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

Dmitrij Golovenko, Elena Manakova, Linas Zakrys, Mindaugas Zaremba, Giedrius Sasnauskas, Saulius Gražulis and Virginijus Siksnys*

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

Received November 15, 2013; Revised and Accepted December 11, 2013

ABSTRACT

The B3 DNA-binding domains (DBDs) of plant transcription factors (TF) and DBDs of EcoRII and BfiI restriction endonucleases (EcoRII-N and BfiI-C) share a common structural fold, classified as the DNA-binding pseudobarrel. The B3 DBDs in the plant TFs recognize a diverse set of target sequences. The only available co-crystal structure of the B3-like DBD is that of EcoRII-N (recognition sequence 5'-CCTGG-3'). In order to understand the structural and molecular mechanisms of specificity of B3 DBDs, we have solved the crystal structure of BfiI-C (recognition sequence 5'-ACTGGG-3') complexed with 12-bp cognate oligoduplex. Structural comparison of BfiI-C–DNA and EcoRII-N–DNA complexes reveals a conserved DNA-binding mode and a conserved pattern of interactions with the phosphodiester backbone. The determinants of the target specificity are located in the loops that emanate from the conserved structural core. The BfiI-C–DNA structure presented here expands a range of templates for modeling of the DNA-bound complexes of the B3 family of plant TFs.

INTRODUCTION

Structural studies of plant transcription factors (TFs) and type II restriction endonucleases (REases) revealed an unexpected link between these seemingly unrelated protein families. Despite the lack of sequence similarity, the B3 DNA-binding domain (DBD) of plant TFs (1,2), the effector domain of EcoRII REase [EcoRII-N (3)] and the DBD of BfiI REase [BfiI-C (4)] share a common fold, classified as the DNA-binding pseudobarrel (SCOP number 101935). The conserved core of these proteins, referred here as ‘B3-like domains’, is comprised of a barrel-shaped 7-stranded β-sheet capped by α-helices at both ends (5,6) (Figure 1). The genes encoding B3-like domains are widespread in plant kingdom from green algae to flowering plants. For example, in Arabidopsis alone there are 118 B3 domain containing proteins that often are involved in the hormone response pathways (1).

An atomic view of DNA recognition by the B3-like proteins was provided by the crystal structure of EcoRII-N–DNA complex (7). In EcoRII-N the β2, β5 and β4 β-strands, located at one edge of the pseudobarrel, the α-helix α2, and the connecting loops make a wrench-like cleft that approaches DNA from the major groove side (7). The DNA-binding interface contains two ‘recognition arms’ which interact with different parts of the 5'-CCTGG-3' target site: the N-arm (strand β2 and helix α2) contacts the 5'-terminal part of the recognition sequence, while the C-arm (strands β5 and β4) interacts with the 3’-terminal part of the recognition site. Structural comparison of other B3-like domains in the DNA-free form indicates a conserved structural core and a wrench-like DNA-binding cleft; moreover, a number of positively charged EcoRII residues involved in DNA-backbone interactions are conserved in this cleft (4,7). Taken together, these findings suggest that the DNA-binding mode identified in EcoRII-N–DNA crystal structure is conserved in other B3-like domains (7). Analysis of the DNA-binding interface of the Arabidopsis thaliana B3-family TFs RAV1 (8) and VRN1 (9) supports this hypothesis.

*To whom correspondence should be addressed. Tel: +370 5260 2108; Fax: +370 5260 2116; Email: siksnys@ibt.lt

Present addresses:
Dmitrij Golovenko, Department of Structural Biology, The Weizmann Institute of Science, 76100 Rehovot, Israel.
Linas Zakrys, Thermo Fisher Scientific, Graičiūnai 8, LT-02241, Vilnius, Lithuania.

© The Author(s) 2014. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
The B3-like domains display remarkable plasticity in terms of the target specificity: they interact with non-specific, pentanucleotide and hexanucleotide DNA sequences (Table 1). The crystal structure of EcoRII-N provided a structural mechanism for the pentanucleotide sequence recognition, however it remained to be established how B3-like domains adapt a conserved structural fold for the recognition of different DNA sites. In order to understand the structural and molecular mechanisms of specificity of B3 DBDs, we have solved the crystal structure of the C-terminal domain (BfiI-C) of BfiI REase from *Bacillus firmus* complexed with a 12-bp cognate oligoduplex. BfiI-C belongs to the B3-like domain family and interacts with the hexanucleotide sequence 5'-ACTGGG-3', which partially overlaps with the EcoRII-N site 5'-CCTGG-3' (the overlapping base pairs are underlined). Comparison of BfiI-C–DNA and EcoRII-N–DNA complexes revealed a structurally conserved pattern of phosphodiester backbone contacts and conserved amino acid residues in the DNA-binding cleft. Taken together, the BfiI-C–DNA structure presented here reveals how a conserved structural core is adapted for the recognition of variable DNA sequences and expands the template range for modeling of the DNA-bound complexes of the B3 family of plant TFs.

**MATERIALS AND METHODS**

**DNA oligoduplexes**

HPLC-grade oligonucleotides were purchased from Metabion (Martinsried, Germany). The oligoduplex 12/12 (top/bottom strand sequences 5'-AGCAGTGGGTCG-3'/3'-TCGTTAGCAG-5', respectively, the BfiI site underlined) for generation of the minimal BfiI-C–DNA complex was assembled as described in (7). Prior to annealing, the bottom strands of oligoduplexes 16SP (5'-AGCGTGAGCCGGGCT-3'/3'-TCGCA-5') and 16NS (5'-AGCGTAGCAGCCGG-3'/3'-TCGGGCGA-5') were radiolabeled using [γ-32P]ATP (Hartmann Analytic, Braunschweig, Germany) and T4 polynucleotide kinase (Thermo Fisher Scientific, Vilnius, Lithuania).

**Preparation of the BfiI-C domain**

BfiI mutant K107A was expressed in *Escherichia coli* strain ER2566 carrying plasmids pET21b-BfiIR6.5-K107A and pBfiIM9.1 and purified as described previously (15). Protein concentration was determined by measuring the absorbance at 280 nm and are expressed in terms of dimer using the extinction coefficient of 99700/M/cm (calculated using the ProtParam tool, http://web.expasy.org/protparam/). The BfiI-C–DNA complex was obtained by limited proteolysis of the full-length BfiI–DNA complex. The full-length K107A BfiI mutant was mixed with oligoduplex 12/12 at the molar protein dimer/DNA ratio 1:2.2 [the stoichiometry of the BfiI-DNA complex is 1:2 (15)] in a buffer containing 10 mM Tris–HCl (pH 7.5 at 25°C), 200 mM KCl, 1 mM DTT, 2 mM calcium acetate and 10% glycerol. Thermolysin was added at 1:100 w/w protease/protein ratio. Limited proteolysis was performed at 37°C for 20h and reaction terminated as described in (14). The reaction mixture was diluted two-fold with 10 mM Tris–HCl (pH 7.5 at 25°C) to reduce the salt concentration and loaded onto a 1-ml Heparin Sepharose column.
NACCESS (24). Protein–DNA contacts were analysed by buried between the two molecules were calculated using PDB ID 3ZI5. DNA geometry in BfiI-C–DNA structure was easily distinguishable in both duplexes within the crystallographic asymmetric unit (Supplementary Figure S1).

Crystallization, data collection and structure determination

BfiI-C complex with oligoduplex 12/12 was concentrated with 3 kDa MW cut-off. Crystals were grown by the crystallization, data collection and structure determination method (14) as a reference.

Dialyzed against 10 mM Tris–HCl (pH 7.5 at 25°C), 100 mM KCl, 1 mM DTT, 10% glycerol and stored at 4°C. Concentration of BfiI-C in the complex was determined by Bradford assay using the apo BfiI-C domain that was purified previously (14) as a reference.

Both models were subjected to 20 cycles of positional replacement using BfiI-C residues 200–347 from the gene. The proteins were purified using Ni2+-charged Ni2+-charged HiTrap HP and HiTrap Heparin HP columns (GE Healthcare) as described in (15) to a very low expression level. The K340A mutant was not isolated due to a very low expression level.

Electrophoretic mobility shift assay

Table 2. Diffraction data and structure refinement statistics

| Data collectiona | Spacegroup |
|-----------------|------------|
| Spacegroup | P63 |
| Unit cell | a = 175.18, b = 175.18, c = 35.79 Å, α = γ = 90.0°, β = 120° |
| Resolution, Å (final shell) | 50.57–3.2 (3.37–3.2) |
| Reflections unique (total) | 10809 (7556) |
| Completeness (%) overall (final shell) | 100 (100) |
| I/σI overall (final shell) | 13.3 (7.2) |
| Rmerge overall (final shell)b | 0.147 (0.240) |
| R(iso) from Wilson, (Å²)c | 24.4 |
| Refinementc | Resolution range (Å) |
| Number of protein atoms | 2642 |
| Number of DNA atoms | 972 |
| Rcryst (Rfree) | 0.176 (0.226) |
| RMS bonds (Å)/angles (°) | 0.003/0.743 |
| Average B factors (Å²), total | 49.1 |
| Main chain | 49.9 |
| Side chains | 49.6 |
| Solvent | 28.8 |
| Ramachandran plot | Favored |
| Allowed | 3.96% |
| Outliers | 0.00% |

4Dataset was collected at 100 K
5Rmerge = ΣhΣkΣl(Ihkl − ⟨I⟩)/ΣhΣkΣlI0, where I0 is an intensity value of i-th measurement of reflection h, h = (h,k,l), sum Σh runs over all measured reflections, and ⟨I⟩ is an average measured intensity of the reflection h. Number nh is a number of measurements of reflection h.
6Test set size 9.6%.

NUCPLOT (25). Graphical representations were produced with MolScript (26,27).

BfiI mutants

His-tagged full-length BfiI mutant variants were generated by QuickChange (28) or ‘megaprimer’ methods (29) using the plasmid pET15b-BfiIR6.5 as the template for PCR. The E. coli strain ER2267 carrying the plasmid pBfiIM9.1 was used as a transformation host. The mutations were confirmed by DNA sequencing of the entire gene. The proteins were purified using Ni2+-charged Ni2+-charged HiTrap HP and HiTrap Heparin HP columns (GE Healthcare) as described in (15) to >90% (~50% for the Y227A) homogeneity as estimated by SDS-PAGE. Concentrations of the WT and mutant proteins were determined as above using extinction coefficients for each mutant calculated by the ProtParam tool (http://web.expasy.org/protparam/) and are expressed in terms of dimer. The K340A mutant was not isolated due to a very low expression level.

Electrophoretic mobility shift assay

DNA binding by the WT BfiI and mutants was analysed by the electrophoretic mobility shift assay (EMSA) using the 16-bp specific and non-specific DNA duplexes 16SP and 16NS, respectively. DNA (final concentration 1 nM) was incubated with BfiI (final concentrations varied from 1 to 1000 nM of dimer) for 10 min in 20 μl of the binding buffer containing 40 mM Tris–acetate (pH 8.3 at 25°C),
Overall structure of the BfiI-C–DNA complex

The asymmetric unit of the crystal contains two BfiI-C–DNA complexes (Supplementary Figure S2A). Despite the relatively large contact surface between the protein chains (~1000Å²), this interaction seems to be biologically irrelevant, since isolated BfiI-C is a monomer that binds only one DNA copy (14). Both monomers in the asymmetric unit are nearly identical and can be superimposed with an RMSD of 0.13Å including protein side chains. The DNA bound in the complex is B–DNA. The secondary structure of the BfiI-C binding cleft includes a continuous left-handed pseudohelix due to the stacking interactions between the symmetry-related oligoduplexes (Supplementary Figure S2B). The secondary structure of the DNA-bound BfiI-C domain is similar to that of apo-BfiI-C (PDB ID 2C1L, residues 193–358) with the only structural elements, namely, the N-loop connecting β-strands β12 and β13 (residues 245–252, Figure 2A), and the C-loop connecting the β-strands β18 and β19 (residues 335-341, Figure 2A).

BfiI-C approaches DNA from the major groove and makes an extensive set of contacts with the DNA backbone (Supplementary Table S1 and Figure S3A). More specifically, a number of residues in the N-arm and the R272 residue in the C-arm make contacts to the phosphates on the top DNA strand of the target site, while the residues in the C-loop and within (or in the vicinity) of the N-arm interact with the phosphates on the bottom strand.

The sequence-specific BfiI-C–DNA interactions are provided by the amino acid residues located in the N- and C-arms. Residues in the N-arm are involved in hydrogen bonding interactions with the bases at the 5'-end of the site (green box in Figure 2B), whereas the residues in the C-arm make specific contacts with the bases at the 3'-end of the site (orange box in Figure 2B). In total, BfiI-C employs nine amino acid residues to make H-bonds and vdW contacts to the bases of the recognition site in the major groove (Figure 2C), and uses a sole R247 residue located in the N-loop for minor groove interactions.

Comparison of apo- and DNA-bound forms of BfiI-C

The conformational changes of BfiI-C occurring upon DNA binding are limited to the N-arm, N-loop and C-arm regions (Supplementary Figure S4). The loop in the N-arm (residues 216–223) moves towards the DNA backbone and brings the DNA-recognition residue T225 closer to the first two base pairs of the recognition site (Figure 2C). Upon DNA-binding unstructured region (amino acid residues 232–237) folds into a 310 helix and the R247 makes an extensive set of contacts with the DNA backbone (Supplementary Figure S4).

The β-strands β14-15 of the C-arm move in respect to the β-barrel core by pulling away from the β18 strand. This conformational change brings the N280 residue ~5Å closer to the DNA bases and enables a base-specific contact to the penultimate C base. Significant conformational change upon DNA binding occurs also in the N-loop (residues 245–252), which is shifted towards DNA along with the C-arm (Supplementary Figure S4). As a result, the R247 residue moves ~7.5Å towards DNA making a H-bond to the 5'-terminal adenine from the minor groove side (Figure 2C).

Mutational analysis of DNA-contacting residues

BfiI-C–DNA co-crystal structure was solved at the relatively low 3.2Å resolution, therefore DNA-binding residues predicted by the structural analysis were subjected to the site-directed mutagenesis to verify their functional importance (Table 3). Three sets of amino acid residues were analysed: (i) residues that make direct H-bonds or non-polar contacts to the bases of the recognition site in the major and the minor grooves (Figure 2C); (ii) positively charged amino acid residues and N245 that
interact with the DNA phosphates (Supplementary Table S1); (iii) residues T252 and Q226 that are conserved in the BfiI family enzymes (Supplementary Figure S5) and overlap with EcoRII-N residues Q37 and N61 involved in DNA binding. The selected residues were replaced by alanine in the context of full-length BfiI, mutant proteins were isolated and DNA binding and cleavage abilities evaluated by EMSA (representative data provided in Supplementary Figure S6) and λ DNA cleavage assay, respectively. Mutations with respect to their effect on DNA binding and cleavage clustered into three groups: (i) ‘unimportant’ mutations that lead to decrease of the cognate DNA-binding affinity ($K_{d}$) or specific cleavage activity up to 10-fold; (ii) ‘important’ mutations that

Figure 2. DNA recognition by BfiI-C and EcoRII-N. (A) The view of the BfiI-C–DNA complex along the long DNA axis (left) and the side view (right). The DNA-recognition site is colored dark grey. The secondary structure elements of the N-arm (α-helix α6, β-strand β11) and the C-arm (β-strands β14 and β15) are colored green and orange, respectively. Spheres of the matching colors represent the Cα atoms of the DNA-recognition residues from the N- and C-arms. The additional DNA-recognition element N-loop (residues 245–252) is colored blue and the C-loop (residues 335–341) is red. A region of the top DNA strand (nucleotides A4-G7) and adjacent recognition residues are shown against their mFo-DFC SIGMAA-weighted-electron density contoured at 2.0 σ level. (B) The sequence and numbering of the cognate 12/12 oligoduplex used in this study. DNA bases that interact with the N- and C-arms are boxed in green and orange, respectively. (C) Recognition of individual base pairs by BfiI-C. Panels for the individual base pairs are arranged following the top strand ACTGGG in the 5'-3' direction. The N- and C-arm residue labels are colored as in panel (A). (D and E) The sequence and numbering of the cognate EcoRII-N oligoduplex and the recognition of individual base pairs by EcoRII-N [PDB ID 3HQF (7)]. The residue labels and boxes encircling EcoRII-N sequence elements are colored as in panels (B and C).
lead to decrease of the cognate DNA-binding affinity ($K_d$) or specific cleavage activity by 1–2 orders of magnitude; and (iii) ‘essential’ mutations, that lead to decrease of the cognate DNA-binding affinity ($K_d$) or specific cleavage activity by more than two orders of magnitude (Table 3).

Nearly all N- and C-arm residues that make direct contacts to the DNA bases are either essential or important (Figure 2C and Table 3) and also show a high degree of conservation in the BfiI family (Supplementary Figure S5). Two of these residues, namely Y227 and D282, are involved in H-bond interactions with the N4 amino groups of cytosines that become methylated because of the missing connectivity between two strands.

Table 3. Mutational analysis of BfiI-C residues

| Mutated residue | Location | Contact with DNA | DNA-binding ability (%) | Impact on DNA binding | Specific activity, (%) | Impact on DNA cleavage |
|-----------------|----------|-----------------|------------------------|-----------------------|-----------------------|------------------------|
| R212 N-arm      | H-bonds to G8 (bottom) and A7 | 2 | Important | <0.5 | Essential |
| Q223 N-arm      | vdW contact to T9 | 100 | Unimportant | 100 | Unimportant |
| T225 N-arm      | H-bonds to A4 and G8 (bottom) | 10 | Important | 10 | Important |
| Y227 N-arm      | H-bond to C5 (top) | 5 | Important | <0.5 | Essential |
| W229 N-arm      | vdW contact to C6 | 1 | Essential | 3 | Important |
| R247 N-loop     | H-bond to A4 | <0.4 | Essential | 20 | Unimportant |
| E276 C-arm      | vdW contact to T6 | 10 | Important | 1 | Essential |
| N279 C-arm      | H-bond to G7 | <0.2 | Essential | <0.5 | Essential |
| N280 C-arm      | H-bonds to G8 (top) and C4 | <0.2 | Essential | <0.5 | Essential |
| D282 C-arm      | H-bonds to C5 (bottom) and C6 | <0.2 | Essential | <0.5 | Essential |
| R284 C-arm      | H-bonds to T6 and G7 | <0.2 | Essential | <1 | Essential |

Contacts with the DNA phosphates

| N245 N-loop | 5'-ApCTGG3'-3' | 100 | Unimportant | 20 | Unimportant |
| K250 N-loop | 5'-ATCTGGG-3' | 2 | Important | 10 | Important |
| R272 C-arm | 5'-pACTGGG-3' | 1 | Essential | 1 | Essential |
| R291 Close to C-arm | 5'-pNACTGGG-3' | 2 | Important | 5 | Important |
| K340 C-loop | 3'-TGACCpC-5' | n.d. | n.d. | n.d. | n.d. |

Other residues conserved in BfiI-C-related proteins

| Q226 N-arm | n.d. | 20 | unimportant |
| T252 N-loop | n.d. | 100 | unimportant |

*The DNA-binding ability is expressed in percent (%) as the ratio of the protein–DNA association constant $K_d$ of full-length BfiI mutants relative to the $K_d$ of WT BfiI dimer ([5 x 10^5]/M).

*The specific activity of BfiI mutants is expressed in percent (%) relative to the activity of WT BfiI.

*Due to very low expression level, the BfiI K340A mutant could not be purified.

Residues Q226 and T252, though they do have structural counterparts in BfiI family enzymes and EcoRII, are not important for BfiI (Table 3).

DISCUSSION

Close structural similarity between B3 domains of plant TFs and the effector domain of EcoRII REase suggests that B3 domains either diverged from a common ancestor, or were horizontally transferred into higher plants from symbiotic or pathogenic bacteria (1,2,33). Target sites for plant TFs of the B3 superfamily vary both in length and sequence (Table 1), suggesting that the DNA-binding pseudobarrel fold exhibits a structural plasticity that enables recognition of different DNA sequences within a conserved structural core. The molecular mechanisms of the target site specificity of the plant TFs yet has to be established since 3D structures of plant B3 domains are solved in the absence of DNA [PDB IDs 1WID (5), 4I1K (9), 1YEL (6)]. The crystal structure of the EcoRII effector domain bound to the oligoduplex containing a 5-bp target sequence provided a first structural template for modeling of DNA complexes of plant TFs (7,9). The BfiI-C–DNA structure presented here reveals for the first time the structural mechanism of the B3-like domain adaption for a different DNA target and provides a new template for the modeling of plant TFs.

Conserved DNA orientation in the B3-like domains

The fold of B3-like domains is termed a ‘pseudobarrel’ because of the missing connectivity between two strands.
Figure 3. Sequence and structure elements involved in protein–DNA interactions in B3-like domains. (A) Structure-based multiple sequence alignment of BfiI-C (PDB ID 2C1L, chain A), EcoRII-N (PDB ID 3HQF), RAV1-B3 (PDB ID 1WID, model 1), VRN1-B3 (PDB ID 4I1K, chain A) and At1g16640-B3 (PDB ID 1YEL, model 1) was generated by MultiProt and Staccato (31). Residues and secondary structure elements are numbered according to the BfiI-C–DNA structure. BfiI-C DNA-binding elements: N-arm, N-loop, C-arm and C-loop are marked by green, blue, orange and red stripes, respectively. The (D/E)XR motif of BfiI-C and EcoRII-N responsible for recognition of the 5'-TGG-3' trinucleotide is marked by black boxes and black circles. Residues contacting the 'clamp' phosphates are marked by magenta boxes and asterisks. The figure was generated using ESPRIT (32). (B) Interaction of B3-like domains with DNA: BfiI-C (this study), EcoRII-N [PDB ID 3HQF, (7)] and RAV1-B3 [PDB ID 1WID, interactions with DNA according to Yamasaki et al. (5,8)]. The bases in the recognition sites are colored yellow; orange, green, blue and red islands encircle residues from the N-arm, C-arm, N-loop and C-loop, respectively. Residues contacting the 'clamp' phosphates are marked by magenta boxes and asterisks. The figure was generated using ESPRIT (32).
(β10 and β11 in BfiI-C and β1 and β2 in EcoRII-N) in the otherwise barrel-shaped 7-stranded β-sheet, which in the case of BfiI-C and EcoRII-N overlap with a RMSD <2Å for 75 Cx atoms [calculated using MultiProt (31)]. The ‘open’ edge of the β-sheet forms the concave wrench-like surface that fits into the DNA major groove and provides base-specific contacts. The spatial positioning of the ‘pseudobarrel’ in respect to the DNA long-helical axis in the BfiI-C–DNA and EcoRII-N–DNA structures is conserved (Supplementary Figure 3A and B), suggesting a similar DNA orientation in respect to the protein core for other B3-like and B3 domains, including RAV1-B3 [PDB ID 1WID, (5)], At1g14660-B3 [PDB ID 1YEL, (6)] and VRN1-B3 [PDB ID 411K, (9)]. A recent high-resolution crystal structure and DNA-binding analysis of plant TF VRN1 are consistent with a similar DNA-binding mode (9).

Plasticity of the DNA-binding interface of the B3-like domains enables recognition of different sequences

The crystal structure of BfiI-C presented here and EcoRII-N structure solved by us earlier (7) provide a unique opportunity to compare structural and molecular mechanisms of sequence recognition by two B3-like proteins interacting with target sites of different length and sequence. BfiI-C is specific for the asymmetric 6-bp sequence 5'-ACTGGG-3', while EcoRII-N recognizes a partially overlapping pseudopalindromic 5-bp sequence 5'-CCWGG-3' (W = A or T, the overlapping bases are underlined). Most of the sequence specific and DNA backbone contacts in BfiI-C and EcoRII-N are made by residues located in the loops that extend from the concave β-sheet. BfiI-C and EcoRII-N share two conserved structural elements called N- and C-arms that make base-specific contacts to DNA in both proteins. The N-arms of both proteins contribute amino acid residues for the specific interactions with bases at the 5'-half of the recognition site, while the residues located in the C-arm contacts the bases in the 3'-half of the target site (Figure 2B and D). Intriguingly, BfiI-C and EcoRII-N bind their partially overlapping recognition sites in the same orientation. Moreover, though BfiI-C and EcoRII-N recognize the first C:G base in the overlapping region differently, both proteins make a similar set of contacts to the 5'-TGG-3' trinucleotide. For example, both proteins use a structurally conservative C-arm arginine (R284 in BfiI-C, R98 in EcoRII-N) to make a H-bond to the O4 atom of the T base, and a conserved carboxylate (D282 in BfiI-C, E96 in EcoRII-N) to make a H-bond to the N4 atom of the cytosine in the last G:C base pair (Figure 2C and E).

Most of the sequence-specific contacts in BfiI are provided by amino acid residues located in the N- and C-arms. Interestingly, the N-arm of BfiI-C is 8 and 11–13 amino acid residues longer than structural equivalents in EcoRII-N and other B3-like domains, respectively (Figure 3A). Despite of the length differences, the N-arm of BfiI-C makes a similar number of base-specific contacts as the N-arm of EcoRII-N (Figure 3B). The extra part of the loop connecting the α-helix x6 and β-strand β11 in the BfiI-C N-arm points away from DNA and does not contribute to the interactions with DNA. Presumably, the length of the N-arm in B3 domains is also sufficient to secure base-specific contacts to DNA (Figure 3).

BfiI-C also contains additional structural elements, namely N- and C-loops, that are involved in interactions with DNA. Amino acid residues in the N-loop contribute to the base-specific interactions at the 5'-end of the target site while amino acid residues in the C-loop make DNA–backbone contacts at the 3'-end of the recognition sequence. The protein–DNA interface area in the BfiI-C–DNA complex therefore is larger than in the EcoRII-N–DNA complex [2800Å² versus 2200Å² (7)]. The structural equivalents of the N- and C-loops in EcoRII-N, and the N-loop in B3 domains from plant TFs are shorter than in BfiI-C and therefore cannot make direct contacts to DNA (Figure 3A).

The BfiI-C recognition sequence is extended to 6 bp by an extra G:C base pair at the 3'-terminus in the respect to the 5-bp EcoRII-N target. Surprisingly, BfiI-C recognizes the extended recognition site using a 5 residues shorter and more compact C-arm (Figure 3A) than EcoRII-N. The side chains of DNA-facing amino acids on the C-arm loop are shorter in BfiI-C than structurally equivalent residues in EcoRII-N (D282 and N280 in BfiI-C versus E96 and R94 in EcoRII-N, Figure 4). BfiI-C positions this compact cluster of amino acids close to the bases at the 3'-end of the recognition site 5'-ACTGGG-3' and makes direct contacts to all three 3'-terminal G:C base pairs (Figures 3B and 4). In contrast, equivalent residues in EcoRII-N (R94 and E96) make direct contacts only to the terminal G:C base pair (underlined) within the 5'-CCTGGG-3' target sequence. In summary, BfiI-C trades two non-specific contacts made by the C-arm of EcoRII-N for a direct base-specific H-bond (N280 main chain oxygen to the N4 atom of the terminal C base in the bottom-strand). The loss of the C-arm and DNA–backbone contacts in BfiI-C is compensated by an increased number of non-specific contacts coming from other structural elements (Figure 3 and Supplementary Table S1). The C-arm in RAV1-B3 is only 1 amino acid shorter than in BfiI-C and would be consistent with specific binding of RAV1-B3 to a 6-bp recognition site (Figure 3). On the other hand, the non-specific DNA-binding protein VRN1-B3 (9) has a C-arm that is 3 residues shorter than in BfiI-C. It would be interesting to see whether the length of N- and C-arms of B3-like domains correlate with the DNA-binding specificity and recognition sequence length.

Conservation of DNA–backbone contacts

Analysis the DNA–backbone contacts in the BfiI-C and EcoRII-N complexes revealed conserved interactions with phosphate groups, referred here as ‘clamp’ phosphates, at the 5'-ends of the top and the bottom DNA strands (Figure 3B and Supplementary Figure S3). The C-arm of BfiI-C is fixed to the 5'-pACTGGG-3' phosphate group via a positively charged side chain of R272 residue (Figure 3B), that is spatially equivalent to the EcoRII-N R81 residue that interacts with the top strand phosphate 5'-pCCTGGG-3'. The impaired DNA-binding
affinity of the R272A BfiI mutant (Table 3), and spatial conservation of this arginine residue in all available B3-like domain structures (Figure 3A) underscores importance of this backbone contact for all B3-like domains. BfiI lacks the direct structural equivalent to the EcoRII-N residue K23, which binds the bottom strand phosphate 3'-GGACCp-5' and is conserved in RAV1-B3, At1g16440-B3 (Figure 3A) and many other B3 domains (5). However, a similar DNA–backbone contact to the clamp phosphate 3'-TGACCp-5' on the bottom strand is provided by the BfiI-C residue K340, which resides on a different structural element (Figure 3 and Supplementary Figure S3). We propose that interactions of the conserved positively charged residues with ‘clamp’ phosphates on the top and bottom DNA strands help to position the DNA molecule in the binding cleft in the conserved orientation and promote formation of base-specific contacts. This observation is in line with the recent analysis of crystal structures of DNA-binding protein complexes (34).

CONCLUSIONS

The crystal structure of the BfiI REase DBD (BfiI-C) in complex with DNA provides a first glimpse into the mechanism of 6-bp sequence recognition by a B3-like protein. The BfiI-C–DNA structure confirms the conserved DNA-binding mode inferred previously for B3-like domains (7) and reveals a conserved set of non-specific and specific interactions with DNA. Two positively charged BfiI-C residues, namely K340 and R272, make contacts to the ‘clamp’ DNA phosphates at the opposite termini of the DNA-recognition site, thereby anchoring the wrench-like DNA-binding surface in the DNA major groove. Spatial conservation of these residues in majority of B3 domains implies a similar DNA ‘clamping’ mechanism. Furthermore, both BfiI-C and EcoRII-N use N- and C-arms for making base-specific contacts to their 6- and 5-bp recognition sequences, respectively. The amino acid residues located in the loops within the C-arms determine the length and specificity of the target site. The loops in the N- and C-arms of plant B3 domains show great variability in length and sequence consistent with the diversity of their DNA-recognition sequences. The BfiI-C–DNA structure presented here opens the way for modeling of DNA-bound B3 domains of plant TFs using a novel template.

ACCESSION NUMBERS

3Z15.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Arūnas Lagunavičius for sharing BfiI plasmids and data on R246A and R247A mutants. Authors acknowledge Egle Rudokienė for sequencing and Dr Zita Liutkevičiūtė for the mass spectrometry. Authors also thank Dr Garib Murshudov and Dr Haim Rozenberg for suggestions and help during the structure refinement.

FUNDING

Research Council of Lithuania [MIP-086/2011 to D.G. and E.M.]; European Union Research Potential Call FP7-REGPOT-2009-1 (245721 ‘MoBiLi’ project). Funding for open access charge: Research Council of Lithuania.

Conflict of interest statement. None declared.
REFERENCES

1. Swaminathan,K., Peterson,K. and Jack,T. (2008) The plant B3 superfamily. *Trends Plant Sci.*, 13, 647–655.
2. Romanel,E.A.C., Schrago,C.G., Couhago,R.M., Russo,C.A.M. and Alves-Ferreira,M. (2009) Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One*, 4, e5791.
3. 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.
4. 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. USA*, 102, 15797–15802.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Leslie,A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESRF-EAMCB Newsletter on Protein Crystallography*, 26, 27–33.
13. Leslie,A.G.W. (2006) The integration of macromolecular diffraction data *Acta Crystallogr. D Biol. Crystallogr.*, 62, 48–57.
14. Evans,P. (2006) Scaling and assessment of data quality *Acta Crystallogr. D Biol. Crystallogr.*, 62, 72–82.
15. 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.
16. Suzuki,M., Kao,C.Y. and McCarty,D.R. (1997) The conserved B3 DNA-binding protein, binds to bipartite recognition sequence specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl Acad. Sci. USA*, 102, 15797–15802.
17. Yamasaki,K., Kigawa,T., Inoue,M., Watanabe,S., Tateno,M., Yamasaki,T., Tamulaitiene,G., Huber,R. and Siksnys,V. (2005) Structure of the 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.
18. Barik,S. (1995) Site-directed mutagenesis by double polymerase chain reaction. *Mol. Biotechnol.*, 3, 1–7.
19. Sapranaukas,R., Sasnauskas,G., Lagunavicius,A., Vilkaitis,G., Lubyts,A. and Siksnys,V. (2000) Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of Salmonella typhimurium. *J. Biol. Chem.*, 275, 30878–30885.
20. Shatsky,M., Nussinov,R. and Wolfson,H.J. (2004) NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.*, 25, 4940–4945.
21. Huang,C.J. (2005) Site-directed mutagenesis by double polymerase chain reaction. *Mol. Biotechnol.*, 3, 1–7.
22. Gouet,P., Robert,X. and Courcelle,E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, 31, 3320–3323.
23. Yamasaki,K., Kigawa,T., Inoue,M., Watanabe,S., Tateno,M., Seki,M., Shinozaki,K. and Yokoyama,S. (2008) Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.*, 46, 394–401.
24. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, 79, 233–269.