

DNA cleavage by CgII and NgoAVII requires interaction between N- and R-proteins and extensive nucleotide hydrolysis

Mindaugas Zaremba^{1,*}, Paulius Toliūsis^{1,†}, Rokas Grigaitis¹, Elena Manakova¹,
Arunas Silanskas¹, Giedre Tamulaitiene¹, Mark D. Szczelkun² and Virginijus Siksnys^{1,*}

¹Department of Protein–DNA Interactions, Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241 Vilnius, Lithuania and ²DNA–Protein Interactions Unit, School of Biochemistry, Medical Sciences Building, University of Bristol, Bristol BS8 1TD, UK

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

ABSTRACT

The stress-sensitive restriction-modification (RM) system CgII from *Corynebacterium glutamicum* and the homologous NgoAVII RM system from *Neisseria gonorrhoeae* FA1090 are composed of three genes: a DNA methyltransferase (M.CgII and M.NgoAVII), a putative restriction endonuclease (R.CgII and R.NgoAVII, or R-proteins) and a predicted DEAD-family helicase/ATPase (N.CgII and N.NgoAVII or N-proteins). Here we report a biochemical characterization of the R- and N-proteins. Size-exclusion chromatography and SAXS experiments reveal that the isolated R.CgII, R.NgoAVII and N.CgII proteins form homodimers, while N.NgoAVII is a monomer in solution. Moreover, the R.CgII and N.CgII proteins assemble in a complex with R₂N₂ stoichiometry. Next, we show that N-proteins have ATPase activity that is dependent on double-stranded DNA and is stimulated by the R-proteins. Functional ATPase activity and extensive ATP hydrolysis (~170 ATP/s/monomer) are required for site-specific DNA cleavage by R-proteins. We show that ATP-dependent DNA cleavage by R-proteins occurs at fixed positions (6–7 nucleotides) downstream of the asymmetric recognition sequence 5'-GCCGC-3'. Despite similarities to both Type I and II restriction endonucleases, the CgII and NgoAVII enzymes may employ a unique catalytic mechanism for DNA cleavage.

INTRODUCTION

Restriction-modification (RM) systems protect bacterial cells against invasion by foreign DNA (e.g. phage or plas-

mid DNA). RM systems are composed of two principal enzymatic activities: an endonuclease and a methyltransferase. Restriction endonucleases (REases) cleave DNA that do not contain characteristic modifications (methyl groups) at defined sites within a specific recognition sequence. The host DNA is protected against nucleolytic attack because it is methylated within the recognition sequences by the cognate methyltransferase activity of the RM system. According to the genetic organization, subunit composition, cofactor requirement and mode of action, the RM systems were subdivided into four types: Type I, II, III or IV (1). Here we investigated a new class of RM system, exemplified by CgII and NgoAVII, which does not readily conform to the classical Types.

The CgII RM system from *Corynebacterium glutamicum* (ATCC 13032) is inactivated under various stress conditions (heat, organic solvents, pH shift or detergents), enabling the transfer of plasmid DNA from a donor (e.g. *Escherichia coli*) by either conjugation or phage infection (2,3). Inactivation of CgII may allow the host to more readily adapt to environmental changes by allowing acquisition of foreign genetic material (4). Due to its potential for inactivation, the CgII system was named as stress-sensitive restriction system. The CgII genes are located within the large prophage CGP3, which can be induced under SOS response conditions (5,6). It remains to be established whether prophage excision from the *Corynebacterium glutamicum* genome is linked to the inactivation of the CgII RM system under stress conditions (7).

Previously the CgII system was assigned as Type II (REBASE, <http://rebase.neb.com>) (8). However, its genetic organization differs from typical Type II RM systems that are usually composed of only REases and methyltransferase genes (1). The CgII system is instead organized as three genes: a methyltransferase (M.CgII), a putative PLD-family endonuclease (R.CgII) and a pre-

*To whom correspondence should be addressed. Tel: +370 5 2602111; Fax: +370 5 2602116; Email: zare@ibt.lt

Correspondence may also be addressed to Virginijus Siksnys. Tel: +370 5 2602108; Fax: +370 5 2602116; Email: siksnys@ibt.lt

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

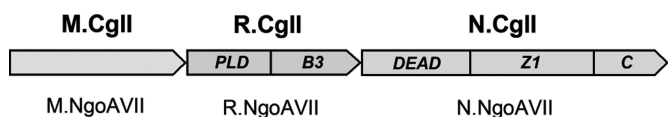


Figure 1. Organization of stress-sensitive restriction-modification systems from *Corynebacterium glutamicum* (CgII) and *Neisseria gonorrhoeae* FA1090 (NgoAVII). M.CgII (ORF NCgII1703) and M.NgoAVII (Ngo0365) are C5-methyl DNA methyltransferases. R.CgII (NCgII1704) and R.NgoAVII (Ngo0364) are putative REases with the PLD-superfamily nucleolytic and B3-like DNA binding domains. N.CgII (NCgII1705) and N.NgoAVII (Ngo0363) are predicted SF2 helicases/ATPases containing uncharacterized Z1-superfamily and C-terminal domains.

dicted DEAD-family helicase/adenosine triphosphatase (ATPase) (N.CgII) (Figure 1). The NgoAVII RM system, identified by bioinformatics in *Neisseria gonorrhoeae* FA1090 (ATCC 700825) was also assigned as a stress-sensitive restriction system (BLAST), and correspondingly is composed of three genes homologous to the CgII RM system: M.NgoAVII, R.NgoAVII and N.NgoAVII, respectively (Figure 1, Supplementary Figures S1, S2 and S3). Given the unusual stress response and presence of a putative ATPase component, the CgII and NgoAVII RM systems may use a new, uncharacterized mechanism for DNA cleavage.

The putative REases R.CgII and R.NgoAVII contain predicted PLD-superfamily HXK nuclease active site motifs within their N-terminal domains and C-terminal B3-type DNA-binding domains (Figure 1, Supplementary Figure S2) (see the accompanying paper) (9). An N-terminal PLD-superfamily domain is also characteristic of the well-studied Type IIS REase BfiI (recognition sequence 5'-ACTGGG-3'), in which two N-terminal nucleolytic domains dimerize to form a single active site that acts sequentially to make double-strand DNA breaks at fixed positions downstream of the recognition sequence (10,11). Although the R.CgII and R.NgoAVII proteins hydrolyzed the artificial substrate bis(*p*-nitrophenyl) phosphate, in contrast to BfiI (12), they showed no DNA cleavage activity (Stonyte, Toliusis and Siksnys, unpublished data) most likely because the third protein (N.CgII or N.NgoAVII, respectively) is also required. A necessity for cooperation between R and N protein can also be inferred based on *in vivo* phage-resistance and conjugation experiments using CgII (3,13) and from *in vitro* DNA cleavage experiments with the homologous BceSIV RM system from *Bacillus cereus* (7).

In silico amino acid sequence analysis predicted that N.CgII and N.NgoAVII are composed of three domains: an N-terminal DEAD-superfamily helicase domain (Superfamily 2, SF2); a domain similar to the Z1-superfamily; and a distinct C-terminal domain (Figure 1, Supplementary Figure S3). The uncharacterized Z1 domain is often found associated with a SF2 helicase domain (14). The N-proteins possess the characteristic ATP- and Mg²⁺-binding amino acid sequence motifs, consistent with ATPase activity (Supplementary Figure S3). Similar DEAD domains are found in SF2 helicases involved in ATP-dependent RNA or DNA unwinding/translocation (15). However, the N-protein helicase motifs are only distantly-related to those characteristic of Type I and Type III RM enzymes (data not shown).

The M.CgII and M.NgoAVII proteins have high amino acid sequence similarity (62% identity and 76% similarity) and are Type II 5-methylcytosine methyltransferases that recognize the 5'-GCSGC-3' DNA sequence (where S represents either G or C) (Supplementary Figure S1) (16,17). Like typical Type II methyltransferases, they are active as isolated proteins (Supplementary Figure S5).

In this paper, we report a biochemical characterization of the putative REases R.CgII and R.NgoAVII (also referred to as 'R-proteins') and cognate ATPases N.CgII and N.NgoAVII (also referred to as 'N-proteins'). Following recombinant gene expression in *E. coli*, we purified individual proteins and their domains to apparent homogeneity. Using co-expression of R.CgII and N.CgII, the RN.CgII complex with an R₂N₂ stoichiometry was also isolated. We demonstrate that both R.CgII and R.NgoAVII are endonucleases that require ATP hydrolysis by the corresponding N-proteins to produce site-specific DNA cleavage at fixed positions downstream of the asymmetric recognition sequence 5'-GCCGC-3'. On the basis of their activities, we propose that CgII and NgoAVII represent a new sub-class of RM enzyme with characteristics of both Type I and II RM systems.

MATERIALS AND METHODS

Expression and purification of CgII and NgoAVII proteins

Corynebacterium glutamicum ATCC 13032 and *Neisseria gonorrhoeae* FA1090 genomic DNA (both from LGC Standards) were used to produce recombinant genes of R.CgII/N.CgII, R.NgoAVII/N.NgoAVII or DNA fragments corresponding to individual domains that later were ligated into expression and co-expression vectors (Supplementary Tables S1 and S2). Complete gene sequences were confirmed. His₆- or StrepII-tagged proteins were expressed and purified using liquid chromatography as described in Supplementary Tables S1 and S2. The homogeneity of the protein preparations (>90%) was determined using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and gel densitometry. The identity of the purified proteins was confirmed by mass spectrometry (data not shown). Protein concentrations were determined by measuring absorbance at 280 nm using appropriate extinction coefficient (Supplementary Tables S1 and S2).

Mutagenesis

The R.CgII (H105A), N.CgII (D158A+E159A), R.NgoAVII (H104A) and N.NgoAVII (D150A+E151A) mutants were produced by QuikChange Site-Directed Mutagenesis (18). Sequencing of the entire gene for each mutant confirmed that only the designed mutations had been introduced.

Gel filtration

Gel filtration was carried out at room temperature on an AKTA Avant system using a Superdex 200 HR column (GE) pre-equilibrated with 20 mM Tris-HCl (pH 8.0 at 25°C), 300 mM NaCl and 5 mM 2-mercaptoethanol. Protein samples at 20 μM (monomer) loading concentration

were prepared in 50 μ l of the above buffer. Elution from the column was monitored by measuring absorbance at 280 nm. The apparent molecular weights of proteins were evaluated from the elution volume using a series of standards (Gel Filtration Calibration Kit from GE Healthcare).

Small angle X-ray scattering (SAXS) experiments

The wild-type (wt) RN.CgII complex was dialyzed against 20 mM Tris-HCl (pH 8.0 at 25°C), 400 mM KCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM dithiothreitol (DTT). R.CgII, R.NgoAVII, N.CgII and N.NgoAVII were transferred into the same buffer using desalting NAP columns (GE Healthcare). Proteins were concentrated using centrifugal concentrators (Amicon Ultra-0.5, Millipore). All samples were processed at 15 000 *g* for 15–20 min.

Small angle X-ray scattering (SAXS) data were collected at P12 EMBL beam-line at PETRAIII storage ring of the DESY synchrotron in Hamburg (Germany). The sample-to-detector (Pilatus 2M) distance used in the measurements was 3.1 m. Data were collected for the momentum transfer s ($s = 4\pi \sin(\theta)/\lambda$, where 2θ is a scattering angle and λ is x-ray wavelength, 0.124 nm) range from 0.07579 to 4.66533 nm^{-1} , except the R.CgII samples, which were measured in s range from 0.02906420 to 4.491070 nm^{-1} . Data were truncated at the first point of the Guinier region, as calculated by AUTORG (19). Real space R_g and $I(0)$ were calculated using GNOM (20). D_{max} was estimated manually by interactive GNOM runs and compared with the value calculated by DATGNOM (19). Unprocessed scattering data $I(s)$ with subtracted buffer scattering, Kratky plots and distance distribution functions $P(r)$ of protein samples are presented in Supplementary Figure S4. Structural parameters derived from the SAXS measurements are presented in Supplementary Table S3.

Molecular weight estimations of individual proteins and the corresponding complexes are presented in Table 1. Molecular weights were calculated from the Fourier-transformed scattering data by the method described in (21) using the MoW program run on the server (<http://www.if.sc.usp.br/~saxs/saxsmow.html>).

Gel mobility-shift assay

DNA binding by the CgII and NgoAVII proteins was analyzed by the electrophoretic mobility shift assay (EMSA) using a 33 bp cognate duplex (5'-TACAATCGATAATAGCCGCCGTGCGAGCCCATT-3'/3'-ATGTTAGCTATTATCGGCCGCACGCTCGGGTAA-5', the recognition sequence is underlined) or a 33 bp non-cognate DNA duplex (5'-TACAATCGATAATACCAGGCCGTGCGAGCCCATT-3'/3'-ATGTTAGCTATTATGGTCCGCACGCTCGGGTAA-5'). DNA (final concentration 1 nM) was incubated with proteins (final concentrations from 1 to 500 nM monomer) for 15 min in 20 μ l of 40 mM Tris-acetate (pH 8.3 at 25°C), 0.1 mM EDTA, 0.1 mg/ml bovine serum albumin (BSA) and 10% (v/v) glycerol at room temperature. Free DNA and protein-DNA complexes were separated by electrophoresis using 8% (w/v) acrylamide gels (29:1

acrylamide/bisacrylamide in 40 mM Tris-acetate, pH 8.3 at 25°C and 0.1 mM EDTA). In the case of the NgoAVII proteins, EDTA in the binding buffer and loading gels was replaced by 5 mM Ca-acetate. Samples were separated at room temperature for 3 h at ~ 6 V/cm. Radiolabeled DNA was detected and quantified using the Cyclone phosphorimager and the OptiQuant software (Packard Instrument). Apparent K_d values for DNA binding were determined as described in (22).

ATPase assay

ATPase reactions were performed at 37°C for 1 h in 33 mM Tris-acetate (pH 7.0 at 25°C), 0.1 mg/ml BSA, 10 mM Mg-acetate and 150 mM K-acetate (only for CgII proteins), supplemented with 50 μ M [α - 32 P]ATP (Hartmann Analytic, Braunschweig, Germany) and 0.01 μ g/ μ l of various DNAs (as indicated): (i) phage λ DNA (dam⁺dcm⁺), (ii) pBR322, (iii) pBR322 methylated at the CgII/NgoAVII recognition sites, (iv) cognate fragment, (v) noncognate fragment, (vi) cognate oligoduplex (5'-GATCGAGCCGCCTACG-3'/3'-CTAGCTCGGCCGATGC-5', the recognition sequence is underlined), (vii) non-cognate oligoduplex (5'-GATCGAC CAGGCTACG-3'/3'-CTAGCTGGTCCGATGC-5') and (viii) single-stranded M13mp18 (Supplementary Data). Reactions were initiated by adding the CgII (40 nM) or NgoAVII (40 or 100 nM) proteins and were stopped by addition of EDTA to 30 mM. An aliquot (1 μ l) was spotted onto a polyethyleneimine-cellulose thin-layer plate, the substrate ATP and product ADP separated by chromatography in a 0.325 M KH₂PO₄ (pH 3.5 at 25°C) and the plates visualized using Fujifilm FLA-5100 fluorescent image analyzer (Fujilm, Tokyo, Japan).

Malachite green assays (BioAssay Systems) were used to measure the ATP hydrolysis rate by detection of liberated free phosphate. Reactions were performed at 37°C in the same reaction buffer as in the radioactivity assay, and containing 0.02 μ g/ μ l phage λ DNA, ATP (1, 2, 4 or 6 mM), 1 mM (d)NTP and 2.5 or 10 mM divalent metal ions (Mg²⁺, Mn²⁺, Ca²⁺ or Ni²⁺) for NgoAVII and CgII, respectively. Reactions were initiated by adding N.CgII or N.NgoAVII (10 nM) plus R.CgII or R.NgoAVII (200 nM). Aliquots were removed at fixed time intervals and the reactions stopped by adding EDTA to 9.5 mM. The relationship between the absorbance and phosphate concentration was established by using KH₂PO₄ as a standard. ATPase rates are quoted as the mean of three independent experiments.

DNA cleavage assay

DNA cleavage reactions were performed for 1 h at 37°C in reaction buffer: 33 mM Tris-acetate (pH 7.0 at 25°C); 0.1 mg/ml BSA; either 50 mM (for NgoAVII) or 150 mM (for CgII) K-acetate; either 2.5 mM (for NgoAVII) or 10 mM (for CgII) divalent metal ions (Mg²⁺, Mn²⁺ or Ca²⁺); 1 mM DTT; 2 or 4 mM adenylyl-imidodiphosphate (AMP-PNP) or ADP; 2 or 4 mM (d)NTP; 25 mM phosphocreatine; and 5 U/ml creatine kinase. The reactions were supplemented with 0.01 μ g/ μ l of various dsDNAs (as indicated): (i) phage λ DNA (dam⁺dcm⁺), (ii) pBR322, (iii) pBR322 methylated at the CgII/NgoAVII recognition sites, (iv) pUC19 or (v) 0-site plasmid (without CgII/NgoAVII recognition sequences

Table 1. Oligomeric state of the CglI and NgoAVII proteins and their domains

Protein/domain/complex	M _w , kDa			Oligomeric state
	Calculated ^a	Gelfiltration, M _{w(exp)} / M _{w(calc)}	SAXS, M _{w(exp)} / M _{w(calc)}	
R.CglI ^b	41.2	85.0 2.1	84.7 2.1	Dimer
R.CglI-PLD	21.8	30.4 1.5		Dimer
R.CglI-B3	23.1	26.3 1.1		Monomer
N.CglI ^c	71.9	128.8 1.8	136.1 1.9	Dimer
N.CglI-DEAD	28.4	29.5 1.0		Monomer
N.CglI-DEAD-Z1	55.3	102.8 1.9		Dimer
N.CglI-Z1	28.1	42.4 1.5		Dimer
N.CglI-Z1-C	46.7	91.2 2.0		Dimer
N.CglI-C	19.8	17.9 0.9		Monomer
RN.CglI ^c	113.2 ^c	243.6 2.2	231.8 2.0	Heterotetramer (R₂N₂)
R.NgoAVII ^c	42.0	69.2 1.7	87.8 2.1	Dimer
R.NgoAVII-B3	22.8	22.9 1.0		Monomer
N.NgoAVII ^c	72.2	81.6 1.1	86.1 1.2	Monomer
N.NgoAVII-DEAD	28.7	35.3 1.2		Monomer
N.NgoAVII-DEAD-Z1	55.2	57.2 1.0		Monomer
N.NgoAVII-Z1	27.7	22.3 0.8		Monomer
N.NgoAVII-Z1-C	46.7	46.8 1.0		Monomer
N.NgoAVII-C	20.2	20.4 1.0		Monomer

^aTheoretical molecular weights were calculated using the ProtParam tool, <http://web.expasy.org/protparam>.

^bFull-length proteins and their complexes are shown on grey background.

^cThis value indicates the sum of theoretical molecular weights of the R- and N-protein monomers.

and generated from pMDS36a by the site-directed mutagenesis (23)). Reactions were initiated by adding N.CglI (200 nM) plus R.CglI (200 nM) or N.NgoAVII (100 or 200 nM) plus R.NgoAVII (100 or 200 nM) to the reaction mixture. Reactions were stopped by addition of 1:1 phenol-chloroform. The aqueous fraction was mixed with loading solution (75 mM EDTA, 0.025% (w/v) bromophenol blue, 50% (v/v) glycerol, 0.3% (w/v) SDS, pH 8.0 at 25°C) and incubated at 70°C for 15 min. The products were separated by electrophoresis through 0.8 or 1.0% (w/v) agarose gels and visualized under ultraviolet light following ethidium bromide staining.

To analyze the DNA cleavage position of CglI and NgoAVII, the 1-site plasmid (10 or 2 nM, respectively), generated from pMDS36a by the site-directed mutagenesis (23), was incubated with N.CglI (500 nM) plus R.CglI (500 nM) or N.NgoAVII (100 nM) plus R.NgoAVII (100 nM) for 1 h at 37°C in reaction buffer. Linearized plasmids were purified following agarose gel electrophoresis us-

ing the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Vilnius). The linear purified DNA was sequenced directly with the primers 5'-GAACTCTGTAGCACC GCC-3' [top strand] and 5'-CACTCAAAGGCGGTAATACGG-3' [bottom strand].

RESULTS

Expression and purification of proteins

The R- and N-proteins and their domains were expressed with either His₆- or StrepII-tags (see 'Materials and Methods' section, Supplementary Tables S1 and S2). A set of alanine-substitution mutations of the putative active sites within the R- and N-proteins was generated: H105 (R.CglI) and H104 (R.NgoAVII) belonging to the predicted HXX active site motif of the PLD-superfamily nucleolytic domain; and D158+E159 (N.CglI), and D150+E151 (N.NgoAVII) located in the DEAD domain and involved in Mg²⁺ coordination (Supplementary Figures S2 and S3).

Also, both R- and N-proteins from either system were co-expressed using a single recombinant vector (see 'Materials and Methods' section, Supplementary Table S1). All proteins were expressed in *E. coli* and were purified to an apparent homogeneity (>90%) by liquid chromatography. The co-expressed R.CgII (containing a StrepII-tag) and N.CgII (containing a His₆-tag) formed the RN.CgII complex that was purified using serial affinity chromatographic columns (HisTrap HP and StrepTrap HP) (Supplementary Table S1). In the case of the co-expressed R.NgoAVII and N.NgoAVII proteins, a complex was not obtained (data not shown). The identity of all purified proteins was confirmed by mass spectrometry (data not shown).

Oligomeric assembly

The oligomeric assembly of the individual R- and N-proteins and their domains was analyzed in solution by gel filtration (Table 1, Supplementary Figures S6 and S7). Both R.CgII and its isolated PLD domain eluted from the column as dimers, while the isolated R.CgII-B3 domain eluted as a monomer (Supplementary Figure S6, Table 1). N.CgII, its isolated DEAD-Z1 domain, its isolated Z1 domain and its isolated Z1-C domain all eluted as dimers, while the isolated DEAD and C-domains eluted as monomers (Supplementary Figure S6, Table 1). Full-length R.NgoAVII also eluted as a dimer, while the R.NgoAVII-B3 domain eluted as a monomer (Supplementary Figure S7, Table 1). N.NgoAVII and all its isolated domains eluted as monomers (Supplementary Figure S7, Table 1). SAXS experiments confirmed the same oligomeric state for the full-length R- and N-proteins as determined by gel filtration (Supplementary Figure S4, Table 1).

The molecular weight of the RN.CgII complex, obtained by co-expression of both the R.CgII and N.CgII proteins, was determined by gel filtration and SAXS to be close to the sum of molecular weights of the R.CgII and N.CgII dimers (Table 1, Supplementary Figures S4 and S6). Therefore, the R.CgII and N.CgII dimers assemble into a protein complex with a R₂N₂ stoichiometry. In addition, the stoichiometry of the RN.CgII complex was examined using densitometric analysis of the purified complex following SDS/PAGE electrophoresis (Supplementary Figure S8). The calculated molar ratio of the R.CgII and N.CgII proteins within the complex was 1:0.75. Since both proteins are dimers in solution, this ratio supports the R₂N₂ stoichiometry. The R₂N₂ CgII complex was also obtained by mixing the individual R.CgII and N.CgII proteins *in vitro* (data not shown). Higher-order complex formation was also observed upon mixing of the full-length R.CgII protein with either the N.CgII-DEAD domain or the N.CgII-DEAD-Z1 domain (Supplementary Figure S9). In contrast, a higher-order complex was not obtained if the PLD domain or B3 domain of R.CgII were mixed with the corresponding domains of N.CgII (data not shown). As with the co-expression study, a stable complex between the R.NgoAVII and N.NgoAVII was not observed upon mixing the proteins *in vitro* (Supplementary Figure S7, data not shown).

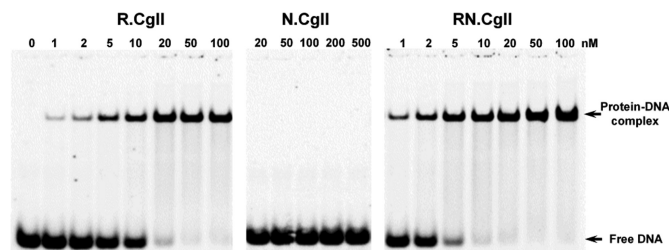


Figure 2. Cognate DNA binding by CgII proteins. The reactions contained 1 nM of the ³³P-labeled cognate oligoduplex and the protein at the concentrations indicated above each lane. After 15 min at room temperature, the samples were subjected to native PAGE for 3 h and analyzed as described in 'Materials and Methods' section. The calculated *K*_d values are 9.2 ± 2.8 nM for R.CgII and 1.9 ± 0.6 nM for RN.CgII.

DNA binding studies

DNA binding by the CgII and NgoAVII proteins was studied by EMSA using ³³P-labeled cognate and non-cognate oligoduplexes (Figure 2, Supplementary Figures S10, S11 and S12). In the case of the non-cognate oligoduplex, little or no DNA binding was observed with the R- or N-proteins, or with the RN-complex (Supplementary Figures S10 and S11). The R-proteins and their corresponding isolated B3 domains formed a specific complex with cognate DNA, while the N-proteins did not (Figures 2, Supplementary Figures S10, S11 and S12). The complete RN.CgII complex bound the cognate oligoduplex into a protein–DNA complex with the same electrophoretic mobility as the DNA complex formed by the isolated R.CgII protein (Figure 2). We rationalize these results as showing that the RN.CgII complex dissociates at the nanomolar protein concentrations used in the EMSA assay (as compared to the micromolar protein concentrations used in the gel filtration and SAXS assays). The observed complex is therefore between the dissociated R-protein and the DNA. On the basis of the EMSAs we propose that the R-proteins, and in particular the B3 domains, are responsible for the interaction with the DNA recognition sequence.

ATP hydrolysis studies

The presence of the characteristic SF2 helicase/ATPase motifs in the primary sequence of the N-proteins predicts that the proteins would be DNA-dependent ATPases (Supplementary Figure S3). To test this, the ATPase activity of the N-proteins was first analyzed by thin-layer chromatography using radioactive [³²P]ATP as a substrate (Figure 3A and Supplementary Figure S9A). In the presence of the N-proteins alone, ATPase activity was not detected, independent of the presence of DNA. Residual ATP hydrolysis was observed with N.NgoAVII and double-stranded bacteriophage λ DNA (Supplementary Figure S13A). However, the double-stranded DNA-dependent ATPase activity of the N-proteins increased significantly in the presence of the corresponding R-proteins (Figure 3A). The active site mutants of N.CgII (D158A+E159A) and N.NgoAVII (D150A+E151A) did not have any detectable ATP hydrolysis activity. In contrast, mutations in the N-terminal PLD domains of R.CgII (H105A) and R.NgoAVII (H104A) had

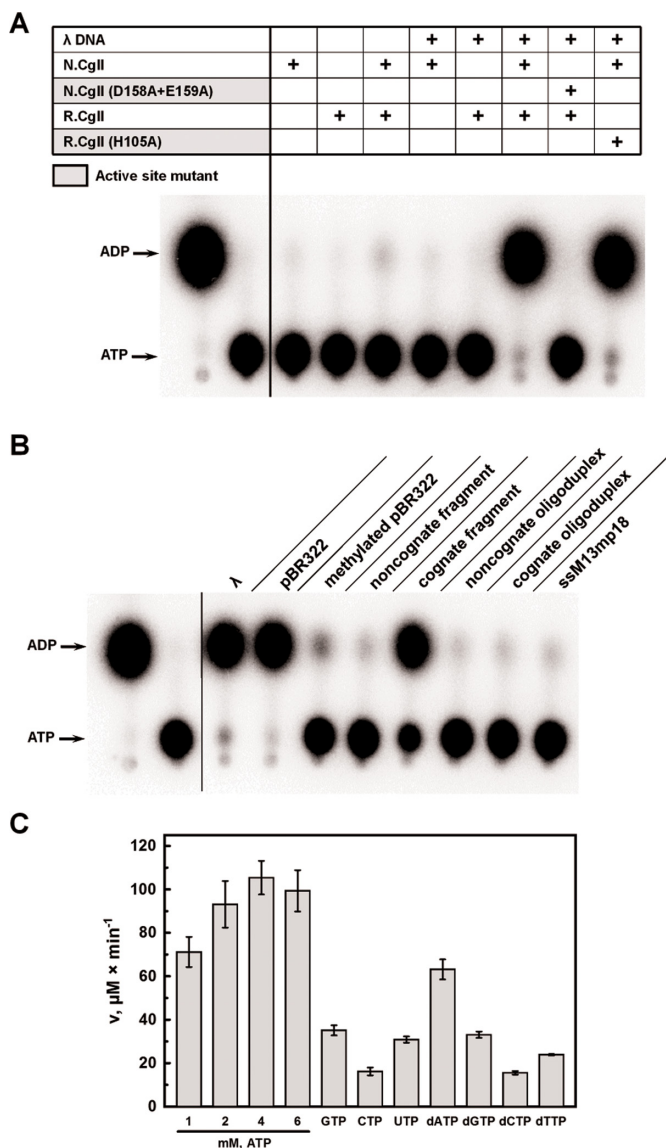


Figure 3. N.CgII ATPase activity. (A) Radioactive ATPase assay. ATPase reactions contained 50 μM [α - ^{32}P]ATP, 0.01 $\mu\text{g}/\mu\text{l}$ phage λ DNA, 40 nM N.CgII or R.CgII and were conducted as described in 'Materials and Methods' section. Reaction products were separated using thin-layer chromatography and visualized using a phosphorimager. (B) Dependence of CgII ATPase activity on different DNAs. Reactions were performed as in (A) using 0.01 $\mu\text{g}/\mu\text{l}$ DNA concentration (see details in the text). (C) NTP hydrolysis rates. Reactions contained 0.02 $\mu\text{g}/\mu\text{l}$ phage λ DNA, 10 nM N.CgII, 200 nM R.CgII, 1–6 mM ATP or 1 mM (d)NTP, as indicated, and were conducted as described in 'Materials and Methods' section. The malachite green assay was used to measure ATP hydrolysis through the detection of liberated free phosphate.

no effect on the ATPase activity of the respective RN-complexes. These results confirm that the N-proteins are ATPases, which require both DNA and a catalytically active or inactive (see below) endonuclease subunit for maximal activity.

The ATPase activity of the RN-complexes was further analyzed in the presence of: various circular or linear double-stranded DNA substrates (see 'Materials and Methods' section and Supplementary Data), either lacking or

containing different numbers of recognition sites (unmodified or modified by M.NgoAVII where stated); or circular single-stranded M13mp18 (Figures 3B and Supplementary Figure S13B). The RN.CgII complex showed ATPase activity in the presence of linear phage λ DNA, circular pBR322 plasmid or a linear DNA fragment containing a single recognition site (Figure 3B). Interestingly, a 16 bp cognate oligoduplex containing the CgII recognition sequence did not support ATP hydrolysis. Little or no ATP hydrolysis was observed in reactions containing pBR322 with methylated recognition sites, with a non-cognate DNA fragment, with a non-cognate oligoduplex, or with single-stranded DNA (Figure 3B). Similar results were obtained with N.NgoAVII, except that more pronounced ATP hydrolysis was observed with methylated pBR322 and with the non-cognate DNA (Supplementary Figure S13B). Thus, ATPase activity of the R₂N₂ complex is only supported by double-stranded DNA containing one or more unmodified recognition sites and with a length > 16 bp.

In order to test the nucleotide triphosphate specificity of the RN-complexes, we compared the catalytic activity using either ribo- and deoxyribonucleotide cofactors (Figure 3C and Supplementary Figure S13C). For these experiments the malachite green assay was used to measure (d)NTP hydrolysis through detection of liberated free phosphate. The RN-complexes had a partial preference for ATP and dATP, but could nonetheless hydrolyze both GTP and dGTP efficiently (~2-fold slower compared to ATP). Other dNTPs/NTPs could be utilized but with a lower hydrolysis rate (up to 5-fold slower compared to ATP) (Figure 3C and Supplementary Figure S13C). According to the hydrolysis rate measured at saturating ATP concentration (4 mM) and optimal concentrations of DNA and R.CgII, one molecule of N.CgII hydrolyses $\sim 176 \pm 13$ ATP/s whilst N.NgoAVII hydrolyses $\sim 174 \pm 25$ ATP/s.

We also studied the divalent metal ion dependence on the ATPase activity. The highest ATP hydrolysis rate was observed with Mg^{2+} , whilst a lower rate (up to 2.5-fold in comparison to Mg^{2+}) was determined with Mn^{2+} or Ca^{2+} (Supplementary Figure S14). No significant ATPase activity was detected in reactions containing EDTA (Supplementary Figure S14).

DNA cleavage studies

DNA cleavage by the R-proteins was investigated using linear phage λ DNA as a substrate (which contains 181 CgII/NgoAVII recognition sites) (Figure 4 and Supplementary Figure S15A). The R-proteins alone did not cleave the DNA to any detectable extent. Reactions with supercoiled pBR322 (which contains 21 recognition sites) did not produce any detectable increase in nicked DNA, indicating that the isolated R-proteins cannot introduce a single-stranded DNA break (data not shown). However, both phage λ DNA (Figure 4) and pBR322 (data not shown) were fragmented by the R-proteins in the presence of their cognate N-proteins and ATP. Active site mutations either within the nucleolytic PLD-superfamily domain of the R-protein or within the SF2-helicase domain of the N-protein abolished DNA cleavage. DNA fragmentation was not observed in the absence of ATP or in the presence of ADP

DNA cleavage by the R-proteins

The R-proteins possess an N-terminal nucleolytic PLD-superfamily domain, also identified in the dimeric Type IIS REase BfiI (Supplementary Figure S2) (12,34). As observed with BfiI (34), we show here that the isolated R-proteins can dimerize (Table 1). They also specifically interact with the cognate sequence 5'-GCCGC-3', suggesting that their B3 domains are responsible for target recognition, in analogy to BfiI (Figure 2 and Supplementary Figure S12). The crystal structure of apo-R.NgoAVII revealed that R.NgoAVII and BfiI are indeed structurally very similar proteins (see the accompanying paper) (9,10). As seen with BfiI, two PLD domains within the R.NgoAVII dimer form a single nuclease active site. The isolated B3 domains of R.NgoAVII (and BfiI) specifically interact with the recognition sequence as monomers (see the accompanying paper) (9,35). The R-proteins also hydrolyze the artificial substrate bis(*p*-nitrophenyl) phosphate, as demonstrated for BfiI. However, in contrast to BfiI, the R-proteins dimers in isolation were unable to cleave cognate DNA (Stonyte, Toliuis and Siksny, unpublished results) (12). We show here that, in contrast to BfiI, they require interaction with the catalytically active ATPase subunits N.CgII or N.NgoAVII and ATP hydrolysis (Figure 4A and Supplementary Figure S15A). All natural deoxy- and ribonucleotides ((d)NTPs) could be utilized for DNA cleavage (data not shown). The dependence on divalent metal ions is likely due to the requirements of the N-proteins, since the nucleolytic PLD-superfamily domains of the R-proteins should be active without metal ions (12). Mutations of the conserved amino-acids residues from the HXK active site motifs within the nucleolytic PLD-superfamily domain (mutation H105A in R.CgII and H104A in R.NgoAVII) abolished DNA cleavage.

Both R-proteins cleaved DNA at fixed positions downstream of the asymmetric recognition sequence 5'-GCCGC-3': R.CgII cuts the top and bottom DNA strands 7 and 6/7, R.NgoAVII - both strands 7 nucleotides downstream of the target, respectively (Figure 4B and Supplementary Figure S15B). The cleavage positions resemble that of BfiI, which cuts the top and bottom DNA strands 5 and 4 nucleotides downstream of its recognition sequence 5'-ACTGGG-3', respectively (36). Thus the cleavage pattern can be said to be Type IIS-like.

Functional complex

According to the experimental data presented here, the R- and N-proteins act together as a heterotetrameric complex to hydrolyze ATP and thus cleave DNA. Based on the crystal structure of R.NgoAVII (see the accompanying paper) (9), the gel filtration data and the SAXS data, we suggest a model for the RN.CgII complex (Figure 5); although a stable NgoAVII R₂N₂ complex was not obtained *in vitro*, we suggest a similar complex is likely exist *in vivo*. R.CgII dimerizes through its PLD domains, to form a single nuclease active site. The B3 domains within this dimer are situated on the opposite sides of the PLD domain core and make no contacts to each other. N.CgII uses its Z1 domains for dimerization, with the DEAD domains responsible for interaction with R.CgII. On cognate DNA, an R-protein

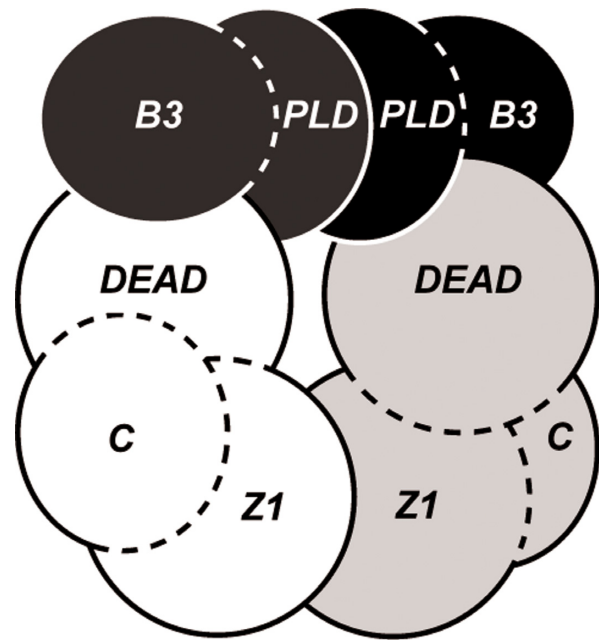


Figure 5. Model of the RN.CgII complex. R.CgII is composed of the PLD and B3 domains (colored black and dark gray). N.CgII contains the DEAD, Z1 and C-terminal domains (colored white and light gray).

within the complex will first should bind the recognition sequence. This triggers the N-protein to hydrolyze ATP; we suggest that the ATP hydrolysis levels are consistent with translocation of the complex along DNA. DNA cleavage should proceed after translocation because the catalytically inactive R-protein stimulates the ATPase activity of the N-protein and not *vice versa*.

Interestingly, DNA cleavage occurs at fixed position near to the recognition sequence as in the case of ATP-independent BfiI. This raises the question: how can translocation away from the site then lead to cleavage close to the site? According to the crystal structure of R.NgoAVII, the R₂N₂ complex should contain a single nuclease active site that could produce a double-stranded DNA break through sequential strand cleavage. Instead one could speculate that two RN-complexes are required to interact and form a catalytically competent supercomplex (possibly one translocating along DNA and a second bound to the recognition sequence). In this model the two interacting R-protein dimers would each cut one strand to introduce a double-stranded DNA break near to the occupied target site. Such a mechanism is reminiscent of that suggested for the Type III restriction enzymes, although communication between the sites in the Type III case is likely by diffusion and not translocation (33,37). Experimental studies are now underway to elucidate the cleavage and possible communication mechanisms for CgII and NgoAVII endonucleases.

Homologs of the CgII and NgoAVII RM systems

Seven more homologs of the CgII and NgoAVII RM systems were previously identified in sequenced microbial genomes (7). Two of them, the BceSIV and Bce14579I RM systems from *B. cereus* (ATCC 10987 and 14579, respec-

Table 2. Comparison of the CgII/NgoAVII and Type I, II and III RM systems

Feature	CgII/NgoAVII	Type I e.g. EcoR124I	Type IIS e.g. BfiI	Type III e.g. EcoPI
Proteins	M (methylase) R (endonuclease) N (ATPase)	HsdS (target recognition) HsdM (methylase) HsdR (helicase-endonuclease)	M1 (methylase) M2 (methylase) R (endonuclease)	Mod (target recognition, methylase) Res (helicase-endonuclease)
Methyltransferase	M	M ₂ S ₁ , R ₂ M ₂ S ₁	M1, M2 ^b	Mod ₂ Res ₁ , Mod ₂
Endonuclease	R ₂ N ₂ ^a	R ₂ M ₂ S ₁	R ₂	Mod ₂ Res ₁
Nuclease domain	PLD	PD-(D/E)XK	PLD	PD-(D/E)XK
Helicase domain	SF2 (DEAD)	SF2 (DEAD)		SF2
Recognition sequence	Short, asymmetric GCCGC	Asymmetric, bipartite GAANNNNNNR ₂ TCG	Short, asymmetric ACTGGG	Short, asymmetric AGACC
DNA cleavage position	Fixed	Variable	Fixed	Fixed
Cofactors required for DNA hydrolysis	Mg ²⁺ , ATP	Mg ²⁺ , ATP, AdoMet		Mg ²⁺ , ATP
ATPase rate	~170 ATP/s/monomer (37°C)	~998 ATP/s/motor (30°C)		~0.7 ATP/s/complex (37°C)

Data for EcoR124I was collated from (31,38). Data for EcoPI was collated from (29,38,39) and for BfiI was collated from (12,34,36).

^aComplex stoichiometry was determined only for CgII.

^bThere are two independent BfiI methylases each of them modifying bases on the opposite strands of the asymmetric recognition site (12).

tively), have been biochemically characterized (7). Analogous to the CgII and NgoAVII enzymes, BceSIV recognizes the 5'-GCWGC-3' sequence (W stands for A and T) and acts as a complex of a PLD-superfamily nuclease and an ATPase. ATP or GTP hydrolysis was required for DNA cleavage by BceSIV. According to product analysis of partially digested plasmid using run-off sequencing it was determined that BceSIV cuts DNA on both sides of the target site in somewhat asymmetrical manner (/N9-11 GCWGC N5-7/), although this remains to be verified. According to bioinformatic analysis over 95 homologs of CgII/NgoAVII were identified in which the PLD-nuclease and ATPase genes are fused (7). We propose that the CgII/NgoAVII enzymes are prototypes of a new class of RM systems with a shared and novel catalytic mechanism.

Comparison with Type I, II and III RM systems

The CgII and NgoAVII RM systems were previously assigned as Type II RM systems (REBASE, <http://rebase.neb.com>) (8). From the data presented in this paper, we can reconsider the previous assignment and compare them with the classical Type I, II and III RM systems (Table 2). The methyltransferases M.CgII and M.NgoAVII recognize a short DNA sequence (5'-GCSGC-3') and are active separately from the REases consistent with a Type II classification (Supplementary Figure S5). They are also share sequence homology with *bona fide* Type II 5-cytosine MTases (16,17). Also consistent with the original classification, the R-proteins cleave DNA at fixed positions downstream of their asymmetric recognition sequences, similarly to Type IIS REases as exemplified by the structurally-related BfiI (Table 2) (see the accompanying paper) (9). However, DNA cleavage by CgII and NgoAVII is dependent on ATP hydrolysis performed by the helicase-like N-proteins (Table 2). The Type I and III restriction enzymes also contain SF2 helicase-like motifs and require ATP hydrolysis for DNA cleavage (Table 2) (38,39). The ATP consumption levels sug-

gest that CgII and NgoAVII may function similarly to Type I restriction enzymes. However, their subunit composition and cleavage patterns are quite different (Table 2). Taken together, the CgII and NgoAVII REases are structurally distinctive and appear to employ a new catalytic mechanism for DNA cleavage. Further mechanistic analysis will reveal whether these enzymes can be shoehorned into one of the classical types, or whether they are indeed prototypes of an entirely distinct class of restriction-modification system.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The authors are grateful to Dr Violeta Stonytė-Morin for sharing pET15b-R.NgoAVII plasmid. Authors acknowledge Dr Zita Liutkevičiūtė for the mass spectrometry and colleagues at the Vilnius Department of Protein–DNA Interactions for valuable discussions and suggestions for this work.

FUNDING

Research Council of Lithuania [MIP-029/2012 to M.Z.]; Wellcome Trust [084086 to M.D.S.]. Funding for open access charge: Wellcome Trust [084086 to M.D.S.].
Conflict of interest statement. None declared.

REFERENCES

- Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T.F., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Schafer,A., Kalinowski,J. and Puhler,A. (1994) Increased fertility of *Corynebacterium glutamicum* recipients in intergeneric matings with *Escherichia coli* after stress exposure. *Appl. Environ. Microbiol.*, **60**, 756–759.

3. Schafer, A., Schwarzer, A., Kalinowski, J. and Puhler, A. (1994) Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from *Corynebacterium glutamicum* ATCC 13032 and analysis of its role in intergeneric conjugation with *Escherichia coli*. *J. Bacteriol.*, **176**, 7309–7319.
4. Dang, T.N., Zhang, L., Zollner, S., Srinivasan, U., Abbas, K., Marrs, C.F. and Foxman, B. (2013) Uropathogenic *Escherichia coli* are less likely than paired fecal *E. coli* to have CRISPR loci. *Infect. Genet. Evol.*, **19**, 212–218.
5. Frunzke, J., Bramkamp, M., Schweitzer, J.E. and Bott, M. (2008) Population Heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J. Bacteriol.*, **190**, 5111–5119.
6. Nanda, A.M., Heyer, A., Kramer, C., Grunberger, A., Kohlheyer, D. and Frunzke, J. (2014) Analysis of SOS-induced spontaneous prophage induction in *Corynebacterium glutamicum* at the single-cell level. *J. Bacteriol.*, **196**, 180–188.
7. Xu, S.Y., Nugent, R.L., Kasamkattil, J., Fomenkov, A., Gupta, Y., Aggarwal, A., Wang, X., Li, Z., Zheng, Y. and Morgan, R. (2012) Characterization of type II and III restriction-modification systems from *Bacillus cereus* strains ATCC 10987 and ATCC 14579. *J. Bacteriol.*, **194**, 49–60.
8. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
9. Tamulaitiene, G., Silanskas, A., Grazulis, S., Zaremba, M. and Siksnys, V. (2014) Crystal structure of the R-protein of the multisubunit ATP-dependent restriction endonuclease NgoAVII. *Nucleic Acids Res.*
10. Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R. and Siksnys, V. (2005) Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15797–15802.
11. Sasnauskas, G., Halford, S.E. and Siksnys, V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 6410–6415.
12. Sapranuskas, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, 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.
13. Hu, Y., Li, T., Yang, Z., Zhang, B. and Li, Y. (2008) Phage resistance of *Corynebacterium crenatum* conferred by the restriction and modification system cglI. *Sheng Wu Gong Cheng Xue Bao*, **24**, 760–765.
14. Iyer, L.M., Abhiman, S. and Aravind, L. (2008) MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Direct*, **3**, 8.
15. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
16. Stein, D.C., Gunn, J.S., Radlinska, M. and Piekarowicz, A. (1995) Restriction and modification systems of *Neisseria gonorrhoeae*. *Gene*, **157**, 19–22.
17. Schafer, A., Tauch, A., Droste, N., Puhler, A. and Kalinowski, J. (1997) The *Corynebacterium glutamicum* cglIM gene encoding a 5-cytosine methyltransferase enzyme confers a specific DNA methylation pattern in an McrBC-deficient *Escherichia coli* strain. *Gene*, **203**, 95–101.
18. Zheng, L., Baumann, U. and Reymond, J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
19. Petoukhov, M.V., Konarev, P.V., Kikhney, A.G. and Svergun, D. (2007) ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.*, **40**, s223–s228.
20. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.
21. Fischer, H., de Oliveira Neto, M., Napolitano, H.B., Polikarpov, I. and Craievich, A.F. (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.*, **43**, 101–109.
22. 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.
23. Peakman, L.J., Antognozzi, M., Bickle, T.A., Janscak, P. and Szczelkun, M.D. (2003) S-adenosyl methionine prevents promiscuous DNA cleavage by the EcoPII type III restriction enzyme. *J. Mol. Biol.*, **333**, 321–335.
24. Cordin, O., Banroques, J., Tanner, N.K. and Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17–37.
25. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
26. Szczelkun, M.D. (2013) Roles for helicases as ATP-dependent molecular switches. *Adv. Exp. Med. Biol.*, **767**, 225–244.
27. Obarska-Kosinska, A., Taylor, J.E., Callow, P., Orłowski, J., Bujnicki, J.M. and Kneale, G.G. (2008) HsdR subunit of the type I restriction-modification enzyme EcoRI24I: biophysical characterisation and structural modelling. *J. Mol. Biol.*, **376**, 438–452.
28. Sisakova, E., Stanley, L.K., Weiserova, M. and Szczelkun, M.D. (2008) A RecB-family nuclease motif in the Type I restriction endonuclease EcoRI24I. *Nucleic Acids Res.*, **36**, 3939–3949.
29. Janscak, P., Sandmeier, U., Szczelkun, M.D. and Bickle, T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoPII and EcoPI5I. *J. Mol. Biol.*, **306**, 417–431.
30. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.
31. Seidel, R., Bloom, J.G., Dekker, C. and Szczelkun, M.D. (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoRI24I. *EMBO J.*, **27**, 1388–1398.
32. Stanley, L.K., Seidel, R., van der Scheer, C., Dekker, N.H., Szczelkun, M.D. and Dekker, C. (2006) When a helicase is not a helicase: dsDNA tracking by the motor protein EcoRI24I. *EMBO J.*, **25**, 2230–2239.
33. Schwarz, F.W., Toth, J., van Aelst, K., Cui, G., Clausing, S., Szczelkun, M.D. and Seidel, R. (2013) The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science*, **340**, 353–356.
34. 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.
35. Golovenko, D., Manakova, E., Zakrys, L., Zaremba, M., Sasnauskas, G., Grazulis, S. and Siksnys, V. (2014) Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal structure of the C-terminal fragment of BfiI restriction enzyme. *Nucleic Acids Res.*, **42**, 4113–4122.
36. Vitkute, J., Maneliene, Z., Petrusyte, M. and Janulaitis, A. (1998) BfiI, a restriction endonuclease from *Bacillus firmus* S8120, which recognizes the novel non-palindromic sequence 5'-ACTGGG(N)5/4-3'. *Nucleic Acids Res.*, **26**, 3348–3349.
37. van Aelst, K., Toth, J., Ramanathan, S.P., Schwarz, F.W., Seidel, R. and Szczelkun, M.D. (2010) Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to-head and tail-to-tail inverted repeat. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 9123–9128.
38. Smith, R.M., Diffin, F.M., Savery, N.J., Josephsen, J. and Szczelkun, M.D. (2009) DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. *Nucleic Acids Res.*, **37**, 7206–7218.
39. Butterer, A., Pernstich, C., Smith, R.M., Sobott, F., Szczelkun, M.D. and Toth, J. (2014) Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Res.*, **42**, 5139–5150.