 Type II REases. Among these structures, all but two possess catalytic domains of the common PD-(D/E)XK nuclease fold. Two structures are unrelated to the others: BfiI possesses the phospholipase D fold nuclease domain[10], while PabI has a new fold termed "half-pipe"[56]. Bioinformatics analysis indicates that 69% of experimentally characterized Type II REases contain a PD-(D/E)XK domain[3].

The second most common nuclease domain is the HNH domain (8% of members). The catalytic domains in Type II REases can be fused with different domains (e.g., HTH, methyltransferase) to generate a great variety of domain architectures[3].

Type II REases are a diverse group in structural organization as well as in mode of DNA cleavage[5]. The current nomenclature[4] classifies Type II REases into 11 subtypes according to recognition sequence, subunit composition, cafactor requirement and need of two targets for DNA cleavage (Table 1). As some of the criteria are based on the sequence cleaved and others on the structure of the enzymes themselves, the same REase can fall into more than one subtype, e.g. BcgI is both Type IIB and Type IIH[4].

Table 1. Subtypes of Type II restriction enzymes[4]

	Defining feature	Examples
A	Asymmetric recognition sequence	FokI, GGATG(9/13)
B	Cleavage at both sides of target on both strands	BcgI, (10/12)CGANNNNNNTGC(12/10)
C	R and M functions in one polypeptide	GsuI, CTGGAG(16/14)
E	Two targets; one cleaved, one an effector	EcoRII, ↓CCWGG
F	Two targets, both cleaved coordinately	SfiI, GGCCNNNN↓NGGCC
G	Affected by AdoMet	Eco57I, CTGAAG(16/14)
H	Similar to Type I gene structure	BcgI, (10/12)CGANNNNNNTGC(12/10)
M	Methylated target	DpnI, G^{m6}A↓TC
P	Symmetric target and cleavage sites	EcoRI, G↓AATTC
S	Asymmetric target and cleavage sites	FokI, GGATG(9/13)
T	Heterodimeric organization	BslI, CCNNNNN↓NNGG

1.3.2.1 Single-domain REases

Dimers. Most of Type IIP enzymes are composed of two identical subunits which each contains a single PD-(D/E)XK domain. Crystal structures were determined for 11 of such REases (Table 2). All these enzymes share the common core which contains the catalytic centre. The "outside'' of the core is decorated by additional subdomains, which are more diverse in structure than the catalytic core and often are involved in DNA binding and dimerization[1].

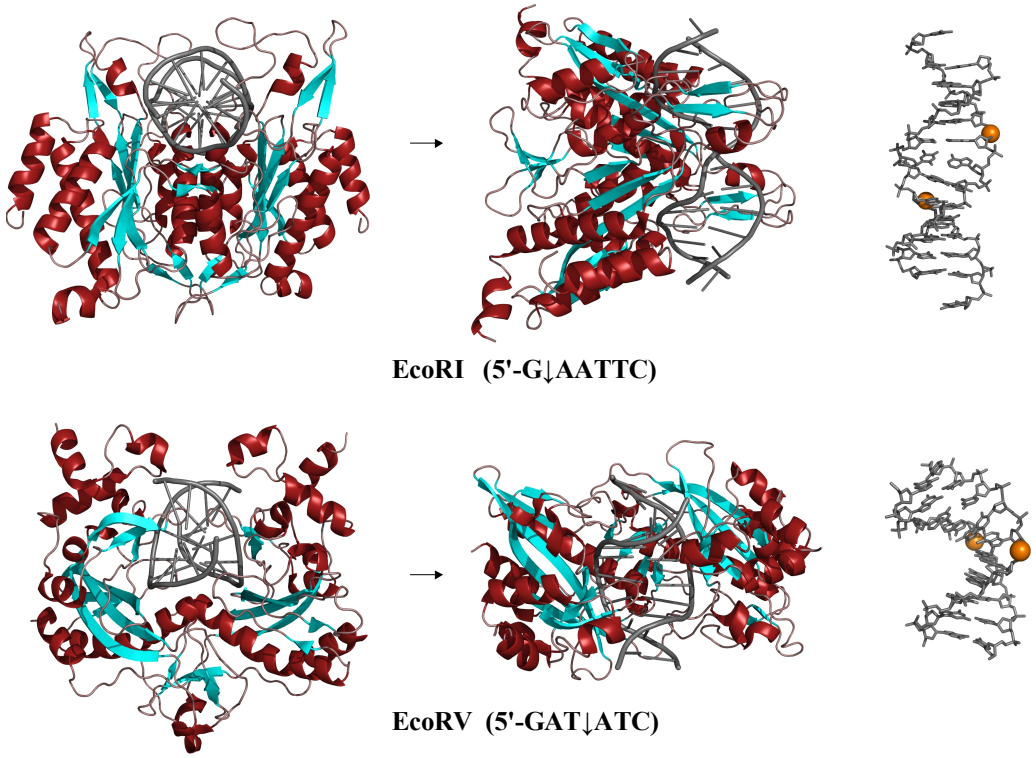

Figure 4. *Dimeric REases. The structures of EcoRI (PDB:1eri)[20] and EcoRV (PBD:4rve)[21] bound to cognate DNA (grey) are shown in two different orientations. Helices and strands are colored in red and cyan, respectively. The left panels show the DNA in the EcoRI and EcoRV complexes with the scissile phosphates highlighted as orange spheres. The different distances between the scissile phosphates affect the dimerization mode of the enzymes.*

Homodimeric REases interact symmetrically with palindromic recognition sequences, one protomer making base-specific contacts with a half-site[1, 5].

Because of the different distances between the scissile phosphates, enzymes that generate 5'-overhangs (e.g. EcoRI, 5'-G↓AATTC) bind the DNA in the major groove, while the enzymes that generate blunt ends (e.g. EcoRV, 5'-GAT↓ATC) or 3'-overhangs (e.g. BglI, 5'-GCCNNNN↓NGGC) approach the DNA from the minor groove side[1, 5]. This has consequences on the positioning of the two active sites and, therefore, for the dimerization mode: restriction enzymes that cleave DNA giving 5'-overhangs dimerize via four helices, while the dimerization contacts of enzymes producing blunt ends are made by other structural elements protruding away from the common structural core (Figure 4).

Table 2. Dimeric single-domain REases with determined crystal structures

Enzyme	Recognition sequence	Reference
BamHI	5'-↓GATCC	[28]
BglI	5'-GCCNNNN↓NGGC	[57]
BglII	5'-A↓GATCT	[27]
BsoBI	5'-C↓YCGYG	[58]
BstYI	5'-R↓GATCY	[59]
EcoO109I	5'-RG↓GNCCY	[20]
EcoRI	5'-G↓AATTC	[21]
EcoRV	5'-GAT↓ATC	[60]
MunI	5'-C↓AATTG	[60]
PvuII	5'-CAG↓CTG	[61]
PspGI	5'-↓CCWGG	[62]

REases producing 5'-overhangs and blut ends (or 3'-overhangs) also show different topology of the conserved structural core (Figure 1A), in particular the orientation of the 5-th β-strand (parallel in EcoRI, antiparallel in EcoRV), and are referred as EcoRI-like and EcoRV-like enzymes[1, 5].

Tetramers. Some of Type II REases function as tetramers. The crystal structures of Bse634I (5'-R↓CCGGY)[34], Cfr10I (5'-R↓CCGGY)[32], NgoMIV (5'-G↓CCGGC)[33] and SfiI (5'-GGCCNNNN↓NGGCC)[63] revealed that they are dimers of primary dimers with two DNA binding clefts on the opposite sides of the tetramer (Figure 5). In contrast to dimeric enzymes, these REases require two recognition sites for optimal activity[33, 34, 64, 65]. The substrates with two sites are converted directly to products cleaved at both sites, bypassing intermediates cut at one site. The current nomenclature classifies the enzymes that interact with, and cleave coordinately, two copies of their recognition sequence, as Type IIF[4]. Single mutation at the Bse634I tetramerization interface converts the tetrameric restriction enzyme into a dimer[66]. The dimeric mutant cleaves plasmid DNA containing one or two sites with the same efficiency as the tetramer cleaves the two-site plasmid, but the disruption of the tetramer compromises the stability of the enzyme[66].

SgrAI (5'-CR↓CCGGYG) and Ecl18kI (5'-↓CCNGG) are dimers in solution[67, 68], but could associate to tetramers in presence of specific DNA.

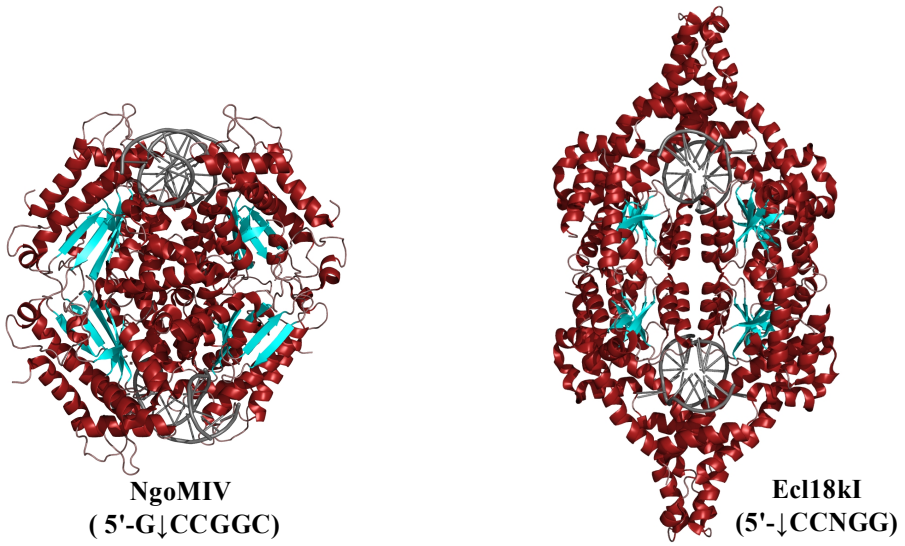

Figure 5. *Tetrameric REases. The structures of NgoMIV (PDB:1fiu)[33] and Ecl18kI (PBD:2gb7)[35] bound to cognate DNA (grey) are shown with helices and strands colored in red and cyan, respectively.*

SgrAI aggregates to high molecular mass species when bound to its specific DNA sequence[67]. As tetrameric enzymes, it cleaves DNA with two sites more rapidly than DNA with one site[69]. Surprisingly, although the structure of SgrAI dimer is similar to those of Cfr10I, Bse634I and NgoMIV, SgrAI is a dimer in the crystal structure[36]. Ecl18kI forms a tetramer in the crystal structure of the protein-DNA complex (Figure 5)[35]. It remains to be determined if the Ecl18kI tetramer is functionally relevant.

Monomers. In contrast to classical Type IIP enzymes that form symmetric enzyme-DNA complexes, MspI (5'-C↓CGG)[30], HinP1I (5'-G↓CGC)[70], MvaI

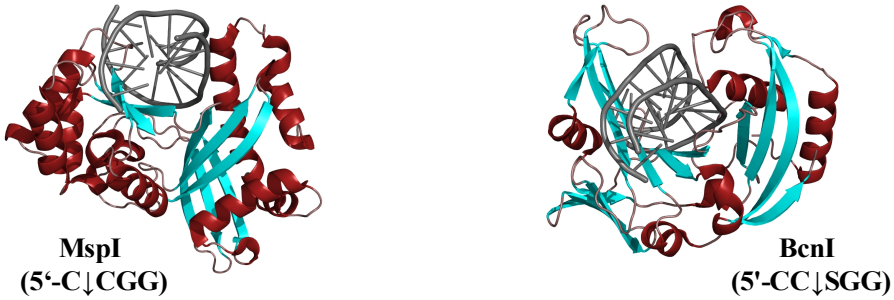

Figure 6. *Monomeric REases. The structures of MspI (PDB:1sa3)[30] and BcnI (PBD:2odi)[72] bound to cognate DNA (grey) are shown with helices and strands colored in red and cyan, respectively.*

(5'-CC↓WGG)[71] and BcnI (5'-CC↓SGG)[72] bind to palindromic DNA sequences as monomers, making specific contacts with all the bases of the target sequence and not just with the half-site (Figure 6). Since the monomer harbours only one active site, the question arises how these enzymes manage to cut both DNA strands. Several explanations are possible: (a) there is a monomer-dimer equilibrium in the presence of DNA, and the initial monomer–DNA complex might recruit another monomer, either from solution or bound to another site, for the cleavage reaction; (b) two strands are cut sequentially by a flip of the enzyme around the DNA after the first cleavage reaction. Sedimentation velocity experiments suggest for MspI monomer-dimer equilibrium[30]. In contrast, preliminary data indicate that BcnI and MvaI enzymes introduce double-strand breaks by sequentially nicking individual DNA strands[71, 72].

Modes of DNA recognition by single-domain REases. Specific DNA recognition by restriction enzymes is often accompanied by more or less pronounced distortions of the DNA and conformational changes in the protein that often involve repositioning of the subunits and subdomains[1, 5]. A characteristic feature of the specific protein-DNA complex is a highly co-operative hydrogen bond network comprising contacts to the target bases as well as to the sugar-phosphate backbone[1, 5]. In the single-domain restriction enzymes, DNA recognition is strictly coupled to catalysis. The change of a contacting residue usually results in a large drop in the catalytic activity but not in a change of the specificity, thus efforts to change the specificity by mutagenesis have been unsuccessful[5, 73].

EcoRI-like enzymes that produce 5'-overhangs recognize DNA via an α-helix and a loop, while EcoRV-like enzymes (producing blunt ends or sticky ends with 3'-overhangs) mainly use a ß-strand or a β-like turn for DNA recognition[5]. REases that recognize similar targets can share a similar mechanism for DNA recognition, even though low similarity in protein

sequences. For example, EcoRI (5'-G↓AATTC) and MunI (5'-C↓AATTG) contact the common core sequence AATT by homologous structural elements[60]. BglI (5'-GCCNNNN↓NGGC) is a dimer, while SfiI (5'-GGCCNNNN↓NGGCC) is a tetramer, but their modes of DNA recognition are unusually similar[63]. However, BamHI (5'-G↓GATCC) and BglII (5'-A↓GATCT) display considerably different protein–DNA contacts even at the common target site base pairs. Also, disparate DNA conformations are observed in these two complexes[74]. The contacts to the central CCGG bases are nearly identical in the NgoMIV (5'-G↓CCGGC)[33] and SgrAI (5'-CR↓CCGGYG)[36] protein/DNA complexes, mediated by a common segment RSDR in both the proteins. The motif RSDR has structural equivalents in the REases Bse634I (5'-R↓CCGGY)[34] and Cfr10I (5'-R↓CCGGY)[32] solved in the DNA-free form, suggesting that all the four enzymes recognize the common CCGG tetranucleotide by the same mechanism. This similarity in DNA recognition is also shared by the Ecl18kI (5'-↓CCNGG)[68] and PspGI (5'-↓CCWGG)[75] REases recognizing interrupted sequences.

The crystal structure revealed that Ecl18kI flips both the central nucleotides within the target sequence and buries the extruded bases in pockets within the protein[35]. The flip of the central bases reduces the length of a 5 bp stretch to the length of a 4 bp stretch, making the distances between scissile phosphates in the Ecl18kI and NgoMIV recognition sequences nearly identical (Figure 7)[35]. Similar DNA conformation also is observed in the PspGI complex with cognate DNA[62]. Thus, Ecl18kI and PspGI flip nucleotides to achieve the specificity for the target site.

1.3.2.2 REases with an extra structural domain

In some REases the conserved catalytic core is decorated by an extra domain that mainly serves a structural role. HincII (5'-GTPy↓PuAC) contains an inserted domain, topologically similar to the ‘‘saposin’’ domain of the

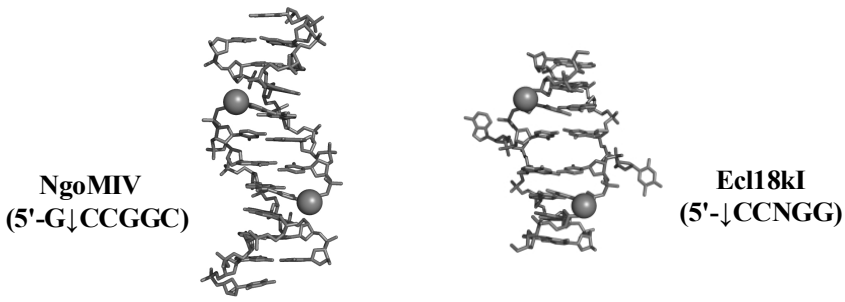

***Figure** 7. The DNA conformation in complex with NgoMIV (PDB:1fiu)[33] and Ecl18kI (PBD:2gb7)[35]. Ecl18kI flips the central bases to achieve the recognition of the interrupted sequence. Due to the base flipping the distance between the scissile phosphates (shown as spheres) in the Ecl18kI complex is nearly identical to that in the NgoMIV complex.*

bacterial protein NK-lysin[76]. This extra domain makes contacts to the backbone of the DNA flanking the recognition sequence. Moreover, in contrast to EcoRV and similar enzymes, the HincII dimerization interface is formed by a long coiled coil located on the major groove of the DNA (Figure 8A)[76].

The N-terminal part of NotI (5'-GC↓GGCCGC) forms a compact metal-binding domain that coordinates an iron ion by a Cys4 motif[77]. This domain is tightly associated with the nuclease domain by its "clamp" region, but does not appear to be coupled to the enzyme's active site (Figure 8A). The metal binding domain of NotI makes nonspecific DNA-backbone contacts and participates in the positioning of the nearby protein elements for DNA recognition. Sequence analysis suggests that a similar domain is present in a large number of other bacterial endonucleases[77].

1.3.2.3 REases with an effector domain

Several REases contain an additional domain recognizing the target sequence. Therefore, they interact with two DNA sites at the same time. In contrast to tetrameric enzymes, which coordinately cleave DNA at two sites, they cleave only one target at the time, the other being an allosteric effector[65]. The current nomenclature classifies such enzymes as Type IIE[4], with NaeI (5'-GCC↓GGC) and EcoRII (5'-↓CCWGG) as archetypal ones.

The crystal structure of NaeI shows a dimeric molecule. The NaeI monomer is composed of an N-terminal PD-(D/E)XK domain and a C-terminal HTH domain (Figure 8B)[78]. The NaeI dimer binds two molecules of DNA: one in the cleft between the catalytic domains, and another in the cleft between the effector HTH domains. The two NaeI domains recognize the same target sequence 5'-GCCGGC very differently[79].

The C-terminal catalytic domain of EcoRII has a PD-(D/E)XK fold and shares the conserved active site architecture and DNA binding elements with NgoMIV and Ecl18kI[80, 81]. The N-terminal effector-binding domain was identified as a novel DNA recognition fold[80], which was also found in REase BfiI and plant transcription factors of B3 superfamily[10]. The crystal structure of

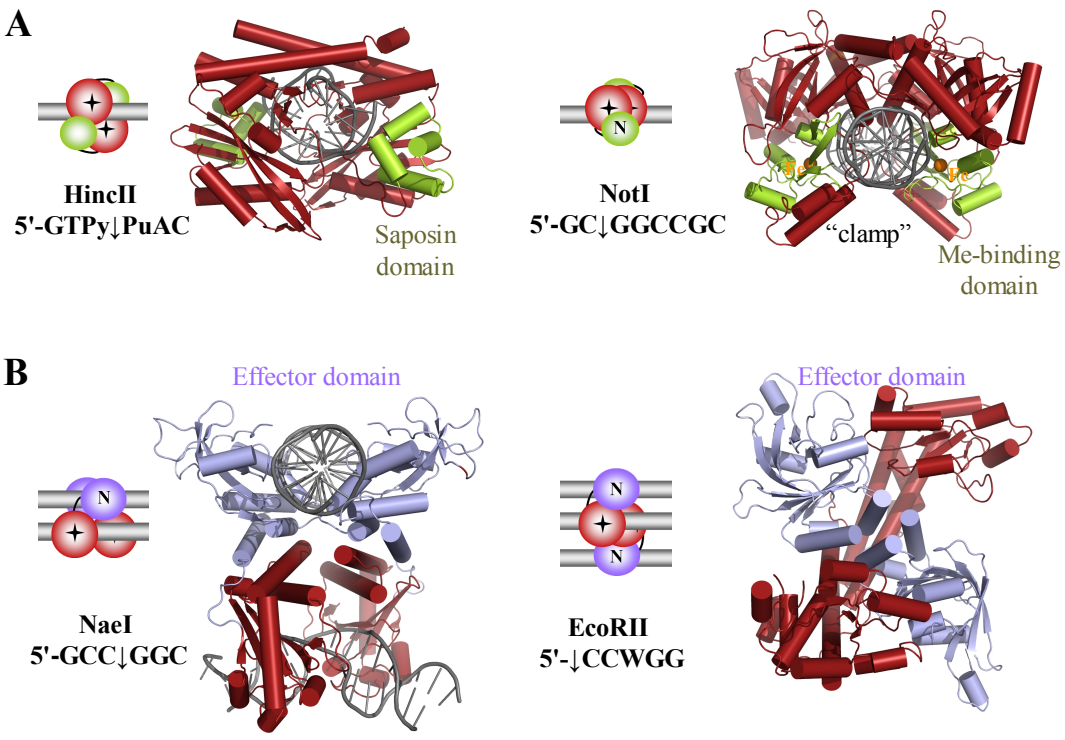

Figure 8. *(A) The structures of REases with extra structural domain, HincII (PDB:1kc6)[76] and NotI (PDB:3c25). Nuclease domains are shown in red, structural domains are green, DNA is shown in grey, and iron ions are represented by orange spheres. The both enzymes are dimeric. (B) The structures of NaeI (PDB:1iaw)[76] and EcoRII (PDB:1na6)[35] that contain and effector domain (shown in blue). The color coding in the cartoons representing protein-DNA complexes is the same as in the corresponding structures. The structure of EcoRII does not contain DNA, but it was shown that EcoRII can bind three copies of the target sequence[81, 82].*

EcoRII was solved without substrate[80]. Biochemical analysis indicates, however, that EcoRII contains three putative DNA binding interfaces: one at the C-terminal domain dimer and two at each of the N-terminal domains (Figure 8B)[81]. Indeed, EcoRII requires simultaneous binding of three rather than two recognition sites in cis to achieve concerted DNA cleavage at a single site[82]. Moreover, the synaptic EcoRII-DNA complexes involving three DNA binding sites were visualized by AFM[83].

Sau3AI (5-↓GATC) also binds two targets simultaneously and displays a strong preference for substrates with two copies of the recognition site over substrates with a single copy[84]. Sequence analysis indicates that Sau3AI is a pseudo-dimer composed of two similar PD-(D/E)XK fold domains in a single monomer[84]. This hypothesis was confirmed by the crystal structure of the Sau3IA C-terminal domain[85]. It was also shown that it binds to one copy of the recognition sequence as a monomer, but has no cleavage activity. Thus, the C-terminal domain is an allosteric effector domain, and Sau3AI belongs to Type IIE restriction enzymes[85].

1.3.2.4 REases with separate domains for recognition and catalysis

The FokI restriction endonuclease. REases that recognize asymmetric sequences and cleave DNA at a distant site from the recognition sequence (Type IIS)[4] are assumed to have separate domains for recognition and cleavage. The archetypal Type IIS enzyme is FokI (GGATG (9/13)). FokI is a monomer in solution[86] and in the crystal structure of the protein-DNA complex[8]. The monomer comprises two domains: an N-terminal DNA-binding domain, composed of three winged HTH subdomains (D1, D2 and D3), and a C-terminal catalytic domain, similar to other restriction enzymes. The subdomains D1 and D2 of the recognition domain make almost all of the base-specific contacts over the whole target sequence. The catalytic domain is packed alongside the recognition domain and positioned away from the site of

cleavage, therefore the crystal structure does not represent the catalytically active FokI complex (Figure 9A).

The FokI catalytic domain has only one active site. To cleave both strands at one site, the two monomers associate into a dimeric complex in presence of divalent metal ions and the cognate DNA[87, 88]. The crystal structure of the apo-enzyme revealed that the FokI dimerization interface is between the cleavage domains. However, this interface is small (800 $Å^2$) compared to that of other dimeric restriction enzymes, such as BamHI, where the interface is about 1550 $Å^2$. The relatively small dimer interface of FokI explains why it exists as a monomer in solution[89]. The FokI endonuclease cleaves plasmids with two copies of its recognition sequence more rapidly than plasmids with a single copy[88], as NaeI and tetrameric restriction enzymes. However, instead of DNA-protein association being more favourable with a two-site substrate, FokI displays a more favourable protein-protein association on a two-site DNA[90]. The reaction pathway for FokI on the DNA with two target sites can be dissected into the following steps: the initial binding of FokI monomers to individual recognition sites, the subsequent association of the two monomers bound to their separate sites to form a dimer and, finally, phosphodiester hydrolysis (Figure 9B)[91].

Chimeric REases. The modular architecture of FokI enabled the creation of chimeric REases by fusion of the FokI non-specific catalytic domain with various DNA binding domains[92-94]. The most important group of such chimeric nucleases is the zinc finger nucleases, which are becoming important tools to modify plant and mammalian genomes via homology-directed repair of a targeted double-stranded break[95-97]. The zinc- finger motif is specific for a DNA base triplet, which can be easily manipulated experimentally. Moreover, these motifs can be combined in tandem to allow recognition of a wide variety of extended DNA sequences[96, 98].

In zinc-finger nucleases, the FokI catalytic domain is fused to three or four

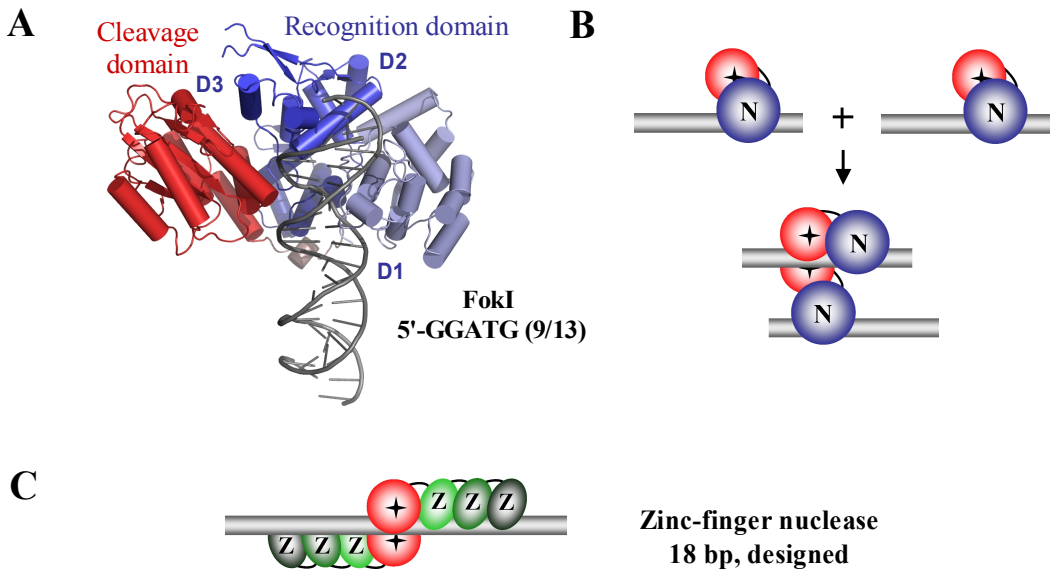

Figure 9. *(A) The structure of the FokI monomer bound to DNA (PDB:1fok)*[8]*. The recognition domain is colored blue, the catalytic domain is red, and DNA is grey. The catalytic domain is positioned away from the DNA. (B) A scheme for DNA cleavage by FokI. The color-coding is the same as above. Two monomers bound to separate target sites associate by their catalytic domains to introduce a double-stranded break*[87, 88]*. (C) A cartoon representation for the three-finger chimeric nuclease bound to DNA. The FokI catalytic domain (red) is fused to three zinc-finger domains (green). To introduce a double stranded break, this nuclease needs two copies of a 9 bp target sequence in an inverted orientation*[95]*.*

zinc fingers which form a DNA binding domain (Figure 9C)[95, 96]. As FokI, the zinc finger nucleases require dimerization of the catalytic domain to achieve a double-stranded break. To facilitate dimerization, these nucleases are typically designed to recognize two copies of the target sequence in a tail-to-tail conformation (Figure 9C). Since a three-finger nuclease requires two copies of a 9 bp recognition site, it effectively has an 18 bp recognition site, which is long enough to specify a unique genomic address in plants and mammals[96, 98].

The BspD6I/MleI/PlyI family enzymes. Another enzyme, composed of a PD-(D/E)XK nuclease domain fused to an HTH fold recognition domain, is the nicking endonuclease Nt.BspD6I (Figure 10A)[99]. Nt.BspD6I recognizes the 5'-GAGTC sequence and cleaves only one DNA strand, containing the target sequence, at the distance of four nucleotides from the recognition site[100]. Initially, Nt.BspD6I was considered as a naturally occuring nicking enzyme[100].

However, it was shown that the protein, encoded by the open reading frame adjacent to the end of Nt.BspD6I gene, exhibits high homology with the catalytic C-terminal domain of the nickase and comprises all the catalytic amino acid residues[101]. When it is added to reaction mixture, cleavage of both DNA strands near the sequence recognized by Nt.BspD6I occurs. Thus, this protein and Nt.BspD6I are the small and large subunits of the heterodimeric REase, BspD6I (5'-GAGTC(4/6))[101].

BspD6I is closely related to the Type IIS enzymes MlyI (5'-GAGTC (5/5)) and PleI (5'-GAGTC (4/5)), which recognize the same sequence as BspD6I, but cleave at different positions[102]. MleI and PlyI contain a single active site, and most likely they form a homodimer like FokI to cleave both DNA strands[102]. Indeed, it was shown that the disruption of the dimerization interface converts MlyI into a nicking enzyme[103].

The SdaI restriction endonuclease. In contrast to other enzymes recognizing palindromic sequences, which are single-domain proteins, SdaI (5'-CCTGCA↓GG) has a modular architecture[15]. The SdaI monomer is composed of two structural domains: the N-domain contains a winged-helix DNA binding motif, while the C-domain shows a typical PD-(D/E)XK fold (Figure 10B). According to mutational data, residues located on the putative recognition helix of the wHTH motif are crucial for the DNA binding specificity of SdaI. As HTH proteins usually interact with symmetric recognition sites as dimers, the dimerization of the N-domains of SdaI may be required to achieve the recognition of its palindromic 8 bp site[15].

1.3.2.5 REases with two catalytic domains

Separate domains for recognition and catalysis allow FokI and other Type IIS enzymes to cleave DNA at distant sites from the recognition sequence[7, 9]. However, REases that cleave both DNA strands very close to the recognition site are unlikely to have a FokI-like architecture because of the

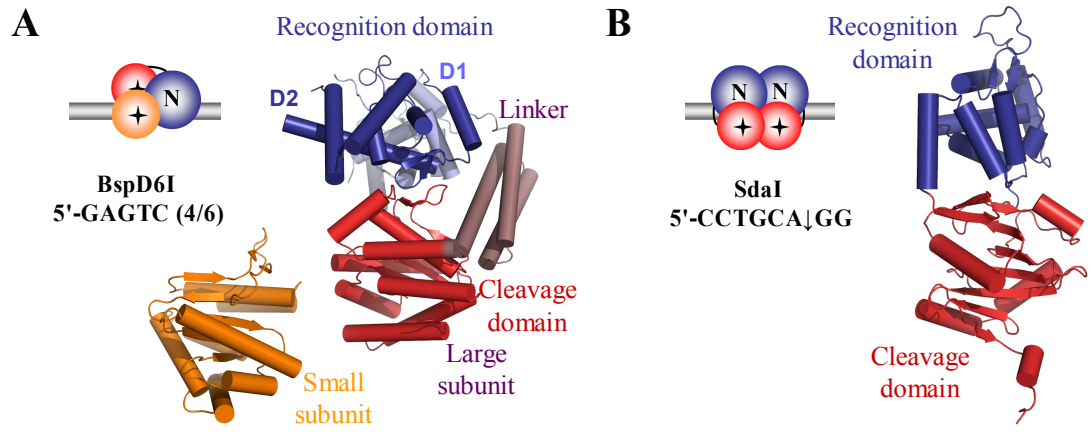

Figure 10. *(A) The subunits of the heterodimeric REase BspD6I. The large subunit (PDB:2ewf)[99] contains the recognition (blue) and catalytic domains (red), while the small subunit (PDB:2p14[99]; orange) contains an active site and is required to achieve the double-stranded break[101]. (B) The structure of SdaI (PDB:2ixs)[15]. The recognition domain is colored blue, the catalytic domain is red. It was proposed that SdaI dimerizes to recognize its palindromic 8 bp site[15]. The color coding in the cartoons representing the models of protein-DNA complexes is the same as in the corresponding structures.*

spatial constraints, provided by the proximity of the target sequence and the cleavage site. Mva1269I (GAATGC(1/-1)) is a monomer both in solution and upon binding of cognate DNA[104]. Fold-recognition analysis revealed that Mva1269I comprises two PD-(D/E)XK domains. The N-terminal domain is related to the REase EcoRI, whereas the C-terminal one resembles the nonspecific nuclease domain of FokI. The biochemical characterization of Mva1269I and its mutants indicates that the EcoRI-like domain incises the bottom strand within the target, and the FokI-like domain completes the cleavage near the target sequence[104].

1.3.2.6 REases with methyltransferase activity

Another group of REases are multi-functional proteins with both methyltransferase and endonuclease activities, but belong to Type II enzymes because they satisfy the overriding criterion: they cut DNA at specified rather than random positions[4]. Many of such enzymes recognize bipartite sequences and cleave at both sides of a target, excising a ~30 bp fragment. The current

nomenclature classifies the enzymes that cleave at both sides of target on both strands as Type IIB[4].

The best-characterized Type IIB enzyme BcgI (5'-(10/12) $CGAN_6TGC$ (12/10)) is composed of two subunits, A and B. Neither protein cleaves or modifies DNA, but together they form a complex at a ratio A_2B that does both[105, 106]. The regulation of these competing activities is determined by DNA substrates and cofactors[107]: BcgI is an active endonuclease on unmodified DNA substrates, requiring Mg^{2+} and AdoMet cofactors, and an active N6-adenine methyltransferase on hemimethylated DNA substrates, requiring AdoMet and stimulated by Mg^{2+}. Sequence analysis together with mutational data indicate that both the PD-(D/E)XK motif resembling the endonuclease active site and the hallmark motifs of N6-adenine methyltransferases are located in the BcgI A subunit[106, 108]. The B subunit seems to contain two target recognition domains (Figure 11). Its organization resembles the HsdS subunit of Type I restriction enzymes, where each target recognition domian recognizes one component of interrupted recognition sequence[106, 108]. BcgI cleaves two-site plasmids more rapidly than plasmids with one recognition site[109, 110]. This suggests that BcgI bridges two copies of the recognition sequence. Furthermore, it elutes as a hexamer $(A_2B)_2$ in gel-filtration experiments[109]. Therefore, it was proposed that the BcgI complex $(A_2B)_2$ binds two copies of cognate DNA, the heterotrimer A_2B being a minimal working unit that cleaves one recognition site[109, 110].

In contrast to BcgI, AloI (5'-(7/12-13) $GAACN_6TCC$ (12-13/7)) is a large polypeptide that possesses both endonuclease and N6-adenine methyltransferase activities[12]. Sequence analysis for AloI suggests three domains: an N-terminal PD-(D/E)XK nuclease; a middle part corresponding to a methyltransferase; and a C-terminal region containing two target recognition domains (Figure 11)[12]. The two target recognition domains of AloI recognize different parts of the bipartite DNA target. Novel specificities were generated

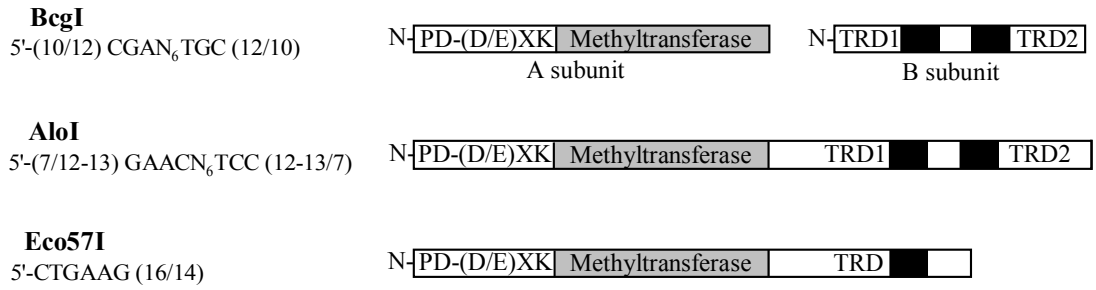

Figure 11. *Sequence regions of the bifunctional Type II enzymes BcgI, AloI and Eco57I. BcgI is composed of two subunits: the A subunit harbours the nuclease and methyltransferase active sites, while the B subunit contains two target recognition domains*[105, 108]*. AloI and Eco57 contain the all three functional regions in one polypeptide chain*[12, 111-113].

by a recombinational reassortment of the target recognition domains of AloI (5'-(7/12-13) GAACN$_6$TCC (12-13/7)) and related enzymes PpiI (5'-(7/12) GAACN$_5$CTC (13/8) and TstI (5'-(8/13) CACN$_6$TCC (12/7))[114].

Another bifunctional enzyme Eco57I (5'-CTGAAG (16/14)), which do not belong to Type IIB, also combines specific DNA recognition, cleavage and methylation activities in a single polypeptide[111, 112]. The PD-(D/E)XK nuclease active site is located near the N-terminus of Eco57I, while the central part contains the conserved motifs of N6-adenine methyltransferases, involved in AdoMet binding and catalysis. The C-terminal part could comprise the DNA recognition domain of Eco57I (Figure 11)[111-113]. The presence of both endonuclease and methytransferase activities in the same polypeptide chain was employed to alter the sequence specificity of Eco57I from 5'-CTGAAG to 5'-CTGRAG[115]. The procedure included three steps: (1) inactivation of the nuclease active site converting of the bifunctional RE into a methyltransferase; (2) mutagenesis and selection of mutants with altered DNA modification specificity; (3) reconstitution of the nuclease.

1.3.3 Mismatch repair endonucleases

In *E.coli* repair of base-base mismatches is performed by two overlapping

biochemical processes, methyl-directed mismatch repair and very short-patch repair[116]. While methyl-directed mismatch repair corrects replication errors, very short-patch repair deals with the C-G mispairs created by 5-methylcytosine deamination to T. The iniciation of both the processes involves PD-(D/E)XK nucleases: MutH[16] in methyl-directed mismatch repair and Vsr[117] in very short-patch repair.

MutH. Methyl-directed mismatch repair depends on the methylation status of the Dam methyltransferase target sites 5'-GATC, which allows to distinguish the parental strand from the daughter strand[118, 119]. After replication, the template strand is methylated in the 5'-GATC sequence, but the newly synthesized strand is not methylated until it is modified by Dam[120]. The repair pathway is initiated by the MutS protein, which recognizes a mismatched base pair. The mismatch-bound MutS associates with MutL and activates the MutH endonuclease[118, 119]. MutH cleaves the unmethylated daughter strand 5' to the hemimethylated 5'-GATC sequence, which can be separated from the mismatch by as much as 1000 bp. The nicked DNA strand containing an error is a substrate for further helicase and exonuclease activity (Figure 12A)[118, 119].

Thus, MutH is a sequence and methylation specific endonuclease. In contrast to restriction enzymes, it cleaves only one DNA strand. The crystal structure[121] revealed that MutH binds to DNA as a monomer, in a cleft between two arms (Figure 12B). MutH binds both hemi- and unmethylated DNAs, but the latter is more bent and distorted than the hemimethylated. The burried surface area in the crystal structures together with sedimentation equilibrium data indicate that hemimethylated DNA is bound more tightly by MutH[121].

Vsr. Vsr endonuclease is a both sequence- and mismatch-specific enzyme. It recognizes a G/T mismatch that results from the 5-methylcytosine deamination to T in the Dcm sites (5'-C(m5C)(A/T)GG), and nicks one DNA strand 5' to the incorrectly paired T[122]. The incision on the T strand by Vsr initiates the repair of the selected strand by DNA polymerase and DNA ligase

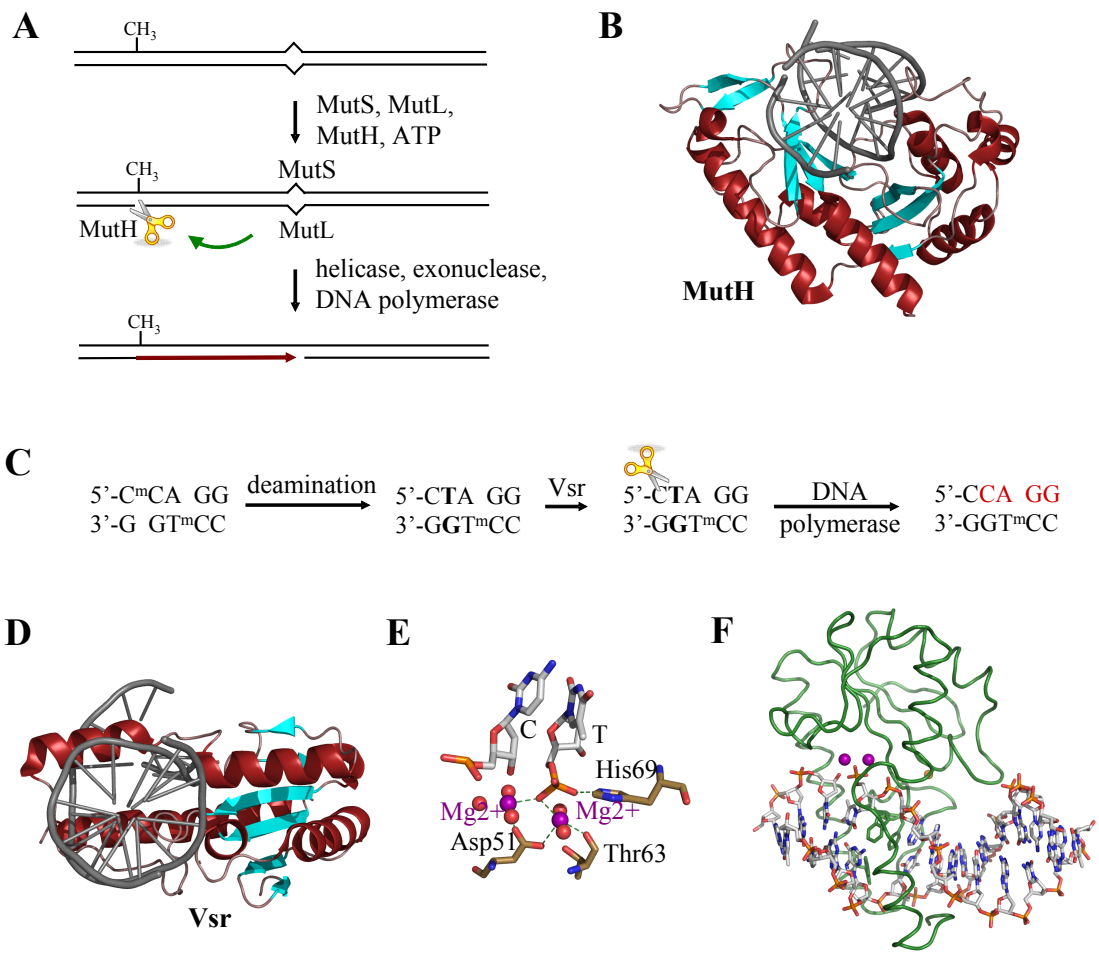

Figure 12. *(A) A generalized mechanism for E.coli methyl-directed mismatch repair. MutS, upon recognition of a mismatch, associates with MutL and activates MutH, which nicks the unmethylated strand 5′ to the hemimethylated 5′-GATC sequence*[119]*. (B) The structure of MutH bound to hemimethylated DNA (PDB:2aor)*[121]*. Helices are colored red, strands are in cyan, DNA is grey. (C) A scheme of very short-patch repair. Vsr nuclease introduces a nick near the incorrectly paired T resulting from deamination in the Dcm sites*[122]*. (D) The structure of the Vsr-DNA complex (PDB:1cwo)*[124]*. The color-coding is the same as above. (E) A close-up view of the unusual PD-(D/E)XK active site in the Vsr nuclease. There are only three direct contacts between the protein and two magnesium ions (purple spheres). The scissile phosphate is coordinated by His69 and both the metal ions. Water molecules are represented by red spheres. (F) Intercalation of Vsr into the DNA. Three aromatic residues, Phe67, Trp68 and Trp86 (shown in stick representation) of Vsr (green) intercalate into the DNA duplex*[124]*. The magnesium and zinc ions are shown as purple and orange spheres, respectively.*

(Figure 12C)[123].

Vsr is a monomeric single-domain protein. The overall fold of Vsr is related to REases (Figure 12D), but the active site seems to be unique to

Vsr[117, 124]. Instead of the highly coordinated catalytic metal ion(s) that are typically found in the PD-(D/E)XK family enzymes, Vsr uses two magnesium-water clusters. There are only three direct protein-Mg^{2+} contacts in the Vsr-DNA structure: the essential residue Asp51 directly coordinates the two magnesium ions, and the mainchain carbonyl of Thr63 coordinates to one of the metal ions. The phosphate at the cleavage termini is coordinated by the both metal ions and His69, another catalytically essential residue (Figure 12E)[124, 125]. The structure of Vsr also contains a zinc atom (Figure 12F), which plays a purely structural role[124].

Unlike most DNA repair enzymes, which use nucleotide flipping, Vsr endonuclease recognizes the TG mismatch as a wobble base pair. A striking feature of the Vsr-DNA complex is the intercalation of three aromatic residues into the major groove of the DNA (Figure 12F) on one side of the mismatch[124, 125].

1.3.4 Other sequence-specific nucleases

TnsA. Tn7 transposase is a heteromeric complex of two proteins, TnsA and TnsB[126]. TnsA mediates DNA cleavage reactions at the 5' ends of the transposon, while TnsB has a dual role: it catalyzes the cleavage at the 3' end and then activates the newly created 3'-OH groups for the subsequent strand transfer reactions (Figure 13A). TnsB appears to be a member of the retroviral integrase superfamily[126]. The DDE active site was also proposed for TnsA[126], but the crystal structure revealed the PD-(D/E)XK fold [127].

TnsA is a monomer in the crystal sructure and contains two domains (Figure 13B)[127]. The N-terminal domain has a PD-(D/E)XK fold and is most similar to the FokI catalytic domain. The C-terminal domain contains an HTH motif. It is unlikely, however, that this domain binds DNA: the helix corresponding to recognition helix of HTH proteins contributes to the interdomain interface and seems to be unable to form protein-DNA contacts;

and surface electrostatic potential of this domain is nearly featureless, with many solvent-exposed hydrophobic patches[127].

The homing endonuclease I-Ssp6803I. Homing is a transfer of an intron or intein to a homologous allele that lacks the recognition sequence[128]. Transfer of group I introns and inteins is initiated solely by a homing endonuclease activity. The cellular mechanisms that repair double-stranded breaks via homologous recombination complete the transfer, using the intron- or intein-containing allele as a template (Figure 13C). Homing endonucleases recognize DNA target sites ranging from 14 to 40 bp in length, balancing between high specifity of sequence recognition and flexibility to permit transfer in diverging

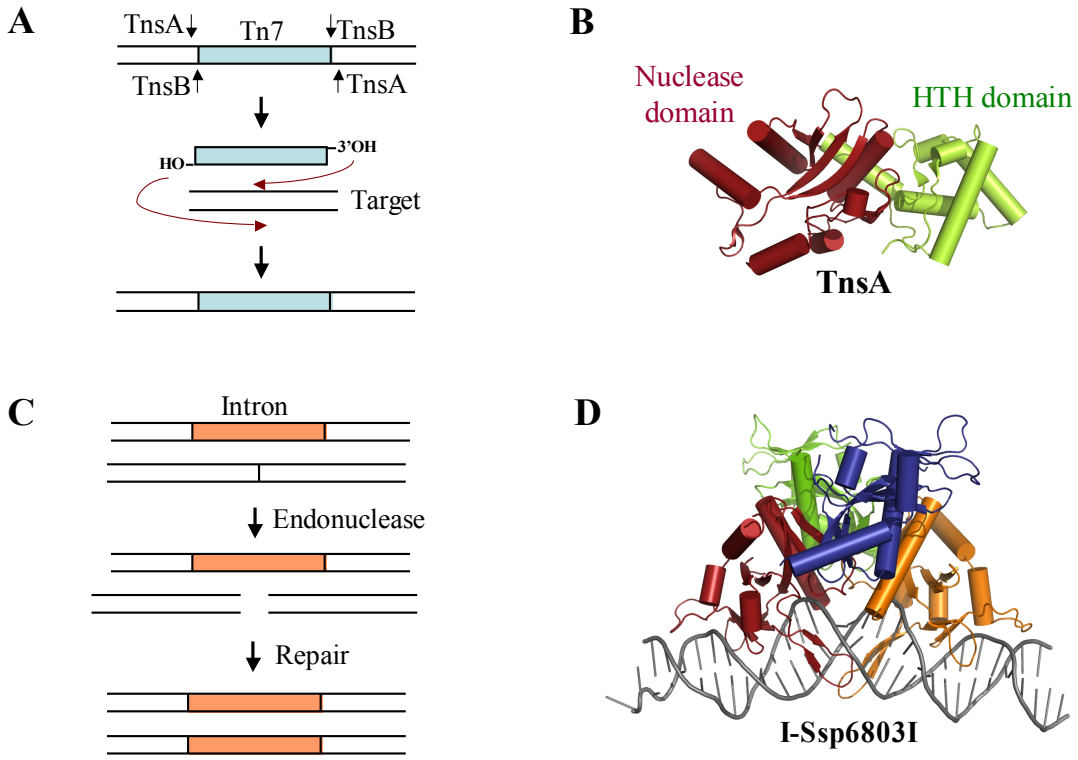

***Figure 13**. (A) A scheme for Tn7 transposase action. TnsA mediates the 5'-incision, while TnsB catalyses the 3'-incision and strand transfer reactions[126, 127]. (B) The structure of TnsA (PDB:1f1z)[127]. The PD-(D/E)XK domain is colored red, the HTH domain is green. (C) A generalized homing mechanism for group I introns. A homing endonuclease introduces a double-stranded break, which is repaired via homologous recombination (HR), using the intron-containing allele as a template[128]. (D) The structure of the homing endonuclease I-Ssp6803I tetramer bound to DNA (PDB:2ost)[127]. The four protomers are colored red, orange, blue and green, DNA is grey.*

targets and hosts[128]. Based on conserved protein sequence motifs, homing endonucleases are grouped into four families: LAGLIDADG, GIY-YIG, HNH and His-Cys box[128]. Surprisingly, the homing endonuclease I-Ssp6803I, which causes the insertion of a group I intron into a bacterial tRNA gene, was shown to be a member of the PD-(D/E)XK family[129, 130].

I-Ssp6803I recognizes an approximately 23 bp target, which corresponds to the pseudopalindrome 5'-TCGTCGGGCTCAT↓AACCCGAAGG, and cleaves both the DNA strands leaving 3 bp 3'-overhangs[129, 131]. The structure of the I-Ssp6803I/DNA complex shows a protein tetramer bound to one DNA duplex (Figure 13D). Such DNA:protein binding stoichiometry was also confirmed by isothermal titration calorimetry. The DNA is bound by two of the four I-Ssp6803I protomers. The DNA-bound protomers are much more loosely associated with one another than subunits of restriction endonucleases. Due to the loose association of the I-Ssp6803I structural cores, their elaborations can reach the distal ends of the long target site. The tetrameric assembly of the homing endonuclease seems to stabilize the functional dimer, allowing the core PD-(D/E)XK domains to be far apart on the long target[129].

1.3.5 Structure-specific nucleases

In contrast to sequence-specific nucleases, structure-specific nucleases recognize a DNA junction that contains a double stranded (ds)DNA or a single-stranded (ss)DNA branch from dsDNA: 5'- and 3'-flaps, bubbles, replication forks and Holliday junctions (Figure 14)[17]. These irregular structures arrise during DNA replication, repair and recombination, and must be resolved to duplex DNA upon termination of the process. In most cases, the resolution of the junction requires recognition and cleavage by a structure-specific nuclease[17].

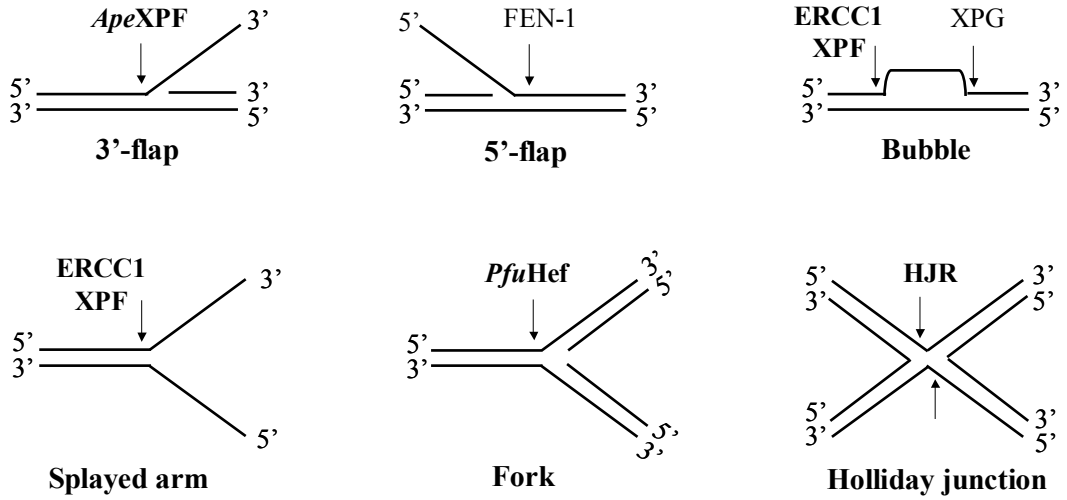

Figure 14. *Branched DNA structures and enzymes that recognize and cleave these structures. The cleavage positions are indicated by arrows. Nucleases that belong to the PD-(D/E)XK superfamily are highlighted in bold. The figure was adapted from Nishino et al*[17]

1.3.5.1 Holliday junction resolvases

Homologous recombination is a crucial process for generating of genetic diversity and repair of DNA lesions. The key intermediate in homologous recombination is the Holliday junction (Figure 14), a four stranded DNA structure that links two DNA duplexes[132, 133].

Holliday junction resolving nucleases were isolated from many organisms[17, 134]. They specifically recognize the four way structure and introduce two nicks, located symmetrically across the point of strand exchange (Figure 14), to create the duplex products. All Holliday junction resolvases are highly basic (pI ≥ 9.4) dimeric enzymes. However, they show high variability in sequence and structure[17, 134]. Bioinformatics approach classifies most of the Holliday junction resolving enzymes into two groups: the group that includes integrases and ribonuclease H, and the group of PD-(D/E)XK nucleases. Endonuclease VII of phage T4 and *E.coli* RusA fit in neither group[134, 135].

The crystal structures were determined for five Holliday junction resolvases belonging to the PD-(D/E)XK superfamily: endonuclease I from

bacteriophage T7[136], Hjc from *Pyrococcus furiosus* (*Pfu*Hjc)[137], Hjc and Hje from *Sulfolobus solfataricus* (*Sso*Hjc and *Sso*Hje)[138, 139] and RecU, the general Holliday junction resolving enzyme in Gram-positive bacteria[140]. Although these nucleases show topological similarity to restriction endonucleases, their dimerization mode is completely different. The quarternary structure also differs among Holliday junction resolvases, suggesting distinct mode of substrate recognition[17, 134]. The archaeal Hjc enzymes (*Pfu*Hjc and *Sso*Hjc) and RecU from *Bacillus subtilis* maintain a similar dimerization interface, formed mainly by the N-terminal portion of the main ß-sheet. RecU also harbors a "stalk" structure, formed by two loops from different monomers, which protrudes from the main body of the enzyme[140]. Contrary, the *Sso*Hje nuclease, which generates distinct cleavage pattern from *Sso*Hjc, shows different relative orientation of the partner subunits (Figure 15A)[138]. The homodimer of T7 endonuclease I is unique between PD-(D/E)XK nucleases. The enzyme is arranged in two well-separated domains, connected by a ß-bridge. Each domain of the T7 endonuclease I is composed of elements from both the subunits, and the catalytic residues are contributed by both the protomers of the dimer (Figure 15A)[136, 141].

The PD-(D/E)XK fold resolvases recognize the four-way junctions without a pronounced sequence specificity, except for RecU, which has a preference for cutting within the sequence 5'-$^{G}/_{T}$G↓C$^{A}/_{C}$[134, 142]. Recognizing the structure of the four-way junction, the resolving enzymes distort it in a pronounced way. The crystal structure[143] of T7 endonuclease I in complex with a DNA junction revealed that it binds the helical arms of the junction in two almost perpendicular channels, formed by the front of one domain and the back of the other (Figure 15B). The channels are positively charged, and they form extensive contacts with the phosphate backbone. These interactions ensure that only branched structure can be tightly bound. The pairs of the DNA helical arms in the T7 endonuclease I/junction structure are essentially coaxial, except

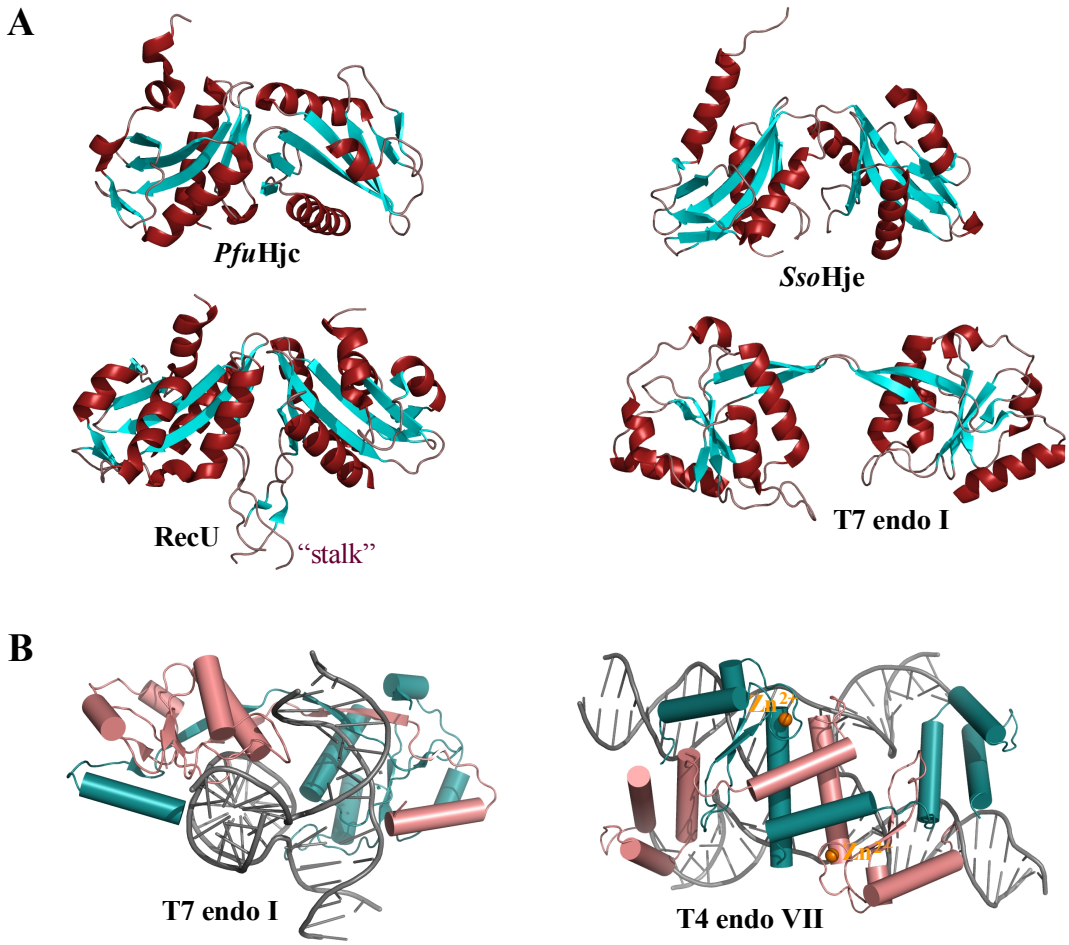

Figure 15. *Holliday junction resolving enzymes. (A) The structures of PfuHjc (PDB:1gef)*[137]*, SsoHje (PDB:1ob8)*[139]*, RecU (PDB:1zp7)*[140] *and T7 endonuclease I (PDB:1fzr)*[136]*. Helices are colored red, strands are in cyan. (B) The structures of T7 endonuclease I (PDB:2pfj)*[143] *and T4 endonuclease VII (2qnc)*[148] *bound to Holliday junction. The two protomers of the dimer are colored red and cyan, DNA is in grey. The zinc ions present in the T4 endo VII structure are shown as orange spheres.*

for the central bases that are unstacked[143]. Such conformation of the junction contrasts with that in the other available structure of a resolvase/substrate complex. In complex with T4 endonuclease VII, the junction is distorted but it adopts an approximately planar structure (Figure 15B)[143, 144].

The distortion of the junction could be connected to DNA cleavage. Experiments using cruciform structures in supercoiled DNA molecules (a self-limiting substrate) revealed that T7 endonuclease I cleaves both DNA strands within the lifetime of the complex[141, 144]. A model was proposed to explain this

observation: the initial complex containing the intact junction is strained, and the cleavage of the first DNA strand is followed by a relaxation that accelerates the second strand cleavage.

1.3.5.2 The XPF/Mus81 family enzymes

Many eukaryal and archaeal nucleases that recognize 3'-flap structures (Figure 14) belong to the XPF/Mus81 family[17]. Members of the XPF/Mus81 family have two characteristic domains[145]: a restriction-endonuclease like catalytic domain with a signature motif GD...ERK (a variant of PD-(D/E)XK), and a helix-hairpin-helix (HhH) domain (Figure 16A,F). The HhH domain binds DNA in a sequence-independent manner[146]. Although related to bacterial nucleases, the XPF/Mus81 family enzymes are present throughout the eukaryal and archaeal kingdoms but not in prokaryotes[17, 145]. XPF (Xeroderma pigmentosum group F) associates with a noncatalytic partner ERCC1 (excision repair cross complementarity group 1) to form a structure-specific endonuclease which is responsible for the 5'-incision during the mammalian nucleotide excision repair[147]. Mus81 is a related eukariotic nuclease, which cleaves stalled replication forks and other branched DNA substrates[145]. It acts in complex with a noncatalytic partner, Eme1(Mms4). In contrast to eukaryal enzymes, archaeal XPF homologs do not associate with noncatalytic subunits. They cleave replication fork or 3'-flap structures as homodimers[17, 145].

Archaeal nucleases. In archaeal XPF homologs, the catalytic domain is followed by two consecutive HhH motifs[145]. Hef (helicase associated endonuclease for forked structure) from *Pyrococcus furiosus* also contains a helicase region, which precedes the catalytic domain in the sequence. The Hef helicase is active and seems to work cooperatively with nuclease in processing of stalled replication forks[149]. *Sulfolobus solfataricus* and *Aeropyrum pernix* contain a different form of XPF homologs (*Sso*XPF and *Ape*XPF) that lacks the helicase region, but requires the sliding clamp PCNA for DNA cleavage

(Figure 16A)[145].

The crystal structure of the separated Hef catalytic domain showed that it folds into a single-domain structure, mostly similar to Vsr and Hjc (Figure 16B)[29]. The Hef HhH domain is a compact structure of two HhH motifs, bridged by a fifth connecting α-helix (Figure 16C)[150]. Both the catalytic and HhH domains of Hef nuclease independently form dimers. The structures of individual domains together with the biochemical data, including site-directed hydroxyl-radical footprinting of a protein-DNA complex, led to a model[150] for how Hef nuclease recognizes the fork-structured DNA. According to this model, the catalytic domain binds in the vicinity of the junction centre, while the HhH domains bind to the duplex regions at more distant positions (Figure 16E).

The independently solved structures of intact *Ape*XPF and its complex with DNA duplex[151] revealed the catalytic and HhH domains, similar to those of Hef nuclease. The *Ape*XPF–DNA structure comprises the protein dimer, asymmetrically bound to one DNA duplex. Only one protomer of *Ape*XPF (protomer A) contacts the DNA by its nuclease and HhH domains (Figure 16D). Moreover, this DNA-bound protomer is more compact than DNA-free protomer B and has an ordered interdomain linker. The HhH domain of the protomer B is also capable of binding DNA. The modelling of a DNA duplex bound to the HhH domain of the protomer B puts it at 90° to the experimentally observed DNA, one end of the modelled duplex being close to the active site of the nuclease domain of protomer A[151]. This leads to the same model for interaction with branched DNA, as was proposed for the Hef protein (Figure 16E)[152].

The model[150-152] explains the substrate specifity of Hef and *Ape*XPF. First, in order for both the HhH domains to bind DNA simultaneously, the substrate would have to bend by almost 90°. As DNA duplexes cannot readily be distorted in this manner, XPF's may recognise ds/ssDNA junctions by their

compressibility. Also, Hef and *Ape*XPF do not interact with the region

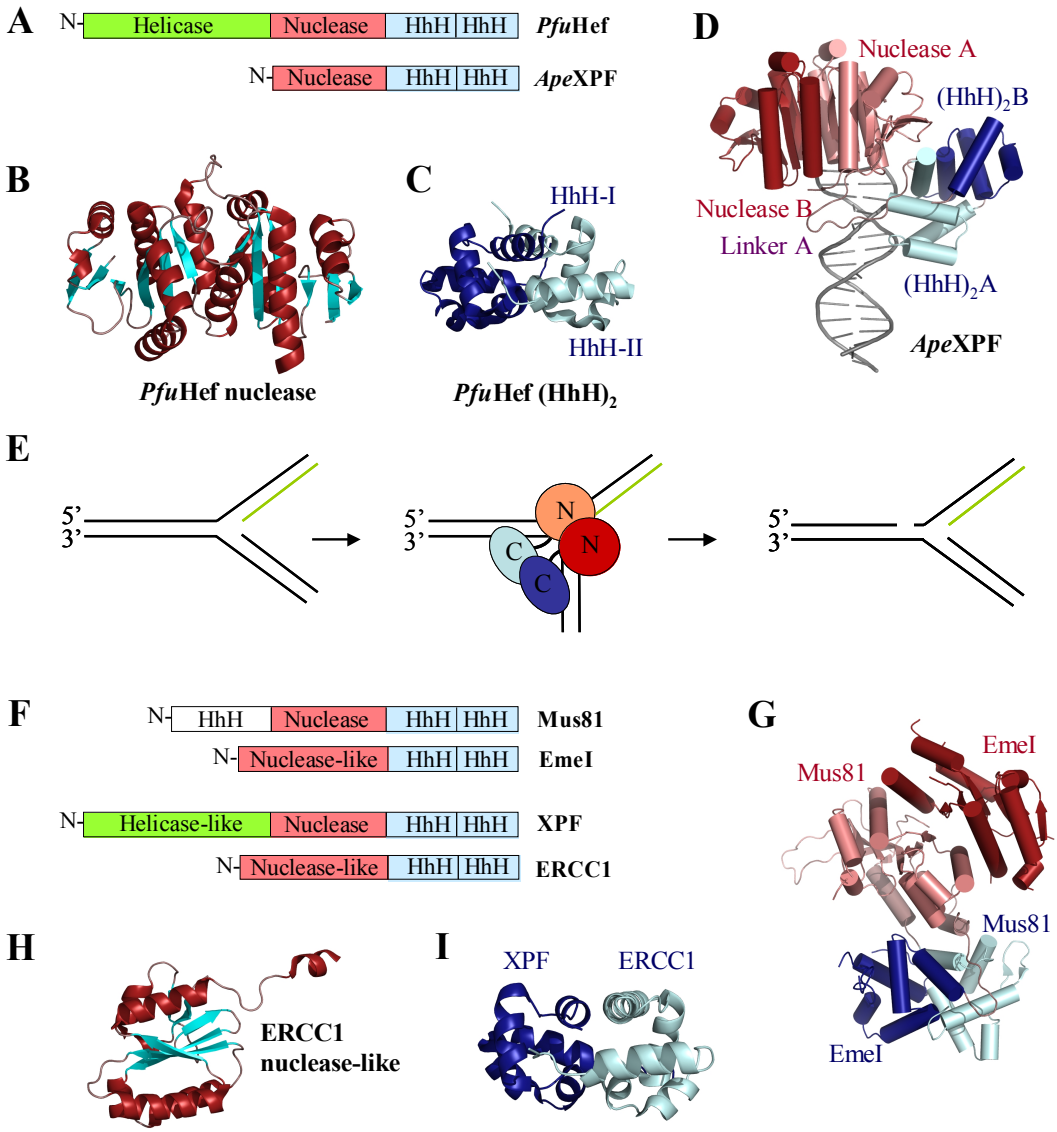

Figure 16. *The Mus81 family enzymes. Nuclease domains are highlighted in red, HhH domains are blue. (A) The domain organization of archaeal XPF enzymes. (B) The structure of the PfuHef nuclease domain (PDB:1j23)[29]. Helices are colored red, strands are in cyan. (C) The structure of the PfuHef HhH domain dimer (PDB:1x2i)[150]. Here and below, different protomers are shown in different shades. (D) The structure of ApeXPF bound to DNA duplex (PDB:2bgw)[151]. (E) A mechanism for branched DNA recognition by PfuHef and ApeXPF[150-152]. The HhH domains bind to the duplex arms bending the substrate, while the nuclease binds and cleaves near the center of the junction. The strand shown in green is missing in 3'-flap substrates recognized by ApeXPF. (F) The domain organization of eukaryal XPF enzymes. Mus81-EmeI and XPF-ERCC1 are heterodimers. (G) The structure of Mus81-EmeI (PDB:1ziu)[155]. (H) The structure of the ERCC1 nuclease-like domain (PDB:2a1i)[156]. Helices are colored red, strands are in cyan. (I) The structure of the ERCC1-XPF HhH heterodimer (PDB:2a1j)[157].*

downstream of the cleaved strand, which can be variable in fork and flap structures (Figure 16E). Actually, both Hef and *Sso*XPF (closely related to *Ape*XPF) nucleases can effectively cleave nicked DNA substrates, completely lacking this region[153, 154].

Although archaeal XPF's are arranged as homodimers and contain two active sites, they specifically cleave only one of the strands near the center of the junction. It was shown that the heterodimer of Hef nuclease, containing one intact and one inactivated catalytic domain, was fully active[150]. Thus, only one of the two active sites is required for forked-DNA cleavage by the Hef dimer[150, 152].

Eukaryal nucleases. Most eukaryotes have at least four XPF family members that assemble into two heterodimeric nucleases with distinct specifity[17, 145]: XPF-ERCC1, participating in the nucleotide excision repair pathway, and Mus81-EmeI, which cleaves stalled replication forks and other branched structures. Both XPF-ERCC1 and Mus81-EmeI are composed of catalytic (XPF and Mus81) and noncatalytic (ERCC1 and EmeI) subunits. The catalytic forms share the structural core, composed of the catalytic and HhH domains, with archaeal nucleases. This structure can be decorated by additional motifs and domains: an additional N-terminal HhH motif in Mus81, and an innactivated helicase domain in XPF (Figure 16F). The noncatalytic subunits, ERCC1 and EmeI, retain the same modular structure, but their counterparts of the catalytic domain lack the active site residues[17, 145].

The structure of truncated Mus81, corresponding to the intact structural core, was solved in complex with EmeI but without DNA[155]. This structure of the Mus81-EmeI heterodimer resembles *Ape*XPF, although the overall orientation of the catalytic and HhH domains is different (Figure 16G). Moreover, the interdomain linker regions of Mus81 and Eme1 are α-helical and not flexible as in archaeal XPF's. Biochemical data indicate that both the HhH domains of Mus81 and Eme1 interact with dsDNA simultaneously and the

linker plays an important role in DNA binding[155].

The separately solved structures of the ERCC1 catalytic domain (Figure 16H) and the heterodimer of the XPF-ERCC1 HhH domains (Figure 16I) also show similarity to the structures of archaeal nucleases[156, 157]. However, DNA binding studies indicate that both the catalytic and HhH domains of XPF-ERCC1 associate with single-stranded rather than double-stranded DNA. This could explain the preference of XPF-ERCC1 for DNA bubble-like substrates over 3'-flaps (Figure 14), recognized by archaeal enzymes[156, 157].

1.3.6 NTP-driven molecular machines

The PD-(D/E)XK nuclease fusions with helicase domains are found in Type I and Type III restriction endonucleases, McrBC and RecBCD. All these enzymes are complex molecular machines requiring hydrolysis of ATP (or GTP) for nuclease activity.

1.3.6.1 Type I REases

Type I REases[158-160] are more complex than Type II enzymes but historically they were the first to be detected and purified. They recognize asymmetric bipartite sequences (e.g. 5'-AACNNNNNNGTGC, EcoKI) and act as methyltransferases or endonucleases depending on the methylation status of the target. Type I REases methylate specific adenine residues within the recognition sequence, or if the target is unmethylated, translocate DNA in an ATP-dependent manner and cleave it thousands of base pairs from the recognition sequence. Methylation is dependent on AdoMet and Mg^{2+}, while restriction requires three co-factors: Mg^{2+}, AdoMet, and ATP. The oligomeric complexes of Type I REases are composed of three subunits, HsdS, HsdM and HsdR (or S, M and R, respectively)[158-160].

HsdS is a specificity subunit that recognizes the target sequence and also

provides a scaffold for other subunits to bind. Two variable regions in the HsdS sequence form two target recognition domains. Each domain indepedently binds to one part of the bipartite target sequence, therefore new specifities can be generated by their recombination[158-160]. The crystal structures of the S subunits from *Methanococcus jannaschii*[161] and *Mycoplasma genitalium*[162] show that the target recognition domain folds into a globular structure, mostly similar to the small domain of the Type II N6-adenine methyltranferase TaqI. The two globular target recognition domains are separated by two long antiparallel α-helices, which form a coiled-coil structure (Figure 17A). This region might act as a molecular ruler that provides a proper physical space between the recognition domains to accommodate the nonspecific region of the target sequence[161, 162].

The HsdM subunit is responsible for AdoMet binding and methyltranferase catalysis. Binding of two M subunits to one S subunit results in a M_2S_1 complex, which is a functional methyltransferase[158-160]. According to a model of M.EcoKI-DNA complex[163] based on electron microscopy data, DNA is bound to the target recognition domains of HsdS. The substrate contains a sharp bend in the non-specific spacer part of the target sequence. The N-terminal domains of HsdM bridge over the DNA duplex and form a clamp-like structure encircling the DNA. This model also suggests that the C-terminal region of HsdM interacts with the coiled-coil in HsdS (Figure 17B)[163].

The methyltransferase core recruits two HsdR subunits to form an endonuclease complex ($R_2M_2S_1$), which is capable of both methylation and cleavage[158-160]. The R subunit of EcoKI was shown to comprise several functional regions[164]. A PD-(D/E)XK catalytic domain is located at the N-terminus of EcoKI HsdR, while the central part contains the sequence motifs, characteristic for superfamily II helicases. The C-terminal region of HsdR seems to participate in protein-protein interactions[164]. The crystal structure of the EcoR124I motor subunit revealed four globular domains: an

N-terminal PD-(D/E)XK domain, RecA-like helicase domains 1 and 2 and a C-terminal helical domain, which has no apparent structural relatives (Figure 17C)[165].

As a large number of other superfamily II helicases, the R subunits of Type I REases move along the intact duplex and do not directly unwind DNA. Thus, although classified as helicases, they actually are dsDNA translocases[166]. EcoR124I was shown to translocate along the 3'→5' strand of intact dsDNA in steps of 1–2 bp, consuming ~1 ATP for each base pair moved[167, 168]. The two R subunits in the $R_2M_2S_1$ endonuclease complex translocate DNA independently.

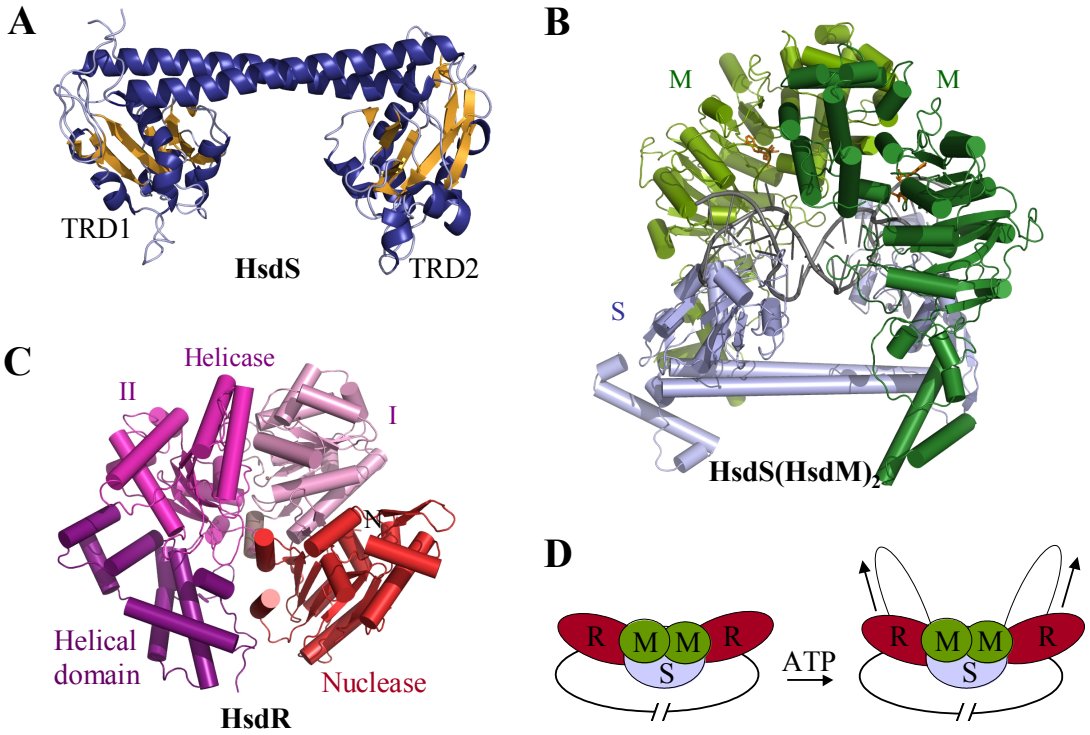

***Figure 17**. Type I REases. (A) The structure of the HsdS subunit from Methanococcus jannaschii (PDB:1yf2)[161]. Helices are colored blue, strands are in orange. Two globular target recognition domains (TRDs) are separated by a coiled-coil structure. (B) The model of the EcoKI methyltransferase core bound to DNA[163]. The HsdS subunit is colored blue, while two protomers of HsdM are shown in different shades of green. DNA is grey, AdoMet bound by HsdM is shown as orange sticks. The coordinates of the model were downloaded from ftp://ftp.genesilico.pl/iamb/models/MTases/EcoKI/. (C) The structure of the EcoR124I HsdR subunit (PDB:2w00)[165]. The nuclease domain is red, the two helicase domains and the C-terminal helical domain are colored in different shades of magenta. (D) A schematic representation of DNA translocation by Type I enzymes[169]. Two R subunits translocate DNA in opposite directions extruding two DNA loops.*

The translocating subunits remain bound to the methyltransferase core, which in turn is bound to the recognition sequence. Consequently, two DNA loops are extruded, one on each side of the occupied target site (Figure 17D)[169].

It is assumed that stalling of the translocation triggers DNA cleavage[170]. Linear substrates with one target site are not efficiently cleaved by Type I REases, in contrast to circular one-site DNA or linear DNA with two targets. According to a model of Type I enzymes action[170], circular one-site DNA can be cut when the entire circle has been translocated, while the collision of two translocating complexes could stimulate the cleavage of a two-site substrate. However, AFM experiments[171] revealed that EcoKI rapidly dimerizes when bound to a plasmid containing two recognition sites. The dimerization is folloved by the ATP-dependent bi-directional translocation and cleavage of the DNA. The cleavage could occur when the translocation is stalled because of either excessive DNA supercoiling or maximal contraction of the DNA loop the two bound enzyme molecules[171].

The $R_2M_2S_1$ complex of Type I enzymes contains two nuclease active sites at the N-terminus of the R subunits. However, according to the current model, they are positioned at distant DNA sites and could not cooperate in cleavage of two DNA strands[159, 160]. Thus, it was proposed that a double-stranded break results from cooperation between a translocating HsdR subunit and another HsdR subunit, either free in solution or as a part of an assembled enzyme[172].

1.3.6.2 Type III REases

Type III REases also are multi-functional complexes. They recognize asymmetric sequences and cleave DNA ~25-27 bp downstream of the target[158-160]. The best-characterized Type III enzymes EcoP1I (5'-AGACC) and EcoP15I (5'-CAGCAG) are composed of two subunits, Mod and Res, at a ratio $(Mod)_2(Res)_2$[173]. The Mod subunit contains the AdoMet binding and catalytic motifs, characteristic for N6-adenine methyltransferases[174], and can function

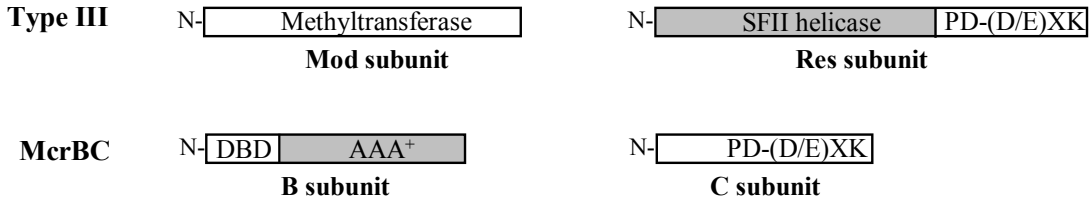

***Figure 18**. Sequence regions in Type III REases and McrBC. Type III enzymes are composed of two subunits: the Mod subunit harbours the methyltranferase motifs, while the Res subunit contains the functional motifs of superfamily II helicases and the PD (D/E)XK active site*[173-175]*. In the McrBC complex, DNA binding domain and AAA+ ATPase functional motifs are located in the B subunit, while the active site is located in the C subunit*[181, 182]*.*

independently of Res subunits. The Res subunit, required for DNA cleavage, is composed of two domains: a large N-terminal domain, containing functional motifs of superfamily II helicases, and a small C-terminal domain, which harbours a PD-(D/E)XK catalytic center (Figure 18)[173, 175].

For efficient cleavage Type III enzymes require two unmethylated target sites in a head-to-head orientation, e.g. 5'-CAGCAG$N_{0\text{-}3500}$CTGCTG for EcoP15I[176]. The cleavage can be blocked by Lac represor bound between two recognition sites, thus it was proposed that the collision of two translocating complexes triggers nuclease activity. However, the rate of ATP hydrolysis by Type III enzymes is very low (~1%) when compared to that of Type I complexes[177]. Unless Type III REases are the most energy-efficient DNA translocases known, it is unlikely that the entire distance between two target sites is translocated. Indeed, AFM imaging showed that EcoP15I, when bound to its recognition site, forms DNA loops not only via translocation, but also by making nonspecific contacts with other DNA sites[178, 179]. Thus, EcoP15I seems to use both translocation and diffusive looping to communicate between two distal DNA sites.

1.3.6.3 McrBC

McrBC from *E.coli* is the best-characterized enzyme among restriction

endonucleases that cleave DNA containing modified bases[4, 159]. This is a fundamentally different way to discriminate between foreign and self DNA: a REase recognizes the modification pattern of foreign DNA, while self DNA is not cleaved as it is not modified. Consequently, these enzymes are not associated with methyltranferase activity[52]. McrBC recognizes the 5'-R^{m}C sequence, where ^{m}C can be 5-methylcytosine, 5-(hydroxymethyl)-cytosine, or N4-methylcytosine, and cleaves DNA at multiple positions between two 5'-R^{m}C sites. Another interesting feature of McrBC is its requirement of GTP hydrolysis for restriction activity[180].

The McrBC system is composed of two subunits, McrB and McrC[180]. McrB contains an N-terminal DNA recognition domain and a C-terminal region, responsible for GTP hydrolysis and interaction with McrC[181]. The PD-(D/E)XK catalytic site is located in the McrC subunit[182] (Figure 18). The GTP-binding region of McrB shows similarity to the AAA^{+} family of proteins (AAA^{+} stands for ATPases associated with various cellular activities), which often perform chaperone-like function in assembly, function or disassembly of protein complexes. The AAA^{+} family proteins oligomerize into hexameric or heptameric ring structures[183, 184]. In presence of GTP, McrB also associates into heptameric complexes that can stack to form tetradecameric cylindrical structures[185]. McrC stabilizes the tetradecameric structure of McrB. Up to two McrC molecules can bind to the McrB tetradecamer, but the exact stoichiometry and the location of McrC in the McrBC complex is unclear[185].

1.3.6.4 RecBCD

RecBCD from *E.coli* is a molecular machine that processes free duplex DNA ends[186]. Linear dsDNA can generated by various DNA-damaging agents, such as ionizing radiation, or form indirectly as a consequence of replication through imperfect templates. RecBCD recognizes blunt dsDNA and cleaves it to produce single-stranded 3'-overhangs for homologous recombination. In

addition to the iniciation of recombinatorial DNA repair, RecBCD degrades foreign DNA that has a free dsDNA end. The two contradictory functions of the RecBCD enzyme require a mechanism to distinguish between self and foreign DNA. This is based on the recognition of the Chi sequence (5-GCTGGTGG), the most over-represented octamer in the *E.coli* genome[186, 187]. Initially, RecBCD unwinds double-stranded DNA and degrades the 3'-terminal strand, producing short oligonucleotide fragments, while the 5'-terminal strand is cut less frequently. The interaction with the Chi sequence attenuates the cleavage of the 3'-terminal strand, thus enhancing the recombination function of the RecBCD enzyme[188].

RecBCD is a heterotrimer composed of three subunits: RecA, RecB and RecC[189]. The RecB and RecD subunits harbour the characteristic motifs of superfamily I helicases[166] and can independently unwind DNA: RecB is a 3'→5' helicase, and RecD is a 5'→3' helicase. Both the helicases are active in the RecBCD complex, travelling with opposite polarities, but in the same direction, on each strand of the antiparallel DNA duplex[190].

The RecB subunit contains a PD-(D/E)XK catalytic site at the C-terminus. This is the single nuclease active site of RecBCD, which is responsible for DNA degradation in both directions during the reaction with double-stranded DNA[191, 192]. The isolated C-terminal domain of RecB efficiently cleaves both linear and circular single- and double-stranded DNA[193]. As endonuclease activity on dsDNA was never seen for intact RecBCD, the catalytic activity of the nuclease domain should be modulated when it is a part of the RecBCD holoenzyme.

The crystal structure of the entire RecBCD complex bound to the DNA substrate[19] revealed the architecture of this complex machine. The N-terminal part of RecB folds into two core helicase domains (1A and 1B) and two auxilary domains: 1B, containing an 'arm' structure, and 2B. The RecB nuclease domain is connected to the remainder of the protein by a long linker

region of about 70 amino acids. The RecD subunit forms three domains: the helicase motor domains 1A and 1B, and an N-terminal, which contacts RecB (Figure 19A)[19]. Although the RecC subunit lacks functional helicase and nuclease motifs, its three-dimentional structure resembles the RecB helicase

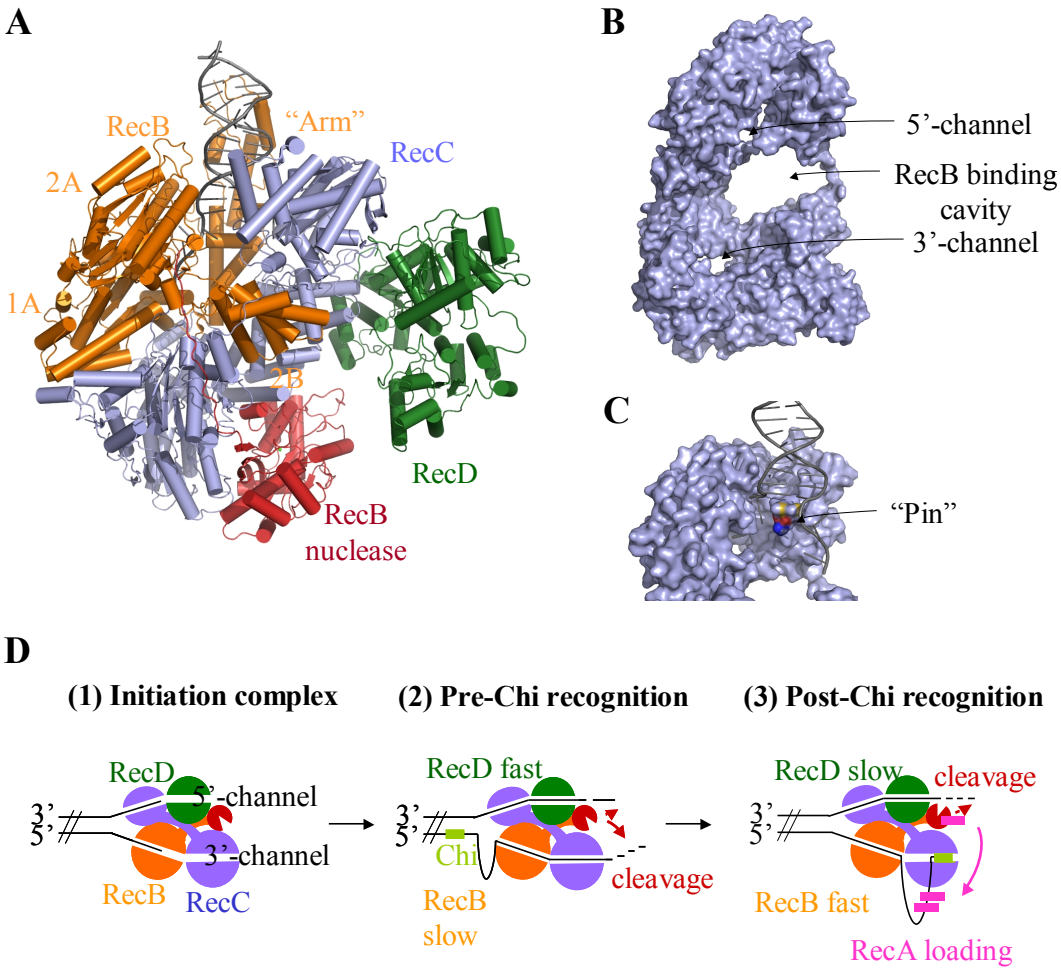

Figure 19. *The RecBCD complex. (A) The structure of RecBCD bound to DNA (PDB:1w36)[19]. RecB is colored orange with the nuclease domain highlighted in red, RecC is blue, RecD is green, and DNA is grey. The DNA is contacted by the the RecB and RecC subunits. (B) The surface of the RecC subunit, showing a hole for RecB 2B domain and two channels for ssDNA. (C) A close-up view of the RecC methionine "pin", which splits DNA duplex. (D) A mechanism for RecBCD action. The color-coding is the same as above. (1) In the initiation complex, seen in the crystal structure, the unwound DNA duplex end is bound by the RecB and RecC subunits. (2) Before Chi recognition RecBCD is a bipolar helicase, RecD travelling faster than RecB. The 3'-tail is hydrolyzed much more vigorously than the 5'-DNA tail. (3) Upon Chi (green rectangle) recognition, RecB is the lead motor and the 5'-tail is cleaved more frequently, as the 3'-tail remains bound by RecC. The RecB nuclease domain loads RecA protein (magenta rectangle) on the growing ssDNA loop downstream of the Chi sequence. The figure was adapted from Dillingham and Kowalczykowski[186].*

domains at the N-terminus, and the C-terminus is similar to the RecB nuclease domain[19, 194]. The inactivated RecC helicase and nuclease domains form a large hole and two channels running through the protein (Figure 19B). The hole accommodates the 2B domain of RecB, while the two channels are for single-stranded DNA[19, 186].

The RecBCD complex binds to the DNA duplex end, which is unwound by 4 bp in the crystal structure. The substrate makes extensive contacts with the RecB and RecC subunits. The 'arm' structure of RecB interacts with the duplex DNA region (Figure 19A), while the "pin" structure of RecC seems to act as a wedge that splits the duplex at the junction between the ssDNA and dsDNA (Figure 19C). The single-stranded tails are inserted into different tunnels running through the complex: the 3'-tunnel goes along RecB and RecC, while the 5'-tunnel goes along the RecD motor subunits. RecD does not contact the DNA in the crystal structure, but the 5'-tail should reach its motor subunits as the enzyme translocates (Figure 19D). Both the tunnels emerge adjacent to the nuclease domain of RecB. The nuclease active site is located so that the 3'-tail can be processively hydrolysed as it emerges from the RecC subunit, while the 5'-tail would be positioned less favourably for digestion[19, 186].

The RecBCD complex unwinds dsDNA very rapidly (1,000 to 2,000 bp/s) and processively (~30,000 bp)[186]. Electron microscopy revealed that the two motors in the bipolar RecBCD helicase move at different speeds: RecD is a fast helicase on the 5'-strand and RecB is a slow helicase on the 3'-strand[195]. As the fast motor RecD acts on the 5'-strand, it forms a long ssDNA tail, while the 3'-strand forms a ssDNA loop before the slow RecB motor (Figure 19D). Hence, only the lead motor RecD is a true helicase that unwinds DNA duplex, whereas the slow motor RecB simply translocates ssDNA[195].

The part of the 3'-tunnel in the RecC subunit seems to be a Chi scanning site (Figure 19D)[19]. The Chi sequence is recognized as ssDNA arising from dsDNA unwound by the enzyme. Indeed, the RecBCD enzyme requires only

the sequence information in the 5-GCTGGTGG-containing strand to be regulated by Chi[196]. It was suggested that the Chi sequence remains tightly bound in the 3'-tunnel. This prevents the further translocation of the 3'-tail into the nuclease domain (Figure 19D). Therefore, the lack of competition could upregulate the digestion of the 5'-tail[19, 186].

The interaction with the Chi sequence affects not only nuclease activity, but also DNA translocation of the RecBCD complex. Single-molecule experiments revealed that it pauses precisely at Chi and then continues translocating, but at approximately one-half of the initial rate[197]. This velocity change is related to a switch of motor usage at Chi: initially, RecD is the lead motor in the RecBCD complex, but after Chi recognition the slower RecB motor leads[198]. Although the speed of RecD is attenuated, it is not ejected from the holoenzyme at Chi[199]. As RecB continues translocating the 3'-terminating strand with the Chi sequence bound to the RecC subunit, a ssDNA loop is formed between the Chi recognition site and the translocating RecB motor (Figure 19D)[186].

Upon Chi recognition, RecBCD also facilitates loading of the RecA protein on the 3'-terminating strand (Figure 19D). This results in the formation of the RecA nucleoprotein filament on the Chi-containing ssDNA, which subsequently invades homologous dsDNA[187]. Surprisingly, the interaction between the RecBCD complex and RecA seems to be mediated by the RecB nuclease domain[200, 201]. In addition to nuclease function, this PD-(D/E)XK fold domain was shown to form stable complexes with RecA. The putative RecB-RecA interaction interface is blocked by the RecC subunit in the crystal structure of the RecBCD complex. It was proposed that the conformational change triggered by the Chi recognition also includes repositioning of the nuclease domain[19, 201].

2 Materials and Methods

2.1 Plasmids and bacterial strains

The pAL-BpuJI and pACYC-MBpuJI plasmids were obtained from dr. K.Stankevičius (Institute of Biotechnology). The pACYC-MBpuJI is a pACYC184 derivative with the inserted *bpuJIM1* and *bpuJIM2* genes; the pAL-BpuJI is a pAL4A derivative with the inserted *bpuJIR* gene. The following *E.coli* strains were used in the study:

ER2267 F^- e14$^-$(MrcA$^-$) *endA1 supE44 thi-1 relA1? RfbD1? SpoT1 Δ(mcrC-mrr)114::IS10 Δ(argF-lac)U169 recA1/F'proA*$^+$*B*$^+$ *lacI*^{q}*Δ(lacZ)M15 zzf::*mini-Tn10 (Kan^r).

ER2566 F^- λ^- *fhuA2 [lon] ompT lacZ ::T7 genel gal sulA11 D(mcrC-mrr)114::IS10 R(mcr 73::*mini Tn*10)2 R(zgb-210::Tn10)1* (Tet^s) *endA1 [dcm]*.

Bacteria were grown according to standart protocols, transformation was carried out by the $CaCl_2$ method[202].

2.2 Oligonucleotides and radiolabelling

All the oligonucleotides used in the study were synthesized at Metabion. Oligoduplexes (Table 3) were obtained by slow annealing. Where indicated, oligonucleotides were 5'-labelled with [γ-^{33}P] ATP and T4 polynucleotide kinase or 3'-labelled with [α-^{32}P] dideoxy ATP and terminal polynucleotidyl transferase (both the enzymes from Fermentas), following the manufacturer's instructions. Radiolabelled DNA was detected and quantified using a Cyclone Storage Phosphor System (PerkinElmer) with the OptiQuant program.

Table 3. Oligoduplexes used in the study

Abbreviation	Sequence
12/12(SP)	5' GGTA**CCCGT**GGA* 3' CCAT**GGGCA**CCT
16/16(SP)	5' TCGGTA**CCCGT**GGATC 3' AGCCAT**GGGCA**CCTAG
26/26(SP)	5' GAGCTCGGTA**CCCGT**GGATCCTCTAG 3' TCGAGCCAT**GGGCA**CCTAGGAGATCT
16/16	5' AGCGTAGCACTGGGCT 3' TCGCATCGTGACCCGA
37/37	5' GTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGT 3' ACTTAAGCTCGAGCCATGGGCCCCTAGGAGATCTCAT
37/25	5' GTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGT 3' ACTTAAGCTCGAGCCATGGGCCCCT
25/37	5' GTGAATTCGAGCTCGGTACCCGGGG 3' ACTTAAGCTCGAGCCATGGGCCCCTAGGAGATCTCAT
25/25	5' GTGAATTCGAGCTCGGTACCCGGGG 3' ACTTAAGCTCGAGCCATGGGCCCCT

*The BpuJI recognition sequence is shown in bold; SP – specific.

2.3 Chromatography

Heparin Sepharose, Blue Sepharose, AH Sepharose media and pre-packed HiTrap Heparin, HiTrap Chelating, Superdex 75 HR, Superdex 200 HR columns were from GE Healthcare; a TSK-GEL HPLC column SuperSW 2000 was from Tosoh Bioscience. The columns were prepared and run according to the manufacturer's instructions.

2.4 Expression and purification of BpuJI

BpuJI REase was purified from ER2267 cells, bearing the pACYC-MBpuJI and pAL-BpuJI plasmids (see 2.1). The bacteria were grown to the mid-log phase (A_{600}~0.5-0.7) at 30°C in the LB medium[202] supplemented with kanamycin (25 μg/ml), chloramphenicol (20 μg/ml) and carbenicillin (100 μg/ml). Protein expression was induced by heating at 42°C for 2 min. The bacteria were further cultivated at 37°C for 4 hours.

The protein purification was carried out at +4°C. The cells harvested by

centrifugation, were resuspended in a chromatography buffer A [10 mM potassium phosphate (pH 7.0), 0.2 M KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 5% glycerol, 0.025% Triton X-100] and sonicated. After removing of the insoluble material by centrifugation, the crude extract was applied onto a Heparin Sepharose column and eluted using a KCl gradient. The fractions containing active endonuclease were pooled and dialysed against the chromatography buffer A. Further protein purification was achieved by subsequent chromatography on Blue Sepharose and AH Sepharose columns.

The fractions containing purified BpuJI were pooled and dialysed against a storage buffer [10 mM Tris–HCl (pH 7.5 at 25°C), 0.2 M KCl, 0.1 mM EDTA, 1 mM DTT, 0.025% Triton X-100, 50% (v/v) glycerol] and kept at -20°C. Protein concentration was determined from A_{280} readings and expressed in terms of a monomer.

2.5 Limited proteolysis

Optimal reaction conditions for the proteolysis of BpuJI by chymotrypsin and thermolysin (both the proteases were from SIGMA) were found experimentally. The samples were analysed by SDS-PAGE according to Laemmli protocol with Coomasie Brilliant Blue staining[203]. For N-terminal sequencing, the proteolytic fragments were transfered onto a polyvinylidene fluoride microporous membrane (Immobilon-P, Millipore)[203] in the CAPS buffer [10 mM 3-[cyclohexylamino]-1-propane sulfonic acid (pH 11)]. The N-terminal sequencing was performed in the University of Bristol Proteomics Facility using an Applied Biosystems 477A sequencer. The molecular masses of proteolytic fragments were determined by electrospray ionization mass spectrometry.

2.5.1 Preparation of the N- and C-terminal proteolytic fragments

The chymotrypsin digestions were carried out in a proteolysis buffer A [10 mM potassium phosphate (pH 7.4), 100 mM KCl] for 1 h and terminated by adding phenylmethyl-sulphonyl fluoride to a final concentration of 1 mM. The concentration of BpuJI in reaction mixtures was 0.4 mg/ml, the mass ratio of chymotrypsine to BpuJI was 1:800 and 1:16 to obtain the N-terminal and C-terminal fragments, respectively.

The N-terminal fragment was purified from proteolysis the mixture by gel-filtration on a Superdex 75 HR column, equilibrated by the proteolysis buffer A. To isolate the C-terminal fragment, the proteolysis mixture was loaded onto a HiTrap heparin column and eluted using a KCl gradient in the chromatography buffer A (see 2.4). The pooled fractions containing BpuJI proteolytic fragments were dialysed against the storage buffer (see 2.4). The protein concentration was determined by Bradford assay with NanoQuant kit (Roth) and expressed in terms of a monomer.

2.5.2 Preparation of the N-domain/DNA complex

BpuJI was mixed with the cognate oligoduplex 16/16(SP) or 12/12(SP) (Table 3) to a final protein concentration of 0.3 mg/ml and 1.1-fold molar ratio of the DNA to the protein monomer in a proteolysis buffer B [10 mM Tris–HCl (pH 7.5 at 25°C), 100 mM KCl and 2 mM $CaCl_2$]. The protein/DNA complex was digested with thermolysin at a 1:10 mass ratio of the protease to BpuJI for 1 h. The reaction was quenched by adding EDTA to a final concentration of 10 mM.

The resulting proteolysis mixture was loaded onto a HiTrap heparin column, pre-equilibrated with a chromatography buffer B [50 mM Tris–HCl (pH 7.5 at 25°C), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.02% NaN_3]. The flow-through was concentrated by ultrafiltration using centrifugal filter

units with a molecular-weight cutoff of 5000 Da (Millipore) and loaded onto a Superdex 200 HR column. The pooled fractions containing the N-terminal domain/DNA complex were concentrated and immediately used for crystallization.

2.6 Bioinformatics analysis of the BpuJI sequence

Theoretical pI, molecular weight and extinction coeficient for BpuJI and its domains were determined with the ProtParam tool (http://www.expasy.org/tools/protparam.html).

Fold recognition (an attempt to match sequence with known structures) was performed using the HHpred server (http://protevo.eb.tuebingen.mpg.de/hhpred)[204], which is based on a comparison of profile Hidden Markov Models (HMMs)[205]. Detection of sequence homologs of the BpuJI C-domain was performed with saturated BLAST[206], which is based on multiple intermediate (transitive) sequence searches. Each search consisted of PSI-BLAST[207] run of up to eight iterations using low complexity filtering and a conservative E-value cutoff (10^{-4}) for sequence inclusion into the profile. Multiple sequence alignment was constructed using combination of PCMA[208] and ProbCons[209] followed by some manual adjustments to improve the hydrophobic packing of the core structure of the PD-(D/E)XK fold. The identification of additional conserved domains in the BpuJI C-domain homologs was done using reverse position specific BLAST (RPS-BLAST) searches against the Conserved Domain Database[210].

2.7 Crystallization and diffraction data collection

Samples for crystallization were concentrated by ultrafiltration using centrifugal filter units (Millipore) with a molecular-weight cut-off of 10 kD and 5 kDa for BpuJI and proteolytic fragments, respectively. The crystallization experiments were carried out at +19°C by the sitting drop

vapour diffusion technique[211]. 1 µl of a sample was mixed with 1 µl of a reservoir solution, containing crystallizing agent, and equilibrated against 500 µl of the reservoir solution in the Cryschem crystallization plates (Hampton Research). Initial trials were set up using screens from Hampton Research, further optimization used home-made solutions.

Diffraction data were collected at EMBL/DESY, Hamburg and processed using the MOSFLM[212], SCALA[213] and TRUNCATE[214] programs.

2.7.1 Crystallization of BpuJI

Prior to crystallization, small aliquots of BpuJI were dialysed against a crystallization buffer A [10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.02% NaN_3] and concentrated to 2-5 mg/ml of the protein. Thin needles were obtained under various conditions in presence of PEG. Further optimization resulted in crystals, grown in drops containing 100 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 9% PEG8000 and 10 mM $BaCl_2$.The crystals were polycrystalline and diffracted X-rays only to ~7 Å.

For crystallization of the BpuJI/DNA complex, the 16/16(SP) oligoduplex (Table 3) was added to concentrated BpuJI at 1.2:1 molar ratio of the DNA to the protein monomer. No crystals of the protein/DNA complex were obtained.

2.7.2 Crystallization of the BpuJI N-domain/DNA complex

The N-domain/DNA complex was purified from the proteolysis mixture (see 2.6) and concentrated to 6 mg/ml of protein. The buffer was changed into a crystallization buffer B [10 mM Tris–HCl (pH 7.5 at 25 °C), 0.1 mM EDTA, 1 mM DTT, 0.02% NaN_3] for crystallization trials. The first crystals grew in drops consisting of 1 µl of the BpuJI N-domain in complex with the 16 bp cognate DNA 16/16(SP) (Table 3) and 1 µl of a reservoir solution containing 100 mM sodium citrate (pH 4.8), 20% PEG6000. The crystals diffracted X-rays to ~2 Å; however, all the crystals tested were polycrystalline and

unsuitable for data collection. All attemps to get single crystals under similar conditions using various additives failed. The problem was solved by using the 12 bp oligoduplex 12/12(SP) (Table 3) instead of the 16/16(SP). The best crystals of the BpuJI N-domain complex with the 12/12(SP) DNA, obtained in the presence of 0.2 M ammonium tartrate and 20% PEG3350, were single crystals diffracting X-rays to 1.3 Å.

2.7.3 Crystallization of the BpuJI C-domain

The BpuJI C-domain obtained by limited proteolysis (see 2.6) was concentrated to 1-2.5 mg/ml of the protein, changing the buffer into a crystallization buffer C [10 mM Tris–HCl (pH 8.0 at 25°C), 200 mM KCl, 0.1 mM EDTA and 0.02% NaN_3]. The BpuJI C-domain gave nice-looking crystals in drops containing 0.4 M potassium sodium tartrate or 0.1 M MES (pH=6.5), 1.2 M sodium acetate, however the diffraction limit of these crystals was ~10 Å. Thin plates, obtained in the presence of 0.1 M sodium citrate (pH 3.8) and 5% PEG6000, diffracted X-rays better (~3.0Å) and were suitable for data collection.

2.7.4 Data collection

The crystals of the BpuJI N-domain/DNA complex were cryoprotected by soaking for ~30 min in a cryoprotecting buffer A [0.2 M potassium sodium tartrate, 0.2 M NH_4Cl, 20% PEG4000 and 25% PEG400]. To obtain a mercury derivative for phasing, the crystals were soaked in a mercury solution [2 mM $HgCl_2$, 0.2 M ammonium tartrate, 20% PEG3350, 15% PEG400] for about one week. Complete diffraction datasets for the native N-domain/DNA complex and the Hg derivative were collected at the X12 beamline with excellent statistics (Table 4).

The crystals of the BpuJI C-domain were transferred into a cryoprotecting buffer B [0.1 M sodium citrate (pH 3.8), 10% PEG6000, 25% PEG400] for ~1 h prior to flash cryo-cooling. A 3.0 Å dataset was collected at the X11

beamline; however, data processing failed, probably due to poor data quality and/or crystal twinning.

Table 4. Data collection statistics

Dataset	Native BpuJI N-domain/DNA complex	Hg-derivative of BpuJI N-domain/DNA complex
Temperature	100 K	
Wavelength	1.000 Å	
Spacegroup	$P2_12_12$	
Cell constants (Å)	56.2; 167.2; 43.8	59.0; 168.5; 43.8
Resolution range (Å)	83.6-1.30	84.2-1.80
Completeness (%)	98.2 (96.1)[a]	99.9 (99.9)
Multiplicity	10.8 (8.0)	15.5 (15.3)
I/σ_I	4.7 (2.4)	3.7 (2.6)
R_{merge}[b]	0.082 (0.31)	0.11 (0.24)

[a] Values in parentheses refer to data in the highest resolution shell.

[b] $R_{merge}=\sum|I_{hi}-\langle I_h\rangle|/\sum I_{hi}$, where I_{hi} is an intensity value of the i-th measurement of reflection h and $\langle I_h\rangle$ is the average measured intensity of the reflection h.

2.8 Structure determination and analysis

The structure of the BpuJI N-domain/DNA complex was solved by the single-wavelength anomalous dispersion (SAD) method. Mercury positions were identified on the anomalous Patterson map Harker sections using the HARA program (S. Gražulis, unpublished data). The mercury peaks were brought to the same hand and origin by anomalous Fourier synthesis. The atom positions were refined, and phases were calculated in the MLPHARE program[215]. The resulting maps were solvent flattened with DM[216]. Map connectivity was used to identify the correct hand and resulted in a DNA helix with the correct handedness.

Initial model building was performed with the ARP/wARP program[217]. Manual rebuilding with Coot[218] alternated with refinement using CNS[219]. The partially refined structure of the Hg derivative was used for molecular replacement with AMORE[215] to solve the native structure. Final refinement steps were carried out with REFMAC[220]. The quality of the model was verified

with the PROCHECK[221] and WHAT IF[222] programs. Statistics concerning refinement and the quality of the final model are presented in Table 5.

Table 5. Refinement statistics

Resolution range used for refinement (Å)	83.6–1.3
No. of reflections	91,374
R-factor	0.138
Free R-factor (10% set)	0.166
No. of non-hydrogen atoms:	
Protein/DNA	2486/486
Solvent	461
Ligand/Ion	40/1
Average B-factors (Å^2)	13.9
r.m.s. deviations from ideal values:	
Bonds (Å)	0.014
Angles (°)	1.1
Ramachandran analysis:	
Most favored (%)	93.0
Allowed (%)	7.0

Secondary structure was assigned with STRIDE[223]; structure comparisons used the DALI server[224] and the TOP3D program[225]; solvent accessible surface areas were calculated with the program NACCESS[226]; and protein–DNA contacts were analyzed with NUCPLOT[227]. Figures were prepared with MOLSCRIPT[228] and Raster3D[229].

The coordinates and structure factors were deposited in the Protein Data Bank under accession code 2VLA.

2.9 Cloning and mutagenesis

PCR, restriction digestions, dephosphorylation of vector DNA by the calf intestine alkaline phosphatase and DNA ligation used enzymes from Fermentas, following the manufacturer's instructions. DNA fragments were separated by electrophoresis in agarose gels[203] in the TAE buffer [40 mM Tris-acetate, 20 mM acetic acid, 2 mM EDTA, pH~8.5] and extracted from gels using the PerfectPrep® gel cleanup kit (Eppendorf).

Plasmid DNA was isolated from the ER2267 cells (see 2.1) by the alkaline

lysis procedure followed by phenol/chloroform extraction[202], or using the GeneJET™ plasmid miniprep kit (Fermentas). The target gene regions in all constructs were verified by seqeuncing in the Sequencing Centre (Institute of Biotechnology).

2.9.1 Site-specific mutagenesis of the active site residues

Site-directed mutagenesis of the active site residues was performed by the megaprimer PCR method[230], using the pAl BpuJI as a template. A megaprimer, obtained with a mutagenic primer and the primer I, was used for a second PCR together with the primer II (Table 6). The resulting PCR products were ClaI/Bsp1407I digested and ligated into the pAl BpuJI cleaved by the same enzymes. The pAl BpuJI has two ClaI sites, therefore the vector was prepared by partial restriction.

2.9.2 Cloning of the BpuJI N domain

The region coding for the BpuJI N-domain (residues 1–285) was PCR-amplified from the pAL-BpuJI, introducing a silent mutation into the NdeI site present in the BpuJI gene by the megaprimer PCR method[230] (see Table 6 for the primers). The product of the second PCR was inserted into the pET21b (Novagen) between the NdeI and XhoI restriction sites. The resulting construct pET-BpuJINH overexpresses the BpuJI N-domain with a hexahistidine tag at the C-terminus.

2.9.3 Mutagenesis of the specifity-determining residues

Site-directed mutagenesis of the specifity-determining regions was carried out by the one-step method[231]. The pET-BpuJINH was amplified by PCR using the *Pfu* DNA polymerase (Fermentas) and two partially overlapping primers containing a desired mutation (Table 6). The pAl-BpuJI was used as a template to introduce the K121A mutation into the full-length BpuJI gene. Following PCR, the products were treated with DpnI to digest the parental template.

Table 6. Oligonucleotide primers for cloning and mutagenesis

No.	Sequence	Comment
I.	5'-atcattgattacatacatag	PCR I primer for the active site mutagenesis
II.	5'-gcttacaagctcgcagtaat	PCR II primer for the active site mutagenesis
III.	5'-agggtatgctattcaatcag	mutagenic primer, D348A
IV.	5'-agggtatgctattcaatcag	mutagenic primer, E367A
V.	5'-acatagcagtaaaatcaacc	mutagenic primer, K369A
VI.	5'-atatcatatgaactataatccagaggaac	PCR I primer for N-domain cloning, +NdeI
VII.	5'-gtttgcataagcgggtaaca	mutagenic primer, -NdeI
VIII.	5'-ttctcgagggatttgtgtgaatttacatttg	PCR II primer for N-domain cloning, +XhoI
IX.	5'-caaaagcaactctagacaatcaccgaactg	forward mutagenic primer, K63A
X.	5'-gattgtctagagttgcttttgtagtttcctc	reverse mutagenic primer, K63A
XI.	5'-ctctagacgctcaccgaactgaaattgctg	forward mutagenic primer, N67A
XII.	5'-cagttcggtgagcgtctagagttttttttg	reverse mutagenic primer, N67A
XIII.	5'-caatcaccgaactgcaattgctggtaaattatttg	forward mutagenic primer, E71A
XIV.	5'-gcaattgcagttcggtgattgtctagagttttttttg	reverse mutagenic primer, E71A
XV.	5'-ctattattgccgggaaagccaaaaatatgc	forward mutagenic primer, R15A
XVI.	5'-ggctttcccggcaataatagtacatctg	reverse mutagenic primer, R15A
XVII.	5'-gaggcgccgccaaaaatatgcttgataac	forward mutagenic primer, K17A
XVIII.	5'-gcatatttttggcggcgcctctaataatag	reverse mutagenic primer, K17A
XIX.	5'-catcccgggaaagagacagcctactcaatgcaac	forward mutagenic primer, S204A
XX.	5'-gtaggctgtctctttcccgggatgtctaacttttttttg	reverse mutagenic primer, S204A
XXI.	5'-ctactccatggcacatattagagagcaattg	forward mutagenic primer, Q208A
XXII.	5'-ctaatatgtgccatggagtaggatgtctc	reverse mutagenic primer, Q208A
XXIII.	5'-ggatgctctagataaggttattgaaaaagttg	forward mutagenic primer, K121A
XXIV.	5'-caataaccttatctagagcatccattccattc	reverse mutagenic primer, K121A

2.9.4 Expression and purification of mutant proteins

The C-$(His)_6$-tagged N-domain and its mutants were expressed in *E.coli* ER2566 cells, bearing the pET-BpuJINH and pACYC-MBpuJI (see 2.1). The cells were grown to the mid-log phase (A_{600}~0.5-0.7) at 37°C in the LB medium[202] supplemented with chloramphenicol (20 μg/ml) and carbenicillin (100 μg/ml). The protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After the induction, the bacteria were further incubated at 37°C for 4 hours and harvested by centrifugation.

The cells were resuspended in a chromatography buffer C [20 mM sodium

phosphate (pH 7.4), 0.5 M NaCl] and sonicated. The crude extract obtained by centrifugation was loaded onto a Hi Trap chelating column and eluted with an imidazol gradient. The pooled fractions containing the target protein were dialyzed against the chromatography buffer C to get rid of imidazol, and afterwards, against the storage buffer (see 2.4). The protein concentration was determined from A_{280} readings and expressed in terms of monomer.

The active site mutants and the full-length K121A mutant were expressed and purified as the wt protein (see 2.4).

2.10 DNA cleavage assays

All DNA cleavage reactions by BpuJI, its domains and mutants were performed at 25°C in a reaction buffer [50 mM Tris–HCl (pH 7.5 at 25°C), 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml of BSA].

2.10.1 Plasmid and phage DNA cleavage assay

Thr phage λ, phage ΦX174 and pUC57 DNA were purchased from Fermentas. Reaction mixes contained 20 μg/ml of the DNA and 25-1000 nM of the enzyme. Reactions were quenched with a loading dye A [75 mM EDTA (pH 9.0), 0.01% bromphenol blue, 50% (v/v) glycerol, 0.1% SDS]. The samples were heated at 65°C for 20 min and analysed by electrophoresis through agarose[203] in an electrophoresis buffer A [0.1 M sodium borate (pH 8.2), 15 mM acetic acid, 2 mM EDTA, 0.5 μg/ml ethidium bromide].

To determine cleavage positions, the reactions were quenched with phenol/ chloroform, DNA was precipitated from the aqueous phase and used as a template for run-off sequencing in the Sequencing Centre (Institute of Biotechnology). The primers used for the run-off sequencing are presented in Table 7.

Table 7. Oligonucleotide primers used to otain PCR fragments and for run-off sequencing

No.	Sequence	Comment
I.	5'-gccagggttttcccagtcacga	direct primer located upstream of the multiple cloning site of pUC plasmids
II.	5'-gcttccggctcgtatgttgtg	reverse primer located dowstream of the multiple cloning site of pUC plasmids
III.	5'-gttttcgccccgaagaacg	direct primer located upstream of the BpuJI target in the pUC57
IV.	5'-agatgcttttctgtgactggtg	reverse primer located downstream of the BpuJI target in the pUC57
V.	5'-ctgaccgctctcgtgc	direct primer located upstream of the BpuJI target in the ΦX174 DNA
VI.	5'-gctttgagtcttcttcgg	reverse primer located downstream of the BpuJI target in the phage ΦX174 DNA

2.10.2 PCR fragment cleavage assay

The specific (203 bp) and non-specific (179 bp) fragments were PCR-amplified from the pUC57 or pUC19 DNA, respectively, using the primers I and II (Table 7) and the *Pfu* DNA polymerase. The primers III and IV were used to synthesize a specific PCR fragment with different flanking sequences from the pUC57 template. Before the PCR reactions one of the primers was 5'-radiolabelled (see 2.2). The fragments were purified on S-400 microspin columns (GE Healthcare).

Reaction mixes contained 1-2 nM of the DNA fragment and 0.2-3 μM of the enzyme. In oligonucleotide activation experiments, the equimolar amount of the specific 16/16(SP) oligoduplex (Table 3) was added into the reaction mixtures. Aliquots were removed at timed intervals and mixed with phenol/chloroform; DNA was precipitated from the aqueous phase. To map cleavage sites on the complementary strand, DNA samples were split into two parts, and one part was treated with the T4 DNA polymerase (Fermentas).

The samples were mixed with a loading dye B [95% (v/v) formamide, 20 mM EDTA (pH 8.0), 0.01% bromophenol blue], heated at 95ºC for 5 min

and analyzed in the standard sequencing gels[203]. The products of dideoxy sequencing reactions obtained with the CycleReader™ DNA sequencing kit (Fermentas) were used as size markers.

2.10.3 Oligoduplex cleavage assay

Radiolabelled oligoduplexes were prepared as discribed above (see 2.2). The radiolabelled DNA (2 nM) was incubated with the BpuJI C-domain at a final concentration of 100 nM. Aliquots were removed at timed intervals, mixed with the loading dye B and heated at 95ºC for 5 min. Products were separated in the denaturing polyacrylamide/urea/TBE gels[203]. The 37/37 oligoduplex (Table 3) cleaved by XbaI, Acc65I, EcoRI, BamHI, Ecl136II and S1 nuclease (Fermentas) gave size markers.

2.11 Gel mobility shift assay for DNA binding

The ^{33}P-labelled oligoduplexes (see 2.2) at a final concentration of 0.1-2 nM were incubated with 1-2000 nM of protein in a binding buffer [30 mM MES (pH 6.5), 30 mM histidine, 10% (v/v) glycerol, 0.2 mg/ml of BSA]. To analyze synaptic complexes or to determine K_d of low-affinity mutants, 10-100 nM of the unlabelled DNA was added to the binding mixes. The samples were analyzed by native electrophoresis through 8% polyacrylamide gels in an electrophoresis buffer B [30 mM MES (pH 6.5), 30 mM histidine] for 2-4 hours at 6 V/cm.

K_d values were calculated by fitting data to the equation:

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

where y is a concentration of protein/DNA complex (in terms of nM) at a total protein concentration x, s_0 is a total DNA concentration in the binding mixture, and K_d is a dissociation constant. Data were analyzed using the

KyPlot program[232].

2.12 Analytical gel-filtration

Gel-filtration of BpuJI and its proteolytic fragments was carried out on a TSK-GEL HPLC column SuperSW 2000 in a gel-filtration buffer A [0.1 M sodium phosphate (pH 6.7), 0.1 M sodium sulphate]. The samples contained 1.5 μM of the protein. Elution from the column was monitored by measuring absorbance at 215 nm.

Gel filtration of protein/DNA complexes was carried out on a Superdex 200 HR column in a gel-filtration buffer B [50 mM Tris-HCl (pH 7.5 at 25°C), 0.1 M KCl, 2 mM $CaCl_2$]. The samples contained 3 μM of the protein and 3 μM of the 16/16(SP) (Table 3); the control samples were prepared without the DNA or protein. Elution from the column was monitored by measuring absorbance at 280 nm.

The calibration curves were generated by measuring the elution volumes of a series of standard proteins (Bio-Rad).

2.13 Analytical ultracentrifugation

Sedimentation equilibrium experiments were performed with a Beckman/Coulter XLA analytical ultracentrifuge, in six-channel centrepieces allowing simultaneous investigation of 24 samples in a single run. Runs were done at 20°C and 12000 rpm, absorbance readings at 280 nm were taken in 1 hour intervals. Samples at initial loading concentrations 1.6 μM, 3.2 μM and 6.4 μM of BpuJI in an ultracentrifugation buffer [10 mM Tris–HCl (pH 7.5 at 25°C), 0.2 M KCl, 0.1 mM EDTA, 0.1 mM DTT] were spun for at least 18 hours. It was assumed that equilibrium had been reached when the measured concentration profiles did not change measurably over the following 12 hours. For signal smoothing, all profiles measured during this 12 h period at

equilibrium were averaged. Apparent molar masses were calculated from the averaged profiles as described earlier[233], using a partial specific volume of $7.386 . 10^{-4}$ m^3kg^{-1}. The latter was calculated from the amino acid composition using the values tabulated in SEDNTERP (http://www.rasmb.bbri.org).

3 Results and discussion

3.1 Functional characterization of BpuJI

Restriction endonuclease BpuJI from the *Bacillus pumilus* RFL1458 strain is specific for the asymmetric sequence 5'-CCCGT (K. Stankevičius, unpublished data). The BpuJI R-M system contains a single gene for restriction enzyme *bpuJIR* (GenBank:EF409421), flanked by two genes encoding 5-methylcytosine methyltransferases, *bpuJIM1* and *bpuJIM2* (Figure 20A) (K. Stankevičius, unpublished data). Such gene organization is similar to that of Type IIS R-M systems, which typically have one REase gene and two MTase genes to modify both strands of the asymmetric recognition sequence[4].

Like most of Type II restriction enzymes, BpuJI requires Mg^{2+} ions for DNA cleavage. The phage lambda DNA, which contains 103 5'-CCCGT sequences, gives a complex cleavage pattern upon BpuJI treatment. On the other hand, the pUC57 plasmid and phage ΦX174 DNA that have 6 and 2 recognition sites, respectively, give a distinct DNA fragmentation that corresponds to the BpuJI cleavage at 5'-CCCGT sequences (Figure 20B). To determine the BpuJI cleavage site the linear products obtained upon ΦX174

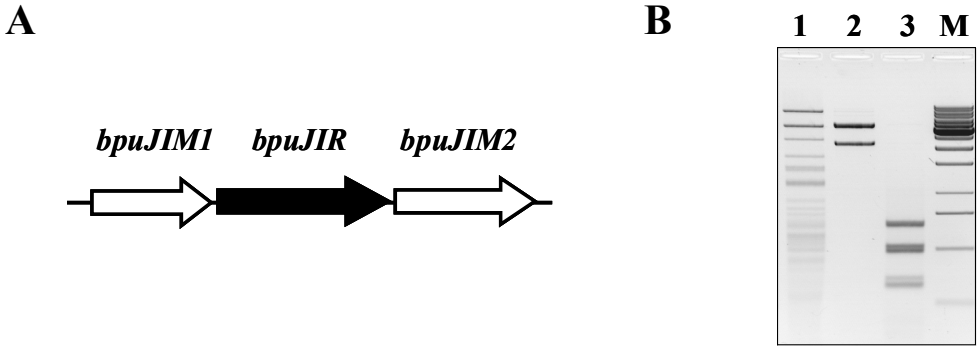

***Figure 20**. (A) The gene organization of the BpuJI R-M system. The bpuJIR gene coding for restriction endonuclease is flanked by the bpuJIMI and bpuJIM2 genes for 5-methylcytosine methyltransferases. (B) BpuJI digests of the phage λ (lane 1), phage ΦX174 (lane 2) and pUC57 (lane 3) DNA. 20 μg/ml of the DNA was incubated with 100 nM of BpuJI for 1 h at 37˚C, products were separated by electrophoresis through 1% agarose (see 2.10). Lane M – DNA size marker.*

DNA cleavage were subjected to run-off sequencing, however no clear termination was observed (data not shown).

3.1.1 Variability of the BpuJI cleavage position

To determine the cleavage position of BpuJI, we employed the 203 bp DNA fragment containing a single 5'-CCCGT sequence positioned approximately in the middle. The fragment was obtained by PCR using a pair of primers flanking the 5'-CCCGT sequence in the multiple cloning sites region of the pUC57. BpuJI cuts this fragment rather slowly (data not shown) therefore high concentrations of enzyme were used to achieve complete cleavage.

1 nM of the 203 bp PCR fragment, ^{33}P-labelled at the 5'-terminus either at the top or bottom strand, was incubated with 3 μM of BpuJI, aliquots were removed at different time intervals and sizes of the resulting fragments were analysed in the sequencing gels. We found that BpuJI reaction on the top DNA strand yields multiple cuts located at varying distances on the both sides of the recognition sequence (Figure 21A). Moreover, the short time (5 and 20 min) and the prolonged (2 and 5 h) reactions generate DNA fragments of different length. The short time incubation yields a mixture of fragments that correspond to the top strand cleavage from 2 to 25 nt downstream the recognition site. The major products (corresponding to the strongest bands in the gel) match to the BpuJI cut 2 nt and 11 nt downstream of the target site. The predominant product resulting from the prolonged incubations correspond to the BpuJI cut 18 nt upstream of the recognition site. Thus, BpuJI cleaves the top strand on both sides of the recognition site, but the first cuts are downstream of the target.

Mapping of the BpuJI breaks on the bottom strand also reveals cleavage at multiple sites (Figure 21B). The short time incubation (5 and 20 min) results in the bottom strand cleavage 6, 7, 13, 15, 16, 19 and 23 nt downstream of the

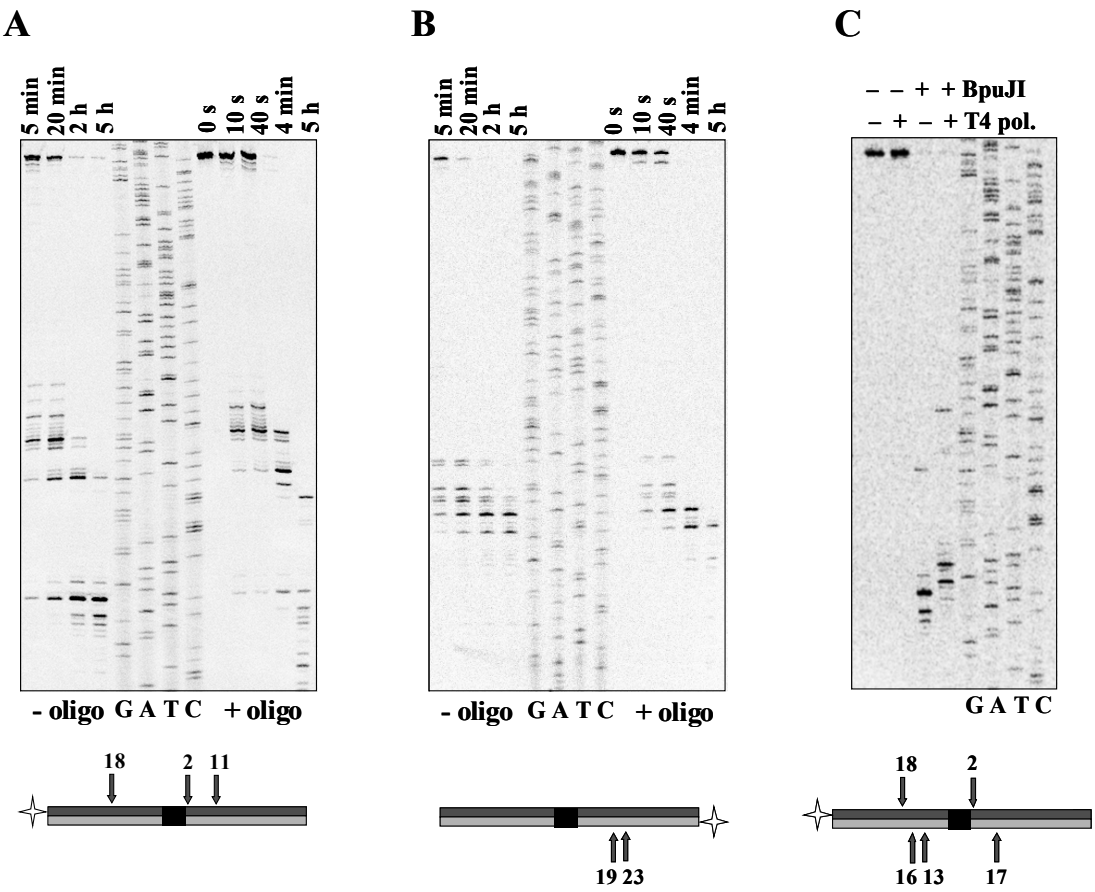

Figure 21. *The cleavage of the specific PCR fragment by BpuJI. 1 nM of the PCR fragment, 5'-labelled either in the top strand (A) or in the bottom strand (B), was incubated with 3 μM of BpuJI in the presence or absence of the 16/16(SP) oligoduplex. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through a standard sequencing gel. (C) The T4 DNA polymerase action on the BpuJI digests. 1 nM of the specific PCR fragment, 5'-labelled in the top strand, was incubated with 1.5 μM of BpuJI for 5 h. The reaction was quenched with phenol/chlorophorm. The DNA was precipitated, divided into two aliquots, and of the aliquots was treated with the T4 DNA polymerase. The products were analysed by electrophoresis through the standard sequencing gel. The lanes G, A, T, C are sequence ladders. In the cartoons below the gels, the BpuJI recognition sequence is indicated by a rectangle, ^{33}P is indicated by a star. The arrows show the major cleavage positions, the numbers above the arrows indicating the distance between the recognition sequence and the cleavage position.*

recognition site. However, the prolonged incubations generate a set of shorter products that correspond to the BpuJI cuts 19 and 23 nt downstream of the target.

The prolonged incubations reveal BpuJI induced breaks on the top strand

upstream of the recognition site. To find out if cleavage of the bottom strand occurs on both sides of the recognition sequence, we mapped the BpuJI breaks on the bottom strand, using a method based on the ability of the T4 DNA polymerase to process DNA ends. If a cleavage generates a 5'-overhang, the top strand will be extended by polymerase; if a cleavage generates a 3'-overhang, the top strand will be digested back; if cleavage generates a blunt end, no change in length will be observed. When the fragment, 5'-labelled in the top strand, was cleaved with BpuJI and afterwards treated with the T4 DNA polymerase, the fragment corresponding to bottom strand cleavage downstream of the recognition sequence was seen. Also, short fragments were observed, indicating that complementary strand is also cleaved 13 and 16 nt upstream the recognition site (Figure 21C).

A similar variable cleavage pattern was observed at another BpuJI site in the pUC57 flanked by different sequences. However, the distances of the major cuts in respect to the recognition sequence varied at different sites (data not shown).

3.1.2 BpuJI needs two recognition sites for its optimal activity

Many of Type IIS REases display their maximal catalytic activities when bound to two copies of their recognition sequences[234-236]. Cleavage of a single recognition site containing DNA by an enzyme interacting with two recognition sites often can be stimulated by the addition of an oligoduplex that carries its recognition sequence[236-238]. In these cases, the oligoduplex *in trans* provides a second site for the enzyme. To find out if BpuJI needs two sites for its optimal catalytic activity, we compared the BpuJI cleavage rates in presence or absence of a cognate oligoduplex. The addition of the 16/16(SP) oligoduplex, carrying the BpuJI recognition sequence, considerably accelerated the cleavage of the 203 bp PCR fragment. Indeed, the cleavage pattern on the

top and bottom strands of the 203 bp fragment, obtained in the presence of the cognate 16/16(SP) oligoduplex after the 40 s incubation, was very similar to the pattern obtained in the absence of the oligoduplex after 2 h incubation (Figure 21). The cognate oligoduplex stimulated the BpuJI cleavage of the 203 bp fragment, but it did not change the cleavage positions: the both strands were still cut at multiple sites. A non-specific oligoduplex had no effect on the cleavage rate (data not shown).

3.1.3 The 3'-end-directed nucleolytic activity

Alongside to the cleavage in the vicinity of the recognition site 5'-CCCGT, the BpuJI treatment of the 5'-labelled 203 bp PCR fragment in the presence of the cognate oligonucleotide yielded extra products. These extra products correspond to the cleavage a few nucleotides away from the 3'-end of the fragment (Figure 21). They also are formed upon BpuJI action on the top DNA strand in the absence of the activating oligonucleotide, albeit at a much slower rate.

To examine if BpuJI possesses a similar activity on non-cognate DNA, the radioactively labelled 179 bp PCR fragment lacking the recognition site was incubated with BpuJI in the presence or absence of the cognate 16/16(SP) oligoduplex. In the absence of the activating oligonucleotide, no cleavage of the top strand occurred after 5 and 20 min. However, the prolonged incubation yielded weak bands that correspond to the cut a few nucleotides away from the 3'-end of the fragment (Figure 22A). A similar cleavage pattern was observed on the bottom strand (Figure 22B). Analysis of the products, resulting from the BpuJI cleavage of nonspecific fragment, in a sequencing gel revealed that the top strand was cut 3 and 6 nt from the 3'-end, while the bottom strand was cut 4 nt from the 3'-end. This suggests that BpuJI possesses an end-directed nuclease activity. When the products of the non-specific nuclease activity were treated with the T4 DNA polymerase, the full-length strand was synthesized.

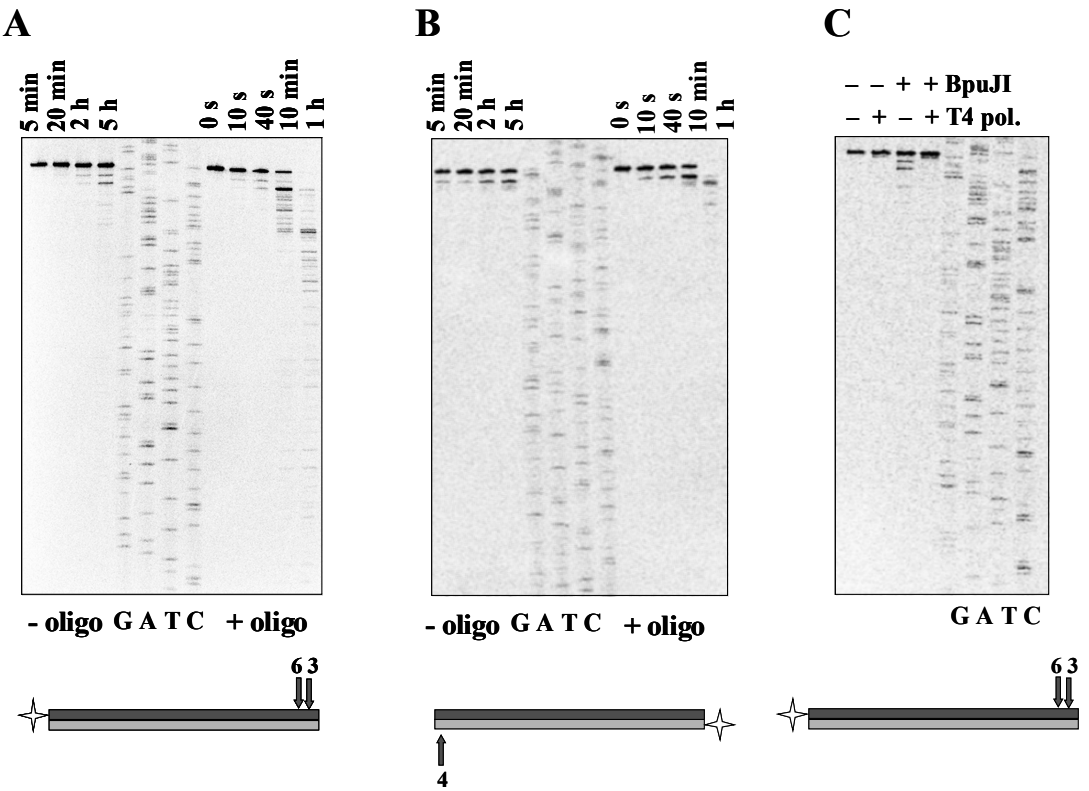

Figure 22. *The cleavage of the non-specific PCR fragment by BpuJI. 1 nM of the PCR fragment, 5'-labelled either in top strand (A) or in the bottom strand (B), was incubated with 3 μM of BpuJI in the presence or absence of the 16/16(SP) oligoduplex. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through a standard sequencing gel. (C) T4 polymerase action on BpuJI digest. 1 nM specific PCR fragment, 5'-labelled in top strand was incubated with 1.5 μM of BpuJI for 5 h, reaction was quenched with phenol/chlorophorm. DNA was precipitated, divided into two aliquots and of with T4 polymerase. Products were analysed by electrophoresis through the standard sequencing gel. The lanes G, A, T, C are sequence ladders. In the cartoons below the gels,* ^{33}P *is indicated by a star, the arrows show the positions of the initial cleavage, the numbers above the arrows indicating the distance from the 3'-terminus.*

This confirms that cleavage occurrs at the 3'-terminus of the radiolabelled strand, while the 5'-terminus of the complementary strand is not cleaved (Figure 22C). The addition of the cognate oligoduplex strongly stimulates the 3'-end-directed activity. Indeed, after 10 min incubation, most of the 179 bp PCR fragment was cut into shorter fragments, producing a DNA ladder (Figure 22). The cleavage of the bottom strand had a similar pattern. Hence, in addition to the sequence-directed nuclease activity, BpuJI also possesses the end-directed nuclease activity, which is activated by the recognition of the

target site. The two activities lead to the complicated pattern around the target site. Also, BpuJI shows the end-directed cleavage of DNA without recognition site, which is very pronounced in presence of cognate DNA in trans.

3.1.4 BpuJI binds two DNA copies simultaneously

DNA binding by BpuJI were studied by gel shift analysis using the ^{33}P-labelled 16/16(SP) and 16/16 oligoduplexes. To avoid cleavage, the binding experiments were performed in the absence of Mg^{2+} ions that are necessary for the catalysis. The BpuJI binding to the 16/16(SP), containing a single recognition site, yielded two different complexes, named 1 and 2 (Figure 23A). No complex was observed with the non-cognate 16/16 oligoduplex, confirming that the complexes 1 and 2 result due to the specific BpuJI binding. The titration of a fixed amount of the 16/16(SP) oligoduplex with increasing concentrations of BpuJI yielded: first, at low concentrations of the protein, the form with the higher mobility, the complex 1; then at increased protein concentrations, the less mobile species, the complex 2. The "oligonucleotide activation" of the BpuJI cleavage suggests that it can interact with two copies of the recognition sequence. Thus, the complexes 1 and 2 could represent the BpuJI protein bound to one and two DNA molecules, respectively, as was demonstrated previously for the REases SfiI and BfiI[236, 239].

To find out if the BpuJI protein can concurrently bind to two copies of specific DNA, oligoduplexes of different lengths were employed. BpuJI was added to the separate tubes containing mixtures of the oligoduplexes 16/16(SP) and 26/26(SP) at different ratios, but keeping the fixed 1:1 total DNA:protein monomer ratio. Traces of either the ^{33}P-labelled 16/16(SP) or the ^{33}P-labelled 26/26(SP) were added into the binding mixtures for detection purpose. Gel shift analysis revealed that BpuJI, when bound to either to the 16/16(SP) or 26/26(SP) DNA, forms a single complex, which moves in a gel with a characteristic electrophoretic mobility (Figure 23B). The incubation of BpuJI

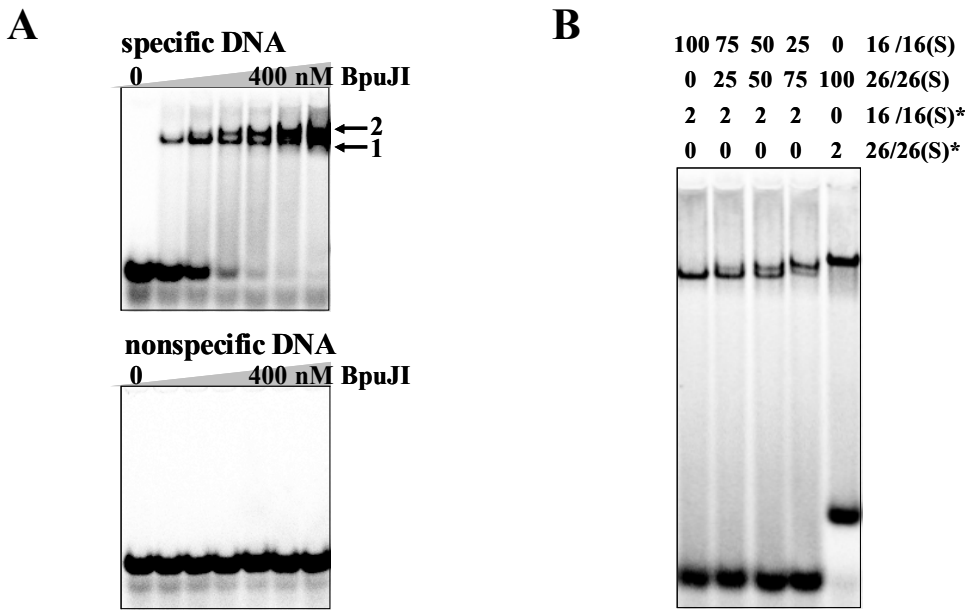

Figure 23. *DNA binding by the BpuJI REase. (A) Gel shift analysis of the BpuJI binding. The binding reactions contained BpuJI (0, 8, 16, 40, 80, 160 or 400 nM) and 2 nM of either the ^{33}P-labelled specific 16/16(SP) oligoduplex (upper gel) or the nonspecific 16/16 oligoduplex (lower gel). The samples were subjected to non-denaturating PAGE for 2 h (see 2.11). (B) Analysis of the BpuJI–DNA complex. The reactions contained 100 nM of BpuJI, the ^{33}P-labelled (the label is indicated by a star) and unlabelled specific oligoduplexes, 16/16(SP) and 26/26(SP). The concentrations of the oligoduplexes (in nM) are indicated above the relevant lanes. The samples were subjected to non-denaturating PAGE for 4 h (see 2.11).*

with the ^{33}P-labelled 16/16(SP) oligoduplex in the presence of the unlabelled cognate 26/26(SP) DNA resulted in a novel complex with an intermediate electrophoretic mobility. The yield of this complex varied depending on the ratio of the unlabelled 16/16(SP) and 26/26(SP). This intermediate complex must represent BpuJI bound to both the ^{33}P-labelled 16/16(SP) and the unlabelled 26/26(SP) DNA. Thus, it provides a direct evidence of BpuJI binding to two recognition sites at the same time.

3.2 Structural organization of BpuJI

3.2.1 BpuJI is a dimer in solution

Most of Type IIS REases are monomers in solution[86, 240, 241] and dimerize upon DNA binding[87, 234], except for BfiI, which is a dimer both in solution and in the protein-DNA complex[236]. The amino acid sequence of the BpuJI subunit predicts a molecular mass of 53.9 kDa. Analytical ultracentrifugation experiments were performed by prof. Claus Urbanke (Hannover Medizinische Hochshule). Sedimentation equilibrium at 12000 rpm for BpuJI yielded a molecular mass of 109 kDa (Figure 24), which corresponds to the BpuJI dimer. Thus, BpuJI is a dimer in solution.

3.2.2 The modular architecture of BpuJI

The archetypal Type IIS enzyme FokI is composed of separate domains for DNA recognition and catalysis[7]. BfiI also was shown to have a modular architecture[9]. To examine the possible domain organization of BpuJI, the protein was subjected to limited proteolysis. Two proteolytic fragments with

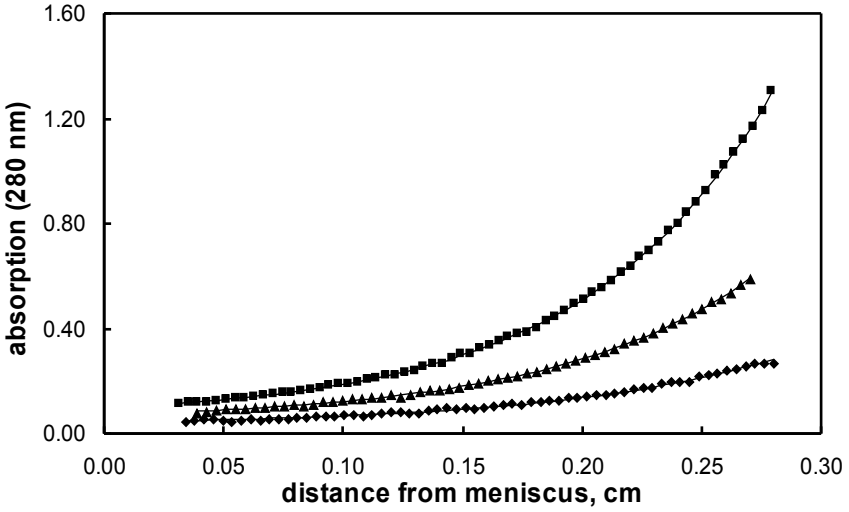

Figure 24. *Sedimentation equilibrium of BpuJI at 12000 rpm. The loading concentrations (monomer) and the meniscus positions were 1.6 µM and 5.84 cm (♦), 3.2 µM and 6.34 cm (▲), and 6.4 µM and 6.83 cm (■). The solid lines represent the concentration distributions calculated for all the three traces with a molar mass of 109 kg and a partial specific volume of 7.386 · 10^{-4} m^3kg^{-1}.*

apparent molecular masses of ~35 kDa and ~20 kDa (Figure 25A) were isolated from chymotrypsin digests, and their N-terminal amino acids were determined by microsequencing. The N-terminal amino acid sequence of ~35 kDa fragment matched exactly the sequence at the N-terminus of BpuJI, while the N-terminal sequence of the ~20 kDa fragment started from the T303 residue in the intact protein. The N-terminal proteolytic fragment was less protease-resistant than the C-terminal one; however, both the fragments resisted further proteolysis in the presence of the cognate 16/16(SP) oligoduplex (data not shown).

The apparent mass values for the N-terminal and the C-terminal proteolytic fragments, estimated from their elution volumes during size-exclusion chromatography (Figure 25B), were ~38 kDa and ~49 kDa, respectively. These values suggest that the N-terminal fragment is a monomer in solution, while the C-terminal fragment is a dimer. Hence, BpuJI appears to be organized into the N-terminal and C-terminal structural domains, the dimerization interface being formed by the C-domains.

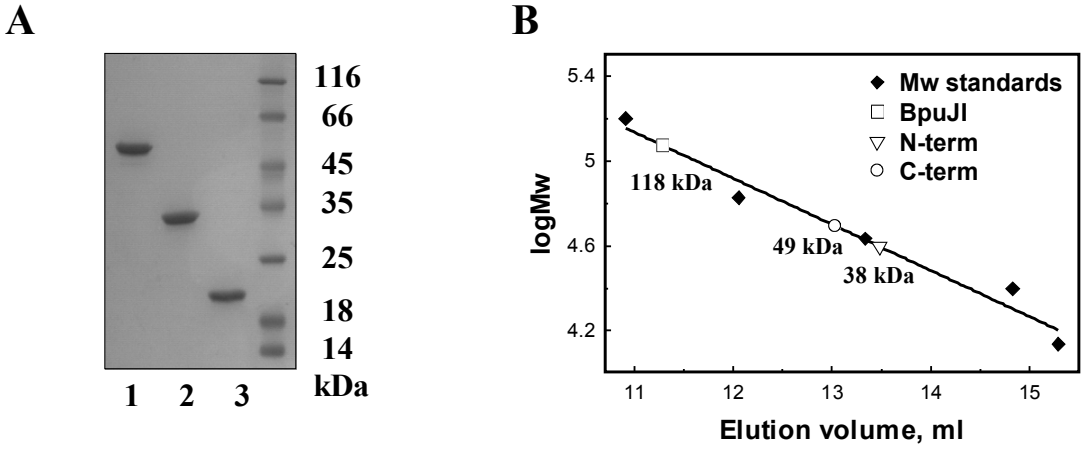

Figure 25. *(A) The BpuJI proteolytic fragments. 1 µg of BpuJI (lane 1), the N-terminal fragment (lane 2) and the C-terminal fragment (lane 3) were loaded on an SDS-PAGE gel. (B) Gel-filtration of BpuJI and its proteolytic fragments. The chromatography was performed on a SuperSW 2000 column (see 2.12). The apparent mass values for BpuJI (□), the N-terminal fragment (▽) and the C-terminal fragment (○) were calculated from a calibration curve (♦) obtained using a set of standart proteins.*

The isolated proteolytic fragments were subjected to DNA cleavage and binding assays to identify the function of the individual BpuJI domains. The DNA cleavage was examined on the phage λ DNA, which has multiple BpuJI recognition sites. In contrast to intact BpuJI, which cleaves DNA into a discrete set of fragments, the isolated C-domain gives a smear, characteristic to a non-specific nuclease (Figure 26A). The isolated N-domain has no detectable nuclease activity on the phage λ DNA.

In the gel shift assays, the isolated N-domain of BpuJI bound readily to the cognate 16/16(SP) oligoduplex, but not to the DNA lacking the recognition site (Figure 26B). The C-domain showed no binding to either oligoduplex under the same experimental conditions. These results indicate that the BpuJI domains have distinct functions: the N-domain recognizes the 5'-CCCGT sequence and the C-domain possesses the nuclease activity.

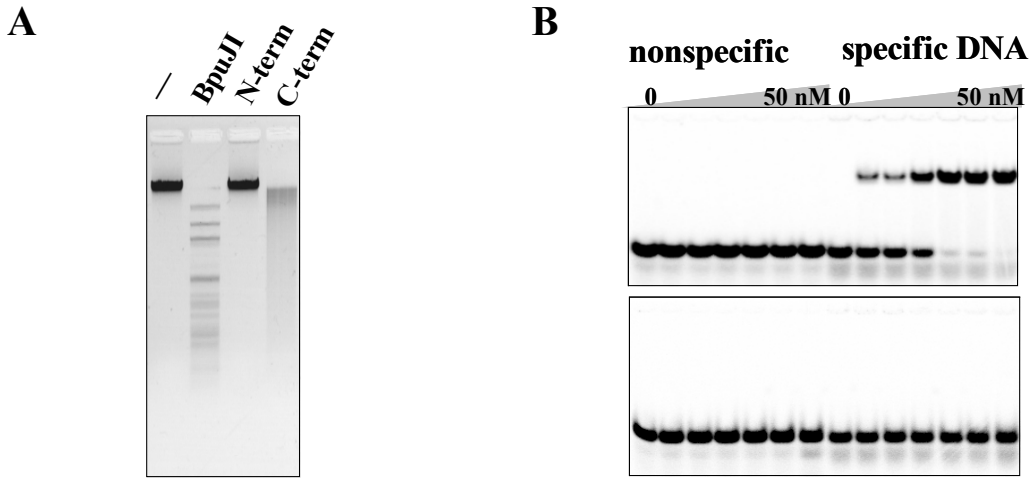

Figure 26. *Functional analysis of the isolated BpuJI domains. (A) The λ DNA cleavage assay. The reactions contained 20 μg/ml of the DNA and 25 nM of either BpuJI, the N-domain or the C-domain, respectively. After 1 h incubation at 37ºC, the products were analyzed by electrophoresis through 1% agarose. (B) The DNA binding assay. 2 nM of ^{33}P-labelled DNA, either the cognate 16/16(SP) or the non-cognate 16/16, were incubated with increasing amounts (0, 1, 2, 5, 10, 20, 50 nM) of either the N-domain (upper gel) or the C-domain (lower gel). The samples were subjected to non-denaturating PAGE for 2 h (see 2.11).*

3.2.3 The C-domain possesses the 3'-end-directed nucleolytic activity

The nuclease activity of the isolated BpuJI C-domain was further investigated using the 5'-labelled 179 bp PCR fragment (Figure 27). The C-domain of BpuJI initially cut the radiolabelled bottom DNA strand 4 nt from the 3'-terminus and afterwards processed it further to shorter fragments. This cleavage pattern is very similar to that observed for the full-length BpuJI on the same fragment in the presence of the specific DNA (Figure 22). Thus, the end-directed nuclease activity of the intact protein results from the C-domain.

To characterize the end-directed activity in more detail, a set of non-cognate oligoduplexes with different DNA ends was employed. Oligonucleotides, 37 nt and 25 nt in length, were annealed to form the DNA

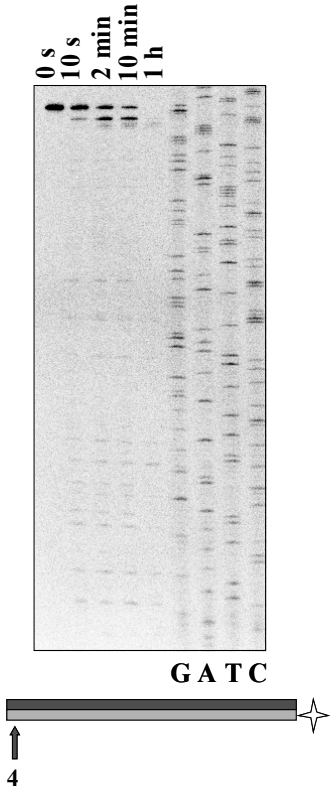

Figure 27*. Cleavage of the PCR fragment by the BpuJI C-domain. 1 nM of the PCR fragment, 5'-labelled in bottom strand, was incubated with 3 μM of the C-terminal domain at +4°C. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analysed by electrophoresis through the standard sequencing gel. Lanes G, A, T, C are sequence ladders.*

substrates: blunt-ended 37/37 and 25/25, a 37/25 oligoduplex with a 3'-overhang, and a 25/37 oligoduplex with a 5'-overhang (Table 3). The oligoduplexes were either 3'- or 5'-labelled in the top strand and treated with the isolated BpuJI C-domain (Figure 28). The C-domain cut the 3'-labelled strand in the blunt-ended 37/37 oligoduplex to yield a single ~3 nt product indicating the initial cleavage from the 3'-end. The analysis of the cleavage of the 37/37 oligoduplex radiolabelled at the 5'-terminus reveals further digestion of the top strand, resulting in the accumulation of smaller DNA fragments.

The top strand in the 37/25 oligoduplex with the protruding 3'-end was also cleaved by the C-domain, but more slowly than the blunt-ended 37/37 substrate. The cut site was located on the top strand within the duplex region of the oligonucleotide. The top strand in the 25/37 oligoduplex with the recessed 3'-end was the poorest substrate for the C-domain. However, the same strand in the blunt-ended 25/25 oligoduplex was cleaved ~4 nt from the 3'-terminus at

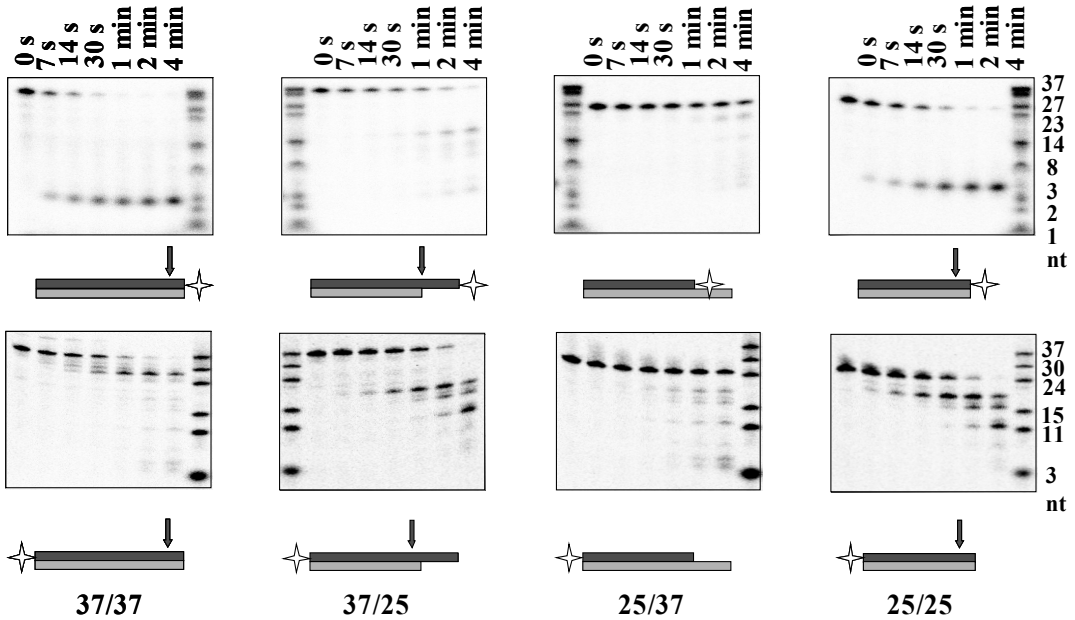

Figure 28. *Cleavage of oligoduplexes by the BpuJI C-domain. 2 nM of the oligoduplex, 3'-labelled (upper gels) or 5'-labelled (lower gels) in the top strand, was incubated with 100 nM of the C-domain at 25°C. Aliquots were removed at timed intervals (indicated above the relevant lanes) and analyzed by electrophoresis through denaturing 20% polyacrylamide. In the cartoons below the gels, the position of the label is indicated by a star, the arrows show the position of initial cleavage.*

a respective rate and further processed to smaller products. Thus, the C-domain of BpuJI shows preference for blunt-ended DNA and first cuts ~3 nt from the 3'-end.

3.2.4 The C-domain has a PD-(D/E)XK fold

To find out whether the BpuJI catalytic domain is related to any known nuclease structures, dr. Česlovas Venclovas (Institute of Biotechnology) performed fold recognition analysis. It revealed that the best-matching structures are those of archaeal Holliday junction resolving (AHJR) enzymes

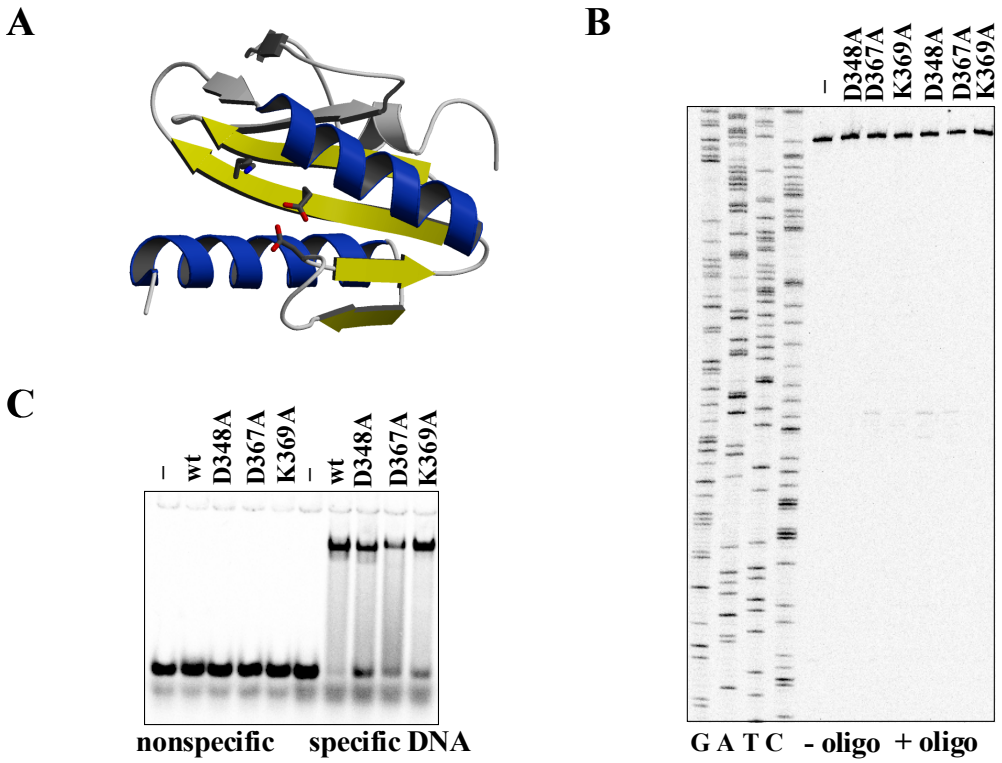

Figure 29. *(A) A schematic representation of the archaeal Holliday junction resolvase Hjc (PDB:1gef)[137], the closest structural relative of the BpuJI C-domain. Four β-strands and two α-helices that form a conserved structural core of the PD-(D/E)XK superfamily are shown in color. Side chains of the active site residues are represented as sticks. (B) Cleavage of the specific PCR fragment by the BpuJI D348A, E367A and K369A mutants. 1 nM of the PCR fragment, 5'-labelled in the top strand, was incubated for 14 hours with 3 μM of protein, in presence or absence of the 16/16(SP) oligoduplex. Products were analysed by electrophoresis through the standard sequencing gel. Lanes G, A, T, C are sequence ladders. (C) DNA binding by the BpuJI mutants. The reactions contained 100 nM of protein, 2 nM of ^{33}P-labelled and 100 nM of unlabelled 16/16(SP) or 16/16 oligoduplex in the binding buffer. The samples were subjected to the non-denaturating PAGE for 2 h (see 2.11).*

Hjc/Hje[137-139]. These archaeal resolvases are members of the PD-(D/E)XK nuclease superfamily[134, 135] and represent some of the closest approximation of the minimal core fold (Figure 29A). Moreover, the catalytic triads of Holliday junction resolvases perfectly aligned with Asp348, Glu367and Lys369 of the BpuJI C-terminal domain, suggesting that the residues of the

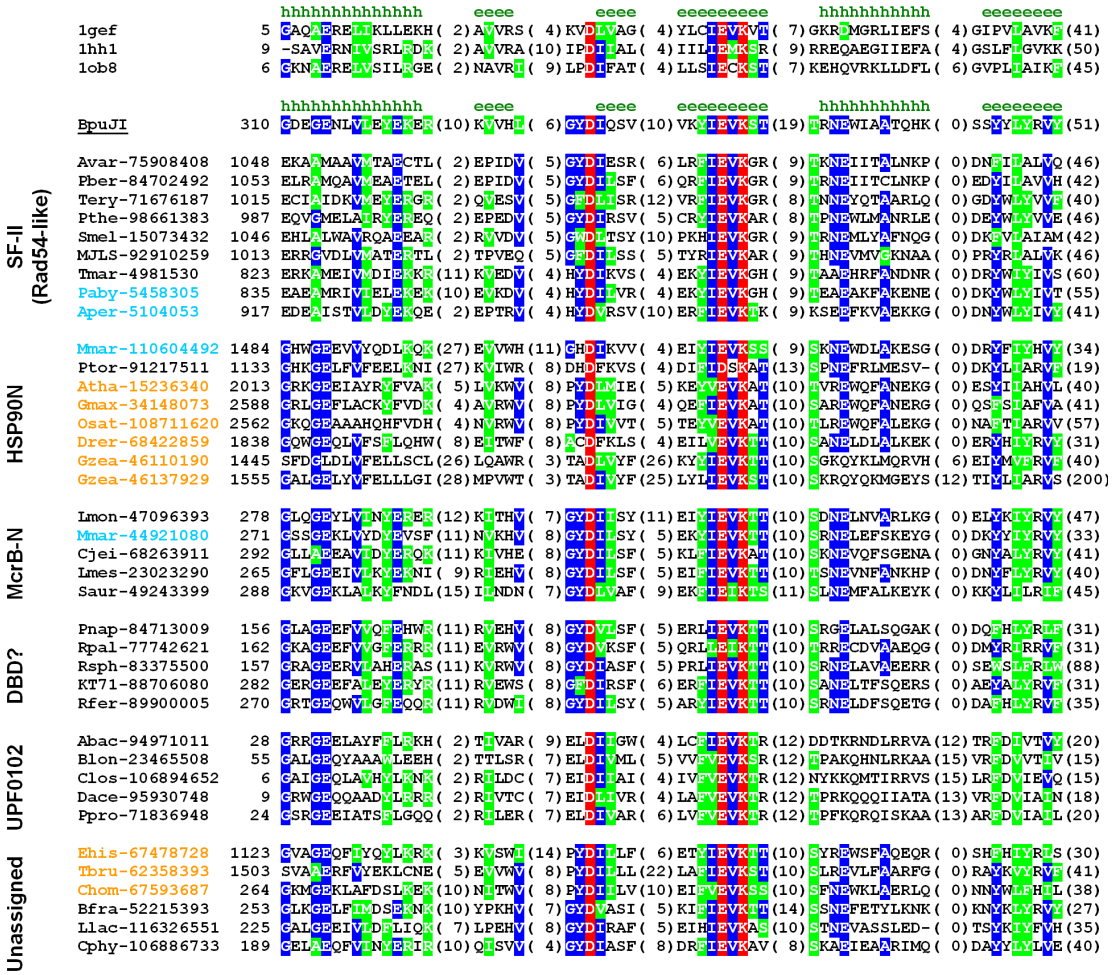

Figure 30. *Multiple sequence alignment between the archaeal Holliday junction resolvases of known structure, the BpuJI C-domain, and other homologs, grouped by the presence of additional conserved domains. Only the regions forming the conserved structural core of the PD-(D/E)XK superfamily (Figure 30A) are included. Secondary structure for 1gef and predicted secondary structure for the BpuJI C-domain are displayed above the respective sequences. Both known and putative active site residues are indicated with red shading. Blue and green colors indicate identical or similar residues,respectively, present in a given position in more than half of the sequences. Sequence names are composed of the organism abbreviation and GI number. Names of bacterial sequences are black, archaeal – blue and eukaryotic – orange (DBD – DNA binding domain).*

$YD^{348}...E^{367}VK^{369}$ motif form the active site of BpuJI (Figure 30).

To test this prediction, D348, E367 and K369 were targeted to alanine replacement mutagenesis. The single amino acid substitutions D348A, E367A and K369A strongly impaired the catalytic activity of the enzyme. The mutants showed only trace of the sequence-directed nuclease activity and no end-directed nuclease activity on the specific PCR fragment (Figure 29B), both in the absence and presence of the cognate oligonucleotide duplex. In a gel shift assay, the D348A, E367A and K369A mutants bound to the 16/16(SP) oligoduplex with the BpuJI recognition site but not to the 16/16 oligoduplex (Figure 29C). Thus, as expected for active site mutants, the D348A, E367A and K369A proteins are defective in catalysis, but not in DNA recognition.

3.2.5 The N-domain binds a single DNA copy as a monomer

According to the gel filtration data, the N-domain of BpuJI is a monomer in solution (see 3.2.2). However, one cannot exclude that dimerization of the BpuJI N-domain is required for cognate DNA binding. For example, the N-terminal domains of SdaI REase dimerize to bind a single copy of cognate DNA[15]. Therefore, the stoichiometry of the BpuJI N-domain binding to the cognate oligoduplex was analysed by size exclusion chromatography in the presence of the cognate DNA. The isolated N-domain eluted at a volume that corresponds to an apparent Mw of ~40 kDa, while the 16/16(SP) oligoduplex eluted with an apparent Mw of ~27 kDa (Figure 31). The latter value is ~2.5 times higher than the actual molecular mass of a 16 bp oligoduplex (10.7 kDa), most likely due to cylindrical shape and much higher frictional ratio compared to spherically-shaped standart proteins. When the 16/16(SP) oligoduplex was added to the BpuJI N-domain at equimolar ratio, the complex eluted from the gel filtration column as a single peak. The apparent mass value for this complex was ~54 kDa, close to that expected for the monomer of the N-domain bound to the DNA duplex. Thus, the BpuJI N-domain binds DNA as

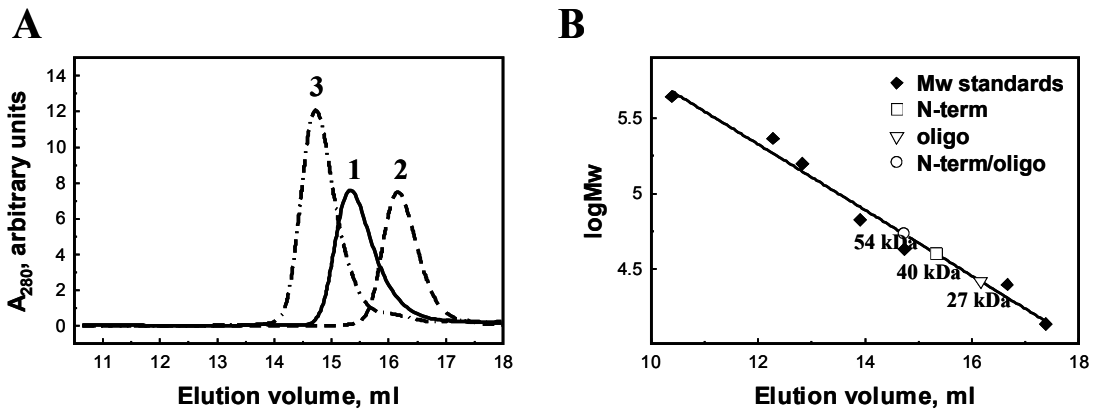

Figure 31. *Gel-filtration of the BpuJI N-domain and its complex with the 16/16(SP) oligoduplex. 6 μM of the N-domain (1), 2 μM of the 16/16(SP) oligoduplex (2) and 2 μM of the N-domain mixed with 2 μM of the 16/16(SP) (3) were loaded on a Superdex 200 HR column (see 2.13). The apparent mass values for the N-domain (□), the N-domain/DNA complex (○) and the 16/16(SP) oligoduplex (▽) were calculated by interpolation from a calibration curve (♦) obtained using a set of standard proteins.*

a monomer.

3.2.6 Crystallization of BpuJI and its proteolytic fragments

To get a more detailed view of the BpuJI structure, we tried to crystallize apo-BpuJI and its complex with cognate DNA; however, the best crystals of BpuJI diffracted X-rays only to ~7 Å. The poor diffraction could result from the conformational flexibility between the recognition and catalytic domains. Therefore, we focused on the crystallization of the separate BpuJI domains

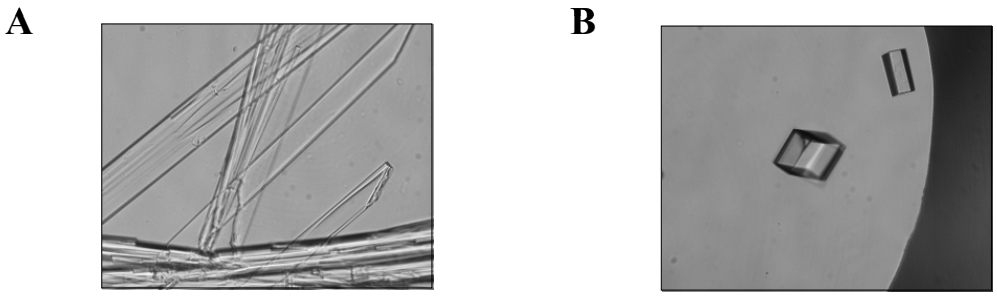

Figure 32. *Crystals of the BpuJI N-domain/DNA complex. (A) Polycrystals of the BpuJI N-domain with the 16/16(SP) oligoduplex, diffracting X-rays to ~2 Å. The crystals grew in 100 mM sodium citrate (pH 4.8), 20% PEG6000. (B) Single crystals of the BpuJI N-domain in complex with the 12/12(SP), grown in 0.2 M ammonium tartrate and 20% PEG3350. The crystals diffract X-rays to 1.3 Å.*

obtained by limited proteolysis. The BpuJI catalytic C-domain also gave crystals that were unsuitable for structure determination, but the N-terminal recognition domain was more promising in crystallization. The N-domain (residues 1-285) purified in the presence of the 16 bp cognate DNA 16/16(SP) (Table 3) gave polycrystals difracting to ~2 Å (Figure 32A). Single crystals (Figure 32B) were obtained using the 12 bp oligoduplex 12/12(SP) instead of the 16/16(SP). These crystals diffracted X-rays to near-atomic resolution and enabled to determine the crystal structure of the BpuJI N-domain/DNA complex. The crystallization results are summarized in Table 8.

Table 8. The crystals of BpuJI and its proteolytic fragments

Protein	DNA	Crystals*	Resolution, Å*
BpuJI	-	polycrystals	~7
BpuJI	16/16(SP)	no crystals	-
C-domain	-	single crystals	~10
C-domain	-	thin plates	~3
N-domain	16/16(SP)	polycrystals	~2
N-domain	**12/12(SP)**	**single crystals**	**1.3**

*See 2.7 for the detailed crystallization conditions and data collection statistics.

3.2.7 The crystal structure of the N-domain/DNA complex

The structure of the BpuJI recognition domain in complex with the DNA was determined by the single-wavelength anomalous dispersion (SAD) method, using a mercury derivative, and refined to 1.3 Å resolution with free R-factor of 16.6% and working R-factor of 13.8%. The electron density shows all the DNA bases and the amino acid residues, except of the five C-terminal residues 281-285. The asymmetric unit contains one protein monomer bound to the DNA duplex, in accordance with the gel-filtration data (see 3.2.5).

The BpuJI N-domain (Figure 33A) contains two winged-helix subdomains[39], D1 and D2, which are interspaced by a DL subdomain. The D1 subdomain (residues 1-112) contains an N-terminal arm, five helices (H1 to H5) and three strands (B1 to B3) (Figure 33B). The helices H2 and H3 form an HTH structure, which is followed by a signature wing region (B2-B3) and the

helices H4-H5. The loop between the helices H1 and H2 adopts an extended configuration and is incorporated as a third strand B1 in the sheet. The D2 subdomain (residues 137-245) is composed of six helices (H7 to H12) and two strands (B6-B7) (Figure 33B). D2 also bears an HTH structure, formed by the helices H8 and H11. The HTH motif is followed by a β-hairpin wing (B6-B7) and the helix H12. In contrast to canonical transcription factors[39], an expected

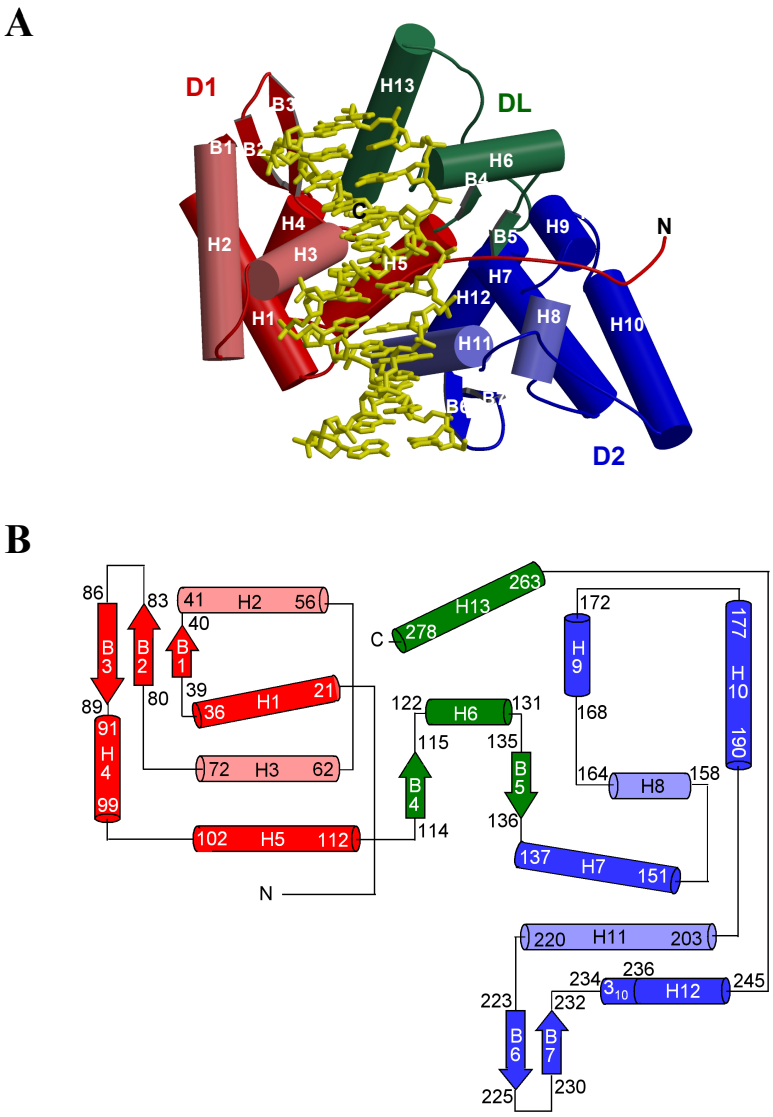

Figure 33. *(A) The BpuJI N-domain (1-285) bound to DNA. The subdomains D1, D2 and DL are shown in red, blue and green, respectively, DNA is represented by yellow sticks. The HTH motifs of the D1 and D2 subdomains are in light red and blue, respectively. (B) A topological diagram of the BpuJI N-domain. Helices are shown as cylinders, β-strands as block arrows. The secondary structure elements are colored as above (A).*

'turn' in the HTH motif of D2 is replaced by a 35 amino acid insertion which contains the helices H9-H10.

A 24 amino acid linker region with the chain topology B4-H6-B5 connects the subdomains D1 and D2 (Figure 33). The helix H13 forms a C-terminal extension which stands apart from the the D2 subdomain and is close to the linker region. The solvent inaccessible surface area of 1663 $Å^2$ buried at the interface between the linker and the helix H13 is larger in comparison to the surface buried between the linker-H13 fragment and the D1 or D2 subdomains (1192 $Å^2$ and 1140 $Å^2$, respectively). Thus, the internal contacts between the linker and the H13 residues are more extensive than the external contacts to the D1 and D2 subdomains. Based on this criterion, the linker and the C-terminal extension including the helix H13 could be classified as an interrupted subdomain DL.

The DNA in complex with the BpuJI recognition domain maintains a canonical B-DNA conformation with no major bends or kinks (Figure 33A). The recognition helices of the both the HTH motifs (H3 and H11) are inserted into the major groove on the opposite sides of the DNA. The N-terminal arm lies in the major groove and makes numerous contacts with the DNA bases and sugar phosphate backbone. The DL subdomain also is involved in protein-DNA interactions: the linker segment makes contacts with the DNA in the minor groove, while the C-terminal helix interacts with the sugar phosphate backbone upstream of the recognition site.

3.2.8 DNA recognition by the BpuJI N-domain

The BpuJI N-terminal domain recognizes the entire target sequence 5'-CCCGT by direct readout in the major groove. The interactions between the protein and the target bases are mediated by residues located solely in the recognition helices of the HTH motifs (H3 and H11) and the N-terminal arm (Figure 34).

The recognition helix of the BpuJI D1 subdomain (H3) recognizes the two CG base pairs at the 5'-end of the recognition sequence (CCCGT). The contacts with the DNA bases are exclusively mediated by Lys63, Asn67 and Glu71 residues (Figure 34): Lys63 donates a hydrogen bond to the O6 of the guanine G_5* (bases are numbered as in Figure 34); Asn67 forms hydrogen bonds to the N4 of the cytosine C_1 and the O6 of the guanine G_4*; the Glu71 accepts a hydrogen bond from the N4 of the cytosine C_2. The inner CG base

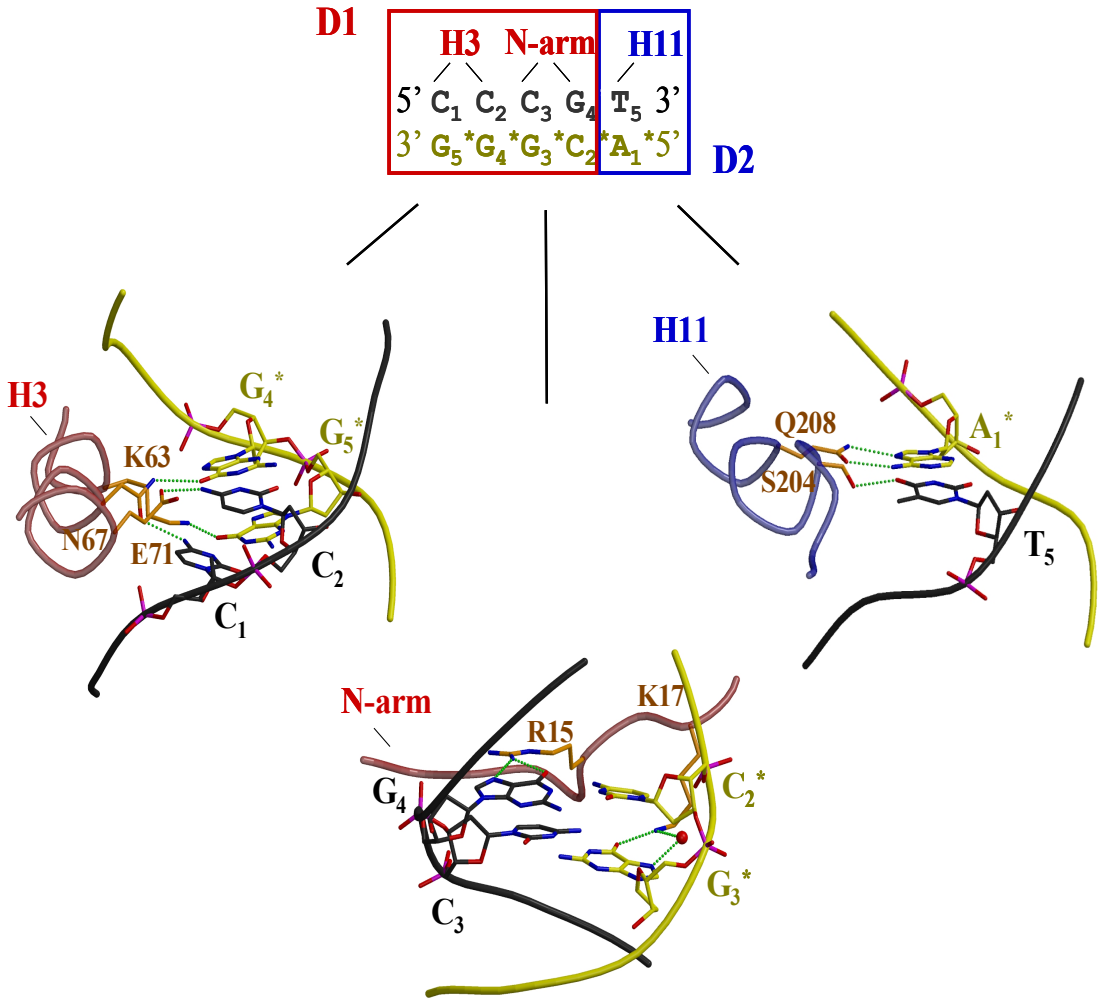

Figure 34. *Target recognition by the BpuJI N-domain. The DNA base pairs recognized by the D1 and D2 subdomains are delineated in red and blue, respectively; the secondary structure elements contacting the target bases are indicated above the respective base pairs. A detailed view of the recognition helix H3, the N-terminal arm and the recognition helix H11 interactions in the major groove is presented below. Water molecules are represented by red spheres, hydrogen bonds are shown as dotted lines.*

pairs (CCCGT) are recognized by Arg15 and Lys17 residues, located in the N-terminal arm (Figure 34). Lys17 donates a hydrogen bond to the O6 of the guanine G_3* and interacts via a water molecule with the N7 of the same guanine, while Arg15 donates a bidentate hydrogen bond to the guanine G_4. The D2 subdomain recognizes only one TA base pair at the 3'-end of the recognition sequence (CCCGT), and contacts with the DNA bases are mediated by the N-terminal portion of the recognition helix H11: Gln208 forms a bidentate hydrogen bond to the adenine and Ser204 donates a hydrogen bond to O4 of the thymine (Figure 34).

The backbone phosphate groups form an extensive set of hydrogen bonds and salt bridges with the residues located in the recognition helices (H3 and H11) and in the loops of the HTH motifs: there are four direct and three water-mediated contacts with the HTH motif of the D1 subdomain, four direct and four water-mediated contacts with the HTH motif of the D2 subdomain (Figure 35). The positive helix dipoles at the N-terminus of the recognition helices H3 and H11 could be in favorable interaction with the negatively charged phosphate groups of the DNA backbone. There also are two salt bridges, nine water-mediated hydrogen bonds and two van der Waals contacts between the backbone phosphates and the residues located in the N-terminal arm (Figure 35).

The DL subdomain interacts with the DNA upstream of the target sequence. The residues of the linker segment interact with DNA in the minor groove, making one direct and two water-mediated contacts with the backbone phosphate groups (Figure 35). Lys121 forms a hydrogen bond and a water mediated contact to the DNA bases upstream of the recognition site. The Lys266 located in the C-terminal helix H13 forms a water-mediated contact to the phosphate located 3 bp upstream of the target site (Figure 35), while Glu263, Lys266 and Lys267 form salt bridges with the phosphates of a symmetry-related DNA molecule (not shown).

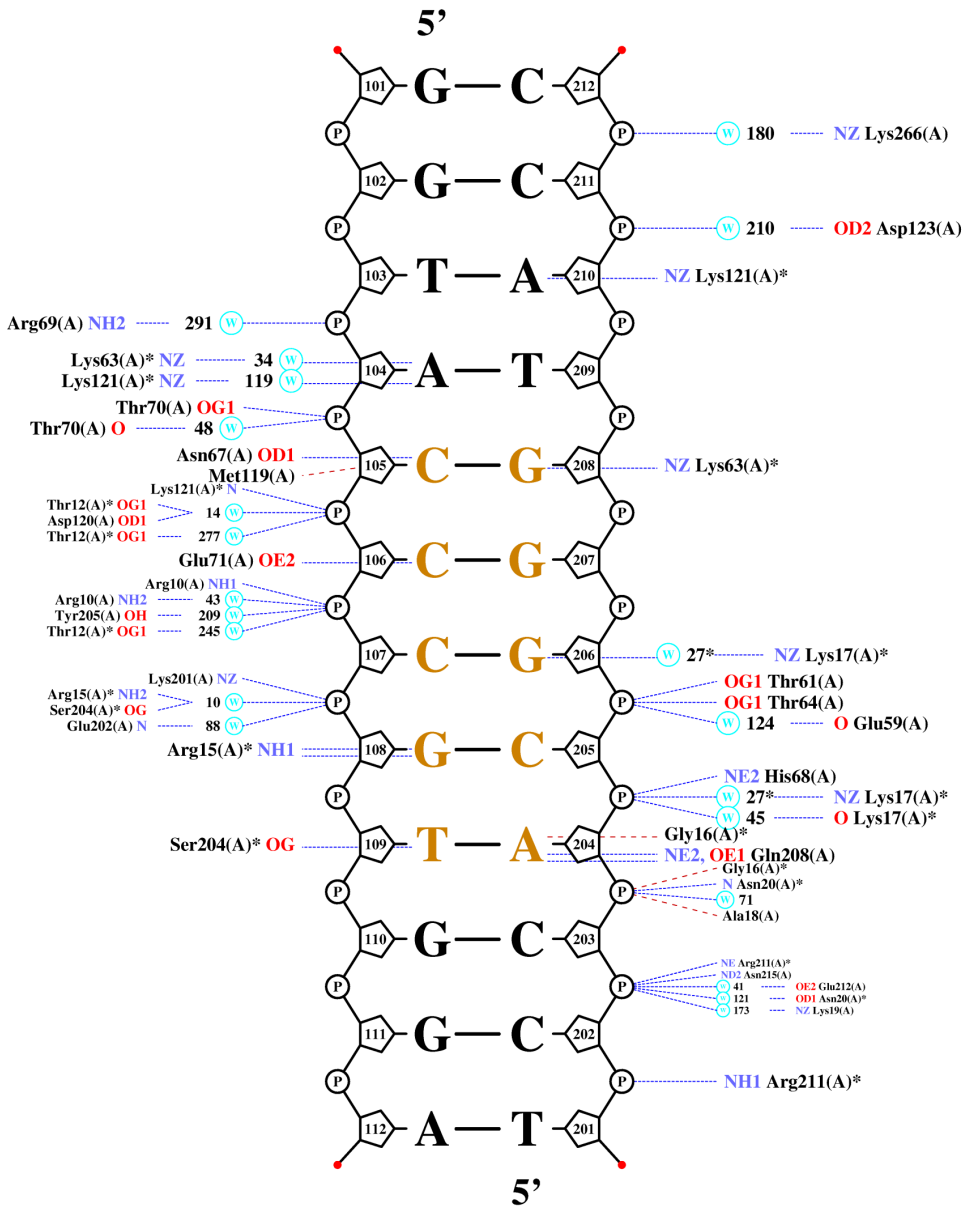

Figure 35. *A schematic diagram summarizing DNA contacts by the BpuJI recognition domain. Hydrogen bonds and salt bridges (<3 Å) are indicated with blue lines, van der Waals contacts (<3.35 Å) are indicated with red lines. Non-bridging water molecules are not shown. P - phosphate group, w - water molecule, * - residue or water on plot more than once.*

3.2.9 Mutational analysis of the DNA binding interface

The protein-DNA contacts observed in the crystal structure were tested by site-directed mutagenesis. The residues Lys63, Asn67, Glu71, Arg15, Lys17, Ser204 and Gln208, which form hydrogen bonds with the target bases in the

major groove, as well as Lys121 interacting with the flanking DNA bases in the minor groove, were separately mutated to alanines to evaluate these residues for their individual contributions to DNA binding.

Both the wild type (wt) and mutant proteins were expressed and purified as the N-domain (1-285) fusions with a C-terminal hexahistidine tag, and their DNA binding properties were studied by gel shift analysis. The ^{33}P-labelled 16/16(SP) and 16/16 oligoduplexes were used in the DNA binding experiments as cognate and non-cognate DNA, respectively.

The gel shift analysis demonstrates that the BpuJI recognition domain binds to the cognate DNA with high affinity (K_d=7.3±0.3 nM), while the non-cognate DNA binding remains very low (Figure 36). The ability to bind the cognate DNA is almost completely abolished by the single N67A, E71A, R15A or Q208A mutations and highly compromised by the K63A (K_d=48±7.4 nM) or K17A (K_d=67±14 nM) mutations. On the other hand, the S204A mutation has only minor effect on the DNA binding (K_d=9.6±1.0 nM). Taken together, the mutational analysis data are consistent with the major groove interactions observed in the crystal structure.

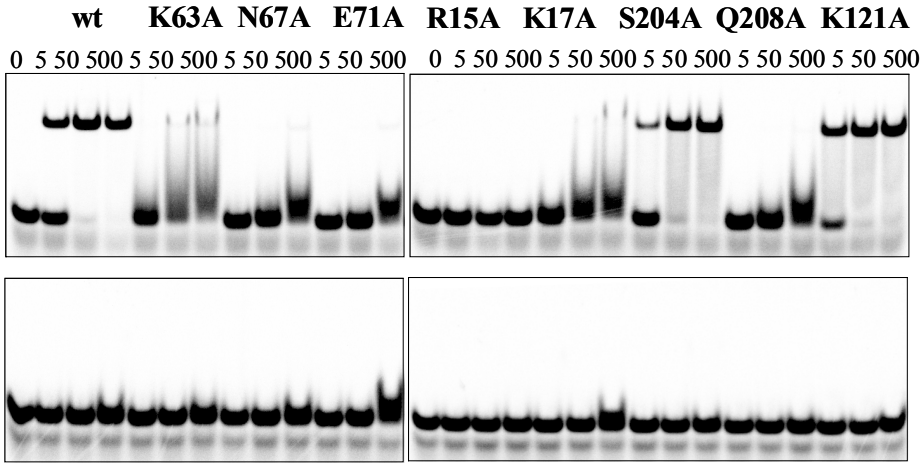

Figure 36. *DNA binding by the BpuJI N-terminal domain and its mutants. The binding reactions contained 2 nM of ^{33}P-labelled DNA, either the specific 16/16(SP) (upper gels) or the non-specific 16/16 (lower gels), and 0-500 nM of the wt or mutant protein (indicated above the relevant lanes). The samples were subjected to non-denaturating PAGE for 2 h (see 2.11).*

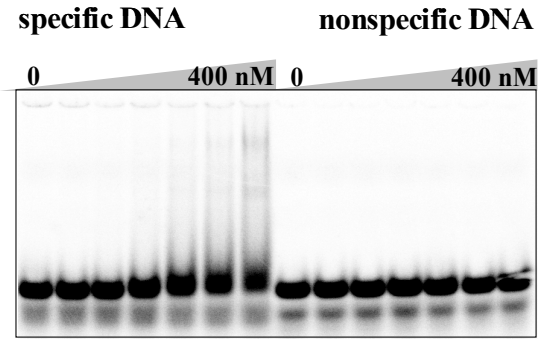

Figure 37*. DNA binding by the BpuJI K121A mutant. The binding reactions contained 2 nM of ^{33}P-labelled DNA, either the specific 16/16(SP) or the non-specific 16/16, and the K121A mutant (0, 10, 20, 40, 100, 200, 400 nM). The samples were subjected to non-denaturating PAGE for 2 h (see 2.11).*

Surprisingly, alanine replacement of Lys121, which interacts with the TA dinucleotide upstream of the BpuJI target site 5'-CCCGT (Figure 35), increased the binding affinity of the BpuJI N-domain to the cognate DNA nearly 10 times (K_d=0.83±0.03 nM). To test if the Lys121 mutant shows the same binding phenotype in the context of the full length BpuJI, the K121A mutation was introduced into the full-length BpuJI gene. In contrast, the purified full-length K121A mutant showed significantly lower binding affinity (K_d ~ 200 nM) to the cognate DNA (Figure 37) than wt BpuJI (K_d ~ 25 nM) (Figure 23). Also, the K121A mutant cleaved the phage ΦX174 DNA ~3 times more slowly (data not shown). These results suggest that Lys121 could be important in the BpuJI interactions with cognate DNA; however, the exact role of the Lys121 residue remains unclear, as no structure of the full-length BpuJI/ DNA complex is available.

3.3 A model of BpuJI action and its similarity to other proteins

3.3.1 A model of BpuJI action

The data provided here demonstrate that BpuJI consists of two physically separate domains: the N-domain with two winged-helix motifs recognizes the

5'-CCCGT sequence as a monomer, while the C-domain contains a single PD-(D/E)XK active site. The two functional domains in BpuJI are presumably connected by a flexible linker. BpuJI is a dimer in solution, and the dimerization interface is formed by the C-domains (Figure 38, A). Therefore, BpuJI has two surfaces for specific DNA binding, provided by the N-terminal domains, and forms synaptic complexes by binding concurrently to two separate segments of specific DNA (Figure 38, B). In full-length BpuJI, the catalytic domain has very a low activity. It can be activated either by proteolytic cleavage or specific DNA binding. The activated catalytic C-domain of BpuJI becomes a nuclease that has no defined sequence specificity, but shows a 3'-end directed endonucleolytic activity and

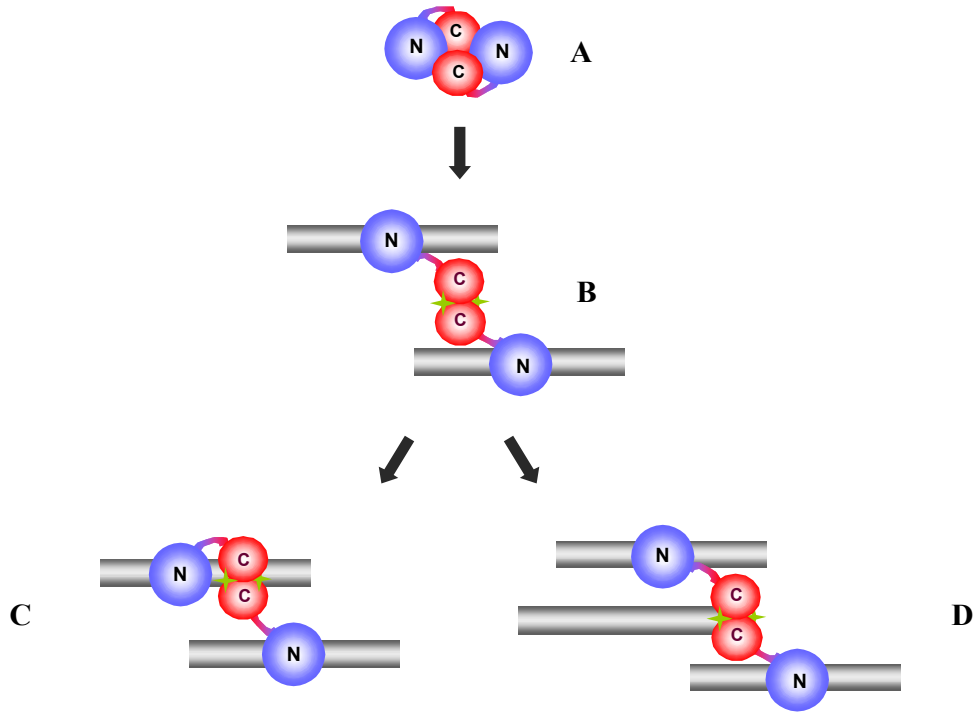

Figure 38. *A cartoon representation of the BpuJI action. BpuJI is a dimer and contains two domains: the N-domain and the C-domain (shown in blue and red, respectively). The C-domain possesses one active site, represented as a green star. There are two active sites in the dimer. In the absence of the target sequence, the nuclease activity of the C-terminal domain is repressed (A). Binding of the cognate DNA by the N-terminal domains triggers a conformational change (B) and activates the catalytic domain. The activated C-domain nicks DNA at a random position in the vicinity of the target site (C). In addition, the activated C-terminal domain possesses the 3'-end directed nuclease activity and can cleave another DNA molecule (D).*

preferentially cuts ~3 nt from the 3'-terminus of the blunt-ended DNA. The isolated catalytic domain of BpuJI is a dimer, and each monomer contains a single PD-(D/E)XK active site. The preferential cut of only one strand at the 3'-terminus by the activated C-domain suggests that only one active site per dimer is used for cleavage.

In the absence of the target sequence, the endonucleolytic activity of the catalytic domain in BpuJI is repressed. BpuJI shows only trace activity on the non-cognate DNA. The cognate DNA binding by the N-terminal domains of BpuJI presumably triggers a conformational change that relieves the autoinhibition and activates the C-terminal domain. If a BpuJI recognition site is embedded within a long DNA fragment, the activated catalytic C-domain may become capable of cleaving phosphodiester bond(s) at a random position (determined by the linker length or/and sequence context) in the vicinity of the recognition site (Figure 38, C). Independent cleavage of the second strand nearby of the first cut will be required, however, to generate the double strand break, if only one active site per dimer is used for the cleavage. A linear DNA resulting from BpuJI cleavage at the recognition site may be further subjected to the 3'-end directed endonucleolytic cleavage by the activated C-domain, as well as another DNA molecule, present in the solution (Figure 38, D). The BpuJI cleavage near the recognition site followed by the 3'-end directed endonucleolytic action results in a complicated cleavage pattern.

3.3.2 The place for BpuJI in the classification of restriction enzymes

BpuJI cleaves DNA on both sides of the recognition sequence, and therefore it could be classified as a Type IIB restriction enzyme. However, Type IIB REases have bipartite recognition sequences and cleave DNA at specified positions on both sides of their target sequences to excise a short fragment containing the recognition site[2], while BpuJI makes multiple cuts at

both DNA strands at distances varying from 2 to 23 nt. Furthermore, Type IIB enzymes are multi-functional proteins with both methyltransferase and restriction endonuclease activities, requiring Mg^{2+} and in many cases AdoMet for DNA cleavage[110]. The organization of the BpuJI R-M system and cofactor requirements make it distinct from Type IIB REases. BpuJI also differs from Type I and Type III enzymes,which cleave at multiple sites but are multi-subunit and multi-functional proteins. These enzymes use translocation mechanisms and require ATP to cleave DNA at sites distant from their binding sites[159].

The asymmetric pentanucleotide recognition sequence and the organization of the BpuJI R-M system (a single REase and two separate MTases for different strands) suggest that it could be categorized as a Type IIS system. The BpuJI structural organization also is similar to that of the archetypal Type IIS REase FokI. Both the enzymes are composed of two domains: the N-terminal domain binds to the target sequence as a monomer, and the C-terminal domain has one catalytic center. The main difference in the structural organization is that BpuJI is a dimer in solution, while FokI is a monomer and forms a dimer on cognate DNA to cleave both DNA strands[87, 90, 91]. However, according to the current nomenclature[4], the overriding criterion for inclusion as a Type II enzyme is a defined fragmentation pattern and cleavage either within or close the recognition sequence at a fixed site or with known and limited variability. Therefore, the current definition still allows classification of HphI[242] and BfiI[11] REases, which show some defined variability in the cleavage positions, as Type IIS enzymes, but creates an uncertainty in the case for BpuJI due to its unspecified cuts. It is interesting to note that chimeric REases, created by fusion of the FokI nuclease domain to zinc-finger or other DNA binding domains, cleave DNA near the target site but at not at well-defined positions[92-94, 243, 244].

3.3.3 Domains, similar to BpuJI C-domain, are present in diverse sequence contexts

Taken together, the fold recognition and experimental data confidently assigns the BpuJI C-domain to the diverse PD-(D/E)XK superfamily with distinct affinity to AHJRs (see 2.3.4). Despite of the fold similarity, the saturated PSI-BLAST search[206] starting with the BpuJI C-domain sequence did not retrieve AHJR sequences, but it identified BpuJI CTD-like domains in quite a diverse pool of proteins. This suggests that homologs detected with PSI-BLAST are more closely related to the BpuJI C-domain than AHJRs. Dr. Česlovas Venclovas identified several protein families on the basis of common domain architecture (Figure 39A).

Family I includes uncharacterized bacterial and archaeal proteins (~900-1200 a.a.) that represent SWI2/SNF2-like superfamily II helicase fused to the BpuJI CTD-like domain (Figure 39A). This type of fusion has been noted before[135]; however, no putative protein function was suggested. We noticed that the domain organization in Family I proteins is the same as in the Res subunits of characterized Type III R-M enzymes[159] (Figure 39B). Moreover, in absolute majority of cases we have detected an adenine-specific DNA methyltransferase gene in close proximity. Therefore, both the domain organization and the genome context suggest that most if not all Family I proteins are Res subunits of Type III R-M enzymes. Yet, there are two major differences that distinguish Family I proteins from the currently characterized Type III R-M systems. First, the helicase region in the Family I proteins is more closely related to Rad54, while the corresponding region in the Res subunits of conventional Type IIIA and IIIB enzymes is closer to Rad25 (Figure 39B). Second, the putative Mod subunits co-localized with Family I sequences and those of the characterized Type IIIA and IIIB systems belong to different orthologous clusters (COG1743 and COG2189, respectively).

Family II proteins, in addition to the BpuJI CTD-like domain, feature a

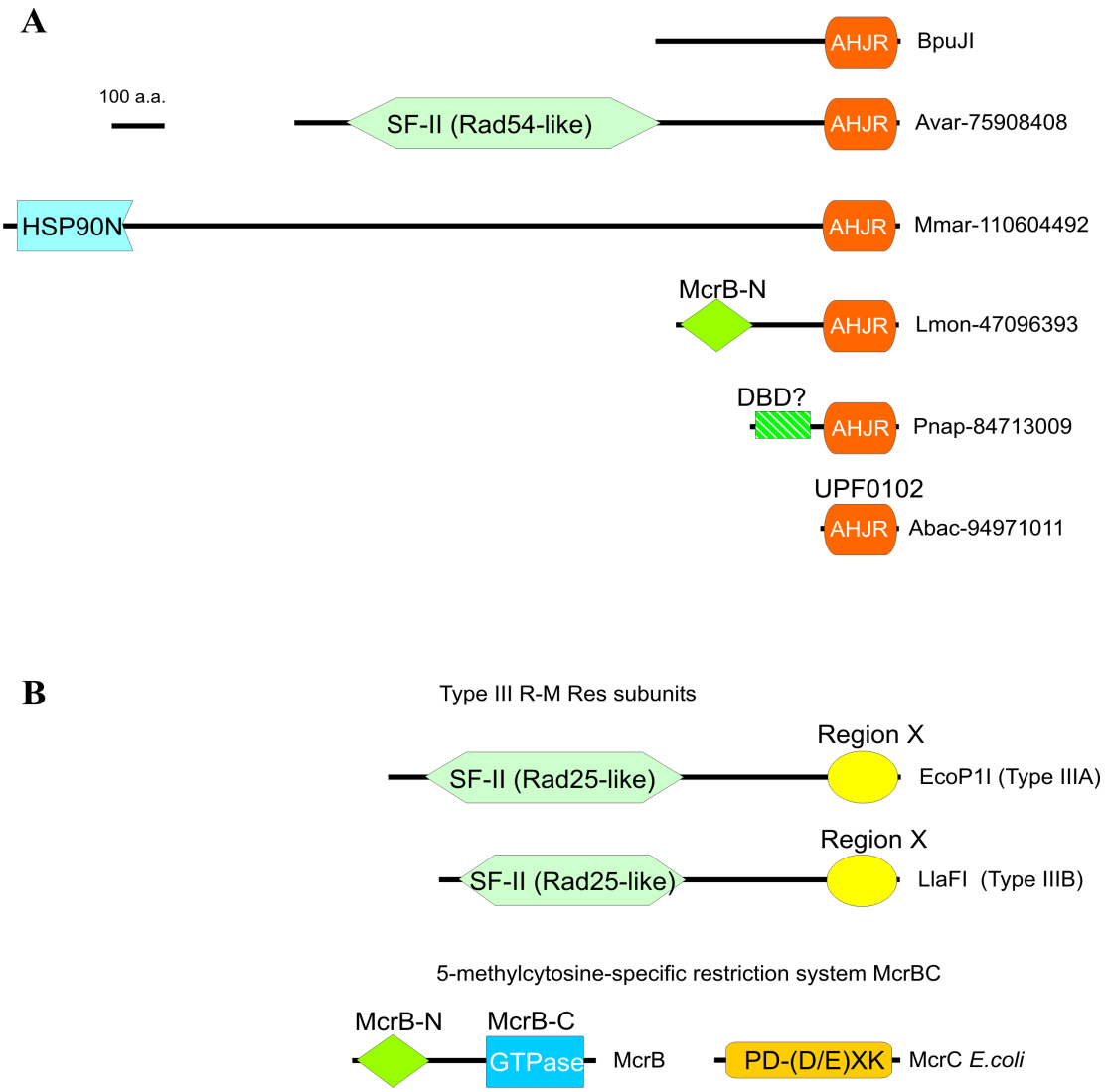

Figure 39. *The domain architectures of (A) the protein families sharing homology with the BpuJI nuclease domain identified in this study and (B) the previously characterized Type III Res subunits and the McrBC system. AHJR, the archaeal Holliday junction resolvase domain similar to the C-domain of BpuJI; for domains of superfamily II helicases, SF-II, the closest homolog of known structure is indicated; HSP90N, the N-terminal domain of HSP90; McrB-N and McrB-C, respectively the N-terminal DNA binding and the C-terminal GTPase domains of the McrB subunit of 5'-methylcytosine-specific restriction enzyme; DBD?, unclassified putative DNA binding domain; UPF0102, uncharacterized protein family with the predicted single-domain nuclease fold related to AHJRs; Region X and PD-(D/E)XK, endonuclease domains of Type III and McrC restriction enzymes, respectively. Each architecture in (A) is illustrated with a single representative and corresponds to the grouping in the alignment (Figure 31).*

common region that is most closely related to the N-terminal nucleotide binding domain of HSP90 family of molecular chaperones (COG0326)

(Figure 39A). This group includes large (~1100-2600 a.a.) proteins of unknown function.

The third group (Family III) is comprised of archaeal and bacterial proteins (~400 a.a.) sharing a conserved N-terminal region. Unexpectedly, we have found that this region shows similarities to the N-domain of McrB, the GTPase subunit of the 5'-methylcytosine-specific restriction enzyme McrBC (Figure 39B). In the *E.coli* McrB the N-domain is responsible for the recognition of the methylated target DNA sequence[181]. By inference, the Family III proteins could be REases acting on 5'-methylcytosine DNA targets. However, unlike the two-component restriction system McrBC, these putative REases have both recognition and nuclease domains encoded within a single polypeptide chain.

Family IV consists of bacterial proteins (~250-400 a.a.) that have a conserved N-terminal region fused with a BpuJI CTD-like nuclease domain (Figure 39A). We have found that this conserved N-terminal region in some other proteins is fused to the H-N-H nuclease domain instead of the BpuJI CTD-like domain. Although this conserved region does not show any apparent sequence similarity to characterized domains, its presence together with either nuclease domain suggests its likely involvement in DNA binding/recognition. However, in most cases we failed to find proximal MTase genes in genomic DNA sequences suggesting that this group of proteins do not function as restriction enzymes. Alternatively, they might act on methylated DNA targets, similarly to McrA.

There also is a statistically significant relationship (E-value<0.005) between the BpuJI C-domain and a large family of small (~130 a.a.) uncharacterized bacterial proteins (Pfam: UPF0102) (Family V, Figure 39A). The members of this family were computationally identified as bacterial counterparts of AHJR enzymes earlier[135], but their biological function has not been demonstrated.

3.3.4 Structural similarity of the BpuJI N-domain to other DNA binding domains

A DALI search against the PDB database revealed significant similarity of the BpuJI N-domain to other DNA-binding proteins (Table 9).

Table 9. Proteins showing similarity to the BpuJI recognition domain

Search target	PDB accession code	Protein	Z-score
BpuJI-Nt (1–280 a.a.)	2ewf	Nicking endonuclease Nt.BspD6I	13.4
	2qww	Transcriptional regulator, MarR family	5.6
	2nnn	Probable transcriptional regulator	5.5
	1yg2	Gene activator AphA	5.4
	1t98	Chromosome partition protein MukF fragment	5.3
BpuJI-D1 (20–131 a.a.)	2ewf	Nicking endonuclease Nt.BspD6I	6.4
	2fok	FokI REase	5.5
	2qww	Transcriptional regulator, MarR family	5.4
	1t98	Chromosome partition protein MukF fragment	5.4
	2nnn	Probable transcriptional regulator	5.2
BpuJI D2-DL (91–280 a.a.)	2ewf	Nicking endonuclease Nt.BspD6I	4.7
	2it0	Iron-dependent repressor	3.1
	1w5s	DNA replication iniciation protein	3.0
	2isz	Iron-dependent repressor	2.8
	2gau	crpFNR family transcriptional regulator	2.3

*Only five proteins with the highest Z-scores are shown

Most of these proteins were transcriptional regulators harboring a winged-helix motif. Surprisingly, the closest structural homolog of the BpuJI recognition domain and its separate subdomains was the nicking endonuclease Nt.BspD6I. The BpuJI D1 subdomain also showed significant similarity to the REase FokI, which was not found using the entire structure as a DALI search target. Nt.BspD6I[99] contains two winged-helix subdomains (D1 and D2) as BpuJI, while the FokI[8] recognition domain is composed of three winged-helix units (D1, D2 and D3).

The D1 subdomains of BpuJI, BspD6I and FokI can be superimposed along major structural elements (Figure 40A). The main difference is that the FokI winged-helix motif has an α-helical insertion, while the D1 subdomains of

BpuJI and Nt.BspD6I are more compact. The D2 subdomains of all the three proteins are extensively modified by large α-helical insertions upstream the recognition helices. The BpuJI and BspD6I D2 subdomains also can be superimposed along almost all structural elements. However, only the trihelical bundles and "wings" superimpose on the equivalent structural elements in

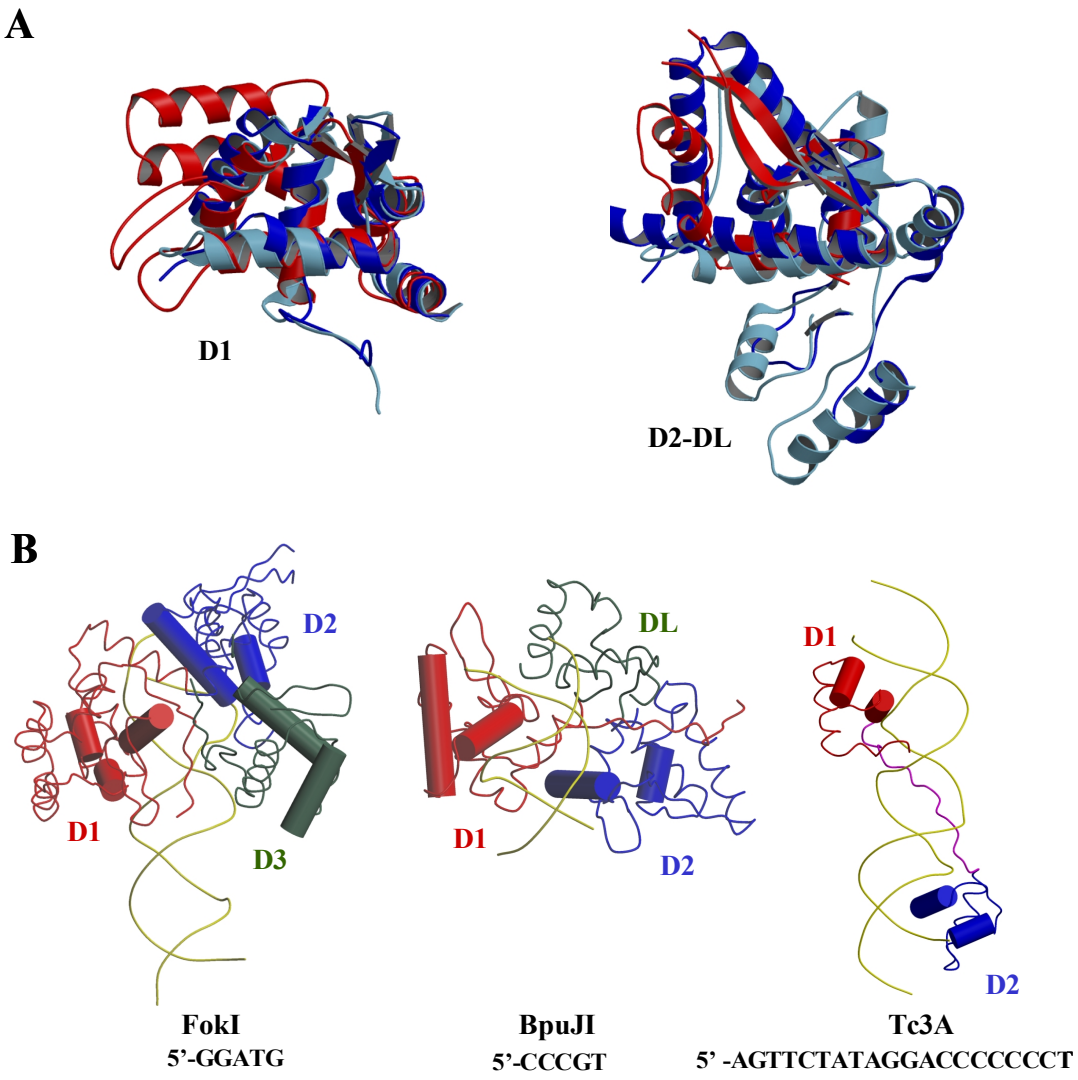

Figure 40. *(A) TOP3D[225] structural overlays of the BpuJI (light blue), Nt.BspD6I (PDB:2ewf)[99] (blue) and FokI (PDB:1fok)[8] (red) subdomains. The BpuJI and BspD6I subdomains can be superimposed along almost all structural elements. (B) Comparison of the FokI (PDB:1fok)[8], BpuJI and Tc3 transposase Tc3A (PDB:1u78)[247] recognition domains bound to DNA. The D1, D2 and D3 (DL in BpuJI) subdomains are shown in red, blue and green, respectively; the linker segment of Tc3A is shown in magenta; the HTH motifs are shown as cylinders. The binding sites of FokI, BpuJI and Tc3 transposase Tc3A are presented below the corresponding structures.*

FokI, while the α-helical insertions share no topological similarity. The similarity of the BpuJI D2 subdomain to FokI was not revealed by a DALI search using the C-terminal part of the protein as a search target (Table 9). The linker between the D1-D2 subdomains and the C-terminal extension form the interrupted DL subdomain in BpuJI. Nt.BspD6I also has a corresponding region, although it is less ordered than that of BpuJI. Taken together, the BpuJI and Nt.BspD6I recognition domains show more structural similarity to each other than to FokI.

The comparison to the BpuJI and FokI complexes with cognate DNA revealed that both BpuJI and FokI employ D1 and D2 subdomains to interact with their target sequences. The D3 subdomain of FokI, which is missing in BpuJI, mediates protein-protein rather than protein-DNA interactions[8]. Although the BpuJI and FokI subdomains share structural similarity, their orientation in the protein-DNA complexes differs significantly (Figure 40B). In BpuJI, both the recognition helices of the HTH motives pack against the major groove with their helical axes perpendicular to the long DNA axis, while in the FokI-DNA complex only the D1 recognition helix is perpendicular to the DNA axis (Figure 40B). The recognition helix of the FokI D2 subdomain juts away from the DNA, and its helical axis is tilted by ~35° with respect to the plane of the base pairs[8]. Also, the BpuJI D1 subdomain contacts DNA bases at the 5'-end of the recognition sequence (CCCGT), and the D2 contacts the base pair at the 3'-end (CCCGT). In contrast, the FokI D1 recognizes the 3'-end of the target sequence (GGATG), and the D2 subdomain recognizes the 5'-end (GGATG)[8].

Duplications of HTH units are found in a large number of protein families, such as sigma factors[245], POU family domains[246], the paired domains of the Pax and Prd transcription factors[14] and the recombinase DNA-binding domains[247]. The HTH units of bipartite domains usually bind to different regions of an extended binding site. In certain cases, an unstructured linker region fills in the

minor groove between domains, as shown for the DNA-binding domain of the Tc3 transposase Tc3A bound to 26 bp DNA[247] (Figure 40B). In contrast, the bipartite recognition domains of the restriction endonucleases BpuJI and FokI bind to the pentanucleotide target sequences. A wHTH motif also was identified in the REase SdaI, which recognizes the continuous palindromic 8 bp sequence[15]. However, the wHTH unit of SdaI forms a single globular domain and dimerizes to achieve the recognition of the 8 bp palindromic target site.

3.3.5 Comparison of BpuJI, BspD6I and FokI

The BpuJI recognition domain shows structural similarity to the archetypal Type IIS enzyme FokI and the nicking endonuclease Nt.BspD6I. The all three enzymes have a modular architecture with a recognition domain in the N-terminus and a cleavage domain in the C-terminus (Figure 41). The cleavage domains belong to the PD-(D/E)XK superfamily, while the recognition domains are comprised of several HTH subdomains. Like FokI, BpuJI and BspD6I recognize asymmetric pentanucleotide sequences and cleave DNA near the target sequence, but at different positions. Also, these enzymes have a different oligomeric organization (Figure 41). FokI (5-GGATG (9/13)) is a monomer in solution and when bound to the target sequence[8, 86]. As the FokI monomer has one catalytic center, it dimerizes in presence of cognate DNA to cleave both DNA strands[87]. BpuJI, which recognizes the 5'-CCCGT sequence and cuts DNA at multiple sites in the vicinity of the target sequence, in contrast to FokI, is a dimer in solution. The dimerization interface is formed by the C-terminal domains, therefore the BpuJI dimer has two surfaces for specific DNA binding provided by the N-terminal domains and forms a synaptic complex with cognate DNA. The nicking endonuclease Nt.BspD6I recognizes the 5-GAGTC sequence and cleaves only one DNA strand containing the target sequence, at the distance of four nucleotides from the recognition site. Actually, Nt.BspD6I is a large subunit of the heterodimeric REase BspD6I

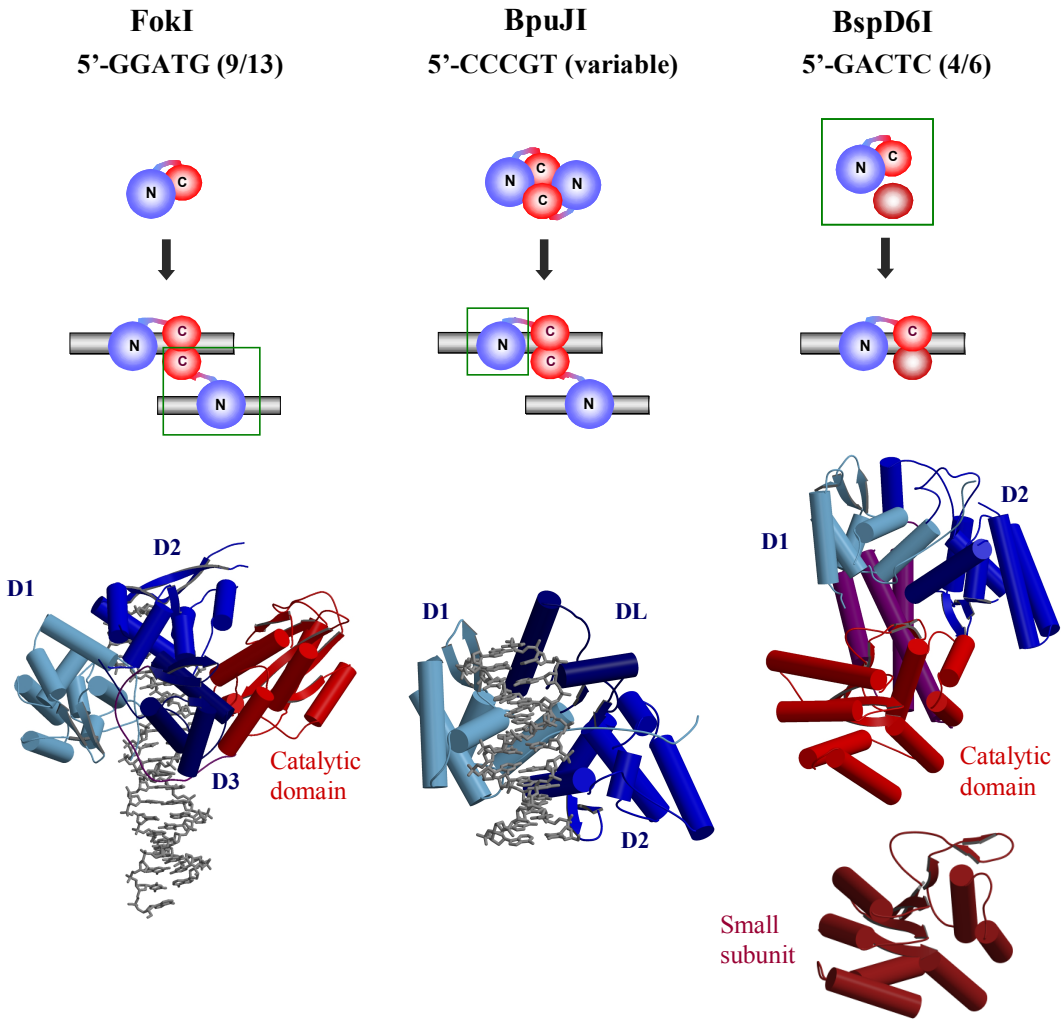

Figure 41. *The structural organization of FokI, BpuJI and BspD6I. The recognition and cleavage domains are shown in blue and red, respectively, the parts with crystal structures available are marked by green rectangles. FokI is a monomer in solution*[86] *but dimerizes to cleave both DNA strands*[87]*; BpuJI is a dimer in solution and forms a synaptic complex with cognate DNA. BspD6I is composed of two subunits: the large subunit is a nicking enzyme, the small subunit harbours an active site and is required to introduce a double-stranded break*[101]*. The crystal structures of the FokI/DNA complex (PDB:1fok)*[8]*, the BpuJI recognition domain/DNA complex and the BspD6I large (pdb:2ewf)*[99] *and small (pdb:2p14)*[99] *subunits are presented below the corresponding schemes. The subdomains D1, D2 and D3 (or DL) are shown in different shades of blue, the catalytic domains are red, DNA is grey.*

(5'-GACTC (4/6)). The small subunit of BspD6I, which is responsible for the cleavage of the bottom strand, shows similarity to the catalytic domain of Nt.BspD6I and harbours the active site residues[101].

3.3.6 A model of the Nt.BspD6I/DNA complex

The crystal structure of Nt.BspD6I solved without DNA[99] can be combined with the structural information for the BpuJI N-terminal domain/DNA complex to get insight into DNA recognition by BspD6I. The Nt.BspD6I complex with cognate DNA was modeled as follows. A B-DNA helix containing the cognate sequence 5'-GAGTC was generated. The DNA helix was superimposed onto the DNA fragment of the BpuJI/DNA complex so that the recognition sequences would be in the equivalent position. Although the recognized DNA sequences differ, their length (five base pairs) is the same. Next, the structure of Nt.BspD6I was superimposed onto the BpuJI N-terminal domain.

The model of the Nt.BspD6I N-terminal domain/DNA complex revealed that the DNA and the protein fit each other well with no steric clashes. Like in BpuJI, there are three putative specificity-determining regions in the Nt.BspD6I N-terminal domain. These include an extended N-terminal arm and the two recognition helices of the HTH motifs,which are facing the recognized base pairs within the major groove of DNA (Figure 42). From this model of the Nt.BspD6I/DNA complex, it is difficult to deduce the exact atomic interactions that are responsible for the recognition of the specific DNA sequence. However, the model enables to substantially reduce the number of possible candidate residues acting as specificity determinants.

In the model of Nt.BspD6I/DNA complex, the scissile phosphate (4 nucleotides following the recognition sequence) appears to be close to the nuclease active site. The distance from the carboxylate of Asp456, which was shown to be an active site residue[102], to the phosphorus atom is less than 7 Å (Figure 42). However, this distance is too large to represent the functional nuclease/DNA complex. In addition, there appears to be significant steric clashes between the DNA helix and a mostly helical motif of BspD6I (residues 394-415) (Figure 42). This implies that Nt.BspD6I would not be able to bind a canonical B-DNA without either kinking the DNA, rearranging the mutual

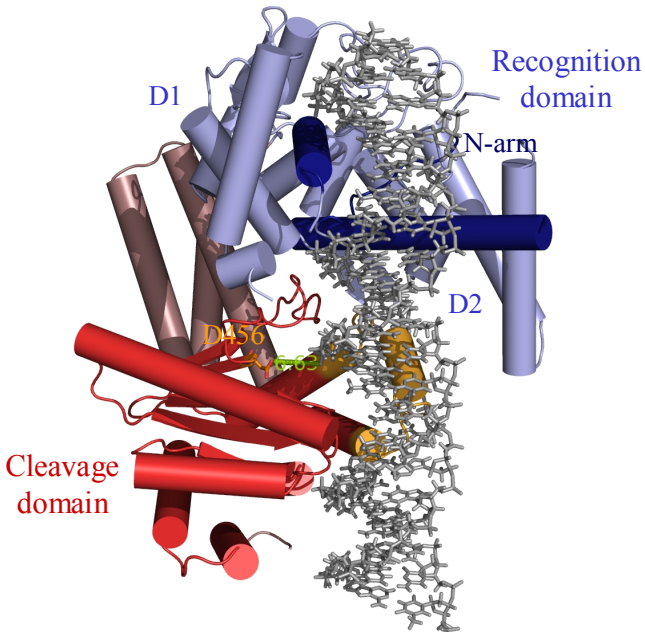

Figure 42. *A model of the Nt.BspD6I/DNA complex. The recognition domain is colored blue, the catalytic domain is red, DNA is grey. The putative specificity-determining regions (the N-terminal arm and two recognition helices of the HTH motifs) are highlighted in dark blue, the catalytic Asp456 is shown in stick representation, and the region of BspD6I that clashes with the DNA in the model is shown in orange.*

orientation between the N- and C-terminal domains or both.

Conclusions

1. BpuJI is a dimer in solution and consists of two domains: the N-domain binds to the target sequence as a monomer, while the C-domain is responsible for nuclease activity and dimerization.
2. The catalytic domain of BpuJI possesses an end-directed nuclease activity and preferentially cuts ~3 nt from the 3'-terminus of blunt-ended DNA. This activity leads to a very variable cleavage pattern.
3. Bioinformatics and mutational analysis revealed that the C-domain of BpuJI has a PD-(D/E)XK fold and is structurally related to archaeal Holliday junction resolvases.
4. The crystal structure of the N-domain of BpuJI bound to cognate DNA showed that the BpuJI recognition domain folds into two winged-helix subdomains. The specificity-determinants are located in the N-terminal arm and the two recognition helices of the HTH motifs.
5. The BpuJI recognition domain shows structural similarity to the nicking endonuclease Nt.BspD6I. The modelling suggests that Nt.BspD6I could share the specificity-determining regions with BpuJI.

Acknowledgement

The work presented in this thesis was carried out in the Institute of Biotechnology. The financial support from BIOCEL, Max Planck Society and Lithuanian State Science and Studies Foundation is gratefully acknowledged.

I am grateful to my supervisor prof. Virginijus Šikšnys for invaluable suggestions, discussions and help with the preparation of the doctoral thesis. I thank dr. Saulius Gražulis for crystallography lessons, scripts and help in the structure solution. I am obliged to my former supervisor dr. Arūnas Lagunavičius for guiding me in the biochemical studies and teaching to work in the laboratory.

My thanks to all the colleagues in the Laboratory of Protein-DNA Interactions, especially to dr. Giedrė Tamulaitienė, dr. Gintautas Tamulaitis, dr. Mindaugas Zaremba and dr. Giedrius Sasnauskas, for advice and discussions. I thank prof. Gervydas Dienys for the encouragement to start doctoral studies and continuous support.

I would like to express my appreciation to dr. Kornelijus Stankevičius for the cloning of the BpuJI R-M system, dr. Česlovas Venclovas for bioinformatics analysis and prof. Claus Urbanke (Hannover Medizinische Hochshule) for analytical ultracentrifugation. I thank UAB Fermentas for the BpuJI R-M system clone and access to some of laboratory equipment.

I would also like to thank dr. Donatas Vaitkevičius and dr. Zita Manelienė (UAB Fermentas) initial experiments with BpuJI, dr. Will Mawby (University of Bristol) for N-terminal sequencing, dr. Mattias Bochtler (International Institute of Molecular and Cell Biology, Warsaw) for initial testing of the crystals, dr. Manfred Weiss (EMBL Hamburg) for help with operating the X12 beamline and dr. Gražvydas Lukinavičius for mass-spectrometry.

Many thanks to my family and friends for their concern and support.

List of publications

The thesis is based on the following publications:

1. Sukackaite, R., Lagunavicius, A., Stankevicius, K., Urbanke, C., Venclovas, C. and Siksnys, V. (2007) Restriction endonuclease BpuJI specific for the 5'-CCCGT sequence is related to the archaeal Holliday junction resolvase family. Nucleic Acids Res, 35, 2377-2389.

2. Sukackaite R., Grazulis S., Bochtler, M., Siksnys, V. (2008) The recognition domain of the BpuJI restriction endonuclease in complex with cognate DNA at 1.3 Å resolution. J. Mol. Biol, 378, 1084–1093.

Summary

Type IIS restriction endonucleases recognize asymmetric DNA sequences and cleave both DNA strands at fixed positions downstream of the recognition site. Restriction endonuclease BpuJI recognizes the asymmetric sequence 5'-CCCGT, but it cuts at multiple sites in the vicinity of the target sequence. This study shows that BpuJI is a dimer in solution and consists of two separate domains. The N-terminal domain binds to the target sequence as a monomer, while the C-terminal domain is responsible for nuclease activity and dimerization. The BpuJI dimer has two DNA binding surfaces provided by the N-terminal domains and displays optimal catalytic activity when bound to two copies of the recognition site. The isolated BpuJI catalytic domain possesses an end-directed nuclease activity and preferentially cuts ~3 nt from the 3'-terminus of blunt-ended DNA. The nuclease activity of the C-domain is repressed in the apo-enzyme and becomes activated upon specific DNA binding by the recognition domains. The activated catalytic domain can cleave DNA at variable position near the target site. A linear DNA resulting from BpuJI cleavage at the recognition site may be further subjected to the 3'-end directed endonucleolytic cleavage by the activated C-domain. This leads to a complicated pattern of specific DNA cleavage in the vicinity of the target site. In addition, another DNA molecule (even lacking the BpuJI target site) present in the solution can also be subjected to the end-directed nuclease activity of the activated C-terminal domain.

Bioinformatics and mutational analysis revealed that the BpuJI catalytic domain harbours a PD-(D/E)XK active site and is structurally related to archaeal Holliday junction resolvases. Bioinformatics analysis identified a domain similar to the BpuJI C-domain in functionally uncharacterized proteins that could be Res subunits of Type III enzymes and metyl-directed restriction endonucleases. The crystal structure of the BpuJI recognition domain bound to

cognate DNA was solved at 1.3 Å resolution. It revealed two "winged-helix" subdomains, D1 and D2. The recognition of the 5'-CCCGT target sequence is achieved by BpuJI through the major groove contacts of the amino acid residues located on both the HTH motifs and the N-terminal arm. The role of these interactions in the DNA recognition is also corroborated by mutational analysis. The BpuJI DNA recognition domain is most similar to Nt.BspD6I, a large subunit of the hetedimeric restriction endonuclease. The modelling suggests that Nt.BspD6I, which was solved without DNA, could share the specificity-determining regions with BpuJI. The recognition domains of BpuJI and Nt.BspD6I also share some structural similarity with the archetypal Type IIS enzyme FokI. However, the orientation of the D1 and D2 subdomains in the BpuJI and FokI protein-DNA complexes is different

References

1 Pingoud A. and Jeltsch A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res* **29**, 3705-3727.

2 Roberts R.J., Vincze T., Posfai J. and Macelis D. (2007) REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* **35**, p. D269-70.

3 Orlowski J. and Bujnicki J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res* **36**, 3552-3569.

4 Roberts R.J., Belfort M., Bestor T., Bhagwat A.S., Bickle T.A., Bitinaite J., Blumenthal R.M., Degtyarev S.K., Dryden D.T.F., Dybvig K., Firman K., Gromova E.S., Gumport R.I., Halford S.E., Hattman S., Heitman J., Hornby D.P., Janulaitis A., Jeltsch A., Josephsen J., Kiss A., Klaenhammer T.R., Kobayashi I., Kong H., Krüger D.H., Lacks S., Marinus M.G., Miyahara M., Morgan R.D., Murray N.E., Nagaraja V., Piekarowicz A., Pingoud A., Raleigh E., Rao D.N., Reich N., Repin V.E., Selker E.U., Shaw P., Stein D.C., Stoddard B.L., Szybalski W., Trautner T.A., Van Etten J.L., Vitor J.M.B., Wilson G.G. and Xu S. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**, 1805-1812.

5 Pingoud A., Fuxreiter M., Pingoud V. and Wende W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* **62**, 685-707.

6 Anderson J. (1993) Restriction endonucleases and modification methylases *Curr. Opin. Struct. Biol.* **3**, 24-30.

7 Li L., Wu L.P. and Chandrasegaran S. (1992) Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A* **89**, 4275-4279.

8 Wah D.A., Hirsch J.A., Dorner L.F., Schildkraut I. and Aggarwal A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100.

9 Zaremba M., Urbanke C., Halford S.E. and 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.

10 Grazulis S., Manakova E., Roessle M., Bochtler M., Tamulaitiene G., Huber R. and Siksnys V. (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.

11 Sasnauskas G., Halford S.E. and 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.

12 Cesnaviciene E., Petrusyte M., Kazlauskiene R., Maneliene Z., Timinskas A., Lubys A. and Janulaitis A. (2001) Characterization of AloI, a restriction-modification system of a new type. *J Mol Biol* **314**, 205-216.

13 Bujnicki J.M. Molecular phylogenetics of restriction endonucleases. In *Restriction endonucleases*. Pingoud A. (Ed.). 2004. pp. 63-94.

14 Murzin A.G., Brenner S.E., Hubbard T. and Chothia C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536-540.

15 Tamulaitiene G., Jakubauskas A., Urbanke C., Huber R., Grazulis S. and Siksnys V. (2006) The crystal structure of the rare-cutting restriction enzyme SdaI reveals unexpected domain architecture. *Structure* **14**, 1389-1400.
16 Ban C. and Yang W. (1998) Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J* **17**, 1526-1534.
17 Nishino T., Ishino Y. and Morikawa K. (2006) Structure-specific DNA nucleases: structural basis for 3D-scissors. *Curr Opin Struct Biol* **16**, 60-67.
18 Kovall R.A. and Matthews B.W. (1998) Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases. *Proc Natl Acad Sci U S A* **95**, 7893-7897.
19 Singleton M.R., Dillingham M.S., Gaudier M., Kowalczykowski S.C. and Wigley D.B. (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**, 187-193.
20 Kim Y.C., Grable J.C., Love R., Greene P.J. and Rosenberg J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307-1309.
21 Winkler F.K., Banner D.W., Oefner C., Tsernoglou D., Brown R.S., Heathman S.P., Bryan R.K., Martin P.D., Petratos K. and Wilson K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J* **12**, 1781-1795.
22 Venclovas C., Timinskas A. and 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.
23 DeLano W. (2002) The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA.
24 Niv M.Y., Ripoll D.R., Vila J.A., Liwo A., Vanamee E.S., Aggarwal A.K., Weinstein H. and Scheraga H.A. (2007) Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res* **35**, 2227-2237.
25 Horton J.R., Blumenthal R.M., Cheng X. Restriction endonucleases: Structure of the conserved catalytic core and the role of metal ions in DNA cleavage. In *Restriction endonucleases*. A. Pingoud (Ed.). 2004. pp. 361-392.
26 Newman M., Lunnen K., Wilson G., Greci J., Schildkraut I. and Phillips S.E. (1998) Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J* **17**, 5466-5476.
27 van der Woerd M.J., Pelletier J.J., Xu S. and Friedman A.M. (2001) Restriction enzyme BsoBI-DNA complex: a tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure* **9**, 133-144.
28 Newman M., Strzelecka T., Dorner L.F., Schildkraut I. and Aggarwal A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269**, 656-663.
29 Nishino T., Komori K., Ishino Y. and Morikawa K. (2003) X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure* **11**, 445-457.
30 Xu Q.S., Kucera R.B., Roberts R.J. and Guo H. (2004) An asymmetric complex of restriction endonuclease MspI on its palindromic DNA recognition site. *Structure* **12**, 1741-1747.
31 Yang Z., Horton J.R., Maunus R., Wilson G.G., Roberts R.J. and Cheng X.

(2005) Structure of HinP1I endonuclease reveals a striking similarity to the monomeric restriction enzyme MspI. *Nucleic Acids Res* **33**, 1892-1901.
32 Bozic D., Grazulis S., Siksnys V. and Huber R. (1996) Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 A resolution. *J Mol Biol* **255**, 176-186.
33 Deibert M., Grazulis S., Sasnauskas G., Siksnys V. and Huber R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat Struct Biol* **7**, 792-799.
34 Grazulis S., Deibert M., Rimseliene R., Skirgaila R., Sasnauskas G., Lagunavicius A., Repin V., Urbanke C., Huber R. and Siksnys V. (2002) Crystal structure of the Bse634I restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res* **30**, 876-885.
35 Bochtler M., Szczepanowski R.H., Tamulaitis G., Grazulis S., Czapinska H., Manakova E. and Siksnys V. (2006) Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J* **25**, 2219-2229.
36 Dunten P.W., Little E.J., Gregory M.T., Manohar V.M., Dalton M., Hough D., Bitinaite J. and Horton N.C. (2008) The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res* **36**, 5405-5416.
37 Skirgaila R., Grazulis S., Bozic D., Huber R. and 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.
38 Kosinski J., Feder M. and Bujnicki J.M. (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**, p. 172.
39 Aravind L., Anantharaman V., Balaji S., Babu M.M. and Iyer L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**, 231-262.
40 Tucker-Kellogg L., Rould M.A., Chambers K.A., Ades S.E., Sauer R.T. and Pabo C.O. (1997) Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 A resolution: structural basis for enhanced affinity and altered specificity. *Structure* **5**, 1047-1054.
41 Cook W.J., Kar S.R., Taylor K.B. and Hall L.M. (1998) Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J Mol Biol* **275**, 337-346.
42 Albright R.A. and Matthews B.W. (1998) Crystal structure of lambda-Cro bound to a consensus operator at 3.0 A resolution. *J Mol Biol* **280**, 137-151.
43 Xu H.E., Rould M.A., Xu W., Epstein J.A., Maas R.L. and Pabo C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* **13**, 1263-1275.
44 Gajiwala K.S. and Burley S.K. (2000) Winged helix proteins. *Curr Opin Struct Biol* **10**, 110-116.
45 Wintjens R. and Rooman M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol* **262**, 294-313.
46 Suzuki M., Brenner S.E., Gerstein M. and Yagi N. (1995) DNA recognition code of transcription factors. *Protein Eng* **8**, 319-328.

47 Kuzminov A. (1999) Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**, p. 751-813, table of contents.
48 Kovall R. and Matthews B.W. (1997) Toroidal structure of lambda-exonuclease. *Science* **277**, 1824-1827.
49 Subramanian K., Rutvisuttinunt W., Scott W. and Myers R.S. (2003) The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Res* **31**, 1585-1596.
50 Matsuura S., Komatsu J., Hirano K., Yasuda H., Takashima K., Katsura S. and Mizuno A. (2001) Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field. *Nucleic Acids Res* **29**, p. E79.
51 van Oijen A.M., Blainey P.C., Crampton D.J., Richardson C.C., Ellenberger T. and Xie X.S. (2003) Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235-1238.
52 Bickle T.A. and Krüger D.H. (1993) Biology of DNA restriction. *Microbiol Rev* **57**, 434-450.
53 Arber W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* **24**, 1-7.
54 Jeltsch A. (2003) Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene* **317**, 13-16.
55 Kobayashi I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* **29**, 3742-3756.
56 Miyazono K., Watanabe M., Kosinski J., Ishikawa K., Kamo M., Sawasaki T., Nagata K., Bujnicki J.M., Endo Y., Tanokura M. and Kobayashi I. (2007) Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res* **35**, 1908-1918.
57 Lukacs C.M., Kucera R., Schildkraut I. and Aggarwal A.K. (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.
58 Townson S.A., Samuelson J.C., Vanamee E.S., Edwards T.A., Escalante C.R., Xu S. and Aggarwal A.K. (2004) Crystal structure of BstYI at 1.85A resolution: a thermophilic restriction endonuclease with overlapping specificities to BamHI and BglII. *J Mol Biol* **338**, 725-733.
59 Hashimoto H., Shimizu T., Imasaki T., Kato M., Shichijo N., Kita K. and Sato M. (2005) Crystal structures of type II restriction endonuclease EcoO109I and its complex with cognate DNA. *J Biol Chem* **280**, 5605-5610.
60 Deibert M., Grazulis S., Janulaitis A., Siksnys V. and Huber R. (1999) Crystal structure of MunI restriction endonuclease in complex with cognate DNA at 1.7 A resolution. *EMBO J* **18**, 5805-5816.
61 Cheng X., Balendiran K., Schildkraut I. and Anderson J.E. (1995) Crystal structure of the PvuII restriction endonuclease. *Gene* **157**, 139-140.
62 Szczepanowski R.H., Carpenter M.A., Czapinska H., Zaremba M., Tamulaitis G., Siksnys V., Bhagwat A.S. and Bochtler M. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res* **36**, 6109-6117.
63 Vanamee E.S., Viadiu H., Kucera R., Dorner L., Picone S., Schildkraut I. and Aggarwal A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J* **24**, 4198-4208.

64 Nobbs T.J. and Halford S.E. (1995) DNA cleavage at two recognition sites by the SfiI restriction endonuclease: salt dependence of cis and trans interactions between distant DNA sites. *J Mol Biol* **252**, 399-411.
65 Embleton M.L., Siksnys V. and Halford S.E. (2001) DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. *J Mol Biol* **311**, 503-514.
66 Zaremba M., Sasnauskas G., Urbanke C. and Siksnys V. (2005) Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: oligomeric structure-stability-function correlations. *J Mol Biol* **348**, 459-478.
67 Daniels L.E., Wood K.M., Scott D.J. and Halford S.E. (2003) Subunit assembly for DNA cleavage by restriction endonuclease SgrAI. *J Mol Biol* **327**, 579-591.
68 Tamulaitis G., Solonin A.S. and Siksnys V. (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett* **518**, 17-22.
69 Bilcock D.T., Daniels L.E., Bath A.J. and Halford S.E. (1999) Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J Biol Chem* **274**, 36379-36386.
70 Horton J.R., Zhang X., Maunus R., Yang Z., Wilson G.G., Roberts R.J. and Cheng X. (2006) DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res* **34**, 939-948.
71 Kaus-Drobek M., Czapinska H., Sokołowska M., Tamulaitis G., Szczepanowski R.H., Urbanke C., Siksnys V. and Bochtler M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res* **35**, 2035-2046.
72 Sokolowska M., Kaus-Drobek M., Czapinska H., Tamulaitis G., Szczepanowski R.H., Urbanke C., Siksnys V. and Bochtler M. (2007) Monomeric restriction endonuclease BcnI in the apo form and in an asymmetric complex with target DNA. *J Mol Biol* **369**, 722-734.
73 Alves J. and Vennekol, P. Protein engineering of restriction enzymes. In *Restriction endonucleases*. A. Pingoud (Ed.). 2004. pp. 393-413.
74 Galburt E.A. and Stoddard B.L. (2000) Restriction endonucleases: one of these things is not like the others. *Nat Struct Biol* **7**, 89-91.
75 Pingoud V., Conzelmann C., Kinzebach S., Sudina A., Metelev V., Kubareva E., Bujnicki J.M., Lurz R., Lüder G., Xu S. and Pingoud A. (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.
76 Horton N.C., Dorner L.F. and Perona J.J. (2002) Sequence selectivity and degeneracy of a restriction endonuclease mediated by DNA intercalation. *Nat Struct Biol* **9**, 42-47.
77 Lambert A.R., Sussman D., Shen B., Maunus R., Nix J., Samuelson J., Xu S. and Stoddard B.L. (2008) Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure* **16**, 558-569.
78 Huai Q., Colandene J.D., Chen Y., Luo F., Zhao Y., Topal M.D. and Ke H. (2000) Crystal structure of NaeI-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J* **19**, 3110-3118.
79 Huai Q., Colandene J.D., Topal M.D. and Ke H. (2001) Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat Struct Biol* **8**, 665-669.

80 Zhou X.E., Wang Y., Reuter M., Mücke M., Krüger D.H., Meehan E.J. and Chen L. (2004) Crystal structure of type IIE restriction endonuclease EcoRII reveals an autoinhibition mechanism by a novel effector-binding fold. *J Mol Biol* **335**, 307-319.
81 Tamulaitis G., Mucke M. and Siksnys V. (2006) Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett* **580**, 1665-1671.
82 Tamulaitis G., Sasnauskas G., Mucke M. and 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.
83 Shlyakhtenko L.S., Gilmore J., Portillo A., Tamulaitis G., Siksnys V. and Lyubchenko Y.L. (2007) Direct visualization of the EcoRII-DNA triple synaptic complex by atomic force microscopy. *Biochemistry* **46**, 11128-11136.
84 Friedhoff P., Lurz R., Lüder G. and 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.
85 Xu C., Yu F., Xu S., Ding Y., Sun L., Tang L., Hu X., Zhang Z. and He J. (2009) Crystal structure and function of C-terminal Sau3AI domain. *Biochim Biophys Acta* **1794**, 118-123.
86 Kaczorowski T., Skowron P. and Podhajska A.J. (1989) Purification and characterization of the FokI restriction endonuclease. *Gene* **80**, 209-216.
87 Bitinaite J., Wah D.A., Aggarwal A.K. and Schildkraut I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10570-10575.
88 Vanamee E.S., Santagata S. and Aggarwal A.K. (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol* **309**, 69-78.
89 Wah D.A., Bitinaite J., Schildkraut I. and Aggarwal A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10564-10569.
90 Catto L.E., Ganguly S., Milsom S.E., Welsh A.J. and Halford S.E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* **34**, 1711-1720.
91 Catto L.E., Bellamy S.R.W., Retter S.E. and Halford S.E. (2008) Dynamics and consequences of DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* **36**, 2073-2081.
92 Kim Y.G. and Chandrasegaran S. (1994) Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A* **91**, 883-887.
93 Kim Y.G., Cha J. and Chandrasegaran S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-1160.
94 Kim Y.G., Smith J., Durgesha M. and Chandrasegaran S. (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem* **379**, 489-495.
95 Durai S., Mani M., Kandavelou K., Wu J., Porteus M.H. and Chandrasegaran S. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* **33**, 5978-5990.
96 Wu J., Kandavelou K. and Chandrasegaran S. (2007) Custom-designed zinc finger nucleases: what is next? *Cell Mol Life Sci* **64**, 2933-2944.
97 Carroll D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther* **15**, 1463-1468.

98 Papworth M., Kolasinska P. and Minczuk M. (2006) Designer zinc-finger proteins and their applications. *Gene* **366**, 27-38.
99 Kachalova G.S., Rogulin E.A., Yunusova A.K., Artyukh R.I., Perevyazova T.A., Matvienko N.I., Zheleznaya L.A. and Bartunik H.D. (2008) Structural analysis of the heterodimeric type IIS restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit. *J Mol Biol* **384**, 489-502.
100 Zheleznaya L.A., Perevyazova T.A., Alzhanova D.V. and Matvienko N.I. (2001) Site-specific nickase from bacillus species strain d6. *Biochemistry (Mosc)* **66**, 989-993.
101 Yunusova A.K., Rogulin E.A., Artyukh R.I., Zheleznaya L.A. and Matvienko N.I. (2006) Nickase and a protein encoded by an open reading frame downstream from the nickase BspD6I gene form a restriction endonuclease complex. *Biochemistry (Mosc)* **71**, 815-820.
102 Higgins L.S., Besnier C. and Kong H. (2001) The nicking endonuclease N.BstNBI is closely related to type IIs restriction endonucleases MlyI and PleI. *Nucleic Acids Res* **29**, 2492-2501.
103 Besnier C.E. and Kong H. (2001) Converting MlyI endonuclease into a nicking enzyme by changing its oligomerization state. *EMBO Rep* **2**, 782-786.
104 Armalyte E., Bujnicki J.M., Giedriene J., Gasiunas G., Kosiński J. and Lubys A. (2005) Mva1269I: a monomeric type IIS restriction endonuclease from Micrococcus varians with two EcoRI- and FokI-like catalytic domains. *J Biol Chem* **280**, 41584-41594.
105 Kong H., Morgan R.D., Maunus R.E. and Schildkraut I. (1993) A unique restriction endonuclease, BcgI, from Bacillus coagulans. *Nucleic Acids Res* **21**, 987-991.
106 Kong H., Roemer S.E., Waite-Rees P.A., Benner J.S., Wilson G.G. and Nwankwo D.O. (1994) Characterization of BcgI, a new kind of restriction-modification system. *J Biol Chem* **269**, 683-690.
107 Kong H. and Smith C.L. (1997) Substrate DNA and cofactor regulate the activities of a multi-functional restriction-modification enzyme, BcgI. *Nucleic Acids Res* **25**, 3687-3692.
108 Kong H. (1998) Analyzing the functional organization of a novel restriction modification system, the BcgI system. *J Mol Biol* **279**, 823-832.
109 Kong H. and Smith C.L. (1998) Does BcgI, a unique restriction endonuclease, require two recognition sites for cleavage? *Biol Chem* **379**, 605-609.
110 Marshall J.J.T., Gowers D.M. and Halford S.E. (2007) Restriction endonucleases that bridge and excise two recognition sites from DNA. *J Mol Biol* **367**, 419-431.
111 Janulaitis A., Vaisvila R., Timinskas A., Klimasauskas S. and Butkus V. (1992) Cloning and sequence analysis of the genes coding for Eco57I type IV restriction-modification enzymes. *Nucleic Acids Res* **20**, 6051-6056.
112 Janulaitis A., Petrusyte M., Maneliene Z., Klimasauskas S. and Butkus V. (1992) Purification and properties of the Eco57I restriction endonuclease and methylase--prototypes of a new class (type IV). *Nucleic Acids Res* **20**, 6043-6049.
113 Rimseliene R. and Janulaitis A. (2001) Mutational analysis of two putative catalytic motifs of the type IV restriction endonuclease Eco57I. *J Biol Chem* **276**, 10492-10497.

114 Jurenaite-Urbanaviciene S., Serksnaite J., Kriukiene E., Giedriene J., Venclovas C. and Lubys A. (2007) Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc Natl Acad Sci U S A* **104**, 10358-10363.
115 Rimseliene R., Maneliene Z., Lubys A. and Janulaitis A. (2003) Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *J Mol Biol* **327**, 383-391.
116 Bhagwat A.S. and Lieb M. (2002) Cooperation and competition in mismatch repair: very short-patch repair and methyl-directed mismatch repair in Escherichia coli. *Mol Microbiol* **44**, 1421-1428.
117 Tsutakawa S.E., Muto T., Kawate T., Jingami H., Kunishima N., Ariyoshi M., Kohda D., Nakagawa M. and Morikawa K. (1999) Crystallographic and functional studies of very short patch repair endonuclease. *Mol Cell* **3**, 621-628.
118 Modrich P. (1991) Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* **25**, 229-253.
119 Iyer R.R., Pluciennik A., Burdett V. and Modrich P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev* **106**, 302-323.
120 Barras F. and Marinus M.G. (1989) The great GATC: DNA methylation in E. coli. *Trends Genet* **5**, 139-143.
121 Lee J.Y., Chang J., Joseph N., Ghirlando R., Rao D.N. and Yang W. (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Mol Cell* **20**, 155-166.
122 Hennecke F., Kolmar H., Bründl K. and Fritz H.J. (1991) The vsr gene product of E. coli K-12 is a strand- and sequence-specific DNA mismatch endonuclease. *Nature* **353**, 776-778.
123 Dzidic S. and Radman M. (1989) Genetic requirements for hyper-recombination by very short patch mismatch repair: involvement of Escherichia coli DNA polymerase I. *Mol Gen Genet* **217**, 254-256.
124 Tsutakawa S.E., Jingami H. and Morikawa K. (1999) Recognition of a TG mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell* **99**, 615-623.
125 Tsutakawa S.E. and Morikawa K. (2001) The structural basis of damaged DNA recognition and endonucleolytic cleavage for very short patch repair endonuclease. *Nucleic Acids Res* **29**, 3775-3783.
126 Sarnovsky R.J., May E.W. and Craig N.L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J* **15**, 6348-6361.
127 Hickman A.B., Li Y., Mathew S.V., May E.W., Craig N.L. and Dyda F. (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol Cell* **5**, 1025-1034.
128 Stoddard B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49-95.
129 Zhao L., Bonocora R.P., Shub D.A. and Stoddard B.L. (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J* **26**, 2432-2442.
130 Orlowski J., Boniecki M. and Bujnicki J.M. (2007) I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA

recognition. *Bioinformatics* **23**, 527-530.
131 Bonocora R.P. and Shub D.A. (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol* **39**, 1299-1306.
132 Kowalczykowski S.C., Dixon D.A., Eggleston A.K., Lauder S.D. and Rehrauer W.M. (1994) Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev* **58**, 401-465.
133 Liu Y. and West S.C. (2004) Happy Hollidays: 40th anniversary of the Holliday junction. *Nat Rev Mol Cell Biol* **5**, 937-944.
134 Lilley D.M. and White M.F. (2001) The junction-resolving enzymes. *Nat Rev Mol Cell Biol* **2**, 433-443.
135 Aravind L., Makarova K.S. and Koonin E.V. (2000) SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res* **28**, 3417-3432.
136 Hadden J.M., Convery M.A., Déclais A.C., Lilley D.M. and Phillips S.E. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. *Nat Struct Biol* **8**, 62-67.
137 Nishino T., Komori K., Tsuchiya D., Ishino Y. and Morikawa K. (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure* **9**, 197-204.
138 Bond C.S., Kvaratskhelia M., Richard D., White M.F. and Hunter W.N. (2001) Structure of Hjc, a Holliday junction resolvase, from Sulfolobus solfataricus. *Proc Natl Acad Sci U S A* **98**, 5509-5514.
139 Middleton C.L., Parker J.L., Richard D.J., White M.F. and Bond C.S. (2004) Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic Acids Res* **32**, 5442-5451.
140 McGregor N., Ayora S., Sedelnikova S., Carrasco B., Alonso J.C., Thaw P. and Rafferty J. (2005) The structure of Bacillus subtilis RecU Holliday junction resolvase and its role in substrate selection and sequence-specific cleavage. *Structure* **13**, 1341-1351.
141 Déclais A.C., Hadden J., Phillips S.E. and Lilley D.M. (2001) The active site of the junction-resolving enzyme T7 endonuclease I. *J Mol Biol* **307**, 1145-1158.
142 Ayora S., Carrasco B., Doncel-Perez E., Lurz R. and Alonso J.C. (2004) Bacillus subtilis RecU protein cleaves Holliday junctions and anneals single-stranded DNA. *Proc Natl Acad Sci U S A* **101**, 452-457.
143 Hadden J.M., Déclais A., Carr S.B., Lilley D.M.J. and Phillips S.E.V. (2007) The structural basis of Holliday junction resolution by T7 endonuclease I. *Nature* **449**, 621-624.
144 Déclais A. and Lilley D.M. (2008) New insight into the recognition of branched DNA structure by junction-resolving enzymes. *Curr Opin Struct Biol* **18**, 86-95.
145 Ciccia A., McDonald N. and West S.C. (2008) Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem* **77**, 259-287.
146 Doherty A.J., Serpell L.C. and Ponting C.P. (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res* **24**, 2488-2497.
147 Sijbers A.M., de Laat W.L., Ariza R.R., Biggerstaff M., Wei Y.F., Moggs J.G., Carter K.C., Shell B.K., Evans E., de Jong M.C., Rademakers S., de Rooij J., Jaspers

N.G., Hoeijmakers J.H. and Wood R.D. (1996) Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**, 811-822.
148 Biertümpfel C., Yang W. and Suck D. (2007) Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **449**, 616-620.
149 Komori K., Hidaka M., Horiuchi T., Fujikane R., Shinagawa H. and Ishino Y. (2004) Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *J Biol Chem* **279**, 53175-53185.
150 Nishino T., Komori K., Ishino Y. and Morikawa K. (2005) Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure* **13**, 1183-1192.
151 Newman M., Murray-Rust J., Lally J., Rudolf J., Fadden A., Knowles P.P., White M.F. and McDonald N.Q. (2005) Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J* **24**, 895-905.
152 Van Duyne G.D. (2005) Bending and cutting forks and flaps. *Structure* **13**, 1092-1093.
153 Komori K., Fujikane R., Shinagawa H. and Ishino Y. (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* **77**, 227-241.
154 Roberts J.A. and White M.F. (2005) An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. *J Biol Chem* **280**, 5924-5928.
155 Chang J.H., Kim J.J., Choi J.M., Lee J.H. and Cho Y. (2008) Crystal structure of the Mus81-Eme1 complex. *Genes Dev* **22**, 1093-1106.
156 Tsodikov O.V., Enzlin J.H., Schärer O.D. and Ellenberger T. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl Acad Sci U S A* **102**, 11236-11241.
157 Tripsianes K., Folkers G., Ab E., Das D., Odijk H., Jaspers N.G.J., Hoeijmakers J.H.J., Kaptein R. and Boelens R. (2005) The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure* **13**, 1849-1858.
158 Murray N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**, 412-434.
159 Dryden D.T., Murray N.E. and Rao D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res* **29**, 3728-3741.
160 Bourniquel A.A. and Bickle T.A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* **84**, 1047-1059.
161 Kim J., DeGiovanni A., Jancarik J., Adams P.D., Yokota H., Kim R. and Kim S. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc Natl Acad Sci U S A* **102**, 3248-3253.
162 Calisto B.M., Pich O.Q., Piñol J., Fita I., Querol E. and Carpena X. (2005) Crystal structure of a putative type I restriction-modification S subunit from Mycoplasma genitalium. *J Mol Biol* **351**, 749-762.
163 Kennaway C.K., Obarska-Kosinska A., White J.H., Tuszynska I., Cooper L.P., Bujnicki J.M., Trinick J. and Dryden D.T.F. (2008) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res* **37**, 762-70.
164 Davies G.P., Martin I., Sturrock S.S., Cronshaw A., Murray N.E. and Dryden

D.T. (1999) On the structure and operation of type I DNA restriction enzymes. *J Mol Biol* **290**, 565-579.
165 Lapkouski M., Panjikar S., Janscak P., Smatanova I.K., Carey J., Ettrich R. and Csefalvay E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat Struct Mol Biol* **16**, 94-95.
166 Singleton M.R., Dillingham M.S. and Wigley D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**, 23-50.
167 Stanley L.K., Seidel R., van der Scheer C., Dekker N.H., Szczelkun M.D. and Dekker C. (2006) When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *EMBO J* **25**, 2230-2239.
168 Seidel R., Bloom J.G.P., Dekker C. and Szczelkun M.D. (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J* **27**, 1388-1398.
169 van Noort J., van der Heijden T., Dutta C.F., Firman K. and Dekker C. (2004) Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion. *Nucleic Acids Res* **32**, 6540-6547.
170 Janscak P., MacWilliams M.P., Sandmeier U., Nagaraja V. and Bickle T.A. (1999) DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J* **18**, 2638-2647.
171 Berge T., Ellis D.J., Dryden D.T., Edwardson J.M. and Henderson R.M. (2000) Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys J* **79**, 479-484.
172 Jindrova E., Schmid-Nuoffer S., Hamburger F., Janscak P. and Bickle T.A. (2005) On the DNA cleavage mechanism of Type I restriction enzymes. *Nucleic Acids Res* **33**, 1760-1766.
173 Janscak P., Sandmeier U., Szczelkun M.D. and Bickle T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J Mol Biol* **306**, 417-431.
174 Ahmad I. and Rao D.N. (1996) Functional analysis of conserved motifs in EcoP15I DNA methyltransferase. *J Mol Biol* **259**, 229-240.
175 Wagenführ K., Pieper S., Mackeldanz P., Linscheid M., Krüger D.H. and Reuter M. (2007) Structural domains in the type III restriction endonuclease EcoP15I: characterization by limited proteolysis, mass spectrometry and insertional mutagenesis. *J Mol Biol* **366**, 93-102.
176 Meisel A., Bickle T.A., Krüger D.H. and Schroeder C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467-469.
177 Meisel A., Mackeldanz P., Bickle T.A., Krüger D.H. and Schroeder C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J* **14**, 2958-2966.
178 Crampton N., Yokokawa M., Dryden D.T.F., Edwardson J.M., Rao D.N., Takeyasu K., Yoshimura S.H. and Henderson R.M. (2007) Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoP15I is capable of DNA translocation and looping. *Proc Natl Acad Sci U S A* **104**, 12755-12760.
179 Crampton N., Roes S., Dryden D.T.F., Rao D.N., Edwardson J.M. and Henderson R.M. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J* **26**, 3815-3825.
180 Sutherland E., Coe L. and Raleigh E.A. (1992) McrBC: a multisubunit GTP-

dependent restriction endonuclease. *J Mol Biol* **225**, 327-348.
181 Pieper U., Schweitzer T., Groll D.H. and Pingoud A. (1999) Defining the location and function of domains of McrB by deletion mutagenesis. *Biol Chem* **380**, 1225-1230.
182 Pieper U. and Pingoud A. (2002) A mutational analysis of the PD...D/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from Escherichia coli. *Biochemistry* **41**, 5236-5244.
183 Neuwald A.F., Aravind L., Spouge J.L. and Koonin E.V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27-43.
184 Erzberger J.P. and Berger J.M. (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* **35**, 93-114.
185 Panne D., Müller S.A., Wirtz S., Engel A. and Bickle T.A. (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J* **20**, 3210-3217.
186 Dillingham M.S. and Kowalczykowski S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* **72**, p. 642-71, Table of Contents.
187 Kowalczykowski S.C. (2000) Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* **25**, 156-165.
188 Dixon D.A. and Kowalczykowski S.C. (1993) The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell* **73**, 87-96.
189 Taylor A.F. and Smith G.R. (1995) Monomeric RecBCD enzyme binds and unwinds DNA. *J Biol Chem* **270**, 24451-24458.
190 Dillingham M.S., Spies M. and Kowalczykowski S.C. (2003) RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**, 893-897.
191 Yu M., Souaya J. and Julin D.A. (1998) The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from Escherichia coli. *Proc Natl Acad Sci U S A* **95**, 981-986.
192 Wang J., Chen R. and Julin D.A. (2000) A single nuclease active site of the Escherichia coli RecBCD enzyme catalyzes single-stranded DNA degradation in both directions. *J Biol Chem* **275**, 507-513.
193 Sun J., Julin D.A. and Hu J. (2006) The nuclease domain of the Escherichia coli RecBCD enzyme catalyzes degradation of linear and circular single-stranded and double-stranded DNA. *Biochemistry* **45**, 131-140.
194 Rigden D.J. (2005) An inactivated nuclease-like domain in RecC with novel function: implications for evolution. *BMC Struct Biol* **5**, p. 9.
195 Taylor A.F. and Smith G.R. (2003) RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423**, 889-893.
196 Bianco P.R. and Kowalczykowski S.C. (1997) The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGGTGG-3'. *Proc Natl Acad Sci U S A* **94**, 6706-6711.
197 Spies M., Bianco P.R., Dillingham M.S., Handa N., Baskin R.J. and Kowalczykowski S.C. (2003) A molecular throttle: the recombination hotspot chi controls DNA translocation by the RecBCD helicase. *Cell* **114**, 647-654.
198 Spies M., Amitani I., Baskin R.J. and Kowalczykowski S.C. (2007) RecBCD

enzyme switches lead motor subunits in response to chi recognition. *Cell* **131**, 694-705.
199 Handa N., Bianco P.R., Baskin R.J. and Kowalczykowski S.C. (2005) Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after chi recognition. *Mol Cell* **17**, 745-750.
200 Churchill J.J. and Kowalczykowski S.C. (2000) Identification of the RecA protein-loading domain of RecBCD enzyme. *J Mol Biol* **297**, 537-542.
201 Spies M. and Kowalczykowski S.C. (2006) The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. *Mol Cell* **21**, 573-580.
202 Sambrook J., Fritsch E. and Maniatis T. Molecular Cloning. A Laboratory manual. Cold Spring Harbor Laboratory Press, 1989.
203 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Sfruchl, K.. Shorts protocols in molecular biology, A compedium of methods from current protocols in molecular biology, 3nd ed. John Wiley & Sons, Inc, Canada, 1995.
204 Söding J., Biegert A. and Lupas A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, p. W244-8.
205 Söding J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951-960.
206 Li W., Pio F., Pawłowski K. and Godzik A. (2000) Saturated BLAST: an automated multiple intermediate sequence search used to detect distant homology. *Bioinformatics* **16**, 1105-1110.
207 Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.
208 Pei J., Sadreyev R. and Grishin N.V. (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* **19**, 427-428.
209 Do C.B., Mahabhashyam M.S.P., Brudno M. and Batzoglou S. (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* **15**, 330-340.
210 Marchler-Bauer A., Anderson J.B., Cherukuri P.F., DeWeese-Scott C., Geer L.Y., Gwadz M., He S., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Liebert C.A., Liu C., Lu F., Marchler G.H., Mullokandov M., Shoemaker B.A., Simonyan V., Song J.S., Thiessen P.A., Yamashita R.A., Yin J.J., Zhang D. and Bryant S.H. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**, p. D192-6.
211 Ducruix A. and Giege R. Crystallization of Nucleic Acids and Proteins. A practical approach. Oxford University Press, New York, 1992.
212 Leslie A.G.W. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57.
213 Evans P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.
214 French, G.S. & Wilson, K.S. (1978) On the treatment of negative intensity observations *Acta Crystallogr.* **A34**, 517-525.
215 Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography *Acta Crystallogr.* **D50**, 760-763.

216 Cowtan K. (1994) Dm: An automated procedure for phase improvement by density modification *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.
217 Morris, R.J., Perrakis, A. & Lamzin, V.S. (2002) Arp/warp's model-building algorithms. I. the main chain *Acta Crystallogr.* **D58**, 968-975.
218 Emsley P. and Cowtan K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
219 Brünger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T. and Warren G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.
220 Murshudov G.N., Vagin A.A. and Dodson E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.
221 Laskowski R.A., Moss D.S. and Thornton J.M. (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **231**, 1049-1067.
222 Vriend G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, p. 52-6, 29.
223 Frishman D. and Argos P. (1995) Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579.
224 Holm L. and Sander C. (1996) Alignment of three-dimensional protein structures: network server for database searching. *Methods Enzymol* **266**, 653-662.
225 Lu G. (2000) TOP: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallogr.* **33**, p. 176–183.
226 Hubbard,S.J. & Thornton, J.M. (1993) 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.
227 Luscombe N.M., Laskowski R.A. and Thornton J.M. (1997) NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res* **25**, 4940-4945.
228 Kraulis P. (1991) Molscript - a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
229 Merritt E.A. and Murphy M.E. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr D Biol Crystallogr* **50**, 869-873.
230 Barik S. (1995) Site-directed mutagenesis by double polymerase chain reaction. *Mol Biotechnol* **3**, 1-7.
231 Zheng L., Baumann U. and Reymond J. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, p. e115.
232 Yoshioka K. (2002) KyPlot – a user-oriented tool for statistical data analysis and visualization. *CompStat* **17**, 425-437.
233 Siksnys V., Skirgaila R., Sasnauskas G., Urbanke C., Cherny D., Grazulis S. and Huber R. (1999) The Cfr10I restriction enzyme is functional as a tetramer. *J Mol Biol* **291**, 1105-1118.
234 Soundararajan M., Chang Z., Morgan R.D., Heslop P. and Connolly B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J Biol Chem* **277**, 887-895.
235 Bath A.J., Milsom S.E., Gormley N.A. and Halford S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J*

Biol Chem **277**, 4024-4033.
236 Lagunavicius A., Sasnauskas G., Halford S.E. and 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.
237 Oller A.R., Vanden Broek W., Conrad M. and Topal M.D. (1991) Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species. *Biochemistry* **30**, 2543-2549.
238 Wentzell L.M., Nobbs T.J. and Halford S.E. (1995) The SfiI restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J Mol Biol* **248**, 581-595.
239 Embleton M.L., Williams S.A., Watson M.A. and Halford S.E. (1999) Specificity from the synapsis of DNA elements by the Sfi I endonuclease. *J Mol Biol* **289**, 785-797.
240 Sektas M., Kaczorowski T. and Podhajska A.J. (1992) Purification and properties of the MboII, a class-IIS restriction endonuclease. *Nucleic Acids Res* **20**, 433-438.
241 Tucholski J., Skowron P.M. and Podhajska A.J. (1995) MmeI, a class-IIS restriction endonuclease: purification and characterization. *Gene* **157**, 87-92.
242 Kleid D.G. (1980) Purification and properties of the HphI endonuclease. *Methods Enzymol* **65**, 163-166.
243 Smith J., Bibikova M., Whitby F.G., Reddy A.R., Chandrasegaran S. and Carroll D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**, 3361-3369.
244 Minczuk M., Papworth M.A., Miller J.C., Murphy M.P. and Klug A. (2008) Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic Acids Res* **36**, 3926-3938.
245 Campbell E.A., Muzzin O., Chlenov M., Sun J.L., Olson C.A., Weinman O., Trester-Zedlitz M.L. and Darst S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**, 527-539.
246 Jacobson E.M., Li P., Leon-del-Rio A., Rosenfeld M.G. and Aggarwal A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev* **11**, 198-212.
247 Watkins S., van Pouderoyen G. and Sixma T.K. (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res* **32**, 4306-4312.