STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE OLD FAMILY NUCLEASES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Carl Schiltz

May 2019

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE OLD FAMILY NUCLEASES

Carl Schiltz, Ph. D.

Cornell University 2019

Overcoming **l**ysogenization **d**efect (OLD) proteins were first discovered in bacteriophage P2, where presence of the *old* gene plays a role in restricting co-infection of lambda phage in the *E. coli* host. These enzymes are well conserved and present in a wide array of bacteria, archaea, and viruses, yet their function and mechanism in other contexts remain a mystery.

Structurally, these enzymes contain an N-terminal ATPase domain and a C-terminal Toprim domain common amongst topoisomerases, DnaG primases, gyrases, RecR recombination proteins, and 5S rRNA maturases. The Toprim-containing C-teriminal region of the OLD proteins is sufficient for nuclease activity and acts on both circular and linear DNA. Crystal structures of the C-terminal region of OLD proteins from *Burhkolderia pseudomallei* and *Xanthomonas campestris* pv campestris revealed that OLD proteins possess a topologically unique Toprim domain and utilize a two-metal catalytic mechanism to hydrolyze DNA.

In order to understand the role of the N-terminal ATPase domain and the overall architecture of OLD proteins, I purified and solved the crystal structure of a

Class 1 full-length OLD protein from *Thermus scotoductus* (Ts). The ATPase domain of TsOLD is topologically similar to the ABC-type ATPases, whose members include the ABC transporters, SMC proteins, and the DNA repair protein Rad50. The structure highlights unique active site resides used to bind and hydrolyze ATP that distinguish the OLD proteins from other ABC ATPases. Together these data provide novel insights into the mechanism and function of OLD family nucleases.

BIOGRAPHICAL SKETCH

Carl Schiltz attended Gustavus Adolphus College in St.Peter MN where he received his Bachelor of Arts degree in Biochemistry in the spring of 2010. He enrolled in the Biochemistry, Molecular, and Cell Biology program at Cornell in the fall of 2010. He joined the lab of Joshua Chappie in May of 2011 and remained so until his commencement in 2019.

ACKNOWLEDGMENTS

This work and the data it contains could have been possible without the strong scientific guidance and patience of my mentor Joshua Chappie. I am also thankful for the constant moral support and friendship of my lab mates Chris Hosford and Myfanwy Adams. Finally I would like to thank my thesis committee, John Helmann, Ailong Ke, and Robert Weiss, for their continued support and advice.

TABLE OF CONTENTS

Chapter 1. Introduction

Bacteriophage and Phage Exclusion

Lambda phage was discovered in 1951 by Esther Lederberg (Lederberg EM) and further characterized in 1953 (Lederberg EM, Lederberg J). She observed that upon irradiation with UV light, plaques formed on her plates of E. coli. This finding highlighted the properties of the lytic and lysogenic lifecycles of the temperate bacteriophage. In the case of lambda phage, it must decide when to remain in the host genome and when to enter into the lytic cycle. Lambda and many other temperate phages rely on a repressor system to accomplish this. In the case of lambda phage, the CI protein binds to two operator sequences that allow for repression of lytic genes in one polarity, while still supporting expression of the CI repressor along with a few other nonessential genes (Casjens SR, 2015). Two of these non-essential genes are *rexA* and *rexB*. These genes have been shown to mediate a phenomenon called “phage exclusion”. It was observed in 1955 by Seymour Benzer that certain mutants of T4 phage were unable to plate on lambda phage lysogens. He characterized these mutants as rapid lysis mutants or rII. The distinct plaque forming characteristics and inability to plate on lambda made the T4rII mutants uniquely valuable for genetic experiment. Benzer was able to utilize the tools he developed to define our current understanding of a gene as a linear stretch of DNA which he called a cistron (Benzer S, 1955). The rIIB mutant was shortly after used to determine the triplet nature of the genetic code (Crick FH, 1961). This was one of the earliest observations of the “phage exclusion” phenotype.

Just as bacteria have an evolutionary pressure to avoid being infected by virus, prophages face an identical pressure. Should a lysogen be infected by a virulent phage such as T4, it is quite possible that the virulent phage could replicate, assemble, and lyse the cell before the lysogen has a change to enter into the lytic cycle and generate its own viral particles. For this reason, systems have been developed by temperate phages to protect themselves from such an insult. The aforementioned RexA and RexB proteins are one such system that is at least somewhat effective mitigating superinfection from a number of virulent phage. Though the exact mechanism is not completely understood, current models suggest that RexA and RexB work together to induce “altruistic cell death” of the host prevent superinfection (Shinedling S, 1987).

Bacteriophage P2, another temperate phage, has at least two other means by which to hobble competing bacteriophage. The exclusion systems of P2 seem to be aimed at disrupting the replication of the insulting virus. Like *rexAB* from lambda the *tin* gene of P2 is capable of excluding T4 phage. Work on this system has identified that Tin interferes with T-even bacteriophage protein Gp32, a ssDNA binding protein necessary for stabilizing the viral replication fork (Mosig G, 1997). The other phage exclusion gene, *old*, lies directly upstream of *tin* (Mosig G, 1997).

The *old* phenotype was originally observed by Sironi while analyzing *lyd* mutants of *E. coli* (Sironi G, 1969). The experiments were undertaken after an observation by L Bertani that a mutant of P2 was capable of lysogenizing *E.*

coli, but not *Shigella*, which is highly similar to *E. coli* and would otherwise support lysogeny (Bertani LE, 1960). The experiments were predicated on the idea that both host and viral factors were important for integration and replication of temperate bacteriophage. Mutants were generated in *E. coli* and tested for infection by phage P2. These mutants were named *lyd* for *lysogenization defective*. In addition to being unsupportive of P2 lysogenization, the *lyd* mutants were also highly sensitive to UV and showed very low rates of recombination in both bacterial crosses as well as P1 transduction (Sironi G, 1969). The *lyd* mutants were hypothesized to be related to the recombination-deficient (*rec*) mutants that had been recently characterized (Clark AJ, 1965 and Willets NS, 1969). This was confirmed by Lindahl et al. who determined that the *lyd* phenotype was a consequence of mutations in the *recB* and *recC* genes, but not mutations in *recA* (Lindahl G, 1970). Spontaneous mutants of P2 phage that were able to lysogenize the *lyd* mutants were deemed *old – overcoming lysogenization defect* (Sironi G, 1969). In addition to preventing lambda phage from infecting *recBC,* the *old* gene was also the cause of the much earlier observation that lambda phage cannot plate on P2 lysogens (Fredericq P, 1953). Lindahl et al. determined that the *old* gene was the cause of this phenotype and that at least two genes from lambda itself are necessary for exclusion (Lindahl G, 1970). At this point it became apparent that there was some significant amount of interplay between P2 *old,* lambda phage replication, and the recombination machinery of *E. coli.*

Recombination in *E. coli* and lambda phage

RecA

Bacterial genetic recombination has been studied most extensively in *E. coli.* The initial identification of the genes responsible for recombination occurred in the mid-1960s and 1970s (Clark AJ, 1965; Howard-Flanders P, 1966; Emmerson PT, 1968; Willets NS, 1969; Zeig J, 1978). The first locus to be identified and extensively characterized was deemed *recA* (Willets NS, 1969). RecA is the center of most recombination and homologous repair in *E. coli* as it mediates homology searching and strand exchange in template-directed double-strand break repair and homologous recombination. It is a structural and functional homolog of the eukaryotic protein Rad51 and the archaeal protein RadA (Cox MM, 2000; Krogh BO 2004; Steitz EM, 1998). RecA is a P-loop ATPase that binds to ssDNA. Binding of RecA causes a global stretching of the DNA molecule, however, at a local level (i.e. 3 bases) the ssDNA maintains the rise and pitch of B-form DNA (Chen Z, 2008). This allows the ssDNA/RecA complex to perform a highly accurate homology search that favors only Watson-Crick base pairing (Savir Y, 2010). RecA-mediated homology search eventually results in the formation of a D-loop, wherein the original ssDNA filament has invaded the dsDNA of a homologous sequence. Following this, replicative DNA polymerases can use the terminus of the broken DNA strand to commence template-directed repair (Bell J, 2016).

RecBCD

In *E. coli* the enzyme complex RecBCD is the premier initiator of double-strand break repair. Double strand breaks DSB represent a particularly dangerous form of DNA damage that can easily lead to cell death. The ability of RecA to mediate homologous recombination is essential for accurate repair of DBS; however, there is also a need to generate an appropriate substrate for RecA as well as to stimulate the formation of the RecA filament.

The RecBCD complex (also known as Exonuclease V) is a heterotrimeric complex that has both helicase and nuclease activity (Smith GR, 1988; Yu M, 1998; Dillingham MS, 2003). The complex preferentially binds blunt DNA or DNA with no more than 25nt between the 5' and 3' ends (Taylor AF, 1985). This makes it optimal for detecting and acting on DBS that have yet to undergo extensive degradation; however, there are mechanism in *E*. *coli* to process most all DSB products into targets for RecBCD; however, these mechanisms will be discussed later in this work.

Upon recognition of its DNA substrate the enzyme translocates in an ATP-dependent manner along both strands of DNA. RecB moves with 3'-5' polarity while RecD moves 5'-3' (Boehmer PE, 1992; Dillingham MS, 2003; Wu CG, 2010). The RecB subunit has a single active site that nicks both strands of DNA as the complex translocates along dsDNA, generating oligonucleotides. Initially the strand with 3'-5' polarity is more rapidly degraded (Taylor AF, 1995 and Anderson DG, 1997). Upon recognition of a specific DNA sequence called the crossover initiation hotspot (chi), the RecBCD

complex undergoes a change in behavior. Chi sites (5'-GCTGGTGG-3') occur approximately every four- to five-thousand base pairs. At this point the 3'-5' helicase activity of the RecB subunit becomes dominant in translocation and digestion of the 5'-3' strand is favored. This leaves a 3' ssDNA tail (Taylor AF, 1992; Taylor AF, 1995; Anderson DG, 1997).

The mechanism that causes the drastic change in RecBCD function is not completely understood, though the major model posits upon recognition of the chi sequence by RecC, the helicase activity of RecD is eliminated which subsequently causes RecB to preferentially cut the ssDNA stand with 5'-3' polarity (Amundsen SK, 2007). It has also been suggested that the nuclease domain of RecB undergoes a conformational change and binds to RecC near where the ssDNA of 5'-3' polarity exits, conferring strand specificity (Taylor AF, 2014). There is a flexible linker that connects the nuclease and helicase domains of RecB that could be responsible for allowing the proposed conformational changes. Significant changes in linker length or flexibility strongly inhibited chi-dependent activity of the RecBCD complex (Amundsen SK, 2019).

In addition to generating the 3' ssDNA, RecBCD also facilitates loading of RecA onto the ssDNA. Mutations that abolish the RecA loading function abolish recombination in vivo (Amundsen SK, 1990; Anderson DG, 1997; Arnold DA, 2000). It has been proposed that the same conformational change that alters the nuclease activity of RecB also exposes a site for RecA interaction. (Taylor AF, 2014 and Amundsen SK, 2019). Refer to Figure 1 for

cartoon illustration of RecBCD-mediated repair.

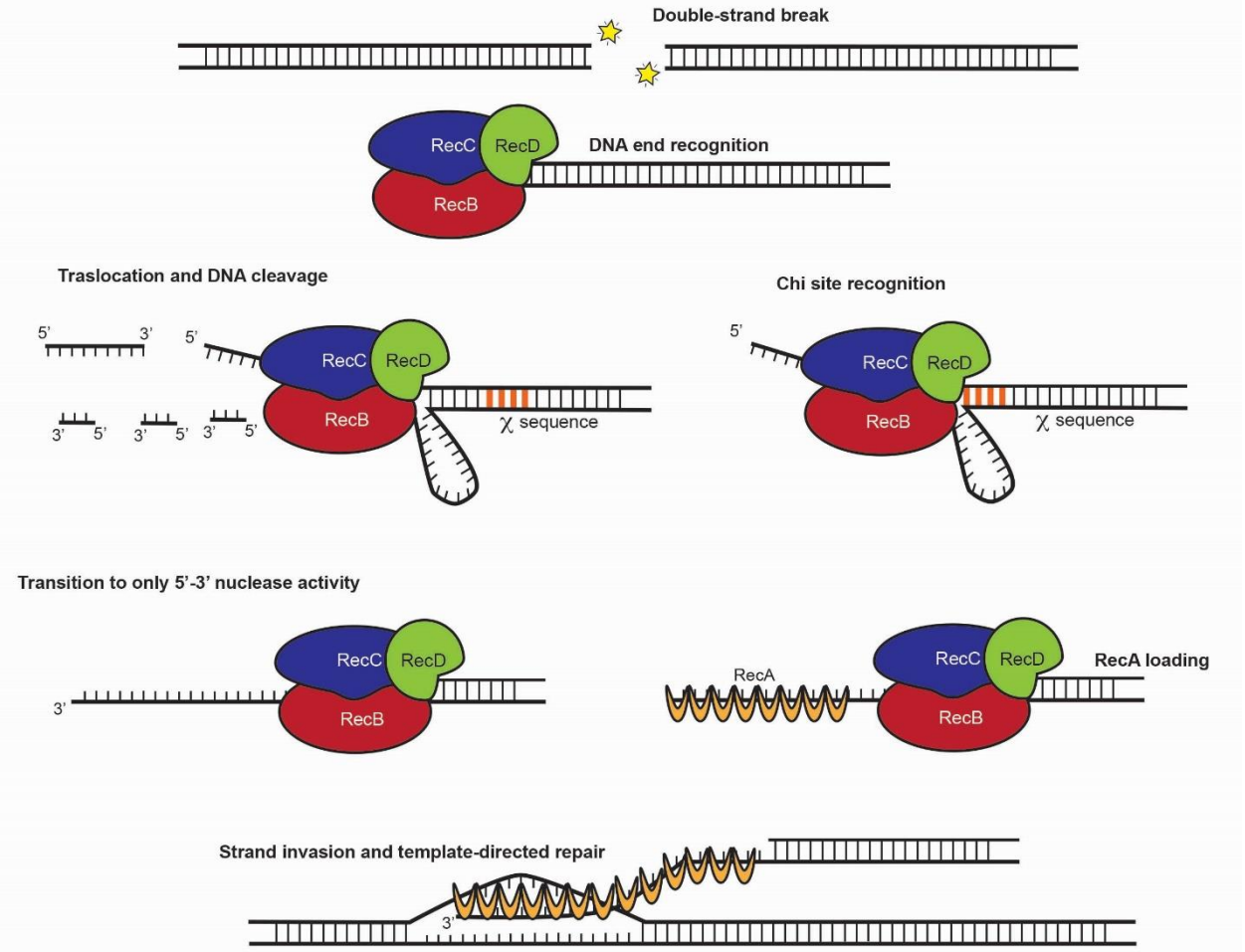


Figure 1. Schematic of RecBCD recombination.

RecFOR

The other major recombination pathway in *E. coli* is the RecFOR system. While some organisms such as *Deinococcus radiodurans* utilize RecFOR for double-strand break repair, its role in *E. coli* is generally relegated to single-strand gap repair (Bentchikou E, 2010). Collapse of a replication fork due to a genetic lesion that blocks progression of DNA polymerase can produce long stretches of ssDNA (Kowalczykowski SC, 2000). The RecF pathway acts via recombination to repair these gaps (Wang TV, 1984). Repair by the RecFOR

pathway begins with the 5'-3' exonuclease RecJ, which begins to resect ssDNA from the point of damage. The helicase RecQ stimulates this activity and also helps to resolve any topological impediments that may have resulted from replication fork collapse. Single-strand binding protein (SSBP) accumulates on the exposed strand of DNA, guarding against additional DNA damage and self-complementation, but also preventing repair by homologous recombination. RecF recognizes junctions between single- and double-stranded DNA with a 5' terminus, just as is generated by RecJ (Morimatsu K, 2003). RecF forms a complex with RecR and RecO at the DNA junctions, where RecR and RecO displace the SSBP and begin loading RecA (Handa N, 2009). RecOR is also capable of acting without RecF. The complex instead interacts with the C-terminus of SSBP, bypassing the need for a specific DNA architecture for recombination (Sakai A, 2009). Refer to Figure 2 for a cartoon representation of the RecFOR and RecOR pathways.

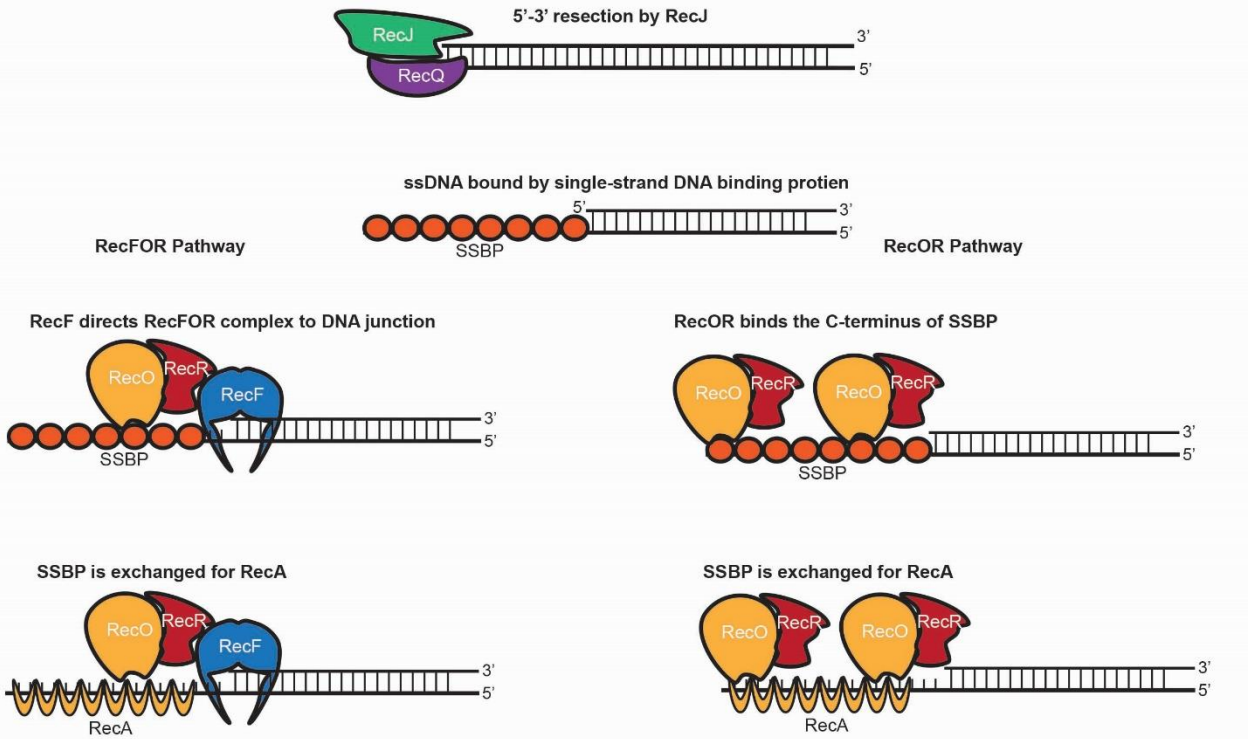


Figure 2. DNA break repair through the RecFOR and RecOR pathways.

Suppressors of recBC-

Given that the RecFOR pathway both generates ssDNA and stimulates the formation of a ssDNA-RecA complex, it would seem that RecFOR could effectively repair DSBs in the absence of RecBCD. While this reasoning holds true for organisms like *D. radiodurans*, mutation of the *recBC* in *E.coli* causes severe defects in DNA damage repair and recombination (Clark 1965). This "favoritism" for the RecBCD pathway of DSB repair is a product of two enzymes: ExoI and SbcCD. The *sbcB* gene was originally identified as a suppressor of the *recBC*- phenotype (Templin A, 1972). It was discovered that mutation of *sbcB* was capable of suppressing both the DNA repair and recombination deficiency of the *recBC*- cells. Moreover, mutation of *sbcB* eliminated the activity of Exonuclease I, a 3'-5' exonuclease (Kushner SR,

1972). Similarly, mutations in the *sbcC* and *sbcD* genes were also capable of suppressing *recBC*- (Lloyd RG, 1985 and Gibson FP, 1992). The SbcCD complex is also a 3'-5' nuclease that is a structural homolog of the Mre11-Rad50 DNA repair protein present in archaea and eukaryotes (Connelly JC, 1998). Given the nuclease activity of these two proteins, it becomes apparent that the preponderance of RecBCD-mediated repair of DSB is through disfavoring the RecFOR pathway. As stated above, RecF specifically recognizes a ssDNA-dsDNA junction with a 5' terminus. Through the actions of ExoI and SbcCD, the vast majority of processed DNA ends are instead blunt or have DNA junctions with a 3' terminus. Figure 3 diagrams the functions of Sbc proteins in favoring RecBCD recombination.

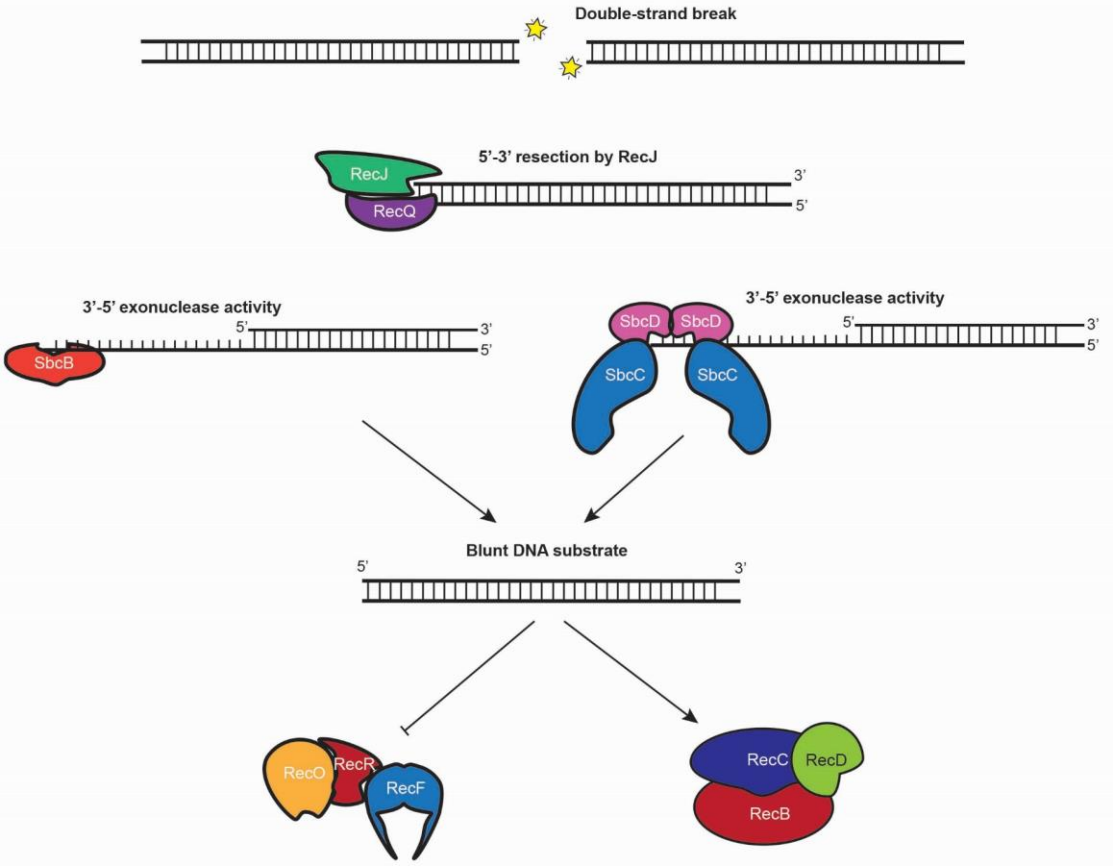


Figure 3. Actions of the Sbc proteins process DNA ends to favor recombination repair by RecBCD rather than RecFOR.

RecET

While it acts outside of the traditional *E. coli* DNA repair machinery, it is necessary to note *sbcA* as well. As described above, the RecA protein is absolutely necessary for recombination mediated by the RecBCD and RecFOR pathways; however, RecA is not the sole protein capable of catalyzing homology searching and strand exchange. Many strains of *E. coli* carry the defective prophage Rac. Much of its genome is intact, however, it no longer possesses the ability to replicate itself or generate capsid proteins. The

initial mapping of the *sbcA* gene determined that it was located in the Rac genome (Lloyd RG, 1974). Mutation of the *sbcA* gene allows for expression of the proteins RecE and RecT (Gillen JR, 1981). RecE is also known as Exonuclease VIII. It is a 5'-3' exonuclease that has high affinity for double-stranded DNA over single-stranded (Kushner SR, 1974). RecT plays the role of the recombination mediator, binding to the ssDNA generated by RecE and facilitating the homology search and strand invasion (Kolodner R, 1994). See Figure 4 for a cartoon diagram of these events.

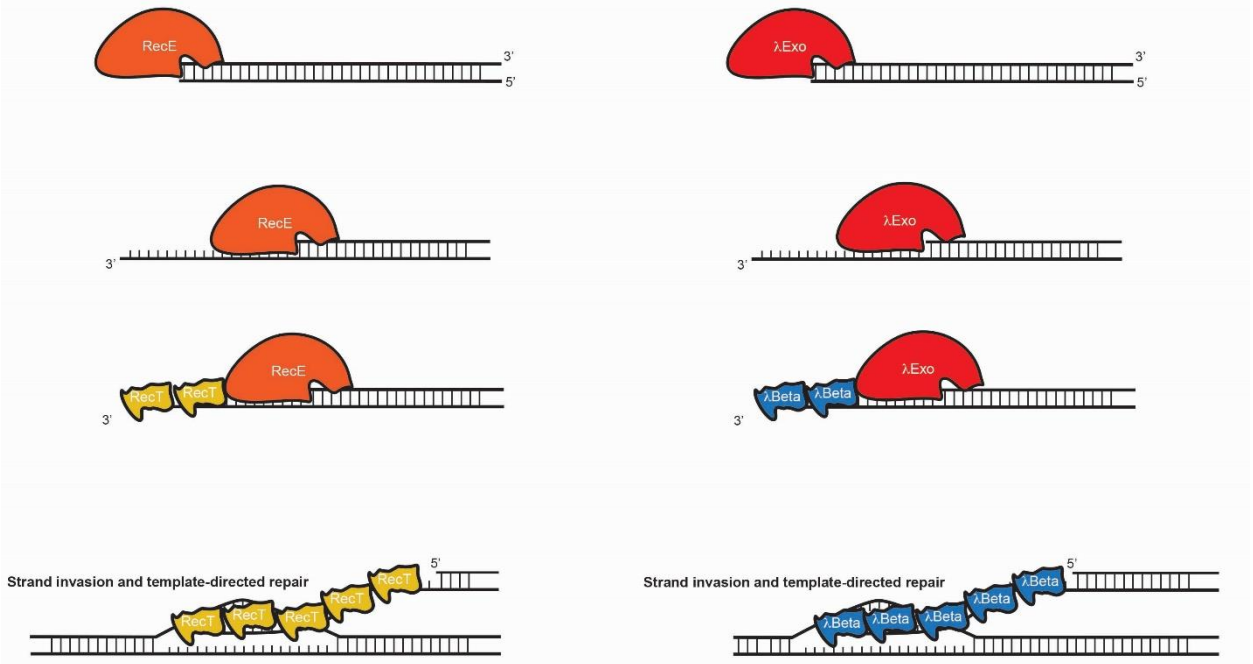


Figure 4. Diagram of both the Rac prophage RecET recombination pathway as well as the lambda *red* recombination pathway.

Lambda red recombination

Much like the RecET system of the defective phage Rac, lambda also encodes for a very effective recombination system. Lambda is also capable of undergoing recombination in the absence of RecA. Mutants of lambda unable

to undergo recombination in *recA- E. coli* were deemed *red-* for *recombination deficient* (Signer ER, 1968). The first protein to act in the Red system is unsurprisingly an exonuclease. The lambda exonuclease (*exo*) is another 5'-3' exonuclease that leaves and extended strand of ssDNA with 3' polarity (Little JW, 1967). The annealing protein in this case is called Beta (*bet*). It binds to ssDNA to catalyze homology searching and strand exchange (Li Z, 1998). The Exo and Beta proteins are sufficient to catalyze recombination; however, an additional gene *gam* is expressed with them. The Gam protein binds to and inhibits the RecBCD enzyme. Gam structurally resembles DNA and is able to insert into the DNA binding groove of RecBC (Murphy KC, 2007). The expression and activity of Gam is important for productive replication of the lambda phage genome.

Replication of lambda phage DNA

While tomes have been written about the entire lambda lifecycle, this discussion will be limited to DNA replication during the lytic cycle, as that is pertinent to the discussion of the *old* phenotype. The lambda-encoded genes O and P are essential to being replication. The four dimers of the O protein bind to four sequence motifs in the lambda origin, which forms the "O-some". Similarly to DnaA in *E. coli* replication, binding of O to the origin sequence helps to melt the DNA to promote replication initiation (Zahn K, 1985). The other essential replication protein P binds to the *E. coli* replicative helicase DnaB, which is then recruited to the O-some (LeBowitz JH, 1984). Once there,

DnaB must be released from P to become active. This is accomplished through the host chaperones DnaK, DnaJ, and GrpE. Once freed from P, DnaB recruits the other members of the replication initiation complex and DNA synthesis can begin (Konieczny I, 1995).

The initial rounds of replication are called *theta* or bidirectional replication – two replication forks move from the origin in opposite direction. Theta replication results in closed circular chromosomes. At some point, however, replication of the lambda genome switches to *sigma*, or rolling circle replication. During this mode of replication DNA polymerase uses the same template; however, the elongation is unidirectional, resulting in a very long piece of ssDNA containing several copies of the genome. Lagging strand synthesis on the ssDNA generates the full dsDNA genome. While important for lambda replication, the switch from *theta* to *sigma* remains a mystery (Wegrzyn G, 2005). The long DNA concatemers of DNA generated from *sigma* replication are more easily packaged into the phage heads (Enquist LW, 1978). However, the free DNA ends of the long concatemers are readily recognized by the RecBCD complex. While chi sites protect the *E. coli* genome from reckless DNA degradation, the lack of any chi sequences in the lambda genome leave it vulnerable to complete digestion. Therefore, the Gam protein mentioned earlier becomes a very important for inhibiting the action of RecBCD and maintaining the integrity of the nascent lambda genomes (Murphy KC, 2007). Figure 5 shows a cartoon of lambda DNA replication.

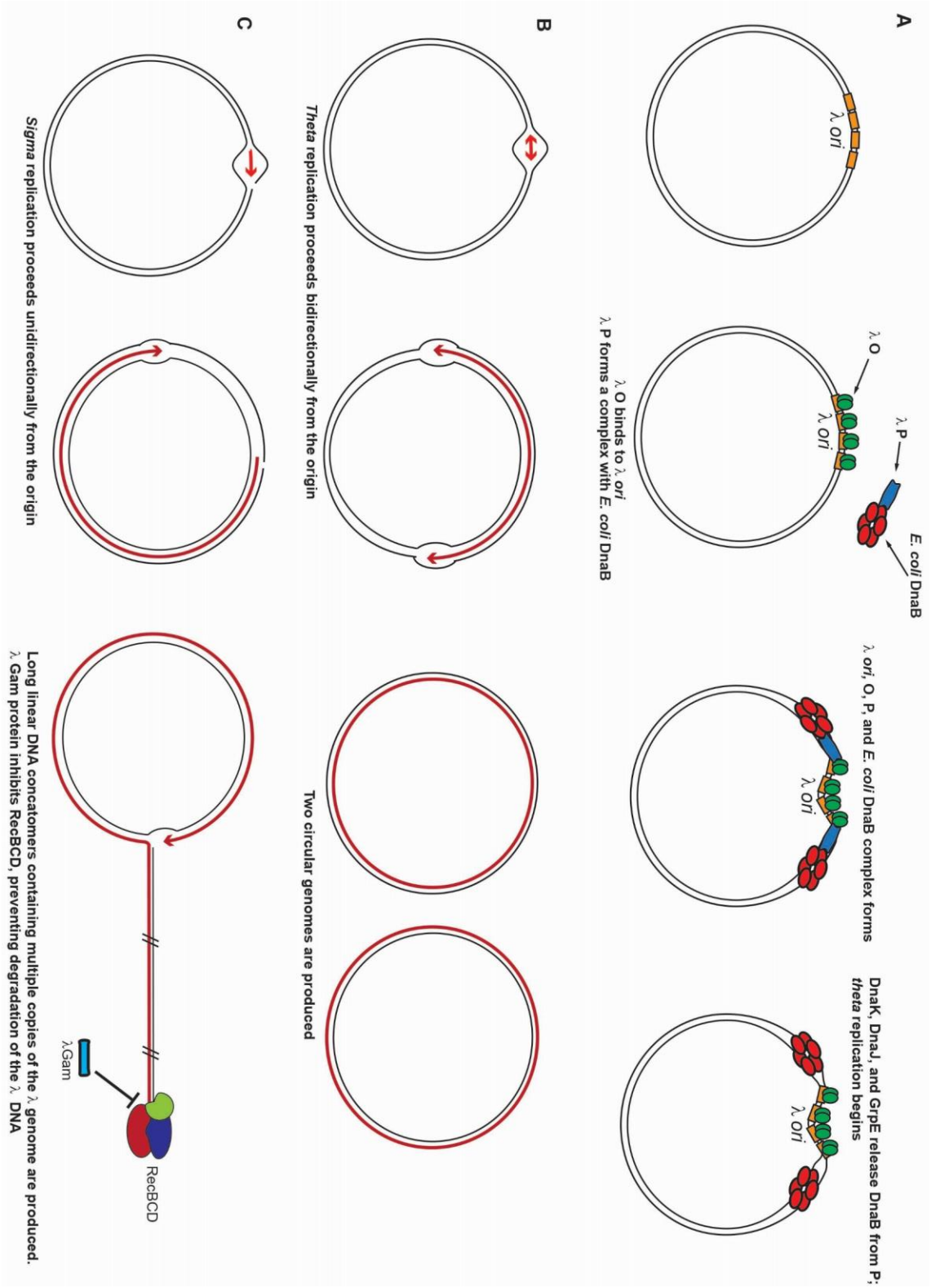


Figure 5A. Lambda phage replication initiation. The origin DNA is represented

by orange rectangles. Green ovals represent the lambda O protein, blue rectangles represent the lambda P protein, and the red ovals are *E. coli* DnaB. B. *Theta* mode of DNA replication. Nascent DNA is colored red. C. *Sigma* mode of DNA replication. Nascent DNA is red. The RecBCD complex is labeled as is the lambda Gam protein. Black slack marks on the DNA signify the start of another lambda phage genome.

P2 *old* in the context of lambda phage and *recBC- E. coli*

As mentioned above, it was established in the mid 1960s-1970s, that presence of the *old* gene from P2 phage was capable of preventing the growth of lambda phage as well as causing the death of *recBC- E. coli.* The mutants of lambda phage that were able to overcome the *old* phenotype were called *Spi-* (sensitive to P2 interference). Full *Spi-* lambda mutants were able to form large clear plaques on P2 lysogens. The *Spi-* mutations were mapped to the region of the genome containing the lambda *red* genes (Lindahl G, 1970). It was quickly noticed, however, that deletion of *exo, bet,* and *gam* was not sufficient to generate a full *Spi-* phenotype. It was originally supposed that an additional gene called *delta* needed to be deleted as well, though further genetic analysis revealed that the mutation giving rise to the full *Spi-* phenotype was actually the introduction of a chi site in the lambda genome (Malone RE, 1975). Even without considering the action of the *old* gene, one can surmise that deletion of the *red* genes (and particularly *gam*) allowed for the full activity of the RecBCD complex, which would digest much of the replicating lambda DNA. However, introduction of a chi site would reverse this

effect. Lambda would now become a proper substrate for RecBCD such that it would facilitate recombination between the lambda genomes, fulfilling the role of *exo* and *bet.* While this data is reasonable when just assessing lambda replication and RecBCD function, it does not explain the role of P2 *old.*

Work by Bregegere sought to more fully understand the role the OLD protein plays in the disruption of lambda replication. He discovered that P2 *old* did not merely prevent lambda phage from replicating, but rather caused the death of the host cell shortly after lambda began replicating. The genes *O* and *P* were absolutely required for both lambda interference as well as host cell killing (Bregegere F, 1975). However, in the presence of P2 *old* no phage DNA is replicated, which may suggest a role for the OLD protein in disrupting replication initiation (Lindahl G, 1970).

The limited biochemical data we have about the P2 OLD protein suggests that it acts as a 5'-3' exonuclease with some limited endonuclease activity, and that this activity is stimulated in the presence of ATP (Myung H, 1995). It has also been reported that mutation of *sbcA* or *sbcB* makes RecBC- cells insensitive to the P2 OLD protein (Barbour SD, 1970 and Myung H, 1995). Mutations in *sbcA* or *sbcB* allow the need for RecBCD for double strand break repair to be bypassed, activating the RecET and RecFOR repair pathways, respectively. We can therefore surmise that P2 OLD is not processing DNA ends to favor RecBCD activity, but rather P2 OLD is likely introducing some sort of damage, perhaps at active replication forks or at the origin of replication. Understanding the biochemical mechanisms of P2 OLD

will hopefully elucidate the role P2 is playing *in vivo.*

OLD proteins in a broader light

Though P2 OLD has been by far the most characterized, OLD proteins are actually very common in bacteria, archaea, and some other bacteriophage. There is a mention of the OLD protein from *Salmonella enterica* having some impact on organism fitness during heat stress; however, this observation is not immediately compelling nor does it offer insight into the function of OLD proteins (Khatiwara A, 2012). It is therefore important to pursue understanding the OLD proteins, not only to clarify an observation from the past, but to characterize a relatively unexplored class of nucleases.

References

Amundsen SK, Smith GR. The RecB helicase-nulcease tether mediates Chi hotspot control of RecBCD enzyme. Nucleic Acids Res. 2019: 47: 197-209.

Amundsen SK, Taylor AF, Reddy M, Smith GR. Intersubunit signaling in RecBCD enzyme, a complex protein machine regulated by Chi hotspots. Genes Dev. 2007: 21: 3296-3307.

Amundsen SK, Neiman AM, Thibodeaux SM, Smith GR. Genetic dissection of the biochemical activities of RecBCD enzyme. Genetics. 1990: 126: 25-40.

Anderson DG, Kowalczykowski SC. The recombination hot spot χ is a

regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev. 1997: 11: 571-581.

Anderson DG, Kowalczykowski SC. The translocating RecBCD enzyme stimulates recombination by directin gRecA prote nonto ssDNA in a χ regulated manner. Cell. 1997; 90: 77-86.

Arnold DA, Kowalczykowski SC. Facilitated loading of RecA protein is essential for recombination by RecBCD enzyme. J Biol Chem. 2000: 275: 12261-12265.

Bell JC, Kowalczykowski SC. RecA: regulation and mechanism of a molecular search engine. 2016. Trends Biochem Sci. 14; 491-507.

Bentchikou E, Servant P, Coste G, Sommer S. A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans.* 2010. PLos Genet; 6: e10000774.

Bertani LE. Host-dependent induction of phage mutants and lysogenization. Virology. 1960: 7: 553-569.

Benzer S. Fine structure of a genetic region in bacteriophage. 1955. PNAS; 41; 344-354.

Clark AJ, Margulies AD. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. PNAS. 1965: 53: 451-459.

Boehmer PE, Emmerson PT. The RecB subunit of the *Escherichia coli* RecBCD Enzyme Couples ATP Hydrolysis to DNA Unwinding. J Biol Chem. 1992: 267: 4981–4987.

Bregegere F. Bacteriophage P2-lambda interference. II. Efects on the host under the control of lambda genes *O* and *P.* 1975. J Mol Biol; 104: 411-420.

Casjens SR, Hendrix RW. Bacteriophage lambda: early pioneer and still relevant. 2015. Virology; 0: 310-330.

Clark AJ, Margulies AD. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. PNAS. 1965; 53: 451-459.

Connelly JC, Kirkham LA, Leach DR. The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. 1998. PNAS; 95: 7969-7974.

Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. 1961. Nature; 30: 1227-1232.

Dillingham MS, Spies M, Kowalczykowski SC. RecBCD enzyme is a bipolar DNA helicase. Nature. 2003; 423: 893-897.

Enquist LW, Skalka AM. Replication of bacteriophage lambda DNA. 1978. Trends Biochem Sci; 3: 279-283

Fredericq P. Genetic transfer of lysogenic properties in *E. coli.* 1953. C R Soc Biol; 147, 2046-2048.

Gibson FP, Leach DR, Lloyd RG. Identification of *sbcD* mutations as cosuppressors of *recBC* that allow propagation of DNA palindromes in *Escherichia coli* K-12. 1992. J Bacteriol; 174: 1222-1228.

Gillen JR, Willis DK, Clark AJ. Genetic analysis of the RecE pathway of genetic recombination in *Escherichia coli* K-12. 1981. J Bacteriol; 145: 521-532.

Emmerson PT. Recombination mutants of *Escherichia coli* K12 the map between *thyA* and *argA.* Genetics. 1968: 60: 19-30.

Handa N, Morimatsu K, Lovett ST, Kowalczykowski SC. Reconstitution of initial steps of dsDNA break repair by the RecF pathway of *E. coli*. 2009. Genes Dev; 23: 1234-1245.

Howard-Flanders P, Theriot L. Mutants of *Escherichia coli* K-12 defective in DNA repair and in genetic recombination. Genetics. 1966; 53: 1137-1150.

Khatiwara A. *et al.* Genome scanning for conditionally essential genes in *Salmonella enterica* Serotype Typhimurium. 2012. Appl Environ Microbiol; 78:3098–3107.

Kolodner R, Hall SD, Luisi-DeLuca C. Homologous pairing proteins encoded by the *Escherichia coli* recE and recT genes. 1994. Mol Microiol; 11: 23-30.

Konieczny I, Marszalek J. The requirement for molecular chaperones in λDNA replication is reduced by the mutation π in λP gene, which weakens the interaction between λP protein and DnaB helicase. J Biol Chem; 270: 9792-9799.

Koonin EV, Wold YI, Aravind L. Protein fold recognition using sequence profiles and its application in structural genomics. Advan. Protein Chem. 2000: 54: 245-275.

Kowalczykowski SC. Initiation of genetic recombination and recombination-dependent replication. 2000. Trends Biochem Sci; 25: 156-165.

Krogh BO, Symington LS. Recombination proteins in yeast. Annu Rev Geneti. 2004; 38: 233-271.

Kushner SR, Nagaishi H, Clark AJ. Isolation of Exonuclease VIII: the enzyme associated with the *sbcA* indirect suppressor. 1974. PNAS; 71: 3593-3597.

Kushner SR, Nagaishi H, Clark AJ. Indirect suppression of *recB* and *recC* mutations by Exonuclease I deficiency. 1972. PNAS; 69: 1366-1370.

LeBowitz JH, McMacken R. The bacteriophage lambda O and P protein initiators promote the replication of single-stranded DNA. 1984. Nucleic Acids Res; 12: 3069-3088.

Lederberg EM. Lysogenicity in *E. coli* K-12. Genetics. 1951;36:560.

Lederberg EM, Lederberg J. Genetic Studies of Lysogenicity in *Escherichia Coli.* Genetics. 1953;38:51–64.

Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. 2002: 317: 41-72.

Leipe DD, Koonin EV, Aravind L. Evolution and classification of P-loop kinases and related proteins. J. Mol. Biol. 2003: 333:781-815.

Li Z, Karakousis G, Chiu SK, Reddy G, Radding CM. The beta protein of phage lamda promotes strand exchange. 1998. J Mol Biol; 276: 733-744.

Lindahl G, Sironi G, Bialy H, Calendar R. Bacteriophage lambda; abortive infection of bacteria lysogenic for phage P2. 1970. PNAS: 66: 587-594.

Little, JW. An exonuclease induced by bacteriophage λ. II. Nature of the enzymatic reaction. 1967. J Biol Chem; 242: 679-686.

Lloyd RG, Buckman C. Identification and genetic analysis of *sbcC* mutations in commonly used *recBC sbcB* strains of *Escherichia coli* K-12. 1985. J Bacteriol; 164: 836-844.

Lloyd RG, Barbour SD. The genetic location of the *sbcA* gene of *Escherichia coli.* 1974. Molec Gen Genet; 134: 157-171.

Malone RE, Chattoraj DK. The role of Chi mutations in the Spi- phenotype of phage lambda: lack of evidence for a gene delta. 1975. Mol Gen Genet; 143: 35-41.

Morimatsu K, Kowalczykowski SC. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. 2003. Mol Cell; 11: 1337-1347.

Mosig G, Yu S, Myung H, Haggård-Ljungquist E, Davenport L, Carlson K, Calendar R. A novel mechanism of virus-virus interactions: bacteriophage P2 tin protein inhibits phage T4 DNA synthesis by poisoning the T4 single-stranded DNA binding protein, gp32. Virology. 1997: 230: 72-81.

Murphy KC. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. 2007. J Mol Biol; 371: 19-24.

Myung H, Calendar R. The *old* exonuclease of bacteriophage P2. 1995. J

Bacteriol; 177: 497-501.

Sakai A, Cox MM. RecFOR and RecOR as distinct RecA loading pathways. 2009. J Biol Chem; 284: 3264-3272.

Saraste M, Sibbald PR, Wittinghofer A. The P-loop - a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990;15:430–434.

Savir Y, Tlusty T. RecA-mediated homology search as a nearly optimal signal detection system. 2010. Mol Cell; 40: 388-396.

Shinedling S, Parma D, Gold L. Wild-type T4 is restricted by the lambda *rex* genes. 1987. J virol; 1987: 3790-3794.

Signer ER, Weil J. Recombination in bacteriophage λ. I. Mutants deficient in general recombination. 1968. J Molec Biol; 34: 261-271.

Sironi G. Mutants of *Escherichia coli* unable to by lysogenized by the temperate bacteriophage P2. Virology. 1969: 37: 163-176.

Smith GR. Homologous recombination in procaryotes. Microbiol Rev. 1988: 52: 1-28.

Willetts NS, Clark AJ. Characteristics of some multiply recombination-deficient strains of *Escherichia coli*. J. Bacteriol. 1969: 100: 231-239.

Wegrzyn G, Wegrzyn A. Genetic switches during bacteriophage λ development. 2005. Prog Nucleic Acid Res Mol Biol; 79: 1-48.

Zahn K, Blattner FR. Binding and bending of the lambda replication origin by

the phage O protein. 1985. EMBO J; 4: 3605-3616.

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

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

Carl J. Schiltz[1], April Lee[1], Edward A. Partlow[1], Christopher J. Hosford[1], and Joshua S. Chappie[1*]

[1] Department of Molecular Medicine, Cornell University, Ithaca, NY, 14853, USA

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

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 *Xanthamonas campestris* p.v. *campestris* Class 2 OLD proteins at 2.24Å and 1.86Å 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 group2. 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 base9. In this scheme, one metal (metal A) deprotonates the nucleophile while the other (metal B) stabilizes the pentavalent transition state intermediate2. 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 phosphoryl-hydrolases 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 **Top**oisomerase/**prim**ase (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 α1-β1 loop, an invariant glycine following β2, and a conserved DxD motif between α3 and β3 (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 requirements. Toposiomerases 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 **l**ysogenization **d**efect (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 E. 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 *Xanthamonas campestris* p.v. *campestris* (Xcc) and present the crystal structures of their catalytic C-terminal regions at 2.24Å and 1.86Å 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.

MATERIAL 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 OD_{600} 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 $MgCl_2$, and a Roche complete protease inhibitor cocktail tablet. Lysozyme was added to 1 mg/ml and the mixture was incubated for 15 minutes rocking at 4°C. Cells were disrupted by sonication for a total of 4 minutes and the lysate was cleared of debris by centrifugation at 13 000 rpm (19 685 g) for 30 minutes at 4°C. The supernatant was filtered using a 0.45 μm syringe filter, loaded onto a 5 ml HiTrap chelating column charged with $NiS0_4$ and then washed with nickel load buffer. Proteins were eluted with an imidazole gradient from 30mM to 1M. 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 $MgCl_2$, 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.

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 μl and reservoir volume of 65 μl. 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 minutes or 10 mM ethylmercury chloride for 45 minutes. 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 $\alpha = \beta = \gamma = 90°$. 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 12300 eV, 11570 eV, 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_{work}/R_{free}: Xcc^{CTR} Pt, 1.86Å, 0.2115/0.2379; Xcc^{CTR} Hg, 1.95Å, 0.2014/0.2416; Xcc^{CTR} I, 2.30Å, 0.2152/0.2727 (Supplementary Table S1).

Bp^{CTR} was crystallized by sitting drop vapor diffusion in 0.1 M HEPES pH 7.5, 0.23 M $MgCl_2$, 30% PEG 400 and 0.001 M glutathione with a drop size of

1 μL and reservoir volume of 65 μL. Crystals appeared within two to three 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 = 83.256Å b = 105.669 c = 123.764 and $\alpha = \beta = \gamma = 90°$. 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_{work}/R_{free} of 0.2134/0.2598 (Supplementary Table S1). 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 size-exclusion chromatography using a Superdex

200 10/300 gl (GE) column equilibrated in SEC-MALS buffer (20 mM HEPES pH7.5, 150 mM NaCl, 5 mM $MgCl_2$, and 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 µM protein to a final volume of 20 µL in DNA cleavage buffer (20 mM Tris-OAc pH 7.9, 50 mM KCl, 0.1 mg/mL BSA, and 10 mM divalent metal). Reactions were incubated at 37ºC for 60 minutes and quenched with 5 µL of 0.5 M ETDA pH 8.0. Samples were analyzed via native agarose electrophoresis.

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 MG1655 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 S2). 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 S2). Class 1 OLD family members (542/837) – including P2 phage, *Escherichia coli*, and *Salmonella typhimurium* – exist as single, isolated genes (Supplementary Figure 1a). Class 2 OLD homologs (295/837) appear in tandem with a UvrD/PcrA/Rep-like helicase (Supplementary Figure 1B), 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 ~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 S2). 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 S2).

We selected numerous candidates from each class for expression studies. Like P2 OLD, Class 1 homologs behaved poorly during purification. Class 2 homologs, in contrast, were intrinsically more stable and generally provided greater yields of soluble, monodispersed protein. Specifically, the Class2 OLD homologs from *Burkholderia pseudomallei* and *Xanthamonas campestris* p.v. *campestris* could be purified to homogeneity (Figure 1A,B) 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 (Figure 1C). In contrast, truncated constructs containing the C-terminal region of each homolog (Xcc^{CTR}, Bp^{CTR}; Supplementary Figure 1A,B) were each monomeric by SEC-MALS analysis (Figure 1D).

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 1E). Cleavage was measured visually by the disappearance of the DNA substrate, which was rapidly degraded under conditions that promote nuclease function. Bp^{FL} exhibits cleavage in the presence of Mg^{2+} and enhanced activity with Mn^{2+}

(Figure 1E). We also observe weak activity in presence of Zn^{2+} and Co^{2+}. Addition of ATP has no appreciable effect on Bp^{FL} cleavage activity with linear substrates (Figure 1F). Bp^{CTR} similarly shows cleavage with Mg^{2}, Mn^{2+}, and Co^{2+} (Figure 1G). These data indicate that the critical catalytic resides associated with nuclease function reside in the C-terminal half of OLD proteins and that ATP binding and/or hydrolysis 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 1H). Bp^{FL} shows weak nicking activity with all metals as compared to the DNA alone and EDTA controls (Figure 1H), with Mg^{2+}, Mn^{2+}, and Co^{2+} again eliciting the strongest effects. Under these conditions, only Mn^{2+} promotes processive cleavage that significantly degrades the circular substrate (Figure 1H). These activities are largely unchanged by the addition of ATP (Figure 1I). Bp^{CTR}, in contrast, is more active and shows pronounced nicking activity in the presence of every metal tested (Figure 1J). Together these data show that Bp OLD can function as both an exo- and endonuclease with a strong preference for Mn^{2+}. The analogous Xcc constructs exhibited the same general activity profile albeit to a lesser extent (data not shown), suggesting Xcc OLD is a less efficient nuclease.

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 ~4Å 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 2Å and three independent structures were solved using SAD datasets from platinum, mercury, and iodide derivatives (Supplementary Table S1, Supplementary Figures S2 and S3A). These models show strong agreement with an overall RMSD of 0.42-0.44Å 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 S3A). Residues 374 to 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.-2.3Å) 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 ($\alpha6^H$). $\alpha6^H$ extends into a groove along one face of the Toprim's central β-sheet, forming extensive hydrophobic interactions (Figure 2B,C). Helix $\alpha5^H$ and the upper portion of $\alpha6^H$, 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 S4) and together bury a total surface area of 1341Å^2. Similar interactions are observed between the domains in the Xcc^{CTR} (Supplementary Figure S2).

Many Toprim family members contain individual structural inserts into the core Toprim fold (Figure 3, Supplementary Figure S5). 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 and Supplementary Figure S5, Teal). Class 2 OLD proteins show sequence variability across this insert region (Supplementary Figure S4). 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,C) while the rest of the core fold is largely unchanged (Figure 3C). 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 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 (45): C.Esp1396I, Z score: 5.1, RMSD 2.5Å) (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 (46). Crystallographic studies have shown that these proteins are dimeric and α-helical, with each monomer containing a helix-turn-helix motif (47). 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,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,B). C protein dimers bind DNA operator sites cooperatively to exert concentration-dependent switching of promoter activation and repression (46,48). 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.

BpOLD 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 S3B). Consequently, T506 and E508 shift 1.2Å and 1.5Å toward the active site respectively, which facilitates the binding of two magnesium ions in a geometry consistent with two metal catalysis (Supplementary Fig. 3c, Figure 5a). 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 4.9Å 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 αH1 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 S4). Individual substitutions of metal A ligands (E400A, D455A, D457A) and metal B ligands (E404A, T506A, E508A) yielded minimal to moderate effects on cleavage activity. Thus, combinations of mutations (E400A/D455A/D457A, E404A/T506A/E508A) were generated. Mutation of either the metal A or the metal B interacting residues together completely abolishes Bp OLD nuclease activity on linear DNA substrates *in vitro* (Figure

5C). These substitutions impair the processive degradation of circular plasmids in the presence of Mn^{2+}, though some nicking activity was still retained (Figure 5D). We also identify a conserved lysine residue in $\alpha5^{H}$ (K562) that extends toward the active site (Figure 5A,B, Supplementary Figure S4). K562A and K562E mutations similarly impair nuclease activity without perturbing direct interactions with either magnesium (Figure 5C,D). Together these data define the key catalytic machinery of Class 2 OLD nucleases and support a two-metal catalysis mechanism for cleavage.

The organization of the Bp OLD active site is structurally conserved in RNase M5 enzymes and DnaG primases (Figure 6, Supplementary Figure S5). Along with the invariant Toprim glutamate and conserved DxD aspartates, D31 and 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 $\alpha1^{H}$ of the Bp OLD helical domain (Figure 3A, dashed circle; Supplementary Figure S5, inset). DnaG primases contain a similar set of catalytic machinery (16). The analogous C-terminal acidic residue of the DnaG Toprim, however, is directed away from the active site via interaction with a conserved arginine residue 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 DxD 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 toposiomerases (Y782 in *Saccharomyces cerevisiae* topoisomerase II) (49). These differences highlight the evolutionary 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). B form DNA can bind patches 1 and 2 leading into the cleft but sterically clashes with the protein near patches 3 and 4 (Supplementary Figure S6A). 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 (50) (Figure 7A). The presence of a G:T mismatch in this substrate kinks the DNA at a 45 degree angle (Supplementary Figure 6B), allowing it to interact unencumbered with all four basic patches (Figure 7A). The orientation of the modelled substrate

would clash with both the Toprim core α2 and α3 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 points toward 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 *Burkholderia pseudomallei* and *Xanthamonas campestris* pv.*campestris*. Bp OLD catalyzes metal-dependent nicking and cleavage of DNA substrates *in vitro*. The N-terminal ATPase domain is not required for this activity and instead 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 E400 and E508 as being critical metal binding side chains (Figure 5). These residues are absolutely conserved among Class 2 OLD proteins (Supplementary Figure S4) and together coordinate two bound magnesium ions in a geometry that supports two-metal

catalysis (Figures 5 and 7C,D). We also find K562 in the 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. 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. This prohibits the unambiguous identification of Class 1 catalytic machinery by sequence alignment alone. Additional biochemical and structural studies will thus be necessary to determine if Class 1 proteins use a similar mechanism or have evolved a different set of machinery to achieve the same end.

The spatial organization of acidic residues in the Bp OLD active site directly mirrors that of RNase M5 maturases (Figure 6, Supplementary Figure S5). 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 circular permutated. 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 S5). 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.

Computational modeling shows that a bent DNA substrate engages all four basic patches on the surface Bp^{CTR} while B form DNA only interacts with two and sterically clashes with the remainder of the molecule (Figure 7). The favorable orientation of the bent substrate positions one strand in the active site with a phosphate situated directly between the two bound magnesium ions. 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 α2 and α3 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. It remains to be seen whether the DNA structure and/or conformational flexing of the Toprim and helical domains would thus be necessary for efficient cleavage. This may be

an important mechanistic requirement and suggests that Class 2 OLD nucleases may recognize structural rather than sequence features for targeting and function.

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. P2 OLD exhibits exonuclease activity *in vitro* (28) and Bp OLD readily degrades linear lambda DNA in the presence of Mn^{2+} 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. 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 (51-53). 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.

ACCESSION NUMBERS

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 Databank with the accession number 6NK8.

ACKNOWLEDGEMENT

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, 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-I2D-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

None declared.

AUTHOR CONTRIBUTIONS

All experiments were conceived by Schiltz and Chappie. Schiltz was responsible for protein construct design, cloning, protein purification, biochemistry, X-ray data collection, and final structural refinements. Lee assisted in growing crystals of Bp^{CTR}, collecting X-ray data, and performing structure refinements. Partlow assisted in crystallizing Xcc^{CTR}, collecting X-ray data, and performing initial structure refinements. Hosford assisted in collecting X-ray data and performing initial structure refinements. Results were

analyzed and interpreted by Schiltz and Chappie. Schiltz drafted the manuscript with critical revisions by Chappie.

.

REFERENCES

1. 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., F48, 220-228.

2. 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.

3. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. Biochemistry, 41, 13851-13860.

4. Pingoud,A., Fuxreiter,M., Pingoud,V., and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol. Life Sci., 62, 685-707.

5. 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.

6. 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.

7. 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 A resolution. J. Mol. Biol., 186, 417-433.

8. Freemont, P.S., Friedman,J.M., Beese,L.S., Sanderson,M.R. and Steitz,T.A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. Proc. Natl. Acad. Sci. USA, 85, 8924-8928.

9. Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA.

Proc. Natl. Acad. Sci. USA, 90, 6498-6502.

10. Yang,W. (2008) An equivalent metal ion in one- and two-metal-ion catalysis. Nat. Struct. Mol. Biol., 15, 1228-1231.

11. 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.

12. Gao,Y., and Yang,W. (2016) Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. Science, 352, 1334-1337.

13. 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.

14. 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.

15. 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.

16. 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..

17. 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.

18. 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.

19. Yang,W. (2010) Topoisomerases and site-specific recombinases: similarities in structure and mechanism. Crit. Rev. Biochem. Mol. Biol., 45, 520-534.

20. 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.

21. Usenik,A., Renko,M., Mihelič,M., Lindič,N., Borišek,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.

22. 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.

23. Lindahl,G., Sironi,G., Bialy,H., and Calendar, R. (1970) Bacteriophage lambda: abortive infection of bacteria lysogenic for phage P2. Proc. Nati. Acad. Sci. U.S.A., 66, 587-594.

24. Sironi,G. (1969) Mutants of Escherichia coli unable to be lysogenized by the temperate bacteriophage P2. Virology, 37, 163-176.

25. Ghisotti,D., Zangrossi,S., and Sironi,G. (1979) X-ray sensitivity of Escherichia coli lysogenic for bacteriophage P2. Mol. Gen. Genet., 169, 229-35.

26. 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.

27. Bregegere,F. (1974) Bacteriophage P2-lambda interference: inhibition of protein synthesis involves transfer RNA inactivation. J. Mol. Biol., 90, 459-467.

28. Myung,H., and Calendar R. (1995) The old exonuclease of bacteriophage P2. J Bacteriol., 177, 497-501.

29. 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.

30. Hendrickson,WA. (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) 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, W344-50.

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. Holm,L. and Rosenström,P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res., 38, W545-549.

46. 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.

47. 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.

48. 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.

49. Schmidt,B.H., Burgin,A.B., Deweese,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.

50. 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.

51. 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.

52. 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.

53. 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.

TABLES AND FIGURES

Schiltz et al. Figure 1

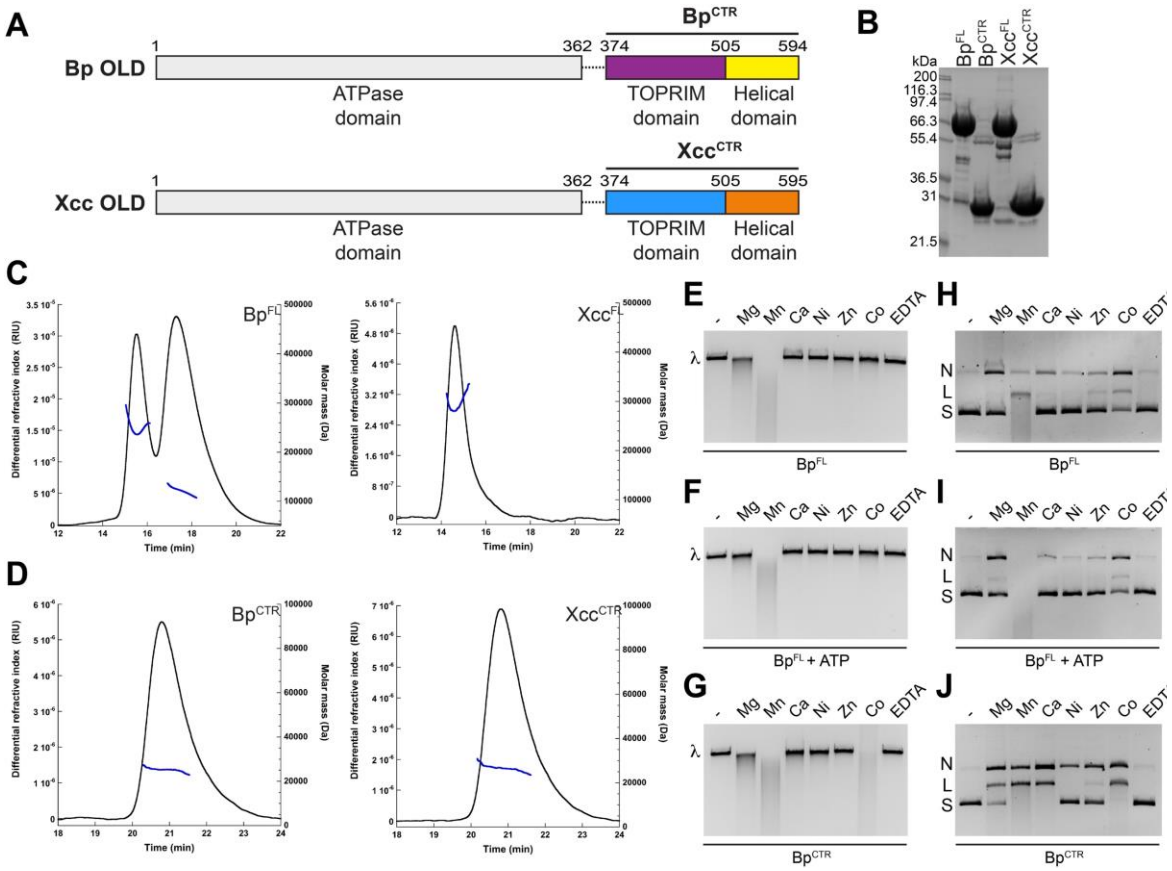


Figure 1. Purification and biochemical characterization of Class 2 OLD proteins. A. Domain architecture of Bp and Xcc OLD proteins. Boundaries of the CTR constructs are labeled. **B.** SDS-PAGE gel showing purified full-length and CTR constructs of Bp and Xcc OLD. **C-D.** SEC-MALS analysis of Bp (C) and Xcc (D) constructs. Black line denotes differential refractive index and blue line denotes measured mass across each peak. **E-F.** Metal-dependent nuclease activity of Bp^{FL} in the absence (E) or presence (F) of ATP on linear λ DNA. 'λ' denotes the position of uncut linear DNA **G.** Metal-dependent nuclease activity of Bp^{CTR} on linear λ DNA. **H-I.** Metal-dependent nuclease activity of Bp^{FL} in the absence (H) or presence (I) of ATP on supercoiled pUC19 DNA. 'N', 'L', and 'S' denote the positions of 'nicked', 'linearized', and

‘supercoiled’ DNA respectively. **J.** Metal-dependent nuclease activity of Bp^{CTR} on supercoiled pUC19 DNA.

Schiltz et al. Figure 2

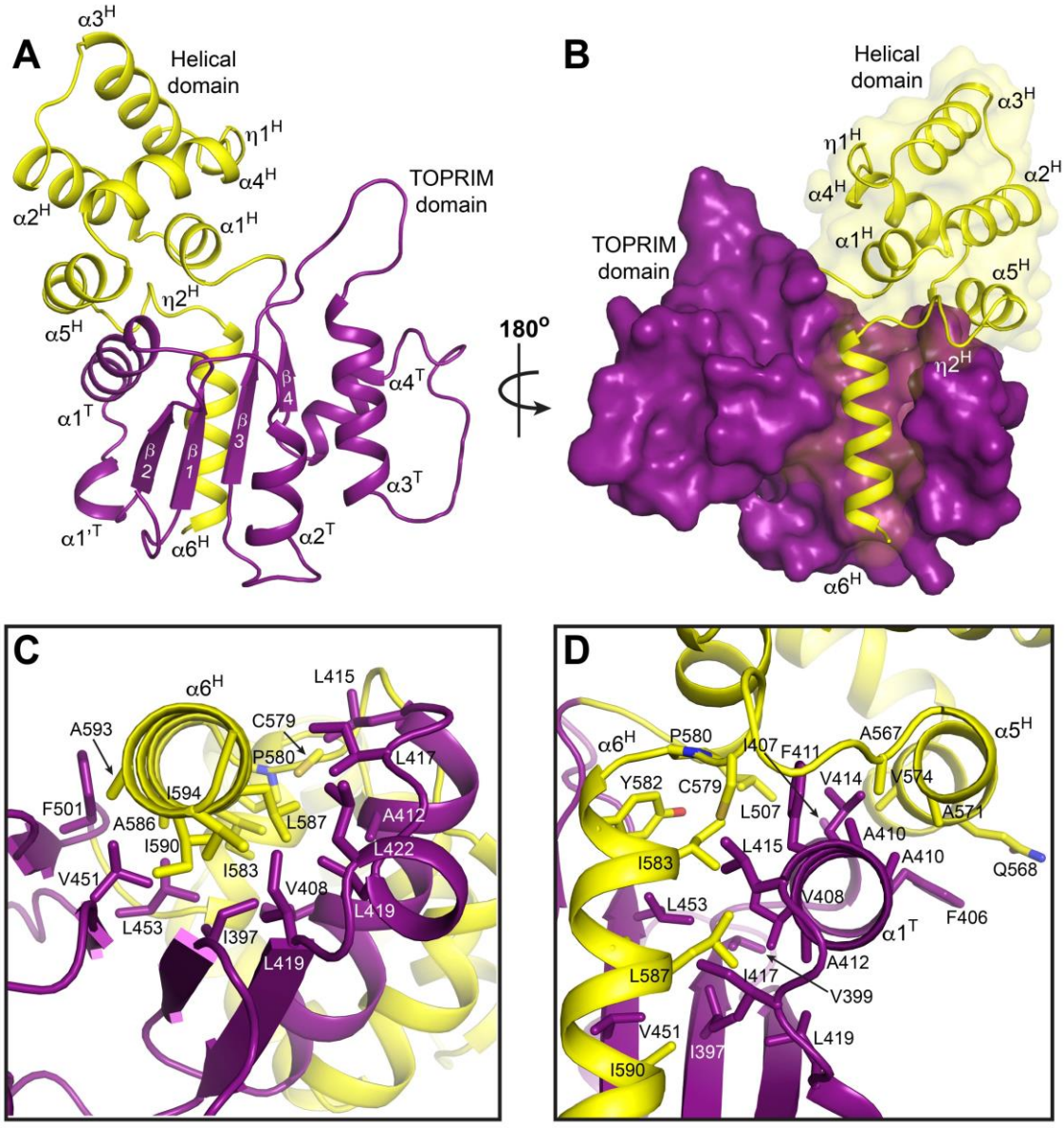


Figure 2. Structure of Bp^{CTR}. **A.** Structure of Bp^{CTR}. Toprim, and helical domains are colored purple and yellow respectively. **B.** Surface representation of domain interactions. **C-D.** Structural interactions between Toprim and helical domains. Side chains involved in stabilizing hydrophobic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

Schiltz et al. Figure 3

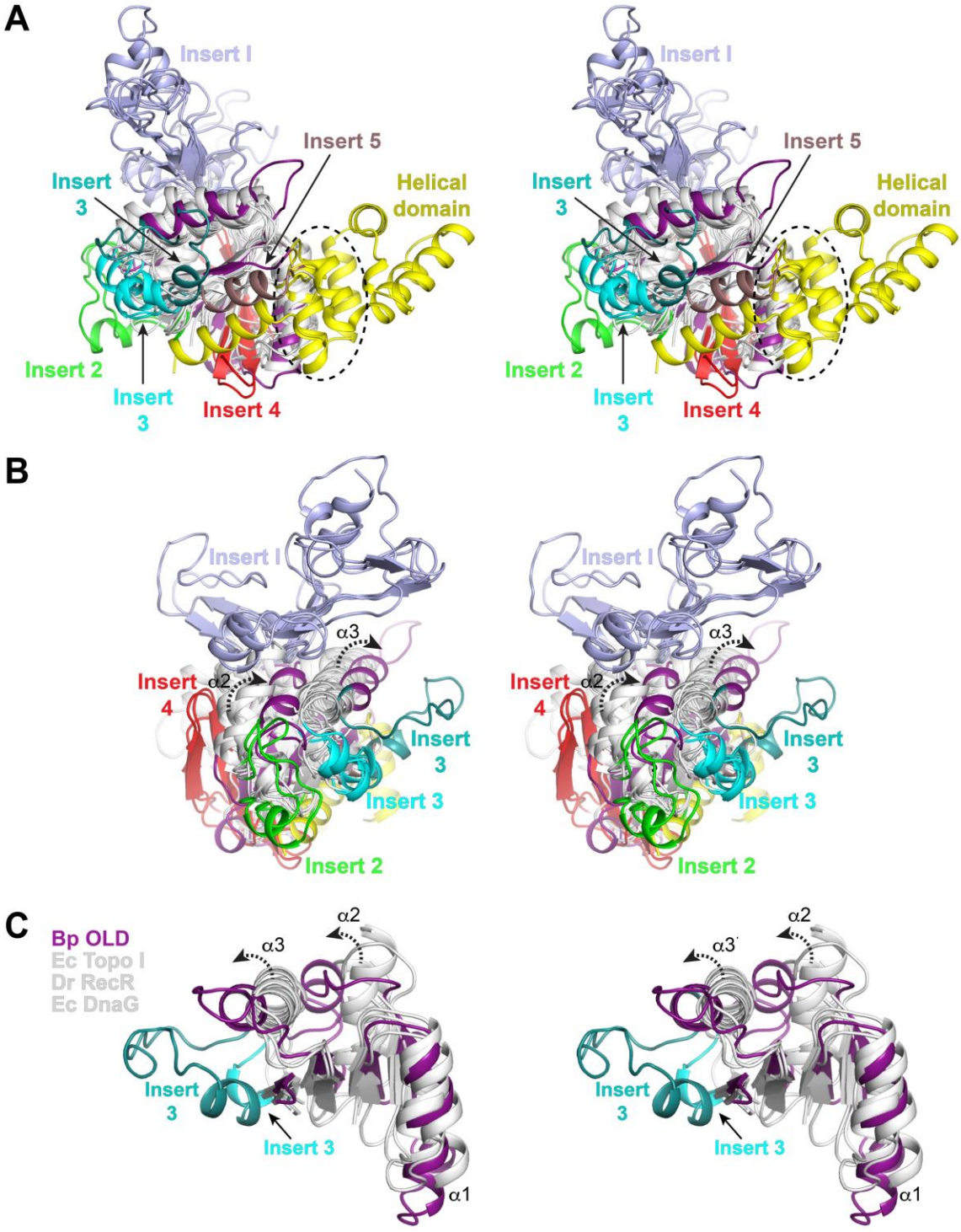


Figure 3. Topological differences in present in Class 2 OLD Toprim domains. A-B, Structural superposition of Toprim domains. End (A) and side

(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 Figure S5). Bp Toprim and helical domains are colored purple and yellow respectively. Dashed circle in b denotes the position of the glutamate helix. **c,** Stereo view of α2 and α3 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.

Schiltz et al. Figure 4

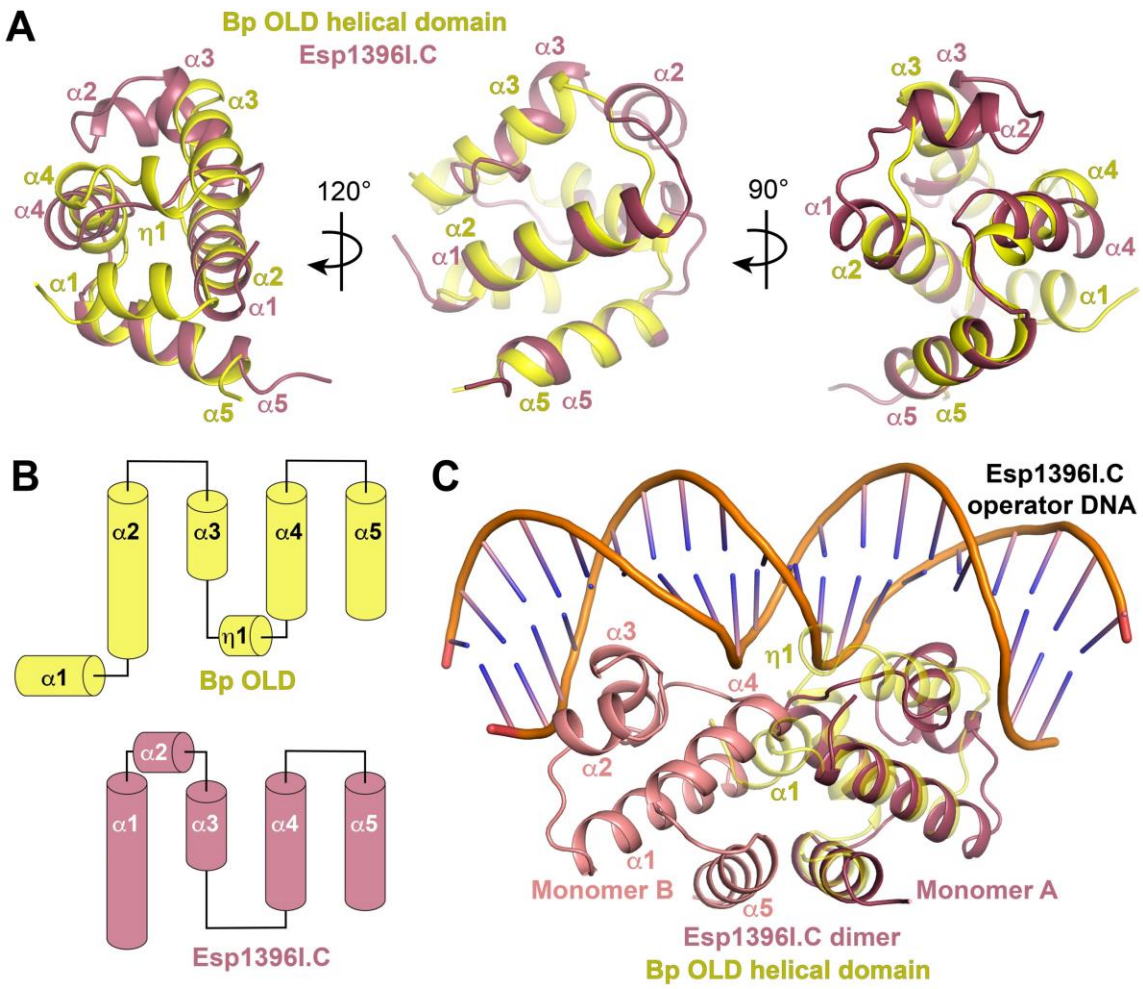


Figure 4. Helical domain in Class 2 OLD proteins shares structural homology with R-M system controller proteins. **A.** Structural superposition of Bp OLD helical domain (yellow) with Esp1396I.C controller protein (raspberry; PDB: 3CLC). **B.** Topology diagrams of Bp OLD helical domain and Esp1361.C monomer. **C.** Bp OLD architecture precludes dimerization as in controller proteins and suggests an alternate mode of DNA binding.

Schiltz et al. Figure 5

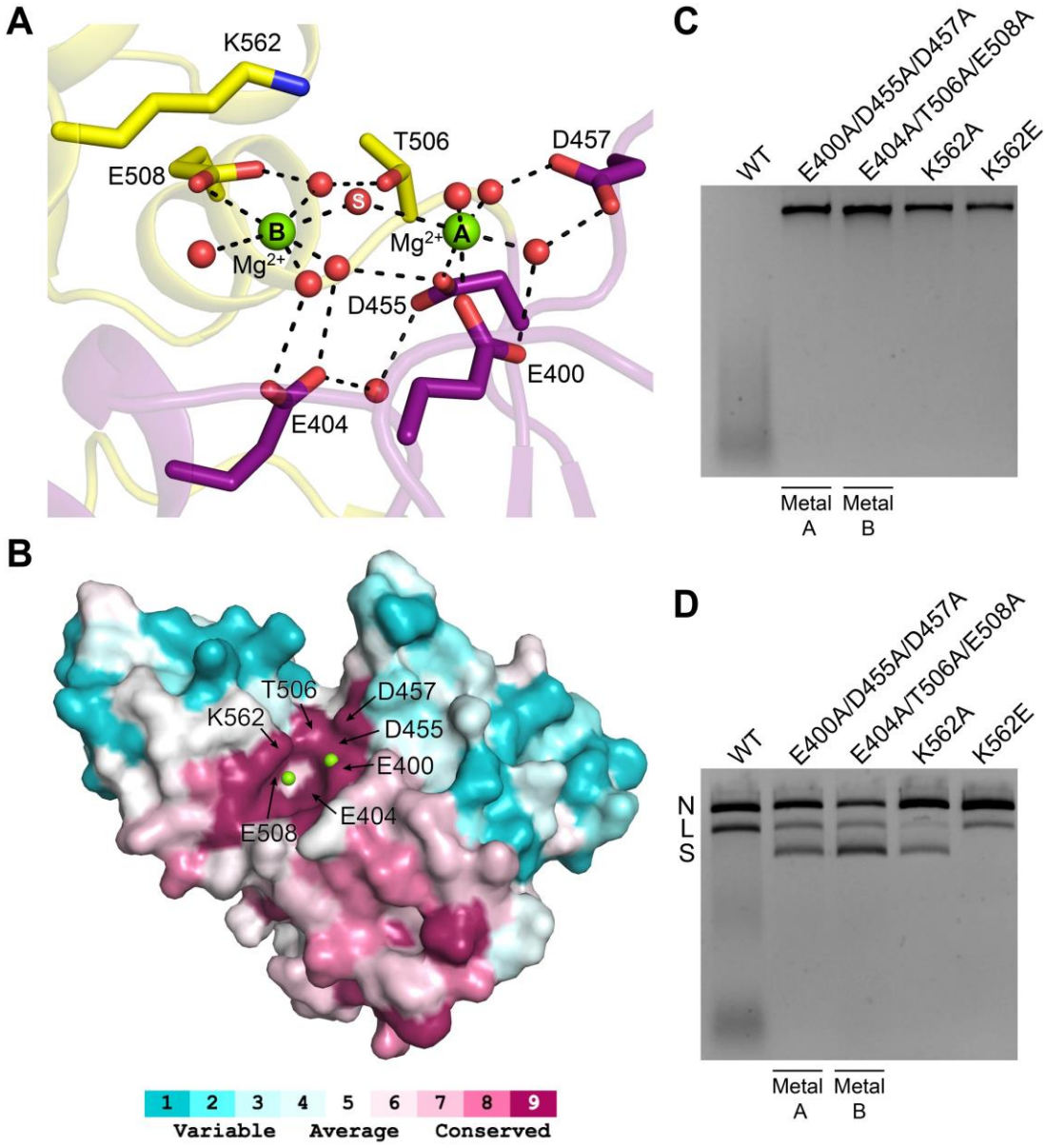


Figure 5. Bp OLD active site geometry is consistent with two metal catalysis. A. Metal coordinating residues in the Bp OLD active site. Ordered magnesium ions (green) and water molecules (red) are shown as spheres. Dashed lines indicate hydrogen bonds. **B.** Conservation of active site residues. Coloring generated using the ConSurf server (38) and the alignment in Supplementary Figure S4. **C.** Cleavage activities of Bp^{CTR} active site

mutants on linear λ DNA. **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 in the presence of manganese as described in the Materials and Methods.

Schiltz et al. Figure 6

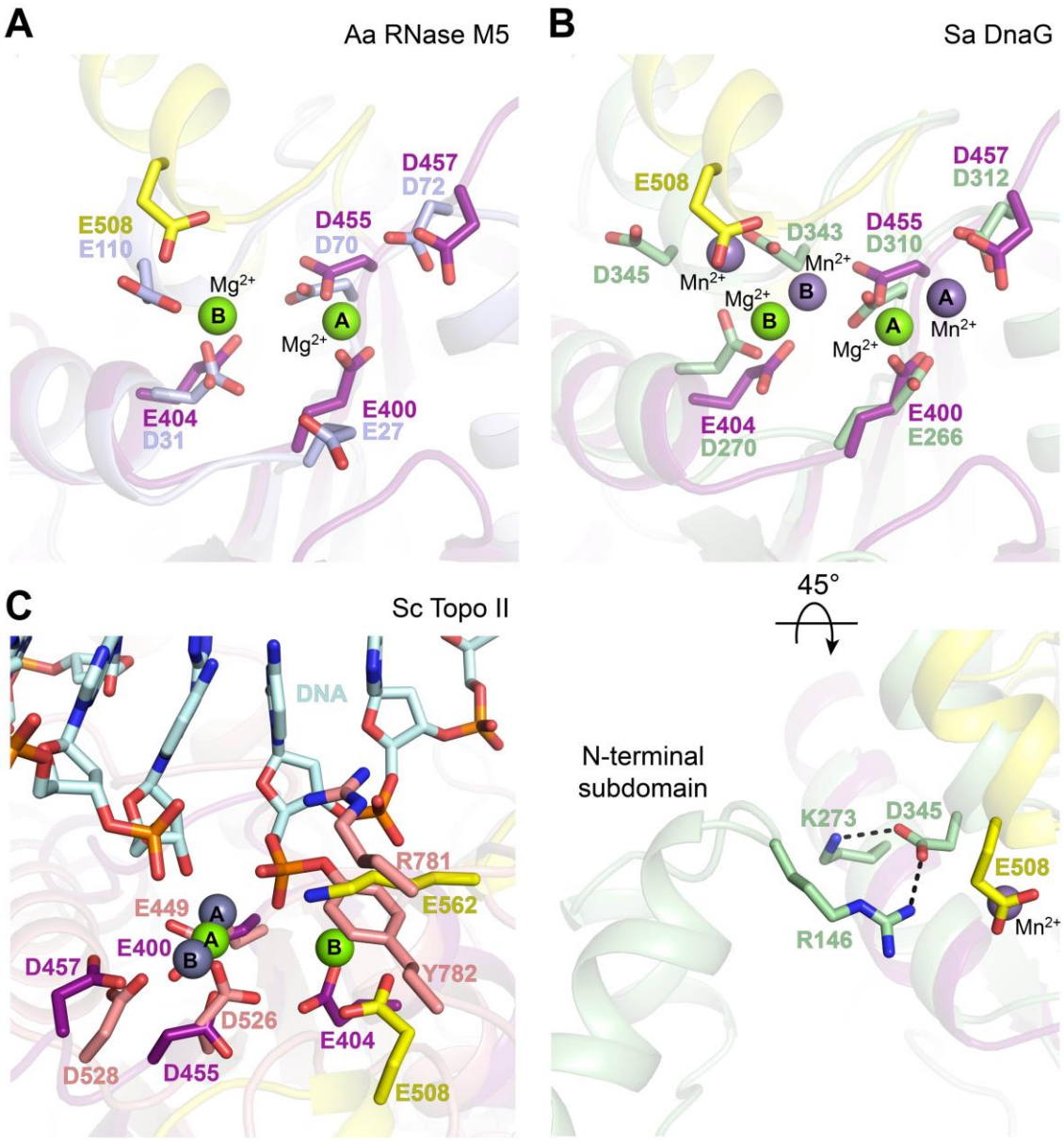


Figure 6. Conservation of active site elements among OLD nucleases, RNase M5 maturases, and DnaG primases. Comparison of Bp^{CTR} active site (Toprim domain, purple; helical domain, yellow) with different Toprim family members. Critical side chains in each protein are labeled and the bound metal A and 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.

Schiltz et al. Figure 7

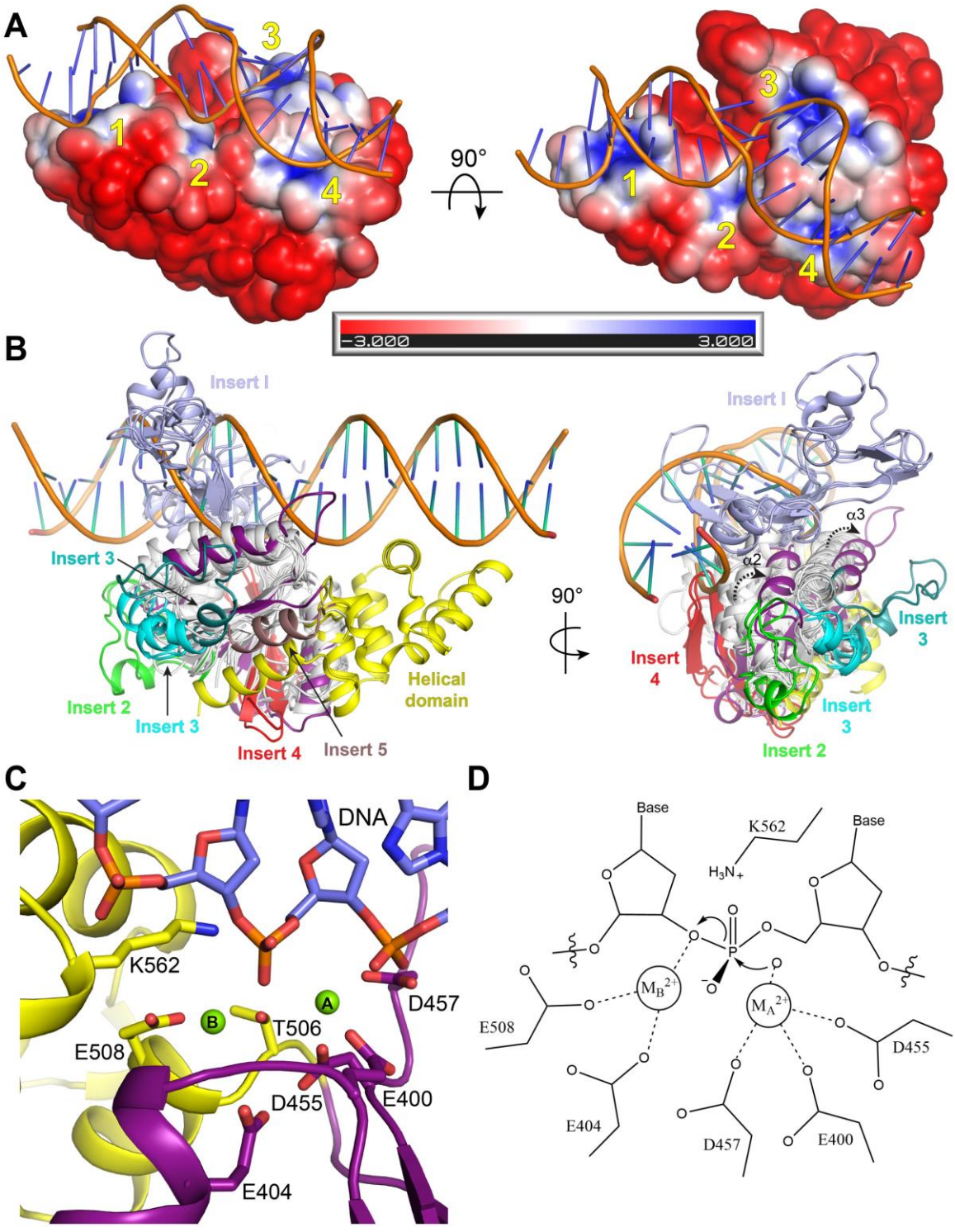


Figure 7. Helical domain orients DNA binding in Class 2 OLD nucleases.

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 -3 K_bT/e_c to +3 K_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.

Supplementary information for:

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

Carl Schiltz[1], April Lee[1], Edward A. Partlow[1], Christopher J. Hosford[1], and Joshua S. Chappie[1,*]

[1] Department of Molecular Medicine, Cornell University, Ithaca, NY, 14853, USA

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

Supplementary Tables

Supplementary Figures

Fig. 3.

Supplementary Figure 6. Structural constraints of Bp^{CTR} DNA binding. Related to Fig. 7.

Supplementary Table 1. X-ray data collection and refinement statistics.

	Xcc^{CTR} Pt	Xcc^{CTR} Hg	Xcc^{CTR} I	Bp^{CTR}
Data collection				
Space group	$P4_3$	$P4_3$	$P4_3$	C 2 2 2_1
Cell dimensions				
a, *b*, *c* (Å)	65.43, 65.43, 63.76	65.19, 65.19, 63.80	64.39, 64.39, 63.39	83.26, 105.67, 123.76
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	45.66-1.86 (1.93-1.86)	46.10-1.95 (2.02-1.95)	45.17-2.30 (2.38-2.30)	61.88-2.24 (2.32-2.24)
R_{sym} or R_{merge}	0.067 (1.41)	0.087 (1.564)	0.078 (0.819)	0.073 (1.183)
$I / \sigma I$	19.5 (1.4)	15.8 (1.3)	16.5 (2.2)	18.2 (1.6)
Completeness (%)	100 (100)	100 (99.9)	99.9 (99.7)	99.5 (97.4)
Redundancy	10.2 (10.4)	10.3 (10.7)	6.7 (6.2)	6.5 (6.4)
Refinement				
Resolution (Å)	45.66-1.86	46.10-1.95	45.17-2.30	61.88-2.24
No. reflections	44373	19604	11622	26561
R_{work} / R_{free}	0.2087/0.2297	0.2019/0.2302	0.2213/0.2766	0.2145/0.2515
No. atoms				3227
Protein	1647	1612	1673	3146
Ligand/ion	4	6	4	6
Water	73	84	36	75
B-factors				
Protein	49.95	51.41	61.59	22.37
Ligand/ion	133.6	47.60	66.34	25.4
Water	52.21	52.94	54.60	21.5
R.m.s deviations				
Bond lengths (Å)	0.008	0.009	0.009	0.008
Bond angles (°)	1.135	0.95	1.39	0.98

*Values in parentheses are for highest-resolution shell. Each dataset was derived from a single crystal.

Supplementary Table 2. Classification of old genes and associated operons.

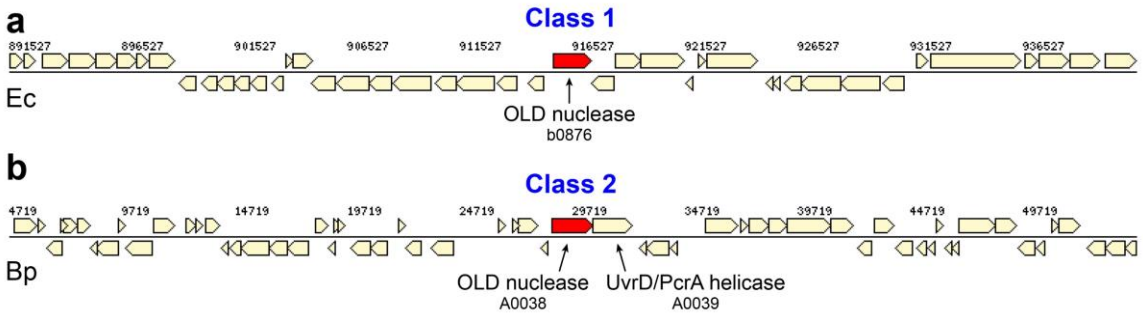
Phylogenetic distribution of OLD nucleases

	Number		
Phylum	**Total**	**Class 1**	**Class 2**
Acidobacteria	21	6	15
Aquificae	1	1	0
Bacteroidetes	29	3	26
Candidatus Cloacimonetes	1	0	1
Candidatus Micrarchaeota	1	1	0
Chlorobi	3	0	3
Chloroflexi	1	0	1
Cyanobacteria	1	0	1
Deinococcus-Thermus	1	1	0
Euryarchaeota	21	8	13
Firmicutes	69	18	51
Fusobactera	3	3	0
Plasmid	4	4	0
Proteobacteria	677	494	183
Thermotogae	1	1	0
Verrucomicrobia	1	0	1
Virus	2	2	0
	837	542	295

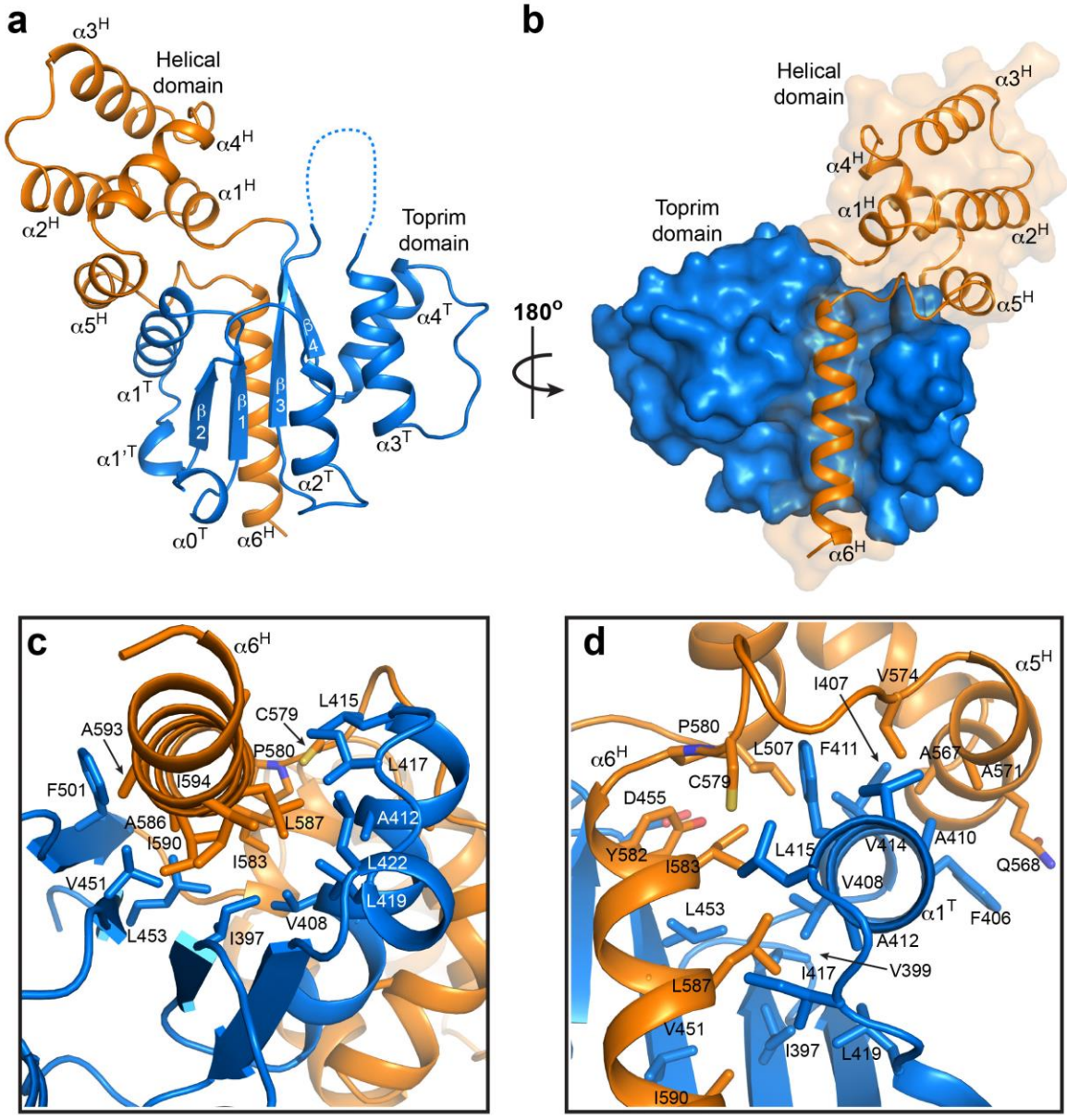
Operons associated with OLD nucleases

	Number		
Function	**Total**	**Class 1**	**Class 2**
Bacterial defense	38	7	31
Biosynthesis	2	1	1
Cell division and chromosomal partitioning	4	0	4
Conjugation and horizontal gene transfer	6	0	6
Isomerase	4	4	0
Macrolide resistance	10	10	0
Membrane transport	52	43	9
Metabolism	17	8	9
Pilus biogenesis and protein secretion	10	7	3
Protein synthesis	2	0	2
Replication and repair	59	35	24

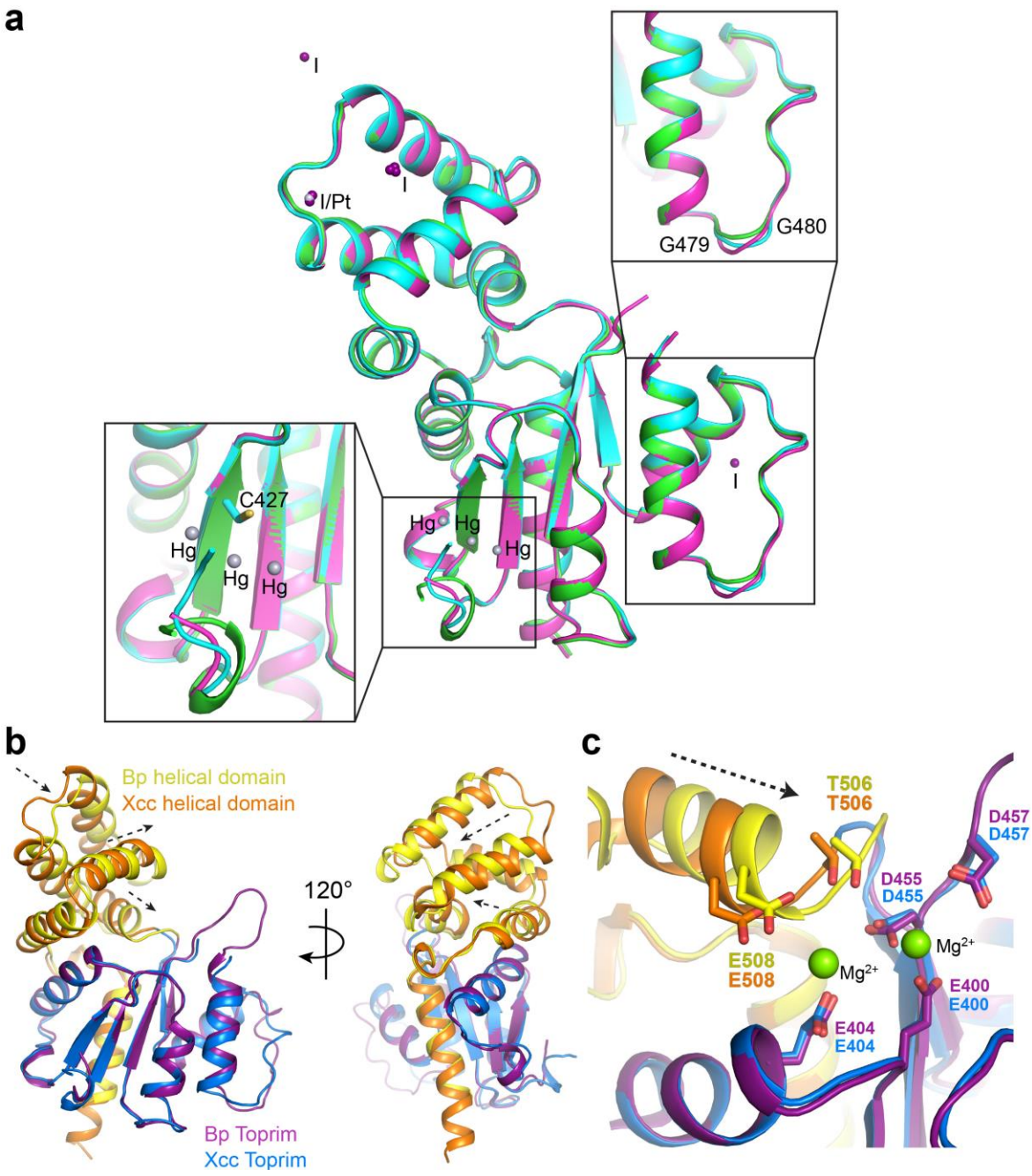
rRNA maturation	3	0	3
Signaling	8	2	6
Stress response	5	5	0
Transcriptional regulation	26	17	9
tRNA biogenesis	20	20	0
Toxin-antitoxin system	19	0	19
Unknown	26	2	24
Virulence	1	1	0



Supplementary Fig. 1: Classification of OLD proteins. a-b, Gene neighborhoods of Class 1 OLD homolog from *Escherichia coli* (a) and Class 2 OLD homolog from *Burkholderia Burkholderia pseudomallei* (b). Gene arrangements and annotations derived from DOE IMG/M database and analysis tools (Chen et al., 2017).

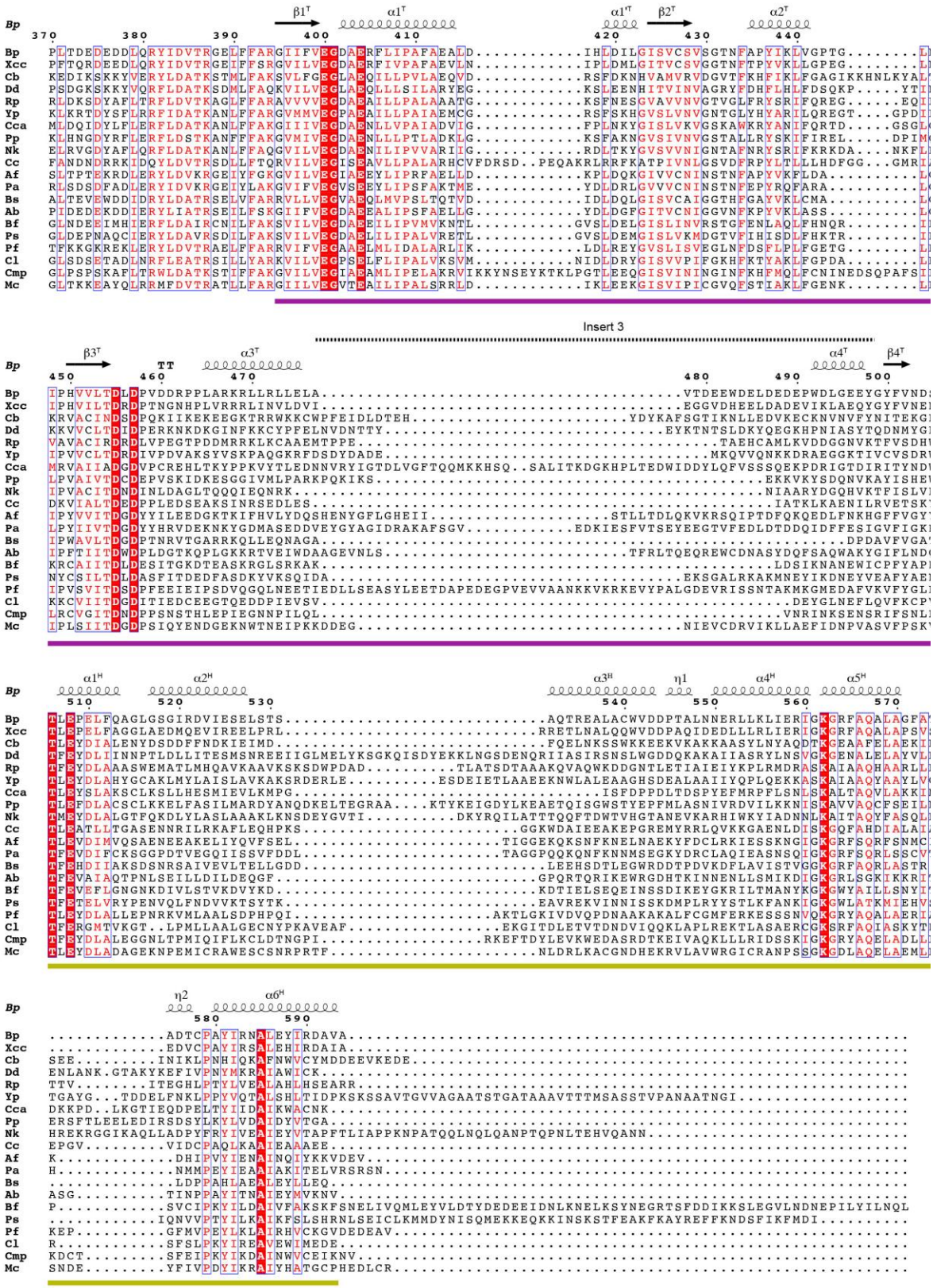


Supplementary Fig. 2. **Structure of Xcc^{CTR}**. **a,** Structure of Xcc^{CTR}. Toprim and helical domains are blue and orange respectively. **b,** Surface representation of domain interactions. **c-d,** Structural interactions between Toprim and helical domains. Side chains involved in stabilizing hydrophobic interactions are labelled.

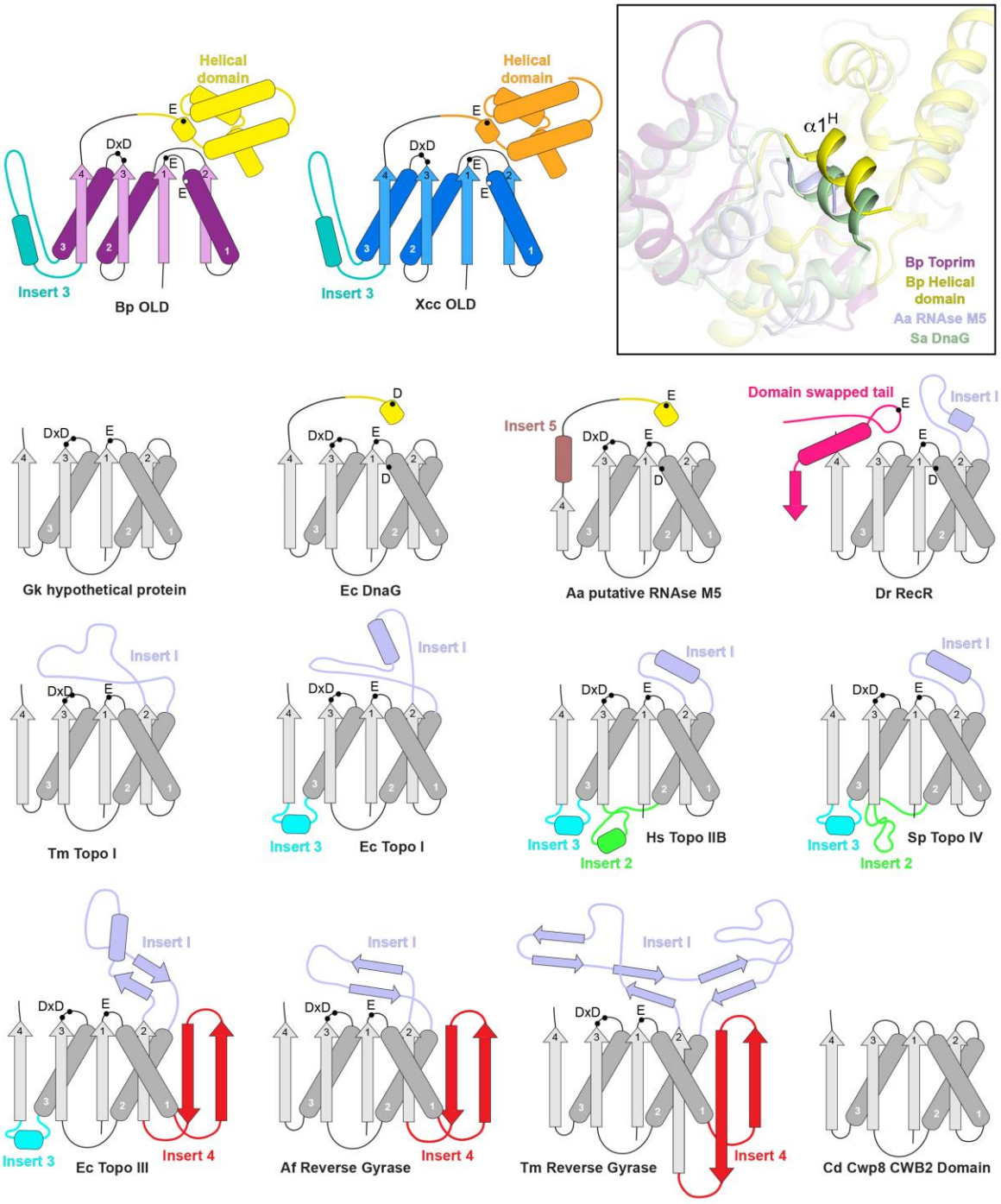


Supplementary Fig. 3. Comparison of OLD CTR structures. a, Structural superposition of Xcc^{CTR} derivative models. Pt, Hg, and I derivatives solved by SAD phasing are colored green, cyan, and magenta respectively. Zoomed inserts show minor structural differences between models. **b-c,** Structural

superposition of Xcc^{CTR} (orange and blue) and Bp^{CTR} (yellow and purple. Movement of the Bp OLD helical domain relative to the Toprim domain (b, dashed arrows) shifts the glutamate helix toward the active site (c, dashed arrow), thereby facilitating the binding of two magnesium ions (c, green spheres).

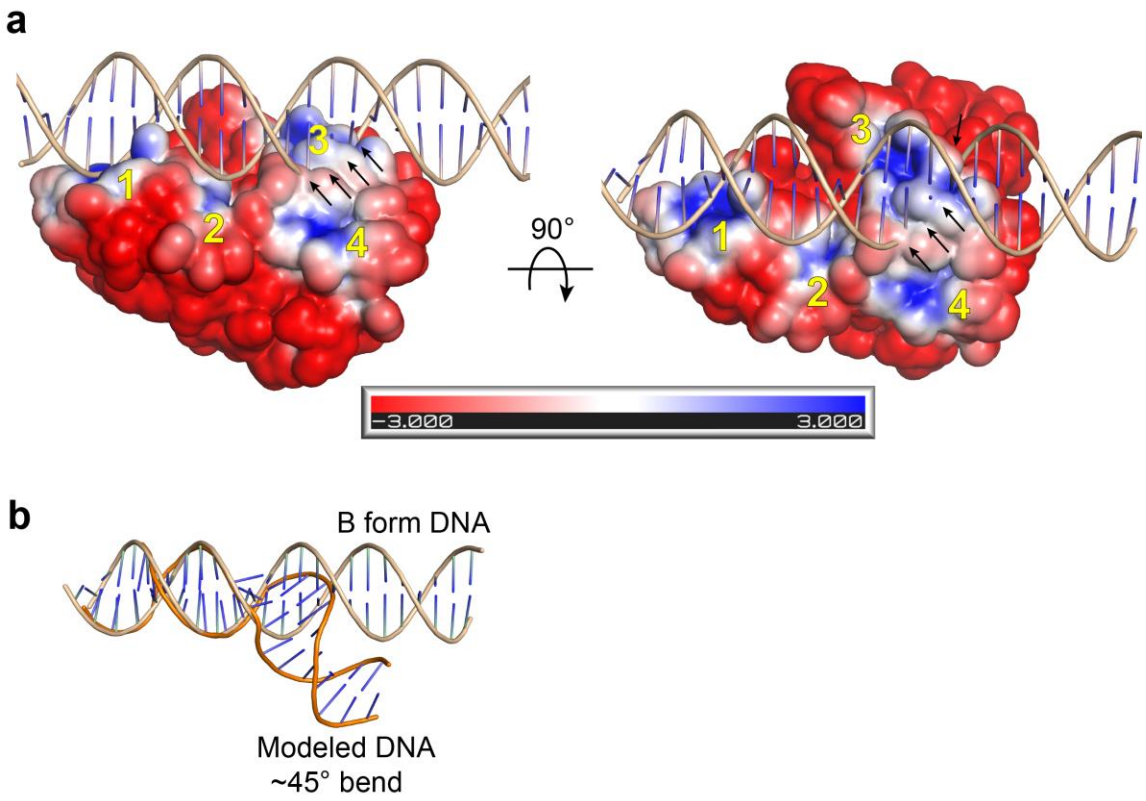

Insert 3

Supplementary Fig. 4: Sequence alignment and conservation of the C-terminal region in Class 2 OLD proteins. Secondary structure of Bp^{CTR} indicated above sequences. Purple and yellow bars below the sequences denote the Toprim and helical domain boundaries respectively. Insertion into the core Toprim fold present in Bp and Xcc OLD (Insert 3) is labeled (see Fig. 3 and Supplementary Fig. 5). Abbreviations are as follows: Bp, *Burkholderia pseudomallei*; Xcc, *Xanthamonas campestris* pv. *campestris* B100; Cb, *Clostridium botulinum* A Hall; Dd, *Desulfitobacterium dichloroeliminans*; Rp, *Rhodococcus pyridinivorans*; Yp,*Yersinia pseudotuberculosis* YPIII; Cca, *Candidatus Cloacimonas acidaminovorans*; Pp, *Paenibacillus peoriae*; Nk, *Niastella koreensis*; Cc, *Corynebacterium callunae*; Af, *Anoxybacillus flavithermus*; Pa, *Peptoclostridium acidaminophilum*; Bs, *Blastococcus saxobsidens*; Ab, *Acinetobacter baumannii* AB0057; Bf, *Bacteroides fragilis* NCTC9343; Ps, *Psychromonas* sp. CNPT3; Pf, *Pseudomonas fluorescens* LBUM223; Cl, *Chlorobaculum limnaeum*; Cmp, *Candidatus Methanosphaerula palustris*; Mc, *Methanosaeta concilii*.



Supplementary Fig. 5: Cartoon depictions of Toprim fold topologies. Bp and Xcc Toprim cores are shown in purple and blue respectively while the corresponding cores of all other proteins are colored gray. β-sheets are

depicted as thick arrows while helices are shown as rounded cylinders, Numbering indicates order and sequence trajectory. Structural elements inserted into the minimal core fold are individually labeled and colored (Inserts 1-5). The Bp and Xcc helical domains are shown as distinct inserts, colored yellow and orange respectively. C-terminal helix present in primases and maturases is shown in yellow. Conserved active site residues are labeled in each model. Abbreviations as follows: Gk,*Geobacillus kaustophilus*; Ec, *Escherichia coli*; Aa, *Aquifex aeolicus*; Dr, *Deinococcus radiodurans;* Tm, *Thermotoga maritima*; Hs, *Homo sapiens*; Sp, *Streptococcus pneumoniae*; Af, *Archaeoglobus fulgidus*; Cd, *Clostridium difficile*. Boxed inset at top right highlights structural conservation of C-terminal helix in RNase M5 maturases and DnaG primases with the aH1 helix in Bp OLD. Superposition includes the Aa RNase M5 (PDB: 1T6T) and *Staphylococcus aureus* (Sa) DnaG (PDB: 4EE1).



Supplementary Figure 6. Structural constraints of Bp^{CTR} DNA binding. a, Electrostatic surface of Bp^{CTR} shown with modeled B-form DNA. Electrostatic potential calculated with APBS (Jurrus et al., 2018). Scale indicates coloring of the potential from -3 K_bT/e_c to +3 K_bT/e^c. The four basic patches around the active site cleft are numbered in yellow. Black arrows indicate steric clashes. **b,** Modeled DNA substrate in Fig. 6a is significantly bent relative to B-form DNA.

Chapter 3. The full-length structure of *Thermus scotoductus* OLD defines the ATP hydrolysis properties and catalytic mechanism of Class 1 OLD family nucleases

The full-length structure of *Thermus scotoductus* OLD defines the ATP hydrolysis properties and catalytic mechanism of Class 1 OLD family nucleases

Carl J. Schiltz[1] and Joshua S. Chappie[1,*]
[1] Department of Molecular Medicine, Cornell University, Ithaca, NY, 14853, USA

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

ABSTRACT

OLD family nucleases are distinguished by an N-terminal ATPase domain and a C-terminal Toprim domain and can be subdivided into two classes based on size and genomic context. Although the accompanying structural study identifies the catalytic machinery controlling nuclease cleavage in Class 2 enzymes, degenerate sequence conservation between the C-termini of Class1 and Class 2 homologs precludes pinpointing the analogous side chains in Class 1 enzymes by alignment alone. Moreover, these models provide no information on the architecture ATPase domain and ATP hydrolysis. Here we present the crystal structure of a full-length Class 1 OLD nuclease from *Thermus scotoductus* (Ts) at 2.21Å resolution. The structure reveals a three domain architecture with a dimerization domain inserted into the N-terminal ABC ATPase domain and a C-terminal Toprim domain. The ATPase domain shares structural homology with genome maintenance proteins and identifies the conserved side chains responsible for Ts OLD ATPase activity. Ts OLD lacks the C-terminal helical domain present in Class 2 OLD homologs yet preserves the spatial organization of the nuclease active site, arguing that OLD proteins use a conserved catalytic mechanism for DNA cleavage.

Moreover, we demonstrate that mutants perturbing ATP hydrolysis or DNA cleavage *in vitro* significantly impair P2 OLD-mediated killing of RecBC- *E. coli* hosts, indicating that both the ATPase and nuclease activities are required for OLD function *in vivo.*

INTRODUCTION

Overcoming lysogenization defect (OLD) proteins comprise a poorly characterized family of nucleases that contain an N-terminal ATPase domain and C-terminal **Top**isomerase/**Prim**ase (Toprim) catalytic domain (1,2). OLD homologs are widely conserved in bacteria, archaea, and some viruses and can be subdivided into two classes based on size and genomic context (Schiltz et al., accompanying manuscript). Class 1 OLD homologs exist as single, isolated genes while Class 2 *old* genes always appear in tandem with a UvrD/PcrA/Rep-like helicase. The coding sequences of Class 1 proteins are also on average ~50 amino acid shorter than their Class 2 counterparts. Despite widespread prevalence across a plethora of species, little is known about the exact biological function of these enzymes. The Class 1 OLD homolog from the temperate bacteriophage P2 remains the best characterized to date. Genetic studies showed *old*$^{+}$ P2 lysogens kill *E. coli recB* and *recC* mutant hosts after infection and specifically interfere with bacteriophage λ growth (3,4). Further *in vitro* characterization of recombinant P2 OLD revealed DNA exonuclease activity and ribonuclease activity (5). A saturating genome-wide transposon screen of *Salmonella typhimurium* indicated that *old* gene is conditionally essential in some instances like temperature stress (6), but the

underlying mechanism for this phenotypic observation has yet to be clarified.

In the accompanying study (Schiltz et al., accompanying manuscript), we demonstrate that the Class 2 OLD homologs from *Burkholderia pseudomallei* and *Xanthamonas campestris* pv. *campestris* function as metal-dependent nucleases *in vitro* and describe the crystal structures of their catalytic C-terminal regions (CTRs). The Class 2 CTRs contain a Toprim domain with altered architecture and a unique helical domain. Side chains in both domains contribute to the nuclease active site and adopt a geometry that supports a two-metal catalysis mechanism for cleavage. The degenerate sequence conservation between the C-termini of Class1 and Class 2 homologs, however, precludes pinpointing the analogous side chains in Class 1 enzymes by alignment alone. Thus, it remains unclear whether Class 1 enzymes like that from P2 utilizes the same mechanism and cleavage machinery.
These Class 2 CTR models also provide no information on the architecture ATPase domain and the ATP hydrolysis machinery, as the N-terminal region (NTR) common to both Classes was removed for crystallization purposes.

Here we present the crystal structure of a full-length Class 1 OLD nuclease from *Thermus scotoductus* (Ts) at 2.21Å resolution. The structure reveals a three domain architecture with a dimerization domain inserted into the N-terminal ABC ATPase domain and a C-terminal Toprim domain. The ATPase domains share unexpected structural homology with genome maintenance proteins, which identifies the critical side chains responsible for Ts OLD ATP hydrolysis and highlights sequence variations that are unique to both classes

of OLD proteins. The orientation of the ATPase domains within the apo Ts OLD dimer differs significantly from the nucleotide-bond conformations of other DNA repair ABC ATPases, suggesting additional nucleotide-dependent conformational rearrangements may occur. Surprisingly, the Ts OLD C-terminus lacks the helical domain present in Class 2 OLD homologs yet preserves the spatial organization of the nuclease active site, arguing that OLD proteins use a conserved catalytic mechanism for DNA cleavage. Moreover, we show that mutants perturbing ATP hydrolysis or DNA cleavage *in vitro* abolish P2 OLD-mediated killing of *recBC* deficient *E. coli* hosts, indicating that both the ATPase and nuclease activities are required for OLD function *in vivo.* Together our data provide novel insights into the structure and mechanisms of Class 1 OLD proteins and further define the basic biological properties of the OLD family as a whole.

MATERIAL AND METHODS

Cloning, expression, and purification of TsOLD

DNA encoding *Thermus scotoductus* (Ts) OLD (UniProt E8PLM2) was codon optimized for *E. coli* expression and synthesized commercially by IDT. A construct containing the full-length protein (residues 1-528) was amplified by PCR and cloned into pET21b, introducing a 6xHis tag at the C-terminus. TsOLD was transformed into BL21(DE3) cells, grown at 37°C in Terrific Broth to an OD_{600} of 0.7-0.9, and then induced with 0.3 mM IPTG overnight at 19°C. Cells were pelleted, washed with nickel loading 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 frozen in liquid nitrogen and stored at -80°C.

Thawed 500 ml pellets of TsOLD constructs were resuspended in 30 ml of nickel loading buffer supplemented with 3mg DNAse, 10mM $MgCl_2$, 10 mM PMSF, and a Roche complete protease inhibitor cocktail tablet. Lysozyme was added to a concentration of 1 mg/ml and the mixture was incubated for 10 minutes rocking at 4°C. Cells were disrupted by sonication and the lysate was cleared via centrifugation at 13 000 rpm (19 685 g) for 30 minutes at 4°C. The lysate was then heated at 65°C for 15 minutes to precipitate heat-labile proteins. Precipitation was removed by centrifugation at 13 000 rpm (19 685 g) for 15 minutes at 4°C. The supernatant was filtered, loaded onto a 5 ml HiTrap chelating column charged with $NiS0_4$, and then washed with nickel loading buffer. Native TsOLD was eluted by an imidazole gradient from 30 mM to 1 M. Pooled fractions were dialyzed overnight at 4°C into Q loading buffer (20 mM Tris pH8, 50 mM NaCl, 1 mM EDTA, 5% glycerol (v:v), and 1 mM DTT). The dialyzed sample was applied to 5 ml HiTrap Q column equilibrated with Q loading buffer, washed in the same buffer, and eluted with a NaCl gradient from 50 mM to 1 M. Peak fractions were pooled, concentrated, and further purified by size exclusion chromatography (SEC) using a Superdex 200 16/600 pg column. Native TsOLD was exchanged into a final buffer of 20mM HEPES pH 7.5, 150mM KCl, 5 mM MgCl2, and 1mM DTT during SEC and concentrated to 20-30 mg/ml. Point mutations were introduced by Quikchange

(Agilent) and all mutant and truncated TsOLD constructs were purified in an identical manner as the wildtype protein.

Size exclusion chromatography coupled to multiangle light scattering

The oligomeric state of TsOLD was determined by SEC coupled to multiangle light scattering (SEC-MALS). TsOLD full-length (residues 1-525), N-terminal region (residues 1-369), and C-terminal region (residues 370-525) constructs were all loaded at 4mg/mL onto a Superdex 200 10/300 Increase column (GE) in size exclusion buffer (20 mM HEPES pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, and 1 mM DTT) at a flow rate of 0.7 ml/min. Eluent from the sizing column flowed directly to a static 18-angle light scattering detector (DAWN HELEOS-II) and a refractive index detector (Optilan T-rEX) (Wyatt Technology) with data collected every second. Molar mass was determined using the ASTRA VI software. Monomeric BSA (Sigma) was used for normalization of light scattering.

Crystallization, X-ray data collection, and structure determination

Native TsOLD at 10 mg/ml was crystallized by sitting drop vapor diffusion at 20°C in 0.1 M NaOAc pH 5.8, 0.3 M $AmSO_4$, 7% PEG MME 2000, and 5 mM $SmCl_3$ or $PrCl_3$. Crystals were of the space group $I2_12_12_1$ with unit cell dimensions a = 83.5Å, b = 101.9Å, c = 203.1Å and α = 90°, β = 90°, γ = 120° and contained a monomer in the asymmetric unit. Heavy metal derivitization was carried out in a solution of 10% PEG MME 2000 and 0.1 M NaOAc pH

5.8. Crystals were soaked in stabilizing solution containing 10 mM potassium tetracyanoplatinate ($K_2Pt(CN)_4$) for one hour. After soaking, samples were cryoprotected by transferring the crystal directly to Parabar 10312 (Hampton Research) prior to freezing in liquid nitrogen. Crystals were screened and optimized using the NE-CAT 24-ID-C and 24-ID-E beamlines. Single-wavelength anomalous diffraction (SAD) data (7) were collected remotely on the tuneable NE-CAT 24-ID-C beamline at the Advanced Photon Source at the platinum edge (λ=1.0718Å) at 100K to a resolution of 2.21Å (Table S1). Data were integrated and scaled via the NE-CAT RAPD pipeline, employing XDS (8) and AIMLESS (9) for data integration and scaling respectively. A total of 4 platinum sites were found using SHELX (10) and used for initial phasing. Density modification and initial model building was carried out using the Autobuild routines of the PHENIX package (11). Further model building and refinement was carried out manually in COOT (12) and PHENIX (11). The resulting model was refined with Rwork/Rfree values of 0.230/0.260 (Supplementary Table S1). The final model contained residues 1-149 and 152-525,13 water molecules, 3 platinum atoms, one sulfate, and one HEPES molecule.

Structural superpositions were carried out in Chimera (13) and conservation based coloring was generated using the ConSurf server (14). All structural models were rendered using Pymol (Schrodinger).

ATPase Assays

ATPase activity was characterized using a colorimetric Malachite green assay that monitored the amount of free phosphate released over time (15). The temperature dependence of Ts OLD ATPase activity was carried out in ATPase reaction buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 5 mM $MgCl_2$) with 1 mM ATP and 4 μM TsOLD. At each time point, 20 μL samples were taken and quenched with 5μL of 0.5M ETDA pH 8.0. 150 μL of filtered Malachite green solution (15) were added to each sample and incubated for 5 minutes. Samples were then visualized in a Thermo Multiskan GO spectrophotometer at 650 nm. Rates were plotted All ATPase mutant assays were carried out in ATPase reaction buffer with the 0.5 mM ATP and 4μM protein at 65ºC.

DNA cleavage assays

100 ng of lambda genomic DNA or pUC19 plasmid was incubated with 4 μM protein to a final volume of 20 μL in DNA cleavage buffer (20mM Tris-OAc pH 7.9, 50mM K-OAc, 0.1mg/mL BSA, and 5 mM divalent metal). Reactions were carried out at 65ºC for 15 minutes and then quenched with 5 μL of 0.5 M ETDA pH 8.0. Samples were analyzed via native agarose electrophoresis. Mutant constructs were assayed in the presence of 5 mM manganese.

***In vivo* characterization of P2 *old* mutants**

Electrocompetent cells of *E. coli* strain SK129 (16) were transformed with the pBAD vector carrying either wildtype or mutant P2 *old* genes and allowed to recover at 30°C to ensure wild type RecBC function. Transformed cells were

grown overnight at 30°C and dilutions of the cultures were then spotted onto LB agar supplemented with either 0.1% arabinose or 0.1% glucose. Plates were incubated overnight at 37°C to ensure reduction of RecBC function. Relative viability of the strains was determined by comparing the cfu/mL of the induced (arabinose) over the repressed (glucose) conditions.

RESULTS

TsOLD functions as a metal-dependent nuclease *in vitro*

To elucidate the architecture and function of the Class I OLD proteins, we screened a multitude of homologs to find a suitable candidate for structural and biochemical studies. Of these, only the thermophilic *Thermus scotoductus* (Ts) showed high levels of expression in *E. coli* and produced protein that remained soluble throughout purification (Figure 1A). Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) revealed that full-length TsOLD (Ts^{FL}) forms stable dimers in solution, regardless of the nucleotide state (Figure 1B, Supplementary Figure S1A,B). The N-terminal region encompassing the ATPase domain (NTR, residues 1-369) also dimerizes on its own (Supplementary Figure S1C) while the CTR containing the Toprim domain (residues 370-525) remains monomeric (Supplementary Figure S1D), consistent with the oligomeric state of Class 2 OLD CTRs (Schiltz et al., accompanying manuscript). These data suggest that the structural features mediating OLD nuclease dimerization reside in the NTR and do not require ATP binding or hydrolysis.

Toprim family enzymes catalyze metal-dependent nicking and cleavage of nucleic acid substrates (17). To determine if Ts OLD acts similarly *in vitro*, we incubated Ts^{FL} with linearized λ phage DNA in the presence of different divalent cations and monitored the disappearance of the substrate as nuclease activity was activated (Figure 1C). Ts^{FL} rapidly degrades linear DNA in the presence of Mg^{2+}, Mn^{2+}, and Ca^{2+}, with weaker activity in the presence of Co^{2+} and Zn^{2+}. This activity profile remained unchanged with the addition of ATP (Figure 1D). Ts^{CTR}, in contrast, shows a strong preference for Mn^{2+} and very weak activity with Co^{2+} (Figure 1E). These data argue that the catalytic machinery responsible for nuclease cleavage is localized to the C-terminal region of Ts OLD and that the ability to degrade linear DNA substrates does not depend on ATP binding and/or hydrolysis.

To characterize nuclease activity further, we assessed how Ts OLD acts on circular plasmids. We incubated Ts^{FL} with supercoiled pUC19 DNA (S) in the presence of different divalent metals and measured activity by the appearance of slower migrating bands as the substrate was nicked (N) and linearized (L) by the enzyme (Figure 1F). Ts^{FL} can nick and linearize supercoiled substrates with Co^{2+}, Zn^{2+}, and Ni^{2+} and exhibits processive cleavage and degradation of circular substrates with Mg^{2+}, Mn^{2+}, and Ca^{2+}. Interestingly, this processive cleavage appears to occur rapidly without an observable linearized intermediate. Addition of ATP elicits strong nicking activity with Ca^{2+}, Ni^{2+}, and Zn^{2+} along with robust processive cleavage with Mg^{2+} and Mn^{2+} (Figure 1G). Weaker process cleavage is also seen with Co^{2+} (Figure 1G). Ts^{CTR} nicks

plasmid DNA with Mg^{2+}, Mn^{2+}, Ca^{2+}, and Co^{2+}, with a linearized band visible with Mn^{2+} and Co^{2+} (Figure 1H). Degradation of circle DNA by Ts^{CTR} is only observed in the presence of Mn^{2+} (Figure 1H). These data suggest Ts OLD can function as both an endo- and exonuclease and likely prefers Mn^{2+} as a cofactor for its nuclease function.

Structural organization of TsOLD

Optimized crystals of native Ts^{FL} routinely diffracted beyond 2.5Å and structure was solved by SAD phasing using a platinum soaked derivative. The resulting model was refined to 2.21Å and contained residues 1-149 and 152-525 (Supplementary Table S1). Consistent with our SEC-MALS data, symmetry-related molecules form dimers in crystal lattice (Supplementary Figure S2). Ts^{FL} monomers are comprised of three structural domains: an N-terminal ATPase domain, a helical dimerization domain, and a C-terminal Toprim domain (Figure 2). The ATPase domain core consists of 12 β-strands organized into an eleven stranded β-sheet ordered β7-6-4-1-2-12-11-3-10-9-8 with the β5 strand packing against the end of the extended β6 strand (Figure 2A, Supplementary Figure S3A). The sheet folds at the nexus of β2 and β12, creating two segments that sandwich a central alpha helix ($\alpha1^A$). Helices $\alpha2^A$ and $\alpha3^A$ flank the sandwich, capping the end of $\alpha1^A$. The remaining five helices line the exterior, with two ($\alpha4^A$ and $\alpha5^A$) contacting the dimerization domain of the opposite monomer *in trans* (Figure 2A-C) and three ($\alpha6^A$, $\alpha7^A$,

and $\alpha 8^{A}$) interacting with the Toprim domain *in cis* (Figure 2A, 2D). The dimerization domain inserts between the $\beta 8$ and $\beta 9$ strands of the ATPase domain and consists of four helices ($\alpha 1^{D}$, $\alpha 2^{D}$, $\alpha 3^{D}$, and $\alpha 4^{D}$) that form a kinked structure (Figure 2B, Supplementary Figure S3A). $\alpha 1^{D}$, $\alpha 2^{D}$, and $\alpha 4^{D}$ from each monomer directly associate *in trans* at the dimer interface while $\alpha 3^{D}$ bends back and rests on top of α4A and α5A (Figure 2B). Together these interactions form an extensive hydrophobic network (Figure 2C) that buries 3479Å^{2} of surface area and maintains the tight association between the monomers.

Each TsOLD Toprim domain contains a canonical four-stranded parallel β-sheet ($\beta 2^{T}$-$\beta 1^{T}$-$\beta 3^{T}$-$\beta 4^{T}$) surrounded by three α-helices ($\alpha 1^{T}$, α2T, and $\alpha 3^{T}$) (Figure 2A, Supplementary Figure S3). As observed in Class 2 OLD nucleases (Schiltz et al., accompanying manuscript), the orientations of α2 and α3 are shifted relative to other Toprim domains (Supplementary Figure S3B). TsOLD, however, lacks an Insert 3 helix and contains an three additional α-helices following the terminal $\beta 4^{T}$ strand (Supplementary Figure S3A and S3C). The amphipathic $\alpha 2^{T}$ helix packs against a hydrophobic groove on the edge of ATPase domain formed by $\alpha 6^{A}$, $\alpha 7^{A}$, $\alpha 8^{A}$, and $\beta 11^{A}$ to bury 920Å^{2} (Figure 2D). A single salt bridge between R353 and E413 adds additional stabilization. The Toprim domains do not participate in dimerization, instead extending outward nearly orthogonal to two-fold symmetry axis (Figure 2A).

Ts^{FL} identifies the determinants of OLD ATP hydrolysis

TsOLD ATP hydrolysis is strongly dependent on temperature, with an optimal activity at 65°C that reflects its the thermophilic nature (Supplementary Figure S4A). We observed a modest basal turnover at this temperature with a k_{cat} of 0.44±0.3 min^{-1} and Km of 180±26 μM (Figure 3A). Despite attempts to co-crystallize with ATP analogs, only a sulfate deriving from the crystallization buffer was present in the Ts^{FL} active site (Supplementary Figure S4B). The DALI alignment algorithm (Holm and Rosenström, 2010) identifies *Methanococcus jannaschii* (Mj) Rad50 (PBD ID: 5dny-D, Z-score = 17.5, RMSD = 3.7 and the *Pyrococcus furiosus* (Pf) Smc-ScpAB complex (PDB ID: 4i99-B, Z-score = 8.2, RMSD = 3.8) as the nearest structural homologs, suggesting TsOLD is a belongs to the SMC/Rad50/RecN/RecF family of DNA repair **A**TP-**b**inding **c**assette (ABC) proteins (18). Structural superposition confirms their overall similarity and identifies residues in Ts OLD that contribute to ATP binding and hydrolysis (Figure 3B and Supplementary Figure S4C-D).

ABC ATPases share a RecA-like nucleotide binding domain (NBD) that dimerizes upon activation, sandwiching two ATP molecules at interface between the associating monomers (19). Six conserved sequence motifs – P-loop/Walker A, Walker B, the ABC signature sequence, Q-loop, D-loop, and H-loop – contribute to nucleotide binding and hydrolysis (Supplementary Figure S5). As in other ATPases, the P-loop/Walker A motif coordinates the α- and β-phosphates of ATP while the Walker B motif stabilizes an essential

magnesium cofactor and coordinates the nucleophilic water (20). The Q-loop also acts *in cis* to coordinate the magnesium and catalytic water, helping to sense the nucleotide state and orchestrate associated conformational changes (21,22). A conserved histidine in the H-loop balances the negative charge of the aspartate/glutamate in Walker B and the γ-phosphate in the transition state (23-25). The signature sequence (LSGGQ) characteristic of all ABC proteins and the D-loop both act *in trans*, where the former contacts the nucleotide and γ-phosphate of the adjacent monomer and the latter serves a dual function in positioning the nucleophilic water molecule and stabilizing the Walker A motif (26-28).

Ts OLD contains a number of sequence variations across these catalytic motifs compared to other SMC/Rad50/RecN/RecF family ABC proteins (Figure 3B, Supplementary Figure S5B). Although K34 and H310 structurally align with conserved P-loop and H-loop side chains, a glutamate (E276) replaces the Walker B aspartate and histidines H140 and H283 substitute for the conserved glutamine and aspartate side chains in the Q-loop and D-loop respectively. These deviations from canonical ABC consensus sites vary in their conservation across other Class 1 OLD homologs (Supplementary Figure S6). For example, H140 in the Q-loop is poorly conserved, with many species having an arginine at this position. Exchange of an asparagine for H310 in the H-loop also occurs at a similar frequency. The remaining residues are largely conserved. Notably, the ABC signature sequence is degenerate in all Class 1 OLD homologs and also absent from the Class 2 homolog from *Burkholderia*

pseudomallei (Supplementary Figure S6).

To elucidate the functional relevance of this putative catalytic machinery, we mutated each residue individually to alanine and tested the effect on ATP hydrolysis. The K34A, H140A, E276A, and H310A substitutions each completely abolish ATPase activity (Figure 3C), underscoring their crucial catalytic function. The D-loop substitution H283A decreases the hydrolysis rate more than 4-fold, suggesting a moderate contribution to the enzymatic activity.

Ts OLD ATPase domains adopt a non-productive conformation

The crystal structures of different SMC/Rad50/RecN/RecF family proteins show that head-to-head dimerization of the NBDs properly positions the catalytic machinery for productive ATP hydrolysis (Supplementary Figure 75A-C) (29-31). In this arrangement, the ABC signature sequence, D-loop, and H-loop are oriented toward the active site of opposing monomer in close approach to P-loop and bound nucleotide. The organization of the ATPase domains in the Ts^{FL} dimer differs from this configuration and would require a significant rotation of one of the subunits to bring the active sites into alignment (Figure 4, Supplementary Figure S7D). This suggests that Ts^{FL} conformation captured in the crystal is non-productive and could represent an intermediate along the ATP hydrolysis reaction pathway.

OLD proteins share a conserved active site organization

Our accompanying structural study (Schiltz et al., accompanying manuscript)

identifies the catalytic machinery of Class 2 OLD proteins and supports a two-metal mechanism for nuclease cleavage. Class 1 proteins are smaller and show significant sequence divergence in the C-terminus relative to their Class 2 counterparts, suggesting possible architectural and mechanistic differences. Superposition of the Ts^{FL} and Bp^{CTR} models reveals a high degree of structural similarity between the Toprim core in both proteins (RMSD: 0.558Å) (Figure 5A). Importantly, the positions of the α2 and α3 helices relative to the central β-sheet remain unchanged between both classes but distinct from every other Toprim containing protein (Supplementary Figure S3B). Ts OLD, however, shows two key structural differences: it lacks both an Insert 3 helix between α3 and β4 that is present in Class 2 OLD proteins and numerous topoisomerases (Schiltz et al., accompanying manuscript) and a C-terminal helical domain (Figure 5A, Supplementary Figures S3 and S6). Despite these topological differences, the critical catalytic side chains responsible for Class 2 nuclease activity are spatially conserved in Ts OLD (Figure 5B). Ts residues E377, D431, and D433 align with the Bp OLD metal A binding residues E400, D455, and D457 while D381, S478, E480 superimpose with the Bp metal B ligands E404, T506, and E508. These putative Ts OLD metal binding residues in are conserved among all Class 1 homologs (Figure 5C, Supplementary Figure S6). Mutation of the these side chains to alanine (E377/D431A/D433A and D381A/S478A/E480A) completely abolishes processive degradation of either linear or supercoiled DNA substrates (Figure 5D,E). As with Class 2 homologs, metal A and metal B mutants still retain the ability to nick and

linearize circular plasmids (Figure 5E; Schiltz et al., accompanying manuscript).

K562 also plays a crucial role in Bp OLD nuclease function, where it acts to stabilize the developing negative charge in the transition state of the phosphoryl transfer reaction and/or protonate the leaving group (Schiltz et al., accompanying manuscript). Superposition shows that guanidinium group of Ts R487 aligns with the amino group of Bp K562. The ariginine sidechain approaches the active site from a different angle, as it residues on in the loop preceding $\alpha 5^{T}$ instead of a separate helical domain. A conserved arginine or lysine residue is present at this position in all Class 1 OLD homologs (Supplementary Figure S6). To validate the functional importance of this residue, we mutated R487 to either an alanine or glutamate and tested how each affected nuclease activity on either linear and supercoiled DNA. R487E completely abrogates processive DNA degradation of both substrates while R487A only partially inhibits (Figure 5D,E). Together, these data argue that Class 1 and Class 2 OLD nucleases use a conserved set of catalytic machinery and a similar two-metal mechanism to cleave DNA.

ATPase and nuclease mutations impair P2 OLD activity *in vivo*

To understand the importance of OLD ATPase and nuclease activities *in vivo*, we leveraged the observation that *old*$^{+}$ P2 lysogens kill *E. coli* harboring *recB* and *recC* mutations (3). The *E. coli* strain SK129 contains temperature-sensitive alleles of both *recB* and *recC* such that RecBC function is impaired at

37°C (16). Overexpression of wildtype P2 OLD from an arabinose inducible pBAD vector mimics the phage phenotype and kills this host strain at the 37°C with a cell viability of 0.003% (Figure 5A). Host killing is supressed at the 30°C where RecBC is fully functional (Supplementary Figure S8). SK129 *E. coli* similarly survive at either temperature when glucose is substituted for arabinose (Figure 6A).

Using this experimental background, we next examined the effect of ATPase and nuclease mutants on P2-mediated host killing. P2 OLD is a Class 1 enzyme like that of Ts, and we identified its putative catalytic side chains based on sequence alignment (Supplementary Figure S6). Mutant constructs containing substitutions in the critical ABC ATPase motifs (K37A, Walker A; N299A, Walker B; H306A, D-loop; H332A, H-loop) and the Toprim metal binding and nuclease residues (E398A, D455A, and D457A, metal A; E402, D541, and E543, metal B; K550A) were cloned into pBAD and overexpressed with arabinose in SK129 *E. coli* at 37°C. Every mutation abolished the lethal effects of P2 OLD expression in the SK129 background and restored the viability at the non-permissive temperature to wildtype levels (Figure 6B). Although D541A rescued viability (Figure 6B), expression of this mutant causes a noticeable slow growth phenotype (Figure 6A). Together these data indicate that the both the ATPase and nuclease activities of P2 OLD are required for function *in vivo* and necessary for host killing when the RecBC recombinational repair pathway is impaired.

DISCUSSION

Here we have described the full-length structure of the Class 1 OLD homolog from *Thermus scotoductus*, which reveals a three domain architecture consisting of an N-terminal ABC ATPase domain, a dimerization domain, and a C-terminal Toprim domain. Our biochemical experiments show that TsOLD functions as a metal-dependent nuclease *in vitro* and exhibits temperature-dependent ATP hydrolysis, with maximal turnover at 65°C.

The TsOLD ATPase domain is topologically related to the NBDs of genome maintenance proteins (Supplementary Figure S4C-D). This structural homology allowed us to pinpoint key catalytic residues that are required for ATP hydrolysis (Figure 3B,C). Importantly, Ts OLD contains a number of variations that deviate from the canonical ABC sequence motifs (Supplementary Figure S5B). These include histidines replacing the conserved glutamine and aspartate side chains in the Q-loop and D-loop respectively and the absence of a defining ABC signature sequence. Although mutation of these side chains abrogates TsOLD ATP hydrolysis *in vitro*, some of these variants are not strictly conserved among Class 1 OLD nucleases (Supplementary Figure S6). It remains to be seen whether homologs like *Stroptococcus suis*, which has an alanine in place of the histidine in the D-loop, retain the ability to hydrolyze ATP. Interestingly, the Class 2 OLD homolog from *Burkholderia pseudomallei* contains a valine at the position corresponding to H140 in the Ts Q-loop and shows no measurable ATP hydrolysis *in vitro* (Schiltz and Chappie, unpublished data). Further

biochemical characterization of representatives from both Classes will be necessary to tweeze out subtle mechanistic differences.

Dimerization of the nucleotide binding domains and subsequent ATP hydrolysis triggers concomitant conformational changes regulate the biological function of ABC ATPases. The cycle of ATP binding, hydrolysis, and release of ADP controls conformational rearrangements in ABC transporters that pump substrates into or out of cells (32). For many recombination and repair proteins, ATPase activity is coupled to nucleic acid substrate recognition and the licensing of nicking and/or cleavage activities that initiate downstream events (31,33-34). Ts^{FL} dimerizes both in solution (Figure 1B, Supplementary Figure S1) and in the crystal lattice (Supplementary Figure S2), with the extensive hydrophobic interface between dimerization domain helices providing the primary stabilizing interactions (Figure 2C). While extended helical coils help to tether monomers in many repair family ABC proteins, productive ATP hydrolysis requires a head-to-head engagement of the NBDs to orient critical catalytic motifs *in cis* and *in trans* (Supplementary Figure S5 and S7). The orientation of the ATPase domains in Ts^{FL} is not compatible with this arrangement, as one monomer is significantly rotated (Figure 4). Whether this represents a non-productive conformation or an intermediate along the hydrolysis reaction pathway remains to be seen. This does, however, imply that Ts OLD must undergo a major conformational to achieve productive turnover. Further structural characterization of TsOLD with different nucleotide analogs will clarify these discrepancies.

The CTRs of Class 2 OLD homologs consist of a Toprim domain with altered architecture and a unique helical domain. Conserved side chains in both domains contribute to the nuclease active site and adopt a geometry that supports a two-metal catalysis mechanism for cleavage (Schiltz et al., accompanying manuscript). Ts OLD shares the altered Toprim domain (Supplementary Figure S3B) but lacks a C-terminal helical extension (Figure 5A). Structural superposition, however, shows that Ts OLD retains the critical metal A and metal B binding residues and has an arginine that could function like K562 in Bp OLD to stabilize developing negative charge in the transition state of the phosphoryl transfer reaction and/or protonate the leaving group (Figure 5B). These residues are conserved among Class 1 homologs, suggesting all OLD family nucleases cleave nucleic acid substrates in the same manner.

Our structural studies represent a major advance in understanding the enzymatic functions of OLD proteins. Despite the novel insights we provide regarding the underlying catalytic mechanisms and machinery, we still know very little about OLD function *in vivo*. Much of our current knowledge derives phage genetics, which established that the P2 *old* gene is dominant, interferes with normal bacteriophage λ growth when present in lysogens, and mediates killing of so-called 'lysogenization defective' *E. coli* hosts that are impaired in recombinational repair (3,4). Here were have demonstrated that P2-mediated killing of these *E. coli recB* and *recC* mutant strains requires both the ATPase and nuclease activities of P2 OLD (Figure 6). We anticipate that these

activities will also be required for the λ interference phenotype. Future genetic experiments examining the behaviour of OLD ATPase and nuclease mutants both in this context as well as in other systems will help to unravel the biological functions of this diverse family of proteins.

ACCESSION NUMBERS

The atomic coordinates and structure factors of the Ts^{FL} structure are deposited in the Protein Databank with the accession number XXXX.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

We thank Drs. Ailong Ke, Robert Weiss, and John Helmann for critical reading of the manuscript and the Northeastern Collaborative Access Team (NE-CAT) beamline staff at the Advanced Photon Source (APS) for assistance with remote X-ray data collection. We thank 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 NE-CAT beamlines (24-ID-C and 24-ID-E) under the under the general user proposals GUP-51113 and GUP-41829 (PI: J.S.C.). 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 departmental funds.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

All experiments were conceived by Schiltz and Chappie. Schiltz carried out all data acquisition. Results were analyzed and interpreted by Schiltz and Chappie. The initial draft of the manuscript was written by Schiltz with critical revisions by Chappie.

REFERENCES

1. 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.

2. 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.

3. Sironi,G. (1969) Mutants of Escherichia coli unable to be lysogenized by the temperate bacteriophage P2. Virology, 37, 163-176.

4. Lindahl,G., Sironi,G., Bialy,H., and Calendar, R. (1970) Bacteriophage lambda: abortive infection of bacteria lysogenic for phage P2. Proc. Nati. Acad. Sci. U.S.A., 66, 587-594.

5. Myung,H., and Calendar R. (1995) The old exonuclease of bacteriophage P2. J Bacteriol., 177, 497-501.

6. 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.

7. Hendrickson,WA. (2014) Anomalous diffraction in crystallographic phase evaluation. Q. Rev. Biophys., 47, 49-93.

8. Kabsch,W. (2010) Integration, scaling, space-group assignment and post-refinement. Acta Cryst. D66,133-144.

9. Evans,P.R. (2006) Scaling and assessment of data quality. Acta Cryst., D62, 72-82.

10. Sheldrick,G.M. (2008) A short history of SHELX. Acta Cryst., A64, 112-122.

11. 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.

12. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. Acta. Crystallogr., D66, 486-501.

13. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C., and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J.
Comput. Chem., 25,1605-1612.

14. 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, W344-50.

15. Leonard,M., Song,B.D., Ramachandran,R., and Schmid,S.L. (2005) Robust colorimetric assays for dynamin's basal and stimulated GTPase activities. Methods Enzymol., 404, 490-503.

16. Kushner,S.R. (1974) In vivo studies of temperature-sensitive recB and recC mutants. J. Bacteriol., 120, 1213-1218.

17. Yang, W. (2011) Nucleases: diversity of structure, function and mechanism. Q. Rev. Biophys., 44, 1-93.

18. Hopfner,K.P. (2016) Invited review: Architectures and mechanisms of ATP binding
cassette proteins. Biopolymers, 105, 492-504.

19. Ye,J., Osborne,A.R., Groll,M., and Rapoport,T.A. (2004) RecA-like motor ATPases – lessons from structure. Biochim Biophys Acta Bioenerg, 1659, 1-18.

20. Davidson,A.L., Dassa,E., Orelle,C., and Chen,J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol. Mol. Biol. Rev., 72, 317-364.

21. Zolnerciks,J.K., Akkaya,B.G, Marjolein,S., Chiba,P., Seelig,A., and Linton,K.J. (2014) The Q-loops of the human multidrug resistane transporter ABCB1 are necessary to couple drug binding to the ATP catalytic cycle. FASEB, 28, 4335-4346.

22. Boswell,Z.K., Rahman,S., Canny,M.D., Latham,M.P. (2018) A dynamic allosteric pathway underlies Rad50 ABC ATPase function in DNA repair. Sci. Rep., 8, 1639.

23. Kerppola, R.E., Shyamala,V.K., Klebba,P., and Ames,G.F. (1991) The membrane-bound proteins of periplasmic permeases form a complex. Identification of the histidine permease HisQMP complex. J. Biol. Chem., 266, 9857-9865.

24. Zaitseva,J., Jenewein,S., Jupmertz,T., Holland,I.B., and Schmitt,L. (2005). H662 is the lynchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. EMBO J., 24, 1901-1910.

25. Oldham,M.L. and Chen,J. (2011) Snapshots of the maltose transporter during ATP hydrolysis.
Proc. Natl. Acad. Sci. USA, 108, 15152-15156.

26. Moncalian,G., Lengsfeld,B., Bhaskara,V., Hopfner,K.P., Karcher,A., Alden,E., Tainer,J.A., and Paull, T.T. (2004) The rad50 signature motif: essential to ATP binding and biological function. J. Mol. Biol., 335, 937-51.

27. Jones,P.M. and George,A.M. (2012) Role of the D-loops in allosteric control of ATP hydrolysis in an ABC transporter. J. Phys. Chem. A., 116, 3004-3013.

28. De La Rosa,M.B. and Nelson,S.W. (2011) An interaction between the Walker A and D-loop motifs is critical to ATP hydrolysis and cooperativity in bacteriophage T4 Rad50. J. Biol. Chem., 286, 26258-26266.

29. Liu,Y., Sung,S., Kim,Y., Li,F., Gwon,G., Jo,A., Kim,A.K., Kim,T., Song,O.K., Lee,S.E., and Cho,Y. (2016) ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. EMBO J., 35,743-758.

30. Diebold-Durand,M.L., Lee,H., Ruiz Avila,L.B., Noh,H., Shin,H.C., Im,H., Bock,F.P., Bürmann,F., Durand,A., Basfeld,A., Ham,S., Basquin,J., Oh,B.H., and Gruber,S. (2016) Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. Mol. Cell., 67, 334-347.

31. Tang,Q., Liu,Y.P., Shan,H.H., Tian,L.F., Zhang,J.Z., and Yan,X.X. (2018) ATP-dependent
conformational change in ABC-ATPase RecF serves as a switch in DNA repair. Sci. Rep. 8, 2127.

32. Hollenstein,K., Dawson,R.J., and Locher,K.P. (2007) Structure and mechanism of ABC transporter proteins. Curr. Opin. Struct. Biol., 17, 412-418.

33. Stracy,M., Jaciuk,M., Uphoff,S., Kapanidis,A.N., Nowotny,M., Sherratt,D.J., and Zawadzki,P. (2016) Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. Nat. Commun., 7,12568.

34. Deshpande,R.A., Lee,J.H., and Paull,T.T. (2017) Rad50 ATPase activity is regulated by DNA ends and requires coordination of both active sites. Nucleic Acids Res. 45, 5255-5268.

Figures

Schiltz and Chappie Figure 1

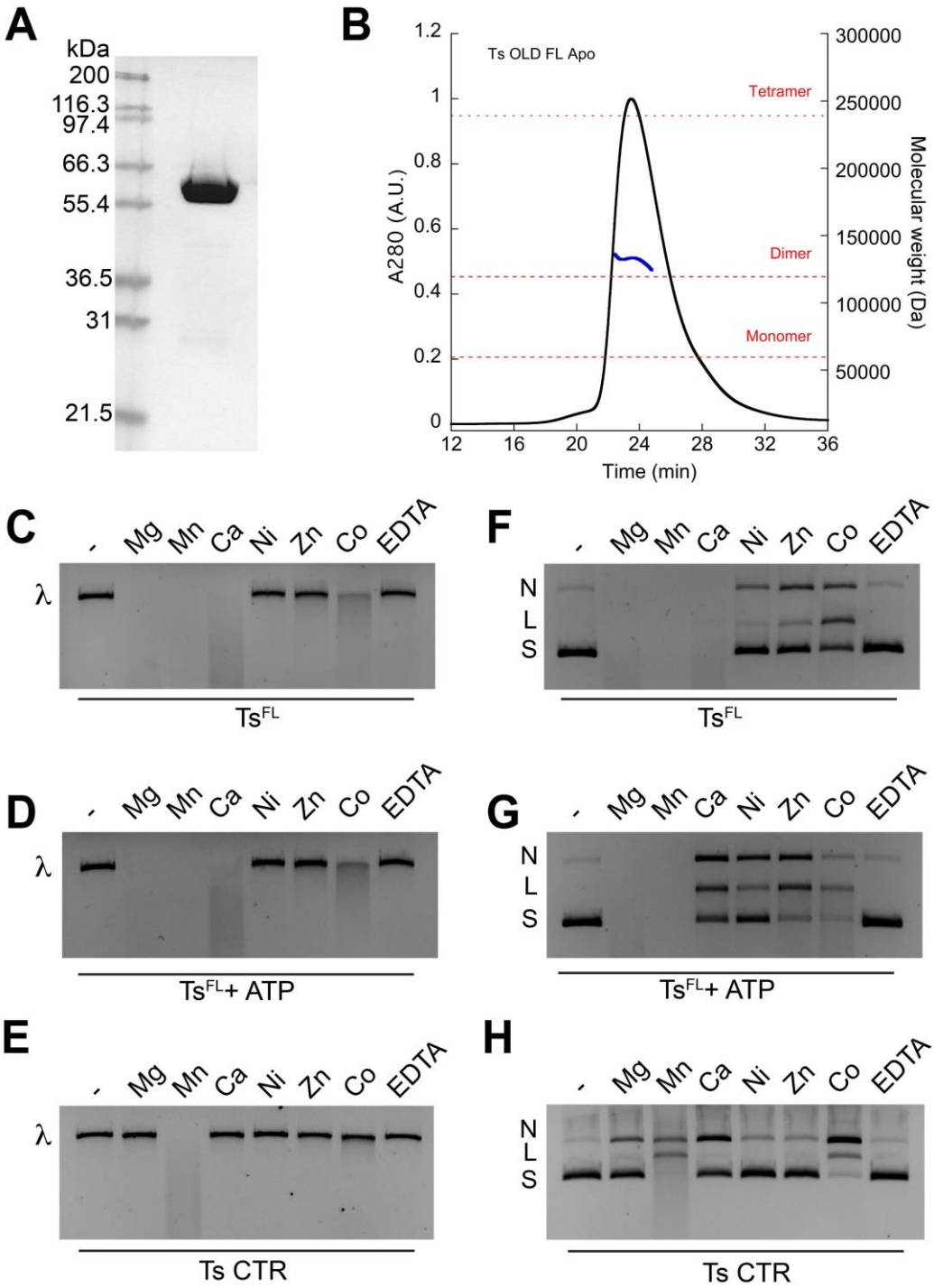


Figure 1. Purification and biochemical characterization of TsOLD. A. SDS-PAGE gel showing purified full-length TsOLD (Ts^{FL}). **B.** SEC-MALS

analysis of Ts^{FL}. UV trace (black) and calculated molecular weight based on light scattering (blue) are shown. **C-D.** Metal-dependent nuclease activity of Ts^{FL} in the absence (c) or presence (d) of ATP on linear λ DNA. **E.** Metal-dependent nuclease activity of TsOLD C-terminal region (Ts^{CTR}) on linear λ DNA. **F-G.** Metal-dependent nuclease activity of Ts^{FL} in the absence (g) or presence (h) of ATP on supercoiled pUC19 DNA. N', 'L', and 'S' denote the positions of 'nicked', 'linearized', and 'supercoiled' DNA respectively.**H.** Metal-dependent nuclease activity of TsOLD C-terminal region (Ts^{CTR}) on supercoiled pUC19 DNA.

Schiltz and Chappie Figure 2

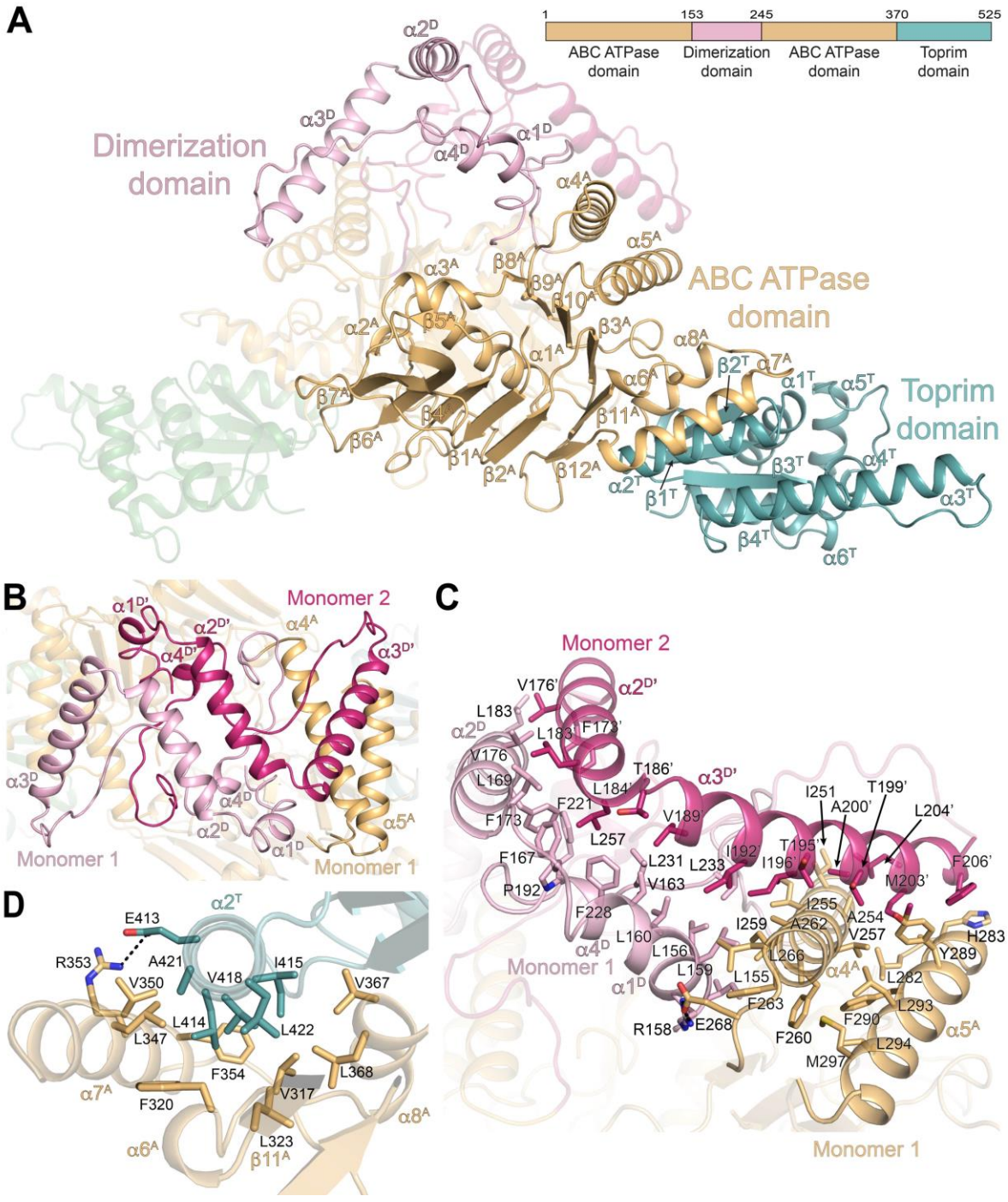


Figure 2. Architectural organization of full-length Ts OLD. **A.** Structure of TsOLD dimer. Coloring is as follows: ATPase domain orange; dimerization domain, light pink; Toprim domain, teal. Dimer is oriented orthogonal to the two-fold axis with one subunit rendered transparent for clarity. Cartoon of

domain the architecture with numbered boundaries is included on the right. Secondary structure elements are numbered sequentially within each domain with the superscripts ‘A’, ‘D’, and ‘T’ denoting the ATPase, dimerization, and Toprim domains respectively (See Supplementary Figure S3A). **B.** Top down wiew of dimerization domain helices. Helices Monomer A and Monomer B are colored light pink and warm pink respectively. **C.** Dimerization domain interactions. Residues contributing to the dimerization interface are shown as sticks and labeled. **D.** Interactions stabilizing the ATPase-Toprim interface. Contributing side chains are shown as stickes and labeled. Dashed black line denotes salt bridge between R353 and E413.

Schiltz and Chappie Figure 3

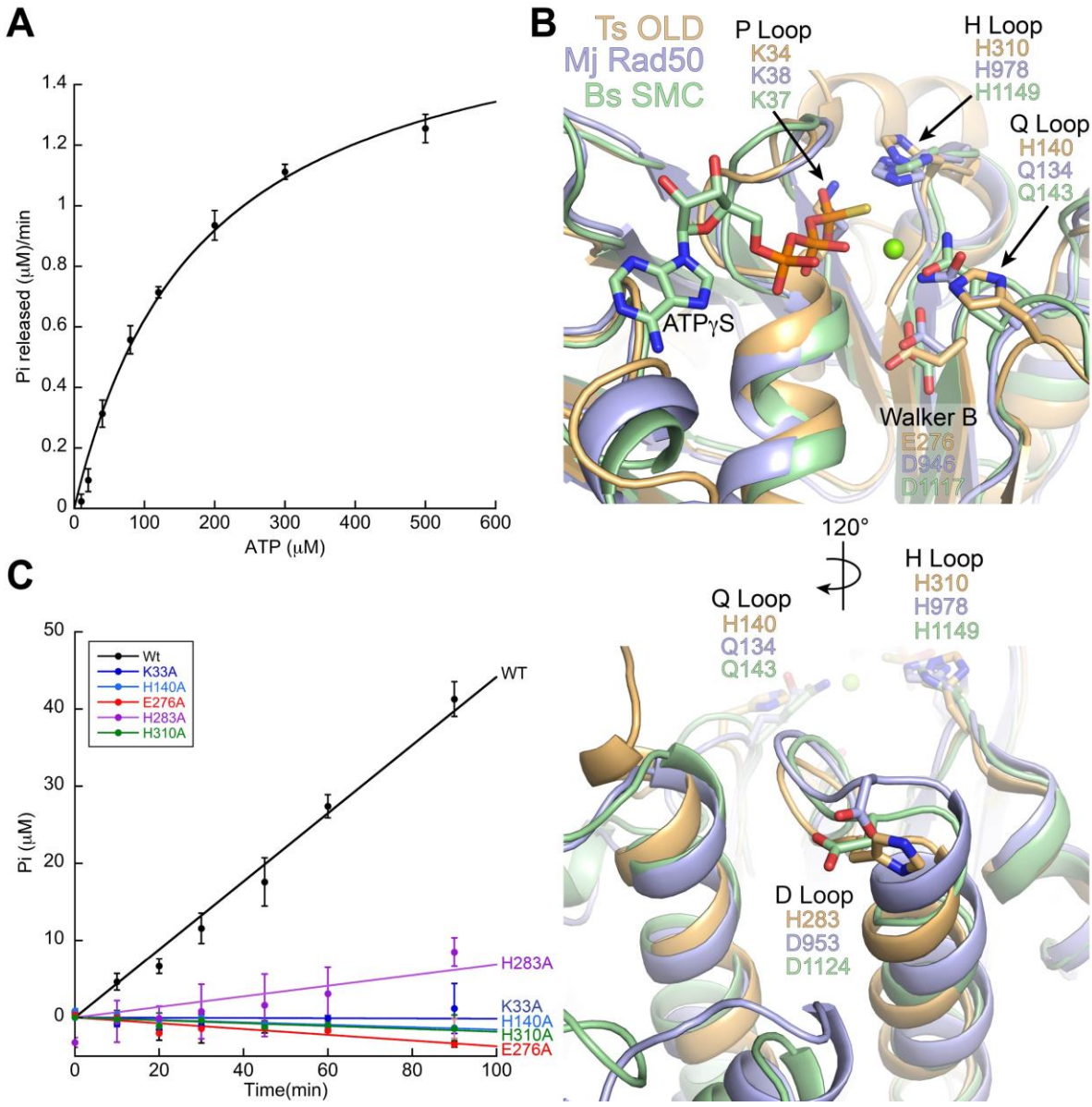


Figure 3. Catalytic determinants of Ts OLD ATP hydrolysis. **A**. Michaelis-menten kinetics of TsFL. **B.** Structural superposition of the ATPase active site loops from Ts OLD (orange), *Methanococcus jannashii* Rad50 (light blue; PDB: 5DNY), and *Bacillus subtilis* SMC (light green; PDB: 5H66). P-loop, Walker B, Q-loop, D-loop, and H-loop are labelled with the identity and position of the catalytic residues for each protein species. **C.** Initial ATP

hydrolysis rates of Wt Ts OLD compared to mutants with single alanine substitutions for each of the residues involved in ATP binding and hydrolysis.

Schiltz and Chappie Figure 4

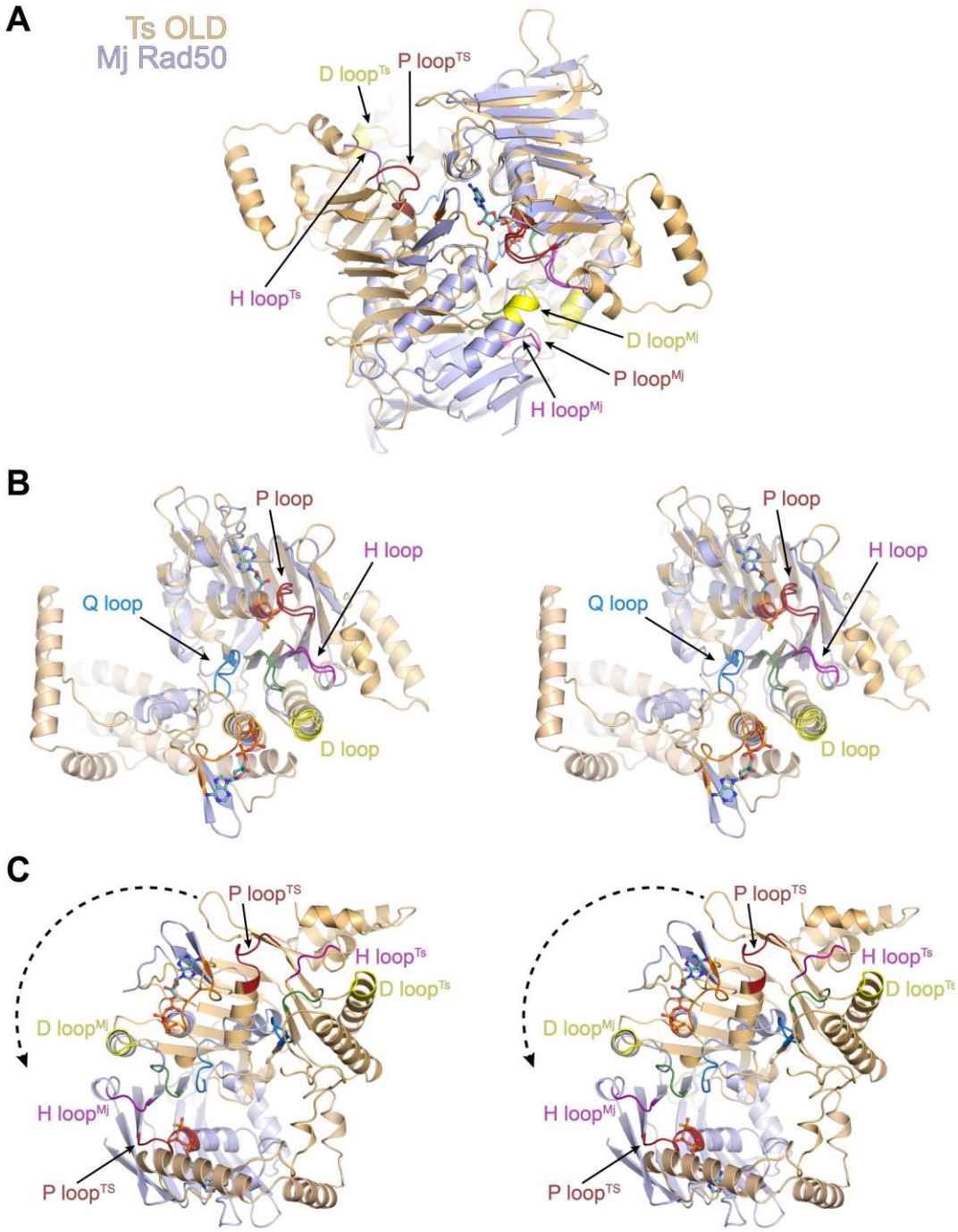


Figure 4. Ts OLD ATPase domains adopt a non-productive conformation. **A.** Structural superposition of the ATPase domains from Ts OLD (orange) and *Methanococcus jannashii* Rad50 (light blue; PDB: 5DNY). Positions of the P-loop (red), D-loop (Yellow), and H-loop (purple) in the non-aligned ATPase domains are labeled. **B.** Stereo view of aligned ATPase domains. Positions of the P-loop, Q-loop (blue), D-loop, and H-loop are labeled. **C.** Stereo view of non-aligned ATPase domains. Positions of the P-loop, D-loop, and H-loop in each model are labeled. Dashed black line indicates rotation required to align the second TsATPase domain with the hydrolysis-competent conformation of MjRad50.

Schiltz and Chappie Figure 5

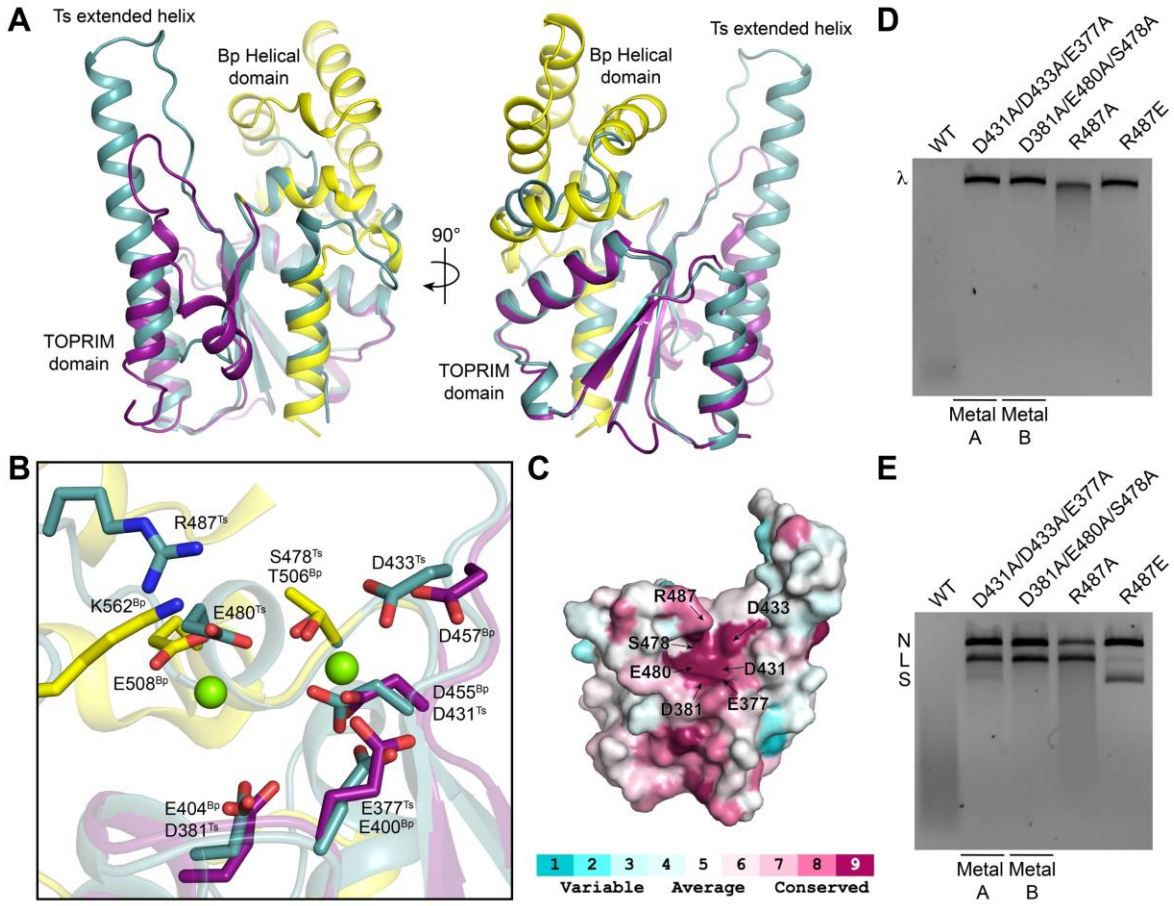


Figure 5. Class 1 and Class 2 OLD proteins share a conserved mechanism for nuclease cleavage. A. Structural superposition of OLD CTR structures from Ts (green) and Bp (yellow and purple). **B.** Zoomed view of active sites identifying catalytic machinery required for nuclease function. Coloring as in A. Bound magnesium ions in Bp^{CTR} are shown as green spheres. **C.** Conservation of active site residues among Class 1 OLD homologs. Coloring generated using the ConSurf server (Ashkenazy et al., 2016) and the alignment in Supplementary Figure S5. **D.** Cleavage activities of TsOLD active site mutants on linear λ DNA. **E.** Nicking and cleavage activities of TsOLD mutants on supercoiled pUC19 DNA. All cleavage assays were performed in the presence of manganese as described in the Methods.

Schiltz and Chappie Figure 6

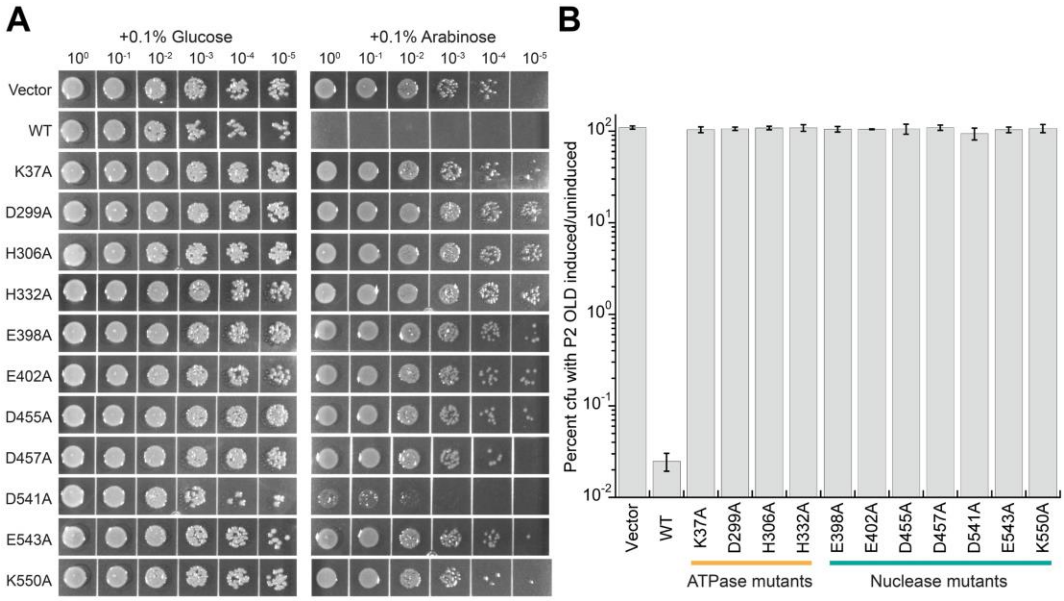


Figure 6. ATPase and nuclease activities are required for P2 OLD function *in vivo*. A Spot assay of *E. coli* carrying a temperature-sensitive *recBC*ts allele transformed with arabinose-inducible P2 OLD wild type as well as P2 OLD carrying mutations in the ATPase and nuclease domains. Strains were grown at 37°C overnight, which is not permissive for the RecBCts function. P2 OLD was induced with 0.1% arabinose or repressed with 0.1% glucose. **B.** Quantification of colony forming units under P2 OLD induction compared to P2 repression.

Supplementary information for:

The full-length structure of *Thermus scotoductus* OLD defines the ATP hydrolysis properties and catalytic mechanism of Class 1 OLD family nucleases

Carl J. Schiltz[1] and Joshua S. Chappie[1,*]

[1] Department of Molecular Medicine, Cornell University, Ithaca, NY, 14853, USA

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

Supplementary Tables

Supplementary Figures

to Figure 3.

Figure S6. Sequence alignment of Class1 OLD family nucleases. Related to Figures 3 and 5.

Figure S7. Organization of ATPase domain dimers. Related to Figure 4.

Figure S8. Spot assay of P2 OLD wildtype and mutants in RecBC^{ts} *E. coli* at 30°C. Related to Figure 6.

Table S1. X-ray data collection and refinement statistics.

	Ts^{FL} Pt PDB: XXXX
Data collection	
Space group	$I2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	83.36, 101.74, 202.75
α, β, γ (°)	90, 90, 90
Resolution (Å)	101.57-2.21 (2.27-2.21)
R_{sym} or R_{merge}	0.095 (1.935)
R_{meas}	0.097
$CC_{1/2}$	0.999
$I / \sigma I$	21.4 (1.1)
Completeness (%)	99.9 (66.4)
Redundancy	25.4 (11.4)
Phasing	
Initial F.O.M.	0.65
Number of sites	4
Refinement	
Resolution (Å)	101.57-2.21
No. reflections	47994
R_{work} / R_{free}	0.230/0.260
No. atoms	
Protein	4153
Ligand/ion	23
Water	13
B-factors	
Protein	97.02
Ligand/ion	170.15
Water	74.42
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.4
Ramachandran statistics	
Favored (%)	95.60
Allowed (%)	3.63
Outliers (%)	0.76

*Values in parentheses are for highest-resolution shell. Each dataset was derived from a single crystal.

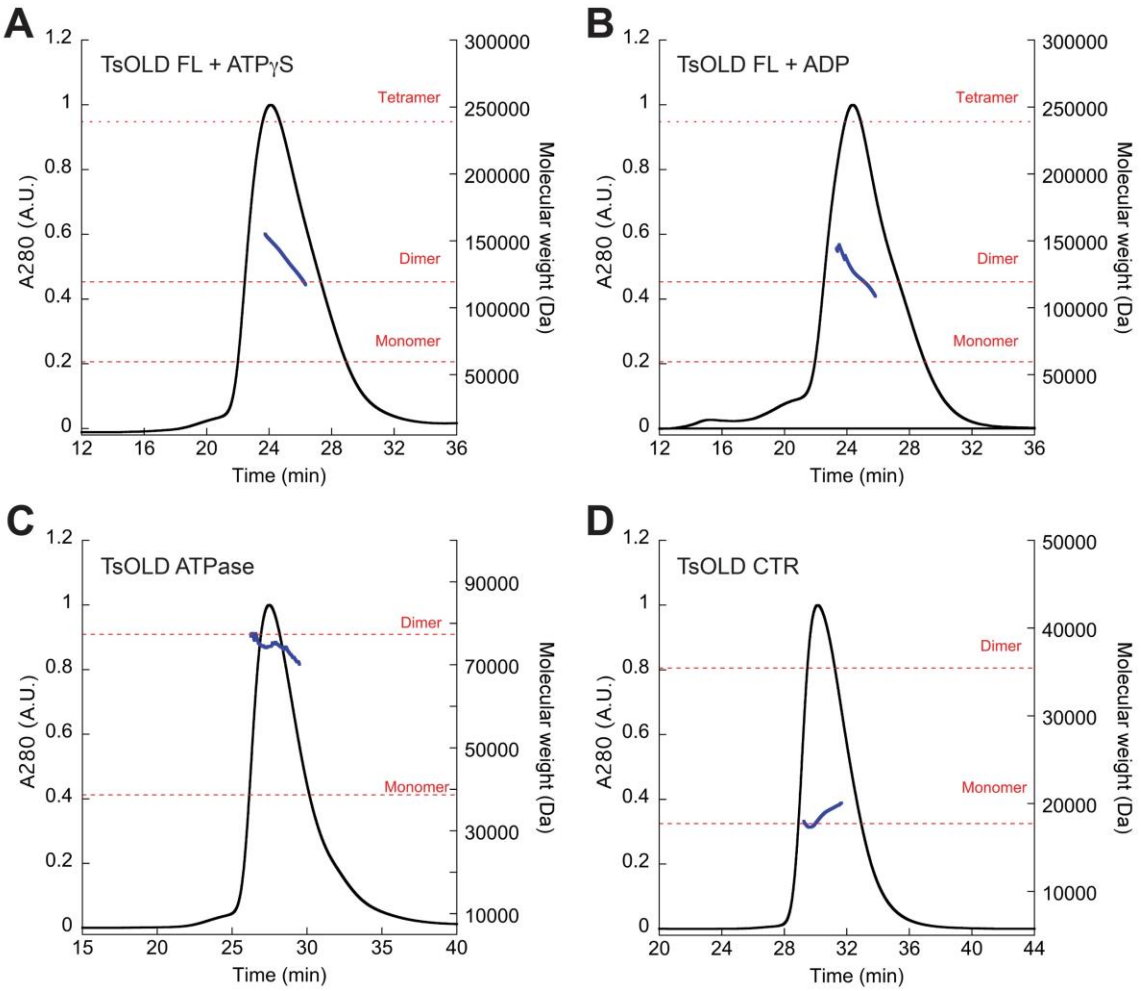


Figure S1. SEC-MALS analysis of Ts OLD constructs. Construct abbreviations are as follows: FL, full-length (residues 1-525); NTR, N-terminal region (residues 1-369); CTR, C-terminal region (residues 370-525). UV trace (black) and calculated molecular weight based on light scattering (blue) are shown. Dashed red lines denote the molecular weight predicted for a monomer, dimer, and/or tetramer of each construct.

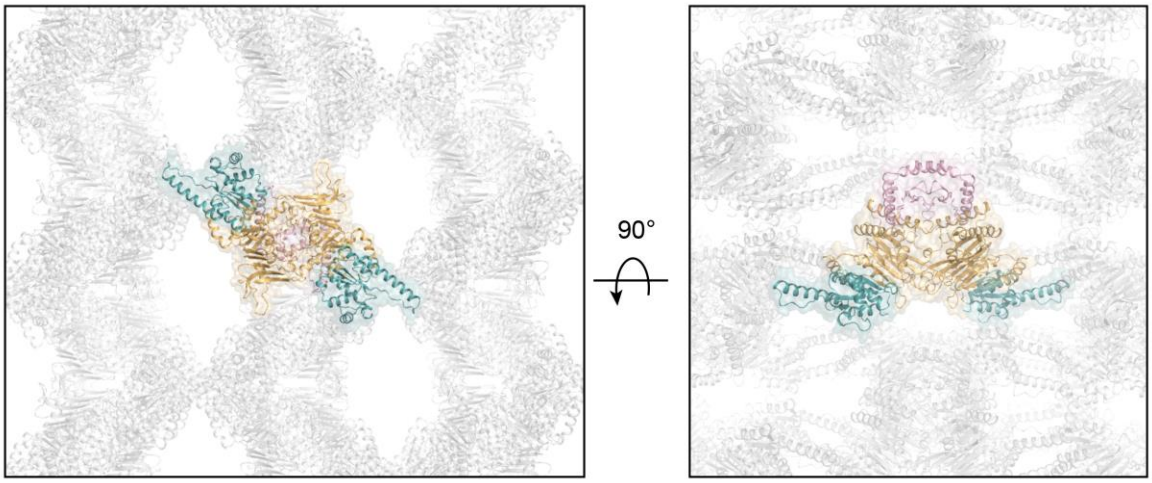


Figure S2. Crystal packing of Ts^{FL}. Individual molecules in the crystal lattices are colored gray with a single dimer highlighted and colored with ATPase domains in light orange, dimerization domains in pink, and Toprim domains in teal.

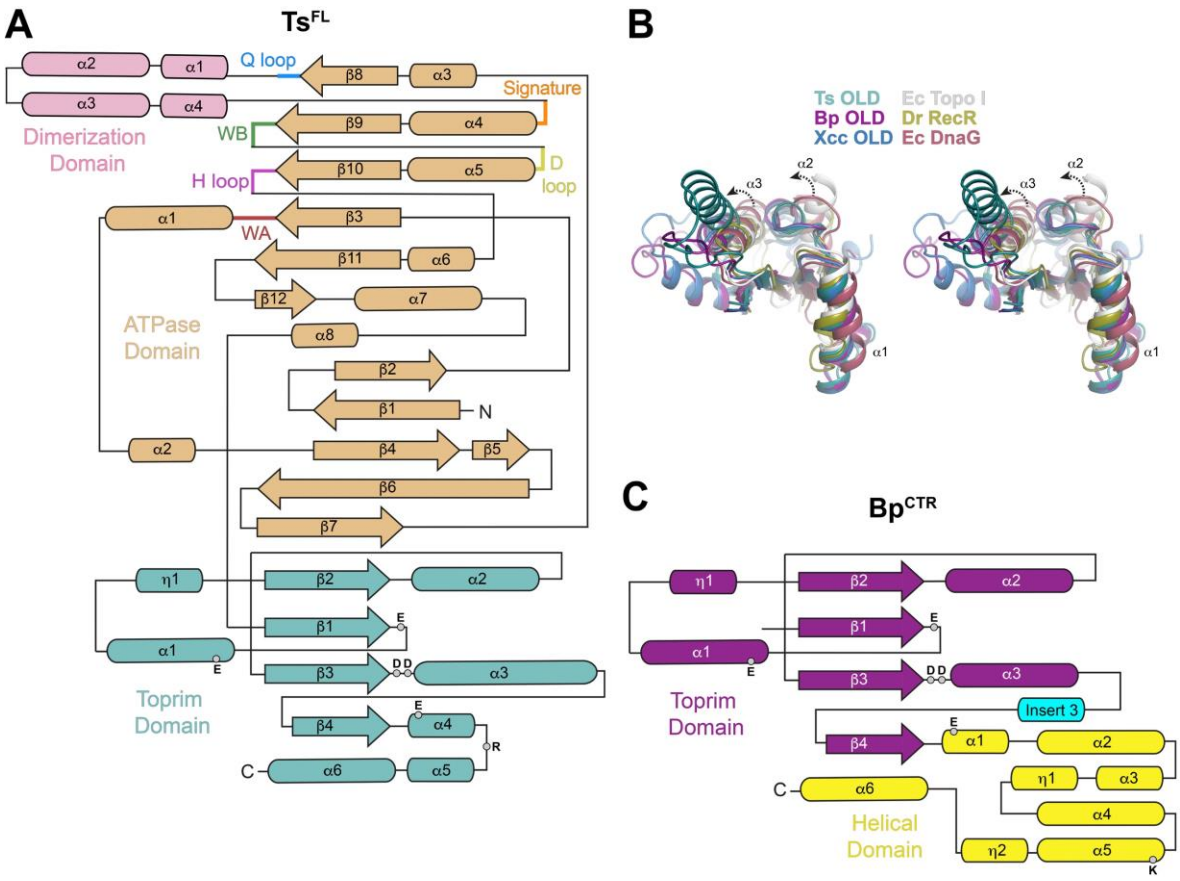


Figure S3. Topology of OLD family nucleases. A. Topology diagram of Ts^{FL} monomer. Coloring as follows: ABC ATPase domain, light orange; dimerization domain, pink; Toprim domain; teal; Ploop/Walker A (WA), red; Q loop, blue; Signature sequence, orange; Walker B (WB), green; D loop, yellow; H loop, purple. Location of Toprim catalytic residues responsible for nuclease cleavage are noted. **B.** Stereo view of structurally conserved α2 and α3 helical shifts (dashed arrows) in OLD family Toprim cores relative to the Toprim central β-sheet. Toprim cores for the following are shown for comparison: Ts OLD, teal; *Burkholderia pseudomallei* (Bp) OLD (PDB: 6NK8), dark purple; *Xanthamonas campestris* p.v. *campestris* (Xcc) OLD (PDB: 6NJW), blue; *Escherichia coli* (Ec) Topoisomerase I (PDB: 1MW9), gray; *Deinococcus radiodurans* (Dr) RecR (PDB: 1VVD), olive; *Escherichia coli* DnaG (PDB: 3B39), raspberry. **C.** Topology diagram of Bp^{CTR}. Toprim and helical domains

are colored dark purple and yellow respectively. Location of catalytic residues responsible for nuclease cleavage are noted. ‘Insert 3’ helix is colored cyan.

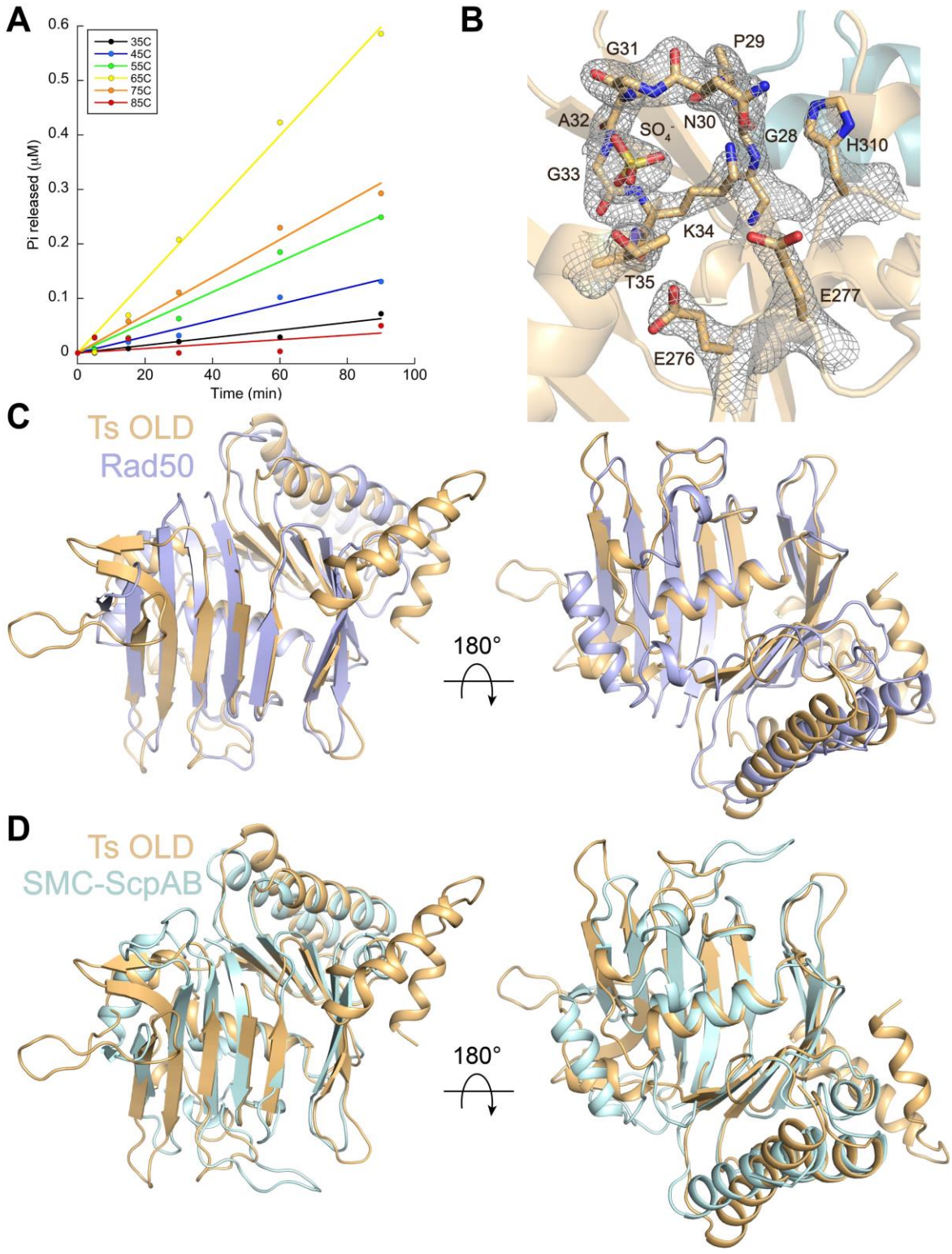


Figure S4. Ts OLD ATPase domain shares structural homology with genome maintenance proteins. A. Temperature dependence of Ts^{FL}

ATPase activity. **B.** 2fo-fc electron density contoured to 2σ (grey mesh) reveals a sulfate ion bound in the Ts^{FL} active site. P-loop, Walker B, and H loop residues are labeled. **C-D.** Structural superposition of the Ts OLD ATPase domain the ATPase domain (light orange) from *Methanococcus jannashii* Rad50 (C, light blue; PDB: 5DNY) and SMC-ScpAB complex from *Pyrococcus furiosus* (D, pale cyan, PDB: 4I99).

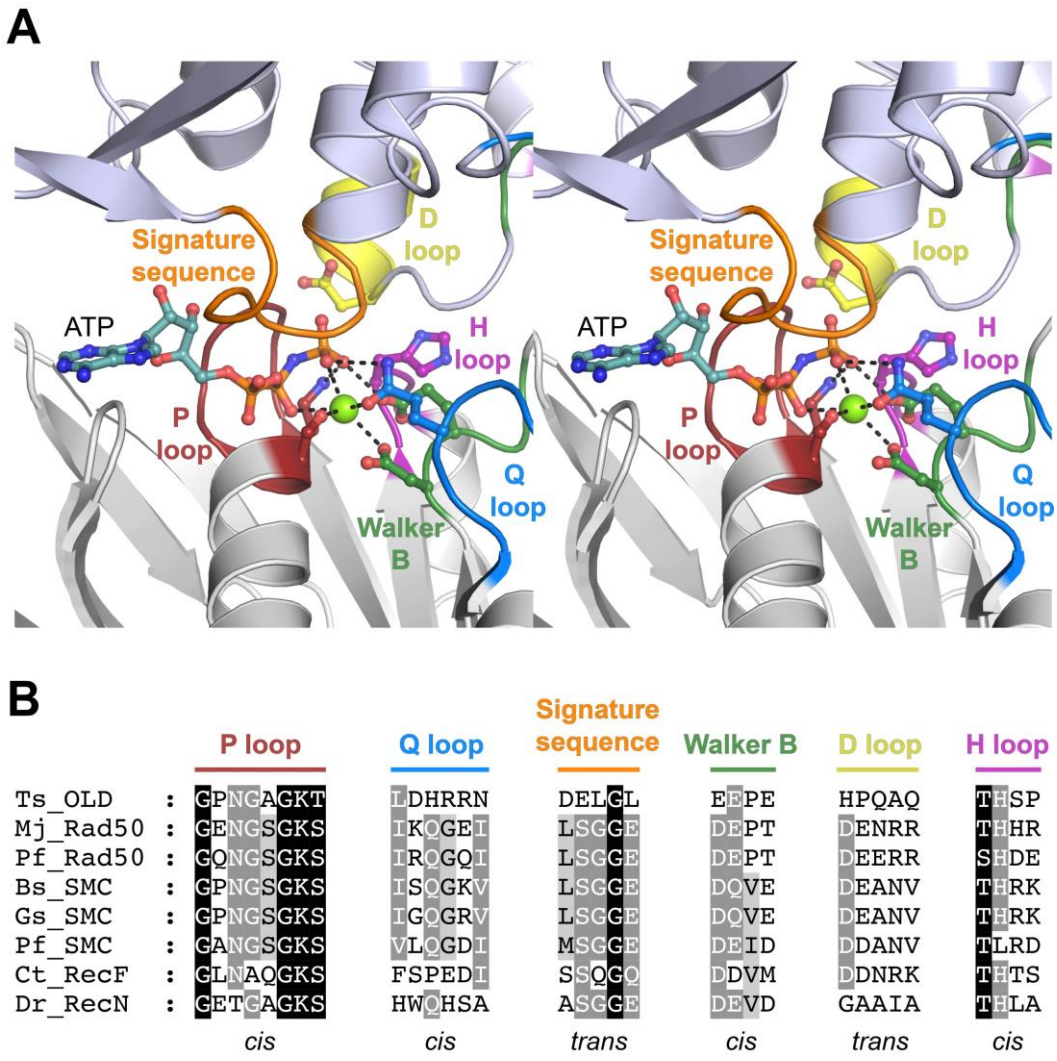
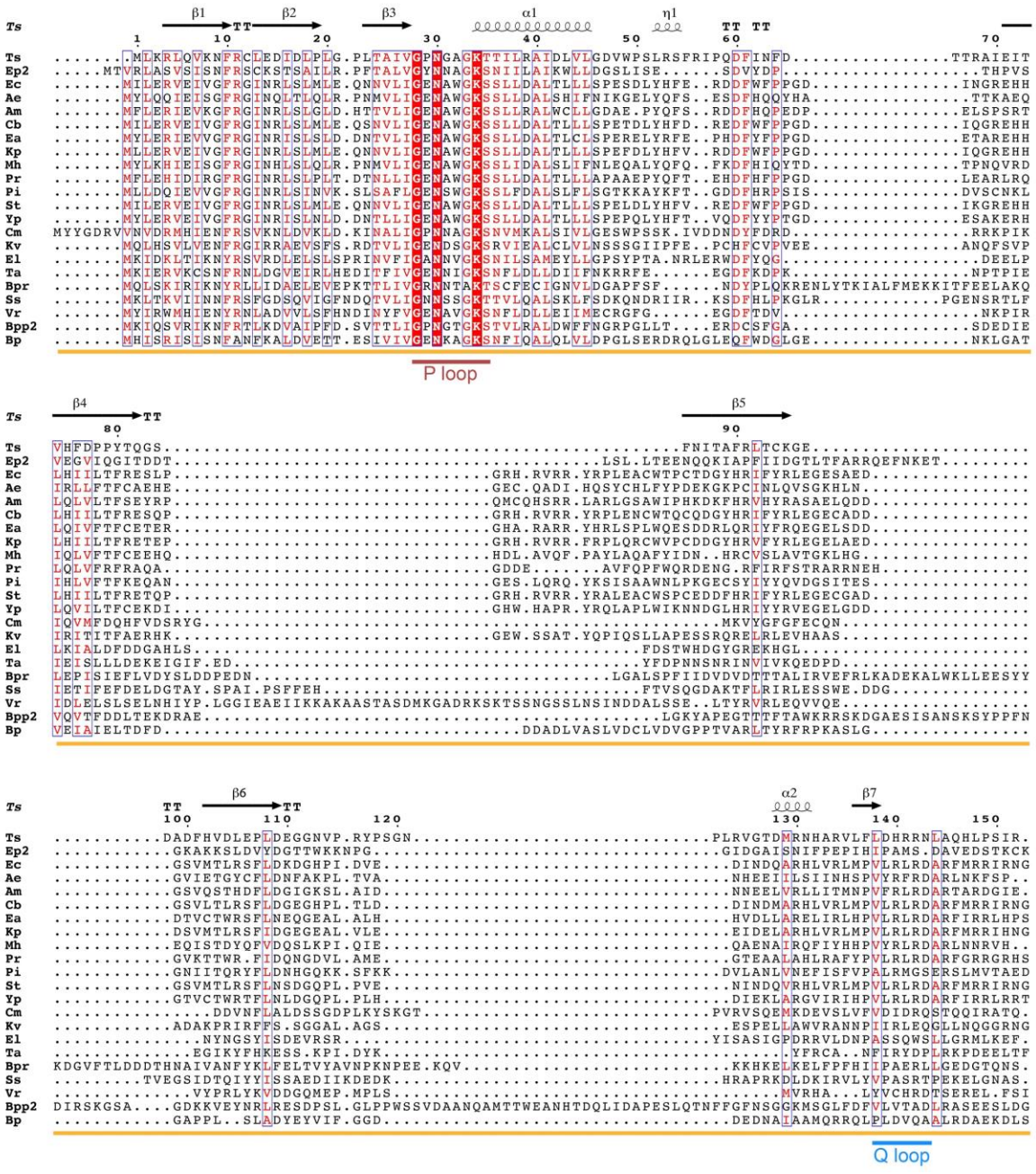
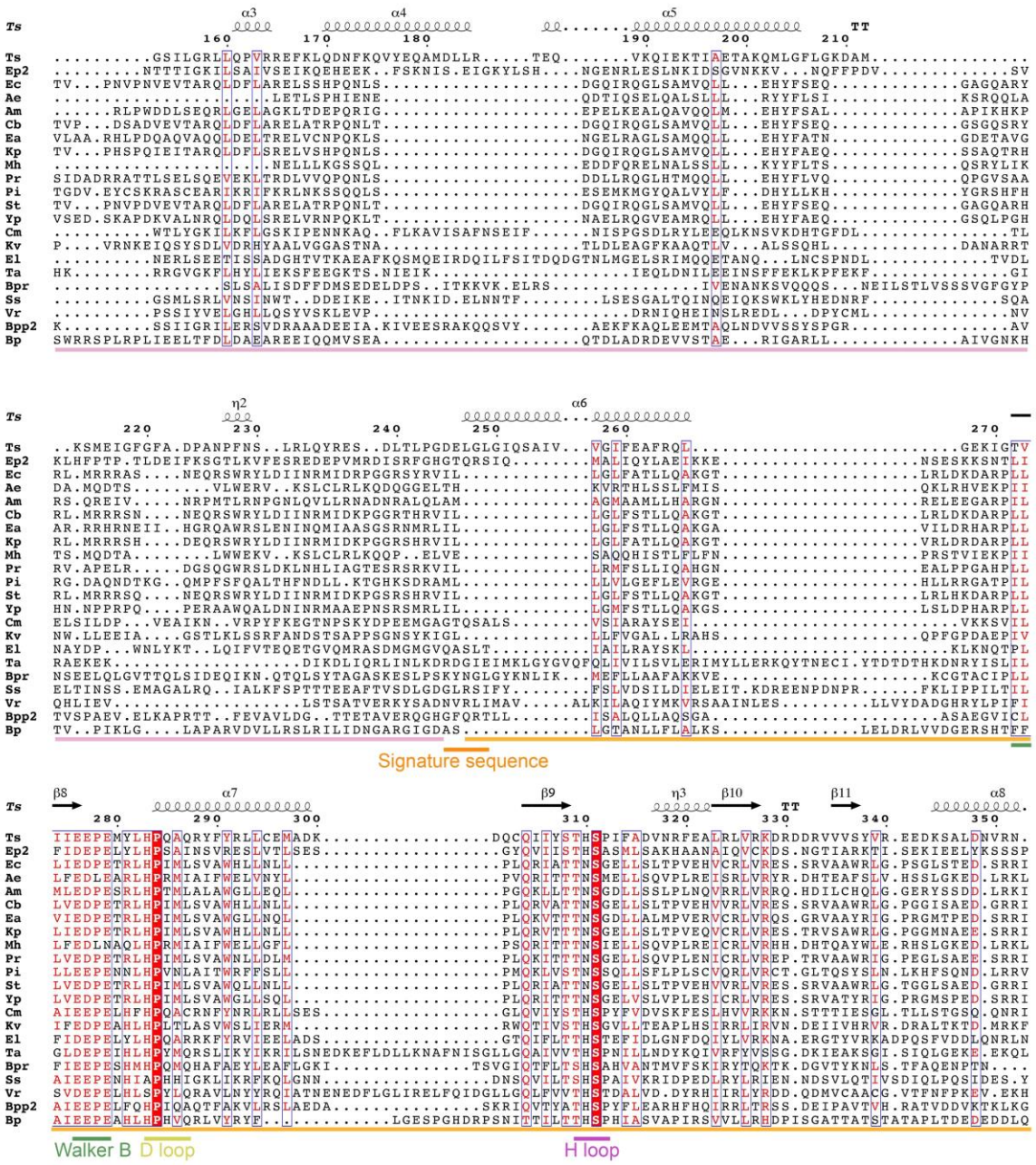
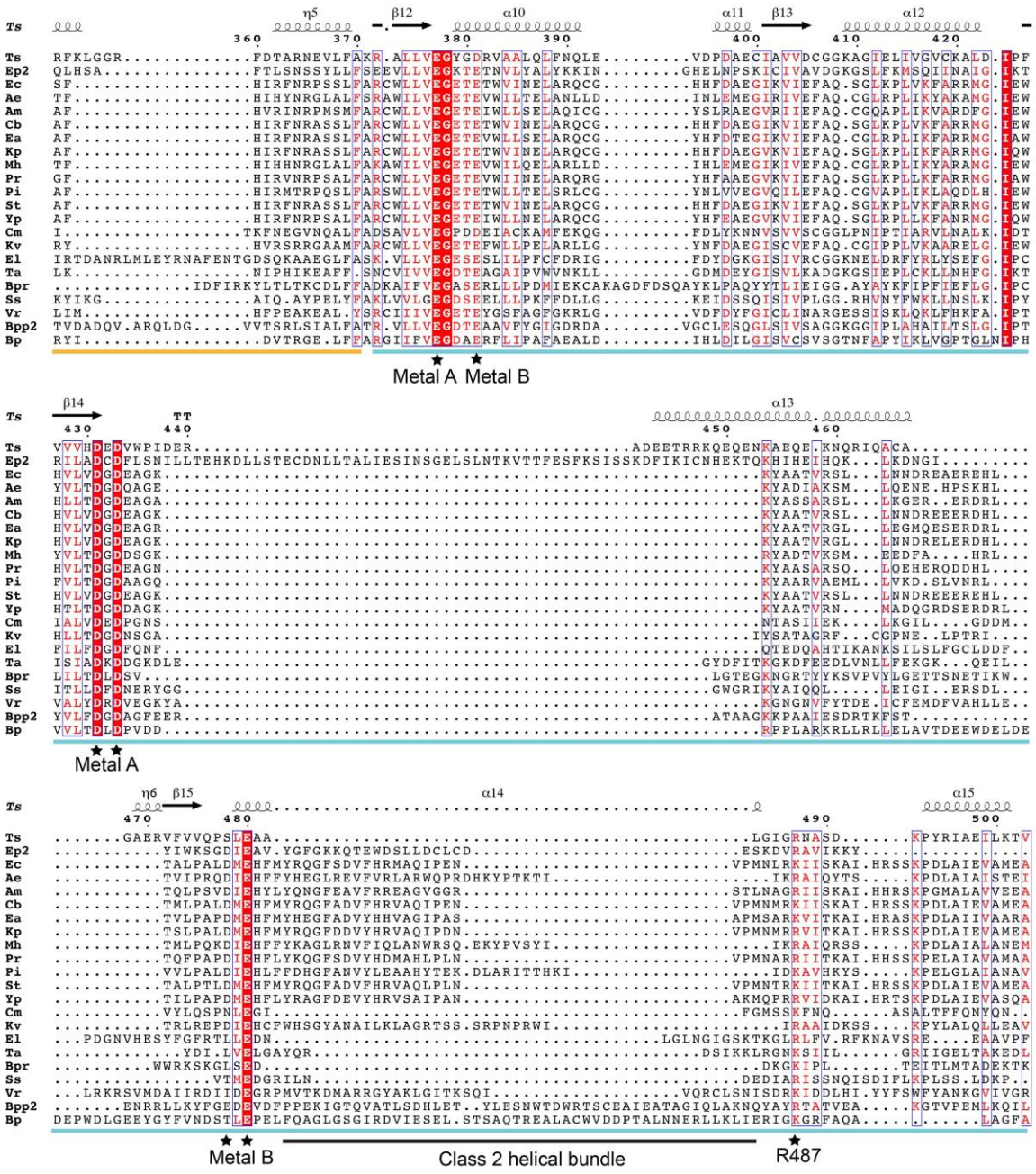


Figure S5. Catalytic motifs of SMC/Rad50/RecN/RecF ABC ATPases. A. Stereo view of *Pyrococcus furiosus* Rad50 (PDB: 3QKU) active site. Monomers are colored white and blue white respectively, with sequence motifs contributing to ATP binding and hydrolysis colored as follows: P loop, red; Q loop, blue; ABC signature sequence, orange; Walker B, green; D loop, yellow; H loop, purple. ATP substrate is colored light teal while bound magnesium cofactor is depicted as a green sphere. Catalytic side chains from each motif are shown as sticks with hydrogen bonding interactions illustrated with black

dashes. **B.** Sequences of ABC ATPase motifs from structurally characterized SMC/Rad50/RecN/RecF family members. Shading denotes conservation between these representatives. Contribution of each motif, either *in cis* or *in trans*, in indicated below. Nomenclature and associated PDB codes are as follows: *Thermus scotoductus* (Ts) OLD; *Methanocaldococcus jannaschii* (Mj) Rad50 (PDB: 5DNY); *Pyrococcus furiosus* (Pf) Rad50 (PDB: 3QKU); Bacillus subtilis (Bs) SMC (PDB: 5XG3); *Geobacillus stearothermophilus* (Gs) SMC (PDB: 5H68); *Pyrococcus furiosus* SMC (PDB: 4I99) *Caldanaerobacter subterraneus subsp. tengcongensis* (Ct) RecF (PDB: 5z68); *Deinococcus radiodurans* (Dr) RecN (PDB: 4ABY).


β1
β2
β3
α1
η1
P loop
β4
β5
β6
α2
β7
Q loop


Signature sequence
Walker B
D loop
H loop


Metal A
Metal B
Metal A
Metal B
Class 2 helical bundle
R487

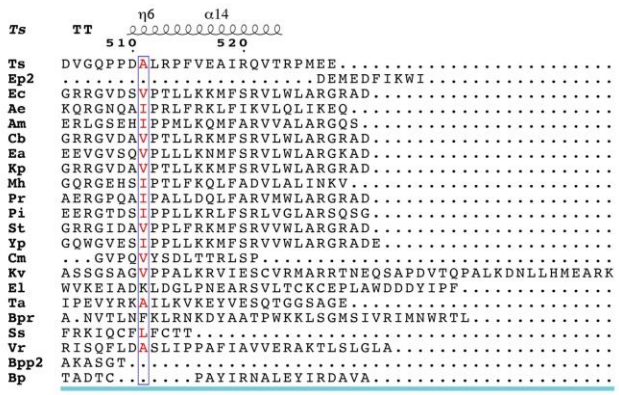


Figure S6. Sequence alignment of Class1 OLD family nucleases. Sequence alignment of representative Class 1 OLD family nucleases. Abbreviations are as follows with accompanying KEGG IDs: Ts, *Thermus scotoductus* (tsc:TSC_c04750)*;* Ep2, *Escherichia coli* phage P2 (vg:1261523); Ec, *Escherichia coli* (ecj:JW0860); Ae, *Actinobacillus equuli* (aeu:ACEE_06610)*;* Am, *Aeromonas media* (amed:B224_2448); Cb, *Citrobacter braakii* (*cbra:A6J81_00640*)*;* Ea, *Edwardsiella anguillarum* (ete:ETEE_0205)*;* Kp, *Klebsiella pneumoniae* (kpp:A79E_3329); Mh, *Mannheimia haemolytica* (mhaq:WC39_10155); Pr, *Pantoea rwandensis* (kln:LH22_15680); Pi, *Psychromonas ingrahamii* (pin:Ping_0201); St, *Salmonella enterica* (sty:STY0935); Yp, *Yersinia pestis* Angola (ypg:YpAngola_A1594); Cm, *Candidatus Micrarchaeota* (marh:Mia14_0874); Kv, *Koribacter versatilis* (aba:Acid345_4083); El, *Eggerthella lenta* (ele:Elen_0801); Ta, *Thermocrinis albus* (tal:Thal_0492); Bpr, Butyrate-producing bacterium SS3/4 (bprs:CK3_32330); Ss, *Stroptococcus suis* (ssv:SSU98_0942); Vr, *Veillonella rodentium* (vrm:44547418_01362); Bpp2, *Burkholderia pseudomallei* phage P2 (vg:10323863). The Class 2 OLD homolog *Burkholderia pseudomallei* (Bp, bpd:BURPS668_A0038) is included for comparison.

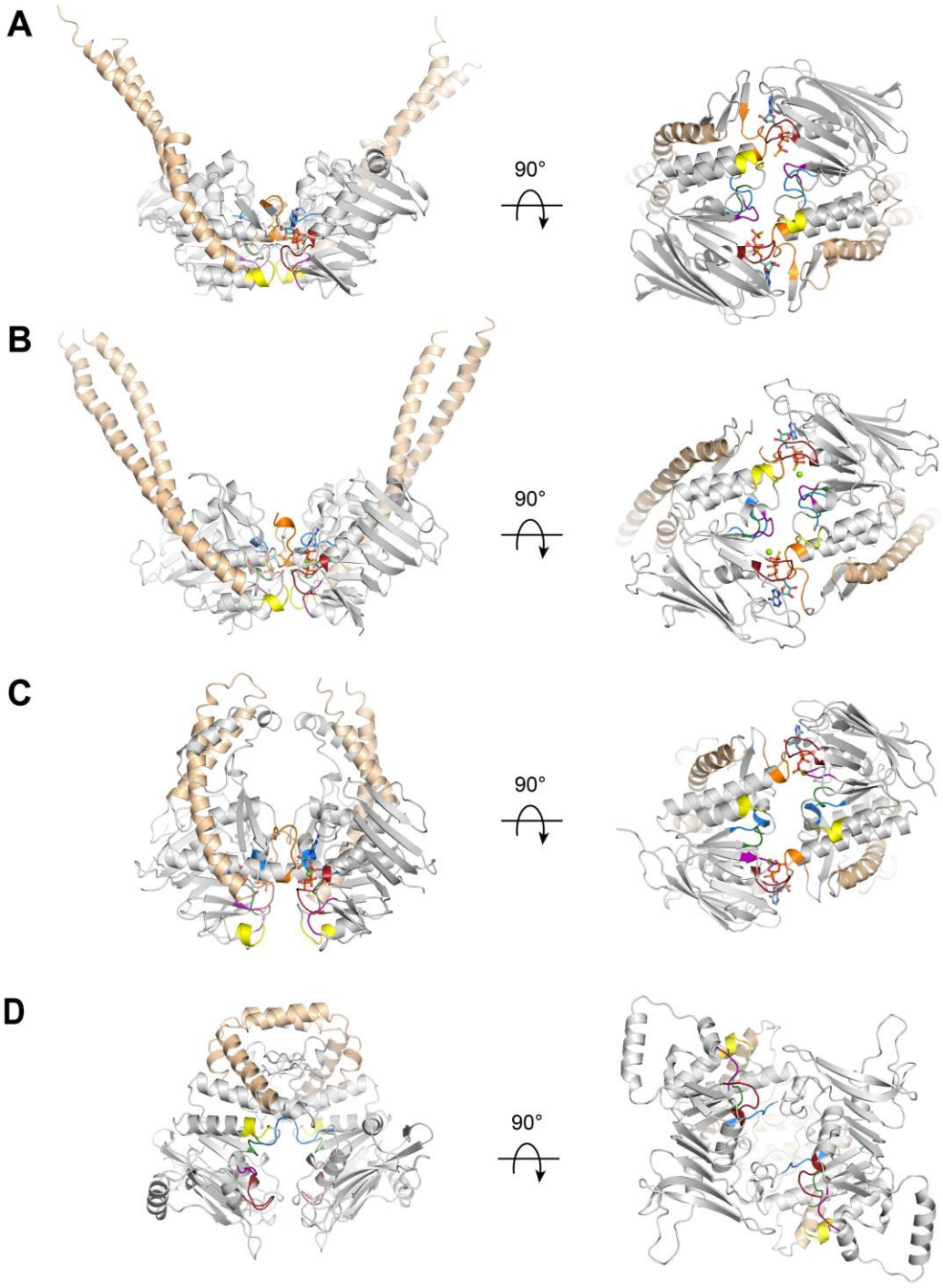


Figure S7. Organization of ATPase domain dimers. A-D. Organization of ATPase domains in the dimers of *Methanocaldococcus jannaschii* Rad50 (PDB: 5DNY) (**A**), *Bacillus subtilis* SMC (PDB: 5XG3) (**B**), *Caldanaerobacter subterraneus subsp. tengcongensis* RecF (PDB: 5z68) (**C**), and *Thermus*

scotoductus (Ts) OLD (**D**). ABC ATP domain core is colored gray with the extended coils colored wheat to illustrate the relative orientation of each monomer. Motifs involved in nucleotide binding and hydrolysis colored as follows: P loop, red; Q loop, blue; ABC signature sequence, orange; Walker B, green; D loop, yellow; H loop, purple. Bound nucleotides are colored light teal while bound magnesium ions are depicted as green spheres. Side (left) and bottom (right) views of each dimer are shown.

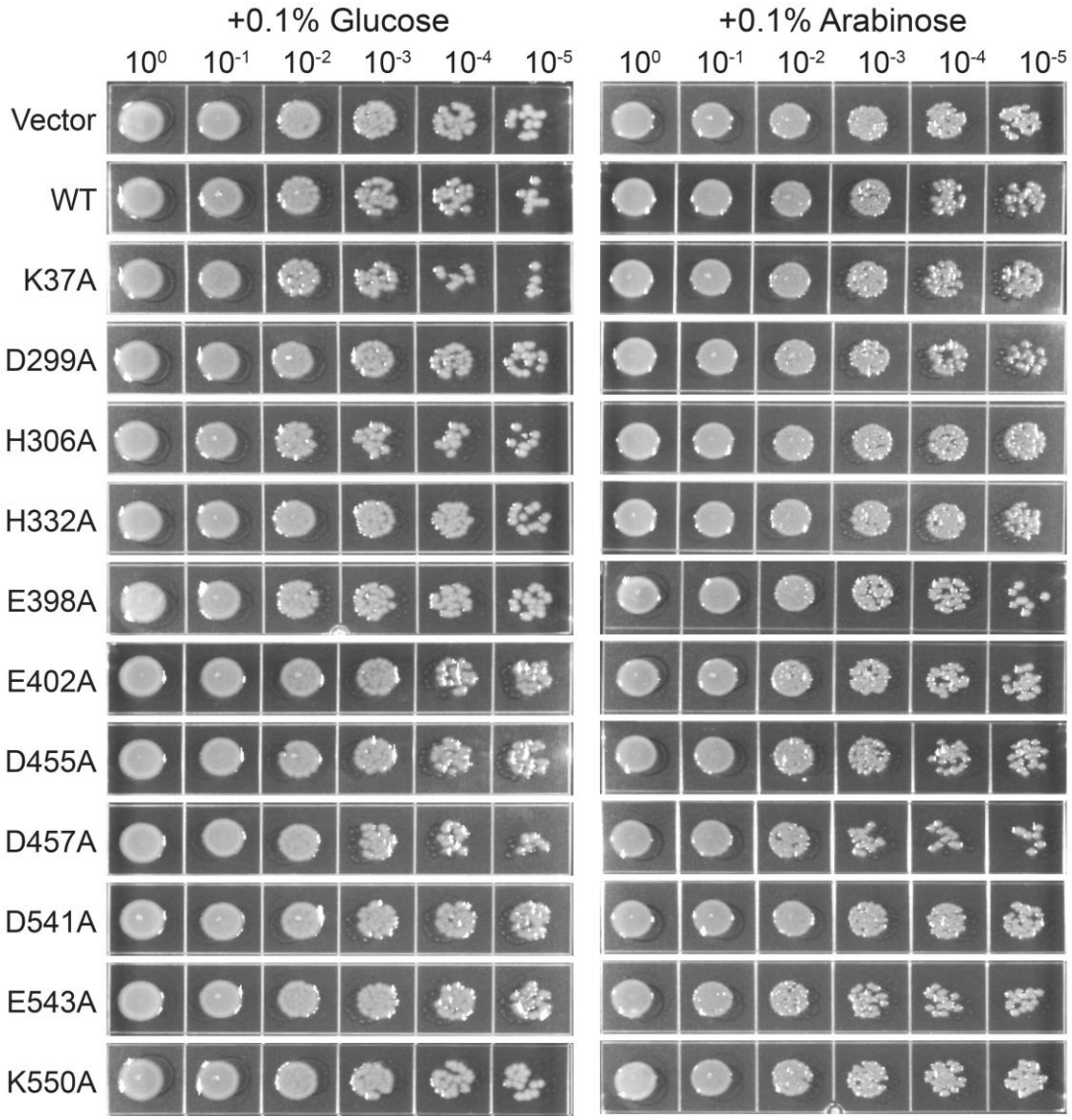


Figure S8. Spot assay of P2 OLD wild type and mutants in RecBC^{ts} *E. coli* at 30°C. E. coli transformed with arabinose-inducible P2 OLD wild type as well as P2 OLD carrying mutations in the ATPase and nuclease domains were grown overnight at 30°C, the temperature permissive for RecBCts function. There is no noticeable loss of cell viability with P2 OLD expression induced (+0.1% arabinose) compared to P2 OLD expression repressed (+0.1% glucose).

Chapter 4. Concluding Remarks and Future Directions

Concluding Remarks

The goal of the work described in this thesis was to broaden our understanding of the OLD proteins, particularly in terms of the structure and biochemical properties. The P2 *old* gene was very well characterized when the molecular biology field was beginning to mature; however, aside from scant biochemical data from the mid-1990s, little had been done to understand the fundamental mechanism of OLD proteins in general. Moreover, almost no work has been done to understand the role of OLD protein in contexts outside of phage exclusion. Though my research has fallen short of revealing the purpose of the OLD proteins, it has nonetheless defined the OLD family as structurally unique among both ABC ATPases as well as Toprim containing proteins. The full length structures of the OLD proteins have also helped to draw unexpected parallels to Rad50 and the SMC proteins. The well-detailed genetic interactions of P2 OLD with the RecBCD recombination pathway coupled with new insights into the nuclease and DNA binding activity of the OLD proteins suggests that they may have an unappreciated role in DNA repair or recombination in bacteria and archaea that has yet to be uncovered.

Given the level of tertiary structural similarity between the OLD proteins and Rad50 it would stand to reason that they function in a similar capacity to recognize DNA substrates and license cleavage by their respective nucleases. One potential confounding factor to this hypothesis is the quaternary arrangement of the TsOLD monomers in the crystal structure described in Chapter 3. However, the contents of Appendix 1 show that the TsOLD dimer

adopts an additional conformation that mimics DNA-bound Rad50. Moreover, Appendix 1 also suggests that TsOLD binds to DNA in an ATP-dependent manner and utilizes similar DNA-binding residues to Rad50.

These results would be consistent the spot assay data presented in Chapter 3. Mutation of the ATPase active site residues in P2 OLD completely abolished the cell death phenotype observed in RecBC- mutants. The simple answer to the result would be that the inability of P2 OLD to hydrolyze ATP would prevent it from binding DNA. I would argue for somewhat more nuanced hypothesis. I suggest that the ATPase activity of P2 OLD and the other homologs is necessary to regulate stable binding of the OLD proteins to a specific DNA architecture (one likely present during replication) and that ATP hydrolysis is vital for the OLD proteins to transition into a nuclease-active state.

What exactly P2 is targeting and why its activity results in replication inhibition and cell death remains quite a mystery. As mentioned in Chapter 1, the P2 OLD nuclease possesses 5'-3' activity, however, this alone is not sufficient to explain the toxic effects of P2 OLD. RecJ, RecBCD, RecE, and lambda exonuclease all possess 5'-3' exonuclease activity, but their expression does not cause such an obvious detrimental phenotype. Cell death in RecBC- cells expressing P2 OLD strongly suggests that double-strand breaks are being formed that cannot be repaired. This is supported by the observations that mutations in *sbcA* and *sbcB* in the RecBC- background are resistant to the killing phenotype of P2 OLD due to the repair activities of

RecET and RecFOR, respectively.

In Chapters 2 and 3 of this manuscript both the endo- and exonuclease activities of Xcc, Bp, and TsOLD were observed. It is quite possible that both activities are necessary for effective OLD protein function. An endolytic nicking event followed by processive cleavage along the *E. coli* genomic DNA or the lambda DNA could explain how replication is inhibited in the presence of P2 OLD. During DNA replication the lagging strand of DNA contains short stretches of RNA-DNA hybrids, nicks, and proper DNA duplexes. If P2 OLD were to recognize an RNA-DNA duplex and proceed to nick and degrade the template DNA, many double-strand breaks along the lagging strand would occur (Figure 1). At this point, however, that mechanism is only a musing. With our current data it is not possible to determine the exact mechanism of P2 OLD; moreover, understanding the role of the bacterial OLD proteins remains a mystery as well. It is quite possible that they perform a similar function as P2 OLD (albeit attenuated as to be beneficial to the host) or the other OLD homologs could be responsible for recognizing and processing completely different DNA structures. While this work has built the foundation to understand the OLD proteins on a structural and biochemical level, there is yet even more to be discovered.

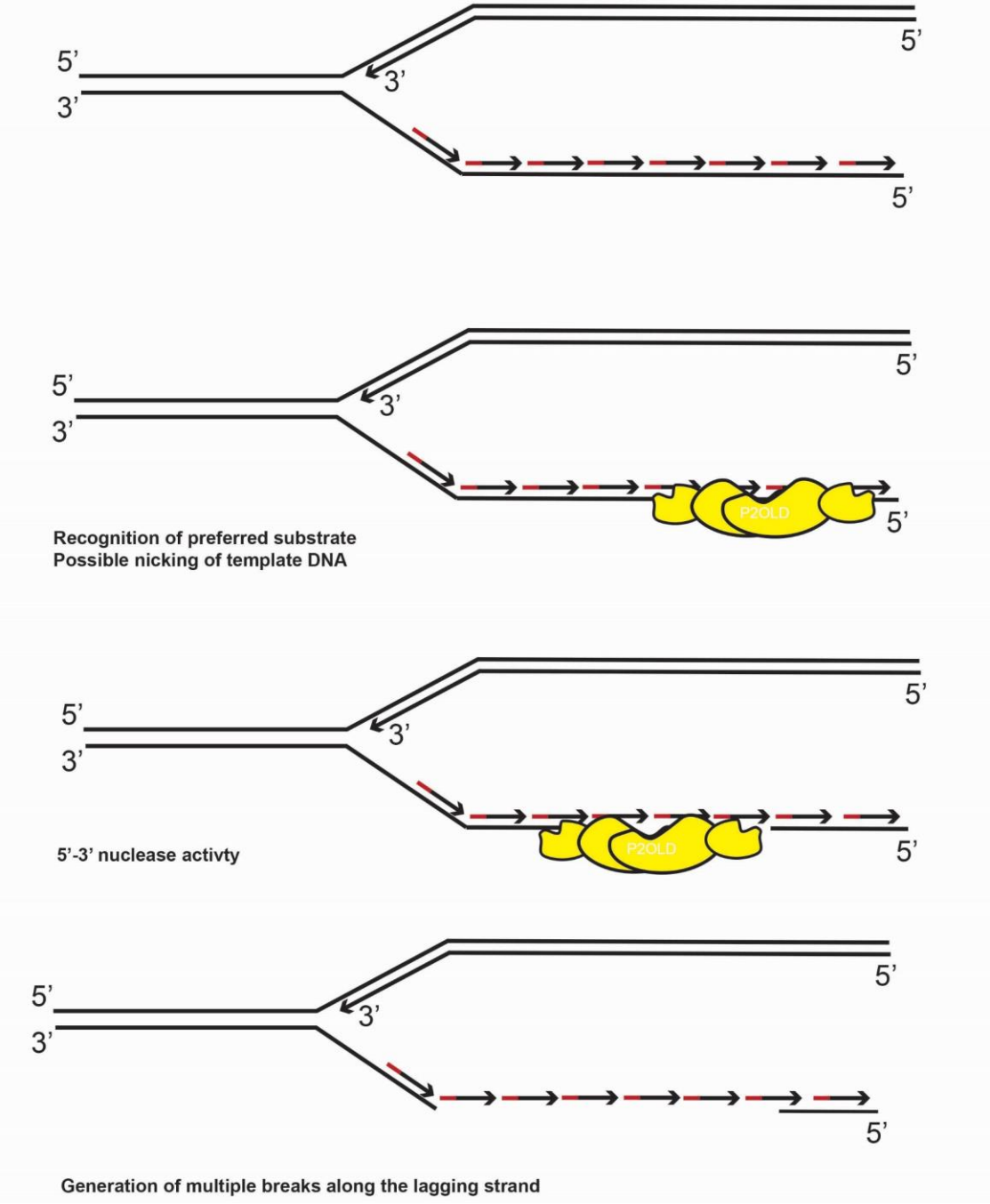


Figure 1. Potential mechanism of P2 OLD in disrupting DNA replication on the lagging strand. DNA is represented with black lines, RNA is represented with red lines.

Future directions

One challenge in working with the thermophilic Ts OLD protein is its dependence on high temperature for ATPase activity and robust DNA binding. Substrate recognition is a key component to many DNA replication and repair systems. While I have made to attempts to determine if Ts OLD recognize blunt DNA ends, 5’ overhangs, or 3’ overhangs, the data has been less than clear. In moving forward, attempts must be made to stabilize the base pairing of the DNA substrates to avoid the ends of the duplexes from melting, giving Ts OLD access to a ssDNA end that it may or may not normally bind. This problem could be alleviated all together if a soluble Class 1 OLD protein can be identified from a more mesophilic source, though that itself has proven a challenging task.

The Class 2 OLD proteins, while exceptionally soluble and easy to purify, remain biochemically frustrating. Their lack of measurable ATPase and DNA binding activity has made it difficult to characterized them much beyond nuclease activity. One avenue to be explored is the UvrD/PcrA-like helicases that usually accompany the *old* genes on the genome. I have made attempts to purify the OLD-associated helicase from *Burkholderia pseudomallei* and *Xanthomonas campestris* pv campestris, but I was unable to obtain soluble protein. Given their proximity to one another on the genome, it is likely that they function together in some capacity. What has yet to be attempted is co-expressing the OLD protein and its helicase in *E. coli.* It is feasible that the helicase requires to be in a complex with an OLD protein in order to be

stabilized. Should this complex actually form, it would likely be a treasure trove of biochemical and structural data.

Finally, understanding the role of the OLD proteins *in vivo* remains the loftiest but most tantalizing goal. Given the knowledge we already have about the function of P2 OLD in exclusion of lambda phage, the simplest place to start would be knock out the endogenous *old* gene in *E. coli* and test sensitivity to a wide variety of bacteriophage. Plaque size and shape would be an easy indicator for any effect that knocking out the OLD protein may have. While more laborious, exploring the role of the *E. coli* OLD protein in the context of other DNA repair systems may be of value. We know the P2 OLD possesses a 5'-3' nuclease activity, so it would not be unreasonable to think that the *E. coli* OLD protein may have a supportive role to the other 5'-3' exonucleases like RecBCD or RecJ. Constructing double mutants with the other exonucleases and then challenging the strains with DNA damaging agents or assessing their recombination efficiency could also parse out OLD's potential function. Look for protein-protein interactions with the OLD protein could also prove fruitful. This could be approached using a genetic tool like a yeast or bacterial two-hybrid experiment; however, co-immunoprecipitation followed by mass spectrometry analysis could also be useful and supplement the genetic approach.

APPENDIX 1. The OLD nuclease from *Thermus scotoductus* is an ATP-dependent DNA-binding protein

The OLD nuclease from *Thermus scotoductus* is an ATP-dependent DNA-binding protein

The OLD nucleases are a conserved family of proteins present in bacteria, archaea, and some bacteriophage. Though the function of these proteins remains elusive for most species, it has been previously established that the OLD protein from bacteriophage P2 is capable of preventing the replication of phage lambda as well as killing *E. coli* deficient in the homologous recombination function of RecBC

The full length structure of the Class 1 OLD protein from the thermophile *Thermus scotoductus* revealed that the N-terminal ATPase domain of the OLD proteins belongs to the ATP-binding cassette (ABC) family of ATPases. Initial characterization of the OLD ATPase domains by sequence alone had been difficult due to the absence of motifs conserved in other ATPase families, but absent in the OLD proteins. ABC ATPases are usually dimeric and characterized by six key motifs: the P-loop (or Walker A), the Walker B, the Q-loop, the H-loop which act in *cis* with the D-loop and signature sequences acting in *trans*. Presence of nucleotide usually stimulates dimerization which generates two composite ATPase sites sandwiched between the two monomers. While OLD proteins have a readily discernable P-loop and Walker B motif, the other elements could not be identified by sequence alone. When the structure of the ATPase domain from TsOLD was compared to the DNA double-strand break protein Rad50, it became apparent that the glutamine of the Q-loop and the aspartate of the D-loop were

substituted with histidine. In addition the signature sequence of canonical ABC ATPases was completely absent. Structure-guided sequence alignment of both the Class 1 and Class 2 OLD proteins showed generally poor conservation of residues in the Q- and H- loop; however, the histidine in the D-loop is absolutely conserved across species and across classes. Substitution of these residues in TsOLD with the canonical ABC ATPase motifs rendered the ATPase domain of TsOLD catalytically inactive.

The other striking difference between the structure of TsOLD and its closest structural homolog was the unusual orientation of the dimer. Structures of ABC ATPases with nucleotide present are generally dimeric with both monomers contributing to the ATPase active sites. The TsOLD structure, though dimeric, was oriented such that the active sites were turned away from one another with no potential for *trans* acting elements to catalyze ATP hydrolysis. In this appendix I provide an additional structure of the Ts OLD protein that adopts a conformation that is more consistent with other members of the ABC family. I also examine additional similarities between the DNA binding and ATPase activity of TsOLD compared to Rad50.

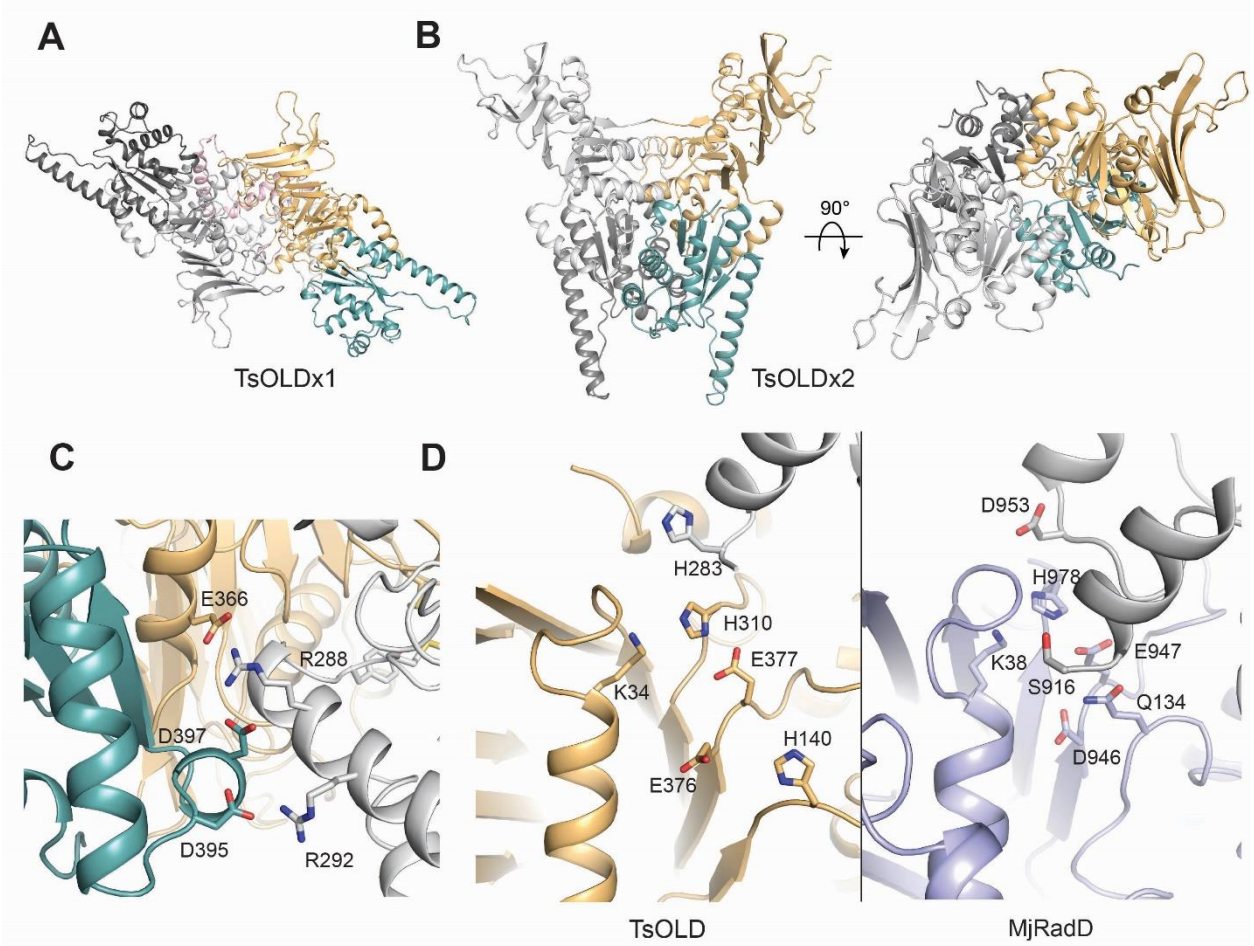


Figure 1. The second crystal form of TsOLD shows a large conformational change. A. Organization of the ATPase domains of the first TsOLD structure solved (TsOLDx1). For one monomer the ATPase domain is colored light orange, the Toprim domain is light teal, and the dimerization is in light pink. The other monomer is shaded in gray. **B.** Two views of the second crystal structure of TsOLD (TsOLDx2). For one monomer the ATPase domain is colored light orange and the Toprim domain is light teal. The other monomer is shaded in gray. **C.** Novel salt bridges formed between the ATPase domain of one monomer (gray) and the Toprim (light teal) and ATPase domain (light orange) of the other. **D.** Comparison of the ATPase active site of TsOLDx2 and the active site of *Methanocaldococcus jannaschii* Rad50 (light purple; PDB:5DNY). Gray residues reflect elements interacting *in trans.*

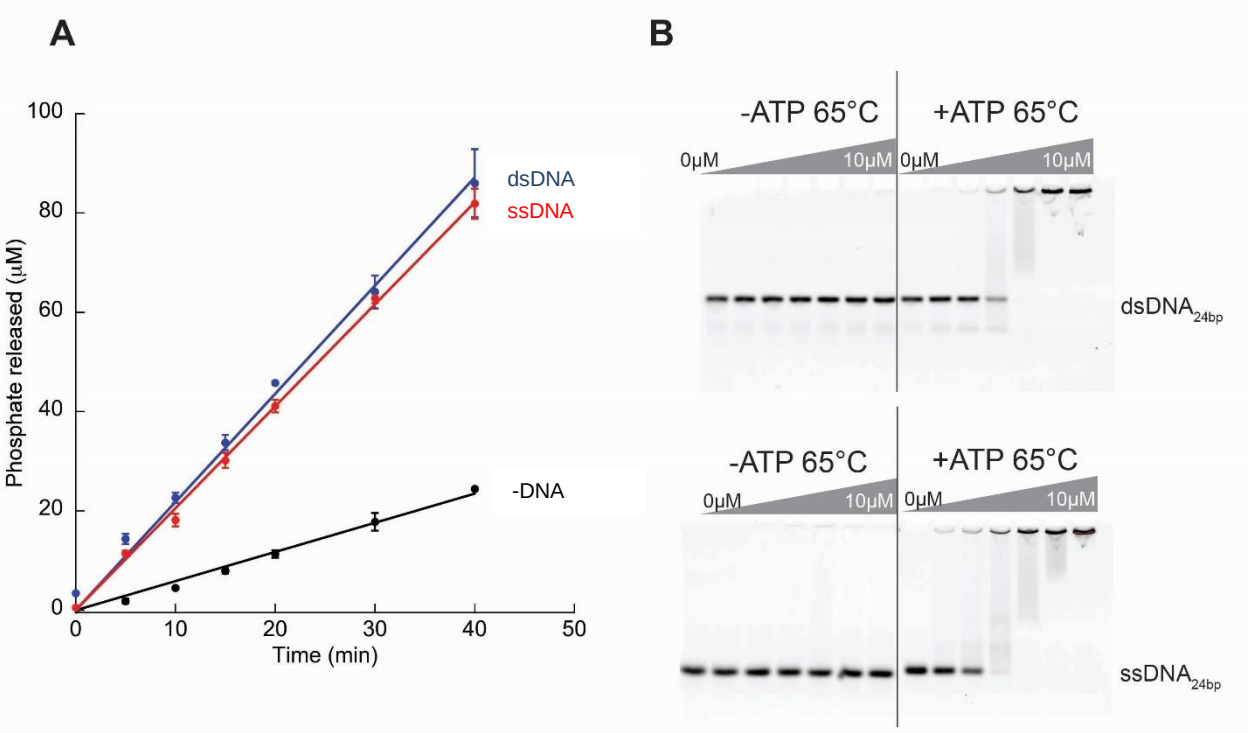


Figure 2. TsOLD binds DNA in an ATP-dependent manner. A. Initial rates of ATP hydrolysis by TsOLD in the presence of ssDNA or dsDNA compared to without DNA. **B.** DNA gel shifts of ss- and dsDNA with increasing amounts of TsOLD with and without ATP present.

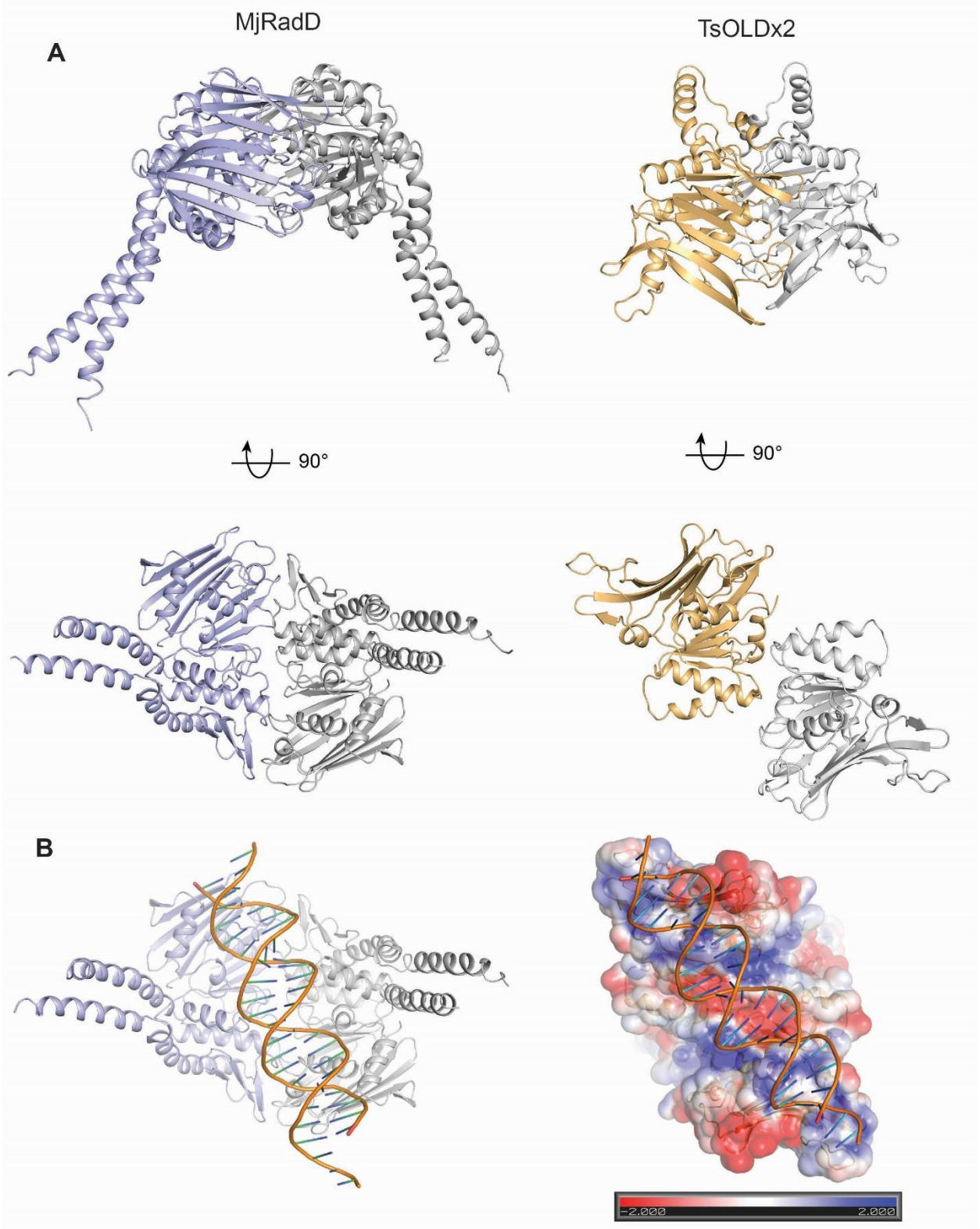


Figure 3. The conformation of TsOLDx2 resembles DNA-bound Rad50. A. Two views comparing the ATPase domain organization of TsOLDx2 (light

orange/gray) to *Methanocaldococcus jannaschii* Rad50 (light purple/gray; PDB:5DNY). **B.** Electrostatic potential map of TsOLDx2 modeled with the DNA substrate from MjRad50.

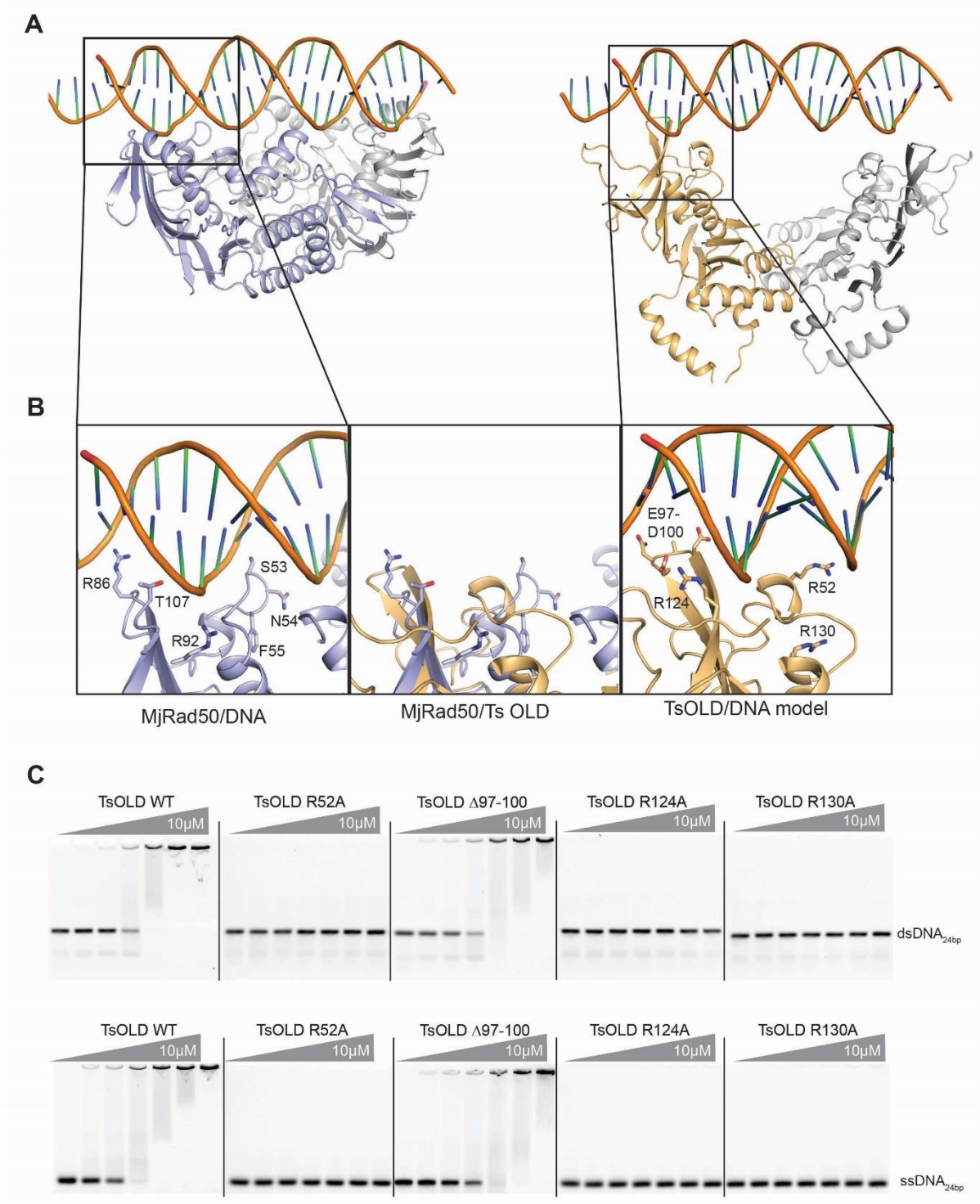


Figure 4. TsOLD uses a similar mechanism of MjRad50 to bind DNA. A. Side view of MjRad50 bound to DNA (light purple/gray) compared to the model of TsOLDx2 ATPase domain (light orange/gray) binding DNA. **B.** Close up view of residues that may interact with DNA in MjRad50. Structural alignment

of MjRad50 and TsOLD suggests residues in TsOLD that may be important for binding DNA. **C.** Gel shift assays with both dsDNA and ssDNA of TsOLD Wt potential DNA binding mutants.

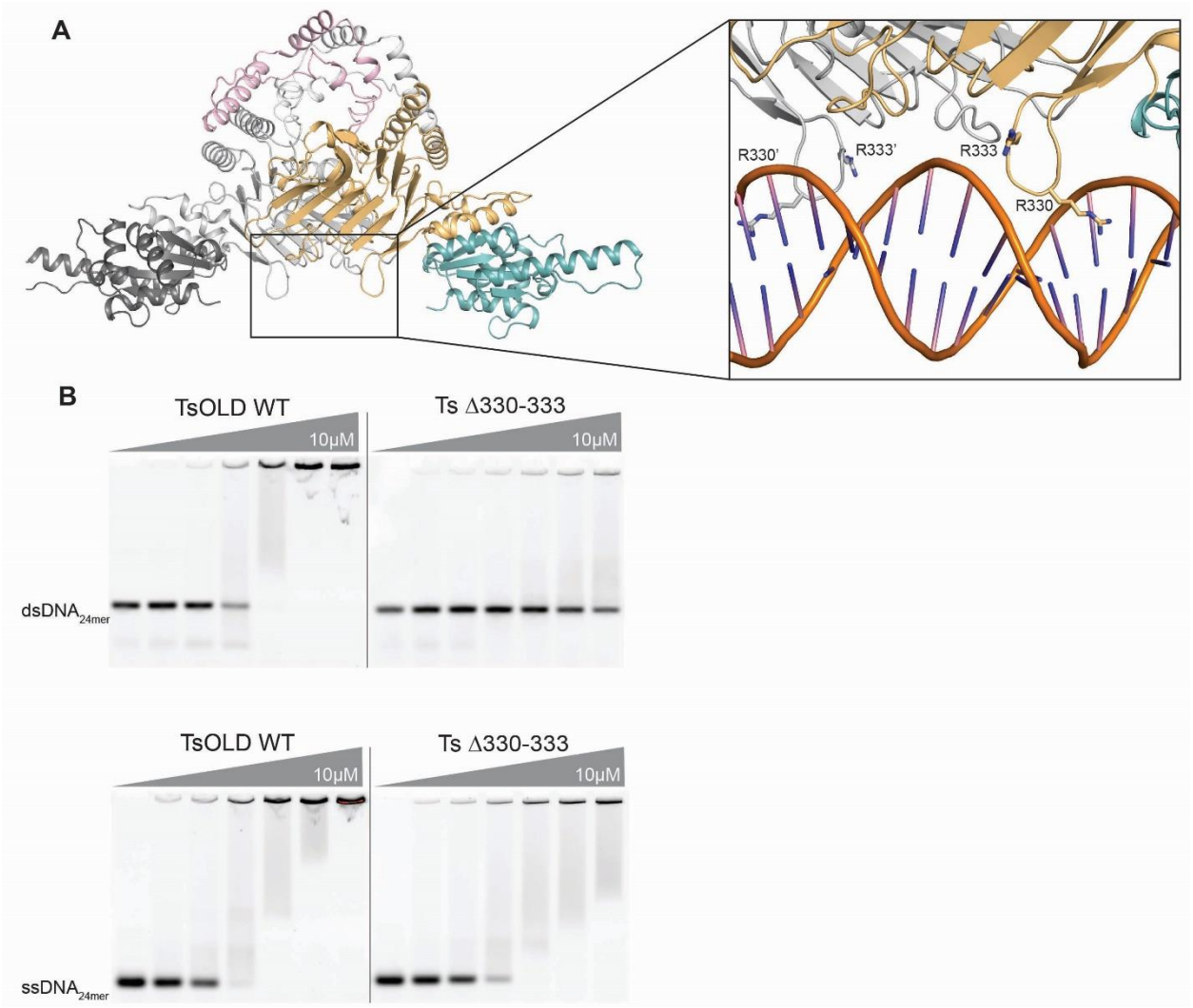


Figure 5. The TsOLD likely has two modes of DNA binding. A. Side view of TsOLDx1 highlighting two arginine-rich loops with DNA modeled between them. **B.** Gel shift assay with dsDNA and ssDNA testing the effect of deleting the loop containing R330 and R333.

Cloning, expression, and purification of TsOLD

DNA encoding *Thermus scotoductus* (Ts) OLD (UniProt E8PLM2) was codon optimized for *E. coli* expression and synthesized commercially by IDT. A construct containing the full-length protein (residues 1-528) was amplified by PCR and cloned into pET21b, introducing a 6xHis tag at the C-terminus. TsOLD was transformed into BL21(DE3) cells, grown at 37°C in Terrific Broth to an OD_{600} of 0.7-0.9, and then induced with 0.3 mM IPTG overnight at 19°C. Cells were pelleted, washed with nickel loading 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 frozen in liquid nitrogen and stored at -80°C.

Thawed 500 ml pellets of TsOLD constructs were resuspended in 30 ml of nickel loading buffer supplemented with 3mg DNAse, 10mM $MgCl_2$, 10 mM PMSF, and a Roche complete protease inhibitor cocktail tablet. Lysozyme was added to a concentration of 1 mg/ml and the mixture was incubated for 10 minutes rocking at 4°C. Cells were disrupted by sonication and the lysate was cleared via centrifugation at 13 000 rpm (19 685 g) for 30 minutes at 4°C. The lysate was then heated at 65°C for 15 minutes to precipitate heat-labile proteins. Precipitation was removed by centrifugation at 13 000 rpm (19 685 g) for 15 minutes at 4°C. The supernatant was filtered, loaded onto a 5 ml HiTrap chelating column charged with $NiS0_4$, and then washed with nickel loading buffer. Native TsOLD was eluted by an imidazole gradient from 30 mM to 1 M. Pooled fractions were dialyzed overnight at 4°C into Q loading buffer (20 mM Tris pH8, 50 mM NaCl, 1 mM EDTA, 5% glycerol (v:v), and 1 mM DTT). The dialyzed sample was applied to 5 ml HiTrap Q column equilibrated with Q loading buffer, washed in the same buffer, and eluted with a NaCl gradient from 50 mM to 1 M. Peak fractions were pooled, concentrated, and further purified by size exclusion chromatography (SEC) using a Superdex 200 16/600 pg column. Native TsOLD was exchanged into a final buffer of 20mM HEPES pH 7.5, 150mM KCl, 5 mM MgCl2, and 1mM DTT during SEC and

concentrated to 20-30 mg/ml. Point mutations were introduced by Quikchange (Agilent) and all mutant TsOLD constructs were purified in an identical manner as the wildtype protein.

Crystallization, X-ray data collection, and structure determination

Native TsOLD at 10 mg/ml was crystallized by sitting drop vapor diffusion at 20°C in 0.1M Tris pH 7.9, 8% PEG 8000, and 5 mM Taurine. Crystals were of the space group $P6_122$ with unit cell dimensions a = 140.7Å, b = 140.7Å, c = 105.6Å and α = 90°, β = 90°, γ = 120° and contained a monomer in the asymmetric unit. Samples were cryoprotected by transferring the crystal directly to Parabar 10312 (Hampton Research) prior to freezing in liquid nitrogen. Data was collected at the NECAT 24-ID-E beamline at the Advanced Photon Source at the Selenium edge (λ=0.9795Å) at 100K to a resolution of 2.23Å (Table S1). Data were integrated and scaled via the NE-CAT RAPD pipeline, employing XDS and AIMLESS for data integration and scaling respectively. The structure was solved using Phaser-MR from PHENIX. The search model consisted of the TsOLDx1 structure with the dimerization domain removed. Further model building and refinement was carried out manually in COOT and PHENIX. The resulting model was refined with Rwork/Rfree values of 0.2248/0.2638 (Supplementary Table 1). The final model contained residues 1-141 and 254-525,one water molecule, one magnesium ion, and one taurine molecule.

Structural superpositions were carried out in Chimera. All structural models were rendered using Pymol (Schrodinger).

Electrophoretic mobility shift assays (EMSAs)

Assays were carried out in EMSA buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 5 mM $MgCl_2$) with 400nM of 5’ 6-FAM labeled DNA substrates ordered from Integrated DNA Technologies. Reactions include 0.5mM of ATP unless otherwise specified. Ts OLD in each reaction was 0, 0.5, 1, 2, 4, 6, and 10 uM.

Samples were incubated at 65°C for 15 minutes before being mixed with 15% Ficoll 400 and loaded onto a 10% non-denaturing acrylamide gel and run in 1xTAE at 200V for 4 hours.

ATPase Assays

ATPase activity was characterized using a colorimetric Malachite green assay that monitored the amount of free phosphate released over time. The DNA dependence of Ts OLD ATPase activity was carried out in ATPase reaction buffer (20 mM Tris-HCl pH 8, 50 mM NaCl, 5 mM $MgCl_2$) with 1 mM ATP, 4 μM TsOLD, with or without 1mM of single- or double-stranded DNA. At each time point, 20 μL samples were taken and quenched with 5μL of 0.5M ETDA pH 8.0. 150 μL of filtered Malachite green solution were added to each sample and incubated for 5 minutes. Samples were then visualized in a Thermo Multiskan GO spectrophotometer at 650 nm. Rates were plotted All ATPase mutant assays were carried out in ATPase reaction buffer with the 0.5 mM ATP and 4μM protein at 65ºC.

Table 1. X-ray data collection and refinement statistics for TsOLDx2

	Ts^{FL} x2 PDB: XXXX
Data collection	
Space group	$P6_122$
Cell dimensions	
a, *b*, *c* (Å)	140.66,140.66 105.56
α, β, γ (°)	90,90,120
Resolution (Å)	2.42
R_{sym} or R_{merge}	0.250 (4.05)
R_{meas}	0.254 (4.112)
$CC_{1/2}$	0.995 (0.506)
$I / \sigma I$	12.5 (0.7)
Completeness (%)	99.7 (96.9)
Redundancy	38.0 (33.7)
Refinement	
Resolution (Å)	2.42
No. reflections	23744
R_{work} / R_{free}	0.2248/0.2638
No. atoms	
Protein	3288
Ligand/ion	8
Water	1
B-factors	
Protein	91.83
Ligand/ion	126.45
Water	85.94
R.m.s deviations	
Bond lengths (Å)	0.018
Bond angles (°)	1.63
Ramachandran statistics	
Favored (%)	92.5
Allowed (%)	6.5
Outliers (%)	1.0