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

Mindaugas Zaremba1,*,†, Paulius Toliusis1,†, Rokas Grigaitis1, Elena Manakova1, Arunas Silanskas1, Giedre Tamulaitiene1, Mark D. Szczelkun2 and Virginijus Siksnys1,*

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 plasmid 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-

© The Author(s) 2014. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
The M.CglI 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'-GCGGC-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.CglI and R.NgoAVII (also referred to as ‘R-proteins’) and cognate ATPases N.CglI 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.CglI and N.CglI, the RN.CglI complex with an R2N2 stoichiometry was also isolated. We demonstrate that both R.CglI 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'-GCGGC-3'. On the basis of their activities, we propose that CglI 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 CglI 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.CglI/N.CglI, R.NgoAVII/N.NgoAVII or DNA fragments corresponding to individual domains that were later 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.CglI (H105A), N.CglI (D158A+E159A), R.NgoAVII (H104A) and N.NgoAVII (D150A+E151A) mutants were produced by QuickChange 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 μl 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.CglI 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.CglI, R.NgoAVII, N.CglI 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 $s = 4 \pi \sin(\theta)/\lambda$, where $2\theta$ is a scattering angle and $\lambda$ is x-ray wavelength, 0.124 nm) range from 0.075759 to 4.66533 nm$^{-1}$, except the R.CglI 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). $D_{\text{max}}$ was estimated manually by interactive GNOM runs and compared with the value calculated by DATGNOM (19). Unprocessed scattering data /s/ with subtracted buffer scattering, Kratyky 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 CglI and NgoAVII proteins was analyzed by the electrophoretic mobility shift assay (EMSA) using a 33 bp cognate duplex (5′-TACATCGATAATAGCGCGTGCAGGCCATT-3′/3′-ATGTTAGCTATTTAGCCGCGCAGACTCGGGTAA-5′, the recognition sequence is underlined) or a 33 bp non-cognate DNA duplex (5′-TACATCGATAATAGCGCGTGCAGGCCATT-3′/3′-ATGTTAGCTATTTAGCCGCGCAGACTCGGGTAA-5′). DNA (final concentration 1 nM) was incubated with proteins (final concentrations from 1 to 500 nM monomer) for 15 min in 20 μl of 40 mM Tris-aceate (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-aceate, 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-aceate. Samples were separated at room temperature for 3 h at ~6 V/cm. Radiolabeled DNA was detected and quantified using the Cyclone phosphorimagery 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-aceate (pH 7.0 at 25°C), 0.1 mg/ml BSA, 10 mM Mg-aceate and 150 mM K-aceate (only for CglI proteins), supplemented with 50 μM [α-32P]ATP (Hartmann Analytic, Braunschweig, Germany) and 0.01 μg/μl of various DNAs (as indicated): (i) phage λ DNA (dam ′ dcm ′ ), (ii) pBR322, (iii) pBR322 methylated at the CglI/NgoAVII recognition sites, (iv) cognate fragment, (v) noncognate fragment, (vi) cognate oligoduplex (5′-GATCGAGCCCGCTACG-3′/3′-CTAGCTCGCGGCGTGC-5′, the recognition sequence is underlined), (vii) non-cognate oligoduplex (5′-GATCGACAGGCTACG-3′/3′-CTAGCTGTCGCGTGC-5′) and (viii) single-stranded M13mp18 (Supplementary Data). Reactions were initiated by adding the CglI (40 nM) or NgoAVII (40 or 100 nM) proteins and were stopped by addition of EDTA to 30 mM. An aliquot (1 μl) 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 μg/μl 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 CglI, respectively. Reactions were initiated by adding N.CglI or N.NgoAVII (10 nM) plus R.CglI 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-aceate (pH 7.0 at 25°C); 0.1 mg/ml BSA; either 50 nM (for NgoAVII) or 150 nM (for CglI) K-aceate; either 2.5 mM (for NgoAVII) or 10 mM (for CglI) divalent metal ions (Mg2+, Mn2+ or Ca2+); 1 mM DTT; 2 or 4 mM adenyl-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 μg/μl of various dsDNAs (as indicated): (i) phage λ DNA (dam ′ dcm ′ ), (ii) pBR322, (iii) pBR322 methylated at the CglI/NgoAVII recognition sites, (iv) pUC19 or (v) 0-site plasmid (without CglI/NgoAVII recognition sequences.
Table 1. Oligomeric state of the CglI and NgoAVII proteins and their domains

<table>
<thead>
<tr>
<th>Protein/domain/complex</th>
<th>Mw, kDa</th>
<th>Calculated&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Gel filtration, M&lt;sub&gt;W(exp)&lt;/sub&gt;/M&lt;sub&gt;W(calc)&lt;/sub&gt;</th>
<th>SAXS, M&lt;sub&gt;W(exp)&lt;/sub&gt;/M&lt;sub&gt;W(calc)&lt;/sub&gt;</th>
<th>Oligomeric state</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.CglI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>41.2</td>
<td>85.0</td>
<td>84.7</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>R.CglI-PLD</td>
<td>21.8</td>
<td>30.4</td>
<td>2.1</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>R.CglI-B3</td>
<td>23.1</td>
<td>26.3</td>
<td>1.1</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.CglI&lt;sup&gt;c&lt;/sup&gt;</td>
<td>71.9</td>
<td>128.8</td>
<td>136.1</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>N.CglI-DEAD</td>
<td>28.4</td>
<td>29.5</td>
<td>1.0</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.CglI-DEAD-Z1</td>
<td>55.3</td>
<td>102.8</td>
<td>1.9</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>N.CglI-Z1</td>
<td>28.1</td>
<td>42.4</td>
<td>1.5</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>N.CglI-Z1-C</td>
<td>46.7</td>
<td>91.2</td>
<td>2.0</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>N.CglI-C</td>
<td>19.8</td>
<td>17.9</td>
<td>0.9</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.CglI&lt;sup&gt;f&lt;/sup&gt;</td>
<td>113.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>243.6</td>
<td>231.8</td>
<td>Heterotetramer (R&lt;sub&gt;N&lt;/sub&gt;N&lt;sub&gt;N&lt;/sub&gt;)</td>
<td></td>
</tr>
<tr>
<td>R.NgoAVII&lt;sup&gt;f&lt;/sup&gt;</td>
<td>42.0</td>
<td>69.2</td>
<td>87.8</td>
<td>Dimer</td>
<td></td>
</tr>
<tr>
<td>R.NgoAVII-B3</td>
<td>22.8</td>
<td>22.9</td>
<td>1.0</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII&lt;sup&gt;f&lt;/sup&gt;</td>
<td>72.2</td>
<td>81.6</td>
<td>86.1</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII-DEAD</td>
<td>28.7</td>
<td>35.3</td>
<td>1.2</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII-DEAD-Z1</td>
<td>55.2</td>
<td>57.2</td>
<td>1.0</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII-Z1</td>
<td>27.7</td>
<td>22.3</td>
<td>0.8</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII-Z1-C</td>
<td>46.7</td>
<td>46.8</td>
<td>1.0</td>
<td>Monomer</td>
<td></td>
</tr>
<tr>
<td>N.NgoAVII-C</td>
<td>20.2</td>
<td>20.4</td>
<td>1.0</td>
<td>Monomer</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Theoretical molecular weights were calculated using the ProtParam tool, http://web.expasy.org/protparam.
<sup>b</sup>Full-length proteins and their complexes are shown on grey background.
<sup>c</sup>This 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) bromphenol 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 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) bromphenol 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.

RESULTS

Expression and purification of proteins

The R- and N-proteins and their domains were expressed with either His<sub>6</sub>- 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 HXK-active site motif of the PLD-superfamily nuclease domain; and D158+159 (N.CglI), and D150+151 (N.NgoAVII) located in the DEAD domain and involved in Mg<sup>2+</sup> 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.CglI (containing a StrepII-tag) and N.CglI (containing a His6-tag) formed the RN.CglI 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.CglI and its isolated PLD domain eluted from the column as dimers, while the isolated R.CglI-B3 domain eluted as a monomer (Supplementary Figure S6, Table 1). N.CglI, 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.CglI complex, obtained by co-expression of both the R.CglI and N.CglI proteins, was determined by gel filtration and SAXS to be close to the sum of molecular weights of the R.CglI and N.CglI dimers (Table 1, Supplementary Figures S4 and S6). Therefore, the R.CglI and N.CglI dimers assemble into a protein complex with a R2N2 stoichiometry. In addition, the stoichiometry of the RN.CglI complex was examined using densitometric analysis of the purified complex following SDS/PAGE electrophoresis (Supplementary Figure S8). The calculated molar ratio of the R.CglI and N.CglI proteins within the complex was 1:0.75. Since both proteins are dimers in solution, this ratio supports the R2N2 stoichiometry. The R2N2 CglI complex was also obtained by mixing the individual R.CglI and N.CglI proteins in vitro (data not shown). Higher-order complex formation was also observed upon mixing of the full-length R.CglI protein with either the N.CglI-DEAD domain or the N.CglI-DEAD-Z1 domain (Supplementary Figure S9). In contrast, a higher-order complex was not obtained if the PLD domain or B3 domain of R.CglI were mixed with the corresponding domains of N.CglI (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).

**DNA binding studies**

DNA binding by the CglI and NgoAVII proteins was studied by EMSA using 32P-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.CglI complex bound the cognate oligoduplex into a protein–DNA complex with the same electrophoretic mobility as the DNA complex formed by the isolated R.CglI protein (Figure 2).

We rationalize these results as showing that the RN.CglI 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 [α32P]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.CglI (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.CglI (H105A) and R.NgoAVII (H104A) had
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.CglI 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 CglI 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.CglI, one molecule of N.CglI hydrolyses ~176 ± 13 ATP/s whilst N.NgoAVII hydrolyses ~174 ± 25 ATP/s.

We also studied the divalent metal ion dependence on the ATPase activity. The highest ATP hydrolysis rate was observed with Mg2+, whilst a lower rate (up to 2.5-fold in comparison to Mg2+) was determined with Mn2+ or Ca2+ (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 CglI/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.
endonuclease and an ATPase, the last two of which are coupled (Figure 1).

ATP hydrolysis by the N-proteins

The N-terminal domains of the N-proteins carry ATPase motifs characteristic of the helicases of the DEAD-superfamily (Supplementary Figure S1) (15). Helicases use ATP hydrolysis to unwind and/or translocate nucleic acids, and/or to remodel nucleic acids or nucleoprotein complexes (15,24,25). The Type I and III RM enzymes both contain domains with SF2 helicase motifs and require ATP hydrolysis for DNA cleavage (see below) (26). However, the DEAD domains within the N-proteins: have helicase motifs distinct to those found in Type I or III enzymes; are associated with uncharacterized Z1 domains that are not found associated with the helicase domains in Type I or III enzymes (14); and do not have the nucleosome domain fused within a single polypeptide as seen in the HsdR subunits of Type I enzymes and Res subunits of Type III enzymes (27–29). We demonstrate here that the N-proteins do have ATPase activity on longer DNA (>16 bp) with non-methylated recognition sequences 5′-GCGGC-3′, but only when in complex with the cognate R-protein (Figure 3A and 3B, Supplementary Figures S13A and S13B). Both N.CglI and N.NgoAVII, like other DEAD-superfamily helicases, required divalent metal ions (Mg2+, Mn2+ or Ca2+) for ATPase activity and were inactive in the presence of EDTA (Supplementary Figure S14) (30). Mutations of the conserved amino-acid residues responsible for Mg2+ coordination (mutations D158A+E159A in N.CglI and D150A+E151A in N.NgoAVII) abolished ATP hydrolysis and cleavage (Figure 3A and Supplementary Figure S13A). However, active site mutations in R.CglI (H105A) or R.NgoAVII (H104A), did not affect the ATPase activity of the cognate N-proteins. The N-proteins showed weak preference for ATP and dATP, although, other deoxy- and ribonucleotides were also hydrolyzed (Figure 3C and Supplementary Figure S13C).

Our ATPase data indicates that the R- and N-proteins from either the CglI or NgoAVII systems act together within a single protein–DNA complex. Indeed, only the R-proteins show a cognate DNA binding activity through their B3 domains (Figure 2, Supplementary Figures S11 and S12), that presumably triggers the DNA-dependent ATPase activity of the N-proteins. Moreover, a CglI complex with an R2N2 stoichiometry could be expressed and isolated (Table 1), or reconstituted in vitro. The maximum rate of ATP hydrolysis of the complexes was ∼176 ± 13 and ∼174 ± 25 molecules per second per monomer for N.CglI and N.NgoAVII, respectively. These values are closer to those reported for Type I restriction enzymes (e.g. EcoR124I ~998 ATP/s/monomer at 30°C) that use ATP hydrolysis (1–1.5 ATP per bp step) for translocation on intact DNA (26,31,32) rather than to those reported for Type III enzymes (~0.3–0.6 ATP/s/monomer), where ATP hydrolysis is used to drive a molecular switches to activate DNA sliding (33).
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 Siksnys, unpublished results) (12). We show here that, in contrast to BfiI, they require interaction with the catalytically active ATPase subunits N.CglI 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.CglI 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.CglI 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.CglII 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.CglI 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.CglI uses its ZI domains for dimerization, with the DEAD domains responsible for interaction with R.CglI. On cognate DNA, an R-protein 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 CglI and NgoAVII endonucleases.

Homologs of the CglI and NgoAVII RM systems

Seven more homologs of the CglI and NgoAVII RM systems were previously identified in sequenced microbial genomes (7). Two of them, the BecSIV and Bec14579I RM systems from B. cereus (ATCC 10987 and 14579, respec-
Comparing the CglI/NgoAVII RM systems, the CglI and NgoAVII enzymes, BeceSIV recognizes the 5′-GCWG-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 BeceSIV. According to product analysis of partially digested plasmid using run-off sequencing it was determined that BeceSIV cuts DNA on both sides of the target site in a somewhat asymmetrical manner (/N9-11 GCWG CN5-7/), although this remains to be verified. According to bioinformatic analysis over 95 homologs of CglI/NgoAVII were identified in which the PLD-nuclease and ATPase genes are fused (7). We propose that the CglI/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 CglI 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.CglI and M.NgoAVII recognize a short DNA sequence (5′-CGGC-3′) and are active separately from the REases consistent with a Type II classification (Supplementary Figure S5). They are also shared 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 CglI 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 suggest that CglI 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 CglI 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


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

<table>
<thead>
<tr>
<th>Feature</th>
<th>CglI/NgoAVII</th>
<th>Type I e.g. EcoR124I</th>
<th>Type IIS e.g. BfiI</th>
<th>Type III e.g. EcoPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteins</td>
<td>M (methylase)</td>
<td>HsdS (target recognition)</td>
<td>M1 (methylase)</td>
<td>Mod (target recognition, methylase)</td>
</tr>
<tr>
<td>Methyltransferase</td>
<td>R (endonuclease)</td>
<td>Msd (endonuclease)</td>
<td>M2 (methylase)</td>
<td>R (endonuclease)</td>
</tr>
<tr>
<td>Endonuclease</td>
<td>N (ATPase)</td>
<td>HsdM (methylase)</td>
<td>R (endonuclease)</td>
<td>Res (endonuclease)</td>
</tr>
<tr>
<td>Recognition sequence</td>
<td>Short, asymmetric</td>
<td>Asymmetric, bipartite</td>
<td>Short, asymmetric</td>
<td>ACTGGGC</td>
</tr>
<tr>
<td>DNA cleavage position</td>
<td>Fixed</td>
<td>Variable</td>
<td>Fixed</td>
<td>Fixed</td>
</tr>
<tr>
<td>DNA hydrolysis</td>
<td>ATPase rate</td>
<td>ATPase/s/monomer</td>
<td>ATPase/s/motor</td>
<td>ATPase/s/complex</td>
</tr>
<tr>
<td>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).</td>
<td>ATPase rate</td>
<td>ATPase/s/monomer</td>
<td>ATPase/s/motor</td>
<td>ATPase/s/complex</td>
</tr>
</tbody>
</table>

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).

Complex stoichiometry was determined only for CglI.

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


