

Structural characterization of Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage

Carl J. Schiltz, April Lee, Edward A. Partlow, Christopher J. Hosford and Joshua S. Chappie *

Department of Molecular Medicine, Cornell University, Ithaca, NY 14853, USA

Received January 07, 2019; Revised July 29, 2019; Editorial Decision July 30, 2019; Accepted July 31, 2019

ABSTRACT

Overcoming lysogenization defect (OLD) proteins constitute a family of uncharacterized nucleases present in bacteria, archaea, and some viruses. These enzymes contain an N-terminal ATPase domain and a C-terminal Toprim domain common amongst replication, recombination, and repair proteins. The *in vivo* activities of OLD proteins remain poorly understood and no definitive structural information exists. Here we identify and define two classes of OLD proteins based on differences in gene neighborhood and amino acid sequence conservation and present the crystal structures of the catalytic C-terminal regions from the *Burkholderia pseudomallei* and *Xanthomonas campestris* p.v. *campestris* Class 2 OLD proteins at 2.24 Å and 1.86 Å resolution respectively. The structures reveal a two-domain architecture containing a Toprim domain with altered architecture and a unique helical domain. Conserved side chains contributed by both domains coordinate two bound magnesium ions in the active site of *B. pseudomallei* OLD in a geometry that supports a two-metal catalysis mechanism for cleavage. The spatial organization of these domains additionally suggests a novel mode of DNA binding that is distinct from other Toprim containing proteins. Together, these findings define the fundamental structural properties of the OLD family catalytic core and the underlying mechanism controlling nuclease activity.

INTRODUCTION

Phosphoryl transfer reactions are critical for the synthesis and processing of nucleic acids (1). DNA and RNA polymerization, nuclease degradation, RNA splicing, and DNA transposition all proceed via the same general reaction scheme involving (i) an SN2 nucleophilic attack on the

scissile phosphodiester bond, (ii) the formation of a pentavalent transition state and (iii) cleavage of the scissile bond leading to stereo inversion of the scissile phosphate and release of the leaving group (2). These steps depend on the presence of a basic moiety to activate the nucleophile, a general acid to protonate the leaving group, and the presence of positively charged groups to stabilize the developing negative charge in the transition state (3,4). The observed catalytic activity of RNA (5,6) coupled with the presence of two metal ions in the refined structures of alkaline phosphatase (7) and the Klenow fragment with DNA (8) led to the generalized mechanistic hypothesis that metals can substitute for protein side chains in phosphoryl transfer reactions and act as the required general acid and base (9). In this scheme, one metal (metal A) deprotonates the nucleophile while the other (metal B) stabilizes the pentavalent transition state intermediate (2). Despite the prevalence of this mechanism, the number of metal cofactors can vary among different enzyme families. Many homing endonucleases, for example, function using one metal (10,11) while *in crystallo* catalytic studies of human DNA polymerase η reveal an essential catalytic role for a third metal during DNA synthesis (12). Structural characterization of phosphorylhydrolases is therefore necessary for understanding the underlying catalytic strategy employed in each case.

Topoisomerases, DnaG primases, gyrases, RecR recombination proteins and 5S rRNA maturases share a conserved catalytic domain that mediates metal-dependent nicking and cleavage of nucleic acid substrates (13,14). This **Topoisomerase/primase** (Toprim) domain consists of a four-stranded parallel β -sheet sandwiched between two pairs of α -helices and contains three key sequence motifs: an invariant glutamate located in the $\alpha 1$ - $\beta 1$ loop, an invariant glycine following $\beta 2$, and a conserved DxD motif between $\alpha 3$ and $\beta 3$ (13,15). Crystallography and mutagenesis have shown the conserved E and DxD motif to be critical for metal binding and catalytic activity in multiple contexts (14,16–18). Additional active site components vary between Toprim family members based on specific functional

*To whom correspondence should be addressed. Tel: +1 607 253 3654; Fax: +1 607 253 3659; Email: chappie@cornell.edu

requirements. Topoisomerases and gyrases contain a catalytic tyrosine that forms a covalent linkage with the DNA (15,19) whereas DnaG primases have extra acidic residues that coordinate multiple metals needed for nucleotide binding and polymerase activity (16,17,20). While most Toprim proteins play important roles in DNA replication, recombination, and repair, recent structural studies revealed the CWB2 cell wall-anchoring module of *Clostridium difficile* proteins Cwp8 and Cwp6 also contains a Toprim fold (21). These domains, however, lack the conserved metal binding side chains and form trimers that act in a purely structural capacity.

Overcoming lysogenization defect (OLD) proteins constitute a family of uncharacterized enzymes that contain a predicted N-terminal ATPase domain and C-terminal Toprim domain (13,22). Much of our present understanding of OLD function derives from bacteriophage P2 genetic and biochemical studies. The P2 *old* gene product interferes with bacteriophage λ growth in P2 lysogens, kills *Escherichia coli* *recB* and *recC* mutants following P2 phage infection, and causes increased sensitivity of P2 lysogens to X-ray irradiation (23–25). These effects appear to be accompanied by a partial degradation of tRNA molecules and inhibition of protein synthesis (26,27). P2 OLD purified as a maltose binding protein fusion exhibits 5'-3' exonuclease cleavage of DNA and ribonuclease activity *in vitro* (28). Recent genetic studies indicate that the *Salmonella typhimurium old* gene becomes critical under certain growth conditions like temperature stress (29), but its mechanism of action and normal physiological functions remain a mystery. Nothing is known about the activities of other homologs and there are currently no structures of OLD proteins.

Here we identify and define two classes of OLD proteins based on differences in gene neighborhood and amino acid sequence conservation. We purify and characterize the Class 2 OLD proteins from *Burkholderia pseudomallei* (Bp) and *Xanthomonas campestris* p.v. *campestris* (Xcc) and present the crystal structures of their catalytic C-terminal regions at 2.24 and 1.86 Å resolution respectively. The structures show a two-domain arrangement containing a Toprim domain with altered architecture and a unique helical domain. Conserved side chains contributed by both domains coordinate two bound magnesium ions in the active site of Bp OLD, which are absolutely required for nuclease activity. The geometry of this catalytic machinery supports a two-metal catalysis mechanism for cleavage and shows unexpected structural conservation with the active sites of DnaG primases and bacterial RNase M5 maturases. The spatial organization of these domains additionally suggests a novel mode of DNA binding that is distinct from other Toprim containing proteins. Together, these findings define the fundamental structural properties of the OLD family catalytic core and the underlying mechanism controlling nuclease activity.

MATERIALS AND METHODS

Expression and purification of OLD constructs

Full-length (Xcc^{FL}, residues 1–595; Bp^{FL}, residues 1–594) and C-terminal region (Xcc^{CTR}, residues 374–595;

Bp^{CTR}, residues 374–594) OLD constructs were cloned into pET21b, introducing a C-terminal 6xHis tag for purification. Constructs were transformed into BL21(DE3) cells, grown at 37°C in Terrific Broth to an OD₆₀₀ of 0.7–0.9, and then induced with 0.3 mM IPTG overnight at 19°C. Cells were harvested, washed with nickel load buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 30 mM imidazole, 5% glycerol (v/v), and 5 mM β -mercaptoethanol), and pelleted a second time. Pellets were typically flash frozen in liquid nitrogen and stored at –80°C.

Pellets from 500 ml cultures were thawed and resuspended in 30 ml of nickel load buffer supplemented with 10 mM PMSF, 5 mg DNase, 5 mM MgCl₂, and a Roche complete protease inhibitor cocktail tablet. Lysozyme was added to 1 mg/ml and the mixture was incubated for 15 min rocking at 4°C. Cells were disrupted by sonication for a total of 4 min and the lysate was cleared of debris by centrifugation at 13 000 rpm (19 685 g) for 30 min at 4°C. The supernatant was filtered using a 0.45 μ m syringe filter, loaded onto a 5 ml HiTrap chelating column charged with NiSO₄, and then washed with nickel load buffer. Proteins were eluted with an imidazole gradient from 30 mM to 1 M. Pooled fractions were dialyzed overnight into TCBg50 buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 5% glycerol, 1 mM DTT) and further purified by anion exchange and size exclusion chromatography (SEC), using a 5 ml HiTrap Q HP column and a Superdex 75 16/600 pg column respectively. Proteins were exchanged into a final buffer of 20 mM HEPES pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT during SEC and concentrated to 10–40 mg/ml.

Active site mutations were introduced via Quikchange and mutants were expressed and purified in the same manner as wildtype.

Inductively coupled plasma atomic emission spectroscopy (ICP-AES)

Bp^{CTR} and Xcc^{CTR} were cloned into the expression vector pASK-IBA3C, introducing a C-terminal Strep-II tag. Strep-tagged CTR constructs were transformed into BL21(DE3) cells, grown at 37°C in Terrific Broth to an OD₆₀₀ of 0.7–0.9, and then induced with 0.3 mM IPTG overnight at 19°C. Cells were harvested and washed in Strep buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM β -mercaptoethanol). Pellets were resuspended in 50 ml of Strep buffer supplemented with 3 mg DNase, 2 mM MgCl₂, 10 mM PMSF, a Roche complete protease inhibitor cocktail tablet, and 1 mg/ml lysozyme. Following a 10 min incubation at 4°C, the cells were sonicated and cleared via centrifugation. The supernatant was filtered, loaded onto a 5 ml StrepTrap column, and washed with Strep buffer. The protein was eluted with Strep buffer supplemented with 2.5 mM *d*-desthiobiotin. The protein was pooled, concentrated, and injected onto a Superdex 75 10/300 GL column. Bp^{CTR} and Xcc^{CTR} were exchanged into a final buffer of 20 mM HEPES pH 7.5 and 50 mM NaCl, which had been first passed through Chelex 100 resin to remove contaminating divalent cations. The final protein sample was concentrated to ~10 mg/ml. Approximately 500 μ l of each protein sample was dried under vacuum and resuspended in 10 ml of 2%

nitric acid. Samples were analyzed with an iCAP 6000 ICP-ES, Thermo. Measurements were done in triplicate. The determined milliequivalents of metal per protein molecule are listed in Supplementary Table S1.

Crystallization, X-ray data collection, and structure determination

Xcc^{CTR} was crystallized by sitting drop vapor diffusion in 0.1 M Bis Tris Propane pH 7.0–8.0, 11–26% PEG 3350, and 0.15–0.2 M sodium iodide with a drop size of 2 μ l and reservoir volume of 65 μ l. Crystals appeared within seven days at 20°C. Native crystals grown in the presence of iodide were stabilized with 0.25 M NDSB-195 or soaked with either 10 mM potassium pentaiodoplatinum IV for 45 min or 10 mM ethylmercury chloride for 45 min. All samples were cryoprotected by transfer to 100% paratone-N, allowing all mother liquor to exit the crystal prior to freezing in liquid nitrogen. Crystals were of the space group $P4_3$ with unit cell dimensions $a = b = 65.4$ Å, $c = 63.8$ Å and $\alpha = \beta = \gamma = 90^\circ$. Crystals were screened and optimized at the MacCHESS F1 beamline at Cornell University and X-ray diffraction data was collected remotely on the tuneable NE-CAT 24-ID-C beamline at the Advanced Photon Source. Single-wavelength anomalous diffraction (SAD) (30) datasets were collected on a Dectris Pilatus 6MF pixel array detector at 100 K for the platinum, mercury, and iodide derivatives at the energies of 12 300, 11 570, and 7500 eV, respectively. Datasets were integrated and scaled with XDS (31) and Aimless (32) via the RAPD pipeline. Heavy atom sites were located using SHELX (33) and phasing, density modification, and initial model building was carried out using the Autobuild routine of PHENIX (34). Initial figures of merit following density modification was 0.62 for Xcc^{CTR} Pt, 0.64 for Xcc^{CTR} Hg, and 0.504 for Xcc^{CTR} I. Further model building and refinement was carried out in COOT (35) and PHENIX (34) respectively. The final models were refined to the following resolutions and $R_{\text{work}}/R_{\text{free}}$: Xcc^{CTR} Pt, 1.86 Å, 0.212/0.238; Xcc^{CTR} Hg, 1.95 Å, 0.201/0.241; Xcc^{CTR} I, 2.30 Å, 0.215/0.275 (Supplementary Table S2).

Bp^{CTR} was crystallized by sitting drop vapor diffusion in 0.1 M HEPES pH 7.5, 0.23 M MgCl₂, 30% PEG 400 and 0.001 M glutathione with a drop size of 1 μ l and reservoir volume of 65 μ l. Crystals appeared within 2–3 days at 20°C. Samples were cryoprotected by transfer to 100% paratone-N, allowing all mother liquor to exit the crystal prior to freezing in liquid nitrogen. Crystals were of the space group $C222_1$ with unit cell dimensions $a = 83.256$ Å, $b = 105.669$, $c = 123.764$ and $\alpha = \beta = \gamma = 90^\circ$. X-ray diffraction data were collected remotely on the NE-CAT 24-ID-C beamline at the Advanced Photon Source at 100 K on a Dectris Pilatus 6MF pixel array detector. The dataset was integrated and scaled using XDS and Aimless via the RAPD pipeline. The structure was solved by molecular replacement in PHASER (36) using the refined Xcc^{CTR} Pt-soaked structure as the search model. Two molecules were found in the asymmetric unit. Model building and refinement were carried out in COOT (35) and PHENIX (34) respectively. The final model was refined to 2.24 Å resolution with an $R_{\text{work}}/R_{\text{free}}$ of 0.213/0.260 (Supplementary Table S2). The model also contained difference density peaks in the active

site that were modeled as two magnesium ions based on the geometry and the components of the crystallization condition.

All structural renderings were generated using Pymol (Schrodinger) and surface electrostatics were calculated using APBS (37). Conservation based coloring was generated using the ConSurf server (38).

Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)

Purified Bp^{FL} (4 mg/ml), Xcc^{FL} (4 mg/ml), Bp^{CTR} (6 mg/ml) and Xcc^{CTR} (6 mg/ml) were subjected to SEC using a Superdex 200 10/300 gl (GE) column equilibrated in SEC-MALS buffer (20 mM HEPES pH7.5, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT). The gel filtration column was coupled to a static 18-angle light scattering detector (DAWN HELEOS-II) and a refractive index detector (Optilab T-rEX) (Wyatt Technology). Data were collected continuously at a flow rate of 0.5 ml/min. Data analysis was performed using the program Astra VI. Monomeric BSA (6.0 mg/ml) (Sigma) was used for normalization of the light scattering detectors and data quality control.

DNA cleavage assays

100 ng of lambda DNA or pUC19 plasmid DNA was mixed with 8 μ M protein to a final volume of 20 μ l in DNA cleavage buffer (20 mM Tris-OAc pH 7.9, 50 mM KCl, 0.1 mg/ml BSA, 10 mM divalent metal). Reactions were incubated at 37°C for 60 min and quenched with 5 μ l of 0.5 M EDTA pH 8.0. Samples were analyzed via native agarose electrophoresis. DNA degradation was quantified using BioRad Image Lab software and assessed by measuring the amount of ethidium bromide signal in each lane and comparing it to the protein-free DNA sample. Bar graphs represent the average of three independent trials with error bars representing the standard error of the mean. Mutant constructs were assayed in the presence of 10 mM MgCl₂ and 1 mM CaCl₂ based on ICP-AES and metal titration results.

Exonuclease assays

The following DNA oligonucleotides for exonuclease assays were synthesized commercially by Integrated DNA Technologies (IDT):

Exo_US (5' or 3' labeled with 6-carboxyfluorescein, 6-FAM)

5'-CTCACTGGTGCTAGGCAACGTTGAAGTGATCGTACGCGGA-3'

Exo_WT_LS

5'-TCCGCGTACGATCACTTCAACGTTGCCTAGCACCAGTGAG-3'

Exo_GT_LS

5'-TCCGCGTACGATCACTTCAACGTTGCCTGGCACCAGTGAG-3'

Lyophilized single-stranded oligonucleotides were resuspended to 1 mM in 10 mM Tris-HCl and 1 mM EDTA and stored at -20°C until needed. Duplex substrates were prepared by heating equimolar concentrations of complementary strands (denoted with suffixes 'us' and 'ls' indicating upper and lower strands) to 95°C for 15 min followed

by cooling to room temperature overnight. Four substrates were prepared: two wildtype substrates (5' or 3' 6FAM-labeled Exo_US each with Exo_WT_LS) and two G:T mismatched substrates (5' or 3' 6FAM-labeled Exo_US each Exo_GT_LS). For each substrate, a 150 μ l reaction containing 8 μ M protein and 75 pmol of labeled double stranded DNA was prepared in exonuclease buffer (20 mM Tris-OAc pH 7.9, 50 mM K-OAc, 0.1 mg/ml BSA, 10 mM MgCl₂, 1 mM CaCl₂) and incubated at 37°C. 20 μ l aliquots were taken at the indicated time points and quenched with 3 \times loading buffer (80% formamide and 1X TBE). Samples were analyzed by a denaturing (8 M urea) 14% polyacrylamide gel and visualized using Bio-Rad ChemiDoc XRS+.

RESULTS

Identification and classification of OLD homologs

Recombinant expression of P2 OLD produced unstable protein that aggregated and/or precipitated, regardless of the tag or conditions employed. We therefore searched the KEGG database (39) to identify OLD homologs more suitable for structural and biochemical characterization. The initial search was carried out using the *E. coli* K12 MG1665 OLD homolog (KEGG ID eco:b0876), which is annotated as the uncharacterized protein YbjD and is 18% identical and 35% similar to P2 OLD. These efforts yielded 833 homologs distributed across numerous kingdoms but absent in eukaryotes (Supplementary Table S3). A further search of mapped plasmid genomes available in the Integrated Microbial Genomes database (40) yielded four additional OLD homologs. We then examined the genetic context of each *old* gene, as inspection of gene neighborhoods has been shown to elucidate unanticipated genetic connections and facilitate new functional predictions (41). Our analyses show that *old* genes segregate into two primary classes (Supplementary Table S3). Class 1 OLD family members (542/837)—including P2 phage, *Escherichia coli*, and *Salmonella typhimurium*—exist as single, isolated genes (Supplementary Figure S1A). Class 2 OLD homologs (295/837) appear in tandem with a UvrD/PcrA/Rep-like helicase (Supplementary Figure S1B), often as an overlapping reading frame. UvrD, PcrA, and Rep are non-hexameric, superfamily 1A helicases that translocate with a 3'-5' polarity and play essential roles in DNA replication, recombination, and repair (42,43). Both classes retain the conserved motifs characteristic of ATPase and Toprim domains, though Class 1 proteins are on average \sim 50 amino acids shorter. Each class appears in a number of different phyla, with examples present in both Gram positive and Gram negative bacteria, archaea, and bacteriophage viruses.

A subset of *old* genes (107/837) exist in species-specific operons (Supplementary Table S3). Neighboring genes within these operons contribute to numerous biological functions including bacterial defense, DNA replication and repair, transcriptional regulation, membrane transport, biosynthesis, metabolism, and signaling (Supplementary Table S3).

We selected numerous candidates from each class for expression studies. Like P2 OLD, most Class 1 homologs behaved poorly during purification. Class 2 homologs, in con-

trast, were intrinsically more stable and generally provided greater yields of soluble, monodispersed protein. Specifically, the Class2 OLD homologs from *B. pseudomallei* and *X. campestris* p.v. *campestris* could be purified to homogeneity (Figure 1A, Supplementary Figure S1C and D) and concentrated to greater than 10 mg/ml without appreciable aggregation or precipitation. Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) indicates that Xcc OLD forms stable tetramers in solution while Bp OLD (Bp^{FL}) exists in equilibrium between dimers and tetramers (Supplementary Figure S1E). In contrast, truncated constructs containing the C-terminal region of each homolog (Xcc^{CTR}, Bp^{CTR}; Figure 1A and Supplementary Figure S1C and D) were each monomeric by SEC-MALS analysis (Supplementary Figure S1F).

Class 2 OLD proteins exhibit metal-dependent DNA cleavage *in vitro*

Metal-dependent nicking and cleavage of nucleic acid substrates is a hallmark of Toprim domain-containing proteins (44). To verify that purified Class 2 OLD proteins share a similar activity *in vitro*, we incubated Bp^{FL} with linearized λ phage DNA in the presence of different divalent cations (Figure 1B). Cleavage activity was quantified by measuring the ethidium bromide signal in each lane and calculating the fraction of DNA digested relative to the untreated substrate, which increased under conditions that promote nuclease function. Bp^{FL} exhibits cleavage in the presence of Mg²⁺, degrading approximately 10% of the substrate within an hour. Activity is enhanced in the presence of Mn²⁺, where 60% of the DNA substrate is degraded (Figure 1B). We also observe weak activity in presence of Zn²⁺ and Co²⁺. Bp^{CTR} similarly shows cleavage with Mg²⁺ (5% degradation) and Mn²⁺ (60%), though it is also highly active in the presence of Co²⁺ (70%) (Figure 1C). Given that Co²⁺ only stimulates activity in Bp^{CTR}, we suspect that this is a construct-specific artifact rather than a general feature of the OLD nucleases. Xcc^{FL} and Xcc^{CTR} similarly can degrade DNA with Mg²⁺ and Mn²⁺ but also are partially active in Zn²⁺ (Supplementary Figure S2A and B). These data indicate that the critical catalytic residues associated with nuclease function reside in the C-terminal half of OLD proteins and that the N-terminal region containing the ATPase domain is not required for DNA binding or nuclease activity.

We next assessed the ability of Bp OLD to nick and cleave circular plasmids. Bp^{FL} was mixed with supercoiled pUC19 DNA (S) in the presence of different divalent metals and activity was evaluated by the appearance of slower migrating bands as the substrate was nicked (N) and linearized (L) by the enzyme (Figure 1D). Bp^{FL} shows weak nicking activity with all metals as compared to the DNA alone and EDTA controls (Figure 1D), with Mg²⁺, Mn²⁺ and Co²⁺ again eliciting the strongest nicking effects. Under these conditions, only Mn²⁺ promotes processive cleavage, degrading 55% of the circular substrate (Figure 1D). Bp^{CTR} shows pronounced nicking activity in the presence of every metal tested, with some processive cleavage stimulated by Mn²⁺ (31% degraded), Co²⁺ (22% degraded), and Ca²⁺ (24% degraded) (Figure 1E). Xcc^{FL} and Xcc^{CTR} show the strongest nicking and cleavage activities on supercoiled DNA with

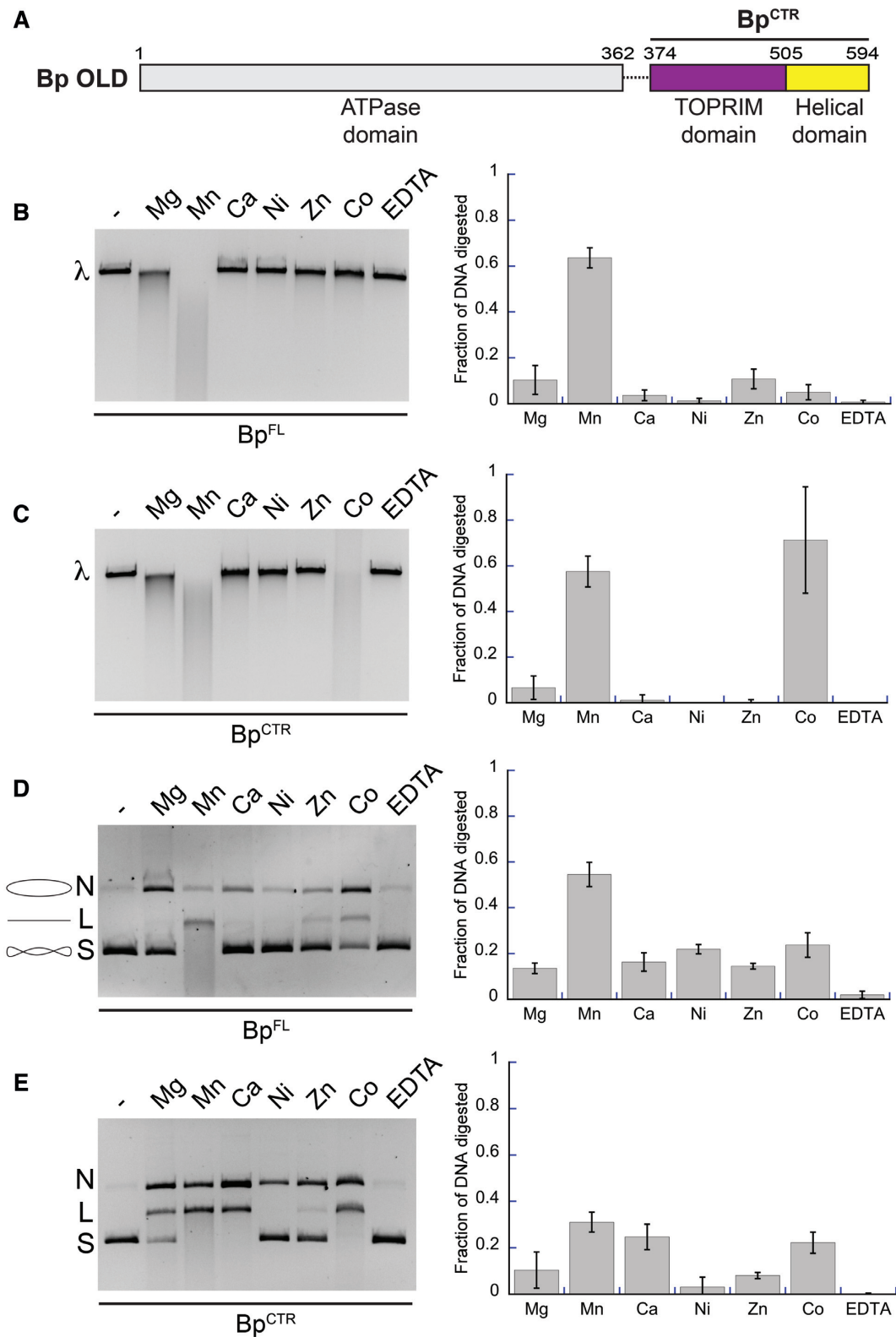


Figure 1. Bp OLD exhibits metal-dependent nuclease activity. (A) Domain architecture of Bp OLD. Boundaries of the CTR construct is labeled. (B, C). Metal-dependent nuclease activity of Bp^{FL} (B) and Bp^{CTR} (C) on linear λ DNA. ‘λ’ denotes the position of uncut linear DNA. (D, E) Metal-dependent nuclease activity of Bp^{FL} (D) and Bp^{CTR} (E) on supercoiled pUC19 DNA. ‘N’, ‘L’, and ‘S’ denote the positions of ‘nicked’, ‘linearized’, and ‘supercoiled’ DNA respectively. Cartoons of these respective products are illustrated in D next to their label. For B–E, lanes labeled with dashed lines indicate samples with no metal added. DNA degradation was quantified using BioRad Image Lab software as described in the Materials and Methods. Bar graphs represent the average of three independent experiments with error bars representing the standard error of the mean.

Mg²⁺, Mn²⁺ and Zn²⁺, though Xcc^{CTR} appears to be able to nick Ca²⁺ as well (Supplementary Figure S2C and D). We note that the extent of cleavage in Xcc is less than Bp overall, suggesting it is a less efficient nuclease.

Given the variation in nuclease function we observed for Bp and Xcc OLD with different metals *in vitro*, we sought to identify which metals are preferentially bound to the CTR constructs *in vivo* using inductively coupled plasma atomic emission spectroscopy (ICP-AES). This technique can measure the type and amount of metal in a given sample with high accuracy (45). Bp^{CTR} and Xcc^{CTR} constructs were purified using Strep-II tags to avoid any confounding results arising from coincidental metal binding to a His tag. ICP-AES showed calcium to be the most abundant metal associated with both Bp^{CTR} (79.35 mEq) and Xcc^{CTR} (100.27 mEq), followed by magnesium (18.38 and 11.39 mEq, respectively), and then by zinc and nickel (Supplementary Table S1). Sparing amounts of cobalt were detected in the Bp^{CTR} sample, suggesting it is not as physiologically relevant. No manganese was found in either sample (Supplementary Table S1). Given the unexpected presence of calcium, we tested if it may play a role in modulating nuclease activity. Presence of Ca²⁺ alone does not promote robust nuclease activity on linear or supercoiled DNA substrates; however, a combination of Ca²⁺ and Mg²⁺ enhances the activities of both Bp^{CTR} and Xcc^{CTR} above either metal alone (Supplementary Figure S3A and B). Nuclease activity is most stimulated with Mg²⁺ in excess and Ca²⁺ between 1 and 2 mM. Under these optimal conditions, the activity of both Bp^{CTR} and Xcc^{CTR} is stimulated more than 10-fold on linear DNA compared to Ca²⁺ or Mg²⁺ alone. Degradation of circular DNA was also enhanced 4–5-fold for Bp^{CTR} and Xcc^{CTR}. Addition of ATP had no appreciable effect on Bp^{FL} cleavage of either substrate (linear versus supercoiled DNA) in the presence of optimal concentrations of Ca²⁺ and Mg²⁺ (Supplementary Figure S3C), further underscoring the notion that the CTR mediates the DNA binding and nuclease functions. Taken together, our results imply that Ca²⁺ acts as an important modulator of OLD family nuclease activity and can potentiate the catalytic effects of these enzymes in the presence of Mg²⁺.

The robust degradation of linear substrates we observe does not explicitly distinguish between exo- and endonuclease activities. To test the exonuclease function and directionality, we incubated Bp^{CTR} with a 40 bp double stranded DNA substrate labeled on the 5' or 3' end with 6-carboxyfluorescein (6-FAM). Bp^{CTR} degrades the 3'-labeled substrate in a stepwise manner (Supplementary Figure S4A) while no intermediates or laddering is observed on the 5'-labeled substrate (Supplementary Figure S4B). These findings indicate Bp^{CTR} can act as an exonuclease that digests DNA in the 5'-3' direction as well as an endonuclease that can act on supercoiled, circular DNA substrates.

Overall structures of the Xcc and Bp OLD C-terminal regions

Although full-length Xcc and Bp OLD proteins crystallize in the presence of different adenine nucleotides, diffraction rarely exceeded ~4 Å and interpretable electron density maps could not be obtained owing to severe radiation damage. Isomorphous crystals were never observed for any

condition screened thereby preventing merging of data. The truncated Xcc^{CTR} construct, in contrast, yielded crystals that routinely diffracted beyond 2 Å and three independent structures were solved using SAD datasets from platinum, mercury, and iodide derivatives (Supplementary Table S2, Supplementary Figures S5 and S6A). These models show strong agreement with an overall RMSD of 0.42–0.44 Å and display only slight deviations at the N-terminus near the mercury-binding site and within a flexible loop containing two adjacent glycines (G479 and G480) (Supplementary Figure S6A). Residues 374–387 and 458–463 are disordered in each structure, though present in the purified construct. Crystals of the analogous Bp^{CTR} construct diffracted to a slightly lower resolution (2.2–2.3 Å) but produced a more complete structural model (residues 390–594; Figure 2).

Bp^{CTR} contains two domains (Figure 2A): a Toprim domain (residues 390–504, purple) and a unique helical domain (residues 505–594, yellow) consisting of a five-helix orthogonal bundle and an additional C-terminal amphipathic helix ($\alpha 6^H$). $\alpha 6^H$ extends into a groove along one face of the Toprim's central β -sheet, forming extensive hydrophobic interactions (Figure 2B and C). Helix $\alpha 5^H$ and the upper portion of $\alpha 6^H$, along with the connecting loop, wrap around the hydrophobic helix $\alpha 1^T$ of the Toprim domain to stabilize the structure further. The contributing hydrophobic side chains are largely conserved among Class 2 OLD proteins (Supplementary Figure S7) and together bury a total surface area of 1341 Å². Similar interactions are observed between the domains in the Xcc^{CTR} (Supplementary Figure S5). Attempts to express Bp^{CTR} and Xcc^{CTR} constructs lacking $\alpha 6^H$ were unsuccessful as deletion of $\alpha 6^H$ rendered the proteins insoluble. This likely reflects the critical stabilizing interactions provided by conserved residues along the $\alpha 6^H$ - β -sheet interface and the exposure of a large hydrophobic surface if this helix is absent.

Many Toprim family members contain individual structural inserts into the core Toprim fold (Figure 3, Supplementary Figures S8 and S9). These include an insertion of variable size and structure between $\beta 2$ and $\alpha 2$ in topoisomerases, gyrases, and RecR (Insert 1, light blue), short helical insertions between $\alpha 2$ and $\beta 3$ (Insert 2 green) and $\alpha 3$ and $\beta 4$ (Insert 3, cyan) in topoisomerases, a two-stranded β -hairpin added between $\alpha 1$ and $\beta 1$ that extends the central β sheet in gyrases and topoisomerase III (Insert 4, red), and an α helix following the shortened $\beta 4$ in the putative RNase M5 from *Aquifex aeolicus* (Insert 5, brown). Bp and Xcc OLD lack most of these embellishments but contain an Insert 3 helix (Figure 3, Supplementary Figures S8 and S9, Teal). Class 2 OLD proteins show sequence variability across this insert region (Supplementary Figure S7). Significantly, structural superposition reveals a shift of the Toprim $\alpha 2$ and $\alpha 3$ helices in OLD proteins relative to all other Toprim family members (Figure 3B and C, Supplementary Figure S8) while the rest of the core fold is largely unchanged (Figure 3C, Supplementary Figure S8). The position of these helices is consistent between the Bp^{CTR} and Xcc^{CTR} structures, arguing it is an intrinsic topological feature and not simply due to crystal packing. This comparison also shows that the OLD helical domain is distinctly separated from all other inserts, localized on the opposite side of the Toprim fold (Figure 3A). We do note that DnaG

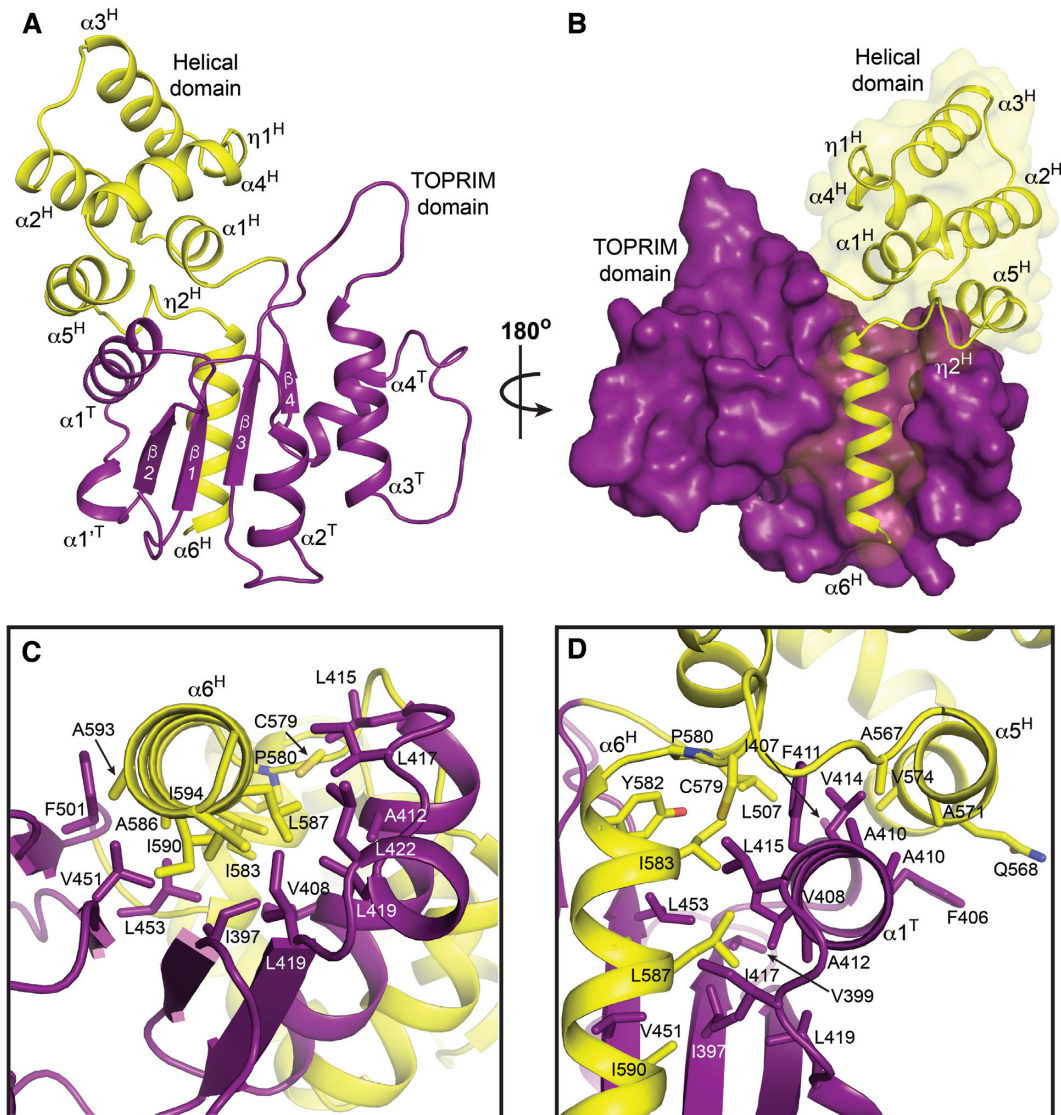


Figure 2. Structure of Bp^{CTR}. (A) Structure of Bp^{CTR}. Toprim, and helical domains are colored purple and yellow respectively. (B) Surface representation of domain interactions. (C, D) Structural interactions between Toprim and helical domains. Side chains involved in stabilizing hydrophobic interactions shown as sticks are labelled.

primases and the putative *A. aeolicus* RNase M5 contain a helix that structurally aligns with the $\alpha 1^H$ helix of the OLD helical domain (Figure 3A, dashed circle).

The helical domain shares structural homology with bacterial controller (C) proteins from restriction–modification (R–M) systems (top hit from the DALI server (46): C.Esp1396I, Z score: 5.1, RMSD 2.5 Å) (Figure 4). C proteins act as transcriptional regulators that tune the expression of R–M methyltransferase and restriction genes to ensure that site-specific nuclease activity is delayed until after a bacterial genome is protected by methylation (47). Crystallographic studies have shown that these proteins are dimeric and α -helical, with each monomer containing a helix–turn–helix motif (48). Structural superposition aligns Bp helices $\alpha 2$, $\alpha 3$, $\alpha 5$ and $\alpha 6$ with $\alpha 1$, $\alpha 3$, $\alpha 4$, and $\alpha 5$ of C.Esp1396I (Figure 4A and B). Bp OLD lacks a helix corresponding to C.Esp1396I $\alpha 2$ and contains two additional helices ($\alpha 1$ and

$\alpha 4$) that localize to the opposite side of the molecule (Figure 4A and B). C protein dimers bind DNA operator sites cooperatively to exert concentration-dependent switching of promoter activation and repression (47,49). In this arrangement, $\alpha 4$ facilitates dimerization while $\alpha 2$ and $\alpha 3$ associate with DNA (Figure 4C). The Bp $\alpha 1$ and $\alpha 4$ helices would sterically block dimerization and DNA interactions respectively, thus preventing OLD proteins from adopting a similar configuration.

Bp OLD active site suggests a two-metal catalysis mechanism

The Xcc^{CTR} derivative structures contain nothing in their active sites. Bp^{CTR} crystallized in a different space group thereby permitting the helical domain to rotate and scrunch closer to the Toprim domain (Supplementary Figure S6B). Consequently, T506 and E508 shift 1.2 and 1.5 Å toward the active site respectively, which facilitates the binding of

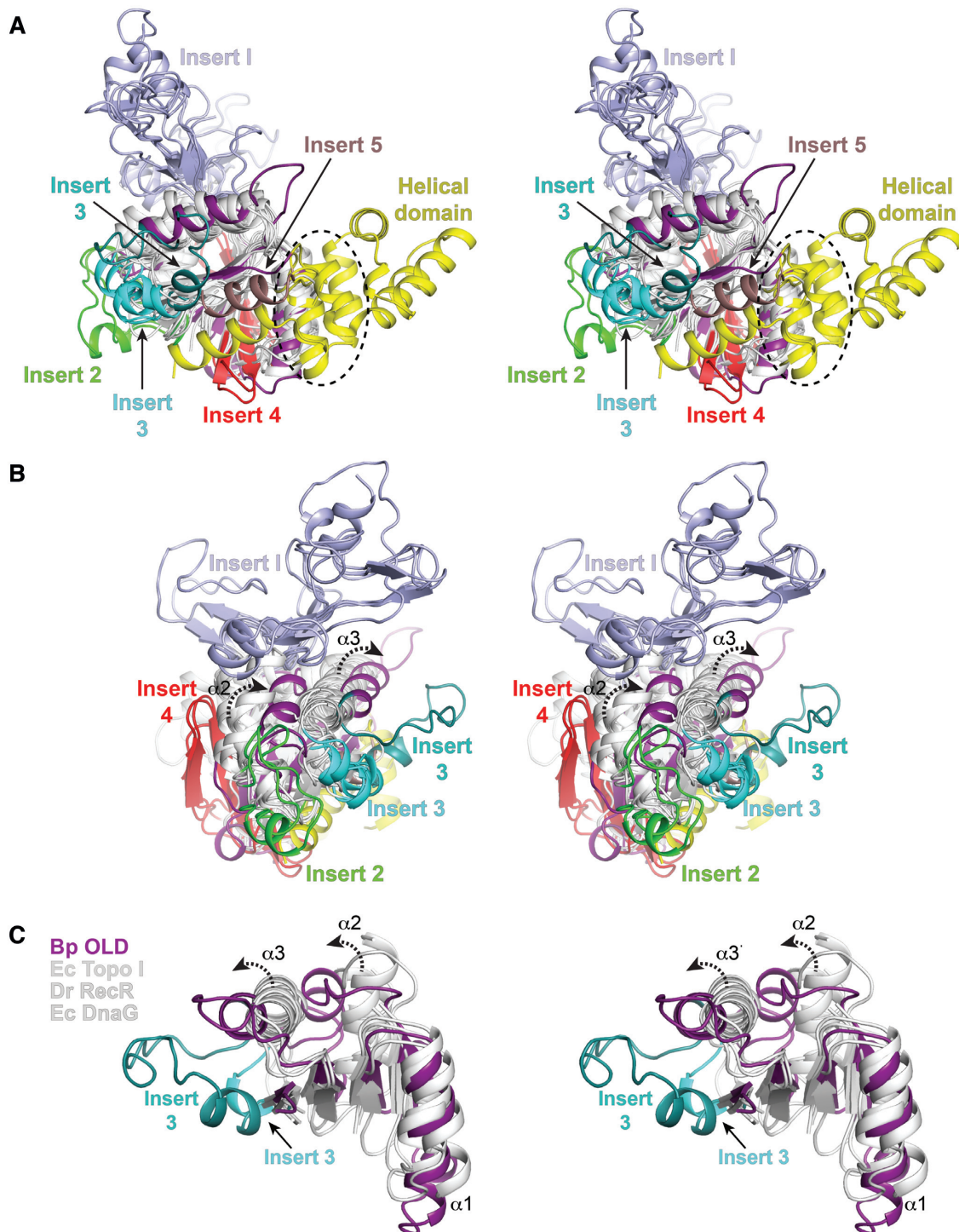


Figure 3. Topological differences in present in Class 2 OLD Toprim domains. (A, B) Structural superposition of Toprim domains. Side (A) and end (B, rotated 90°) views are shown in stereo. Structural alignment includes the following: *Geobacillus kaustophilus* conserved hypothetical protein (PDB: 2FCJ), *Escherichia coli* DnaG primase (PDB: 3B39), *Aquifex aeolicus* putative ribonuclease M5 (PDB: 1T6T), *Deinococcus radiodurans* RecR (PDB: 1VVD), *Thermotoga maritima* Topoisomerase I (PDB: 2GAJ), *Escherichia coli* Topoisomerase I (PDB: 1MW9), *Homo sapiens* Topoisomerase IIB (PDB: 3QX3), *Streptococcus pneumoniae* IV topoisomerase (PDB: 4I3H), *Escherichia coli* Topoisomerase III (PDB: 2054), *Archaeoglobus fulgidus* reverse gyrase (PDB: 1GKU), and *Thermotoga maritima* reverse gyrase (PDB: 4DDU). Toprim cores are colored white with Inserts 1–5 individually labeled (see Supplementary Figures S8 and S9). Bp Toprim and helical domains are colored purple and yellow respectively. Dashed circle in (A) denotes the position of the glutamate helix. (C) Stereo view of $\alpha 2$ and $\alpha 3$ helical shifts (dashed arrows) in Bp OLD Toprim core relative to the Toprim central β -sheet. Central segments of *Escherichia coli* (Ec) Topoisomerase I, *Deinococcus radiodurans* (Dr) RecR and *Escherichia coli* DnaG are shown for comparison. Alternative position of Insert 2 in Bp OLD is also shown.

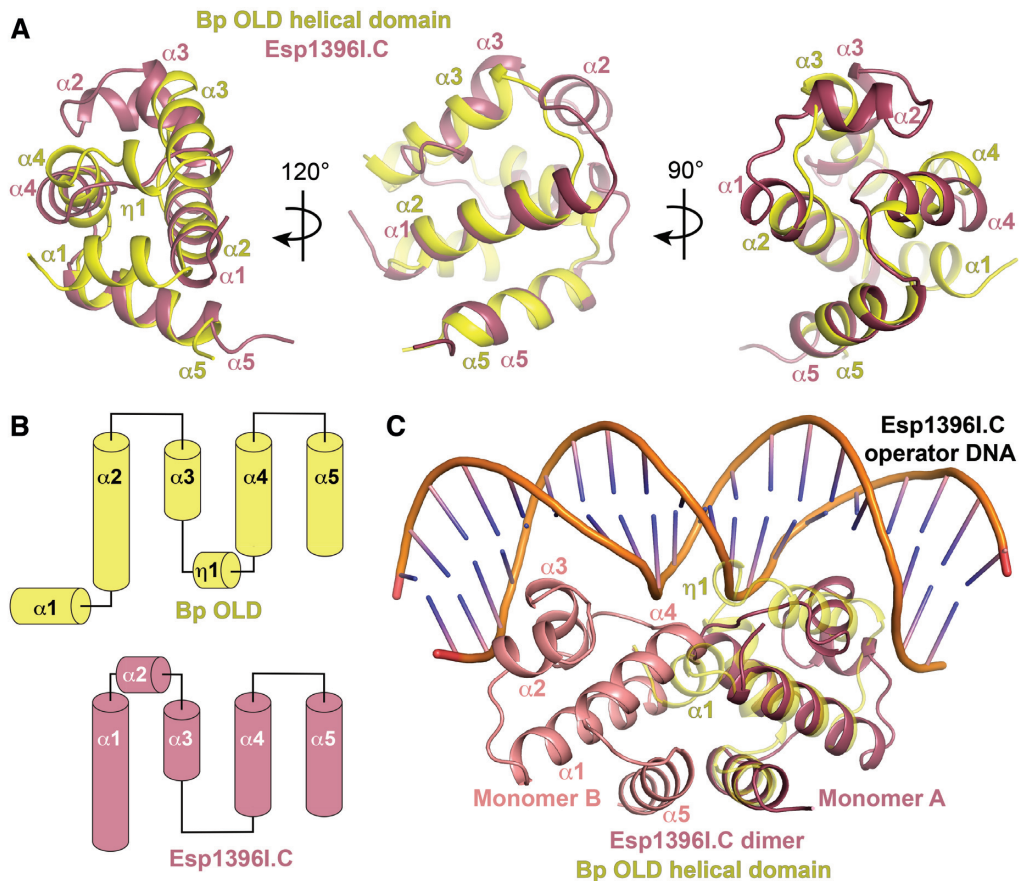


Figure 4. Helical domain in Class 2 OLD proteins shares structural homology with R–M system controller proteins. (A) Structural superposition of Bp OLD helical domain (yellow) with Esp1396I.C controller protein (raspberry; PDB: 3CLC). (B) Topology diagrams of Bp OLD helical domain and Esp1396I.C monomer. (C) Bp OLD architecture precludes dimerization as in controller proteins and suggests an alternate mode of DNA binding.

two magnesium ions in a geometry consistent with two metal catalysis (Figure 5A, Supplementary Figure S6C). Each magnesium is octahedrally coordinated with a water molecule bridging the two metals where the scissile phosphate would normally sit (Figure 5A). The metals are spaced 4.9 Å apart, suggesting they may move closer together once DNA is engaged. The conserved Toprim glutamate (E400) and the first aspartate of the DxD motif (D455) each provide a ligand to the first magnesium (metal A). The second DxD aspartate (D457) hydrogen bonds with two waters that form two additional metal A ligands. E508, located in α H1 of the helical domain, directly coordinates the second magnesium (metal B), while E404 and T506 stabilize additional metal B waters. These side chains are absolutely conserved in Class 2 OLD nucleases (Figure 5B, Supplementary Figure S7). Individual substitutions of metal A ligands (E400A, D455A, D457A) and metal B ligands (E404A, T506A, E508A) yielded no discernible effects on cleavage activity (Figure 5C and D). Thus, combinations of mutations (3A, E400A/D455A/D457A; 3B, E404A/T506A/E508A) were generated. Mutant combinations of either the metal A or the metal B interacting residues together completely abolish Bp OLD nuclease activity on linear DNA substrates *in vitro* (Figure 5C). These substitutions impair the processive degradation of circu-

lar plasmids in the presence of Mg^{2+} and Ca^{2+} , though some nicking activity is still retained (Figure 5D). Simultaneous mutation of both metal A and metal B sites together (2A/2B, D455A/D457A/T506A/E508A) eliminates processive cleavage and significantly reduces nicking activity relative to the 3A and 3B substitutions (Figure 5C and D). This suggests that a single metal in either site can facilitate nicking but both sites are required for processive cleavage and degradation.

We also identify a conserved lysine residue in $\alpha 5^H$ (K562) that extends toward the active site (Figure 5A and B, Supplementary Figure S7), separated from metal A by 5.5 Å and from metal B by 3.8 Å. K562A and K562E mutations similarly impair processive nuclease activity without perturbing direct interactions with either magnesium (Figure 5C and D). K562A and K562E mutations, however, retain the ability to nick DNA (Figure 5D), similar to the perturbation of the individual metal A and B binding sites. Together these data define the key catalytic machinery of Class 2 OLD nucleases and support a two-metal catalysis mechanism for processive nuclease activity.

The organization of the Bp OLD active site is structurally conserved in RNase M5 enzymes and DnaG primases (Figure 6, Supplementary Figure S9). Along with the invariant Toprim glutamate and conserved DxD aspartates, D31 and

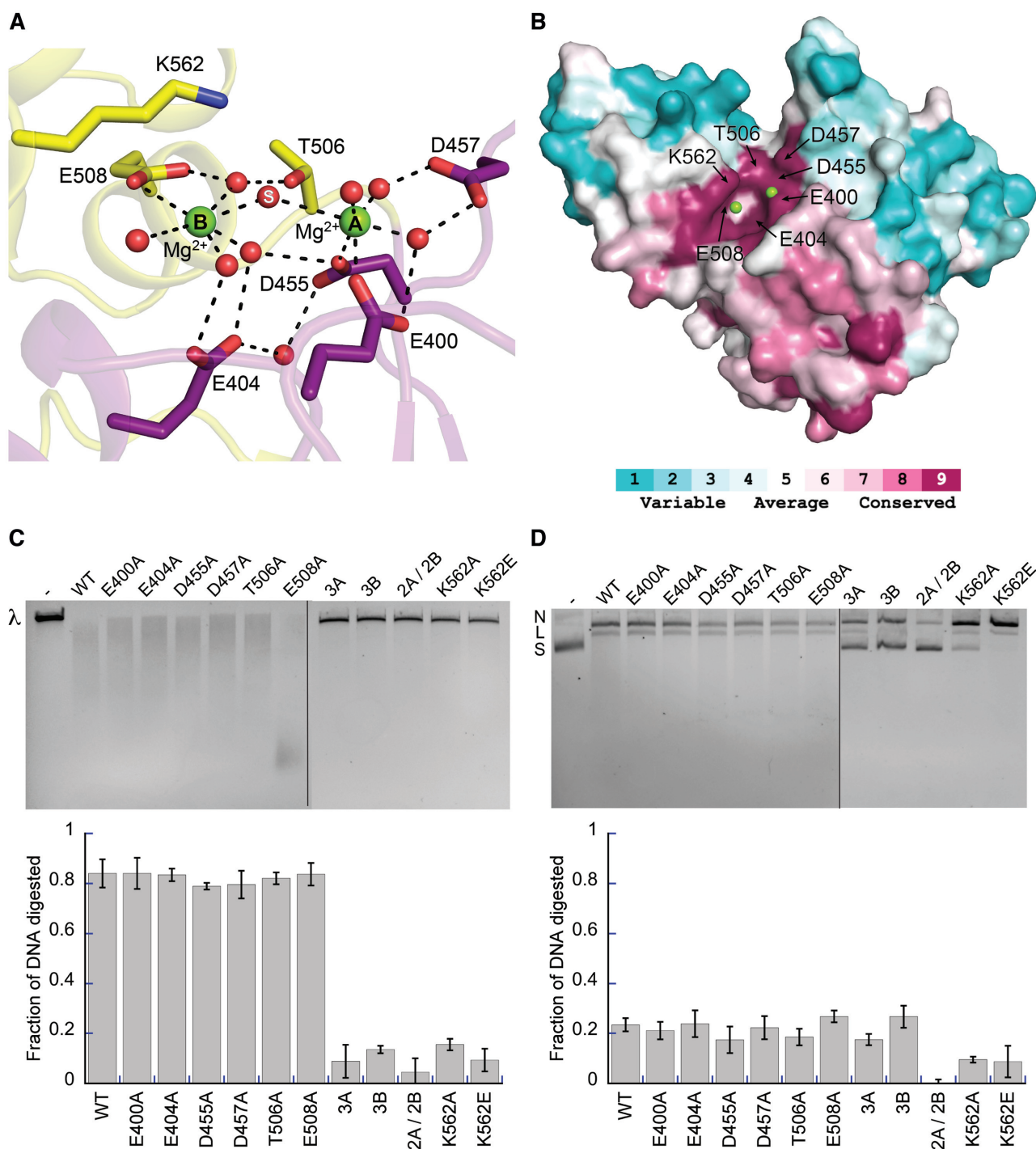


Figure 5. Bp OLD active site geometry is consistent with two-metal catalysis. (A) Metal coordinating residues in the Bp OLD active site. Ordered magnesium ions (green) and water molecules (red) are shown as spheres. Dashed lines indicate hydrogen bonds. (B) Conservation of active site residues. Coloring generated using the ConSurf server (38) and the alignment in Supplementary Figure S7. (C) Cleavage activities of Bp^{CTR} active site mutants on linear λ DNA. 3A denotes triple mutant in the metal A binding site (E400A/D455A/D457A), 3B signifies triple mutant in the metal B binding site (E404A/T506A/E508A), 2A/2B contains mutations in both metal-binding sites (D455A/D457A/T506A/E508A). (D) Nicking and cleavage activities of Bp^{CTR} mutants on supercoiled pUC19 DNA. ‘N’, ‘L’ and ‘S’ denote the positions of ‘nicked’, ‘linearized’, and ‘supercoiled’ DNA respectively. All cleavage assays were performed with 10 mM MgCl₂ and 1 mM CaCl₂. Lanes labeled with dashes in C and D indicate substrate alone controls without protein added. DNA degradation was quantified using BioRad Image Lab software as described in the Materials and Methods as described in the Materials and Methods. Bar graphs show the average of three independent experiments with error bars representing the standard error of the mean.

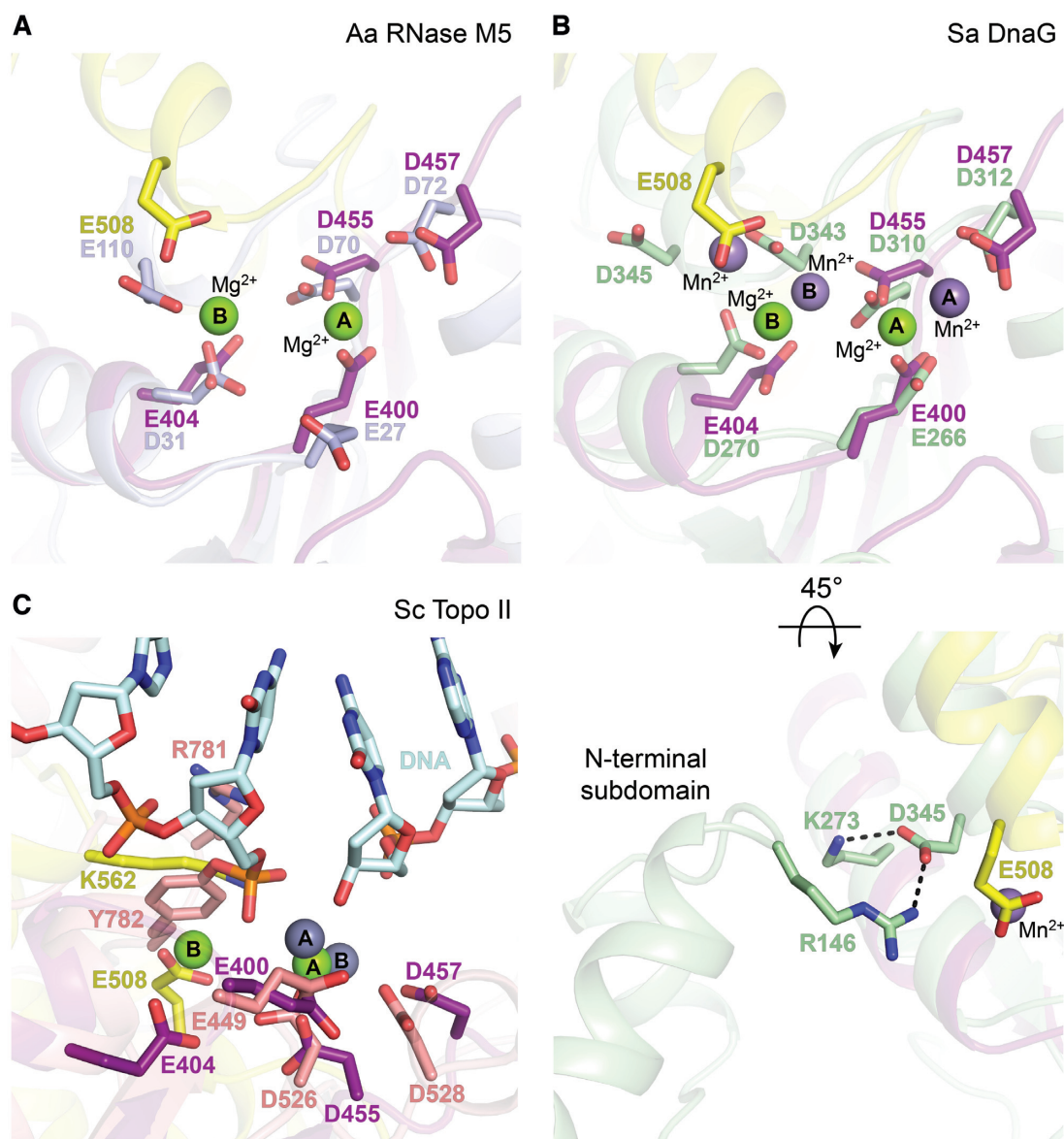


Figure 6. Conservation of active site elements among OLD nucleases, RNase M5 maturases, and DnaG primases. Comparison of Bp^{CTR} active site (Toprim domain, purple; helical domain, yellow) with different Toprim family members. Critical side chains in each protein are labeled and the bound metal A and metal B magnesiums are shown as green spheres. (A) Superposition with the putative RNase M5 from *Aquifex aeolicus* (light blue, PDB: 1T6T). (B) Superposition with DnaG primase from *Staphylococcus aureus* (light green, PDB: 4EE1). Bound manganese ions in the Sa DnaG structure are shown as light purple spheres. Lower panel is rotated and recentered to illustrate interactions between N-terminal subdomain and the Toprim within the Sa DnaG RNA polymerase core. Dashed lines indicate hydrogen bonds. (C) Superposition with *Saccharomyces cerevisiae* topoisomerase II (salmon, PDB: 3L4K). Bound zinc ions in the Sc Topo II structure are shown as gray spheres. Covalently-linked DNA substrate is colored cyan.

E110 in the *A. aeolicus* RNase M5 homolog spatially align with E404 and E508 in Bp OLD (Figure 6A). E110 localizes to a C-terminal helix that superimposes with $\alpha 1^H$ of the Bp OLD helical domain (Figure 3A, dashed circle; Supplementary Figure S9, inset). DnaG primases contain a similar set of catalytic machinery (16). The analogous C-terminal acidic residue of the DnaG Toprim (D345 in *Staphylococcus aureus*), however, is directed away from the active site via interaction with a conserved arginine residue (R146 in *Staphylococcus aureus*) in the adjacent N-terminal subdomain of the RNA polymerase core (Figure 6B). As a result, a third metal (metal C) binds in the position occupied

by E508 in Bp OLD, coordinated by a conserved aspartate residue immediately upstream (D343 in the *Staphylococcus aureus*) (20) (Figure 6B). The arrangement of metals relative to the core catalytic side chains in these enzymes is distinct from the coordination observed in topoisomerases, where metal B is positioned closer to the Dx/D motif in the absence of additional acidic residues (Figure 6C). Metal B and the catalytic lysine in OLD proteins occupy the same position as the catalytic tyrosine that forms a covalent linkage with DNA in topoisomerases (Y782 in *Saccharomyces cerevisiae* topoisomerase II) (50). These differences highlight the evo-

lutionary fine tuning of the Toprim scaffold for unique biological functions.

Structural model for DNA binding

Our attempts to co-crystallize OLD proteins with nucleic acids have thus far been unsuccessful. The robust nuclease activity exhibited by Bp^{CTR} suggests that this fragment alone can associate with DNA in a manner that is competent for cleavage. We therefore computationally modeled DNA onto the Bp^{CTR} structure to gain insight into how OLD nucleases interact with their substrates. Calculation of surface electrostatics identifies four basic patches on one face of Bp^{CTR} that flank a small cleft containing the active site (Figure 7A). Patch 1 lies between $\alpha 3^H$ and $\alpha 4^H$ in the helical domain, formed by R552, K555 and R559 (Supplementary Figure S10A). The catalytic K562 lysine on $\alpha 5^H$ constitutes patch 2. As noted above, this extends into the active site and has a direct role in nuclease activity (Figure 5C and D). Patch 3 localizes along $\alpha 3^T$ in Toprim domain, comprised of R467 and K468, while patch 4 contains R405, which extends from $\alpha 1^T$ toward $\beta 2$ (Supplementary Figure S10A). Modeled B form DNA can bind patches 1 and 2 and part of patch 3, but sterically clashes with the protein beyond the active site cleft (Supplementary Figure S10B). In contrast, we obtain a near optimal fit with a bent DNA substrate taken from a co-crystal structure of the bacterial mismatch repair enzyme MutS (51) (Figure 7A). The presence of a G:T mismatch in this substrate kinks the DNA at a 45° angle (Supplementary Figure S10C), allowing it to interact unencumbered with all four basic patches (Figure 7A). Bp^{CTR} does not show any preference for a substrate containing a G:T mismatch in an exonuclease assay (Supplementary Figure S4C and D).

Mutation of positive residues in patch 1 (R552A/K555A) and patch 3 (R467A/R468A) reduces the DNA cleavage activity of Bp^{CTR} on both supercoiled plasmids and linear lambda DNA (Figure S10D and E), thus indirectly implicating these regions as important for binding. R405A (patch 4) and R559A (patch 1) substitutions do not significantly impair the overall cleavage compared to wildtype (Supplementary Figure S10D and E). We do note, however, accumulation of uncut, supercoiled DNA with every mutant (Supplementary Figure S10E), suggesting each region contributes at least partially to orienting DNA in a manner that promotes endonuclease function. A truncation construct deleting the helical bundle helices $\alpha 1^H$ – $\alpha 5^H$ but retaining the stabilizing $\alpha 6^H$ helix ($\Delta 505$ – 577) severely impairs both nuclease degradation and nicking (Supplementary Figure S10D and E), further highlighting the importance of this region in DNA binding and catalytic function.

The orientation of the modelled substrate would clash with both the Toprim core $\alpha 2$ and $\alpha 3$ helices in their canonical positions and the Insert 1 segments present in topoisomerases and gyrases (Figure 7B), suggesting that OLD nucleases associate with DNA differently than other Toprim proteins. Importantly, this arrangement places one strand directly into the Bp OLD active site cleft with a phosphate residue situated between metal A and metal B (Figure 7C). K562 is 2.8 Å away from the back side of the scissile phosphate, where it would be primed either to stabilize the

charge in the transition state along with metal B and/or protonate the leaving group following cleavage. This favors the proposed catalytic mechanism diagrammed in Figure 7D.

DISCUSSION

Here we have described the structural and biochemical characterization of the Class 2 OLD proteins from *B. pseudomallei* and *X. campestris* pv. *campestris*. Bp and Xcc OLD catalyze metal-dependent nicking and cleavage of DNA substrates *in vitro*. While the N-terminal region containing the ATPase domain is dispensable for these activities, its presence mediates higher ordered oligomerization of Class 2 OLD proteins (Supplementary Figure S1E and F). We suspect that the ATPase domain may act in a regulatory capacity, controlling how and when the catalytic C-terminal region accesses substrates.

The Bp^{CTR} structure elucidates the catalytic machinery of Class 2 OLD proteins. In addition to the canonical invariant glutamate (E400) and DxD aspartates (D455 and D457), we identify E404, T506 and E508 as side chains that play a role metal binding. These residues are absolutely conserved among Class 2 OLD proteins (Supplementary Figure S7) and together coordinate two bound magnesium ions in a geometry that supports two-metal catalysis (Figures 5A and 7C and D). Single point mutations at these sites are tolerated, whereas triple mutant substitutions removing all metal coordination completely abolish processive degradation of substrates (Figure 5C and D). We speculate that a water may be capable substituting as a ligand when a single metal binding residue is mutated, especially since some of the metal contacts in the Bp^{CTR} crystal are water mediated in the absence of substrate.

We also find K562 in the Bp helical domain is critical for efficient catalytic function (Figure 5). K562 is directed toward the putative scissile phosphate in our Bp^{CTR}-DNA bound model (Figure 7C), where it would be poised to stabilize the developing negative charge in the transition state and/or protonate the leaving group. Significant perturbation to one part of the key catalytic machinery (metal A, metal B, or K562) still permits Bp^{CTR} to nick and linearize plasmid DNA; however, processive DNA cleavage is only achieved when the three elements are intact (Figure 5C and D). Truncation of the helical domain ($\Delta 505$ – 577) or simultaneous mutation of both metal sites (2A/2B mutant) impairs both functions (Figure 5C and D, Supplementary Figure S10D and E). Together these results argue that nicking only requires a single metal but full nuclease activity in Class 2 OLD proteins requires proper coordination of two metals and the presence of the conserved lysine. Class 1 OLD proteins are on average ~50 amino acids shorter and diverge from their Class 2 counterparts in portions of the C-terminal region, which prohibits the unambiguous identification of Class 1 catalytic machinery by sequence alignment alone. Structural and biochemical characterization of the Class 1 OLD homolog from *Thermus scotoducts* indicates that the mechanisms and machinery we describe here for nuclease cleavage is conserved (Schiltz and Chappie, *in review*).

The spatial organization of acidic residues in the Bp OLD active site directly mirrors that of RNase M5 maturases

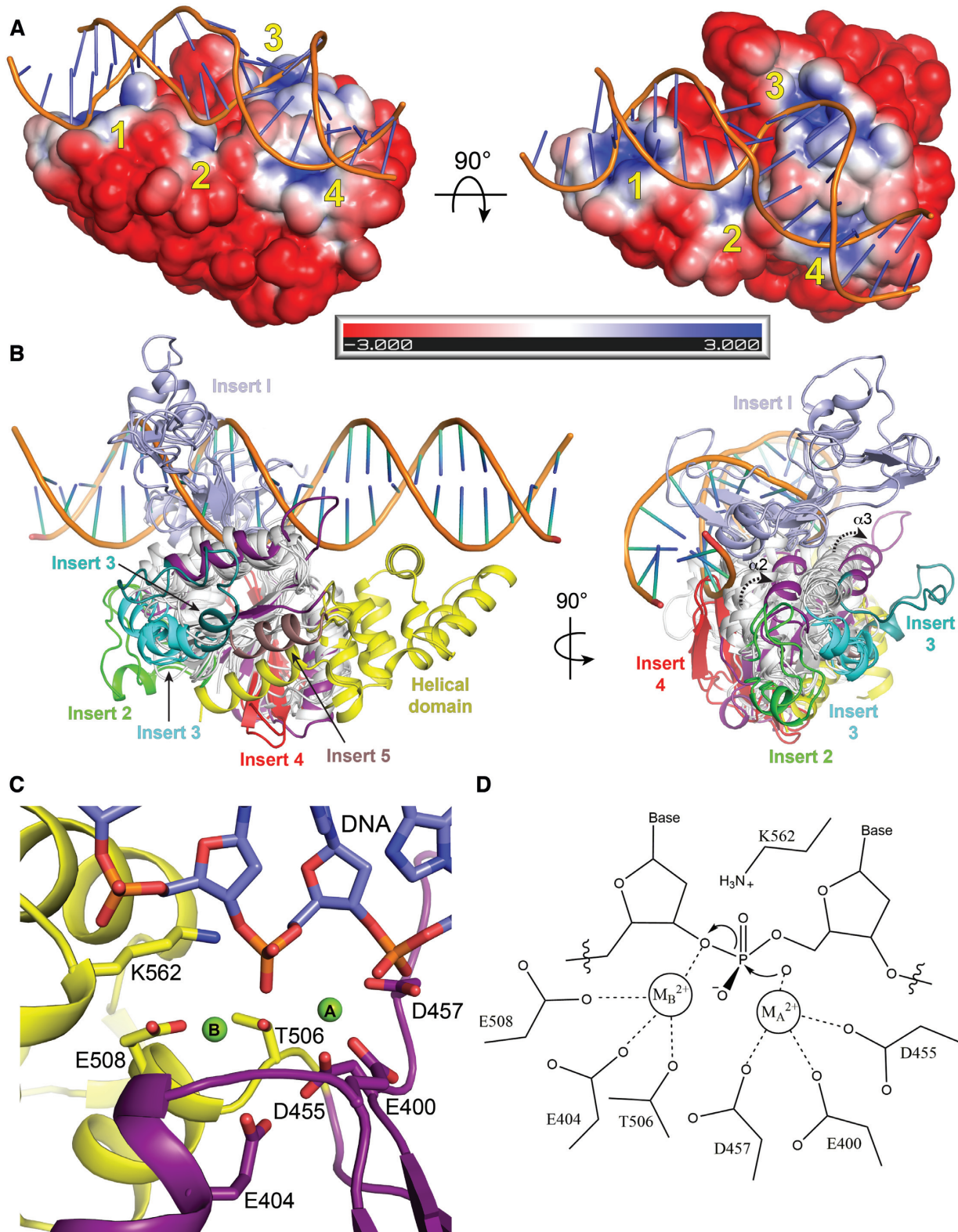


Figure 7. Helical domain orients DNA binding in Class 2 OLD nucleases. (A) Electrostatic surface of Bp^{CTR} with modeled DNA (G:T mismatched substrate taken from PDB: 3K0S). Electrostatic potential calculated with APBS (37). Scale indicates coloring of the potential from $-3K_bT/e_c$ to $+3K_bT/e_c$. The four basic patches around the active site cleft are numbered in yellow. (B) Side and end views of Toprim family superposition with modeled DNA substrate bound to $Bp^{OLD^{CTR}}$. Toprim cores are colored white with Inserts 1–5 individually labeled. Bp Toprim and helical domains are colored purple and yellow respectively. (C) Arrangement of Bp^{CTR} active site with modeled DNA. (D) Proposed two-metal catalysis mechanism for Class 2 OLD nuclease cleavage.

(Figure 6, Supplementary Figure S9). In addition to conserved catalytic residues previously identified through the biochemical characterization of *Bacillus subtilis* RNase M5 (14), our structural comparison with the available *A. aeolicus* RNase M5 structure suggests that a C-terminal glutamate (E96 in *B. subtilis*; E110 in *A. aeolicus*) will also be critical for 5S RNA maturation. Interestingly, *A. aeolicus* RNase M5 appears to be truncated and circularly permuted. Many other homologs including *B. subtilis* contain C-terminal helical extensions (13,14) that could fold into a domain like that observed in Class 2 OLD proteins. DnaG primases also share this conserved arrangement of active site residues (16,17); however, structural constraints imposed by the N-terminal subdomain in the RNA polymerase core prevent the coordination of the Toprim metal B in the same manner. A third metal observed in the *Staphylococcus aureus* DnaG structure (20), which occupies the same position as E508 in Bp OLD, appears to compensate. Importantly, this common active site blueprint is distinct from topoisomerases, gyrases and RecR (Supplementary Figure S9). The overall structural similarity between primases, maturases, and OLD nucleases thus implies a common evolutionary lineage and further segregates the Toprim family into distinct subgroups based on differences in metal coordination, with the distinguishing feature being the presence or absence of additional acidic residues beyond the canonical Toprim glutamate and DxD aspartates.

Our initial biochemistry indicated that Bp^{CTR} and Xcc^{CTR} were more active in Mn²⁺; however, further analysis by ICP-AES analysis revealed that both of the purified constructs preferentially contained bound Ca²⁺ and Mg²⁺ and no Mn²⁺ (Supplementary Table S1). Addition of calcium potentiates Bp^{CTR} activity with magnesium *in vitro*. While calcium typically inhibits most nucleases (2), some enzymes like the *Staphylococcal* nuclease utilize calcium in their active site to cleave DNA (52). Additionally, DNase I is known to be most active in the presence of both magnesium and calcium (53). In the case of DNase I, however, magnesium occupies the active site while calcium binds to other regions of the structure to act as an allosteric enhancer (54). Whether calcium plays a direct role in the active site or modulates activity indirectly, possibly by stabilizing the protein or enhancing DNA binding, remains to be determined. Importantly, the Class 1 OLD homolog from *Thermus scotoductus* exhibits the same general affinity for calcium and magnesium and shows the same stimulatory response (Schiltz and Chappie, in review). This implies that utilization of calcium and magnesium is conserved and functionally relevant among all OLD homologs.

Computational modeling shows that a bent DNA substrate engages all four basic patches on the surface Bp^{CTR} while B form DNA would sterically clash with portions of the Toprim domain (Figure 7A and Supplementary Figure S10B). Mutations in patches 1 and 3 reduce Bp^{CTR} activity on both substrates (Supplementary Figure S10D and E), indirectly supporting a role for these regions in DNA binding. These patches flank the active site cleft and in our model anchor the DNA duplex such that one strand is positioned in the active site with a phosphate situated directly between the two bound magnesium ions (Figure 7). The catalytic K562 sidechain resides in patch 2 and engages the substrate at one

end of this cleft. Although mutation of R405 in patch 4 does not significantly alter nuclease activity, we note an observable accumulation of the uncut, supercoiled substrate compared to wildtype (Supplementary Figure S10E). This implies patch 4 partially contributes to orienting DNA in a manner that promotes endonuclease function.

The orientation of DNA suggested by our model differs significantly from how other Toprim proteins engage their substrates. Importantly, the structural constraints of this arrangement explain (i) the lack of an insert 1 in OLD Toprim domains, (ii) the significant shift in the positions of the $\alpha 2$ and $\alpha 3$ Toprim core helices in Bp^{CTR} and Xcc^{CTR} and (iii) the position of the helical domain on the opposite side of the core Toprim fold. In the absence of DNA bound structure, we cannot rule out that substrate binding induces further structural changes in the OLD CTR, including those that would permit the unhindered association with an extended B form DNA duplex. Conformational rearrangements could also be coupled to ATP hydrolysis in the full-length protein.

Our binding model, however, does not preclude Bp OLD from also binding DNA ends. Here the terminal phosphate would become the scissile phosphate. This arrangement is equally compatible with the catalytic machinery and indeed Bp^{CTR} exhibits 5'-3' exonuclease activity (Supplementary Figure S4). P2 OLD exhibits exonuclease activity *in vitro* (28) and Bp OLD readily degrades linear lambda DNA in the presence of Mn²⁺ or Mg²⁺ and Ca²⁺ as detailed above. Bp and Xcc OLD also can nick and cleave circular plasmids, suggesting a robust endonuclease activity. OLD nucleases thus appear to act as either an endo- or exonuclease depending on the substrate presented (Supplementary Figure S4E). The Mre11 nuclease, which functions in double strand break repair and processing, displays a similar duality: it functions as a 3'-5' exonuclease on double strand DNA and an endonuclease on single strand DNA at protruding 3'- and 5'-ends and 3' branches (55-57). Further biochemical characterization will be necessary to determine how these different modes of cleavage contribute to OLD function *in vivo*.

While the role of P2 OLD in bacteriophage lambda interference is well documented (23), little is known about the function of other OLD homologs *in vivo*. Our bioinformatics data indicate that OLD proteins are widely distributed across bacteria, archaea, and viral genomes. The presence of *old* genes in species-specific operons and on mobile elements suggest they confer a functional advantage. We speculate that these proteins may play a novel role in DNA repair and/or replication based the specific association of UvrD/PcrA/Rep helicase with Class 2 OLD proteins. Future genetic experiments will be necessary to validate this hypothesis and define the biological roles of OLD nucleases more explicitly.

DATA AVAILABILITY

The atomic coordinates and structure factors for the Xcc^{CTR} Hg, Pt, and I derivatives are deposited in the Protein Data Bank with accession numbers 6NJW, 6NJX and 6NJV respectively. The atomic coordinates and structure factors

of the Bp^{CTR} structure are deposited in the Protein Data-bank with the accession number 6NK8.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Richard Cerione and Dr Holger Sondermann for critical reading of the manuscript, NE-CAT beamline staff for assistance with remote X-ray data collection, Angela Kruse and Eric Craft assistance with ICP-AES data collection and analysis, Dr Eric Alani and Dr Ailong Ke for advice about DNA binding and cleavage experiments, Dr John Helmann and Dr Joseph Peters for insightful discussions, and Dr Holger Sondermann for use of his SEC-MALS instrument.

FUNDING

J.S.C. is a Meinig Family Investigator in the Life Sciences. This work is based upon research conducted at the Macromolecular Diffraction facility at the Cornell High Energy Synchrotron Source (MacCHESS) beamlines (A1 and F1) and the Northeastern Collaborative Access Team (NE-CAT) beamlines (24-ID-C and 24-ID-E) under general user proposals GUP-51113 and GUP-41829 (PI: J.S.C.). CHESS is supported by an award from the National Science Foundation [DMR-1332208] and MacCHESS is supported by a grant from the National Institutes of Health [GM-103485]. NE-CAT beamlines are funded by the National Institute of General Medical Sciences from the National Institutes of Health [P30 GM124165]. The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant [S10 RR029205]. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Funding for open access charge: Laboratory startup funds. *Conflict of interest statement.* None declared.

REFERENCES

- Palermo, G., Cavalli, A., Klein, M.L., Alfonso-Prieto, M., Dal Peraro, M. and De Vivo, M. (2015) Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.*, **48**, 220–228.
- Yang, W., Lee, J.Y. and Nowotny, M. (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
- Galburt, E.A. and Stoddard, B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
- Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
- Cech, T.R., Zaug, A.J. and Grabowski, P.J. (1981) In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell*, **27**, 487–496.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
- Sowadski, J.M., Handschumacher, M.D., Murthy, H.M., Foster, B.A. and Wyckoff, H.W. (1985) Refined structure of alkaline phosphatase from *Escherichia coli* at 2.8 Å resolution. *J. Mol. Biol.*, **186**, 417–433.
- Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. and Steitz, T.A. (1988) Cocystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8924–8928.
- Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 6498–6502.
- Yang, W. (2008) An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.*, **15**, 1228–1231.
- Taylor, G.K. and Stoddard, B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189–5200.
- Gao, Y. and Yang, W. (2016) Capture of a third Mg²⁺ is essential for catalyzing DNA synthesis. *Science*, **352**, 1334–1337.
- Aravind, L., Leipe, D.D. and Koonin, E.V. (1998) Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.*, **26**, 4205–4213.
- Allemand, F., Mathy, N., Brechemier-Baey, D. and Condon, C. (2005) The 5S rRNA maturase, ribonuclease M5, is a Toprim domain family member. *Nucleic Acids Res.*, **33**, 4368–4376.
- Sissi, C. and Palumbo, M. (2009) Effects of magnesium and related divalent metal ions in topoisomerase structure and function. *Nucleic Acids Res.*, **37**, 702–711.
- Keck, J.L., Roche, D.D., Lynch, A.S. and Berger, J.M. (2000) Structure of the RNA polymerase domain of *E. coli* primase. *Science*, **287**, 2482–2486.
- Podobnik, M., McInerney, P., O'Donnell, M. and Kuriyan, J. (2000) A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases. *J. Mol. Biol.*, **300**, 353–362.
- Corbett, K.D. and Berger, J.M. (2004) Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 95–118.
- Yang, W. (2010) Topoisomerases and site-specific recombinases: similarities in structure and mechanism. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 520–534.
- Rymer, R.U., Solorio, F.A., Tehranchi, A.K., Chu, C., Corn, J.E., Keck, J.L., Wang, J.D. and Berger, J.M. (2012) Binding mechanism of metal-NTP substrates and stringent-response alarmones to bacterial DnaG-type primases. *Structure*, **20**, 1478–1489.
- Usenik, A., Renko, M., Mihelič, M., Lindič, N., Boršček, J., Perdih, A., Pretnar, G., Müller, U. and Turk, D. (2017) The CWB2 cell Wall-Anchoring module is revealed by the crystal structures of the *clostridium difficile* cell wall proteins Cwp8 and Cwp6. *Structure*, **25**, 514–521.
- Koonin, E.V. and Gorbalenya, A.E. (1992) The superfamily of UvrA-related ATPases includes three more subunits of putative ATP-dependent nucleases. *Protein Seq. Data Anal.*, **5**, 43–45.
- Lindahl, G., Sironi, G., Bialy, H. and Calendar, R. (1970) Bacteriophage lambda: abortive infection of bacteria lysogenic for phage P2. *Proc. Natl. Acad. Sci. U.S.A.*, **66**, 587–594.
- Sironi, G. (1969) Mutants of *Escherichia coli* unable to be lysogenized by the temperate bacteriophage P2. *Virology*, **37**, 163–176.
- Ghisotti, D., Zangrossi, S. and Sironi, G. (1979) X-ray sensitivity of *Escherichia coli* lysogenic for bacteriophage P2. *Mol. Gen. Genet.*, **169**, 229–35.
- Sironi, G., Bialy, H., Lozeron, H.A. and Calendar, R. (1971) Bacteriophage P2: interaction with phage lambda and with recombination-deficient bacteria. *Virology*, **46**, 387–396.
- Bregere, F. (1974) Bacteriophage P2-lambda interference: inhibition of protein synthesis involves transfer RNA inactivation. *J. Mol. Biol.*, **90**, 459–467.
- Myung, H. and Calendar, R. (1995) The old exonuclease of bacteriophage P2. *J. Bacteriol.*, **177**, 497–501.
- Khatiwara, A., Jiang, T., Sung, S.S., Dawoud, T., Kim, J.N., Bhattacharya, D., Kim, H.B., Ricke, S.C. and Kwon, Y.M. (2012) Genome scanning for conditionally essential genes in *Salmonella enterica* Serotype Typhimurium. *Appl. Environ. Microbiol.*, **78**, 3098–3107.
- Hendrickson, W.A. (2014) Anomalous diffraction in crystallographic phase evaluation. *Q. Rev. Biophys.*, **47**, 49–93.

31. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* **D66**, 133–144.
32. Evans, P.R. (2006) Scaling and assessment of data quality. *Acta Cryst.*, **D62**, 72–82.
33. Sheldrick, G.M. (2008) A short history of SHELX. *Acta Cryst.*, **A64**, 112–122.
34. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.*, **D66**, 213–221.
35. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta. Crystallogr.*, **D66**, 486–501.
36. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Cryst.*, **40**, 658–674.
37. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K. *et al.* (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci.*, **27**, 112–128.
38. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, W344–W350.
39. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
40. Chen, I.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M. *et al.* (2017) IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.*, **45**, D507–D516.
41. Rogozin, I.B., Makarova, K.S., Murvai, J., Czabarka, E., Wolf, Y.I., Tatusov, R.L., Szekely, L.A. and Koonin, E.V. (2002) Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res.*, **30**, 2212–2223.
42. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50.
43. Raney, K.D., Byrd, A.K. and Aarattuthodiyil, S. (2013) Structure and Mechanisms of SF1 DNA Helicases. *Adv. Exp. Med. Biol.* **767**, 17–46.
44. Yang, W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.
45. Bolann, B.J., Rahil-Khazen, R., Henriksen, H., Isrenn, R. and Ulvik, R.J. (2007) Evaluation of methods for trace-element determination with emphasis on their usability in the clinical routine laboratory. *Scand. J. Clin. Lab. Invest.* **67**, 353–366.
46. Holm, L. and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
47. McGeehan, J.E., Streeter, S.D., Thresh, S.J., Ball, N., Ravelli, R.B. and Kneale, G.G. (2008) Structural analysis of the genetic switch that regulates the expression of restriction-modification genes. *Nucleic Acids Res.* **36**, 4778–4787.
48. McGeehan, J.E., Streeter, S.D., Papapanagiotou, I., Fox, G.C. and Kneale, G.G. (2005) High-resolution crystal structure of the restriction-modification controller protein C.AhdI from *Aeromonas hydrophila*. *J. Mol. Biol.*, **346**, 689–701.
49. Ball, N.J., McGeehan, J.E., Streeter, S.D., Thresh, S.J. and Kneale, G.G. (2012) The structural basis of differential DNA sequence recognition by restriction-modification controller proteins. *Nucleic Acids Res.*, **40**, 10532–10542.
50. Schmidt, B.H., Burgin, A.B., Dewese, J.E., Osheroff, N. and Berger, J.M. (2010) A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature*, **465**, 641–644.
51. Lebbink, J.H., Fish, A., Reumer, A., Natrajan, G., Winterwerp, H.H. and Sixma, T.K. (2010) Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS. *J. Biol. Chem.*, **285**, 13131–13141.
52. Loll, P.J. and Lattman, E.E. (1989) The crystal structure of the ternary complex of staphylococcal nuclease, Ca²⁺, and the inhibitor pDtp, refined at 1.65 Å. *Proteins*, **5**, 183–201.
53. Pan, C.Q. and Lazarus, R.A. (1999) Ca²⁺-dependent activity of human DNase I and its hyperactive variants. *Protein Sci.*, **8**, 1780–1788.
54. Gueroult, M., Picot, D., Abi-Ghanem, J., Hartman, B. and Baaden, M. (2010) How cations can assist DNase I in DNA binding and hydrolysis. *PLoS Comput. Biol.*, **6**:e1001000.
55. Paull, T.T. and Gellert, M. (1998) The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol. Cell*, **1**, 969–979.
56. Paull, T.T. and Gellert, M. (1999) Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.*, **13**, 1276–1288.
57. Trujillo, K.M. and Sung, P. (2001) DNA structure-specific nuclease activities in the *Saccharomyces cerevisiae* Rad50* Mre11 complex. *J. Biol. Chem.*, **276**, 35458–35464.