doi:10.1016/j.jmb.2008.03.041 J. Mol. Biol. (2008) 378, 1084–1093

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Recognition Domain of the BpuJI Restriction Endonuclease in Complex with Cognate DNA at 1.3-Å Resolution

Rasa Sukackaite[1], Saulius Grazulis[1], Matthias Bochtler[2,3,4] and Virginijus Siksnys[1]*

[1]*Institute of Biotechnology, Graiciuno 8, 02241 Vilnius, Lithuania*

[2]*International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland*

[3]*Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01309 Dresden, Germany*

[4]*Schools of Chemistry and Biosciences, Cardiff University, Park Place, Cardiff CF10 3AT, Wales, UK*

Received 8 February 2008; received in revised form 17 March 2008; accepted 19 March 2008
Available online 28 March 2008

Edited by K. Morikawa

Type IIS restriction endonucleases recognize asymmetric DNA sequences and cleave both DNA strands at fixed positions downstream of the recognition site. The restriction endonuclease BpuJI recognizes the asymmetric sequence 5′-CCCGT; however, it cuts at multiple sites in the vicinity of the target sequence. BpuJI consists of two physically separate domains, with catalytic and dimerization functions in the C-terminal domain and DNA recognition functions in the N-terminal domain. Here we report the crystal structure of the BpuJI recognition domain bound to cognate DNA at 1.3-Å resolution. This region folds into two winged-helix subdomains, D1 and D2, interspaced by the DL subdomain. The D1 and D2 subdomains of BpuJI share structural similarity with the similar subdomains of the FokI DNA-binding domain; however, their orientations in protein–DNA complexes are different. Recognition of the 5′-CCCGT target sequence is achieved by BpuJI through the major groove contacts of amino acid residues located on both the helix–turn–helix motifs and the N-terminal arm. The role of these interactions in DNA recognition is also corroborated by mutational analysis.

Keywords: restriction endonuclease; BpuJI; DNA recognition; helix–turn–helix; crystal structure

Introduction

Type II restriction endonucleases recognize 4- to 8-bp DNA sequences and cleave both strands of the DNA at fixed positions.[1] They follow many different modes of action and are categorized into 11 overlapping subtypes.[2] Type IIS enzymes recognize asymmetric nucleotide sequences and cleave both DNA strands at fixed positions downstream of the recognition site. The archetypal Type IIS enzyme FokI is composed of a specific DNA-binding module fused to a cleavage domain that possesses a conserved restriction endonuclease catalytic core but cuts DNA in a non-specific manner.[3] The FokI recognition domain is made of three smaller subdomains (D1, D2 and D3) that contain helix–turn–helix (HTH) motifs.[4] The first two subdomains are involved in DNA recognition, whereas the third subdomain has been co-opted for protein–protein interactions. Modular architecture is also characteristic for another structurally characterized Type IIS enzyme, BfiI, which is composed of a DNA-binding domain fused to a catalytic module similar to the non-specific nuclease belonging to the phospholipase D family.[5] The DNA recognition domain of BfiI exhibits a barrel-like structure that shows no similarity to FokI.[6]

**Corresponding author.* E-mail address: siksnys@ibt.lt.

Abbreviations used: HTH, helix–turn–helix; wt, wild type; wHTH, winged helix–turn–helix; PEG, polyethylene glycol.

Restriction endonuclease BpuJI, like most Type IIS enzymes, recognizes the asymmetric sequence 5′-CCCGT; however, it cuts at multiple sites in the vicinity of the target sequence.[7] Like FokI and other Type IIS enzymes, BpuJI consists of two physically separate domains: an N-terminal domain that binds to the recognition sequence as a monomer and a C-terminal domain that forms a dimer with non-specific nuclease activity.[7] In contrast to FokI, which is a monomer in solution[8] and dimerizes in the presence of target DNA,[9] full-length BpuJI is a dimer in the DNA-free form.[7] The dimerization interface is formed by the C-terminal domains; therefore, wild-type (wt) BpuJI has two surfaces for specific DNA binding provided by the N-terminal domains. As a consequence, BpuJI binds two copies of cognate DNA, making a synaptic complex that promotes cleavage.[7] The isolated catalytic domain of BpuJI is an active stand-alone nuclease that possesses a non-specific end-directed nuclease activity and shows preference for blunt-ended DNA. In full-length BpuJI, the nuclease activity is repressed and becomes activated only upon specific DNA binding, resulting in a complicated pattern of specific DNA cleavage close to the target site.[7] It is likely that interdomain interactions or the linker connecting catalytic and DNA-binding domains may control the BpuJI cleavage activity, similarly to FokI[4] and BfiI.[6] Fold recognition revealed that the BpuJI C-terminal domain is structurally related to the archaeal Holliday junction resolvases; however, it failed to predict the fold of the N-terminal DNA recognition domain.[7]

Here we report the crystal structure of the BpuJI recognition domain (residues 1–285) bound to a cognate 12-bp oligoduplex at 1.3-Å resolution. The structure reveals two winged-helix subdomains and provides a detailed view of 5′-CCCGT sequence recognition by the restriction endonuclease BpuJI.

Results and Discussion

Overall structure of the protein–DNA complex

Using limited proteolysis and mass spectrometry, we identified a stable N-terminal fragment (residues 1–285) constituting the cognate DNA-binding domain of the BpuJI restriction endonuclease.[7] The BpuJI N-terminal domain–DNA complex obtained by limited proteolysis crystallized in space group $P2_12_12$. The structure of the BpuJI recognition domain in complex with a 12-bp oligoduplex was determined by the single-wavelength anomalous dispersion method using a mercury derivative and refined to 1.3-Å resolution with a free *R*-factor of 16.6% and a working *R*-factor of 13.8%. Electron density shows all the DNA bases and the amino acid residues except the five C-terminal residues 281–285. The asymmetric unit contains one protein monomer bound to the DNA duplex, in accordance with the biochemical data.[7]

The BpuJI N-terminal domain (Fig. 1a) contains two winged-helix subdomains,[10] D1 and D2, that are interspaced by the DL subdomain. The D1 subdomain (residues 1–112) contains an N-terminal arm, five helices (H1–H5) and three strands (B1–B3) (Fig. 1b). Helices H2 and H3 form an HTH structure, which is followed by a signature wing region (B2 and B3) and helices H4 and H5. The loop between 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–H12) and two strands (B6 and B7) (Fig. 1b). D2 also bears an HTH structure, formed by helices H8 and H11. The HTH motif is followed by a β-hairpin wing (B6 and B7) and helix H12. In contrast to canonical transcription factors,[10] an expected "turn" in the HTH motif of D2 is replaced by a 35-aa insertion that contains helices H9 and H10. A 24-aa linker region with the chain topology

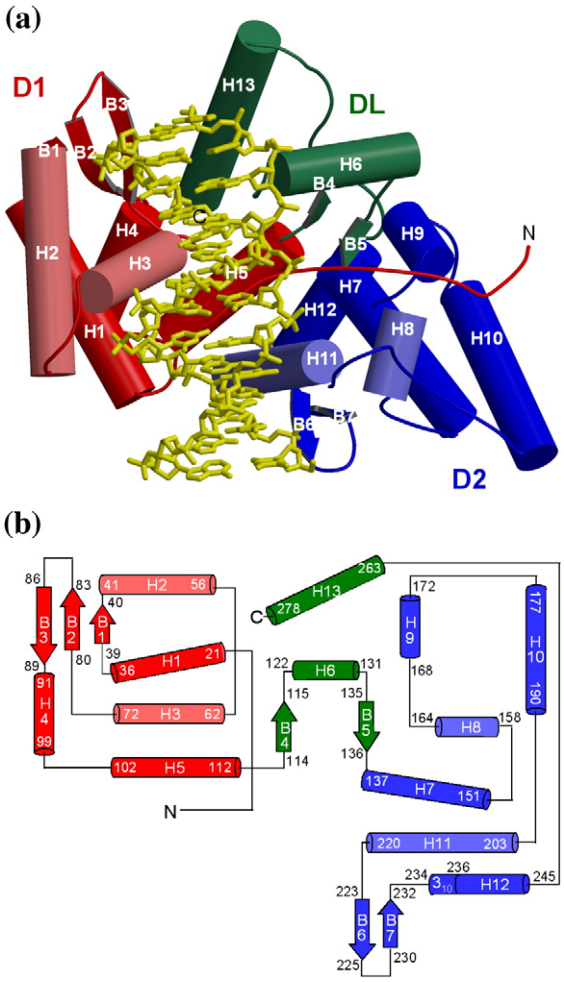

Fig. 1. (a) The BpuJI N-terminal domain (residues 1–285) bound to DNA. The subdomains D1, D2 and DL are shown in red, blue and green, respectively; DNA is represented by yellow sticks. The HTH motifs of the D1 and D2 subdomains are shown in light red and blue, respectively. (b) Topological diagram of the BpuJI N-terminal domain. Helices are shown as cylinders; β-strands, as block arrows. Secondary structure elements are colored as in (a).

B4–H6–B5 connects subdomains D1 and D2 (Fig. 1b). Helix H13 forms a C-terminal extension that stands apart from 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 helix H13 is larger in comparison with the surfaces buried between the linker–H13 fragment and the D1 and D2 subdomains (1192 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 helix H13, are classified as an interrupted subdomain DL.

The DNA in complex with the BpuJI recognition domain maintains a canonical B-DNA conformation with no major bend or kink (Fig. 1a). The recognition helices of both HTH motifs (H3 and H11) are inserted into the major groove on opposite sides of the DNA. The N-terminal arm lies in the major groove and makes numerous contacts with DNA bases and the sugar phosphate backbone. The DL subdomain is also 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.

Comparison with other DNA-binding proteins

A DALI search against the Protein Data Bank (PDB) database, using the entire structure of the BpuJI recognition domain as a target, revealed significant similarity to other DNA-binding proteins (Table 1). Most of the proteins showing similarity to the BpuJI recognition domain harbor a winged-helix motif, some of them, such as reverse gyrase[11] and Werner syndrome protein,[12] have a nuclease function. However, in all cases, the region of similarity is limited to one winged HTH (wHTH) subdomain of BpuJI, and no overall match is observed.

With the use of the structure of the D1 subdomain as a DALI search target, the closest similarity was encountered with the restriction endonuclease FokI (Table 1), which was not found using the entire structure. According to a structural alignment of the TOP3D program,[13] the D1 subdomains of FokI and BpuJI superimpose with an r.m.s. distance of 2.1 Å over 63 C_α atoms. The region of topological similarity spans over the winged-helix motif and the two following helices (Fig. 2a). The winged-helix motif of the BpuJI D1 subdomain is more compact than that of FokI and resembles the domains of the MarR family.[10] The winged-helix motifs of the BpuJI and FokI D2 subdomains also superimpose with an r.m.s. distance of 1.6 Å over 45 C_α atoms (Fig. 2a). They are both extensively modified, with large insertions upstream of the recognition helices, although the insertions show no topological similarity. However, the similarity of the BpuJI D2 subdomain to the FokI D2 subdomain was not revealed by a DALI search using the C-terminal part of the protein as a search target (Table 1).

Table 1. Proteins showing similarity to the BpuJI recognition domain

Search target	PDB accession code	Protein	Z score
BpuJI-Nt (1–280 aa)	1t98	Chromosome partition protein MukF fragment	5.3
	1jgs	Multiple antibiotic resistance protein MarR	5.0
	1gku	Reverse gyrase	4.8
	1fy7	Esa1 histone acetyl-transferase fragment	4.8
	2axl	Werner syndrome protein	4.7
BpuJI-D1 (20–131 aa)	2fok	FokI restriction endonuclease	5.5
	1t98	Chromosome partition protein MukF fragment	5.4
	1yg2	Gene activator AphA	5.2
	1ecl	*Escherichia coli* DNA topoisomerase I	5.0
	1jgs	Multiple antibiotic resistance protein MarR	4.8
BpuJI D2-DL (91–280 aa)	1gln	Aminoacyl-tRNA synthase	3.6
	1kyq	Oxidoreductase, lyase	3.5
	2gau	crpFNR family transcription regulator	3.2
	1gku	Reverse gyrase	3.2
	1w5s	DNA replication initiation protein	3.0

The similarities were revealed by a DALI database search using the BpuJI recognition domain (residues 1–280), the D1 subdomain (residues 20–131) or the C-terminal part of the BpuJI recognition domain (91–280 aa) as a search target. Only five proteins with the highest *Z* scores are shown.

The DNA-binding domain in BpuJI contains two wHTH units, while the FokI recognition domain is composed of three wHTH motifs (Fig. 2b). The BpuJI subdomains show similarity to the respective FokI subdomains D1 and D2 that interact with DNA. The third wHTH subdomain of FokI, which is missing in BpuJI, mediates protein–protein rather than protein–DNA interactions.[4] Thus, both BpuJI and FokI employ two wHTH subdomains to interact with their target sequences.

Although the BpuJI and FokI subdomains D1 and D2 share structural similarity, their orientations in protein–DNA complexes differ significantly (Fig. 2b). In BpuJI, both the recognition helices of the HTH motifs pack against the major groove with their helical axes perpendicular to the long DNA axis and almost parallel with the DNA base pair edges, while in the FokI–DNA complex, only the D1 recognition helix is perpendicular to the DNA axis (Fig. 2b). 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.[4]

Duplications of HTH units are found in a large number of protein families, such as sigma factors,[14] POU family domains,[15] the paired domains of the Pax and Prd transcription factors and the recombinase DNA-binding domains.[16] The HTH units of bipartite domains usually bind to different regions of an extended binding site. In certain cases, the unstructured linker region fills in the minor groove

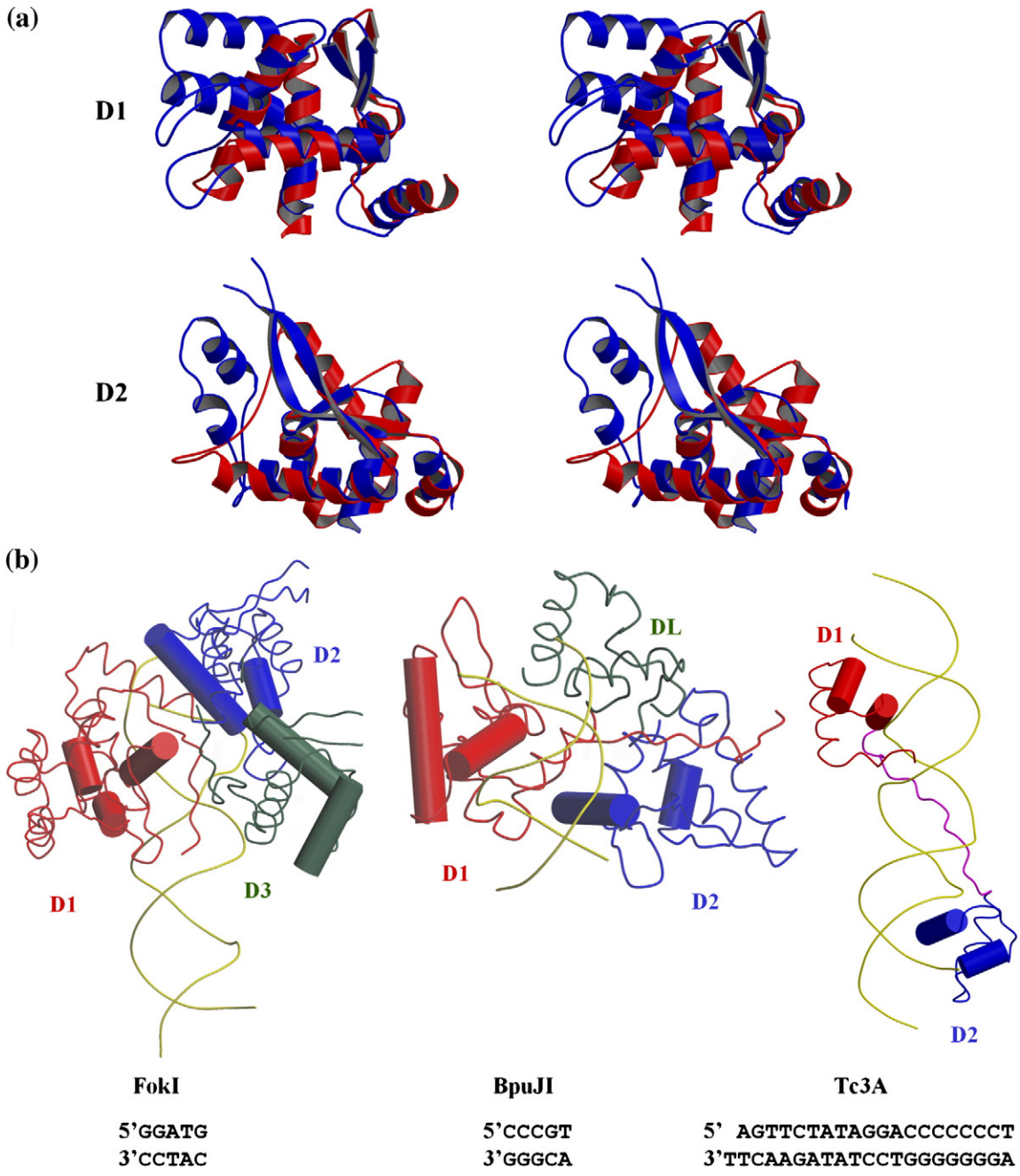

Fig. 2. (a) Stereoview of the TOP3D structural overlays between the BpuJI (red) and FokI (blue) subdomains. The D1 subdomains superimpose with an r.m.s. distance of 2.1 Å over 63 C_α atoms, while the D2 subdomains superimpose with an r.m.s. distance of 1.6 Å over 45 C_α atoms. (b) Comparison of the FokI (PDB accession code 1FOK), BpuJI and Tc3 transposase Tc3A (PDB accession code 1U78) 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; and the HTH motifs are shown as cylinders. The binding sites of FokI, BpuJI and Tc3 transposase Tc3A are presented below the corresponding structures.

between domains, as shown in Fig. 2b for the DNA-binding domain of the Tc3 transposase Tc3A bound to 26-bp DNA.[17] The wHTH units of the bipartite recognition domains of restriction endonucleases BpuJI and FokI bind to the 5-bp target sequences. The wHTH motif was also identified in the restriction endonuclease SdaI, which recognizes the continuous palindromic 8-bp sequence.[18] However, the wHTH unit of SdaI forms a single globular domain and has to dimerize to achieve the recognition of the 8-bp palindromic target site.

DNA recognition by the BpuJI N-terminal domain

Interactions between the BpuJI recognition domain and the target sequence occur along the two HTH motifs and the N-terminal arm, as in the FokI–

DNA complex,[4] although the BpuJI target sequence 5′-CCCGT shows no similarity to the FokI recognition sequence 5′-GGATG. The BpuJI N-terminal domain recognizes the entire target sequence 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 (Fig. 3).

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 (Fig. 3): Lys63 donates a hydrogen bond to O6 of the guanine G_5^* (bases are numbered as in Fig. 3); Asn67 forms hydrogen bonds with N4 of the cytosine C_1 and O6 of the guanine G_4^*; and Glu71 accepts a hydrogen bond from N4 of the cytosine C_2. The inner CG base pairs (CCCGT) are recognized by Arg15 and Lys17 residues, located in the N-terminal arm (Fig. 3). Lys17 donates a hydrogen bond to O6 of the guanine G_3^* and interacts via water with 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 DNA bases are mediated by the N-terminal portion of the recognition helix H11: Gln208 forms a bidentate hydrogen bond with the adenine and Ser204 donates a hydrogen bond to O4 of the thymine (Fig. 3).

The backbone phosphate groups form the 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, and there are four direct and four water-mediated contacts with the HTH motif of the D2 subdomain (Fig. 4). The positive helix dipoles at the N-terminus of the recognition helices H3 and H11 could be in favorable interaction with negatively charged phosphate groups of the DNA backbone. There are also two salt bridges, nine water-mediated hydrogen bonds and two van der Waals contacts between the backbone phosphates and residues located in the N-terminal arm (Fig. 4).

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 (Fig. 4). Lys121 forms a hydrogen bond and a water-mediated contact to the DNA bases upstream of the recognition site. Lys266 located in the C-terminal helix H13 forms a water-mediated contact to the phosphate located 3 bp upstream of the target site (Fig. 4), while Glu263, Lys266 and Lys267 form salt bridges with the phosphates of a symmetry-related DNA molecule (not shown).

Mutational analysis of 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 wt and mutant proteins were expressed and purified as the N-terminal domain (residues 1–285) fusions with a C-terminal hexahistidine tag, and their DNA-

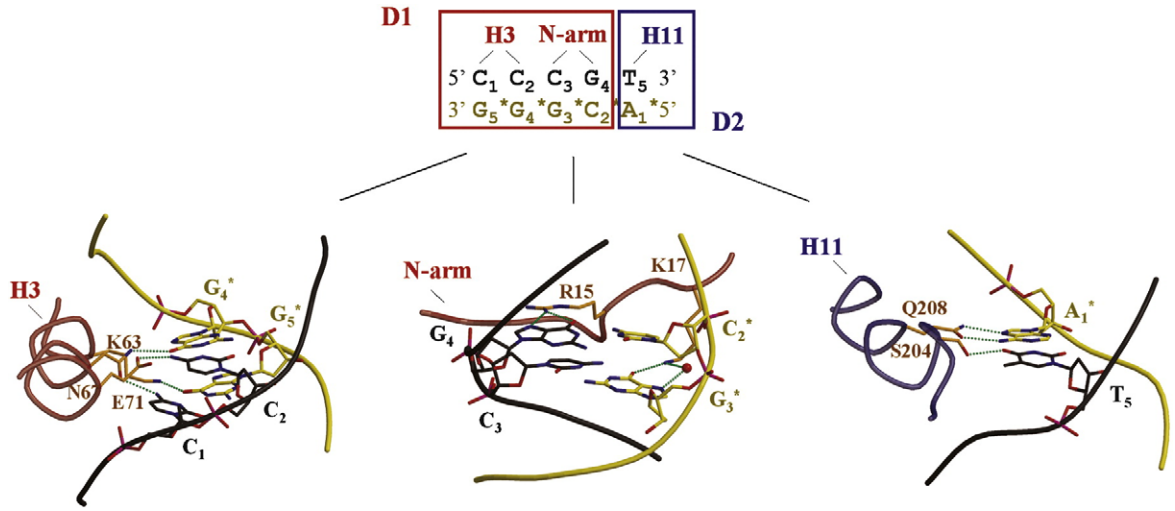

Fig. 3. Target recognition by the BpuJI N-terminal domain. DNA base pairs recognized by the D1 and D2 subdomains are delineated in red and blue, respectively; secondary structure elements contacting the target bases are indicated above the respective base pairs. A detailed view of the recognition helix H3, the N-terminal arm and the recognition helix H11 making hydrogen bonds with the DNA base pairs in the major groove is presented below. Water molecules are represented by red spheres, and hydrogen bonds are shown as dotted lines.

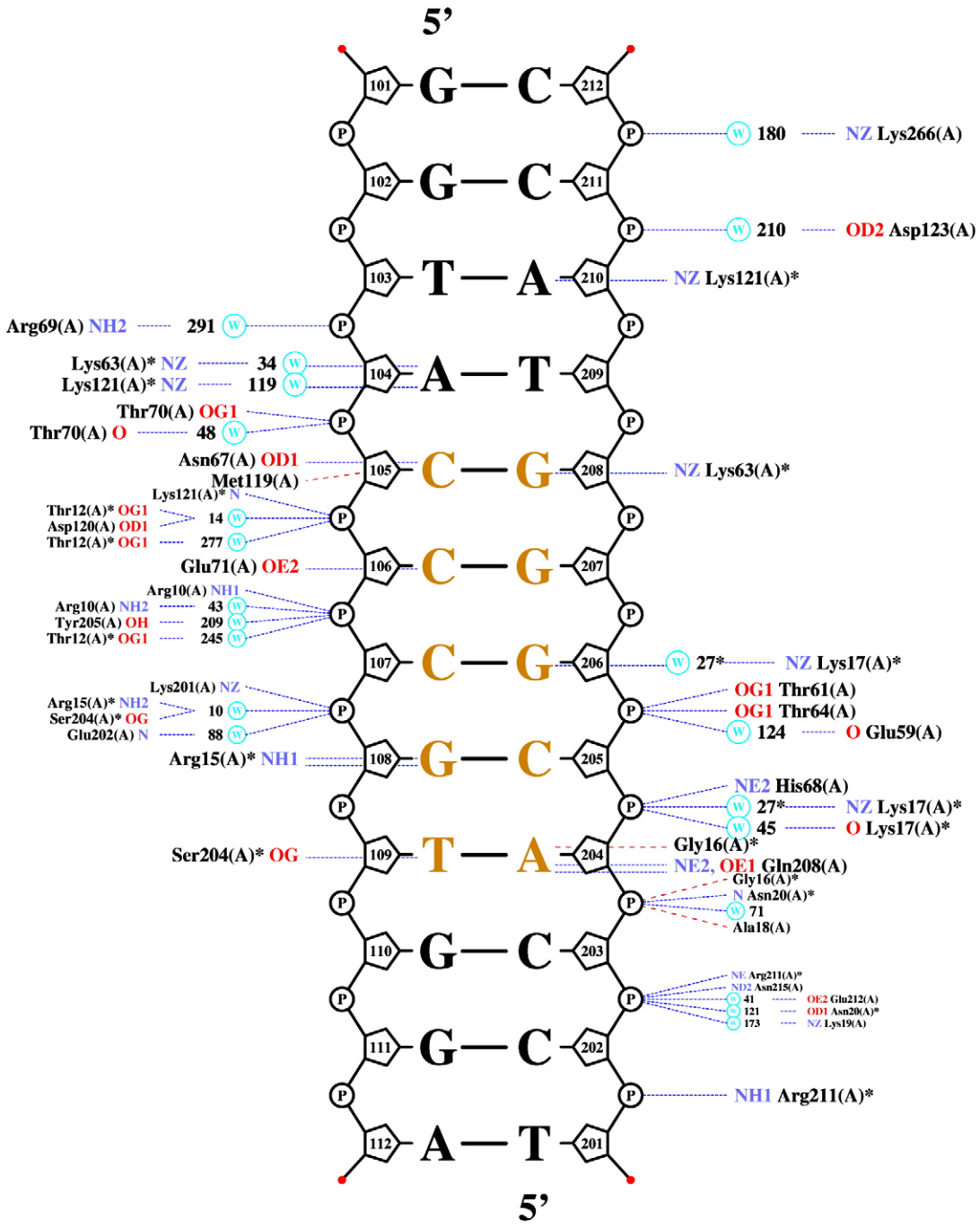

Fig. 4. Schematic diagram summarizing DNA contacts by the BpuJI recognition domain. Hydrogen bonds and salt bridges (<3 Å) are indicated by blue lines, and van der Waals contacts (<3.35 Å) are indicated by red lines. Non-bridging water molecules are not shown. P, phosphate group; w, water molecule; *, residue or water on plot more than once.

binding properties were studied by gel shift analysis. The ^{33}P-labelled 16/16(specific) and 16/16 oligoduplexes (Table 2) were used in DNA-binding experiments as cognate DNA and non-cognate DNA, respectively.

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 non-cognate DNA binding remains very low (Fig. 5). The ability to bind cognate DNA is almost completely abolished by single N67A, E71A, R15A or Q208A mutation and highly compromised by K63A (K_d=48±7.4 nM) or K17A (K_d=67±14 nM) mutation. On the other hand, S204A mutation has a minor effect on DNA binding (K_d=9.6±1.0 nM). Taken together, mutational analysis data are consistent with the major groove interactions observed in the crystal structure. Surprisingly, alanine replacement of Lys121, which interacts with the TA dinucleotide upstream of the BpuJI target site 5′-CCCGT (Fig. 5), increased the binding affinity of the BpuJI N-terminal domain to cognate DNA nearly 10

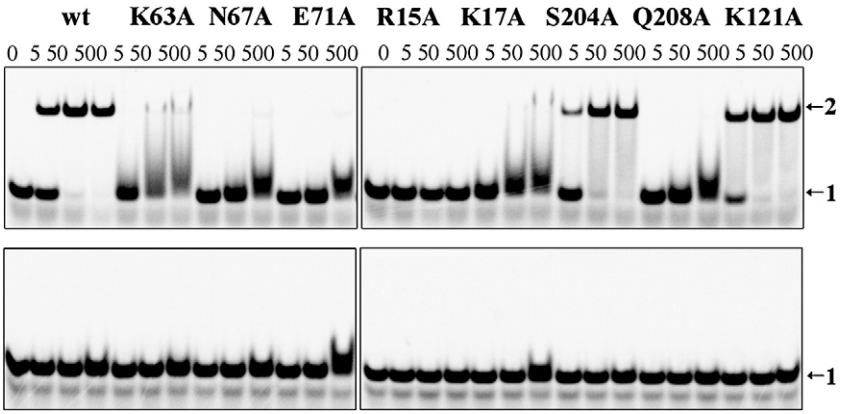

Fig. 5. DNA binding by the BpuJI N-terminal domain and its mutants. Binding reactions contained 2 nM concentration of the ^{33}P-labelled-specific 16/16(SP) (upper gels) or non-specific 16/16 (lower gels) duplex and 0–500 nM concentration of the wt or mutant protein (indicated above the relevant lanes) in the binding buffer. The samples were analyzed by PAGE, as described in Materials and Methods. 1 = free DNA; 2 = protein–DNA complex.

times (K_d = 0.83 ± 0.03 nM). However, it still should be tested if the Lys121 mutant shows the same binding phenotype in the context of the full-length BpuJI.

Conclusions

The restriction endonuclease BpuJI cuts DNA at multiple sites in the vicinity of the asymmetric target sequence.[7] It is composed of two separate domains that perform different functions. The C-terminal catalytic domain cuts DNA non-specifically and is structurally related to the archaeal Holliday junction resolvases, while the N-terminal domain binds to the 5′-CCCGT target site.[7] Protein sequence analysis and fold recognition failed to predict the structure of the N-terminal domain of BpuJI.

In this study, we determined the crystal structure of the BpuJI recognition domain bound to cognate DNA. The structure reveals two winged-helix subdomains, D1 and D2, interspaced by the DL subdomain. The BpuJI D1 and D2 subdomains share structural similarity with the DNA-binding subdomains of the archetypal Type IIS restriction endonuclease FokI.

Both FokI and BpuJI contact the respective target sequences 5′-GGATG and 5′-CCCGT by residues located in the two wHTH units and the N-terminal arm. However, the orientations of the wHTH subdomains in FokI and BpuJI bound to cognate DNA are different (Fig. 2b). Thus, two topologically similar wHTH motifs and an extended N-terminal arm are adapted to recognize different asymmetric pentanucleotide sequences in FokI and BpuJI. The presence of the bipartite wHTH motifs in the two Type IIS enzymes recognizing different sequences demonstrates the versatility of HTH units in protein–DNA interactions. Indeed, while the HTH units of bipartite domains, such as the recognition domain of Tc3 transposase, usually bind to different regions of an extended binding site (Fig. 2b), both the wHTH units of the restriction endonucleases BpuJI and FokI bind to the pentanucleotide target sequences. In addition, the extended N-terminal arm also plays an important role in the sequence-specific DNA binding by BpuJI and FokI. The use of two HTH motifs and an N-terminal arm allows to form a highly cooperative hydrogen bond network that is a characteristic feature of the specific protein–DNA complexes of Type II restriction endonucleases.[1] One may expect that similar structural motifs could be present in other Type IIS enzymes that exhibit modular structures and cleave DNA outside of their target sites.

Table 2. Oligoduplexes used for crystallization and in DNA-binding experiments

12/12(SP)	5′-GGTACCCGTGGA
	3′-CCATGGGCACCT
16/16(SP)	5′-TCGGTACCCGTGGATC
	3′-AGCCATGGGCACCTAG
16/16	5′-AGCGTAGCACTGGGCT
	3′-TCGCATCGTGACCCGA

The BpuJI recognition sequence is underlined. The oligoduplexes were prepared by annealing HPLC-purified oligonucleotides from Metabion.

Materials and Methods

Preparation of the N-terminal domain–DNA complex

The restriction endonuclease BpuJI was purified as described previously.[7] The BpuJI–DNA complex was prepared in buffer A (10 mM Tris–HCl, pH 7.5 at 25 °C, 100 mM KCl and 2 mM $CaCl_2$) by diluting BpuJI and a cognate oligoduplex (Table 2) to a final protein concentration of 0.3 mg/ml and 1.1-fold molar ratio of DNA duplex to protein monomer. The complex was digested with thermolysin (Sigma) at a 1:10 mass ratio of the protease to BpuJI for 1 h at 25 °C. The reaction was quenched by adding ethylenediaminetetraacetic acid (EDTA) to a final concentration of 10 mM. The resulting proteolysis mixture was loaded onto a heparin–Sepharose column (GE Healthcare), pre-equilibrated with buffer B (50 mM Tris–HCl, pH 7.5 at 25 °C, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT and 0.02% NaN_3). The flow-through was concentrated by ultrafiltration using centrifugal filter units with a molecular-weight cutoff of 5000 (Millipore) and loaded onto a Superdex 200 HR column (GE Healthcare). Pooled fractions containing the N-terminal domain–DNA complex were concentrated to 6 mg/ml of protein, changing buffer B into buffer C (10 mM Tris–HCl, pH 7.5 at 25 °C, 0.1 mM EDTA, 1 mM DTT and 0.02% NaN_3).

Crystallization and data collection

Crystallization was carried out by the sitting-drop method using screens from Hampton Research. The first crystals of the BpuJI N-terminal domain were obtained with a 16-bp cognate oligoduplex 16/16(SP) (Table 2); however, all the crystals tested were polycrystalline. The problem was solved using the N-terminal domain complex with a 12-bp DNA duplex 12/12(SP), which gave orthorhombic crystals diffracting to 1.3 Å. The best crystals were grown in drops consisting of 1 μl of the N-terminal domain complex with the 12/12(SP) in buffer C and 1 μl of a reservoir solution [0.2 M ammonium tartrate and 20% polyethylene glycol (PEG) 3350] at room temperature. The crystals were soaked in 2 mM $HgCl_2$ for about 1 week to obtain a Hg derivative. For data collection, the crystals were transferred into a cryo-protecting buffer (0.2 M potassium sodium tartrate, 0.2 M NH_4Cl, 20% PEG 4000 and 25% PEG 400) for 30 min prior to flash cryo-cooling. Complete diffraction data sets for the native N-terminal domain complex and the Hg derivative (Table 3) were collected at the EMBL X12 beamline at the DORIS storage ring (DESY, Hamburg, Germany). Data were processed using the MOSFLM,[19] SCALA[20] and TRUNCATE[21] programs.

Structure determination and analysis

Hg positions were identified on anomalous Patterson map Harker sections using the HARA program (S.G., unpublished data). Mercury peaks were brought to the same hand and origin by anomalous Fourier synthesis. Atom positions were refined, and single-wavelength anomalous dispersion phases were calculated in the MLPHARE program.[22] The resulting maps were solvent flattened with DM.[23] Map connectivity was used to identify the correct hand and resulted in a DNA helix with correct handedness. Initial model building was performed with the ARP/wARP program[24]; manual rebuilding using Coot[25] alternated with refinement using CNS.[26] The partially refined structure of Hg derivative was used for molecular replacement with AMORE[22] to solve the native structure. Final refinement steps were carried out with REFMAC[27]; the quality of the model was analyzed with the PROCHECK[28] and WHAT IF[29] programs.

Table 3. Data collection statistics

Data set	Native	Hg
Temperature (K)	100	
Wavelength (Å)	1.000	
Space group	$P2_12_12$	
Cell constants (Å)		
A	56.2	59.0
B	167.2	168.5
C	43.8	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] $R_{merge} = \Sigma|I_{hi} - \langle I_h \rangle| / \Sigma I_{hi}$, 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 reflection h.

Table 4. 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

Statistics concerning refinement and the quality of the final model are presented in Table 4.

Secondary structure was assigned with STRIDE[30]; structure comparisons used the DALI server[31] and the TOP3D program[13]; and protein–DNA contacts were analyzed with NUCPLOT.[32] Figures were prepared with MOLSCRIPT[33] and Raster3D.[34]

Cloning and mutagenesis

The pAL-BpuJI plasmid-containing *bpuJIR* gene was used as a PCR template to clone the N-terminal domain (residues 1–285) into pET21b (Novagen) between the NdeI and XhoI sites. The megaprimer PCR method[35] was used to introduce a silent mutation into the NdeI site present in the BpuJI gene. The resulting construct pET-BpuJINH overexpresses the BpuJI N-terminal domain with a hexahistidine tag at the C-terminus. Site-directed mutagenesis was carried out by the one-step method[36]; the gene regions were sequenced to confirm that only designed mutations had been introduced. Oligonucleotide primers used in the cloning and mutagenesis experiments are presented in Supplementary Table 1.

The C-$(His)_6$-tagged wt and mutant proteins were expressed in *E. coli* ER2566 (New England Biolabs) cells, harboring the pET-BpuJINH and pACYC184 derivative with the inserted *bpuJIM1* and *bpuJIM2* genes. The cells were grown to an $OD_{600} = 0.7$ in an LB medium at 37 °C and induced by 1 mM IPTG for 3–4 h. The proteins were purified on a Hi-Trap chelating column (GE Healthcare) and dialyzed 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 and 50% (v/v) glycerol]. Protein concentration was determined from A_{280} readings using an extinction coefficient of 33,350 M^{-1} cm^{-1}.

Gel mobility shift assay for DNA binding

The cognate 16/16(SP) and non-cognate 16/16 oligoduplexes (Table 2) were 5′-labelled with [γ-^{33}P]ATP and T4 polynucleotide kinase (Fermentas). The ^{33}P-labelled DNA was incubated with the wt or mutant N-terminal domain in a binding buffer [30 mM Mes, pH 6.5, 30 mM histidine, 10% (v/v) glycerol and 0.2 mg/ml of bovine serum albumin] for 10 min at room temperature. Binding mixes contained 0.1–2 nM concentration of the ^{33}P-labelled oligoduplex and 0.1–1000 nM concentration of the protein. Unlabelled DNA (10–100 nM) was added to binding mixes to analyze binding of low-affinity

mutants. Samples were electrophoresed through 8% polyacrylamide gels in a running buffer (30 mM Mes, pH 6.5, and 30 mM histidine) for 2 h at 6 V/cm. Radiolabelled DNA was detected and quantified using a Cyclone Storage Phosphor System with the OptiQuant program. K_d values were calculated by fitting data to the following equation:

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

where y is the concentration of protein–DNA complex (in terms of nanomolar) at the total protein concentration x, s_0 is the total DNA concentration in the binding mixture and K_d is the dissociation constant. Data were analyzed using the KyPlot program.[37]

PDB accession code

Coordinates and structure factors are deposited in the PDB under accession code 2VLA.

Acknowledgements

Experiments at the Vilnius Institute of Biotechnology were supported by the Max Planck Society. Measurements at the EMBL outstation were supported by the European Community's Research Infrastructure Action under the Sixth Framework Programme "Structuring the European Research Area Specific Programme." We are grateful to Fermentas UAB for the BpuJI restriction–modification system clone. We thank Manfred Weiss for invaluable help with operating the X12 beamline, Grazvydas Lukinavicius for mass spectrometry and Giedre Tamulaitiene for discussions and comments.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.03.041

References

1. Pingoud, A. & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727.
2. Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J. *et al.* (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812.
3. Li, L., Wu, L. P. & Chandrasegaran, S. (1992). Functional domains in FokI restriction endonuclease. *Proc. Natl. Acad. Sci. USA*, **89**, 4275–4279.
4. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
5. Zaremba, M., Urbanke, C., Halford, S. E. & 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.
6. Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R. & 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. USA*, **102**, 15797–15802.
7. Sukackaite, R., Lagunavicius, A., Stankevicius, K., Urbanke, C., Venclovas, C. & 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.
8. Kaczorowski, T., Skowron, P. & Podhajska, A. J. (1989). Purification and characterization of the FokI restriction endonuclease. *Gene*, **80**, 209–216.
9. Bitinaite, J., Wah, D. A., Aggarwal, A. K. & Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. USA*, **95**, 10570–10575.
10. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262.
11. Rodríguez, A. C. & Stock, D. (2002). Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA. *EMBO J.* **21**, 418–426.
12. Hu, J., Feng, H., Zeng, W., Lin, G. & Xi, X. G. (2005). Solution structure of a multifunctional DNA- and protein-binding motif of human Werner syndrome protein. *Proc. Natl. Acad. Sci. USA*, **102**, 18379–18384.
13. Lu, G. (2000). TOP: a new method for protein structure comparisons and similarity searches. *J. Appl. Crystallogr.* **33**, 176–183.
14. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O. *et al.* (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell*, **9**, 527–539.
15. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G. & 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.
16. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
17. Watkins, S., van Pouderoyen, G. & 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.
18. Tamulaitiene, G., Jakubauskas, A., Urbanke, C., Huber, R., Grazulis, S. & Siksnys, V. (2006). The crystal structure of the rare-cutting restriction enzyme SdaI reveals unexpected domain architecture. *Structure*, **14**, 1389–1400.
19. Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 48–57.
20. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62**, 72–82.
21. French, G. S. & Wilson, K. S. (1978). On the treatment of negative intensity observations. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **34**, 517–525.
22. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.

23. 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.
24. Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). ARP/wARP's model-building algorithms: I. The main chain. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 968–975.
25. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
26. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
27. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
28. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067.
29. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics*, 8, 52–56, 29.
30. Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.
31. Holm, L. & Sander, C. (1996). Alignment of three-dimensional protein structures: network server for database searching. *Methods Enzymol.* **266**, 653–662.
32. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (1997). NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions. *Nucleic Acids Res.* **25**, 4940–4945.
33. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
34. Merritt, E. A. & Murphy, M. E. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 869–873.
35. Barik, S. (1995). Site-directed mutagenesis by double polymerase chain reaction. *Mol. Biotechnol.* **3**, 1–7.
36. Zheng, L., Baumann, U. & Reymond, J. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**, e115.
37. Yoshioka, K. (2002). KyPlot—a user-oriented tool for statistical data analysis and visualization. *CompStat*, **17**, 425–437.

Note added in the proof: While the article was being prepared and submitted, the coordinates of the site-specific nicking endonuclease N.BspD6I (PDB:2EWF) in the DNA free form were released. The N-terminal part of N.BspD6I shows significant similarity to the BpuJI N-terminal domain (DALI Z-score=13.4). Further modelling studies should reveal if BpuJI and N.BspD6I share similar mode of target recognition.