Published online 3 September 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 19* ***6613–6624***
doi:10.1093/nar/gkp699

Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII

Dmitrij Golovenko, Elena Manakova, Giedre Tamulaitiene, Saulius Grazulis and Virginijus Siksnys*

Institute of Biotechnology, Graiciuno 8, LT-02241 Vilnius, Lithuania

Received June 12, 2009; Revised August 5, 2009; Accepted August 6, 2009

ABSTRACT

EcoRII restriction endonuclease is specific for the 5′-CCWGG sequence (W stands for A or T); however, it shows no activity on a single recognition site. To activate cleavage it requires binding of an additional target site as an allosteric effector. EcoRII dimer consists of three structural units: a central catalytic core, made from two copies of the C-terminal domain (EcoRII-C), and two N-terminal effector DNA binding domains (EcoRII-N). Here, we report DNA-bound EcoRII-N and EcoRII-C structures, which show that EcoRII combines two radically different structural mechanisms to interact with the effector and substrate DNA. The catalytic EcoRII-C dimer flips out the central T:A base pair and makes symmetric interactions with the CC:GG half-sites. The EcoRII-N effector domain monomer binds to the target site asymmetrically in a single defined orientation which is determined by specific hydrogen bonding and van der Waals interactions with the central T:A pair in the major groove. The EcoRII-N mode of the target site recognition is shared by the large class of higher plant transcription factors of the B3 superfamily.

INTRODUCTION

Orthodox Type IIP restriction endonucleases (REases) like EcoRI or EcoRV, are homodimers which interact with a single copy of the symmetrical (palindromic) recognition site and in the presence of Mg^{2+} ions cut phosphodiester bonds within their target site to generate a double-strand break (1). The Type IIE endonuclease EcoRII differs in two important respects from the orthodox REases. First, it recognizes a pentanucleotide sequence 5′-CCWGG (W stands for A or T) where the perfect target site symmetry is broken in the center. Second, it does not cut DNA that contains a single recognition site (2,3). To trigger DNA cleavage, EcoRII requires interaction with at least two copies of the recognition sequence, one being the actual target of cleavage, the other being the allosteric effector (4). While the interaction with two DNA sites is a prerequisite for EcoRII cleavage, simultaneous binding of three recognition sites is necessary for a concerted cleavage of both DNA strands at a single site (3). EcoRII synaptic complex involving three recognition sites has been recently visualized by atomic force microscopy (5).

The crystal structure of the EcoRII solved in the DNA-free form shows a homodimer (Figure 1B) comprised of two monomers (6). However, in contrast to the orthodox REases which generally are single domain proteins [except of the rare cutters SdaI (7) and NotI (8)], each monomer of EcoRII consists of two structural domains (6). The C-terminal domain (EcoRII-C) is built up of five stranded β-sheet which is arranged similarly to other REases of PD-(D/E)XK superfamily (9). The N-terminal domain (EcoRII-N) is arranged as pseudobarrel, and according to the SCOP database (10) is a founding member of a novel DNA binding fold which was identified later in the B3 family of higher plant transcription factors (11) and the BfiI restriction enzyme (12). Both EcoRII domains are connected by a linker which is cleaved by proteolytic enzymes to yield separate domains which perform different functions (13). Isolated EcoRII-N is a monomer which binds the effector 5′-CCWGG sequence but does not cleave it (13,14). The EcoRII-C is a dimer which functions as an orthodox Type IIP endonuclease and does not require the second copy of the target site for DNA cleavage (13,14). In fact, EcoRII-C follows the same reaction pathway as an orthodox Type IIP REase PspGI which recognizes the 5′-CCWGG sequence but lacks the effector DNA binding domain (14–16).

Here, we present crystal structures of the isolated N- and C-terminal domains of EcoRII in complex with cognate oligodeoxynucleotides containing the 5′-CCTGG sequence. Structural analysis demonstrates that the

*To whom correspondence should be addressed. Tel: +370 5 2602108; Fax: +370 5 2602116; Email: siksnys@ibt.lt

EcoRII restriction enzyme combines two different strategies to interact with the 5′-CCTGG sequence. The N-terminal effector domain monomer binds the 5′-CCTGG sequence asymmetrically in a single defined orientation which is determined by specific interactions with the central T:A pair in the major groove of B-form DNA. The catalytic EcoRII-C dimer uses a radically different mechanism: it flips out the central T:A base pair and makes symmetrical contacts with the CC:GG half-sites.

MATERIALS AND METHODS

EcoRII-N and EcoRII-C purification

Escherichia coli strain JM109 [pQE30 (R + M-), pDK1 (R-M +)] was used for the protein expression (4). The EcoRII-N and the EcoRII Y41A mutant proteins were expressed as the N-terminal $(His)_6$-tag fusions and purified as described (4). EcoRII-C was generated from EcoRII Y41A mutant by limited thermolysin proteolysis and purified as described (5). The protein concentration was determined by measuring the absorbance at 280 nm. Extinction coefficients 22 943 $M^{-1}\,cm^{-1}$ for the EcoRII-N monomer and 43 240 $M^{-1}\,cm^{-1}$ for the EcoRII-C dimer were calculated by the ProtParam tool at: http://www.expasy.ch/.

Oligoduplexes

The 2-aminopurine containing oligonucleotide was purchased from Integrated DNA Technologies, other oligodeoxynucleotides were from Metabion. Oligoduplexes for crystallization (Figures 1A and 2A) were assembled as described in (17). Oligoduplexes for binding and cleavage studies (Supplementary Table SI) were 5′-end-labeled and assembled as described in (18).

EcoRII-N–DNA complex preparation

His-tagged EcoRII-N monomer (13) and 9-bp oligoduplex (Figure 2A) were mixed, treated with trypsin and purified (details are provided in Supplementary Methods). EcoRII-N–DNA complex was concentrated to 18.5 mg/ml using Centricon (Millipore) concentrator with 3-kDa cutoff limit.

EcoRII-C–DNA complex preparation

Equimolar amounts of EcoRII-C dimer obtained by limited proteolysis as described in (5) and 12-bp oligoduplex (Figure 1A) were mixed in buffer 20 mM Tris–HCl (pH 8.0, 25°C), 0.4 M KCl, 10% glycerol, 0.1 mM EDTA and $CaCl_2$ was added to final 5 mM concentration. EcoRII-C–DNA complex was then concentrated to 4.6 mg/ml using Ultrafree-0.5 (Millipore) concentrator with 5-kDa cutoff limit.

Crystallization of EcoRII-N and EcoRII-C complexes with DNA

Crystals were grown by the vapor diffusion technique in sitting drops. EcoRII-N–DNA: 2.0 µl of the EcoRII-N–DNA solution was mixed with 0.5 µl of reservoir buffer A [1 M CH_3COONa (pH 4.5, 25°C), 0.2 M CH_3COOLi, 10% glycerol]. EcoRII-N–DNA crystals appeared after 3 days from a clear drop. Prior to flash-cryocooling, EcoRII-N–DNA crystals were transferred into the 18.5:3.5 mixture of the reservoir buffer A and PEG400. EcoRII-C–DNA: 0.5 µl of the EcoRII-C–DNA solution was mixed with 0.2 µl of reservoir buffer B (26% PEG1500, 25% glycerol). EcoRII-C–DNA crystals appeared after ~1 month from a clear drop. Crystals of EcoRII-C–DNA were flash-cryocooled without extra cryoprotection.

Data collection

EcoRII-N–DNA crystallized in the space group $P4_32_12$ with one monomer and one cognate oligoduplex in the asymmetric unit. EcoRII-C–DNA crystallized in space group $P2_12_12$ with one monomer and one DNA strand in the asymmetric unit. All diffraction data were collected at 100 K at X13 beamline of the DESY in Hamburg, Germany. All data were processed with the programs MOSFLM (19), SCALA (20) and TRUNCATE (21). The working data sets for EcoRII-N and EcoRII-C were to 2.5 Å and to 2.6 Å resolution correspondingly (Table 1).

Structure determination and refinement

The structures of the EcoRII-N–DNA and EcoRII-C–DNA complexes were solved by the molecular replacement method using the apo-EcoRII model (6) residues 4-180 and 181-402, respectively. The molecular replacement was done with programs AMORE (22) and MOLREP (23). Initial phases and electron density maps were calculated with SIGMAA (24) program using molecular replacement solutions followed by manual fitting using the program COOT (25). Model refinement was carried out with CNS (26) and REFMAC (27). Refinement proceeded in several cycles in combination with manual rebuilding. In the EcoRII-N complex 173 a.a. of the protein and DNA 9-mer were fully modeled in the electron density. In the asymmetric unit of the EcoRII-C complex 221 a.a. of protein are visible in the density along with the 12-mer oligonucleotide, which was modelled as two alternative strands with 50% occupancy each. The DNA parameter values were calculated with CURVES (28).

Gel mobility shift assay

Increasing amounts of EcoRII-N were mixed with ^{33}P-labeled oligoduplexes (Supplementary Table SI) at 2.0 nM concentration in the presence of 0.1 mM EDTA and samples were analysed in the polyacrylamide gel as described (14). In the case of EcoRII-C increasing amounts of protein were mixed with ^{33}P-labeled oligoduplexes (Supplementary Table SI) at 0.1 nM concentration in the presence of 5 mM Ca^{2+} and 2 mM DTT and mixtures were analysed in the polyacrylamide gel as described (14). Apparent K_d values were determined as described (14).

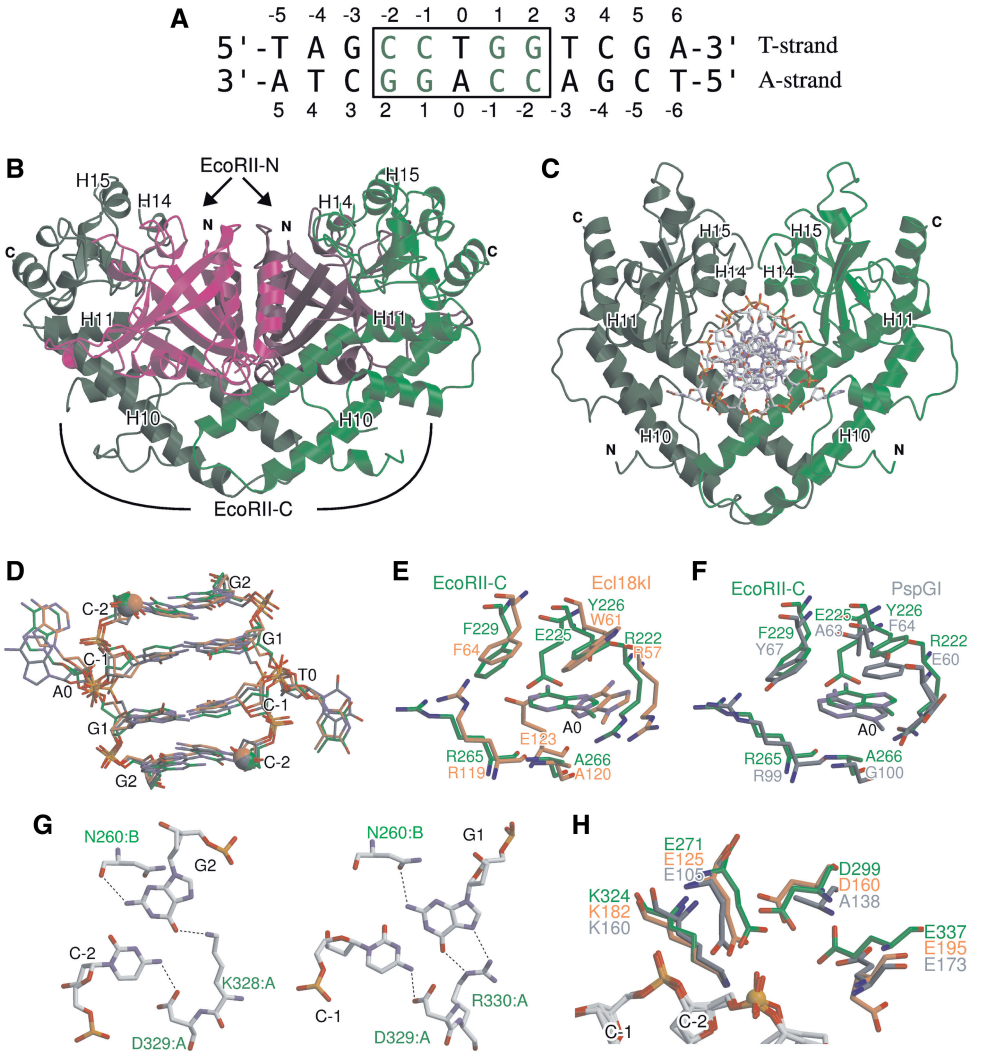

Figure 1. EcoRII-C–DNA complex structure. (**A**) Oligonucleotide used for crystallization of EcoRII-C. Central base pair is shown in black, recognition sequence is boxed, symmetric half-sites within the recognition sequence are shown in green letters, respectively. (**B**) Crystal structure of DNA-free EcoRII [PDBID 1NA6]. EcoRII-N domains of different subunits are shown in dark and light magenta, EcoRII-C domains in dark and light green, respectively. (**C**) Crystal structure of EcoRII-C–DNA complex. DNA is shown as a stick model. (**D**) Stick representation of the 5′-CCTGG sequence conformation in EcoRII-C (green), Ecl18kI (coral) and PspGI (grey) complexes. Superposition was obtained by aligning Ecl18kI-DNA complex structure [PDBID 2FQZ] with the EcoRII-C and PspGI complex [PDBID 3BM3] structures using TOP3D program (45). Scissile phosphates in the EcoRII-C–DNA, Ecl18kI–DNA and PspGI–DNA complexes are shown as green, coral and grey spheres respectively. (**E**) Superposition of binding pockets for the flipped nucleotide between EcoRII-C (green) and Ecl18kI (coral). (**F**) Superposition of binding pockets for the flipped nucleotide between EcoRII-C (green) and PspGI (grey). (**G**) Interactions of EcoRII-C with the CC:GG half-site of the 5′-CCTGG sequence. (**H**) Superimposed active sites of EcoRII-C (green), Ecl18kI (coral) and PspGI (grey). Dimers of Ecl18kI and PspGI were aligned on liganded EcoRII-C with TOP3D program. DNA fragments are shown in stick representation, scissile phosphate of EcoRII-C–DNA complex is shown as an orange sphere. Figures were generated with MOLSCRIPT.

DNA cleavage activity

The DNA cleavage activities of EcoRII-C were monitored using a 16 bp oligoduplex containing a ^{33}P-label at the 5′-end in one of the strands (Supplementary Table SI). Cleavage reactions were conducted at 25°C in the reaction buffer [33 mM Tris–acetate (pH 7.9 at 25°C), 66 mM potassium acetate, 10 mM magnesium acetate, 2 mM DTT, 0.1 mg/ml BSA] using 200 nM of oligoduplex and 1000 nM of protein. Aliquots were removed at timed intervals and quenched by mixing with loading dye [9% (v/v) formamide, 0.01% (w/v) bromphenol blue] before denaturing gel electrophoresis. The samples were analysed and quantified as described in (18).

RESULTS

DNA-bound structure of EcoRII-C

The EcoRII-C domain was obtained by limited proteolysis (see 'Materials and Methods' section) and co-crystallized with a cognate 12-bp oligonucleotide duplex (Figure 1A) containing the EcoRII recognition sequence. The structure was solved by molecular replacement using the

C-terminal domain of EcoRII [PDBID 1NA6] as a model and subsequently refined to 2.6-Å resolution (Table 1). A biological dimer was generated by a 2-fold axis rotation around crystallographic dyad which coincides with the complex dyad. Symmetry of protein–DNA interactions within the recognition site is broken only at the central bases. Due to this ambiguity the electron density for the central bases is an average of T and A. The crystallographic 2-fold axis maps one oligoduplex (Figure 1A) strand onto another but since such mapping is not exact, crystals clearly contain disorder. Namely, DNA in both possible orientations must be present in the crystal to ensure that the crystallographic twofold is satisfied on average. However, since protein pockets seem to accommodate equally well both A and T bases and in the oligoduplex (Figure 1A) the CC-GG recognition site part is symmetric with respect to the twofold which matches purines and pyrimidines in the flanking sequences of both strands, alternative DNA orientations in the crystal seem to be equally probable. Indeed, our attempt to process the data in $P2_1$ space group (i.e. without imposing the DNA 2-fold during processing) did not reveal any predominant DNA orientation. The secondary structure elements of the liganded EcoRII-C form are identical to those in the apo-EcoRII and therefore follow the same labelling (6).

Two EcoRII-C monomers are arranged into a clamp-like structure and completely encircle DNA (Figure 1C). The subunit interface of EcoRII-C dimer is bipartite and buries 2400 $Å^2$ per monomer. In the N-terminal dimer interface α-helices H10 and H11 of two subunits interact. In the C-terminal part intermolecular contacts are made by α-helices H14 and H15 located at the conserved catalytic core. The N-terminal interface is predominantly hydrophobic and buries 1600 $Å^2$ of the accessible surface area. The C-terminal interdomain contacts are dominated by hydrogen bond interactions and bury 840 $Å^2$ at the interface. The intersubunit contacts create a large bipartite DNA binding surface which occludes DNA and buries nearly 4900 $Å^2$ per dimer at the EcoRII-C–DNA interface.

EcoRII-C flips out the central T:A base pair within the 5′-CC(T/A)GG site

DNA in the complex with EcoRII-C has unusual conformation. The central T:A base pair is extruded from the DNA double helix (Figure 1C) followed by phosphodiester backbone compression which transforms the 5-bp sequence into the 4-bp one (Figure 1D). Similar DNA distortion is observed in the complexes of Ecl18kI and PspGI restriction enzymes (16,17). The conserved CC:GG dinucleotides within the 5′-CCNGG (Ecl18kI) and 5′-CCWGG (EcoRII-C/PspGI) recognition sites nearly superimpose (Figure 1D). The extra-helical nucleotides in EcoRII-C–DNA complex are overwound in comparison to Ecl18kI-DNA complex (Figure 1D and E) but spatially coincide with those of PspGI-DNA (Figure 1D and F). The nucleotide flips seen in the EcoRII crystal structure are in an excellent agreement with the solution studies which show the fluorescence enhancement of the 2-aminopurine placed at the center of the recognition sequence (29).

Each unstacked base is accommodated into the EcoRII-C pocket that is formed by Arg222, Glu225, Tyr226, Phe229 residues located on the α-helix H10 and Arg265, Ala266 residues located on the α-helix H11 (Figure 1E). Extra-helical bases are sandwiched between the aromatic ring of Tyr226 and side chain of Arg222. Structural comparison indicates that in the apo-EcoRII the binding pocket for the flipped out base is already preformed and binding of the extra-helical base is accompanied by minor changes of the side chain conformation of pocket residues (data not shown). The EcoRII-C, PspGI and Ecl18kI pockets for the flipped out nucleotides are strikingly similar (Figure 1E and F) raising a question how EcoRII-C/PspGI discriminate 5′-CC(T/A)GG targets from the 5′-CC(G/C)GG sites.

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

EcoRII-C achieves recognition of the symmetric CC:GG half-sites through the direct contacts of the consecutive Lys328, Asp329 and Arg330 residues located at the N-terminus of the symmetry-related α-helices H14 which are dubbed as recognition helices in analogy with other REases (30). Each monomer makes an identical set of contacts to the CC : GG half of the 5′-CC(T/A)GG sequence. Therefore, only one set of contacts to the CC : GG subsite will be discussed. The $O_\delta1$ and $O_\delta2$ carboxylate oxygens of Asp329 accept hydrogen bonds from N4 atoms of two neighboring cytosines (Figure 1G). The amino group of the Lys328 donates a hydrogen bond to O6 of the outer guanine, while Arg330 makes a bidentate hydrogen bond with O6 and N7 atoms of the inner guanine (Figure 1G). Main chain and $O_\delta1$ oxygen atoms of Asn260 residue located on the α-helix H11 of the neighboring monomer are within hydrogen bond distance to the N2 atoms of the inner and outer guanine residues in the minor groove completing the interaction network with the CC:GG half-site (Figure 1G). The structural model of sequence recognition is in agreement with mutational studies of Lys328, Asp329 and Arg330 residues (14). Not surprisingly, these residues are absolutely conserved in the EcoRII-family (Supplementary Figure S1).

In the EcoRII-C–DNA complex the Glu271, Asp299, Lys324 and Glu337 residues from both monomers are located in close proximity to the scissile phosphates on both DNA strands (Figure 1H). Mutational analysis is consistent with the active site function of these residues (14). Moreover, EcoRII-C active site residues are absolutely conserved across the EcoRII family (Supplementary Figure S1). Structural superposition of the EcoRII-C and Ecl18kI in DNA bound form reveals that the active sites residues (Figure 1H) as well as the scissile phosphates of the two complexes (Figure 1D and H) almost coincide. Despite the presence of Ca^{2+} in the crystallization buffer there is no electron density that could be interpreted as Ca^{2+} ion in the vicinity of the active site of EcoRII-C. In the PspGI, the side chain conformation of the Glu173

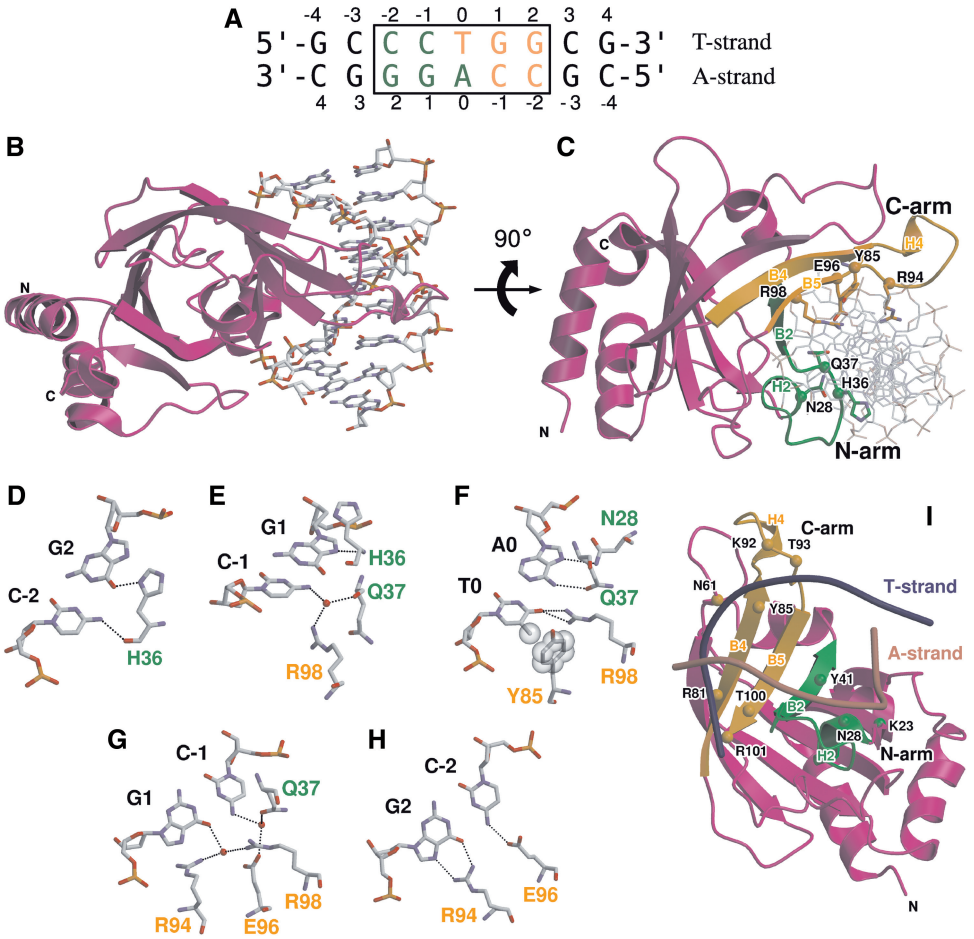

Figure 2. EcoRII-N–DNA complex structure. (**A**) Oligonucleotide used for crystallization of EcoRII-N. Recognition sequence is boxed, pseudosymmetric parts of the target site are shown in orange and green letters, respectively. (**B** and **C**) Two different views of EcoRII-N–DNA complex structure. The secondary structure elements of the N-arm that predominantly interact with CC:GG base pairs of the effector DNA are shown in green; those belonging to the C-arm and predominantly interacting with GG/CC are shown in orange. Residues and their C_α atoms are colored according to the corresponding arm color. (D–H) Recognition of the 5′-CCTGG sequence by EcoRII-N. Panels for individual base pairs are arranged following the T-strand in $5' \rightarrow 3'$ direction. Residues belonging to the N- and C-arms are labelled in green and orange, respectively. (**I**) EcoRII-N contacts to the DNA phosphodiester backbone. Protein C_α atoms of residues that interact with the phosphodiester backbone of DNA are depicted as spheres. Color labelling is consistent with Figure 2C. Residues from N-arm and K23 (shown in green) interact with the A-strand (salmon) while residues from C-arm and N61 (shown in orange) interact with the T-strand (blue).

differs from that of the structurally equivalent residues in EcoRII-C and Ecl18kI (Figure 1H). The active site environment of PspGI, however, may be perturbed due to the Asp138Ala mutation which was introduced to obtain diffracting crystals (16).

DNA sequence recognition by EcoRII-N

The EcoRII-N was crystallized with the 9-bp oligonucleotide duplex (Figure 2A) containing the target sequence (see 'Materials and Methods' section). The crystal contained one protein monomer and a single DNA duplex in the asymmetric unit. The structure was solved by molecular replacement using the N-terminal domain of the EcoRII structure [PDBID 1NA6] as a model (Table 1). The secondary structure of the DNA-bound EcoRII-N domain is identical to that of B chain (residues 4-172) in the apo-EcoRII structure (6) and follows the same labelling.

In the crystal, the EcoRII-N monomer is bound to one copy of the 9-bp oligoduplex (Figure 2A) where the central T:A base pair breaks the perfect 2-fold symmetry of the recognition site. The concave DNA binding cleft of the N-terminal domain resembles a wrench with a skewed U-shaped opening that grips DNA from the major groove side (Figure 2B and C). The N-arm (residues 27-42) of the wrench is formed by α-helix H2, β-strand B2, and a connecting loop, while two antiparallel β-strands (B4 and B5) and a connecting loop comprise the C-arm (residues 78-101) (Figure 2C). Both arms penetrate into the DNA major groove, and amino acid residues located on the arms interact with the base edges of the recognition sequence (Figure 2C). In the EcoRII-N–DNA complex structure ~2200 $Å^2$ of protein surface is buried at the interface.

The DNA in the EcoRII-N complex is not significantly distorted (Figure 2B) except that the major groove

Table 1. Diffraction data and structure refinement statistics

Data set	EcoRII-C	EcoRII-N
Temperature	100 K	100 K
Space group	$P2_12_12$	$P4_32_12$
$a =$, $b =$, $c =$, Å	77.1, 58.0, 61.0	43.2, 43.2, 253.6
Resolution, Å (final shell)	47.84–2.6 (2.74–2.60)	63.5–2.5 (2.64–2.50)
Reflections unique (total)	8835 (119 326)	9055 (62 633)
Completeness (%) overall (final shell)	99.8 (99.8)	99.1 (94.3)
I/σ_I overall (final shell)	13.0 (2.0)	6.9 (7.0)
R_{merge} overall[a] (final shell)	0.05 (0.33)	0.07 (0.08)
B(iso) from Wilson (Å^2)	71	38
Refinement statistics		
Number of protein atoms	1812	1334
Number of DNA atoms	486	363
Number of solvent molecules	67	153
Test set size	10%	10%
R_{cryst} (R_{free})	0.236 (0.293)	0.187 (0.229)
Rmsd bonds/angles	0.006/1.012	0.016/1.916
Average B-factors (Å)	45.7	15.0
Main chain	46.3	14.3
Side chains	48.0	16.9
DNA	39.6	13.3
Solvent	39.7	12.6

[a]$R_{merge} = \Sigma_h\Sigma_i \mid I_{hi} - < I_h > \mid / \Sigma_h\Sigma_1|_{hi}$, were $|_{hi}$ is the i-th observation of the reflection h, while $< I_h >$ is the mean intensity of reflection h.

becomes ~3.0 Å wider in comparison to the canonical B-form. EcoRII-N monomer interacts with the recognition sequence in a single defined orientation where amino acid residues located on the N-arm interact predominantly with the 'left' 5′-CC:GG half-site, while residues in the C-terminal arm make contacts to the 'right' 5′-CC:GG half-site (Figure 2A and C). Both arms contribute amino acid residues to the recognition of the central T:A base pair. Detailed interactions of the EcoRII-N residues in the DNA major groove are depicted in Figure 2D-H following the T-strand from the 5′-terminus (Figure 2A).

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

C-1:G1. A water molecule sandwiched by guanidino group of Arg98 and the main chain carbonyl oxygen of the Gln37 accepts a hydrogen bond from the N4 atom of the cytosine. The main chain nitrogen atom of His36 donates a hydrogen bond to the N7 atom of the guanine (Figure 2E).

T0:A0. The N6 atom of the adenine donates a hydrogen bond to the O_ε atom of Gln37 while the N7 atom accepts a hydrogen bond from the N_δ atom of Asn28 (Figure 2F). The $N_\eta1$ and $N_\eta2$ atoms of Arg98 make hydrogen bonds to the thymine O4 atom. The methyl group the thymine makes van der Waals contact with the aromatic ring of the Tyr85 residue.

G1:C-1. The guanidino groups of Arg94 and Arg98 sandwich the water molecule that donates a hydrogen bond to the O6 atom of guanine. The carboxylate of Glu96 and the O_ε atom of Gln37 sandwich the water molecule that accepts a hydrogen bond from the N4 atom of cytosine (Figure 2G).

G2:C-2. Arg94 nitrogen atoms $N_\eta1$ and $N_\eta2$ make hydrogen bonds to the N7 and O6 atoms of guanine. The carboxylate of Glu96 accepts a hydrogen bond from the cytosine N4 atom (Figure 2H).

Each arm of the wrench also makes a set of contacts with the phosphodiester backbone of one particular DNA strand (Figure 2I). Asn28 and Tyr41 from the N-arm interact with DNA phosphates of the A-strand, while Arg81, Tyr85, Lys92, Thr93, Thr100 and Arg101 from the C-arm interact with DNA phosphates of the T-strand.

Comparison of apo- and DNA-bound forms of EcoRII-N

The conformational changes of EcoRII-N occurring upon DNA binding are limited to the loop regions. The loop in the C-arm, which harbours Arg94 involved in sequence recognition, moves by 3 Å closer to the DNA but does not change its backbone conformation (Figure 3A). Therefore, the C_α atoms of the loop residues 86-95 in the DNA-free and bound forms superimpose with a RMSD of 0.25 Å. In contrast, the loop in the N-arm changes its conformation significantly upon DNA binding resulting in the RMSD value of 1.7 Å between DNA-bound and free forms for residues 31-38 (Figure 3A). These perturbations in the N-arm loop conformation bring Gln37 in the close proximity to the A0 base edge (Figure 3B). Noteworthy, the O_ε atom of Gln37 in the DNA bound form moves ~8 Å in respect to its position in the DNA-free form and makes a hydrogen bond with N6 atom of the central A0 nucleotide (Figure 3B).

Probing specificity determinants of the central T:A pair using base analogues

The van der Waals contact between the methyl group of the central T0 base and Tyr85 and hydrogen bonding interaction between the side chain oxygen of Gln37 and 6-amino group of A0 (Figure 2F) implies the asymmetric EcoRII-N binding in one preferable orientation. Theoretically, N- and C-'arm' interactions with the symmetrical CC:GG half-sites are compatible with an alternative binding orientation; however, the central T:A base pair recognition interface is not (Figure 2F). In the alternative binding mode, both the van der Waals contact between Tyr85 and T0-base methyl group and the hydrogen bond interactions between the side chain carbonyl atom of Gln37 and N6 amino group of A0 will be disrupted. To determine the importance of these interactions for the EcoRII-N specificity we analysed EcoRII-N binding to the modified oligoduplexes containing either uridine instead of T0 or 2-aminopurine instead of A0 (Figure 3C). Gel shift analysis demonstrates that

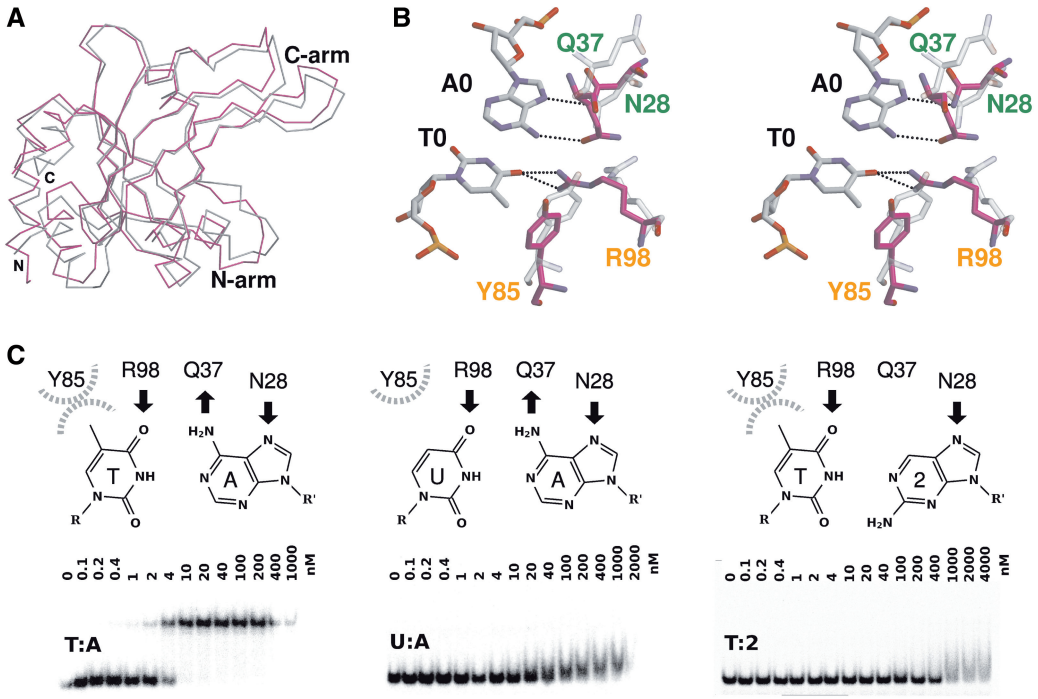

Figure 3. EcoRII-N–DNA interactions. (**A**) Superposition of apo (grey) and DNA bound (magenta) EcoRII-N domains. (B) Stereoview illustrating conformation of residues involved in the central T:A base pair recognition in apo (grey) and DNA bound (magenta) EcoRII-N. Color labelling of residues is consistent with Figure 2F. (C) EcoRII-N interactions with oligoduplexes containing modified central base pair. Binding reactions contained either the specific T:A or modified U:A and T:2 oligoduplexes (Supplementary Table SI), and the protein at increasing concentrations (see 'Materials and Methods' section for details).

elimination of the methyl group by T to U replacement or elimination of the N6 amino group in the case of A to 2-aminopurine replacement completely abolishes DNA binding by EcoRII-N (Figure 3C).

EcoRII-C and EcoRII-N sensitivity to DNA methylation

C5-methylation of the second C in the 5′-CCWGG sequence makes it resistant against the EcoRII REase cleavage (31,32). The crystallographic analysis presented above demonstrates that the N- and C-terminal domains of EcoRII use very different structural mechanisms to interact with the 5′-CCWGG target, raising the question of whether individual domains differ in their sensitivity to cytosine methylation. To find an answer we have analysed EcoRII-N and EcoRII-C binding to a set of oligoduplexes containing hemimetylated or double methylated 5′-CCWGG site (Supplementary Table SI). DNA binding experiments show that C5 methylation of the second C residue either in the A- or T-strand completely abolishes EcoRII-N binding (Supplementary Figure S2A). *In silico* analysis shows that conversion of the second cytosine to 5mC in the A-strand should displace the water molecule which is involved in specific hydrogen bond interactions with Gln37 (Figure 2G) and create a steric clash with the 5′-upstream C base. Introduction of 5mC in the T-strand should result in a steric clash with the oxygen atom of Tyr85 involved in the neighbouring base pair recognition (data not shown).

Previous studies revealed that methylation of the internal cytosine within the 5′-CCWGG sequence makes M13 phage DNA resistant for EcoRII-C cleavage (32). Our data show that EcoRII-C does not bind to the oligoduplex containing 5mC in both strands; however, it still binds to the hemimethylated 5′-CCWGG site. DNA bands in the gel, however, are smeared, suggesting that the complex stability is decreased (Supplementary Figure S2B). Oligoduplex cleavage studies show that methylation of both cytosines decreases EcoRII-C cleavage rate nearly by 5 orders of magnitude. Hemimethylation of the 5′-CCWGG site, however, affects cleavage rates of methylated and non-methylated strands differently (Supplementary Figure S2C). While the cleavage rate of methylated strand goes down by nearly 4 log orders, the nonmethylated strand is cleaved only 100–200-fold slower than the unmethylated oligoduplex. *In silico* analysis shows that introduction of 5mC should disrupt the recognition network at the EcoRII-C interface (data not shown). EcoRII-C cleavage rate differences of methylated and nonmethylated strands suggest that despite the symmetric arrangement of EcoRII-C dimer, the individual monomers act quite independently on separate DNA strands.

DISCUSSION

According to REBASE database (33) the sequence 5′-CCWGG represents one of the most common

recognition sequences for restriction-modification system. Indeed, out of the 3392 RM systems in the REBASE, 225 recognize 5′-CCWGG. REases specific for the 5′-CCWGG sequence fall into two groups which differ in the cleavage position (5′-CC/WGG or 5′-/CCWGG). EcoRII is a prototype of REases which cut before the first C in CCWGG site. Bioinformatics analysis of REases indicates that the EcoRII subfamily includes 45 members (34). Structural information for the EcoRII subfamily so far was limited to EcoRII endonuclease in the DNA-free form (6). The EcoRII dimer consists of three structural units (Figure 1B): a central catalytic core made from two copies of the C-terminal domain (EcoRII-C), and two N-terminal effector DNA binding domains (EcoRII-N). Structures of the EcoRII-N and EcoRII-C in the DNA bound forms provided here demonstrate that both modules use different strategies to interact with the pseudosymmetric 5′-CCWGG sequence.

EcoRII-C recognizes 5′-CCWGG sequence using base flipping mechanism

EcoRII-C forms a dimer and each monomer interacts with the symmetrical CC:GG dinucleotides using three consecutive amino acid residues Lys328, Asp329 and Arg330 located at the N-terminal end of the recognition α-helix H14 (Figure 1C and G). EcoRII recognizes CC:GG half-sites within 5′-CCWGG site analogously to REases that contain symmetrical CCGG subset within their target sites (Supplementary Table SII) (30,35,36). To accommodate an extra base pair in the target site EcoRII-C, like PspGI (16) or Ecl18kI (17), flips out the central nucleotide pair from the DNA helix followed by backbone compression which restores stacking interactions between adjacent CC:GG blocks. The compressed DNA conformation is presumably stabilized by interactions of extruded backbone phosphates with Arg262 residues, which are absolutely conserved within the EcoRII family (Supplementary Figure S1). Alanine mutation of structurally equivalent Arg116 and Arg96 residues in Ecl18kI and PspGI, respectively, completely abolished DNA binding and cleavage functions (37). The conformation of the flipped nucleotides is nearly identical in the EcoRII-C and PspGI structures but differs slightly from that of Ecl18kI (Figure 1D-F). It is unclear whether these differences in the extrahelical base conformation are due to the differences in the central base pair specificity: EcoRII-C and PspGI accepts only A or T nucleotides within their target sites while Ecl18kI shows no specificity for the central nucleotide.

Are extra-helical A and T bases specifically recognized in the binding pocket?

In the EcoRII-C structure, extruded A0 and T0 bases are sandwiched between the aromatic ring of the Tyr226 and side chain of the Arg222 (Figure 1E). Within the binding pocket, the C5 methyl group of T0 is positioned at the van der Waals contact distance (3.7 Å) to the aromatic ring of Phe229 (data not shown). The N6 atom of the flipped out A comes close (3.1 Å) to the $O_{\varepsilon}1$ oxygen atom of Glu225, however the geometry for the hydrogen bond is far from optimal. In theory, these interactions could allow discrimination of T:A base pair from C:G in the EcoRII-C binding pockets. Indeed, C to T replacement should eliminate favorable van der Waals interactions between the T methyl group and the Phe229 aromatic ring, while G to A replacement should perturb hydrogen-bonding interactions with the side chain carboxylate oxygen of Glu225, and introduce a steric clash with the aromatic ring of Phe229 (2.9 Å). The importance of Glu225 and Phe229 residues is supported by nearly absolute conservation among the 13 protein sequences of EcoRII homologues (Supplementary Figure S1). Surprisingly, in Ecl18kI REase, which shows no base preference in the binding pockets, Glu123 and Phe64 are structural equivalents of the conserved Glu225 and Phe229 residues of EcoRII-C (Figure 1E). Moreover, in PspGI, which like EcoRII-C, is specific for the 5′-CC(T/A)GG sequence, Tyr67 is structurally equivalent to Phe229 but Ala63 replaces Glu225 (Figure 1F). Biochemical experiments demonstrate that EcoRII-C and PspGI use a double check mechanism for the central base pair recognition probing both the stability of the central base pair and the identity of the flipped bases in the pockets (18). Further crystallographic and mutational studies are required to determine the mechanism of T : A versus C : G base pair discrimination in the binding pockets of EcoRII-C and PspGI.

EcoRII-N interacts with the 5′-CC(T/A)GG sequence asymmetrically

The effector DNA binding domain EcoRII-N uses a radically different strategy to recognize the 5′-CCWGG sequence. In the crystal EcoRII-N monomer binds oligoduplex containing T:A base pair at the center (Figure 2A) and uses the direct readout mechanism to discriminate the target sequence (Figure 2D-H). These interactions occur in the context of the canonical B-DNA form and the central T:A base pair remains within the double helix. In the crystal, the EcoRII-N monomer is bound to the pseudosymmetrical 5′-CCTGG sequence in one preferable orientation where the N-arm binds to the 'left' half-site (Figure 2A) and the C-arm binds to the 'right' half-site. In theory, interactions of amino acid residues located on the N-and C-arms with the CC:GG subsites within the target site (Figure 2C) are compatible with an alternative binding orientation which would position the N-arm to the 'right' half-site and the C-arm to the 'left' half-site. However, structural and biochemical data indicate that such a binding mode would disrupt EcoRII-N interactions with the central T : A base pair. Indeed, the central T:A base pair is discriminated through a combination of hydrogen bonding and favorable van der Waals interactions, which play a key role in the EcoRII-N binding to cognate DNA (Figure 2F). Elimination of the hydrogen bonding interactions between the 6-amino group and O_{ε} oxygen atom of Gln37 in the modified oligoduplex containing 2-aminopurine (Figure 3C) ablates cognate DNA binding. Nearly absolute conservation (with one exception) of the Gln37 among EcoRII homologues is consistent with its

key role in cognate DNA recognition. Furthermore, the perturbation of van der Waals interactions between the C5 methyl group of T0 and the aromatic ring of Tyr85 through T to U replacement also abolishes DNA binding by EcoRII-N (Figure 3C). Thus, crystal structure suggests that EcoRII-N interacts with the 5′-CCTGG sequence in one preferable orientation which positions the aromatic ring of the Tyr85 residue in the proximity of the methyl group of T0. Such an asymmetric binding pattern explains why the Tyr41 of EcoRII photocross-links with 45% yield to the outer C of the A-strand of the recognition sequence and no cross-link is observed for the 5′-CCTGG strand (38). The key role of Tyr85 in EcoRII-N binding specificity is supported by its absolute conservation among the EcoRII homologues (Supplementary Figure S1).

Implications on the EcoRII mechanism

An autoinhibition mechanism has been proposed for the regulation of EcoRII catalytic activity (6). According to this mechanism, the apo-form EcoRII forms a dimer where the C-terminal catalytic half-sites are blocked by the N-terminal effector DNA binding domains (autoinhibition). Effector DNA must bind to the N-terminal domain first to open the gate for DNA entrance into the C-terminal catalytic cleft (activation). Binding of a single DNA copy is enough to trigger DNA nicking, but both N-terminal domains must bind effector DNA copy for the concerted cleavage of both strands at EcoRII-C (3).

The EcoRII-N and the EcoRII-C structures in the DNA bound form illustrate structural changes that may occur during the transition of EcoRII from the autoinhibited to the active form. Structural rearrangements triggered by DNA binding at the EcoRII-N domain are limited to the loop regions at the N- and C-terminal arms (Figure 3A). These changes bring amino acid residues involved in the sequence recognition into the proximity of the DNA bases (Figure 3A and B). Rigid body superposition of the EcoRII-N in the DNA bound form on the N-terminal domain of wt EcoRII results in a steric clash between the effector DNA and C-terminal domain (data not shown). These unfavorable interactions may destabilize the interface between EcoRII-N and EcoRII-C, and trigger a conformational change that opens DNA access to the catalytic EcoRII-C domain. In the wt EcoRII structure, two N-terminal domains make a dimer but DNA binding clefts are exposed to solution and are available for effector DNA binding. Since EcoRII-N binds DNA as a monomer, wt EcoRII therefore can bind two effector DNA copies through its N-terminal domains. Biochemical studies suggest that simultaneous binding of two effector DNA copies results in the concerted cleavage of the third DNA copy at the catalytic EcoRII-C domain (3).

In the apo-EcoRII structure (Figure 1B), the DNA binding interface of EcoRII-C is not fully assembled (Figure 1C). To create the DNA binding surface at the EcoRII-C, the enzyme dimer must rearrange. The changes occur both at the N-terminal and C-terminal parts of EcoRII-C but to different extent. The rearrangements at the N-terminal EcoRII-C interface (α-helices H10 and H11) are moderate and result in the increase of the buried accessible surface area from 1400 Å^2 in the apo-form to 1600 Å^2 in the DNA-bound form. The most notable structural change is bending of the long H11 α-helix (Figure 1B and C) about 25° at the absolutely conserved Gly267 residue (Supplementary Figure S1). Much larger changes occur at the C-terminal interface. Recognition α-helices H14, which are located far apart from each other in the apo-form (Figure 1B), interact in EcoRII-C–DNA complex (Figure 1C). These newly developed contacts at the C-terminal interface bury extra 840 Å^2, closing the clamp around the DNA. The much smaller and predominantly hydrophilic interface suggests that the C-terminal part of EcoRII-C may act as an entrance gate which closes after encircling DNA.

NaeI restriction enzyme belongs to the Type IIE subtype restriction enzymes and must bind two 5′-GCC GGC recognition sequences to cleave one DNA sequence (39). The NaeI dimer consists of the Endo and Topo domains which bind substrate and effector DNA, respectively (40). The structural mechanisms of effector and substrate DNA recognition by NaeI differs from those of EcoRII. First, the Topo domain of NaeI is arranged as a dimer and binds a single copy of the effector DNA using two symmetry related CAP-like motives (41). Second, the catalytic Endo domain dimer recognizes substrate DNA primarily through direct hydrogen bond interactions between amino acid side chains and donor-acceptor sites on the base edges. The structural differences between NaeI and EcoRII result in different mechanisms of cleavage activation and reaction patterns (3).

Similarities of EcoRII-N to the B3 family transcription factors of higher plants

The crystal structure of the apo-EcoRII revealed that the N-terminal domain (EcoRII-N) represented a fold which was unknown at that time (6). NMR structures of B3 DNA binding domains of RAV1 (11) and At1g16640 (42) plant-specific transcription factors (43) appeared to be strikingly similar to EcoRII-N. Indeed, according to DALI (44), EcoRII-N and B3 domains of RAV1 and At1g16640 show pairwise comparison *Z*-scores of 6.2 and 7.1, respectively. Phylogenetic analysis confirms that EcoRII-N is closely related to the B3 family proteins (11). Therefore, the EcoRII-N–DNA complex structure represents the first structure of the B3 family protein in complex with DNA. Structural superposition of EcoRII-N–DNA and RAV1 B3 domain (Figure 4A) suggests that plant-specific transcription factors of B3 family may bind DNA in a similar mode. In the superimposed structures Arg202, Trp245 and Asn246 residues of RAV1 B3 domain are in position to contact DNA bases, while Lys190, Thr193, Ser195, Arg241 and Thr254 are close to DNA phosphates (Figure 4C).

An NMR titration experiment allowed Yamasaki et al. (11) to build a docking model of the RAV1-B3–DNA complex. Many of the putative DNA binding residues of RAV1-B3 identified by the NMR titration experiment

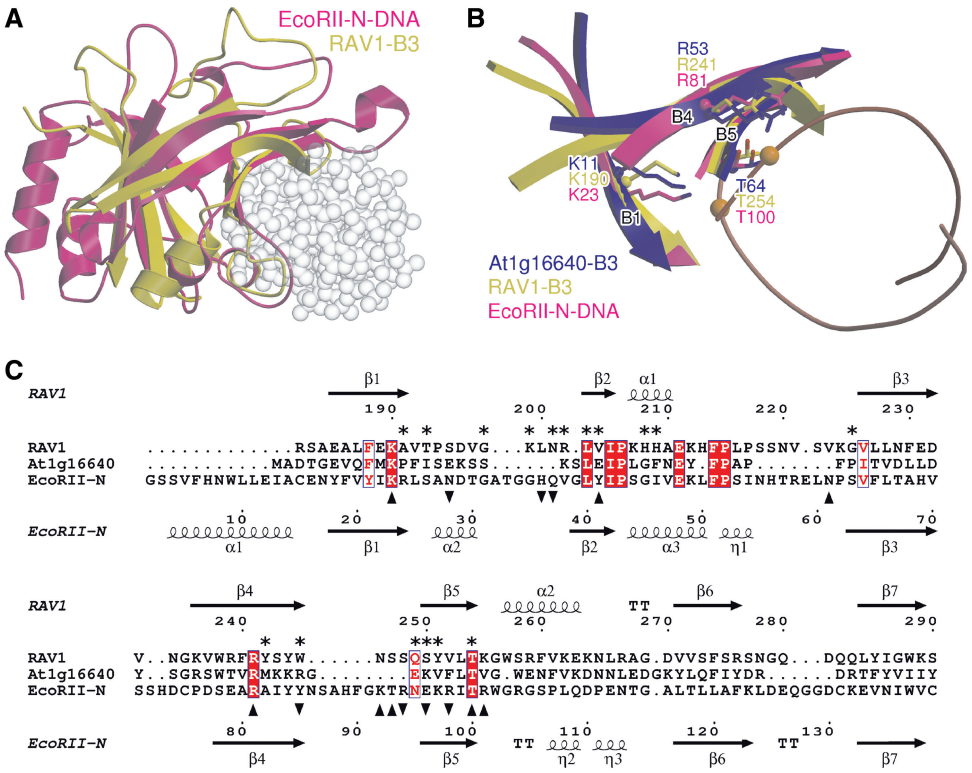

Figure 4. Similarities between EcoRII-N and plant transcription factors of B3 family. (**A**) Superposition of RAV1-B3 (yellow) [PDBID 1WID] on EcoRII-N–DNA complex structure (magenta). DNA is depicted in the spacefill representation and shown in grey. (**B**) Structurally conserved β-strands of EcoRII (magenta), RAV1-B3 (yellow), and At1g16640 (blue) [PDBID 1YEL] bearing conserved amino acid residues involved in contacts with DNA phosphates. DNA backbone is shown in brown, phosphates as orange spheres. (**C**) Alignment of the B3 domains of RAV1 and At1g16640 with EcoRII-N obtained by CLUSTALW (46), ESPRIPT (47). Secondary structures of RAV1 and EcoRII-N (aligned by DALI) are indicated above and below the alignment, correspondingly. EcoRII-N amino acid residues involved in specific and non-specific contacts are denoted by apex down or up triangles, correspondingly. The asterisks mark position of the RAV1 amino acid residues with largest backbone chemical shifts in NMR spectra affected by DNA binding (11).

and structural superposition with EcoRII-N coincide (Figure 4C). Since the coordinates for the RAV1-B3–DNA complex model obtained by Yamasaki, *et al.* are not available, we are unable to provide detailed comparisons. However, visual inspection of both models in the same orientation shows that DNA in the docking model (11) is rotated around the protein-DNA dyad axis nearly by 90° with respect to the DNA in the model obtained by structural superposition with EcoRII-N. Structural conservation of EcoRII-N residues Lys190, Arg241 and Thr254 involved in binding of the DNA backbone phosphates in RAV1-B3, At1g16640 (Figure 4B) and throughout B3 superfamily (Supplementary Table SIII) argues for the same DNA orientation in EcoRII-N and RAV1-B3 complexes.

The B3 domain of RAV1 is specific for the 5′-CACCTG sequence which in part overlaps with the target site 5′-CCTGG of the EcoRII-N (overlapping parts are underlined). The specificity of the transcription factor At1g16640, a member of the REM family in subgroup B, and other members of B3 family is not yet determined. Interestingly, the DNA binding domain of the BfiI restriction enzyme which shares the same fold with EcoRII-N (12) is specific for the 5′-ACTGGG sequence. This illustrates the highly diverse repertoire of DNA recognition specificities that can be adopted by the DNA-binding pseudo-barrel fold represented by EcoRII-N.

Coordinates

Coordinates and structure factors have been submitted to the RCSB Protein Data Bank with accession codes 3HQF for EcoRII-N–DNA and 3HQG for EcoRII-C–DNA.

ACCESSION NUMBERS

3HQF and 3HQG.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Monika Reuter (Institute of Virology at Humboldt University, Berlin, Germany) for sharing EcoRII constructs and to Matthew Groves for assistance with data collection at EMBL X13 beamline at the DORIS storage ring, DESY Hamburg. The authors also thank Gintautas Tamulaitis, Giedrius

Sasnauskas and Mindaugas Zaremba for comments and critical reading of the manuscript.

FUNDING

The EC FP6 Marie Curie Training network 'DNA enzymes'; and EC Research Infrastructure FP6 Action 'Structuring the European Research Area Specific Programme', Contract Number RII3-CT-2004-506008. Funding for open access charge: EC FP6 Marie Curie Training network 'DNA enzymes'.

Conflict of interest statement. None declared.

REFERENCES

1. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
2. Kruger,D.H., Barcak,G.J., Reuter,M. and Smith,H.O. (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res.*, **16**, 3997–4008.
3. Tamulaitis,G., Sasnauskas,G., Mucke,M. and Siksnys,V. (2006) Simultaneous binding of three recognition sites is necessary for a concerted plasmid DNA cleavage by EcoRII restriction endonuclease. *J. Mol. Biol.*, **358**, 406–419.
4. Reuter,M., Kupper,D., Meisel,A., Schroeder,C. and Kruger,D.H. (1998) Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites. *J. Biol. Chem.*, **273**, 8294–8300.
5. Shlyakhtenko,L.S., Gilmore,J., Portillo,A., Tamulaitis,G., Siksnys,V. and Lyubchenko,Y.L. (2007) Direct visualization of the EcoRII-DNA triple synaptic complex by atomic force microscopy. *Biochemistry*, **46**, 11128–11136.
6. Zhou,X.E., Wang,Y., Reuter,M., Mucke,M., Kruger,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.
7. Tamulaitiene,G., Jakubauskas,A., Urbanke,C., Huber,R., Grazulis,S. and Siksnys,V. (2006) The crystal structure of the rare-cutting restriction enzyme SdaI reveals unexpected domain architecture. *Structure*, **14**, 1389–1400.
8. Lambert,A.R., Sussman,D., Shen,B., Maunus,R., Nix,J., Samuelson,J., Xu,S. and Stoddard,B.L. (2008) Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure*, **16**, 558–569.
9. Niv,M.Y., Ripoll,D.R., Vila,J.A., Liwo,A., Vanamee,E.S., Aggarwal,A.K., Weinstein,H. and Scheraga,H.A. (2007) Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res.*, **35**, 2227–2237.
10. Lo Conte,L., Brenner,S.E., Hubbard,T.J.P., Chothia,C. and Murzin,A.G. (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.*, **30**, 264–267.
11. 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.
12. 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.
13. Mucke,M., Grelle,G., Behlke,J., Kraft,R., Kruger,D.H. and Reuter,M. (2002) EcoRII: a restriction enzyme evolving recombination functions? *EMBO J.*, **21**, 5262–5268.
14. Tamulaitis,G., Mucke,M. and Siksnys,V. (2006) Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett.*, **580**, 1665–1671.
15. Pingoud,V., Geyer,H., Geyer,R., Kubareva,E., Bujnicki,J.M. and Pingoud,A. (2005) Identification of base-specific contacts in protein–DNA complexes by photocrosslinking and mass spectrometry: a case study using the restriction endonuclease SsoII. *Mol. Biosyst.*, **1**, 135–141.
16. Szczepanowski,R.H., Carpenter,M.A., Czapinska,H., Zaremba,M., Tamulaitis,G., Siksnys,V., Bhagwat,A.S. and Bochtler,M. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res.*, **36**, 6109–6117.
17. Bochtler,M., Szczepanowski,R.H., Tamulaitis,G., Grazulis,S., Czapinska,H., Manakova,E. and Siksnys,V. (2006) Nucleotide flips determine the specificity of the Ecl18kI restriction endonuclease. *EMBO J.*, **25**, 2219–2229.
18. Tamulaitis,G., Zaremba,M., Szczepanowski,R.H., Bochtler,M. and Siksnys,V. (2008) How PspGI, catalytic domain of EcoRII and Ecl18kI acquire specificities for different DNA targets. *Nucleic Acids Res.*, **36**, 6101–6108.
19. Leslie,A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 48–57.
20. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
21. French,G. and Wilson,K. (1978) On the treatment of negative intensity observations. *Acta Crystallogr*, **A34**, 517–525.
22. Navaza,J. (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 1367–1372.
23. Vagin,A. and Teplyakov,A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 1622–1624.
24. Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr*, **A42**, 140–149.
25. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
26. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
27. Winn,M.D., Murshudov,G.N. and Papiz,M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Meth. Enzymol.*, **374**, 300–321.
28. Lavery,R. and Sklenar,H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.
29. Tamulaitis,G., Zaremba,M., Szczepanowski,R., Bochtler,M. and Siksnys,V. (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res.*, **35**, 4792–4799.
30. Tamulaitis,G., Solonin,A.S. and Siksnys,V. (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett.*, **518**, 17–22.
31. Petrauskene,O.V., Gromova,E.S., Romanova,E.A., Volkov,E.M., Oretskaya,T.S. and Shabarova,Z.A. (1995) DNA duplexes containing methylated bases or non-nucleotide inserts in the recognition site are cleaved by restriction endonuclease R.EcoRII in presence of canonical substrate. *Gene*, **157**, 173–176.
32. Kruger,D.H. and Reuter,M. (2005) Reliable detection of DNA cytosine methylation at CpNpG sites using the engineered restriction enzyme EcoRII-C. *BioTechniques*, **38**, 855–856.
33. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE–enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
34. Orlowski,J. and Bujnicki,J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552–3569.
35. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
36. Dunten,P.W., Little,E.J., Gregory,M.T., Manohar,V.M., Dalton,M., Hough,D., Bitinaite,J. and Horton,N.C. (2008) The

structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.*, **36**, 5405–5416.
37. Pingoud,V., Conzelmann,C., Kinzebach,S., Sudina,A., Metelev,V., Kubareva,E., Bujnicki,J.M., Lurz,R., Luder,G., Xu,S. *et al.* (2003) PspGI, a type II restriction endonuclease from the extreme thermophile Pyrococcus sp.: structural and functional studies to investigate an evolutionary relationship with several mesophilic restriction enzymes. *J. Mol. Biol.*, **329**, 913–929.
38. Mucke,M., Pingoud,V., Grelle,G., Kraft,R., Kruger,D.H. and Reuter,M. (2002) Asymmetric photocross-linking pattern of restriction endonuclease EcoRII to the DNA recognition sequence. *J. Biol. Chem.*, **277**, 14288–14293.
39. Conrad,M. and Topal,M.D. (1989) DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme Nae I. *Proc. Natl Acad. Sci. USA*, **86**, 9707–9711.
40. Huai,Q., Colandene,J.D., Chen,Y., Luo,F., Zhao,Y., Topal,M.D. and Ke,H. (2000) Crystal structure of NaeI-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J.*, **19**, 3110–3118.
41. Huai,Q., Colandene,J.D., Topal,M.D. and Ke,H. (2001) Structure of NaeI–DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat. Struct. Biol.*, **8**, 665–669.
42. 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.
43. Swaminathan,K., Peterson,K. and Jack,T. (2008) The plant B3 superfamily. *Trends Plant Sci.*, **13**, 647–655.
44. Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
45. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
46. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
47. 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.