14022–14030 *Nucleic Acids Research, 2014, Vol. 42, No. 22* *Published online 27 November 2014*
doi: 10.1093/nar/gku1237

Crystal structure of the R-protein of the multisubunit ATP-dependent restriction endonuclease NgoAVII

Giedre Tamulaitiene, Arunas Silanskas, Saulius Grazulis, Mindaugas Zaremba* and Virginijus Siksnys*

Department of Protein–DNA Interactions, Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241 Vilnius, Lithuania

Received September 25, 2014; Revised October 31, 2014; Accepted November 10, 2014

ABSTRACT

The restriction endonuclease (REase) NgoAVII is composed of two proteins, R.NgoAVII and N.NgoAVII, and shares features of both Type II restriction enzymes and Type I/III ATP-dependent restriction enzymes (see accompanying paper Zaremba *et al.*, 2014). Here we present crystal structures of the R.NgoAVII apo-protein and the R.NgoAVII C-terminal domain bound to a specific DNA. R.NgoAVII is composed of two domains: an N-terminal nucleolytic PLD domain; and a C-terminal B3-like DNA-binding domain identified previously in BfiI and EcoRII REases, and in plant transcription factors. Structural comparison of the B3-like domains of R.NgoAVII, EcoRII, BfiI and the plant transcription factors revealed a conserved DNA-binding surface comprised of N- and C-arms that together grip the DNA. The C-arms of R.NgoAVII, EcoRII, BfiI and plant B3 domains are similar in size, but the R.NgoAVII N-arm which makes the majority of the contacts to the target site is much longer. The overall structures of R.NgoAVII and BfiI are similar; however, whilst BfiI has stand-alone catalytic activity, R.NgoAVII requires an auxiliary cognate N.NgoAVII protein and ATP hydrolysis in order to cleave DNA at the target site. The structures we present will help formulate future experiments to explore the molecular mechanisms of intersubunit crosstalk that control DNA cleavage by R.NgoAVII and related endonucleases.

INTRODUCTION

Type II restriction endonucleases (REases) are the components of restriction-modification (RM) systems, and recognize specific DNA sequences, cleaving the DNA at fixed positions at or close to these sequences (1). Usually they require Mg^{2+} ions as a cofactor for nuclease activity, and in addition some are stimulated by S-adenosyl methionine (AdoMet). They may act as monomers, dimers, tetramers or even higher order species, and typically do not need the presence of their companion methyltransferases (MTases) in order to cleave DNA (2). In contrast, Type I and Type III REases are multisubunit protein complexes that can function as both nuclease and methyltransferase, with the former having an absolute requirement for ATP hydrolysis (3). Type I REases are composed of HsdR (or R) restriction subunit, HsdM (or M) modification subunit and HsdS (or S) specificity subunit in an $R_2M_2S_1$ stoichiometry (4,5). The R subunit of Type I restriction enzymes contains PD-(D/E)XK nuclease and motor (RecA helicase-like) domains and is responsible for the DNA cleavage and translocation activities of the holoenzyme. Type III restriction enzymes are composed of two modification (Mod) subunits, each containing a target recognition domain, and a single restriction (Res) subunit containing DNA helicase and endonuclease domains (6,7).

The NgoAVII RM system, identified by bioinformatics in *Neisseria gonorrhoeae* FA1090 (ATCC 700825), is composed of three genes: *ngoAVIIM*, *ngoAVIIR* and *ngoAVIIN*. In the accompanying paper (8) we show that NgoAVII shares features of Type I and II enzymes. NgoAVII recognizes the asymmetric sequence 5′-GCCGC-3′ and cleaves the DNA to produce a blunt end 7 nt away from the 3′-end of the recognition site. The cleavage pattern suggests that NgoAVII can be designated as Type IIS system. However, the NgoAVII REase is a multisubunit protein composed of separate nuclease (R.NgoAVII, hereafter referred to as R-protein) and ATPase (N.NgoAVII, hereafter referred to as N-protein) proteins, and requires ATP hydrolysis for DNA cleavage (8). These features suggest that NgoAVII can instead be designated as a Type I-like enzyme. The amino acid sequence of the R-protein is similar to the Type IIS REase BfiI. BfiI is a well characterized, stand-alone Type IIS enzyme, which does not require ATP for DNA cleavage. BfiI is composed of two domains: an N-terminal PLD nuclease domain and a C-terminal B3-like DNA-binding do-

*To whom correspondence should be addressed. Tel: +370 5 2602111; Fax: +370 5 2602116; Email: zare@ibt.lt
Correspondence may also be addressed to Virginijus Siksnys. Tel: +370 5 2602108; Fax: +370 5 2602116; Email: siksnys@ibt.lt

main (9,10). The PLD domains dimerize to form a single active site (9) whilst the B3-like domains bind specific DNA as a monomer (10). Similar B3 domains are also widespread in plant transcription factors, usually in combination with other domains (11). However, only apo-form structures of plant B3 domains are available (12–14) and all clues about DNA recognition of these domains come from DNA bound B3-like domain structures of the BfiI and EcoRII restriction enzymes (10,15).

In this paper, we report the crystal structures of the R.NgoAVII apo-protein and of the C-terminal domain in complex with a cognate DNA, at 2.25 and 2.7 Å resolution, respectively. The R-protein shares a conserved domain architecture with BfiI (9,10). The DNA-binding interface of the B3 domain is comprised of N- and C-arms that are also conserved in the EcoRII-N and BfiI-C B3-like domains. While the C-arms are of similar size, the N-arm of R.NgoAVII, which makes the majority of contacts to the target site, is longer in comparison to equivalent structures found in EcoRII, BfiI and plant B3 domains.

MATERIALS AND METHODS

Oligonucleotides

All oligonucleotides used in this study were synthesized and HPLC purified by Metabion (Martinsried, Germany). Oligoduplexes for crystallization were assembled by slow annealing from 95°C to room temperature in a buffer (10 mM Tris–HCl (pH 8.0 at 25°C), 50 mM NaCl).

Protein expression and crystallization

R.NgoAVII (with N-terminal His_6-tag) and R.NgoAVII-B3 (with C-terminal His_6-tag, residues 179–345) were expressed and purified as described in the accompanying paper (8). The concentration of the protein monomers was estimated spectrophotometrically using extinction coefficients 50880 $M^{-1}cm^{-1}$ and 29450 $M^{-1}cm^{-1}$, respectively. Proteins were crystallized by the sitting drop vapor diffusion technique. The R-protein in a buffer (20 mM Tris–HCl (pH 7.5 at 25°C), 100 mM NaCl, 1 mM dithiothreitol (DTT), 0.02% NaN_3) was concentrated to 6.4 mg/ml and mixed with an equal volume of a crystallization solution (0.1 M NaCl, 0.1 M Na-Bicine pH 9.0, 20% v/v PEGMME 550). Crystals grew in a week. An Hg-soak was obtained by soaking crystals of R-protein overnight in the crystallization solution supplemented with 1 mM ethylmercury chloride. The R-protein crystals were cryo-protected by adding the crystallization solution supplemented with 5% ethylene glycol, and flash frozen in liquid nitrogen. The R.NgoAVII-B3 in buffer (20 mM Tris–HCl (pH 8.5 at 25°C), 300 mM KCl, 5% glycerol, 0.02% NaN_3) was concentrated and mixed with DNA–1 oligoduplex (14 bp with one nucleotide 5′ overhang, Figure 2B) or DNA–2 oligoduplex (13 bp with one nucleotide 5′ overhang, Supplementary Figure S2C) in 1:1.2 ratio to final protein concentration ∼8 mg/ml; and an equal volume of a crystallization solution was added. Crystallization solution for the complex with DNA–1 oligoduplex was: 0.1 M MES-imidazole pH 6.5, 12.5% w/v PEG1000, 12.5% w/v PEG3350, 12.5% 2-methyl-2,4-pentanediol (MPD), 0.02 M of each Glu, Ala, Gly, Lys, Ser. Crystallization solution for the complex with DNA–2 oligoduplex was: 0.2 M ammonium sulfate, 0.1 M Bis-Tris pH 5.5, 25% w/v PEG3350. Data for the complex with DNA–1 oligoduplex were collected without cryo-protection. The crystals of the complex with DNA–2 oligoduplex were cryo-protected by adding the crystallization solution supplemented with 15% ethylene glycol.

Data collection and structure determination

Data for R-protein crystals (native and Hg-derivative) and B3 domain complex with 13 bp DNA were collected in-house, using a Raxis IV++ detector mounted on a RU-H3R copper rotating anode generator (Rigaku). Crystallographic data for the R.NgoAVII-B3 complex with 14 bp DNA were collected at EMBL Hamburg outstation DESY PETRA III P13 beamline on a Pilatus detector. MOSFLM (for in-house collected data) (16) or XDS (for synchrotron dataset) (17), SCALA and TRUNCATE (18) were used for data processing. The structure of R-protein was solved using the SAD protocol of Auto-Rickshaw (19). The substructure was solved using SHELXCDE (20–22). The occupancy of all substructure atoms was refined and initial phases were calculated using the program MLPHARE (18). The 2-fold non-crystallographic symmetry (NCS) operator was found using the program RESOLVE (23). Density modification, phase extension and NCS-averaging were performed using the program DM (24). A partial alpha-helical model was produced using the program HELICAP (25). The partial model contained 700 residues out of the total number of 730 residues. The model was manually rebuilt using COOT (26), and refined using REFMAC (27) and PHENIX (phenix.refine-1.8.3) (28) programs. Initial phases for complex with 14 bp DNA-1 structure were obtained by molecular replacement using MOLREP (29) and B3-domain (residues 179–227, 235–245) of R-protein as an initial model. Initial phases for complex with 13 bp DNA-2 structure were obtained by molecular replacement using MOLREP and B3-domain from complex with 14 bp DNA structure as an initial model. DNA was built manually in COOT. The models were rebuilt and refined using COOT, REFMAC and PHENIX (phenix.refine-1.8.3) programs. All crystal structures were refined using NCS restraints between two protein subunits (in case of the complexes, also DNA) present in the asymmetric unit. Data collection and refinement statistics are shown in Table 1.

DNA geometry in R.NgoAVII-B3–DNA structure was analyzed with CURVES (30). The contact surfaces buried between the two molecules were calculated using NACCESS (31). Protein–DNA contacts were analyzed by NUCPLOT (32). All molecular scale representations were prepared using MOLSCRIPT (33) and RASTER3D (34) software.

RESULTS

Overall structure of R.NgoAVII

To elucidate structure-function relationships in the NgoAVII REase, we solved the structure of the R-protein. The R.NgoAVII crystals belong to space group $P2_12_12_1$,

Table 1. Data collection and refinement statistics

	R.NgoAVII	R.NgoAVII-Hg	R.NgoAVII-B3–DNA1 complex	R.NgoAVII-B3–DNA2 complex
Data collection statistics				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P4_12_12$	$P2_12_12_1$
A (Å)	75.51	74.79	88.41	63.31
B (Å)	106.80	106.85	88.41	95.46
C (Å)	108.23	110.31	152.44	97.81
Wavelength	1.54	1.54	0.976300	1.54
X-ray source	Cu rotating anode	Cu rotating anode	P13 PETRA III	Cu rotating anode
Total reflections	322521	254521	330637	114931
Unique reflections	42289	22169	17327	16907
Resolution range (Å)	29.7–2.25	28.1–2.95	152.4–2.7	38.7–2.7
Completeness (%) (last shell)	99.9 (100)	99.0 (94.6)	100 (99.7)	100 (100)
Multiplicity (last shell)	7.6 (7.6)	11.5 (9.9)	19.1 (14.8)	6.8 (6.9)
I/σ (last shell)	7.9 (1.8)	7.8 (2.1)	5.4 (1.9)	5.8 (1.1)
R(merge) (%) (last shell)	8.1 (39.2)	8.3 (36.1)	9.7 (34.5)	11.7 (65.7)
B(iso) from Wilson ($Å^2$)	33.5	70.6	47.9	64.2
Refinement statistics				
Resolution range (Å)	27.80–2.25		48.4–2.7	38.7–2.7
Reflections work/test	72142/8094		28455/3242	28326/3079
Protein atoms	5581		2492	2631
DNA atoms			1224	1200
Solvent molecules	348		86	53
R-factor (%)	18.7		21.4	21.8
R-free (%)	21.0		25.0	25.8
R.m.s.d. bond lengths (Å)	0.005		0.006	0.004
R.m.s.d. angles (deg)	1.063		0.794	0.647
Ramachandran core region (%)	98.1		96.8	97.2
Ramachandran allowed region (%)	1.9		3.2	2.8
Ramachandran disallowed region	0		0	0

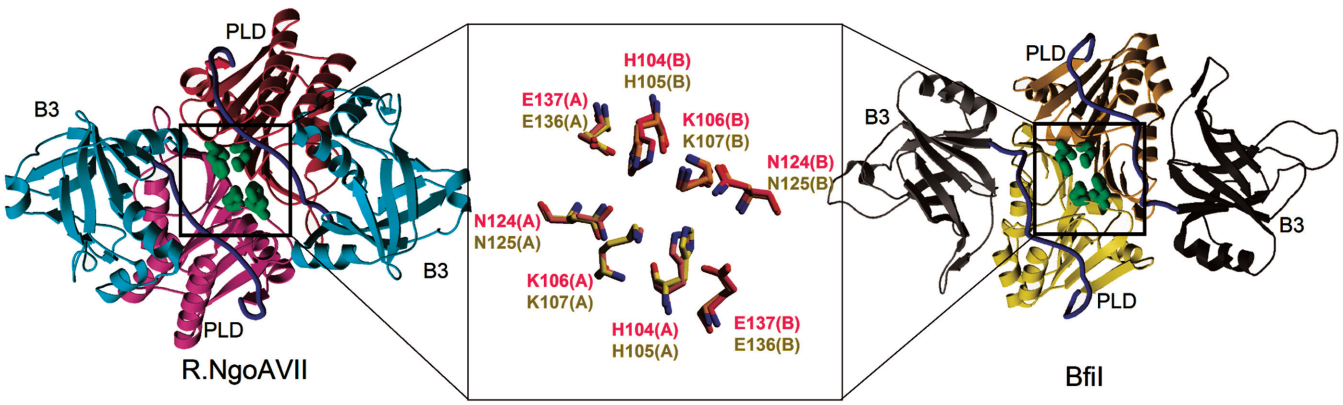

Figure 1. Structure of R.NgoAVII. The R.NgoAVII N-terminal PLD domains (pink) form a dimer with a single active site (active site residues H104 and K106 are depicted in green). The C-terminal B3-like domains (cyan) are connected to the catalytic domains by long linkers (blue) and are positioned on both sides of a dimeric core. The R.NgoAVII dimer is similar to BfiI (PDB ID 2C1L). BfiI PLD domains are colored yellow, B3-like domains are gray, interdomain linkers are blue, active site residues H105 and K107 are shown in green. The central panel shows the superposition of the active site residues of R.NgoAVII (pink) and BfiI (yellow).

contain two protein molecules in the asymmetric unit and diffract X-rays to 2.25 Å resolution. The structure was solved by SAD method using a Hg derivative. The statistics of X-ray diffraction data collection and structure refinement are shown in Table 1.

The crystal structure reveals that R.NgoAVII is a two domain protein, which forms a dimer in the asymmetric unit (Figure 1). The dimerization is achieved through the N-terminal domains (residues 1–164). Long extended linkers (residues 165–187) connect the N-terminal domains to the C-terminal domains (residues 188–345), which are located on both sides of a dimeric core and which make only minor contribution to the dimer interface. Indeed, the surface area buried at the interface between the N-terminal domains is ∼3000 $Å^2$, while the buried surface area between the full-length protein subunits is ∼3100 $Å^2$. The dimerization interface involves 48 residues that form an intricate network of hydrogen bonds and hydrophobic interactions. The linker residues do not contribute much to the dimer interface but do make extensive contacts with the N- and C-domains. The N-terminal linker fragment (residues 165–175) makes 12 hydrogen bonds with a catalytic N-terminal domain. On the

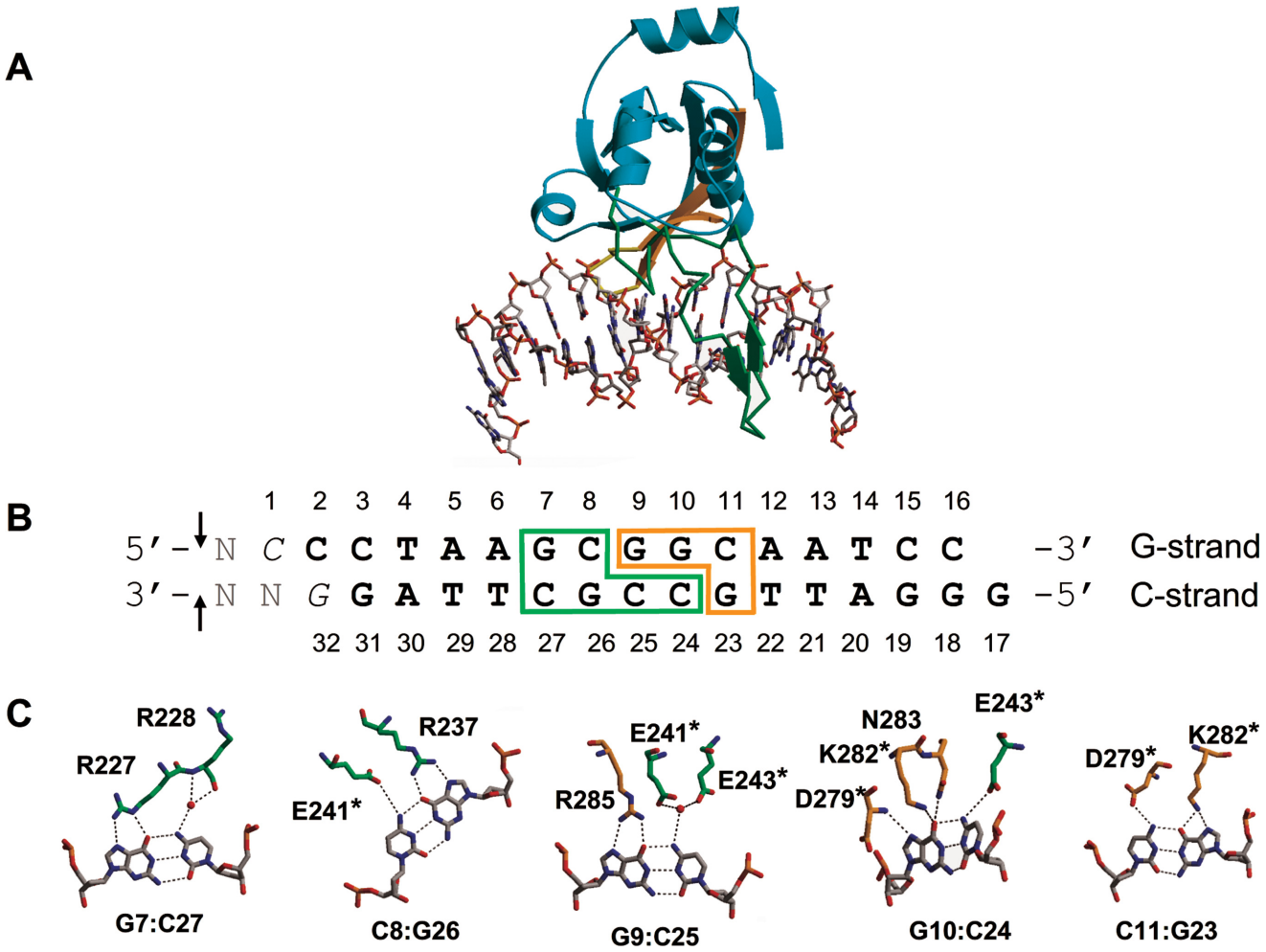

Figure 2. DNA recognition by R.NgoAVII-B3. (**A**) Overall structure of the complex. The N-arm (residues 212–247) is colored green and the C-arm (residues 270–290) is colored orange. (**B**) The cognate oligoduplex DNA–1 used in co-crystallization. The DNA bases of the recognition site are boxed green and orange depending on whether they are recognized by the N-arm or C-arm, respectively. Nucleotides C1 and G32 (showed in italic) are not visible in the structure. The 'N' nucleotides (gray) are added to mark the NgoAVII cleavage sites shown by the arrows (8). To simplify the comparison of DNA binding and recognition patterns between R.NgoAVII-B3 and BfiI/EcoRII B3-like domains, we kept the conserved orientation of the N- and C-arms, therefore the recognition sequence 5′-GCCGC-3′ of R.NgoAVII is shown in the opposite direction. This flip of the recognition sequence conditionally positions the cleavage site on the opposite side of the target sequence. (**C**) Recognition of the individual base pairs by R.NgoAVII-B3. Residues from the N-arm are colored green and residues from the C-arm are colored orange. Residues involved in the recognition of more than one base pair are marked by asterisk.

other hand, the C-terminal fragment of the linker (residues 177–187) interacts with a C-terminal DNA-binding domain predominantly through the hydrophobic interactions. The interface between the N- and C-terminal domains within a monomer (Figure 1) buries ∼1600 Å^2 of the accessible surface area and contains nine hydrogen bonds.

Structure of the nuclease core and active site of R.NgoAVII

The N-terminal domain of R.NgoAVII is composed of an eight-stranded mixed β-sheet (β1–6, β8–9), flanked by one α helix on one side and three α helices on the other side (Figure 1 and Supplementary Figure S1). A DALI search (35) against the Protein Data Bank (PDB) database of protein structures revealed significant structural similarities between the N-terminal domain of R.NgoAVII and several PLD nucleases: BfiI REase N-terminal domain (PDB ID: 2C1L, Z-score 12.6); phospholipase D (PDB ID: 1BYS, Z-score 12.3) and Nuc nuclease (PDB ID: 1BYR, Z-score 12.3). The PLD superfamily is defined by a common sequence motif, HxK(x)$_4$D(x)$_6$GSxN, and includes enzymes involved in signal transduction, lipid biosynthesis and DNA cleavage (36). We therefore conclude that the N-terminal domain (referred to as R.NgoAVII-PLD in the following) is responsible for the R.NgoAVII nuclease activity and could be referred to it as the catalytic domain.

Two R.NgoAVII-PLD domains form a dimer which contains a single active site, composed of residues H104, K106, D124 and E137 from both subunits (Figure 1, central panel). These residues are equivalent to the BfiI active site residues H105, K107, N125 and E136 (9). Similar to the BfiI H105A mutant, the H104A replacement in R.NgoAVII abolished nucleolytic activity of the RN.NgoAVII complex (8,37).

Structure of the B3-like domain of R.NgoAVII

The C-terminal domain of R.NgoAVII comprises residues 188–345, which fold into a barrel-like structure composed of seven β strands, flanked by four α helices (Figure 1 and Supplementary Figure S1). The DALI search revealed clear structural similarity between the C-terminal domain of R.NgoAVII and plant transcription factor B3 domain-containing proteins including: plant transcription factors RAV1 (PDB ID: 1WID, Z-score 6.7) and VRN1 (PDB ID: 4ILK, Z-score 5.7); hypothetical *Arabidopsis thaliana* protein At1g16640.1 (PDB ID: 1YEL, Z-score 5.7); and the DNA-binding domains of BfiI (PDB ID: 2C1L, Z-score 4.9); and EcoRII (PDB ID: 3HQF, Z-score 4.4). Thus, the C-terminal domain of R.NgoAVII is most likely responsible for DNA binding (referred to as R.NgoAVII-B3 in the following). Indeed, the isolated R.NgoAVII-B3 domain binds specific DNA with similar affinity as a full-length R.NgoAVII protein (8).

The overall structure of R.NgoAVII is similar to the Type IIS restriction enzyme BfiI. Both proteins are composed of two domains: the N-terminal PLD nuclease domain responsible for DNA cleavage and the C-terminal B3-like DNA-binding domain (9,10) (Figure 1).

An interdomain linker couples the nuclease and DNA-binding domains

The R.NgoAVII linker (residues 165–187) connecting the PLD and B3-like domains of R.NgoAVII has an extended conformation running near the active site/DNA-binding cleft at the PLD domain (Figure 1) and has four characteristic negatively charged Glu residues (E167, E169, E176 and E182). BfiI possesses a similar linker (Figure 1), which also contains a stretch of negatively charged Asp and Glu residues. It was suggested that the BfiI linker mimics the DNA backbone and acts as an intramolecular inhibitor that represses nucleolytic activity of the catalytic N-terminal domain by occluding the DNA-binding surface (9). One cannot exclude that negatively charged Glu residues in R.NgoAVII play a similar role.

Structure of the B3-like domain of R.NgoAVII bound to DNA

To elucidate the structural mechanism for DNA recognition by R.NgoAVII, we co-crystallized the R.NgoAVII–B3 domain with a cognate oligoduplex (DNA–1) and solved the structure at 2.7 Å resolution (Figure 2). The structure was solved by molecular replacement using the apo R.NgoAVII-B3 domain structure. The R.NgoAVII-B3 domain binds the DNA as a monomer. The asymmetric unit of the crystal contains two R.NgoAVII-B3–DNA complexes. Both monomers in the asymmetric unit are very similar and their Cα atoms can be superimposed with an r.m.s.d. of 0.38 Å. In the crystal, the DNA oligoduplexes form a pseudo-continuous helix due to stacking interactions between symmetry-related DNA oligoduplexes in the asymmetric unit (Supplementary Figure S2A). Interestingly, two terminal nucleotides (one from each strand, C1 and G32 in Figure 2B and Supplementary Figure S2B) are not visible in the density and apparently do not contribute to the pseudo-helix formation. Expecting to obtain better diffracting crystals, we performed a further round of crystallization trials using an oligoduplex (DNA–2) that lacked these terminal nucleotides (Supplementary Figure S2C). However, the crystal packing resulted in a different space group ($P2_12_12_1$ instead of $P4_12_12$) with similar resolution (Table 1). The overall structure of the complex in both space groups is very similar (overlay of the protein Cα atoms results in r.m.s.d. of 0.35 Å, with the DNA conformation differing only slightly outside the binding region); therefore only the first structure will be discussed.

The concave DNA-binding cleft of R.NgoAVII-B3, located on the side of the barrel, is formed by two so-called DNA recognition arms. The N-arm (residues 212–247) is composed of β strands β12–14, whilst the C-arm (residues 270–290) contains β strands β16–17 (Figure 2A, Supplementary Figure S1). The DNA recognition arms approach the DNA from the major groove side (Figure 2A). The DNA in the complex is mainly in a B–form conformation but is slightly bent away (∼22 degrees) from the protein along the entire length. The DNA strand which contains the C nucleotide at the central position is referred to as the C-strand, and the complement as the G-strand (Figure 2B). The protein makes extensive contacts with the DNA backbone, particularly at the 5′ end of the each strand of the recognition site. Most contacts to the backbone of the C-strand are made by residues from N-arm, while the backbone of G-strand is bound by the C-arm. The structure of the DNA-bound R.NgoAVII-B3 domain is similar to that of apo-R.NgoAVII-B3 with the only exception being the orientation of the N-arm tip (residues 225–237) which moves closer to the DNA in the complex (Supplementary Figure S3).

Site-specific DNA recognition by R.NgoAVII

The DNA sequence-specific readout by R.NgoAVII-B3 is provided by the amino acid residues located in both the N- and C-arms. Residues in the N-arm are involved in hydrogen bonding interactions with the bases at the 3′-end of the recognition site (green box in Figure 2B, note the position of the cleavage site), whereas the residues in the C-arm make specific contacts with the bases at the 5′-end of the target site (orange box in Figure 2B). In total, nine R.NgoAVII amino acid residues make H-bonds to the bases of the recognition site in the major groove, and base-specific contacts are not made in the minor groove (Figure 2C). The guanine bases are recognized mainly by Arg and Lys residues, while cytosines make direct or water-mediated contacts to Glu and Asp residues. Contacts to the central C:G base pair determine the orientation of the recognition sequence and the position of the cleavage site. The G base (G9) is unambiguously recognized by bidentate hydrogen bonds to the NH1 and NH2 atoms of R285, located in the C-arm, while the N4 of the C base (C25) makes water-mediated contacts to D241 and D243 from the N-arm (Figure 2C).

DISCUSSION

R.NgoAVII is the nuclease subunit of an atypical RM system (see accompanying paper (8)), which cannot act as a stand-alone endonuclease. DNA cleavage activity requires

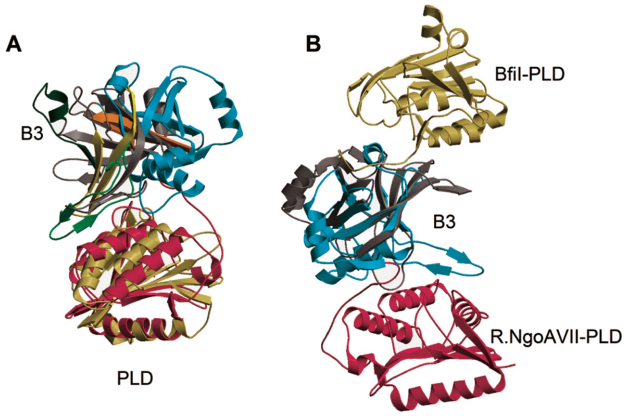

Figure 3. Overlay of R.NgoAVII and BfiI structures. (**A**) R.NgoAVII overlaid on BfiI over PLD domain. R.NgoAVII domains are depicted in pink (PLD) and cyan (B3), BfiI domains are colored khaki (PLD) and gray (B3). DNA recognition arms of R.NgoAVII are colored green (N-arm) and orange (C-arm), while DNA recognition arms of BfiI are colored dark green (N-arm) and yellow (C-arm). (**B**) R.NgoAVII overlaid on BfiI over B3 domain.

complex formation with the cognate N.NgoAVII protein and extensive ATP hydrolysis (8). To establish the structure-function relationship in the atypical NgoAVII REase, we solved the crystal structures of the R.NgoAVII nuclease subunit (R-protein) in the apo-form and the R.NgoAVII-B3 DNA-binding domain in complex with a cognate DNA duplex.

R.NgoAVII is similar to BfiI

Both sequence and structure comparison reveals that BfiI and R.NgoAVII have similar domain organization and overall structures (Figure 1, Supplementary Figure S1). The N-terminal domains of the proteins possess a PLD fold, while the C-terminal domains are similar to B3 plant transcription factor domains (9). Structural overlay of the PLD domains of R.NgoAVII and BfiI proteins reveals that B3-like DNA-binding domains adopt different orientations relative to the N-terminal domains (Figure 3A). The overlay of the proteins using the B3-like domains shows that the R.NgoAVII and BfiI PLD domains are located on the opposite sides of the B3-like domain (Figure 3B). Differently to BfiI, the N-arm of R.NgoAVII-B3 is located close to the PLD domain while the C-arm points away from it (Figure 3A). However, the structure of R.NgoAVII-B3 bound to DNA revealed that the protein uses the N-arm to recognize the 3′ half of the recognition sequence and the C-arm to recognize the 5′ half of the recognition sequence, which is opposite polarity to BfiI (Figures 2 and 4). Accordingly the position of the DNA cleavage site relative to the PLD domain dimer remains conserved in both proteins despite differences in the relative positions of the PLD domain and B3 domain.

The structures determined here do not provide an explanation for why BfiI is a standalone nuclease and R.NgoAVII is not. Biochemical studies of the homologous CglI REase showed that the DEAD helicase-like domain of the ATPase subunit makes contacts with R.CglI. However, the isolated PLD and B3 domains of CglI do not form a stable complex with the CglI DEAD domain (8).

Comparison with the Type I restriction subunit

Type I and Type III restriction enzymes, like RN.NgoAVII, require ATP hydrolysis for DNA cleavage. However, their R subunits contain additional helicase domains, which in the case of the NgoAVII restriction enzyme are provided by the separate N-protein (4,6,8). Extensive ATP hydrolysis by the RN.NgoAVII complex is reminiscent of the Type I systems which can translocate DNA (8). Structures of the R subunits of Type I RM systems EcoR124I and *Vibrio vulnificus* YJ016 are known, and it is tempting to use them to predict the interaction interface of the R.NgoAVII and N.NgoAVII proteins (38,39). However, the structural architecture of R.NgoAVII and the Type I R subunits is very different. First, R.NgoAVII contains the B3 domain, which is responsible for the target sequence recognition; while in Type I systems this function is performed by a separate S subunit which has a different protein fold (40). Second, the nuclease domain of R.NgoAVII belongs to the PLD fold (Figure 1), while Type I (and Type III) nuclease domains belong to the PD-(D/E)XK family (4,6). These differences make it difficult to directly compare R.NgoAVII and Type I R subunit structures, and a structural basis for the interaction between the R- and N-proteins remains to be elucidated experimentally.

DNA binding by B3-like domains of restriction endonucleases

Since only apo-form structures of plant B3 domains are available (12–14), the B3-like domain interaction with DNA is currently exemplified by the DNA-bound structures of restriction endonuclease EcoRII-N and BfiI-C domains, which have many shared features (Figure 4) (10,15). These structures revealed that most of the DNA contacts are made by so-called N- and C-recognition arms, which form a wrench-like DNA-binding cleft (10,15). Structural comparison of EcoRII-N and BfiI-C DNA-bound structures also identified a conserved pattern of contacts to the backbone phosphates, referred to as the 'clamp' phosphates (10). It was proposed that these conserved interactions help to position the DNA in the binding cleft and promote formation of base-specific contacts (10). Conserved 'clamp' phosphate binding residues were also identified in the B3-family plant transcription factors RAV1 and At1g16440 structures, and other B3-like domains (10). The clamp phosphates are bound by the spatially-conserved lysine or arginine residues from the N- and C-arms (Figure 4). Positively charged K23 and R81 residues of EcoRII-N, and R272 and K340 of BfiI-C, interact with the 'clamp' phosphates (Figure 4B and 4C). In R.NgoAVII the structurally-equivalent K212 and K275 residues are positioned in the N- and C-arms, respectively (Figure 4B and 4C). K212 contacts the 5′-'clamp' phosphate of the C-strand. K212 is located on the different structural element compared to K23 of EcoRII and K340 of BfiI; however, the spatial position of these residues is conserved (Figure 4C). The K275 residue makes a conserved contact to the 'clamp' phosphate on the G-strand and overlays well with R81 of EcoRII-N and

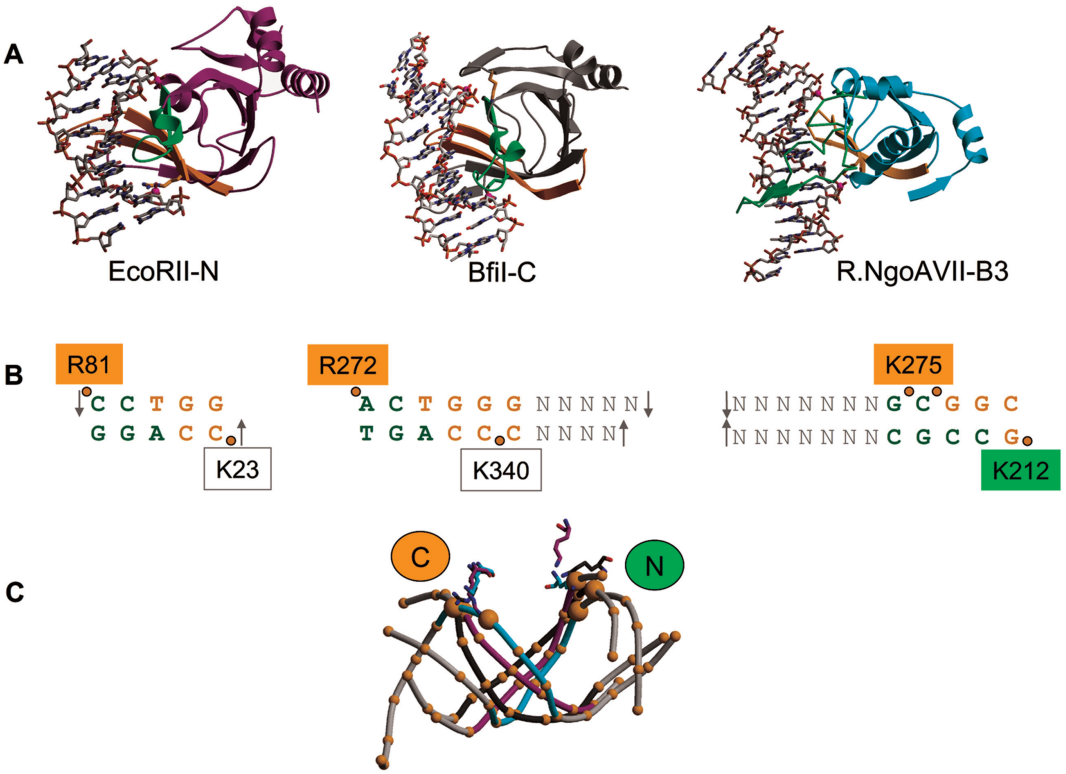

Figure 4. Comparison of DNA-bound structures of restriction endonucleases EcoRII-N, BfiI-C and R.NgoAVII-B3. (**A**) Overall structures of the complexes. DNA recognition N-arms are colored green and C-arms are colored orange. (**B**) Schematic representation of the EcoRII, BfiI and R.NgoAVII binding to DNA. Bases of the recognition sequences which are recognized by the corresponding N-arms are colored green, while those recognized by the C-arms are colored orange. The cleavage sites are shown by gray arrows. Orange circles represent position of the 'clamp' phosphates, residues that make contacts to these phosphates are shown in rectangles colored depending on their location (N- arm - green, C-arm - orange, other-clear). K23 of EcoRII-N is located on the core close to the N-arm (residues 27–42) (15), while K340 of BfiI-C resides on the additional DNA recognition element C-loop (10). (**C**) DNA from the overlaid EcoRII-N (recognition site shown in violet), BfiI-C (dark gray) and R.NgoAVII-B3 (cyan) complexes. The P atoms of the phosphates are shown as orange spheres, with the 'clamp' phosphates highlighted as larger spheres. 'Clamp' phosphates contacting residues are shown and colored using the same schema as for the DNA.

with R272 of BfiI-C (Figure 4C). However, differently from EcoRII-N and BfiI-C, K275 of R.NgoAVII-B3 contacts the second and third phosphates at the 5′ end of the G-strand, rather than the first phosphate (Figure 4B and 4C). The 'clamp' phosphate-contacting residues are also conserved in homologous R.CglI protein sequence: K213 and K278 of R.CglI correspond to K212 and K275 of R.NgoAVII, respectively (Supplementary Figure S1). The structural overlay of B3-like domains shows that the position of DNA in respect to the B3-like domain is slightly different in each case (Figure 4C). This makes it more difficult to predict the base-specific contacts on the basis of amino acid sequence.

DNA recognition interface of B3-like domains

We compared N- and C-arms involved in the target site recognition by R.NgoAVII-B3 with the structural equivalent elements of EcoRII-N and BfiI-C by overlaying the domain structures (Figure 5A). The N-arms of the R.NgoAVII, EcoRII and BfiI B3 domains differ in size and secondary structure, while the C-arms are more similar (Figures 4A and 5A). The N- and C- arms of EcoRII-N each make contacts to one-half of the recognition site (Figure 4B). This symmetry is broken in BfiI-C where the C-arm recognizes a larger part of the target sequence than the N-

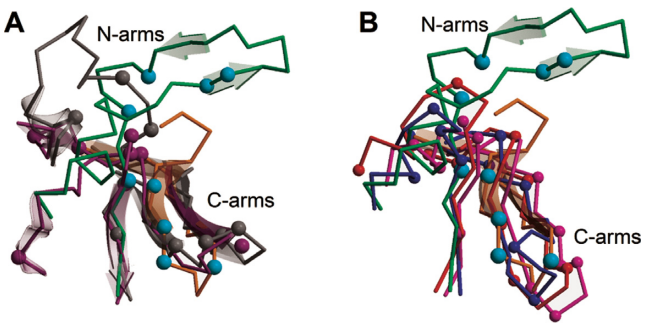

Figure 5. N- and C-arms in R.NgoAVII-B3 and other B3/B3-like domains. (**A**) Superposition of the N- and C-arms of B3-like domains of restriction endonucleases. N-/C-arms of R.NgoAVII-B3 are colored in green (residues 212–247)/orange (270–290), BfiI- gray (N-arm residues 210–231; C-arm residues 266–287), EcoRII - magenta (N-arm residues 27–42, C-arm residues 78–110). Secondary structure elements are transparent. Cα atoms of DNA-binding residues are shown as cyan (R.NgoAVII), gray (BfiI) and magenta (EcoRII) spheres. (**B**) Superposition of R.NgoAVII-B3 arms (coloring as in (A)) with the corresponding elements from RAV1 (blue, residues 193–206, 236–257), VRN1 (red, 42–55, 77–96) and At1g16640.1 (pink, 14–25, 49–68). Cα atoms of residues R, K, D, E, N and Q, located on the N- and C-arms of the plant transcription factors (putative DNA contacting residues) are shown as spheres (the color is same as for the backbone).

arm. In contrast to BfiI-C, the long N-arm of R.NgoAVII-B3 is deeply inserted in the DNA major groove and makes contacts to a larger part of the recognition sequence compared to EcoRII or BfiI. Due to the different orientations of the B3 domains to the PLD domains (which determines the cleavage site and the orientation of the asymmetric recognition sequence 5′-GCCGC-3′), the N-arm of R.NgoAVII contacts the 3′ end while the C-arm - the 5′ end of the recognition site. The DNA contacting residues of EcoRII-N, BfiI-C and R.NgoAVII-B3 are located on the same protein surface; however their Cα atoms do not overlap in most cases (Figure 5A).

We also compared the R.NgoAVII-B3 structure with the B3 plant transcription factor structures RAV1 (12), VRN1 (14) and At1g16440 (13). Structural overlay of the R.NgoAVII-B3 recognition arms with the equivalent structural elements of the plant transcription factors showed a high overall similarity of the C-arms (Figure 5B). However, the N-arms of the plant transcription factors are much shorter compared to R.NgoAVII-B3. To see whether we could predict the DNA recognition residues of the B3 plant transcription factors, we highlighted the Cα atoms of D/E/K/R/N/Q residues that could participate in the DNA recognition (Figure 5B). The DNA-binding surfaces of the B3 plant transcription factors contain multiple putative DNA-binding residues, but few of them overlap with the identified R.NgoAVII residues. Due to the variability of the recognition arm conformations, slight variations in the orientation of DNA, and variable position of the clamp phosphates, it is hard to propose a universal DNA recognition code for the B3 domains.

Model of R.NgoAVII-DNA structure

To obtain a model of the full-length R.NgoAVII-DNA complex, we overlaid the DNA-bound R.NgoAVII-B3 domain on the B3 domain in the apo structure (Figure 6). In the model, the active site (depicted in green) is blocked by the linker (depicted in blue) and the DNA points away from the active site. The distances between the scissile phosphodiester bonds on the 'C' and 'G' strands and the active site are ∼32 and ∼43 Å, respectively. This conformation is incompatible with cleavage within the PLD nuclease active site but may represent R.NgoAVII protein in a DNA target search mode. The switch to the cleavage productive conformation requires a domain rearrangement that opens the DNA-binding cleft in the nuclease domain, displacing the linker and bringing the scissile phosphates into the active site. Importantly, R.NgoAVII requires the presence of an auxiliary N.NgoAVII protein and ATP in order to cleave DNA at the target site. ATP hydrolysis by N.NgoAVII is triggered only when R.NgoAVII is bound to the target site on the DNA (8). Therefore, DNA binding by R.NgoAVII presumably induces a conformational change in N.NgoAVII, enabling ATP hydrolysis and facilitating domain rearrangement into a cleavage proficient conformation (8). Biochemical data show that at least the CglI R- and N-proteins make a heterotetrameric complex. However, the NgoAVII complex cannot be captured by gel-filtration suggesting transient interaction (8). Molecular mechanisms of intersubunit crosstalk that control DNA cleavage and coupled ATP hydrolysis in the NgoAVII REase require further investigation.

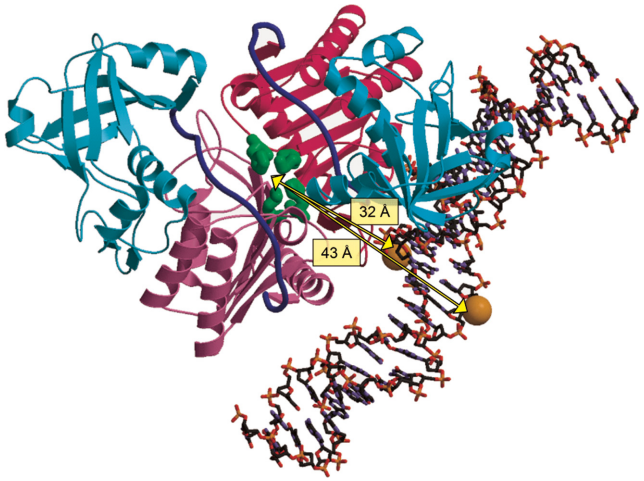

Figure 6. DNA binding by R.NgoAVII. The R.NgoAVII-B3 domain together with its bound DNA and adjacent DNA molecule from the crystal was overlaid onto the B3 domain in the apo structure. Colors are the same as in Figure 1. The orange spheres mark the scissile phosphates, located 7/7 nt away from the recognition site (8). The distances between the active site and scissile phosphates are shown. The domain orientation of R.NgoAVII must change in order to place the scissile phosphate into the active site in R.NgoAVII-PLD domain dimer (pink).

ACCESSION NUMBERS

Coordinates and structure factors of R.NgoAVII and R.NgoAVII-B3-DNA complexes with DNA–1 and DNA–2 oligoduplexes are deposited under PDB IDs: 4RCT, 4RD5 and 4RDM, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The authors would like to thank Rokas Grigaitis for the purified R.NgoAVII-B3 domain. The authors are grateful to Gleb Burenkov for assistance with data collection at EMBL P13 beamline at the PETRA III storage ring, DESY Hamburg. The authors acknowledge Prof. M. D. Szczelkun for a critical reading of the manuscript.

FUNDING

Research Council of Lithuania [MIP-029/2012 to M.Z.]. Funding for open access charge: Research Council of Lithuania.
Conflict of interest statement. None declared.

REFERENCES

1. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.

2. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
3. Bourniquel,A.A. and Bickle,T.A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie*, **84**, 1047–1059.
4. Youell,J. and Firman,K. (2012) Mechanistic insight into Type I restriction endonucleases. *Front. Biosci. (Landmark Ed)*, **17**, 2122–2139.
5. Loenen,W.A., Dryden,D.T., Raleigh,E.A. and Wilson,G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.*, **42**, 20–44.
6. Raghavendra,N.K., Bheemanaik,S. and Rao,D.N. (2012) Mechanistic insights into type III restriction enzymes. *Front. Biosci. (Landmark Ed.)*, **17**, 1094–1107.
7. Butterer,A., Pernstich,C., Smith,R.M., Sobott,F., Szczelkun,M.D. and Toth,J. (2014) Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–5150.
8. Zaremba,M., Toliusis,P., Grigaitis,R., Manakova,E., Silanskas,A., Tamulaitiene,G., Szczelkun,M.D. and Siksnys,V. (2014) DNA cleavage by CgII and NgoAVII requires interaction between N- and R-proteins and extensive nucleotide hydrolysis. *Nucleic Acids Res.*, doi:10.1093/nar/gku1236.
9. 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.
10. Golovenko,D., Manakova,E., Zakrys,L., Zaremba,M., Sasnauskas,G., Grazulis,S. and Siksnys,V. (2014) Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal structure of the C-terminal fragment of BfiI restriction enzyme. *Nucleic Acids Res.*, **42**, 4113–4122.
11. Yamasaki,K., Kigawa,T., Seki,M., Shinozaki,K. and Yokoyama,S. (2013) DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci.*, **18**, 267–276.
12. Yamasaki,K., Kigawa,T., Inoue,M., Tateno,M., Yamasaki,T., Yabuki,T., Aoki,M., Seki,E., Matsuda,T., Tomo,Y. *et al.* (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell*, **16**, 3448–3459.
13. Waltner,J.K., Peterson,F.C., Lytle,B.L. and Volkman,B.F. (2005) Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. *Protein Sci.*, **14**, 2478–2483.
14. King,G.J., Chanson,A.H., McCallum,E.J., Ohme-Takagi,M., Byriel,K., Hill,J.M., Martin,J.L. and Mylne,J.S. (2013) The Arabidopsis B3 domain protein VERNALIZATION1 (VRN1) is involved in processes essential for development, with structural and mutational studies revealing its DNA-binding surface. *J. Biol. Chem.*, **288**, 3198–3207.
15. Golovenko,D., Manakova,E., Tamulaitiene,G., Grazulis,S. and Siksnys,V. (2009) Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII. *Nucleic Acids Res.*, **37**, 6613–6624.
16. Leslie,A.G.W. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 48–57.
17. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
18. CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
19. Panjikar,S., Parthasarathy,V., Lamzin,V.S., Weiss,M.S. and Tucker,P.A. (2005) Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 449–457.
20. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1772–1779.
21. Sheldrick,G.M., Hauptman,H.A., Weeks,C.M., Miller,R. and Uson,I. (2001) In: Arnold,MGRaE (ed). *International Tables for Macromolecular Crystallography*. Kluwer Academic Publishers, Dordrecht, **F**, pp. 333–345.
22. Sheldrick,G.M. (2002) Macromolecular phasing with SHELXE. Z. *Kristallogr.*, **217**, 644–650.
23. Terwilliger,T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 965–972.
24. Cowtan,K. (1994) 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsl. Protein Crystallogr.*, **31**, 34–38.
25. Morris,R.J., Zwart,P.H., Cohen,S., Fernandez,F.J., Kakaris,M., Kirillova,O., Vonrhein,C., Perrakis,A., Lamzin,V.S., Zwart,P.H. *et al.* (2004) Breaking good resolutions with ARP/wARP. *J. Synchrotron Rad.*, **11**, 56–59.
26. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
27. Murshudov,G., Vagin,A. and Dodson,E. (1996) Application of maximum likelihood refinement in the refinement of protein structures. *Proc. Daresbury Study Weekend*, 93–104 .
28. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
29. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
30. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
31. Hubbard,S.J. and Thornton,J.M. (1993) 'NACCESS' Computer Program. *Department of Biochemistry and Molecular Biology*. University College London, London.
32. 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.
33. Kraulis,P. (1991) Molscript-a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.
34. 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.
35. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
36. Stuckey,J.A. and Dixon,J.E. (1999) Crystal structure of a phospholipase D family member. *Nat. Struct. Biol.*, **6**, 278–284.
37. 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.
38. 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.
39. Uyen,N.T., Park,S.Y., Choi,J.W., Lee,H.J., Nishi,K. and Kim,J.S. (2009) The fragment structure of a putative HsdR subunit of a type I restriction enzyme from Vibrio vulnificus YJ016: implications for DNA restriction and translocation activity. *Nucleic Acids Res.*, **37**, 6960–6969.
40. Kim,J.S., DeGiovanni,A., Jancarik,J., Adams,P.D., Yokota,H., Kim,R. and Kim,S.H. (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.