

# CgII cleaves DNA using a mechanism distinct from other ATP-dependent restriction endonucleases

Paulius Toliūsis<sup>1,†</sup>, Mindaugas Zaremba<sup>1,\*</sup>, Arunas Silanskas<sup>1</sup>, Mark D. Szczelkun<sup>2</sup> and Virginijus Siksnys<sup>1,\*</sup>

<sup>1</sup>Department of Protein–DNA Interactions, Institute of Biotechnology, Vilnius University, Sauletekio al. 7, LT-10257, Vilnius, Lithuania and <sup>2</sup>DNA–Protein Interactions Unit, School of Biochemistry, Biomedical Sciences Building, University of Bristol, Bristol, BS8 1TD, UK

Received April 6, 2017; Revised June 20, 2017; Editorial Decision June 21, 2017; Accepted June 28, 2017

## ABSTRACT

The restriction endonuclease CgII from *Corynebacterium glutamicum* recognizes an asymmetric 5'-GCCGC-3' site and cleaves the DNA 7 and 6/7 nucleotides downstream on the top and bottom DNA strands, respectively, in an NTP-hydrolysis dependent reaction. CgII is composed of two different proteins: an endonuclease (R.CgII) and a DEAD-family helicase-like ATPase (H.CgII). These subunits form a heterotetrameric complex with R<sub>2</sub>H<sub>2</sub> stoichiometry. However, the R<sub>2</sub>H<sub>2</sub>.CgII complex has only one nuclease active site sufficient to cut one DNA strand suggesting that two complexes are required to introduce a double strand break. Here, we report studies to evaluate the DNA cleavage mechanism of CgII. Using one- and two-site circular DNA substrates we show that CgII does not require two sites on the same DNA for optimal catalytic activity. However, one-site linear DNA is a poor substrate, supporting a mechanism where CgII complexes must communicate along the one-dimensional DNA contour before cleavage is activated. Based on experimental data, we propose that adenosine triphosphate (ATP) hydrolysis by CgII produces translocation on DNA preferentially in a downstream direction from the target, although upstream translocation is also possible. Our results are consistent with a mechanism of CgII action that is distinct from that of other ATP-dependent restriction-modification enzymes.

## INTRODUCTION

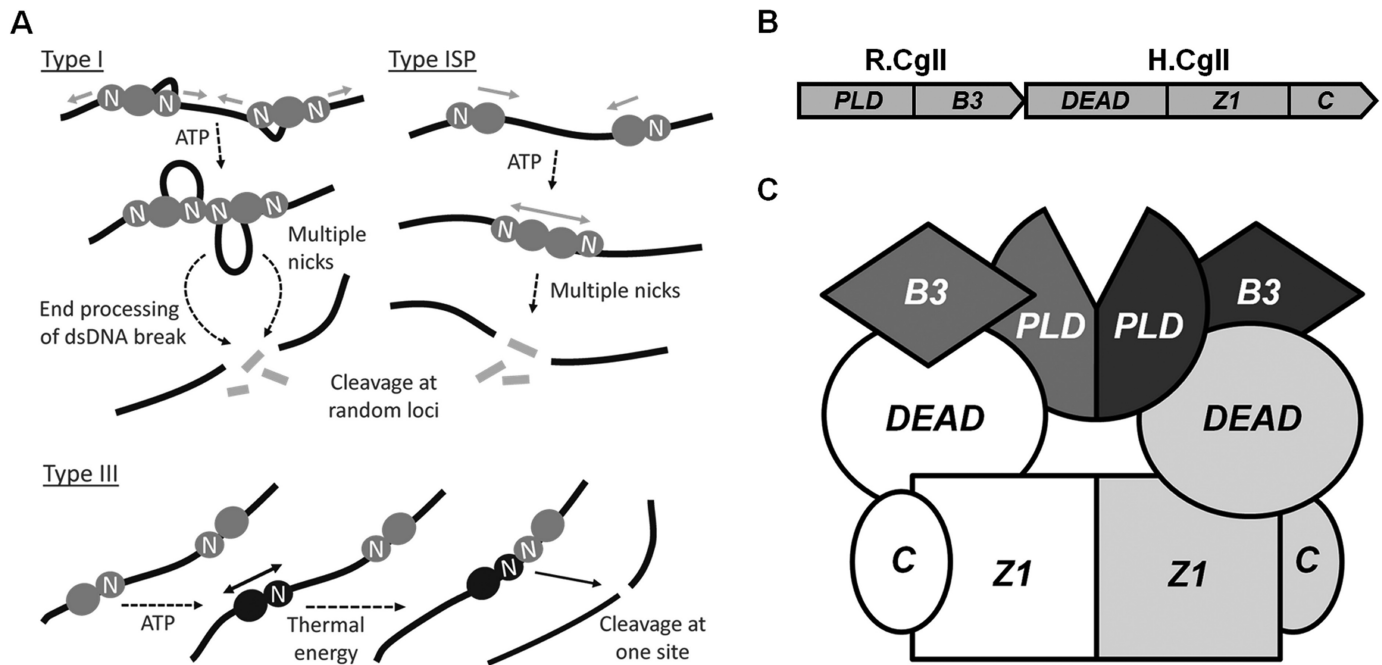
Restriction-modification (RM) enzymes are important player in protecting bacterial cells against invasion by foreign DNA (1–3). The Types I, II and III RM enzymes are

composed of two main enzyme activities: an endonuclease and a methyltransferase. The methyltransferases methylate DNA targets and protect host genomes from self-cleavage by the cognate endonucleases. Foreign DNA that has unmodified targets will be degraded by the introduction of double strand DNA breaks by the endonuclease. Type II restriction enzymes differ from Type I and III restriction enzymes in several ways. A principal difference is that they do not require adenosine triphosphate (ATP) for DNA cleavage (4). In contrast, all Type I, Type for single polypeptide (ISP) and Type III RM enzymes studied to-date use ATP hydrolysis for a long-range communication between enzymes along the one-dimensional (1D) contour of the DNA (5–9). The protein–protein interaction between communicating subunits activates DNA cleavage. In some instances the relative orientation of the asymmetric targets is also important (called ‘site orientation preference’). Examination of the communication mechanism and of the effect of different arrangements of targets on DNA cleavage activity can distinguish between the different mechanisms:

- i) For the classical Type I RM enzymes, there are two independent motors subunits (HsdR) that catalyze step-wise translocation along the DNA up- and downstream from the site; i.e. translocation is simultaneously bidirectional (Figure 1A) (10–12). Each motor consumes at least one ATP to move 1 bp and translocation can be measured using the triplex displacement assay (10,13). DNA cleavage occurs at random non-specific sites when two motors collide along the 1D DNA contour (forming a ‘collision complex’). On a plasmid this can be achieved by two motors from the same complex and so a single site is sufficient to produce cleavage. However, on linear substrate at least two DNA sites are needed; with a single site each motors run to the DNA ends and cannot collide. The Type I target sites are asymmetric but since translocation is bidirectional, there is absolutely no effect of the relative orientation of pairs of sites on

\*To whom correspondence should be addressed. Tel: +370 5 2234 357; Fax: +370 5 2234 367; Email: zare@ibt.lt  
Correspondence may also be addressed to Virginijus Siksnys. Tel: +370 5 2234 359; Fax: +370 5 2234 367; Email: siksnys@ibt.lt

†These authors contributed equally to the paper as first authors.



**Figure 1.** The adenosine triphosphate (ATP)-dependent restriction-modification (RM) enzymes. (A) Mechanisms of DNA cleavage by Type I, ISP and III RM enzymes. Nuclease subunit/domains are labeled 'N'. For Type I enzymes, each complex has two helicase–nuclease subunits which can translocate away each side of the site, drawing in two DNA loops (only one loop is shown per complex for clarity). Upon collision of two converging translocating enzymes at a non-specific site, the nucleases engage and generate a dsDNA break (2). Additional processing leads to DNA shortening and release of small fragments (54,55). For Type ISP enzymes, translocation is unidirectional. Upon collision at a non-specific site, the nuclease domains are held apart. It is proposed that movement of the collision complex results in multiple nicks that generate the dsDNA break (9,40). For the Type III enzymes, ATP hydrolysis leads to a conformation switch into a DNA sliding state. Long-range communication is bidirectional and driven by thermal energy. Upon collision with a second enzyme complex bound to a target site, the nucleases engage and generate a dsDNA break (18). (B) Domain organization of the R- and H.CgII proteins. R.CgII (NCgl1704) is a putative REase with the phospholipase D (PLD)-superfamily nucleolytic and B3-like DNA binding domain. H.CgII (NCgl1705) is predicted superfamily 2 (SF 2) helicase/ATPase containing uncharacterized Z1-superfamily and C-terminal domains. (C) Proposed model for the  $R_2H_2$ :CgII complex. R.CgII is composed of the PLD and B3 domains (colored gray and dark gray). H.CgII contains the DEAD, Z1 and C-terminal domains (colored white and light gray). Figure was made according to (28).

the efficiency or rate of DNA cleavage (i.e. there is no site orientation preference) (10).

- ii) For the single polypeptide Type ISP RM enzymes, there is only a single motor that moves the protein downstream from the asymmetric target site; i.e. translocation is strictly unidirectional (Figure 1A) (14). The motor consumes at least one ATP to move 1 bp and translocation can be measured using the triplex displacement assay (14–16). DNA cleavage by the translocating motor is activated by collision with another Type ISP enzyme. On one site plasmid or linear DNA with sites in direct repeat, the collision is described as a 'rear-end' collision and only inefficient cleavage of one strand (nicking) results. The nicking is located at a target site where the collision occurred. On linear DNA with sites in inverted repeat, efficient dsDNA cleavage is activated when the enzymes can collide 'head-on'. This can only occur where target sites are in head-to-head orientation. Linear DNA with targets in tail-to-tail orientation never produces collision events as the enzymes translocate away from one another (14–17).
- iii) For the Type III RM enzymes, there is a single ATPase subunit (Res) that undergoes a small burst of ATP hydrolysis (10–30 ATPs) while the enzyme is at the site (Figure 1A). This produces a conformation switch into

a 'DNA sliding' conformation (18). When the enzyme subsequently dissociates from the site, it can move both up- and downstream driven purely by thermal motion (19). This motion is independent of ATP hydrolysis and does not produce triplex displacement (in fact the triplex can be by-passed) (20). The original binding orientation of the enzyme is maintained during the sliding and thus cleavage shows site orientation preference (5,8,21). On plasmid or linear DNA with sites in inverted repeat (either head-to-head or tail-to-tail), dsDNA cleavage can be activated by head-on collision between a sliding enzyme and an enzyme bound to a site (5,20,22). The efficiency of cleavage of linear DNA relative to circular DNA is poor as the sliding enzymes can dissociate via the DNA ends of the former (5). Additionally, the relative cleavage efficiency of tail-to-tail sites on linear DNA can be lower if the enzyme binds tightly to its site (22). For the single site DNA or two site head-to-tail (HtT) DNA, in either linear or circular form, DNA cleavage is not observed as all collisions are in the incorrect 'rear-end' orientation and the nuclease domains cannot engage (5,20).

In addition to their important biological roles as innate immune systems against DNA invasion (23), RM enzymes are model nucleoprotein machines and study of their mech-

anisms have added significantly to fundamental knowledge about DNA binding, DNA methylation and DNA cleavage (e.g. (24–26)). For the ATP-dependent RM enzymes, there have been the additional opportunities to study how helicase-like motors utilize ATP hydrolysis for non-classical DNA translocation and switching roles (as noted above). These mechanisms have parallels to ATP-dependent motors in chromatin remodeling (27) and in mismatch repair (6). Previously, it was shown that the CgII enzyme has properties that differ from all Types of RM systems and cannot be readily assigned to any particular type (28). Examining these properties in more detail will add to our understanding of the diversity of roles for helicase-like motors.

CgII requires hydrolysis of ATP (or other nucleotide triphosphates) to produce cleavage of linear bacteriophage DNA with multiple sites in different relative orientations (28). The level of ATP hydrolysis (~170 ATP/s/monomer) is consistent with a stepwise translocation mechanism rather than a Type III-like conformational switch/sliding mechanism (28). The stoichiometry of the CgII enzyme is  $R_2H_2$  and each complex has one complete nuclease active site sufficient to cut one DNA strand, and two helicase-like motor domains and DNA binding domains (Figure 1B and C) (28). Therefore both uni- and bidirectional communication are possible. It is also likely that two  $R_2H_2$  complexes will be required to produce a single dsDNA break since dimerization of the R-subunits will produce only a single active site, akin to BfiI (29). This would also predict a requirement for inter-target communication. Each  $R_2H_2$  complex also contains two B3-like DNA binding domains, which also offers the possibility of communication by looping between pairs of targets.

In this paper, we report the investigation of the CgII catalytic mechanism of DNA cleavage. We demonstrate that CgII efficiently hydrolyzes one-site circular DNA and does not strictly require the presence of a second DNA target *in cis* on the same DNA. However, two-site linear DNA substrates can be cleaved more efficiently than one-site linear DNA substrates. Moreover, on HtT linear DNA, cleavage predominantly occurs at one of the two targets. These cleavage data are consistent with a communication scheme that follows the 1D DNA contour with some directional bias. Using a triplex displacement assay, we confirm that ATP-dependent translocation from a target has partial directionality, with a preference for communication downstream from the asymmetric recognition sequence (5'-GCCGC-3'). We also demonstrate that CgII has the unexpected ability to displace a triplex *in trans*, in a reaction that does not require entry/exit from DNA ends. We demonstrate that *in trans* triplex displacement occurs via simultaneous binding of both DNA molecules (specific and non-specific). On the basis of the DNA cleavage and triplex displacement experiments, we propose that CgII translocates bidirectionally along the 1D DNA contour from its target site and cleaves dsDNA when two  $R_2H_2$ -CgII complexes collide with each other (one translocating and one bound to a target site). In some cases the translocating species can be delivered *in trans* having initiated on a separate DNA molecule. This reaction scheme is quite distinct to that observed for Type I, ISP or III RM enzymes, and adds further to our understanding of

the remarkable diversity of mechanisms that have evolved to combat bacteriophage infection.

## MATERIALS AND METHODS

### Expression and purification of CgII proteins

The wt CgII and its active site mutants were expressed and purified as described earlier (28).

### Oligonucleotides and DNA

Named oligonucleotide sequences are given in Supplementary Table S1. Plasmid substrates and templates are summarized in Supplementary Table S2.

**Substrates for DNA cleavage experiments.** One-site open-circle (OC) and one-site full-length linear (FLL) substrates were obtained from one-site supercoiled circular (SC) p1 plasmid DNA by cleavage of a single DNA strand with a Cas9(H840A)-crRNA complex and linearization with restriction endonuclease NdeI, respectively. The HtN fragment used in DNA cleavage stimulation experiments was obtained from one-site SC p1 DNA by linearization with restriction endonuclease XhoI and labeling of 5' overhangs with [ $\gamma$ <sup>33</sup>P]ATP (Perkin Elmer) and T4 polynucleotide kinase (Thermo Fisher Scientific, Vilnius, Lithuania). The HtT1 and HtT2 fragments were obtained from two-site SC p2.HtT DNA by linearization with restriction endonucleases XhoI and NdeI, respectively. HtH and HtT fragments were obtained from two-site SC p2.TtT DNA by linearization with restriction endonucleases XhoI and NdeI, respectively. A total of 85-bp fragment S85 containing a single target of CgII was synthesized by polymerase chain reaction (PCR) using MZ-318 and AS-41 oligonucleotides as primers and pECFP-ICAD-S(NLS) plasmid as a template (Supplementary Tables S1 and 2). All linear and OC fragments were purified using the GeneJET™ PCR Purification Kit (Thermo Fisher Scientific, Vilnius, Lithuania).

**Substrates for triplex displacement assay.** Three different 278 bp substrates N, H and T were synthesized by PCR using MZ-503 and MZ-504 oligonucleotides as primers and pRMA03\_0, pRMA03\_1 and pRMA03\_1\_IN plasmids as templates (14), respectively (Supplementary Tables S1 and 2). 5'-end biotinylated 278 bp substrate N was obtained from the previously synthesized substrate N by PCR using MZ-948 and MZ-504 oligonucleotides (Supplementary Table S1). A total of 3383 bp CN plasmid substrate was constructed by ligation of the 278 bp N fragment into p0s vector pre-digested with restriction endonuclease EheI. A total of 3383 bp LN substrate was generated from CN plasmid by linearization with restriction endonuclease CaiI. L0 and L1 were obtained from p0s and p1s plasmid linearization with restriction endonuclease NdeI, respectively. All PCR generated fragments were purified using the GeneJET™ PCR Purification Kit (Thermo Fisher Scientific, Vilnius, Lithuania). The 5'-end of triplex forming oligonucleotide (TFO) was labeled with [ $\gamma$ <sup>33</sup>P]ATP (Perkin Elmer) and T4 polynucleotide kinase (Thermo Fisher Scientific, Vilnius, Lithuania).

### DNA cleavage assay

DNA (10 nM) cleavage reactions were performed at 37°C in Reaction Buffer 1 (33 mM Tris-acetate (pH 7.0 at 25°C); 20 mM Mg-acetate; 0.1 mg/ml bovine serum albumin (BSA); 1 mM dithiothreitol (DTT); 4 mM ATP; 20 mM phosphocreatine and 5 U/ml creatine kinase). Reactions were initiated by addition of 1000 nM H.CgII and 1000 nM R.CgII (in terms of monomer) and stopped by addition of 2× Loading solution (75 mM ethylenediaminetetraacetic acid (EDTA), 0.6% (w/v) orangeG, 50% (v/v) glycerol, 0.5% (w/v) sodium dodecyl sulfate (SDS), pH 8.0 at 25°C) and incubation at 70°C for 15 min. The reaction products were separated by electrophoresis through 0.8 or 1.0% (w/v) agarose gels with 0.5 µg/ml ethidium bromide and the relative amount of dsDNA cleavage quantified by gel densitometry using BioDocAnalyze image analyzer (Biometra, Göttingen, Germany) and the images analyzed with OptiQuant software (Packard Instrument). To determine whether the ring size of one-site circular DNA has an effect on cutting efficiency of CgII one-site plasmid DNAs with various ring sizes (p1\_2064, p1\_3105, p1\_4320, p1\_4498, p1\_5722, Supplementary Table S2) were incubated with 1000 nM H.CgII and 1000 nM R.CgII (in terms of monomer) at 25°C in Reaction Buffer 1 (indicated above). Reactions were stopped and products analyzed as described above.

To determine the DNA cleavage position of CgII pHtT or pTtT, plasmid DNA (10 nM) was incubated with 1000 nM H.CgII and 1000 nM R.CgII (in terms of monomer) for 1 h at 25°C in Reaction Buffer 1 (indicated above). Reactions were stopped as mentioned above. Linearized plasmids were purified following agarose gel electrophoresis using GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Vilnius, Lithuania). The linear purified DNAs were sequenced using MZ-438 and MZ-437 oligonucleotides as primers (Supplementary Table S1).

DNA cleavage stimulation reactions contained 1 nM of the HtN fragment and were performed at 20°C in Reaction Buffer 2 (33 mM Tris-acetate (pH 7.0 at 25°C); 10 mM Mg-acetate; 0.1 mg/ml BSA; 1 mM DTT; 4 mM ATP; 150 mM KAc; 20 mM phosphocreatine and 5 U/ml creatine kinase) and in the presence of 10, 50, 100 or 250 nM of the S85 fragment. Reactions were initiated by addition of 1000 nM H.CgII and 1000 nM R.CgII (in terms of monomer) and stopped as mentioned above. The products were separated by electrophoresis through 1.0% (w/v) agarose gels in 40 mM Tris, 20 mM acetic acid and 1 mM EDTA (pH 8.3 at 25°C). Gels were air-dried and scanned in Fujifilm FLA-5100 fluorescent image analyzer (Fujifilm, Tokyo, Japan) and the image analyzed with OptiQuant software (Packard Instrument) to determine the volume of each band, taking into account background readings. The fraction of DNA substrate left in each sample was calculated as  $\text{volume}_{(\text{DNA substrate left})} / (\text{volume}_{(\text{DNA substrate left})} + \text{volume}_{(\text{DNA product})})$ .

### Triplex formation and analysis

A total of 50 nM linear DNA and 25 nM 5'-end <sup>33</sup>P-labeled TFO (for *in trans*: 10 nM linear or plasmid DNA and 5 nM <sup>33</sup>P-labeled TFO) were mixed in buffer MM

(10 mM 2-(N-morpholino)ethanesulfonic acid (MES) (pH 5.5 at 25°C), 12.5 mM MgCl<sub>2</sub>) and incubated at 20°C overnight. The triplexes were stored on ice and then diluted 1/10 into Reaction Buffer 1 before use. To analyze the proportion of bound and free TFO, reactions were quenched using 1:1 GSMB buffer (15% (w/v) glucose, 3% (w/v) SDS, 250 mM 3-(N-morpholino)propanesulfonic acid (MOPS) (pH 5.5 at 25°C), 0.4 mg/ml bromophenol blue) and analyzed in 6 or 8% (w/v) polyacrylamide gels (29:1 acrylamide/bisacrylamide in 40 mM Tris-acetate (pH 5.5 at 25°C), 5 mM Na-acetate, 1 mM MgCl<sub>2</sub>) at ~10 V/cm for 2 h at 4°C or 1% (w/v) agarose gels (40 mM Tris-acetate (pH 5.5 at 25°C), 5 mM Na-acetate, 1 mM MgCl<sub>2</sub>) at ~8 V/cm for 1 h at 4°C. Gels were air-dried and scanned in Fujifilm FLA-5100 fluorescent image analyzer (Fujifilm, Tokyo, Japan) and the image analyzed with OptiQuant software (Packard Instrument) to determine the volume of each band, taking into account background readings. The fraction of triplex in each sample was calculated as  $\text{volume}_{\text{triplex}} / (\text{volume}_{\text{triplex}} + \text{volume}_{\text{free TFO}})$  (10).

### CgII translocation reactions

Triplex displacement reactions were performed at 20°C mixing preformed triplex (see above) in 33 mM Tris-acetate (pH 7.0 at 25°C), 0.1 mg/ml BSA (or 1 mg/ml BSA for reactions with magnetic beads), 1 mM DTT, 10 mM Mg-acetate, 150 mM K-acetate, 4 mM ATP or adenylyl-imidodiphosphate (AMP-PNP), 25 mM phosphocreatine and 5 U/ml creatine kinase. Triplex displacement *in trans* reactions were additionally supplemented with 4, 10 or 50 nM of various dsDNAs. Reactions were initiated by addition of 200 nM H.CgII and 200 nM R.CgII to the reaction mixture. Aliquots were removed at fixed time intervals and the reactions stopped by adding GSMB buffer. Resulting samples were analyzed as above. TFO displacement rates are quoted as the mean of three independent experiments.

### Kinetic data analysis

Quantitative approximation of the rates of TFO displacement and DNA cleavage estimated by fitting the data to first or second degree exponential equations using KyPlot version 2.0 (30). Rate constants ( $k_1$  and  $k_2$ ) for triplex displacement were calculated using following equations:

$$S = 100 \cdot \exp(-k_1 \cdot t) \quad (1)$$

$$S = [A_1] + [100 - A_1] \cdot \exp(-k_1 \cdot t) \quad (2)$$

$$S = [A_1] \cdot \exp(-k_1 \cdot t) + [100 - A_1] \cdot \exp(-k_2 \cdot t) \quad (3)$$

Where S is the percentage of substrate left,  $A_1$  and  $100 - A_1$  are the amplitudes of the first and second phases, respectively,  $k_1$  and  $k_2$  are the rates of the first and second phases, respectively,  $t$  is the reaction time in seconds.

Rate constants ( $k_1$ ) for DNA cleavage were calculated using following equations:

$$S = [A_1] \cdot \exp(-k_1 \cdot t) \quad (4)$$

$$S = [A_1] \cdot \exp(-k_1 \cdot t) + [A_2] \quad (5)$$

$$P = [A_1] \cdot (1 - \exp(-k_1 \cdot t)) \quad (6)$$

Where S is the amount of substrate left (nM), P is the amount of product created (nM),  $A_1$  is the amplitude of the first phase,  $A_2$  is the free amplitude,  $k_1$  is the rate of the first phase,  $t$  is the reaction time in seconds. In all cases the kinetic profiles are actually more complex, with cleavage being preceded by multiple steps. Therefore the exponential fits are to be taken as estimates of the relative rates for comparison purposes, rather than measurements of rate constants that correspond to particular mechanistic steps.

## RESULTS AND DISCUSSION

The stress-sensitive RM system CgII (from *Corynebacterium glutamicum*) was originally assigned as a Type II RM system but has functional and structural similarities to both Type I and III RM systems (23). Here we investigated the mechanisms of DNA hydrolysis and long-range communication of the CgII restriction enzyme. Throughout this work we utilized reaction conditions where the concentration of CgII is significantly in excess of DNA since these are the condition necessary to observe DNA cleavage (28). Super-saturating protein concentration may be required due to a low specific activity but can also be due to the mechanism of the cleavage pathway; for example both Type I and III RM enzymes have been shown to only work at or above ‘single turnover’ conditions as a consequence of long-range communication mechanisms (31–33). Therefore understanding the cleavage mechanism can in part help explain the necessity for certain reaction conditions; the protein–DNA stoichiometry of CgII will be investigated in detail in a further study.

### CgII uses 1D communication to cut DNA

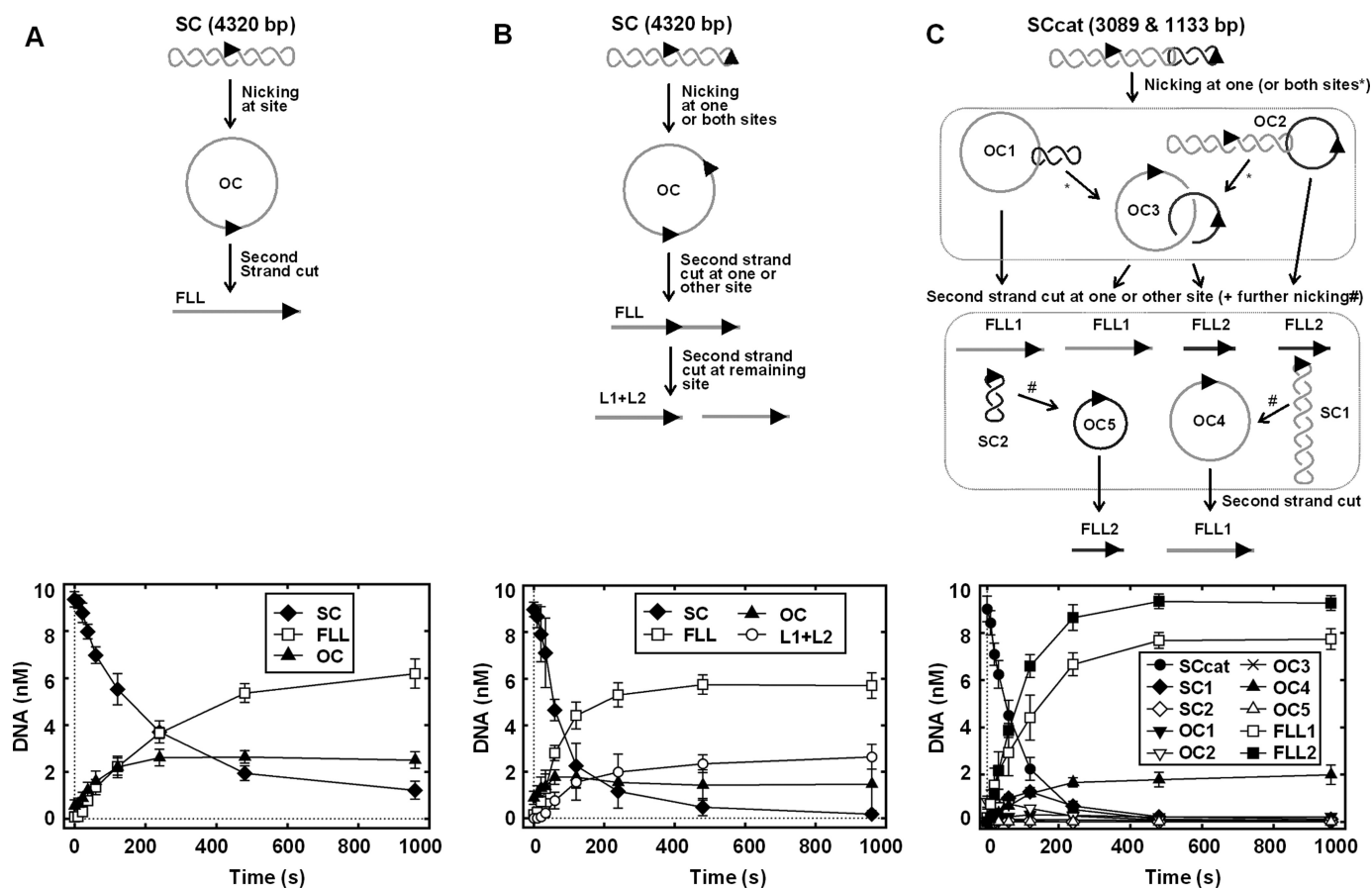
A basic test of the requirement for single or multiple sites for DNA cleavage that has been widely applied to Type II RM enzymes as well as Type I and III enzymes is to follow endonuclease activity on one- and two-site plasmids, and on DNA catenanes (34–36). Simple RM enzymes which only bind a single target site will cleave a two-site plasmid or catenane only 2-fold faster than one-site plasmid, as the increase in site number does not affect the ability to cleave but merely the chance of cleavage occurring at a site at any given point in time (e.g. (37)). In contrast, enzymes that require simultaneous binding of two targets for an efficient DNA hydrolysis will cleave two-site substrates up to 100-fold faster than one-site substrates (e.g. (37)). The two-site catenanes allow two interacting DNA sites to be on separate rings that are topologically-linked, thus holding the sites in proximity to each other. Long-range communication between the pair of sites on the catenane substrate can only occur by three-dimensional (3D) DNA looping rather than via the 1D DNA contour (e.g. by translocation or DNA sliding). If the reaction occurs by a 3D process, such as DNA looping, the reactions with catenane (both sites are held *in trans*) and two site plasmid (both sites are held *in cis*) will occur with similar efficiency (e.g. (38)). Alternatively, if the communication occurs only by 1D process, the reaction of the catenane will occur with the same efficiency as on one site plasmid (e.g. (36)).

First, control DNA cleavage experiments were performed using DNA without a target site (zero-site plasmid) (Supplementary Figure S1); only background levels of DNA cleavage were observed, demonstrating the requirement for a specific CgII target site. Next, one-site plasmid DNA, a two-site HtT (HtT orientation of the asymmetric recognition sequence 5'-GCCGC-3') plasmid DNA and a catenane (generated from the HtT plasmid DNA using the Tn21 resolvase), were used as substrates (Figure 2) (7). The results show that one-site circular DNA is cut only 2-fold slower than the two-site circular DNA, indicating that only one DNA site is necessary for efficient DNA cleavage: the rate of disappearance of the one-site substrate, which represents the first strand break, was only 2-fold slower than the equivalent rate on either of the two-site substrates (Figure 3 and Supplementary Table S3). Note that in all cases the kinetic profiles are actually complex, with cleavage being preceded by multiple steps. Therefore the exponential fits are to be taken as estimates of the relative rates for comparison purposes, rather than measurements of rate constants that correspond to particular mechanistic steps. Additionally we note that the activity of CgII reduces with incubation at 37°C, so rates may slow with time and reactions may not always go to completion.

A simple explanation of the 2-fold difference in the rates (Figure 3 and Supplementary Table S3) is that the initial strand break is twice as likely in the presence of two recognition sites. Disappearance of the two-site catenane substrate band on the gel can be due to cleavage of either the large or small ring. Similarly, the two-site plasmid substrate band can disappear due to cleavage at one or other of the sites. In comparison, disappearance of the one-site plasmid substrate can only occur due to cleavage at one site. Therefore a non-communicating system might show a 2-fold difference between one-site and substrates, regardless of topology. This does not rule out long-range communication, but for CgII a second site *in cis* (plasmid) or *in trans* (catenane) does not necessarily have a large activating effect on DNA cleavage of the circular DNA substrates, where one site can communicate through 1D space with another molecule bound at the same site.

We note that cleavage of the two rings of the two-site catenane to FLL products is more efficient than the two-site plasmid (Figure 2B and C). While the efficiency appears to scale with size of the DNA rings, we found that circular DNA size does not have a clear effect on cleavage rate (Supplementary Figure S2). An alternative explanation is that there is some level of *in trans* activation (e.g. see below) that is possible on the catenane due to the topological linking of the rings.

If only a single  $R_2H_2$  complex were involved in DNA cleavage, then we might expect a nicked DNA intermediate to accumulate since cleavage of the second strand would require rearrangement of the complex. However, CgII leaves max. 25–30% OC and FLL starts to accumulate immediately after initiation of the reaction. While BfiI using a single active site for a double stranded break generation produces ~80% OC and FLL starts to accumulate later (29). We propose that this is consistent with two  $R_2H_2$  complexes (each containing only a single nuclease active site) being required during DNA cleavage and preceded by a rate-limiting step



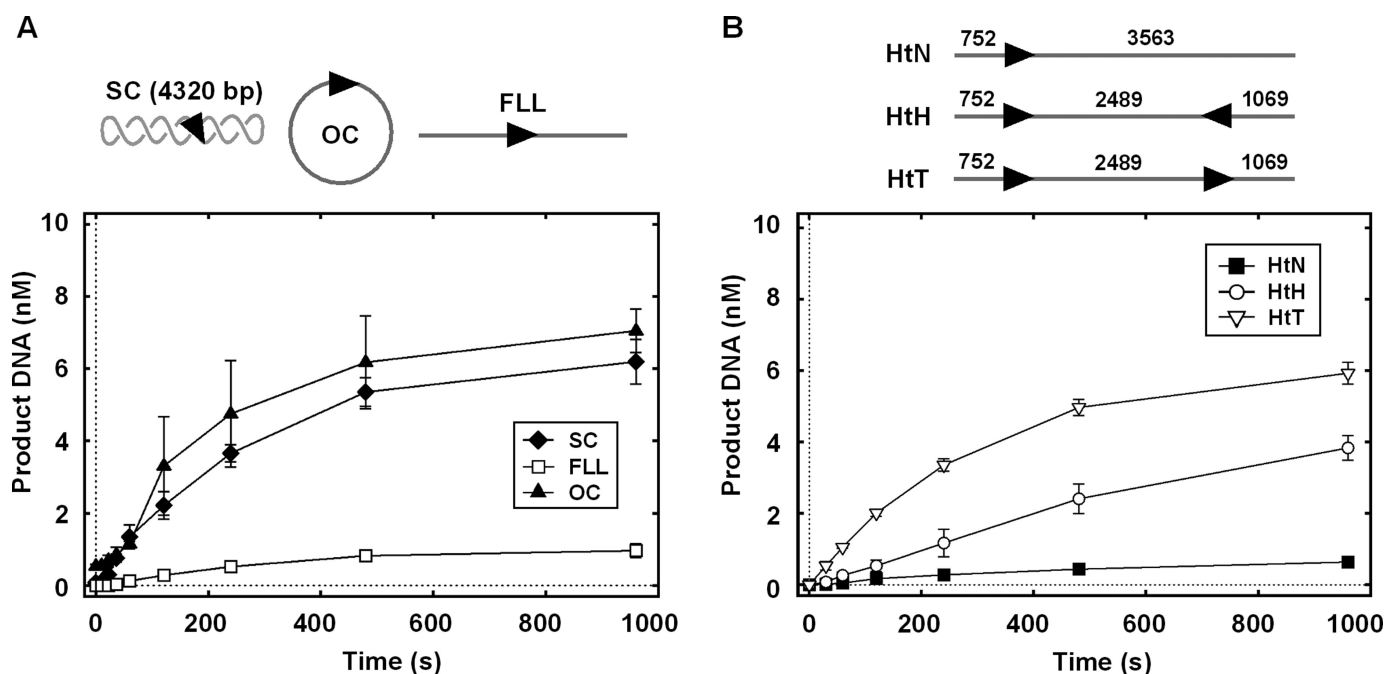
**Figure 2.** One-site plasmid (A), two-site plasmid (B) and two-site catenane (C) cleavage by CgII. The two-site plasmid substrate was used to form catenane in the presence of Tn21 resolvase (7). The CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (▶). Possible DNA species formed when CgII is incubated with the plasmid or catenane substrates are presented above the graphs: SC—supercoiled circular DNA, OC—open circle DNA ('nicked'), FLL—full-length linear DNA, L1+L2—linear DNA cut at both CgII sites, SCcat—SC catenane DNA, OC1—catenane nicked in the large ring only, OC2—catenane nicked in the small ring only, OC3—catenane nicked in both rings, OC4—nicked large plasmid, OC5—nicked small ring, FLL1—full-length linear DNA of the large ring, FLL2—full-length linear DNA of the small ring. Note that final FLL products from SCcat can be generated by different routes, requiring either 2, 3 or 4 consecutive independent cleavage events on a single DNA substrate. Reactions contained 10 nM DNA, 4 mM ATP, 500 nM R<sub>2</sub>H<sub>2</sub>-CgII and were conducted as described in 'Materials and Methods' section. The rate constant values ( $5.3 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$  for (A),  $1.1 \pm 0.1 \times 10^{-2} \text{ s}^{-1}$  for (B),  $1.2 \pm 0.1 \times 10^{-2} \text{ s}^{-1}$  for (C)) were obtained by fitting a single exponential to the time courses of supercoiled form depletion. Points are averages with error bars as standard deviation for at least three repeat reactions.

or series of steps, which in this case, we suggest, is a communication process (see below). The product accumulation pattern and apparent cleavage rate constants were same in the case of two-site plasmid substrates with different target orientations (HtT and TtT) (Supplementary Table S3 and Figure S3). Run-off sequencing analysis of the cleavage products showed that in all cases a double strand break is introduced at the same position near the asymmetric recognition sequence (GCCGCN<sub>6</sub>/N<sub>6-7</sub>) (Supplementary Figure S4). This indicates that the same CgII collision complex must be produced regardless of the number of targets or the relative orientation of pairs of targets.

The DNA cleavage results are not consistent with an enzyme that uses a 3D route to communicate between sites (i.e. tetrameric Type II restriction endonucleases SfiI, Cfr10I, NaeI and NgoMIV require interaction with two recognition sites before displaying their optimal DNA cleavage activities and that their interactions with two DNA sites occur through 3D space) (35,38). Although a simple interpre-

tation of the results is that multiple sites are not required for cleavage, long-range communication via 1D process cannot be ruled out. For example, the ATP-dependent Type I RM enzymes are also able to efficiently cleave one-site plasmid, but must do so by extensive 1D translocation around the DNA circle. On a linearized version of the same DNA, cleavage does not occur as the translocating Type I motors cannot collide. Therefore to test this possibility for CgII, we needed to explore the effect of DNA topology on cleavage activity.

We tested the relative DNA cleavage activity on one-site DNA substrates that were: negatively supercoiled circles, topologically unconstrained relaxed circles or topologically unconstrained full-length linear DNA (Figure 3A). The rate of appearance of dsDNA cleaved product of one site circular DNA was the same as on relaxed (nicked) or supercoiled DNA (Figure 3A). However, the one-site linear DNA only supported a relatively small amount of DNA cleavage (Figure 3A). This result is similar to what has been observed



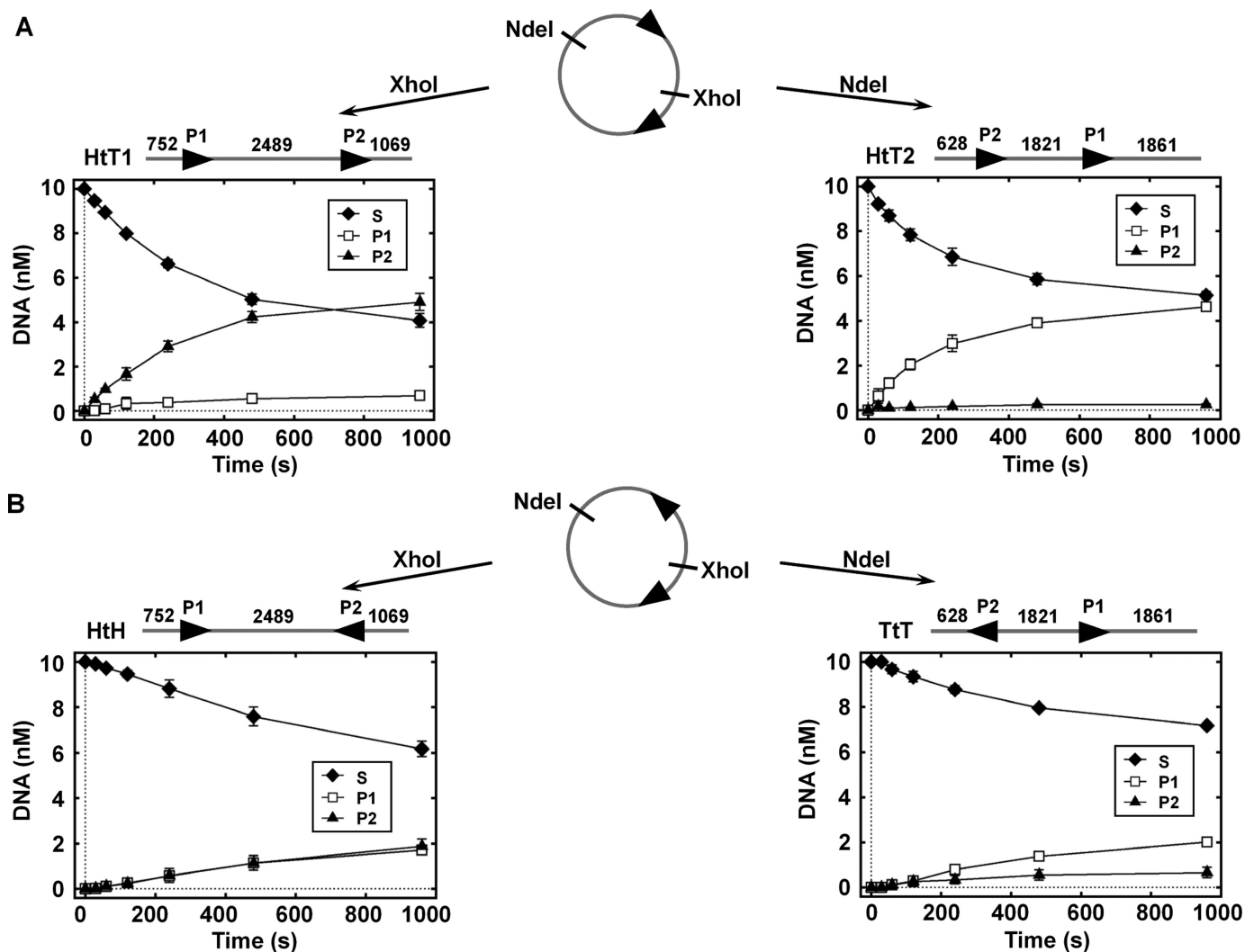
**Figure 3.** (A) Cleavage of the SC, OC and linear one-site plasmid DNA by CgII. Accumulation of the final DNA cleavage products with both DNA strands cleaved at the CgII recognition sequence is shown. The one-site SC DNA was used to form the OC DNA (by cleavage of a single DNA strand with the Cas9–crRNA complex) or the full-length linear DNA (FLL) (both DNA strands were cleaved with the restriction endonuclease NdeI). The CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►). The rate constant values ( $3.6 \pm 0.2 \times 10^{-3} \text{ s}^{-1}$  for SC,  $3.0 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$  for FLL and  $4.2 \pm 0.7 \times 10^{-3} \text{ s}^{-1}$  for OC) were obtained by fitting a single exponential to the time courses of substrates. (B) Cleavage of the linear one- and two-site DNA by CgII. Accumulation of the final DNA cleavage products with both DNA strands cleaved at one CgII recognition sequence is shown. Numbers indicate distances in bp between the CgII targets and DNA ends. The rate constant values ( $2.0 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$  for HtN,  $8.9 \pm 1.9 \times 10^{-4} \text{ s}^{-1}$  for HtH and  $3.3 \pm 0.1 \times 10^{-3} \text{ s}^{-1}$  for HtT) were obtained by fitting a single exponential to the time courses of substrates. All reactions contained 10 nM DNA, 4 mM ATP, 500 nM  $\text{R}_2\text{H}_2\text{-CgII}$  and were conducted as described in 'Materials and Methods' section. Points are averages with error bars as standard deviation for at least three repeat reactions.

with Type I (and Type ISP nicking) reactions (15,39) and supports a communication mechanism along the 1D contour. Therefore we propose that the cleavage of the one-site circular DNA is activated by an enzyme moving around the whole contour of the DNA from the site and two sites are not necessary. However, on linear DNA the same communication will be largely inefficient due to the enzymes dissociating from the ends of the DNA. We propose that the small amount of cleavage of one-site linear DNA results from *in trans* interactions between CgII complexes each bound (or originating from) separate DNA molecules. Only a small fraction of DNA is captured in such complexes, therefore, small amplitude of DNA cleavage is observed. The ability of CgII to occasionally communicate *in trans* is considered below in the translocation assays.

Although cleavage of one-site linear DNA is not efficient, in the presence of multiple targets on linear DNA, cleavage may be activated by 1D interactions between enzymes initially bound at the separate sites. Moreover, a site orientation preference may be revealed using such substrates, as seen for the Type ISP and III RM enzymes (16,17,20). To test this we used four different two-site linear DNA substrates: HtT1 and HtT2—generated from two-site HtT plasmid DNA digestion with XhoI and NdeI, respectively; and, HtH and TtT: generated from two-site TtT plasmid DNA digestion with XhoI and NdeI, respectively (Figure 4).

The cleavage data show that linear DNA with a pair of sites in direct repeat (HtT1 or HtT2) is the most efficiently cut, where cleavage dominates at the second site (downstream to the first as drawn in Figure 4), with a rate that is similar to that of the one-site circular DNA and which is independent of local sequence ( $\sim 10^{-3} \text{ s}^{-1}$ , Supplementary Table S3). This would be consistent with a rear-end collision mechanism for activation of cleavage that is different from the Type I, ISP and III RM enzymes (5,16,20,22,40,41). However, we also observed cleavage of both DNA with sites in inverted repeat (HtH and TtT) with a similar rate ( $\sim 10^{-4} \text{ s}^{-1}$ ) that is 10-fold slower than the direct repeats (Figure 4B and Supplementary Table S3). Therefore, although this is not a strict site-orientation preference as it is seen with Type ISP and Type III RM enzymes, there is still  $\sim 10$ -fold kinetic preference for the sites in direct repeat, which also differs from the Type I RM enzymes. This points to a unique mechanism for CgII. The preference is not observed on the circular DNA as we suggest that single-site communication events occur with a faster rate on these substrates.

One explanation for an ability to cut pairs of the sites without site-orientation preference is that the communication is bidirectional, as shown for the Type I RM enzymes (10–12). A prediction of this would be that on two-site linear DNA there is a possibility to cut both sites. However, for HtT1 and HtT2, there was  $\sim 10$ -fold preference for cleavage at one of the two sites (Figure 4A), while for TtT, there was



**Figure 4.** Cleavage of the two-site linear DNAs with different target orientations by CgII. Cleavage of the HtT substrates is presented in (A), the HtH and TtT substrates—in (B). The two-site supercoiled circular plasmids were used to form the linear substrates by cleavage with the restriction endonucleases XhoI and NdeI. The CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►), 'S' represents the substrates and any nicked intermediates, while P1 and P2 represent a double strand break at site P1 or P2, respectively. Numbers indicate distances in base pair between the CgII targets and DNA ends. All reactions contained 10 nM DNA, 4 mM ATP, 500 nM R<sub>2</sub>H<sub>2</sub>-CgII and were conducted as described in 'Materials and Methods' section. The rate constant values ( $3.3 \pm 0.1 \times 10^{-3} \text{ s}^{-1}$  for HtT1,  $4.4 \pm 0.4 \times 10^{-3} \text{ s}^{-1}$  for HtT2,  $8.9 \pm 1.9 \times 10^{-4} \text{ s}^{-1}$  for HtH and  $2.0 \pm 0.2 \times 10^{-3}$  for TtT) were obtained by fitting a single exponential to the time courses of substrates. Points are averages with error bars as standard deviation for at least three repeat reactions.

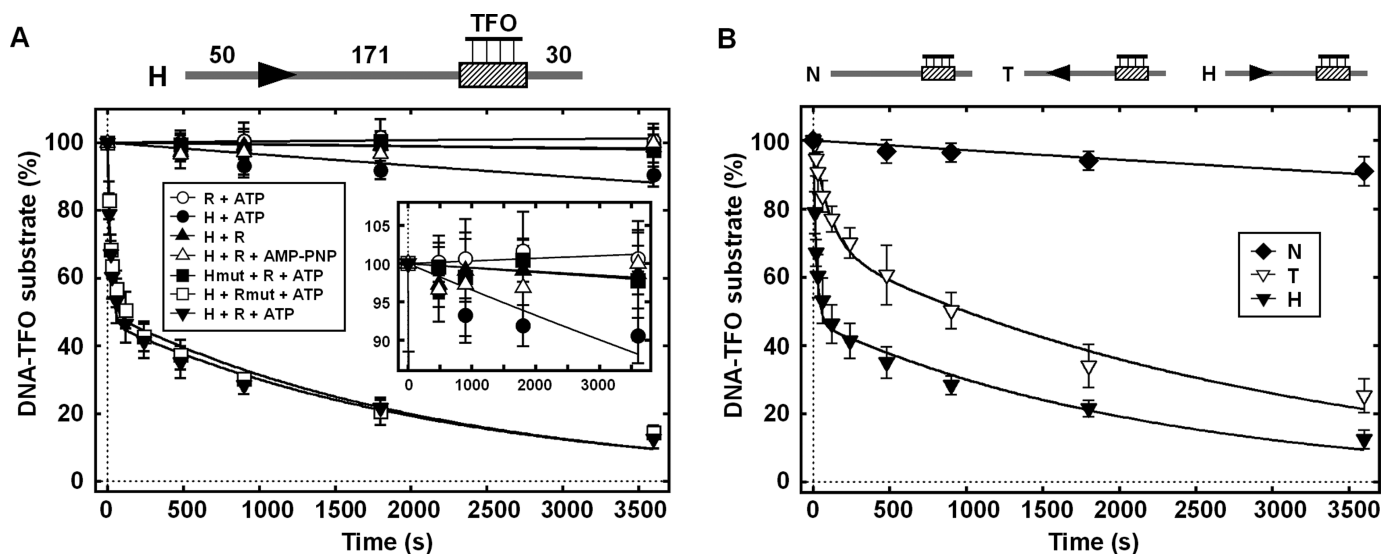
~3-fold preference for one site of the two sites (Figure 4B). In contrast, for HtH, there was no apparent preference between the sites (Figure 4B). This data appear to support alternative DNA cleavage pathways depending on the DNA substrate. The most efficient cleavage occurs where the enzymes can communicate directionally downstream from the 5'-GCCGC-3' site. This would produce a rear-end collision with an enzyme bound at the downstream site, where cleavage is activated (Figure 4A). Alternatively, less efficient cleavage is produced when the sites are in other orientations (Figure 4B). This alternative mechanism needs to accommodate the fact that both head-to-head and tail-to-tail arrangements can produce cleavage. This would predict some sort of bidirectional translocation or an alternative communication mechanism that is independent of the site orienta-

tion preference (such as reversal of directionality of communication).

#### Evidence for bidirectional translocation of CgII on DNA

Previously it was shown that H.CgII carries SF2 helicase motifs (28). These motifs are also found in Type I and III RM enzymes and are required for ATP hydrolysis-driven DNA translocation or activation of DNA sliding (19,28). The rapid DNA cleavage seen above with one site plasmids or with linear DNA with two sites in direct repeat supports a long-range communication mechanism that follows the 1D DNA contour, principally, but not exclusively, downstream of the 5'-GCCGC-3' site. Given the relatively high level of ATP hydrolysis by CgII (28), we would suggest that this communication is due to stepwise DNA translocation rather than sliding. To test this, we used the triplex dis-





**Figure 5.** CgII translocase activity analyzed using the DNA triplex displacement assay. (A) CgII displaces TFO. Scheme of the DNA substrate used in the triplex assay is shown above the graph: DNA is shown as a thick line, the CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►), the triplex binding sequence is shown as a varied rectangle, numbers indicate distances in base pair. Used abbreviations: H—H.CgII, R—R.CgII, Hmut—the H.CgII (D158A + E159A) mutant, Rmut—the R.CgII (H105A) mutant. The *inset* shows zoomed part of the graph to clarify the differences of TFO displacement. (B) TFO displacement by CgII is dependent on the recognition sequence. DNA substrates: N—DNA without the recognition sequence, the T and H DNA fragments contain the recognition sequence in both orientations. All reactions contained 5 nM DNA, 2.5 nM TFO, 4 mM ATP, 200 nM R.CgII, 200 nM H.CgII and were conducted as described in 'Materials and Methods' section. Solid lines are single or double exponential fits to the data. Rate constants for TFO displacement are presented in Supplementary Table S4. Points are averages with error bars as standard deviation for at least three repeat reactions.

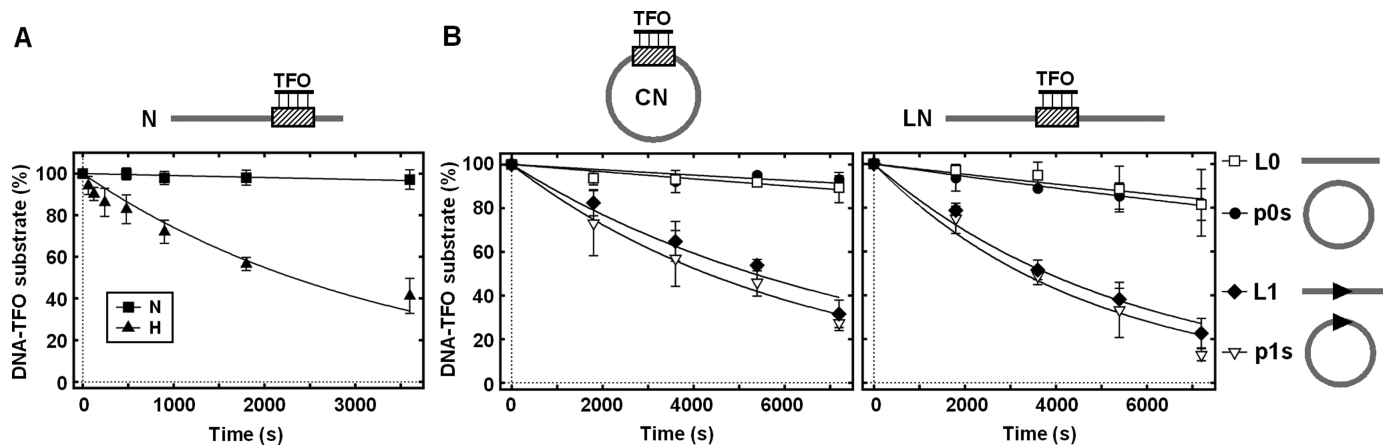
placement assay, which was successfully applied for Type I and ISP RM enzymes (42,43), and other dsDNA translocating motors (44–46). When located downstream of the 5'-GCCGC-3' site, triplex displacement was observed dependent upon ATP hydrolysis and a WT helicase domain (Figure 5A). Mutation of the nuclease motif (R<sub>mut</sub>) had no effect. Triplex displacement is consistent with a translocating motor. There were two distinct phases to the triplex displacement profile; a first phase of ~50% amplitude with a rate  $\sim 10^{-2} \text{ s}^{-1}$  and a second phase of ~50% amplitude with a rate  $\sim 10^{-4} \text{ s}^{-1}$  (Figure 5A and Supplementary Table S4).

To confirm the directionality of communication, we tested non-specific DNA without a CgII site or specific DNA where the triplex was located upstream of the target (Figure 5B). On the non-specific DNA we observed a markedly slower rate of displacement, which most likely reflects non-specific displacement due to protein-dependent and -independent effects that are not related to translocation *per se* (Figure 5A). When the triplex was formed upstream the recognition site, we also observed a biphasic displacement profile, with a first phase of ~30% amplitude and a rate  $\sim 10^{-3} \text{ s}^{-1}$ , and with a second phase of ~70% amplitude and a rate  $\sim 10^{-4} \text{ s}^{-1}$  (Figure 5B and Supplementary Table S4). As observed with the cleavage data, the triplex displacement data appear to support two different communication mechanisms with differing efficiencies. Rapid triplex displacement is due to translocation *downstream* of the site while slower triplex displacement is due to translocation *upstream* of the site. An even slower second phase of displacement was seen on both DNA with a similar rate. We note that we did not observe a clear lag phase characteristic of a stepping motor (47). However, similar profiles have

been observed with other *bona fide* translocases such as Mfd (48), where the absence of a clear lag may be due to events prior to translocation (DNA binding, initiation) being rate-limiting. The biphasic nature of the displacement profiles may reflect dissociation of a subset of translocating enzymes that then reinitiate (9). We also tested translocation using a protein (*Streptococcus thermophilus* CRISPR Cas9 or Cascade) displacement assay (Supplementary Figure S5). This data also support communication both up- and downstream but with a preference for downstream translocation.

#### Formation of protein–DNA sandwich complexes allows translocation activity *in trans*

One explanation for the above data is that CgII can translocate bidirectionally but that the efficiency and translocation rate is different in the downstream and upstream directions. An alternative hypothesis is that CgII can only translocate unidirectionally, but that the motor can switch directions with a slower rate. This directional switch may be due to the enzyme leaving the DNA completely, remaining activated in free solution and then rebinding a new DNA site, either on the same DNA molecule or on a different DNA molecule. This rebinding would be randomized so that subsequent translocation would no longer be dictated by the site orientation. Alternatively, since the R<sub>2</sub>H<sub>2</sub> complex has two DNA binding and two motor domains, there could be a simultaneous binding of a second DNA location without dissociation from the first. This could result in both DNA loops (binding of two sites on the same DNA) and formation of DNA sandwiches (binding of two sites, each on dif-

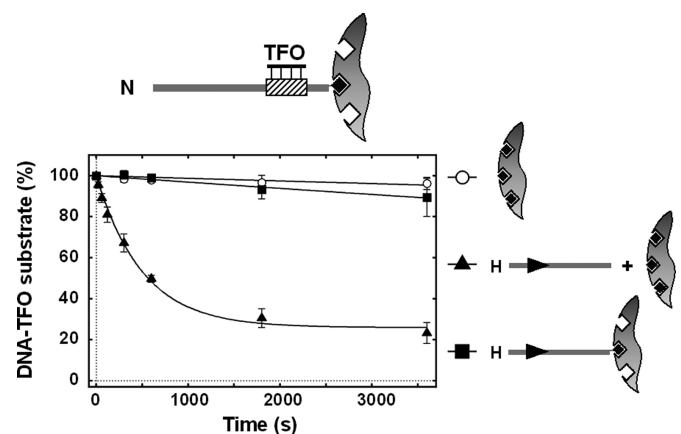


**Figure 6.** TFO displacement by CgII *in trans*. (A) TFO displacement from the linear DNA (N) without the recognition sequence of CgII. DNA substrate (278 bp in length) is shown above the graph. The triplex binding sequence is shown as a varied rectangle. DNA added *in trans* (278 bp in length) are the same as used in Figure 5B. (B) TFO displacement from the circular versus linear DNA. DNA substrates (3383 bp in length) are shown above graph. The CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►). DNA added *in trans* (3105 bp in length) were linear (L0 and L1) and circular (p0s and p1s) without and with recognition sequence of CgII (are shown on the right). All reactions contained 1 nM DNA substrate, 0.5 nM TFO, 4 mM ATP, 200 nM R.CgII (H105A), 200 nM H.CgII, 4 nM (A) or 10 nM (B) DNA added *in trans* and were conducted as described in 'Materials and Methods' section. Solid lines are single exponential fits to the data. Rate constants are presented in Supplementary Table S4. Points are averages with error bars as standard deviation for at least three repeat reactions.

ferent DNA). Both types of event would randomize the direction of subsequent translocation.

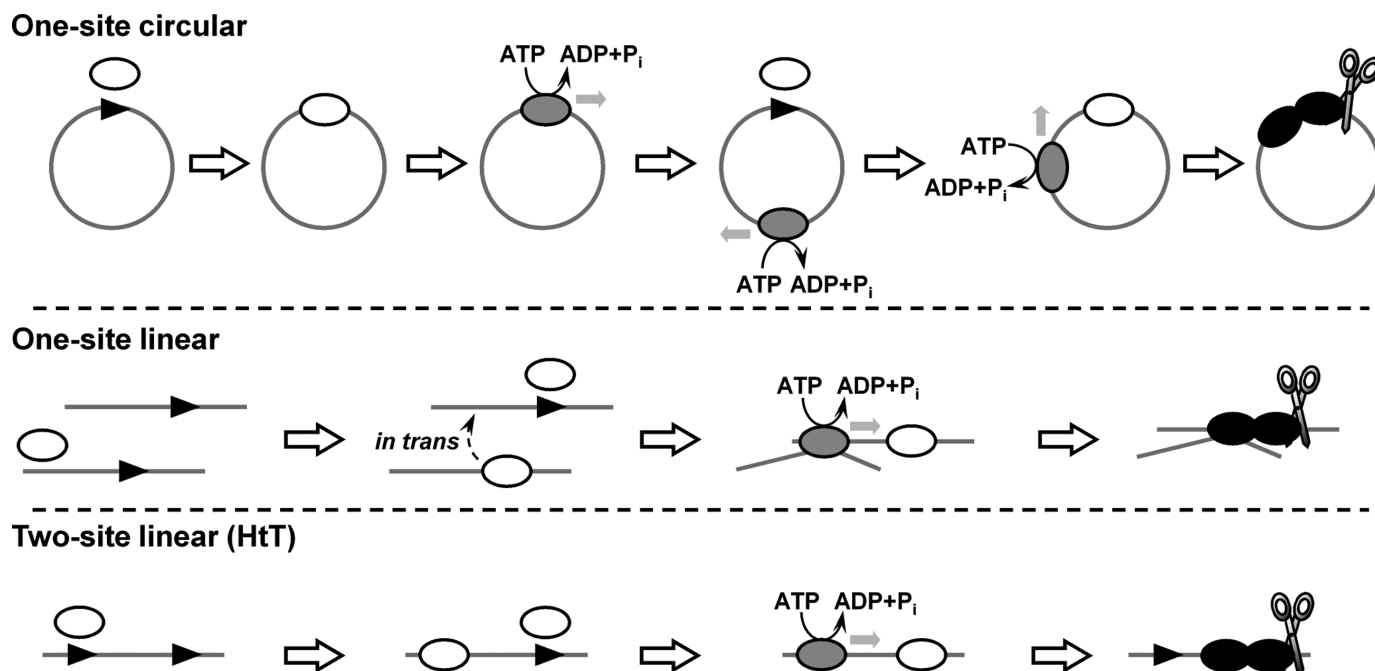
First, we tested whether motor activity can transfer from one DNA molecule to another. In the absence of a CgII target site on a linear DNA, triplex displacement was not observed above background levels (Figure 5B). We then mixed this DNA with a second linear DNA molecule either without a target ('N') or with a target ('H') (Figure 6A). Triplex displacement was observed when a CgII site was added *in trans*, suggesting that the motor activity can switch between DNA molecules, with the second binding event occurring on non-specific DNA. We repeated this experiment using different combinations of circular or linear triplex DNA with circular or linear specific, or with circular or linear non-specific DNA (Figure 6B). Activation of triplex displacement was observed in all cases where a CgII site was provided *in trans*. Therefore, the transfer between DNA molecules does not require a DNA end, either for CgII exit from the specific DNA or for entry on to the non-specific DNA (Figure 6B).

To distinguish between single motor (dissociation/association) and double motor (looping/sandwich) transfer mechanisms, we measured triplex displacement activity using DNA bound to magnetic beads. Where DNA molecules are bound to separate beads, exclusion effects prevent the formation of sandwich complexes between those DNA (49,50). Therefore any transfer must occur via free solution due to dissociation-association events. We tested this by mixing specific or non-specific biotinylated linear triplex DNA, either bound or not bound to streptavidin-coated magnetic beads, with a specific biotinylated linear DNA either bound or not bound to streptavidin coated magnetic beads. The results in Figure 7 indicate that transfer cannot occur via free solution. Also we tried to catch the activated CgII translocase by removing the biotinylated activator DNA (contains the CgII recognition sequence) using streptavidin coated



**Figure 7.** Effect of DNA immobilization on TFO displacement by CgII *in trans*. TFO displacement from the non-specific DNA in the presence of the specific DNA (in solution or immobilized). N, H and T DNA fragments (278 bp in length) are the same as used in Figure 5B. DNA is shown as a thick line, the CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►), the triplex binding sequence is shown as a varied rectangle, biotin is shown as black diamond, magnetic bead is shown as gray circle. All reactions contained 1 nM DNA substrate, 0.5 nM TFO, 4 mM ATP, 200 nM R.CgII, 200 nM H.CgII, 50 nM DNA *in trans* (in solution or immobilized) and were conducted as described in 'Materials and Methods' section. Solid lines are single exponential fits to the data. Rate constants are presented in Supplementary Table S4. Points are averages with error bars as standard deviation for at least three repeat reactions.

magnetic beads and then adding the non-specific DNA triplex, however, no TFO displacement was observed. When non-specific triplex DNA was immobilized on the magnetic beads and mixed with magnetic beads bearing the immobilized specific DNA to prevent simultaneous CgII binding with both specific (activator) and non-specific DNA, triplex displacement was not observed (Figure 7). In contrast, the triplex was displaced when non-immobilized specific DNA was added in solution (Figure 7). This indicates that CgII re-



**Figure 8.** Proposed mechanism of action of CgII on various DNA substrates. DNA is shown as a thick line, the CgII recognition sequence (5'-GCCGC-3') is shown as an arrowhead (►). The  $R_2H_2$ -CgII complex (white ellipse) containing only a single active site binds the recognition sequence, becomes activated (gray ellipse), starts to translocate on DNA using ATP hydrolysis and leaves the target. For clarity here only one of two possible directions of translocation on DNA is shown. The second CgII complex binds to the same recognition sequence. After collision of the target-bound and translocating CgII complexes (black ellipses) a double-stranded break is introduced 7 and 6/7 nucleotides downstream of the 5'-GCCGC-3' site on the top and bottom strands, respectively. In the case of the linear one-site substrate two target bound/activated CgII complexes associate *in trans* introducing a double-stranded break near to the one of the targets (inefficient cleavage). For the linear two-site substrates only a scheme of the HtT fragment cleavage is shown. CgII shows a preferential directionality of translocation on DNA (shown by an arrow) resulting in the predominant DNA cleavage at the second site (downstream to the first).

quires simultaneous binding both specific and non-specific DNA for the TFO displacement *in trans*.

Under equivalent experimental conditions, the rate of triplex displacement ( $\sim 10^{-4} \text{ s}^{-1}$ ) due to *in trans* transfer (Figure 6A) was of the same order as the rates of the second phases of triplex displacement observed in Figure 5, (Supplementary Table S4). Therefore, it is possible that at least some of the observed triplex displacement activity in Figure 5 is due to *in trans* transfers of translocation activity between DNA molecules. Such events would lose the original site orientation. When the triplex was located upstream of the CgII site, the initial triplex displacement rate was faster than the *in trans* transfer (Figure 5B). This could be due to (i) a slower upstream translocation mechanism that is still more efficient than *in trans* transfer; and/or, (ii) a more efficient *in cis* transfer due to DNA looping. For the Type III RM enzymes, a protein roadblock assay was used to rule out looping activities (21). However, the vigorous protein displacement activity of the translocating CgII means that we cannot repeat a similar assay to prove or disprove looping of the motor.

As noted above, the inefficient cleavage of the one-site linear DNA (Figure 3) may be due to inefficient *in trans* transfer events. In principle the *in trans* transfer can be improved by increasing the local concentration of the protein-DNA complex relative to the second DNA. To test this, we measured the cleavage of a one-site linear DNA in the presence of a fixed concentration of CgII and increasing concentra-

tions of a second, smaller one-site DNA (Supplementary Figure S6). An increased efficiency of cleavage was observed as concentration of the second DNA increased, consistent with *in trans* transfer playing a role in cleavage where direct 1D communication is not possible.

Why would an RM enzyme show site orientation preference during long-range communication? For the Type ISP and III enzymes, a simple explanation is that the accompanying methyltransferase activity only modifies one strand (17,51,52). This hemimethylation pattern would result in unmodified sites following semi-conservative replication. However, by requiring pairs of sites in inverted repeat orientation for cleavage, any unmodified site is only aligned with modified sites and cleavage is prevented, allowing maintenance methylation. We currently do not know the methylation pattern for the CgII-like enzymes. However, we note that unlike the Type ISP and III enzymes, there is a less clear-cut site orientation preference for CgII and the results may be more consistent with methylation of both strands which accommodates bidirectional translocation, as for Type I RM enzymes (10).

## CONCLUSION

On the basis of the data presented in this article, we make a number of conclusions and propose alternative cleavage mechanisms to account for the CgII cleavage activity on the different DNA substrates (Figure 8):

- i) A single site plasmid DNA is sufficient for efficient DNA cleavage performed by two CgII complexes. This DNA cleavage requires long-range communication by one of the complexes *in cis* along the 1D contour.
- ii) Communication is most likely driven by a stepwise ATP hydrolysis-dependent translocation mechanism that is active both downstream and upstream of the site (i.e. bidirectionally with some preference downstream of the asymmetric 5'-GCCGC-3' site). The enzyme leaves the site and the original binding orientation of the protein is retained. Motor activity can even produce sufficient force to displace stable R-loop structures of CRISPR effector complexes.
- iii) Unusually, this translocation activity can transfer onto a new DNA substrate with low efficiency. The transfer between DNA molecules does not require a DNA end, either for CgII exit from the specific DNA or for binding to the non-specific DNA, and results from simultaneous interaction between both the DNAs.
- iv) The most efficient cleavage activity is due to rear-end collision between a target-bound complex and a translocating complex. Consequently, circular DNA and linear DNA with sites in direct repeat (i.e. HtT) DNA are the best cleavage substrates.
- v) Cleavage can occur with lower efficiency on linear with a single site or pairs of sites in inverted repeat (either HtH or TtT). This is due to a combination of upstream translocation events and/or motor transfers events *in cis* (looping) or *in trans* (sandwich complexes).
- vi) Regardless of the pathway to activation of the nuclease activity, cleavage always occurs at the same location (GCCGCN<sub>6</sub>/N<sub>6-7</sub>). This indicates that the same CgII collision complex must be produced by different pathways of communication.
- vii) Data from a related enzyme, BceSIV, has been interpreted to show cleavage on both sides of the pseudo-palindromic site GCWGC (53). Although we do not believe that our data support cleavage on both sides of the site, there is the possibility that CgII can occasionally load onto the site in a GCGGC orientation and initiate translocation 'upstream'. This would allow bidirectional translocation but with a strong bias in one direction. Although this is a possibility, the difference in the rates of cleavage of HtH and TtT substrates (Figure 4B) would suggest that there should be more cleavage upstream of the site than observed.

Based on these points, the CgII enzyme does not easily conform to classification as a Type I, ISP or III RM enzyme. Since stepwise translocation is the more likely mode of communication, the closest resemblance is to the Type I enzymes. Single molecule experiments are underway to confirm or refute the bidirectional DNA translocation mechanism of CgII enzyme during the cleavage of DNA.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors are grateful to Dr Tautvydas Karvelis for a kind gift of Cas9 double mutant (D10A+H840A), Dr Tomas Sinkunas for sharing Cascade-crRNA complex and Dr Fiona Diffin for DNA samples. Authors acknowledge colleagues at the Vilnius Department of Protein-DNA Interactions for valuable discussions and suggestions for this work.

## FUNDING

Research Council of Lithuania [MIP-56/2015 to M.Z.]; National Institutes of Health (NIH) [1R01GM111507-01 to M.D.S.]. Funding for open access charge: NIH [1R01GM111507-01 to M.D.S.].

*Conflict of interest statement.* none declared.

## REFERENCES

1. Loenen, W.A., Dryden, D.T., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*, **42**, 3–19.
2. Loenen, W.A., Dryden, D.T., Raleigh, E.A. and Wilson, G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res.*, **42**, 20–44.
3. Rao, D.N., Dryden, D.T. and Bheemanai, S. (2014) Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.*, **42**, 45–55.
4. Pingoud, A. and Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
5. Szczelkun, M.D., Friedhoff, P. and Seidel, R. (2010) Maintaining a sense of direction during long-range communication on DNA. *Biochem. Soc. Trans.*, **38**, 404–409.
6. Szczelkun, M.D. (2011) Translocation, switching and gating: potential roles for ATP in long-range communication on DNA by Type III restriction endonucleases. *Biochem. Soc. Trans.*, **39**, 589–594.
7. Peakman, L.J. and Szczelkun, M.D. (2004) DNA communications by Type III restriction endonucleases—confirmation of 1D translocation over 3D looping. *Nucleic Acids Res.*, **32**, 4166–4174.
8. Ramanathan, S.P., van Aelst, K., Sears, A., Peakman, L.J., Diffin, F.M., Szczelkun, M.D. and Seidel, R. (2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748–1753.
9. Chand, M.K., Nirwan, N., Diffin, F.M., van Aelst, K., Kulkarni, M., Pernstich, C., Szczelkun, M.D. and Saikrishnan, K. (2015) Translocation-coupled DNA cleavage by the Type ISP restriction-modification enzymes. *Nat. Chem. Biol.*, **11**, 870–877.
10. Firman, K. and Szczelkun, M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J.*, **19**, 2094–2102.
11. Bianco, P.R. and Hurley, E.M. (2005) The type I restriction endonuclease EcoR124I, couples ATP hydrolysis to bidirectional DNA translocation. *J. Mol. Biol.*, **352**, 837–859.
12. Bianco, P.R., Xu, C. and Chi, M. (2009) Type I restriction endonucleases are true catalytic enzymes. *Nucleic Acids Res.*, **37**, 3377–3390.
13. Seidel, R., Bloom, J.G., Dekker, C. and Szczelkun, M.D. (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J.*, **27**, 1388–1398.
14. Smith, R.M., Josephsen, J. and Szczelkun, M.D. (2009) The single polypeptide restriction-modification enzyme LlaGI is a self-contained molecular motor that translocates DNA loops. *Nucleic Acids Res.*, **37**, 7219–7230.
15. van Aelst, K., Sisakova, E. and Szczelkun, M.D. (2013) DNA cleavage by type ISP restriction-modification enzymes is initially targeted to the 3'-5' strand. *Nucleic Acids Res.*, **41**, 1081–1090.
16. Sisakova, E., van Aelst, K., Diffin, F.M. and Szczelkun, M.D. (2013) The type ISP restriction-modification enzymes LlaBIII and LlaGI use a translocation-collision mechanism to cleave non-specific DNA distant from their recognition sites. *Nucleic Acids Res.*, **41**, 1071–1080.

17. 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.
18. 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.
19. Szczelkun, M.D. (2013) Roles for helicases as ATP-dependent molecular switches. *Adv. Exp. Med. Biol.*, **767**, 225–244.
20. 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.
21. Toth, J., van Aelst, K., Salmons, H. and Szczelkun, M.D. (2012) Dissociation from DNA of type III restriction-modification enzymes during helicase-dependent motion and following endonuclease activity. *Nucleic Acids Res.*, **40**, 6752–6764.
22. Schwarz, F.W., van Aelst, K., Toth, J., Seidel, R. and Szczelkun, M.D. (2011) DNA cleavage site selection by Type III restriction enzymes provides evidence for head-on protein collisions following 1D bidirectional motion. *Nucleic Acids Res.*, **39**, 8042–8051.
23. Labrie, S.J., Samson, J.E. and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
24. Pingoud, A. and Jeltsch, A. (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.*, **246**, 1–22.
25. Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
26. Taylor, J.D., Badcoe, I.G., Clarke, A.R. and Halford, S.E. (1991) EcoRV restriction endonuclease binds all DNA sequences with equal affinity. *Biochemistry*, **30**, 8743–8753.
27. Narlikar, G.J., Sundaramoorthy, R. and Owen-Hughes, T. (2013) Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, **154**, 490–503.
28. Zaremba, M., Toliuisis, P., Grigaitis, R., Manakova, E., Silanskas, A., Tamulaitiene, G., Szczelkun, M.D. and Siksnys, V. (2014) DNA cleavage by CgII and NgoAVII requires interaction between N- and R-proteins and extensive nucleotide hydrolysis. *Nucleic Acids Res.*, **42**, 13887–13896.
29. 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.
30. Yoshioka, K. (2002) KyPlot—a user-oriented tool for statistical data analysis and visualization. *Comput. Stat.*, **17**, 425–437.
31. Simons, M. and Szczelkun, M.D. (2011) Recycling of protein subunits during DNA translocation and cleavage by Type I restriction-modification enzymes. *Nucleic Acids Res.*, **39**, 7656–7666.
32. Roberts, G.A., Cooper, L.P., White, J.H., Su, T.J., Zipprich, J.T., Geary, P., Kennedy, C. and Dryden, D.T. (2011) An investigation of the structural requirements for ATP hydrolysis and DNA cleavage by the EcoKI Type I DNA restriction and modification enzyme. *Nucleic Acids Res.*, **39**, 7667–7676.
33. 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.
34. Wood, K.M., Daniels, L.E. and Halford, S.E. (2005) Long-range communications between DNA sites by the dimeric restriction endonuclease SgrAI. *J. Mol. Biol.*, **350**, 240–253.
35. Embleton, M.L., Siksnys, V. and Halford, S.E. (2001) DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. *J. Mol. Biol.*, **311**, 503–514.
36. Szczelkun, M.D., Dillingham, M.S., Janscak, P., Firman, K. and Halford, S.E. (1996) Repercussions of DNA tracking by the type IC restriction endonuclease EcoR124I on linear, circular and catenated substrates. *EMBO J.*, **15**, 6335–6347.
37. Marshall, J.J., Gowers, D.M. and Halford, S.E. (2007) Restriction endonucleases that bridge and excise two recognition sites from DNA. *J. Mol. Biol.*, **367**, 419–431.
38. Szczelkun, M.D. and Halford, S.E. (1996) Recombination by resolvase to analyse DNA communications by the SfiI restriction endonuclease. *EMBO J.*, **15**, 1460–1469.
39. Szczelkun, M.D., Janscak, P., Firman, K. and Halford, S.E. (1997) Selection of non-specific DNA cleavage sites by the type IC restriction endonuclease EcoR124I. *J. Mol. Biol.*, **271**, 112–123.
40. van Aelst, K., Saikrishnan, K. and Szczelkun, M.D. (2015) Mapping DNA cleavage by the Type ISP restriction-modification enzymes following long-range communication between DNA sites in different orientations. *Nucleic Acids Res.*, **43**, 10430–10443.
41. Kulkarni, M., Nirwan, N., van Aelst, K., Szczelkun, M.D. and Saikrishnan, K. (2016) Structural insights into DNA sequence recognition by type ISP restriction-modification enzymes. *Nucleic Acids Res.*, **44**, 4396–4408.
42. Seidel, R., Bloom, J.G., van Noort, J., Dutta, C.F., Dekker, N.H., Firman, K., Szczelkun, M.D. and Dekker, C. (2005) Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J.*, **24**, 4188–4197.
43. 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 EcoR124I. *EMBO J.*, **25**, 2230–2239.
44. Blackwood, J.K., Rzechorzek, N.J., Abrams, A.S., Maman, J.D., Pellegrini, L. and Robinson, N.P. (2012) Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex. *Nucleic Acids Res.*, **40**, 3183–3196.
45. Graham, J.E., Sherratt, D.J. and Szczelkun, M.D. (2010) Sequence-specific assembly of FtsK hexamers establishes directional translocation on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20263–20268.
46. Smith, A.J., Szczelkun, M.D. and Savery, N.J. (2007) Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. *Nucleic Acids Res.*, **35**, 1802–1811.
47. McClelland, S.E., Dryden, D.T. and Szczelkun, M.D. (2005) Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases. *J. Mol. Biol.*, **348**, 895–915.
48. Smith, A.J., Pernstich, C. and Savery, N.J. (2012) Multipartite control of the DNA translocase, Mfd. *Nucleic Acids Res.*, **40**, 10408–10416.
49. Zaremba, M., Owsicka, A., Tamulaitis, G., Sasnauskas, G., Shlyakhtenko, L.S., Lushnikov, A.Y., Lyubchenko, Y.L., Laurens, N., van den Broek, B., Wuite, G.J. et al. (2010) DNA synapsis through transient tetramerization triggers cleavage by Ecl18kI restriction enzyme. *Nucleic Acids Res.*, **38**, 7142–7154.
50. Gasiunas, G., Sasnauskas, G., Tamulaitis, G., Urbanke, C., Razaniene, D. and Siksnys, V. (2008) Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family. *Nucleic Acids Res.*, **36**, 938–949.
51. Meisel, A., Mackeldanz, P., Bickle, T.A., Kruger, D.H. and Schroeder, C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.*, **14**, 2958–2966.
52. Meisel, A., Bickle, T.A., Kruger, D.H. and Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, **355**, 467–469.
53. 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.
54. Endlich, B. and Linn, S. (1985) The DNA restriction endonuclease of *Escherichia coli* B. II. Further studies of the structure of DNA intermediates and products. *J. Biol. Chem.*, **260**, 5729–5738.
55. Kimball, M. and Linn, S. (1976) The release of oligonucleotides by the *Escherichia coli* B restriction endonuclease. *Biochem. Biophys. Res. Commun.*, **68**, 585–591.