Structure and function of a Lon protease-like domain of a bacterial RecA paralog, RadA

(RecA パラログタンパク質 RadA が持つ Lon プロテアーゼ様ドメインの構造と機能)

Doctoral Thesis

Masao Inoue

Department of Biological Sciences, Graduate School of Science

Osaka University

2017

CONTENTS

ABBREVIATIONS

CFUs	colony-forming units
dsDNA	double-stranded DNA
EMSA	electromobility shift assay
HR	homologous recombination
LonC	Lon protease-like domain
nt	nucleotides
rLonC	RadA-type LonC
pLonC	Lon protease-type LonC
r.m.s.	root mean square
ssDNA	single-stranded DNA
SSB	single-stranded DNA-binding protein
UV	ultraviolet
ZF	zinc finger

ABSTRACT

Homologous recombination (HR) is one of the most important DNA-repair pathways to maintain genome integrity. RecA/Rad51 paralogs have been recognized as an important factor of HR. Among them, only one bacterial RecA/Rad51 paralog—RadA—is involved in HR as an accessory factor of RecA recombinase. RadA has a unique Lon protease-like domain (LonC) at its C-terminus, in addition to a RecA-like ATPase domain. Unlike Lon protease, RadA's LonC domain does not show protease activity but is still essential for RadA-mediated DNA repair. Reconciling these two facts has been difficult because RadA's tertiary structure and molecular function are unknown. Here, I describe the hexameric ring structure of RadA's LonC domain, as determined by X-ray crystallography. The structure reveals the two positively charged regions unique to the LonC domain of RadA are located at the intersubunit cleft and the central hole of a hexameric ring. Surprisingly, a functional domain analysis demonstrated the LonC domain of RadA binds DNA, with site-directed mutagenesis showing that the two charged regions are critical for this DNA-binding activity. Interestingly, only the intersubunit cleft was required for the DNA-dependent stimulation of ATPase activity of RadA, whereas both of the DNA-binding sites were essential for DNA repair function *in vivo*. These data provide the structural and functional aspects of the LonC domain in RadA-mediated DNA repair. Finally, I conducted a phylogenetic analysis to reconsider the protein family possessing the LonC domain and discovered ComM, a member of genetic recombination machinery, as a candidate for a novel "DNA binding-type" LonC domain-containing protein subfamily.

INTRODUCTION

In living cells, DNA is constantly exposed to various types of DNA damage agent, such as ultraviolet (UV) light, ionizing radiation, reactive oxygen species, and DNA crosslink agents (Schärer, 2003; Altieri *et al*., 2008). The resulting DNA damages cause genetic alterations via mutagenesis and, in the most severe cases, cell death. All organisms developed a variety of different DNA repair machineries to overcome the various types of DNA damages (Morita *et al*., 2010).

Among the DNA-repair machineries, homologous recombination (HR) is important for maintaining genome integrity through DNA double-strand break repair and rescue of stalled replication forks (Fig. 1) (Cromie *et al*., 2001; West, 2003; Krejci *et al*., 2012; Shin *et al*., 2004). On the other hand, HR also generates genetic variation and is recognized as one of the driving forces of evolution through horizontal gene transfer or intrachromosomal recombination (Brigulla & Wackernagel, 2010; Thomas & Nielsen, 2005).

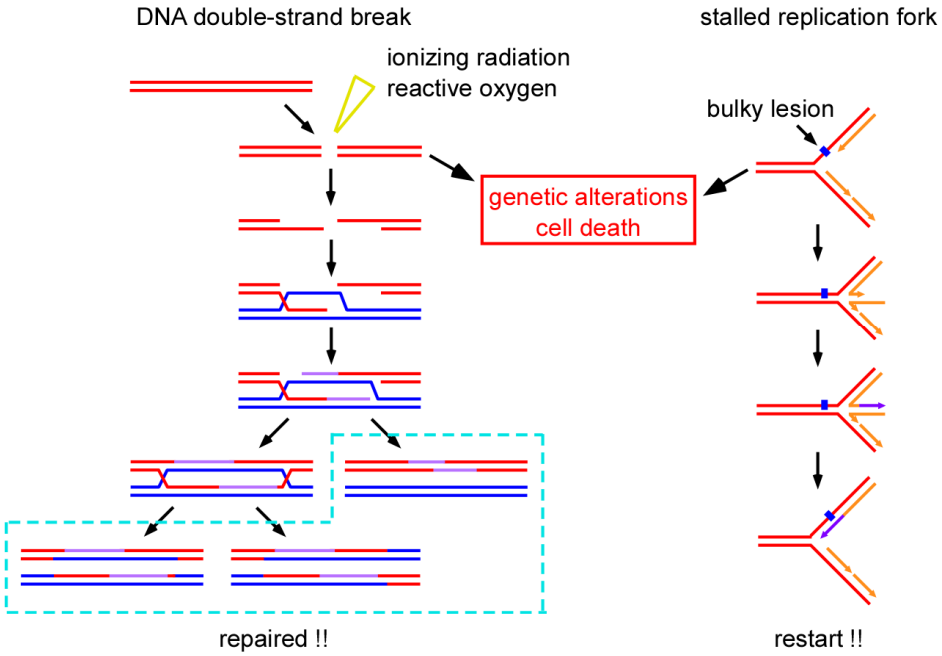

FIGURE 1. A schematic representation of HR system. The pathway of DNA double-strand break repair and the rescue of a stalled replication fork are shown.

To fulfill these contradictory roles, numerous proteins engage in each reaction in the HR system. Among these proteins, the RecA/Rad51 proteins have a central role in HR as a recombinase that searches for a homologous DNA sequence and catalyzes the strand-exchange reaction (Bell & Kowalczykowski, 2016; Cox, 2007; Lin *et al*., 2006; Chen *et al*., 2008). Interestingly, most organisms have one or more RecA/Rad51 paralogs in addition to the main recombinase, RecA/Rad51 (Fig. 2*A*).

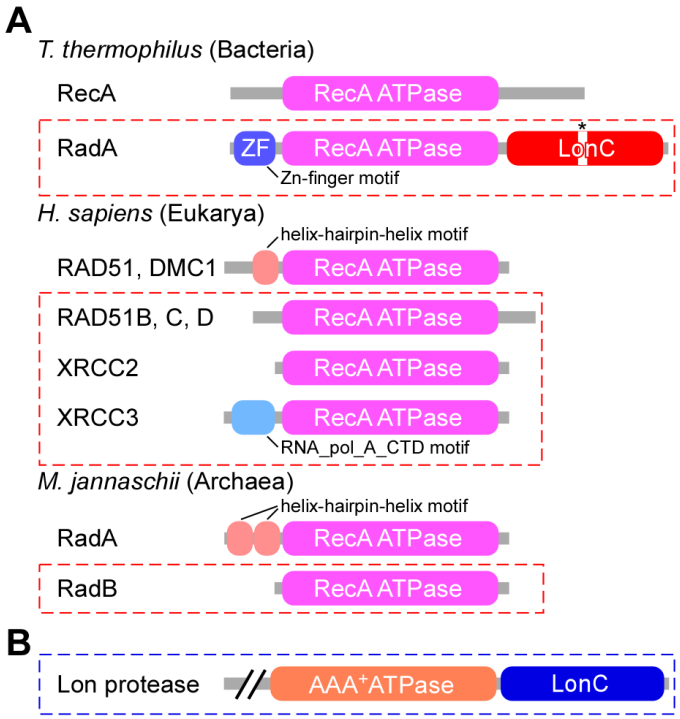

FIGURE 2. Domain organization of RecA/Rad51 superfamily proteins and Lon protease. (A) Schematic representations of domain organization of RecA/Rad51 superfamily proteins in the three kingdoms of life. RecA/Rad51 paralogs lacking the recombinase activity are highlighted by *red dotted rectangles*. It should be noted that archaeal RadA proteins are not orthologs of bacterial RadA proteins. The asterisk indicates that the catalytic serine residue of Lon protease activity is not conserved in RadA. (B) A schematic representation of the domain organization of Lon protease. The long N-terminal region that contains the putative substrate recognition domain is omitted.

The RecA/Rad51 paralogs lack the recombinase activity but are involved in HR in various organisms. The human Rad51 paralogs are known to function as mediators of Rad51 and as tumor suppressors (Suwaki *et al*., 2011; Thacker, 2005). The human Rad51 paralogs are also involved in the late step of HR, such as branch migration and Holliday junction processing (Compton *et al*., 2010; Liu *et al*., 2004). The worm Rad51 paralogs remodel the Rad51 pre-synaptic nucleoprotein filament and stimulate Rad51-catalyzed strand exchange (Taylor *et al*., 2015). On the other hand, the yeast Rad51 paralogs protect the Rad51 pre-synaptic nucleoprotein filament from anti-recombinase Srs2 (Liu *et al*., 2011).

RadA (also known as Sms) is a highly conserved RecA paralog in bacteria (Fig. 2*A*). Genetic studies have shown that *radA* is involved in DNA repair and genetic recombination in various organisms, such as *Escherichia coli*, *Bacillus subtilis*, and *Deinococcus radiodurans* (Diver *et al*., 1982; Song & Sargentini, 1996; Beam *et al*., 2002; Lovett, 2006; Cooper *et al*., 2015; Krüger *et al*., 1997; Carrasco *et al*., 2004; Qing *et al*., 2006; Slade *et al*., 2009). Recently, a biochemical characterization of *E. coli* RadA protein has shown that RadA stimulates RecA-catalyzed strand exchange *in vitro* (Cooper & Lovett, 2016). In this reaction, RadA promotes branch migration in a RecA-dependent manner, implying the action of RadA on the RecA post-synaptic nucleoprotein filament or the recombination intermediates. It has also been indicated that *radA* genetically interacts with genes involved in the late steps of HR such as *recG*, *ruvAB*, and *dnaB* (Beam *et al*., 2002; Lovett, 2006; Cooper *et al*., 2015).

RadA is a multidomain protein consisting of three putative functional regions (Fig. 2*A*). A Cys2/Cys2-type zinc-finger (ZF) domain, a RecA-like ATPase domain, and a Lon protease-like domain (LonC) are located at the N-terminal, central, and C-terminal regions, respectively. Among RecA/Rad51 paralogs, the ZF and LonC domains are unique to RadA and are not found in eukaryotic and archaeal RecA/Rad51 paralogs (Fig. 2*A*). To our knowledge, the tertiary structure of neither the RadA protein nor each domain has not been determined. Genetic studies

have shown that these three putative domains are essential for RadA-mediated DNA repair in *E. coli* and *D. radiodurans* (Cooper *et al*., 2015; Qing *et al*., 2006). It has also been shown that RadA functions in *E. coli* requires a Cys residue that is thought to be necessary for Zn binding in the ZF domain and a Lys residue of Walker A motif in the RecA-like ATPase domain (Cooper *et al*., 2015). *E. coli* RadA protein has a DNA-dependent ATPase activity like a RecA protein (Cooper & Lovett, 2016). It can be supposed that the ZF and RecA-like ATPase domains are involved in DNA-binding ability of RadA; however, the domains or residues responsible for DNA binding have not been identified.

A LonC domain was originally identified in Lon protease as a catalytic domain of protease activity (Amerik *et al*., 1991; Botos *et al*., 2004; Rotanova *et al*., 2004). Lon protease is a member of the ATP-dependent serine proteases found in three kingdoms of life and involved in protein degradation upon various stress responses to change protein abundances and to adapt to the stress conditions (Van Melderen & Aertsen, 2009; Pinti *et al*., 2016). The structural studies show that Lon protease forms a hexameric ring, which has a hole in its center (Cha, *et al*., 2010; Lin, *et al*., 2016; Su, *et al*., 2016). It has been thought that a protein substrate binds to an N-terminal putative substrate recognition domain and is accommodated into a hole of ring, unfolded by a central AAA^+ motor domain and then degraded by the C-terminal LonC domain (Figs. 2*B* and 3).

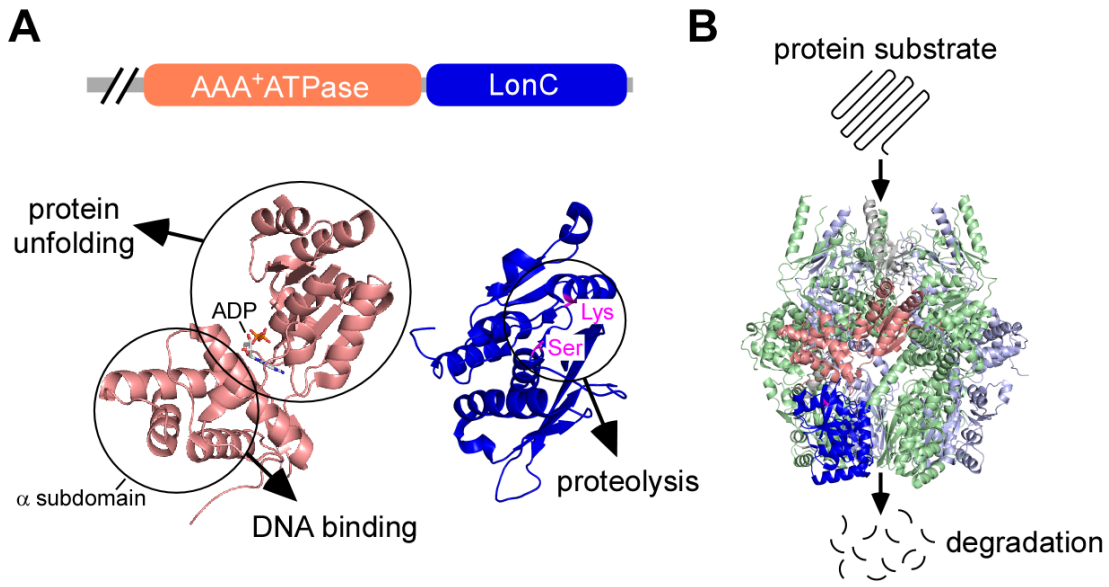

FIGURE 3. Structure of Lon protease. (A) Domain structures of Lon protease. The AAA^+ ATPase and LonC domains (PDB ID: 3K1J) are represented by *salmon pink* and *blue cartoons*, respectively. In this structure, the catalytic Ser and Lys residues (*magenta sticks*) in the LonC domain are replaced with Ala to avoid self-digestion. (B) Schematic model of protein substrate degradation by Lon protease hexameric ring.

In Lon protease, the protease activity of the LonC domain is exerted by a catalytic Ser residue (Amerik *et al*., 1991; Botos *et al*., 2004) (Fig. 3). Interestingly, the catalytic Ser is not conserved in a LonC domain of RadA (Fig. 2*B*), and it has been shown that *E. coli* RadA has no protease activity (Beam *et al*., 2002). These strongly suggest that the LonC domain of RadA is catalytically inactive as a protease. As a matter of convenience, I refer to a LonC domain of Lon protease and that of RadA as pLonC (Lon protease-type) and rLonC (RadA-type), respectively.

Like rLonC, several types of inactive protease-like proteins called "pseudoprotease" have been identified. Pseudoproteases have traditionally been hypothesized to act as regulators of their active homologues. Actually, *Clostridium* CspA has been thought to be involved in the regulation of catalytically active protease paralogs, CspB and CspC (Kevorkian *et al*., 2016). In addition, some pseudoproteases have an extensive and expanding role in biological processes, including regulation and inhibition. *Drosophila* Masquerade protein, a potential extracellular adhesion factor, possesses an inactive protease-like domain at its C-terminus (Murugasu-Oei *et al*., 1995). Derlin-1, an inactive member of the rhomboid intramembrane protease family, has been thought to be involved in protein-protein interaction at the endoplasmic reticulum membrane (Greenblatt *et al*., 2011; Chen *et al*., 2014). As RadA is supposed to interact with proteins involved in the late steps of HR in *E. coli* (Beam *et al*., 2002; Lovett, 2006; Cooper *et al*., 2015), rLonC may be associated with the interaction with those proteins. In contrast to the other two domains, there is no infomation about the function of rLonC of RadA. New studies focusing on the characterization of rLonC are therefore needed.

In this study, I determined the first X-ray crystal structure of rLonC from *Thermus thermophilus* HB8. Surprisingly, rLonC possesses the DNA-binding activity exerted by two positively charged regions, namely the intersubunit cleft and the central hole of the hexameric ring structure. I also showed that these two DNA-binding sites are essential to the RadA-

mediated DNA repair function *in vivo* and have different roles in the DNA-dependent ATPase activity of RadA *in vitro*, suggesting an inter-domain interaction between the rLonC and ATPase domains. Furthermore, combining these results with a phylogenetic analysis highlighted structural and functional relationships between RadA and another LonC family protein, ComM. My findings provide not only the structural and functional aspects of the rLonC domain in RadA-mediated DNA repair but also evolutionary aspects of LonC-family proteins.

EXPERIMENTAL PROCEDURES

Construction of expression plasmids

An expression plasmid of RadA from *T. thremophilus* HB8 was obtained from the RIKEN BioResource Center (Tsukuba, Japan). The plasmid, pET-11a/*ttha0541* (*radA*) (clone name TEx05B04), was constructed by inserting the target gene into pET-11a (Novagen, Madison, WI) using NdeI and BamHI. A fragment containing the *radA* gene digested by NdeI and EcoRI was ligated to a pET-HisTEV vector to create pET-HisTEV/*radA* for the expression of a His_6-tagged RadA protein. The pET-HisTEV vector was constructed based on pET-15b (Novagen) by replacing the recognition sequence of thrombin (LVPRGS) with that of TEV protease (ENLYFQG). DNA fragments coding for domain-truncated mutants of RadA, ΔN53 (amino acid residues 54–423), ΔN261 (residues 262–423, identical to the rLonC domain), Δ54C (residues 1–53, corresponding to the ZF domain), Δ261C (residues 1–260), and ΔN53/261C (residues 54–260, corresponding to the RecA-like ATPase domain) were generated by PCR using pET-11a/*radA* as a template (Fig. 4). The amplified fragments were digested by NdeI and BamHI at primer regions and cloned into the pET-HisTEV vector. Expression plasmids for site-directed mutants, R286A, R305A, R314A, K345A, R385A, R392A, R395A, R399A, R404A, and K412A/R413A were constructed by the QuikChange site-directed mutagenesis method using pET-HisTEV/*radA* as a template (Stratagene, La Jolla, CA). Expression plasmids for R286A/R385A and R305A/R314A/K345A mutants were constructed by repeating the QuikChange mutagenesis process twice and three times, respectively. The primer sequences used for plasmid construction are listed in Table 1.

TABLE 1 Sequences of primers used for this study

Domain truncation[a]	
ΔN53	5'-ATATCATATGTCCCAGGTGGACGAGGCCGA-3'
	5'-ATATAGATCTTCATGCCAGGTACGCCTCCA-3'
ΔN261	5'-TCATATCCATATGTCCGAGGCCTTCCTC-3'
	5'-TATGCTAGTTATTGCTCAGCG-3'
Δ54C	5'-TGTATAGGATCCTTATCAGAGGGCGAGGAGG-3'
	5'-TAATACGACTCACTATAGGGG-3'
Δ261C	5'-TCTATAGGATCCTTATCAGTTCTCCACCTCCAGA-3'
	5'-TAATACGACTCACTATAGGGG-3'
Site-directed mutagenesis	
R286A	5'-CCCTGGCCGGGGAGGCGGCCTTGGCCCTGG-3'
	5'-CCACGGGGCCGAAGCCGTTCTTGGCGCTC-3'
R305A	5'-CCCTTCCCCGCCCCCGCCAGGGTGGTCCAGGG-3'
	5'-CCCTGGACCACCCTGGCGGGGGCGGGGAAGGG-3'
R314A	5'-CCAGGGGCTGGACGGAGCGCGCGTGGACGTGGTC-3'
	5'-GACCACGTCCACGCGCGCTCCGTCCAGCCCCTGG-3'
K345A	5'-CTCGCCGGGGGGCTCGCGGTCCAGGACCCGGG-3'
	5'-CCCGGGTCCTGGACCGCGAGCCCCCCGGCGAG-3'
R385A	5'-GGCCGGGGAGGTGGCGCGGGTGGCGGG-3'
	5'-CCCGCCACCCGCGCCACCTCCCCGGCC-3'
R392A	5'-GGTGGCGGGGCTGGAGGCGAGGCTTAGGGAGGGG-3'
	5'-CCCCTCCCTAAGCCTCGCCTCCAGCCCCGCCACC-3'
R395A	5'-CTGGAGCGGAGGCTTGCGGAGGGGGAGCGGGC-3'
	5'-GCCCGCTCCCCCTCCGCAAGCCTCCGCTCCAG-3'
R399A	5'-CTTAGGGAGGGGGAGGCGGCGGGGTTCGGCCG-3'
	5'-CGGCCGAACCCCGCCGCCTCCCCCTCCCTAAG-3'
R404A	5'-CGGGCGGGGTTCGGCGCCTTCCTCCACCCGGG-3'
	5'-CCCGGGTGGAGGAAGGCGCCGAACCCCGCCCG-3'
K412A/R413A	5'-CCACCCGGGGAACCTTGCGGCCCTTCAGGAGGCTGTG-3'
	5'-CACAGCCTCCTGAAGGGCCGCAAGGTTCCCCGGGTGG-3'
Gene disruption	
1	5'-CATCCTCATCATGACCTCCAAC-3'
2	5'-CAAAGCCAGGTAGACGAAGAG-3'
3	5'-GAGTCCACGAAGAGGACCTC-3'
4	5'-CCAACATGATTAACAATTATTAGAGGTCATCGTTCAA-3'

[a]The recognition sites of the restriction enzymes used for plasmid construction and the start/terminator codons are indicated by *underlining* and *boxes*, respectively.

Protein overexpression and purification

E. coli Rosetta 2(DE3) cells (Novagen, Madison, WI) transformed with pET-HisTEV/*radA* were cultured for 15 h at 37°C in 2 L of an overexpression medium with 1.6% hipolypepton (Nihon Pharmaceutical, Tokyo, Japan), 1.0% yeast extract (Nihon Pharmaceutical), 0.5% NaCl, 0.75% glycerol, 0.05% glucose, 0.2% lactose, 100 μg/mL ampicillin, and 34 μg/mL chloramphenicol, buffered with 25 mM sodium-potassium phosphate at pH 7.4.

The cells were harvested by centrifugation and suspended in 40 mL of buffer I (50 mM Tris-HCl (pH 7.5) and 500 mM NaCl) containing 20 mM imidazole and 1 mM phenylmethylsulfonyl fluoride, and were then lysed by ultrasonication on ice. After centrifugation (8,000 × g) for 1 h at 4°C, the supernatant was loaded onto 2 mL of TALON M*etal* Affinity Resin (Clontech, Mountain View, CA) pre-equilibrated with buffer I containing 20 mM imidazole. The resin was washed with 100 mL of buffer I and then eluted with a 20-mL gradient of 20–300 mM imidazole in buffer I. The fraction containing the His_6-tagged RadA was dialyzed against buffer I; then, the protein was digested by TEV protease for 6 h at 15°C. After complete cleavage of the His_6-tag, the tag-free RadA was precipitated by 30% saturated ammonium sulfate. After centrifugation (8,000 × g) for 30 min at 4°C, the precipitant was dissolved in buffer I, and then the solution was loaded onto a HiLoad 16/60 Superdex 200 pg column (GE Healthcare Biosciences, Uppsala, Sweden) with the same buffer. The eluted RadA was concentrated using a Vivaspin concentrator (Sartorius AG, Göttingen, Germany). The protein concentration was determined on the basis of the absorbance at 280 nm using the molar extinction coefficients (M^{-1} cm^{-1}) calculated through the previously described procedure (Kuramitsu *et al*., 1990) as follows: wild type and site-directed mutants, 19,005; ΔN53, 10,290; ΔN261, 4,410; Δ54C, 8,715; Δ261C, 14,595; and ΔN53/261C, 5,880. The purity of each protein (>95%) was assessed by SDS-PAGE.

Except for the rLonC used for X-ray crystallography, all of the RadA mutants were overexpressed and purified by the same procedures as mentioned above. For crystallization, rLonC was purified in an N-terminal His_6-tagged form without TEV protease cleavage.

X-ray crystal structure determination

A crystal used for X-ray crystallographic analysis was obtained after 6 months of incubation under 4 mg/mL His_6-tagged rLonC in 50 mM Tris-HCl (pH 7.5) and 500 mM NaCl, in a glass test tube at 4°C. The crystal was soaked in 25% glycerol as a cryoprotectant and was cryo-cooled in a nitrogen cryo stream. Data were collected at the wavelength of 1.00000 Å at BL38B1 at SPring-8 (Hyogo, Japan). The diffraction images were processed with the HKL2000 program suite (Otwinowski *et al*., 1997). Using a structural model predicted using the I-TASSER server as a search model (Yang *et al*., 2015), the structure of rLonC was solved by molecular replacement. The MOLREP program in the CCP4 suite was used for the rotation and translation searches (Vagin *et al*., 1997; Winn *et al*., 2011). The model was refined using REFMAC5 (Murshudov *et al*., 2011) in the CCP4 suite and COOT (Emsley *et al*., 2010). The final model was validated by RAMPAGE (Lovell *et al*., 2003) and PROCHECK (Laskowski *et al*., 1993) in the CCP4 suite. The structure coordinates have been deposited into the Protein Data Bank under accession number 5H45. The surface electrostatic potentials were calculated by APBS (Baker *et al*., 2001) and PDB2PQR (Dolinsky *et al*., 2007). The calculation of the molecular contacts and the prediction of the stable multimer were carried out using PISA in the CCP4 suite (Krissinel & Henrick, 2007). All of the molecular graphics were generated by PyMOL (Schrödinger, New York, NY).

Gel-filtration analysis

Gel-filtration analysis was performed using a Superdex 200 10/30 GL column (GE

Healthcare Biosciences) on an ÄKTA explorer system (GE Healthcare Biosciences) at 25°C. First, 5 μM of the full-length RadA or rLonC was incubated in a buffer composed of 50 mM Tris-HCl (pH 7.5), 0.2–2 M NaCl at 25°C for 2 h. Then, 100 μL of the protein solution was loaded on the column and eluted at a flow rate of 0.5 ml/min in the same buffer. The elution profile was monitored by recording the absorbance at 230 nm. The apparent molecular mass was estimated by the calibration curve using thyroglobulin (669,000 Da), β-amylase (200,000 Da), alcohol dehydrogenase (150,000 Da), bovine serum albumin (66,000 Da), ovalbumin (44,000 Da), and carbonic anhydrase (29,000 Da).

Electromobility shift assay (EMSA)

Chemically synthesized 60-mer ssDNA (Bex, Tokyo, Japan), 5′-GGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAATCCATTGGTGATCACTGGTAGCGG-3′, was annealed to complementary 60-mer ssDNA (Bex), 5′-CCGCTACCAGTGATCACCAATGGATTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC-3′, to obtain 60 bp dsDNA. Then, 3 μM (bp) of dsDNA and 1.1 to 20 μM RadA were incubated in a buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mg/mL bovine serum albumin, 5% glycerol, and 0.01% bromophenol blue at 25°C for 30 min. The reaction mixtures (5 μL) were loaded onto a 5% polyacrylamide gel containing 10% glycerol and then electrophoresed in 1 × TBE buffer (89 mM Tris-borate and 2 mM EDTA). The gel was stained with SYBR Gold Nucleic Acid Gel Stain (Molecular Probes, Eugene, OR), and the bands were visualized by a FLA-3000 image analyzer (Fujifilm, Tokyo, Japan). The amounts of free and shifted DNAs were quantified using Multi Gauge version 2.1 software (Fujifilm) to determine the percent shifted. Under conditions with a large excess of RadA over DNA, the apparent dissociation constant K_d was determined from three independent experiments by fitting to the following equation using the Igor Pro 4.03 software (WaveMetrics, Lake Oswego, OR).

$$\% \text{ shifted} = [P]^h/(K_d{}^h + [P]^h) \quad \text{(Equation 1)}$$

where [P], K_d, and h are the total protein concentration, apparent dissociation constant for DNA, and Hill coefficient, respectively.

ATPase assay

To determine the kinetic constants of the ATPase activity, an ATPase assay was performed in reaction mixtures containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/mL bovine serum albumin, 0.5 μM RadA, and 7.8 to 500 μM ATP spiked with 0.1 μCi of [γ-^{32}P]ATP, in the presence and absence of 30 μM (bp) dsDNA (60 bp) or 30 μM nucleotides (nt) of ssDNA (60-mer). Reactions were initiated ATP and carried out at 25°C. To stop the reactions, aliquots of the mixtures were mixed with a final concentration of 25 mM EDTA and 0.1% SDS; then, 0.5 μL of each sample was spotted onto a polyethylenimine cellulose plastic sheet (Merck Millipore, Darmstadt, Germany) and analyzed by thin-layer chromatography with a development buffer comprising 0.1 M formic acid and 0.2 M LiCl at room temperature for 5 min. The plate was dried, placed in contact with an imaging plate, and then analyzed by a FLA-3000 image analyzer. The amounts of ATP and Pi were quantified using Multi Gauge version 2.1 software. The apparent rate constant k_{app} (= $v_0/[E]_0$) was estimated from the time-course analysis. The kinetic parameters k_{cat} and K_m were determined from three independent experiments by fitting to the following equation using the Igor Pro 4.03 software (WaveMetrics, Lake Oswego, OR).

$$k_{app} = k_{cat}[ATP]/(K_m + [ATP]) \quad \text{(Equation 2)}$$

where v_0, $[E]_0$, [ATP], k_{cat}, and K_m are the initial velocity, total enzyme concentration, free ATP concentration, catalytic rate constant, and Michaelis constant, respectively.

To determine the dissociation constants for DNA through the DNA-dependent ATPase activity, an ATPase assay was performed in reaction mixtures containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/mL bovine serum albumin, 0.1 μM RadA, 500 μM ATP spiked with 0.1 μCi of [γ-^{32}P]ATP, 0 to 30 μM (bp) dsDNA (60 bp) or 0 to 30 μM (nt) ssDNA (60-mer). The initiation and termination of the reactions and the analysis by thin-layer chromatography were performed in the same manner as described above. The apparent rate constant k_{app} was estimated from the time-course analysis. The apparent rate constant of the DNA-free enzyme, $k_{unbound}$, was determined from six independent experiments in the DNA-free condition. Under conditions with a large excess of DNA over RadA, the kinetic parameters k_{bound} and K_d were determined from three independent experiments by fitting to the following equation using the Igor Pro 4.03 software.

$$k_{app} = k_{unbound} + (k_{bound} - k_{unbound}) \times [DNA]/(K_d + [DNA]) \quad \text{(Equation 3)}$$

where [DNA], k_{bound}, and K_d are the total DNA concentration, apparent rate constant of the DNA-bound enzyme, and dissociation constant for DNA, respectively.

Culture conditions for *T. thermophilus* HB8

T. thermophilus HB8 was grown at 70°C in TR medium: 0.4% tryptone (Difco, Detroit, MI), 0.2% yeast extract (Nihon Pharmaceutical, Tokyo, Japan), and 0.1% NaCl (pH 7.4, adjusted with NaOH). To make TT plates, 1.5 mM $CaCl_2$, 1.5 mM $MgCl_2$, and 1.5% gellan gum (Nacalai Tesque, Kyoto, Japan) were added to TR medium.

Constructs for gene disruption and complementation

A *T. thermophilus* HB8 strain lacking *radA* (Δ*radA*) was constructed by substituting the target gene with a thermostable kanamycin-resistance gene, HTK, through natural transformation and HR as previously described (Hashimoto *et al*., 2001). A plasmid used for gene disruption was obtained from RIKEN BioResource Center. The plasmid (clone name TDs03G08) is a derivative of pGEM-T Easy (Promega, Madison, WI) and was constructed by inserting HTK flanked by 500 bp upstream and downstream sequences of the target gene. Gene disruption was confirmed by PCR using the isolated genomic DNA as a template and western blotting using cell lysates. The primer sequences used to confirm gene disruption are listed in Table 1.

For *in vivo* complementation analysis, gene expression of *radA* derivatives in *T. thermophilus* HB8 was carried out using a polycistronic transcription of the *radA* gene following a thermostable hygromycin B resistance gene under the isocitric acid dehydrogenase promoter from *T. aquaticus* YT1 as previously described (Fukui *et al*., 2008), with a slight modification. The pMK18::HygΔKm plasmid was constructed by self-ligation after inverse PCR to delete the HTK region of pMK18::Hyg. The pET-HisTEV/*radA*, pET-HisTEV/*radA*_R286A/R385A, and pET-HisTEV/*radA*_R305A/R314A/K345A plasmids were digested by XbaI and HindIII and then ligated into the complement site of the pMK18::HygΔKm plasmid to generate pMK18::HygΔKm::*radA*, pMK18::HygΔKm::*radA*_R286A/R385A, and pMK18::HygΔKm::*radA*_R305A/R314A/K345A. These plasmids and pMK18::HygΔKm, as a negative control, were transformed into the *T. thermophilus* HB8 wild-type and Δ*radA* strains. Protein expression in each strain was confirmed by western blotting.

Examination of the sensitivities to DNA-damaging agents

The sensitivity to DNA-damaging agents was measured as follows. To measure UV sensitivity, *T. thermophilus* HB8 cells in the mid-exponential growth phase were serially diluted by TR medium; then, the cells were spread onto TT plates and irradiated with 254 nm UV light at a dose of 54 J m^{-2} using a SLUV-4 lamp (254 nm; As One, Osaka, Japan). The numbers of colonies were counted after incubation at 70°C for 18 h. To measure mitomycin C sensitivity, serially diluted *T. thermophilus* HB8 cells were mixed with a 300 μg/mL mitomycin C solution (Nacalai Tesque, Kyoto, Japan) to a final concentration of 3 μg/ml. The cells were further incubated at 37°C for 30 min and spread onto TT plates. The numbers of colonies were counted after incubation at 70°C for 18 h. The survival fraction was determined by monitoring the colony-forming units (CFUs) of bacterial cells with UV or mitomycin C treatment and was expressed relative to the CFUs of untreated cells. Statistical analysis was performed on log10-transformed data of the surviving fractions from six independent experiments using Welch's *t*-test and the multiplicity was adjusted by the Holm–Bonferroni method (Holm, 1979).

Western blotting

T. thermophilus HB8 cells at the late-exponential growth phase were harvested by centrifugation. The cells were lysed by ultrasonication in a buffer containing 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 10 mM EDTA, and 10 × protease inhibitor cocktail (Nacalai Tesque, Kyoto, Japan). The cell debris was removed by centrifugation. Total protein concentrations in the cell lysates were measured and corrected by the Bradford method using the Bio-Rad Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). The cell lysates containing 50 μg total protein in each lane were resolved on 11% polyacrylamide gels and electroblotted onto a PVDF membrane (Millipore, Milford, MA). The membrane was incubated in a blocking solution containing 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, and 5% skimmilk for 30 min at room

temperature. After washing with a wash buffer containing 20 mM Tris-HCl (pH 7.5) and 500 mM NaCl, the membrane was immersed into a blocking solution containing rabbit anti-RadA and anti-single-stranded DNA-binding protein (SSB) antisera, and then incubated for 2 h at room temperature. After washing with the wash buffer, the membrane was reacted with Goat Anti-Rabbit IgG-AP Conjugate (Bio-Rad Laboratories) in the blocking solution for 1 h at room temperature. The membrane was washed twice with the wash buffer and then reacted with AP color reagents in a development buffer (Bio-Rad Laboratories) for 15 min at room temperature. The color development was stopped by washing with deionized water.

Phylogenetic analysis

A multiple sequence alignment was prepared by the Muscle program (Edgar, 2004) using 64 LonC domain sequences consisting of 26 Lon protease, 21 RadA, and 17 ComM proteins. Then, a phylogenetic tree was created with the Mega program (Kumar *et al*., 2016) using the neighbor-joining method (Saitou & Nei, 1987). The optimal tree with the sum of branch length = 18.98 is shown in Fig. 15. All of the gaps and missing data were eliminated. The number of total positions was 136 in the final dataset.

RESULTS

Overall structure of rLonC

First, I performed structural prediction and multiple sequence alignment analyses to determine the domain organization of RadA homologs (Fig. 4). Structural prediction of *T. thermophilus* HB8 RadA (423 residues) using the SWISS-MODEL (Biasini *et al*., 2015) and the I-TASSER (Yang *et al*., 2015) servers showed that a region of amino acid residues 256–275 of RadA from *T. thermophilus* HB8 could be a linker region between the RecA-like ATPase domain and rLonC (Fig. 4*D*). Based on this prediction, I prepared rLonC (amino acid residues 262–423) in N-terminal His_6-tagged form for X-ray crystallographic analysis (Fig. 5*A*). The crystal structure of rLonC was determined at 2.7 Å resolution by molecular replacement using the structural model predicted by the I-TASSER server (Yang *et al*., 2015) as a search model. The statistics for the data collection and model refinement are summarized in Table 2.

The overall structure of rLonC was composed of a core structure, ribosomal protein S5 domain 2-type βββαβα fold, and an extra C-terminal α/β subdomain (Fig. 6*A*, *left panel*). The structure of rLonC was similar to that of pLonC from *E. coli* (Botos *et al*., 2004) as denoted by a root mean square (r.m.s.) deviation of 3.9 Å for Cα atoms. The extra C-terminal α/β subdomain of rLonC was shorter than that of pLonC, and two α-helices and one β-sheet in the extra C-terminal α/β subdomain of pLonC were absent in the rLonC structure (Fig. 6).

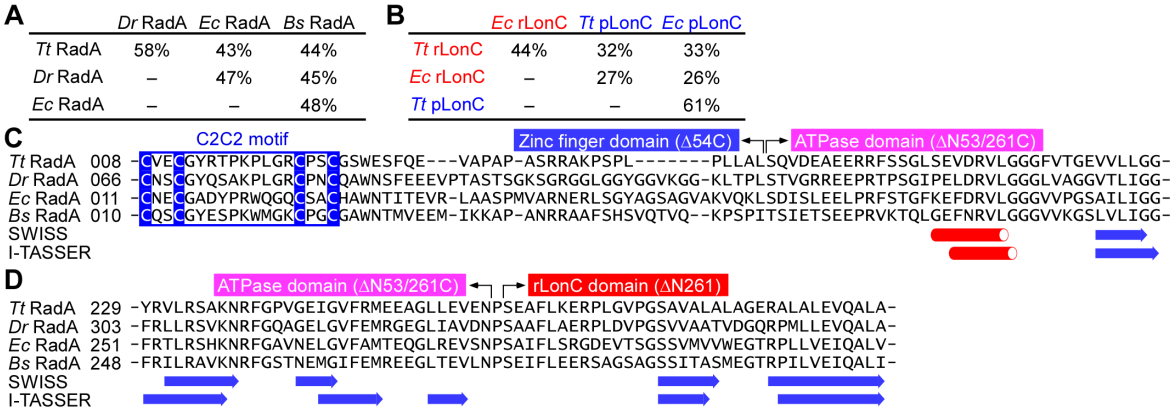

FIGURE 4. Prediction of the interdomain regions of RadA for domain-truncated mutagenesis. (A and B) Sequence identities of RadA homologs (A) and LonC domains (B) are summarized. (C and D) Multiple sequence alignment and predicted structural elements of the interdomain regions between the ZF and ATPase domains (C) and the ATPase and rLonC domains (D) are shown. The alignment was generated by Muscle (Edgar, 2004). The accession numbers of the sequences are as follows: *T. thermophilus* (*Tt*) RadA (YP_143807), *D. radiodurans* (*Dr*) RadA (NP_294829), *E. coli* (*Ec*) RadA (NP_418806), *B. subtilis* (*Bs*) RadA (NP_387968), *Tt* Lon (YP_144036), and *Ec* Lon (NP_414973). The secondary structure elements of *Tt* RadA predicted by the SWISS-MODEL (Biasini *et al*., 2015) and the I-TASSER (Yang *et al*., 2015) servers are shown below the alignments. The cylinders (*red*) and arrows (*blue*) represent α helices and β strands, respectively.

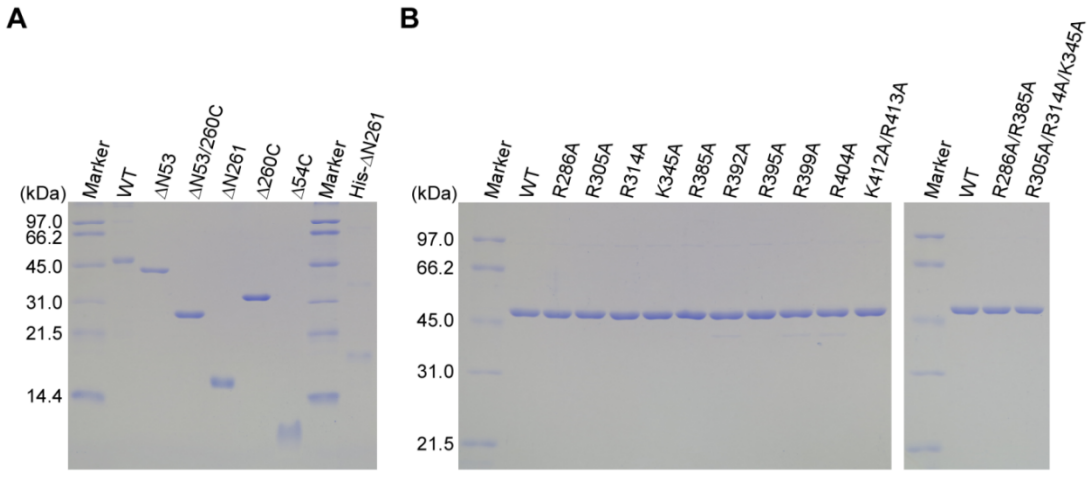

FIGURE 5. Purification of RadA and its derivatives. (A) The domain-truncated mutants were analyzed by 15% SDS-PAGE containing 7 M Urea. Note that ΔN261 and His-ΔN261 are rLonC and His6-tagged rLonC, respectively. (B) The site-directed mutants were analyzed by 11% SDS-PAGE. WT, wild type.

TABLE 2 Data collection and refinement statistics

	rLonC
Data collection[a]	
Beamline	SPring-8 BL38B1
Space group	$P4_332$
Cell dimensions	
a, *b*, *c* (Å)	156.29, 156.29, 156.29
α, β, γ (°)	90.00, 90.00, 90.00
Wavelength (Å)	1.00000
Resolution (Å)	49.4-2.70 (2.75-2.70)[b]
R_{pim}[c]	0.051 (0.486)
$I/\sigma I$	20.4 (1.9)
$CC_{1/2}$[d]	0.625
Completeness (%)	100 (100)
Redundancy	41.3 (41.6)
Refinement	
Resolution (Å)	49.4-2.70 (2.75-2.70)
No. of reflections	18,509
R_{work}/R_{free}[e]	0.197/0.250
No. of atoms	
Protein	2,288
Water	52
B-factors	
Protein	53.3
Water	49.8
r.m.s. deviations[f]	
Bond lengths (Å)	0.0126
Bond angles (°)	1.62
Ramachandran plot (%)[g]	
Favored	98.4
Allowed	1.6
Outlier	0
Protein Data Bank Code	5H45

[a]One crystal was used for each data set.

[b]The values in parentheses are for the highest resolution shell.

[c]$R_{pim} = \sum_{hkl} (N-1)^{-1/2} \sum_i |I_i(hkl) - \langle I_i(hkl) \rangle| / \sum_{hkl} \sum_i I_i (hkl)$, where $I_i(hkl)$ is the intensity of *i* th measurement of reflection *hkl*, $\langle I_i(hkl) \rangle$ is the corresponding mean value of all *i* measurements, and *N* is the multiplicity of each *hkl*.

[d]Values of the highest resolution shell.

[e]$R_{work} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} was calculated as R_{work} using a randomly selected 5.4% of reflection data excluded from refinement.

[f]Root mean square deviations.

[g]The Ramachandran plot was validated by RAMPAGE (Lovell *et al*., 2003).

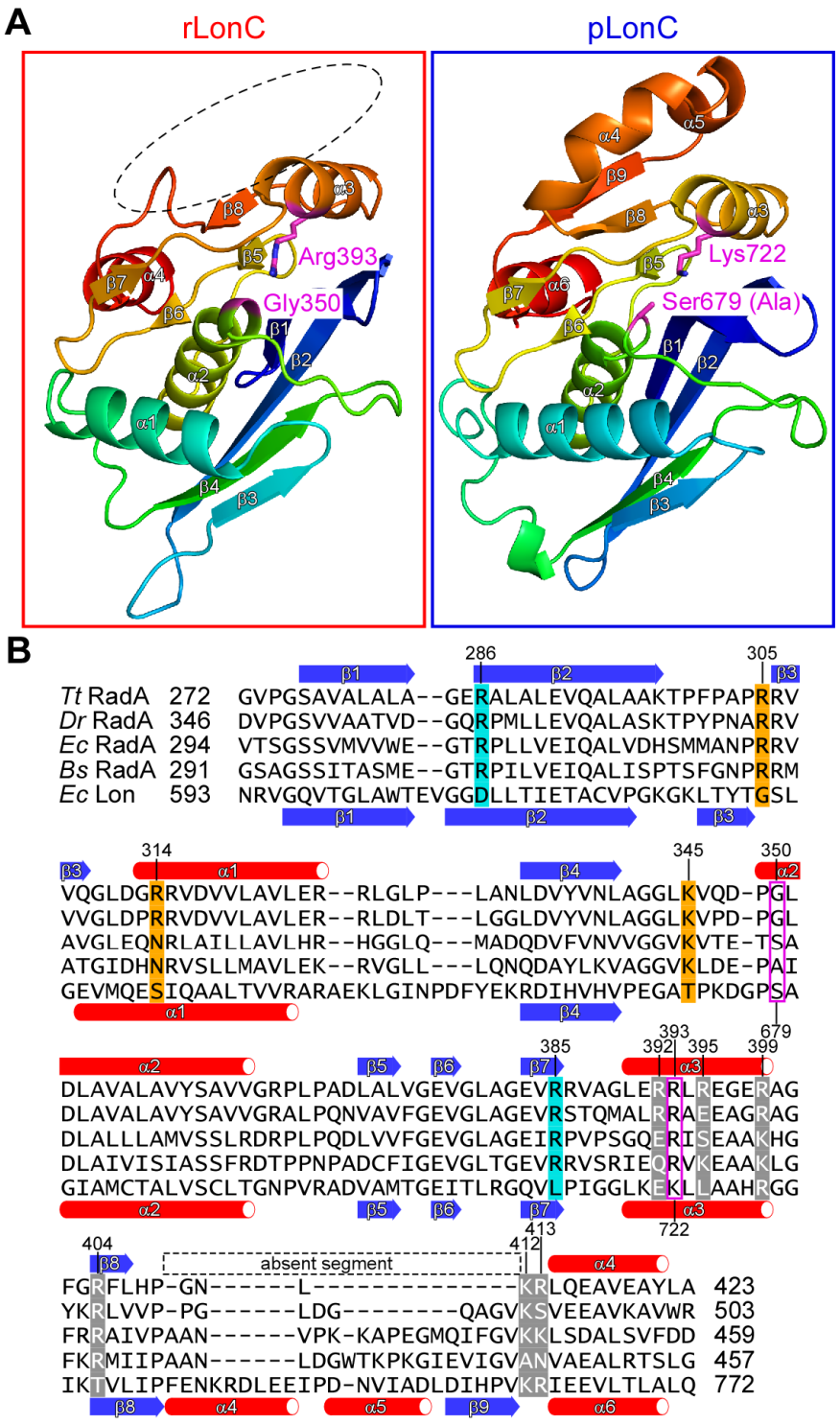

FIGURE 6. Structure of rLonC. (A) The overall structures of *T. thermophilus* rLonC (*left panel*) and *E. coli* pLonC (*right panel*; PDB ID: 1RR9) (Botos *et al.*, 2004) are shown in a cartoon representation, colored in a spectrum from the N-terminus (*blue*) to the C-terminus (*red*). The side chains of the two residues (Ser679 and Lys722) forming the catalytic dyad of *E. coli* pLonC and the corresponding residues (Gly350 and Arg393, respectively) of rLonC are represented as *magenta sticks*. In the *E. coli* pLonC structure, Ser679 is replaced with Ala to

avoid self-digestion. The segment absent in rLonC is indicated as a *dotted oval*. (B) Multiple sequence alignment of the LonC domains with secondary structures. The accession numbers of the sequences are as follows: *T. thermophilus* (*Tt*) RadA (YP_143807), *D. radiodurans* (*Dr*) RadA (NP_294829), *E. coli* (*Ec*) RadA (NP_418806), *B. subtilis* (*Bs*) RadA (NP_387968), and *Ec* Lon (NP_414973). The locations of the secondary structure elements of *T. thermophilus* rLonC and *E. coli* pLonC are shown above and below each amino acid sequence, respectively. The cylinders (*red*) and arrows (*blue*) represent α helices and β strands, respectively. The segment absent in rLonC is indicated as a *dotted rectangle*. The catalytic residues, Ser679 and Lys722, of *E. coli* Lon protease and the corresponding residues in the RadA orthologs are boxed in *magenta*. The residues for the site-directed mutagenesis of *T. thermophilus* RadA and the corresponding residues in other LonC domains are highlighted in color backgrounds as follows (related to Fig. 12): *cyan*, the residues that showed decreased DNA-binding activity by mutation to Ala at the intersubunit cleft; *yellow*, the residues that showed decreased DNA-binding activity by mutation to Ala at the central hole; and *gray*, the residues that showed similar DNA-binding activity to the wild type by mutation to Ala at the outer wall and peripheral regions.

In *E. coli* Lon protease, Ser679 is a catalytic residue of the protease activity and forms a catalytic dyad with Lys722 (Botos *et al*., 2004). On the other hand, in the rLonC structure, Gly350 and Arg393 were located at the positions corresponding to Ser679 and Lys722 of *E. coli* Lon protease, respectively (Fig. 6, *A* and *B*). Furthermore, there was no serine residue around Gly350 or Arg393. These findings are consistent with a previous report that *E. coli* RadA had no protease activity (Beam *et al*., 2002).

Oligomeric structure of rLonC

In a crystal, rLonC forms a hexameric ring structure related by crystallographic 3-fold symmetry of non-crystallographic dimer with an r.m.s. deviation of 0.16 Å for Cα atoms (Fig. 7, *A* and *B*). The difference between the rotation angles of the two subunit interfaces was 1°, suggesting that the subunits were related by an approximate 6-fold symmetry (Fig. 7*B*). It should be noted that the N-terminal residues exhibited high *B*-factor values (Fig. 7*C*). The N-terminal regions including His_6-tags were disordered in two asymmetric molecules in the crystal. The disordered N-terminal region of one molecule in the asymmetric unit was longer by five residues (residues 262–271) than that of the other molecule (residues 262–266). These data imply that this N-terminal region could be a linker between the rLonC and the ATPase domains, in agreement with the prediction of the domain organization.

The subunit arrangement of the hexameric ring structure of rLonC was similar to that of *E. coli* pLonC (Fig. 7*A*) (Botos *et al*., 2004). In addition, like *E. coli* pLonC, the predominant interactions in the subunit interface are hydrophilic, including 11 hydrogen bonds and 6 to 7 salt bridges in each subunit interface. The solvation free energies in respective subunit interfaces formed by the two neighboring molecules were calculated to be −5.5 and −6.5 kJ/mol, which does not include the effect of satisfied hydrogen bonds and salt bridges. The total buried accessible surface area of rLonC upon hexamer formation was 10,046 $Å^2$, comparable to the

9,169 Å^2 for *E. coli* pLonC (Botos *et al*., 2004). The hexameric ring was predicted to be the most stable multimer in the crystal structure by the PISA program (Krissinel & Henrick, 2007). By contrast, gel-filtration analyses revealed that in solution, the rLonC domain and full-length RadA existed as trimers to tetramers and dimers to pentamers, respectively (Fig. 8). In addition, the dissociation of the oligomeric structure was observed in a salt-dependent manner, suggesting the ionic interactions are predominant in the subunit interface. These behaviors in the oligomer formation of RadA are similar to those of Lon proteases, in that they form various types of oligomers (Rudyak *et al*., 2001; Vieux *et al*., 2013; Lee *et al*., 2004; Rasulova *et al*., 1998).

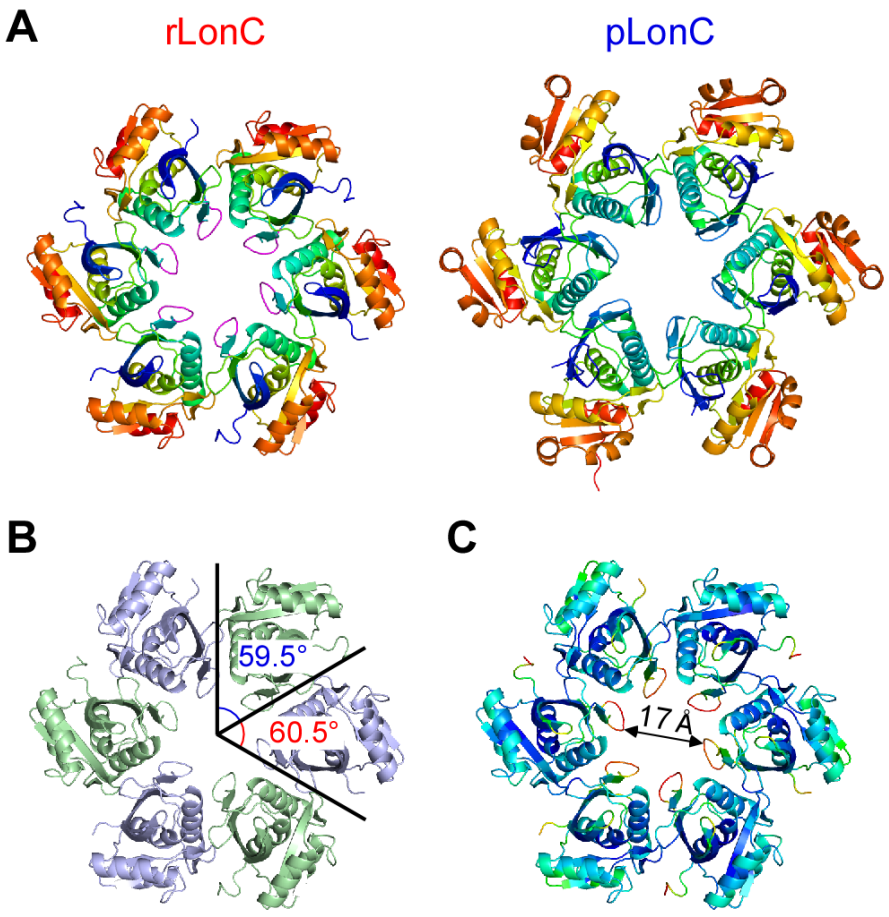

FIGURE 7. The oligomeric structure of rLonC. (A) The hexameric structures of *T. thermophilus* rLonC (*left panel*) and *E. coli* pLonC (*right panel*) are shown in a cartoon representation. Each subunit is colored in a spectrum from the N-terminus (*blue*) to the C-terminus (*red*). The loop containing 299-TPFPAP-304 is colored in *magenta*. (B) The organization of the subunits in the hexamer of rLonC. The asymmetric two molecules are shown in different colors (*blue* and *green*). Rotation angles of two subunit interfaces are indicated. (C) The *B*-factors of the Cα atoms are colored from 31 $Å^2$ (*blue*) to 109 $Å^2$ (*red*). The smallest inner diameter of the central hole of the ring is indicated.

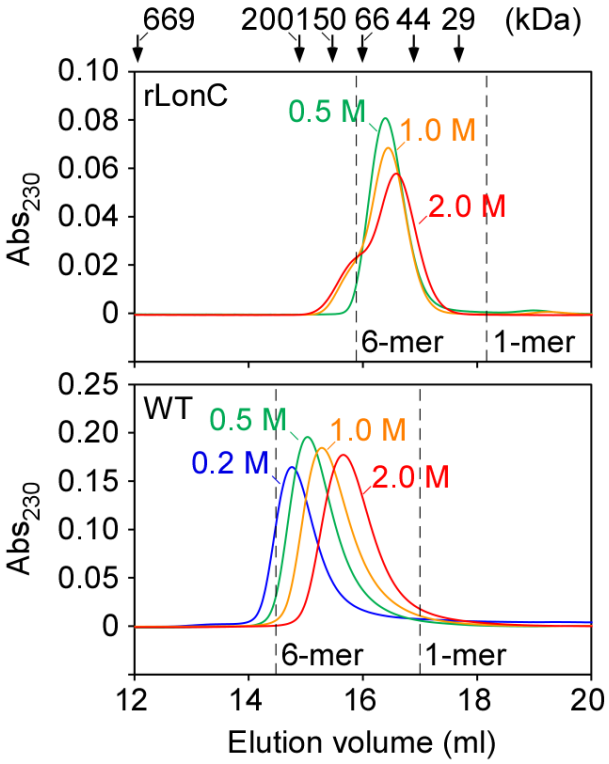

FIGURE 8. Gel filtration analysis of the rLonC domain and full-length RadA. The elution profiles of the rLonC domain (ΔN261) and full-length RadA (WT) in various salt concentrations are shown in the *top* and *bottom panels*, respectively. The protein concentration applied to the column was 5 μM. The NaCl concentrations are labeled. The elution peaks and molecular masses of standard proteins are indicated as *black arrows* and labeled in kDa, respectively. The estimated elution peaks of monomer and hexamer by the calibration curve is indicated as *dotted lines*. Under the condition of 0.2 M NaCl, the rLonC domain was not eluted because of adsorption to the filter or the resin.

Unlike in *E. coli* pLonC (Fig. 9*A*, *right panel*), the surface electrostatic potential map showed that rLonC had positively charged regions at one side of the ring (Fig. 9*A*, *left panel*). At this "basic" side of the ring, a cleft-like structure is formed at the subunit interface (hereinafter referred to as the "intersubunit cleft") (Fig. 9*B*, *left panel*). At the other side of the ring, there was no significant biased distribution of charged residues (Fig. 9*A*, *left panel*). In addition, rLonC had another positively charged region at the central hole of the ring (hereinafter referred to as the "central hole") (Fig. 9*B*, *right panel*). On the basic side, the inner diameter of the entrance of the central hole is approximately 29 Å. The hole is gradually narrower from the basic side to the other side of the ring, and the inner diameter at the exit formed by the loop containing 299-TPFPAP-304 is approximately 17 Å (Figs. 7*C* and 9*B*, *right panel*). On the basis of these unique features, especially the positively charged regions, of rLonC's structure, I hypothesized that rLonC has a DNA-binding ability.

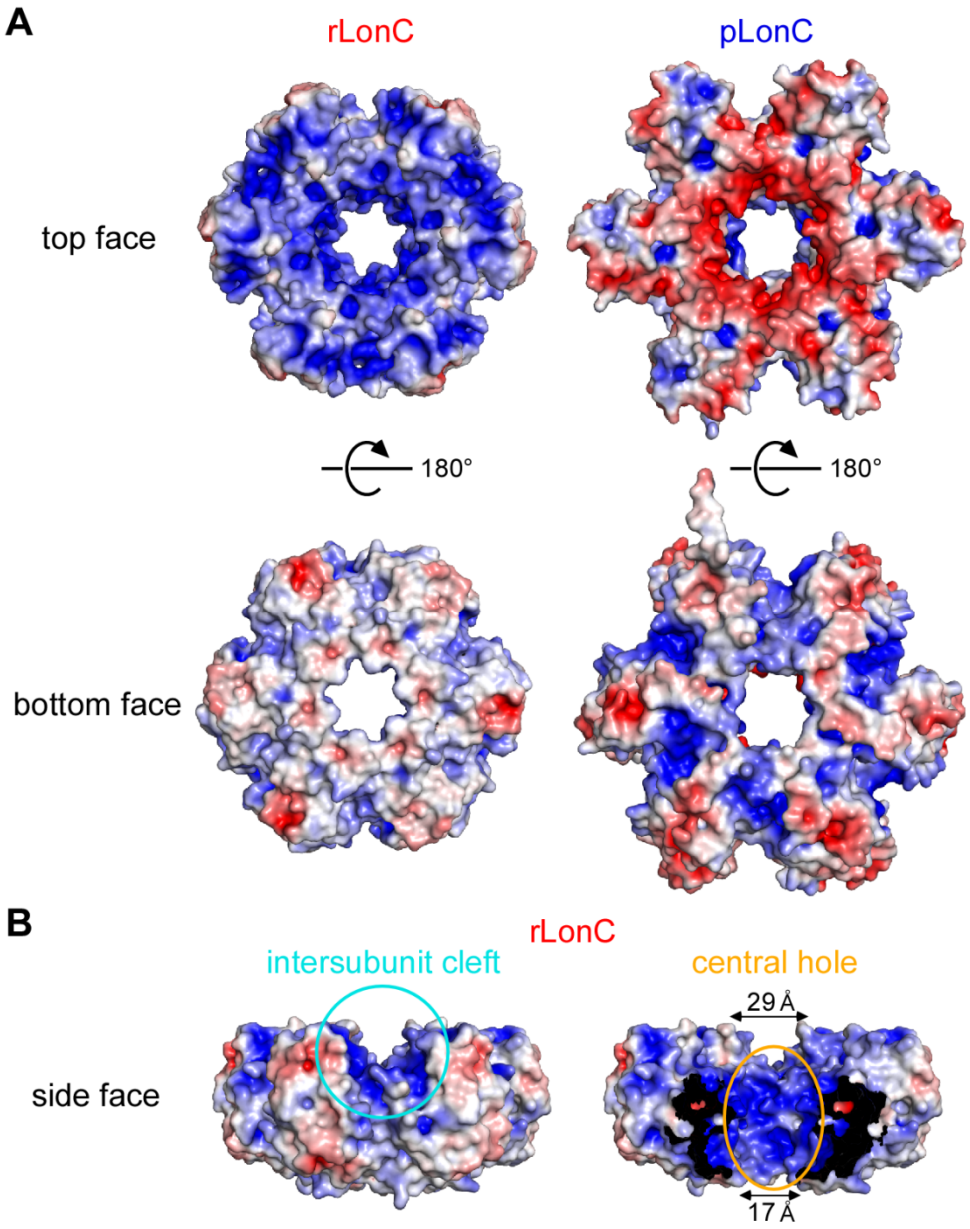

FIGURE 9. The electrostatic potential of rLonC. (A) The surface electrostatic potentials of *T. thermophilus* rLonC (*left panel*) and *E. coli* pLonC (*right panel*) are shown in the same view as in (Fig. 7*A*) and in a view rotated by 180° around the horizontal axis (*bottom panels*), contoured in the range from –5 kT/e (*red*) to +5 kT/e (*blue*). (B) The two unique positively charged regions of *T. thermophilus* rLonC—the intersubunit cleft (*left panel*) and the central hole (*right panel*)—are shown in a view rotated by 90° from (A, *top left panel*) around the horizontal axis. In the right panel, two subunits are removed to show the inner surface of the central hole and the largest and smallest inner diameters are indicated at top and bottom faces, respectively.

rLonC is a novel DNA-binding module

To verify the hypothesis that rLonC has DNA-binding activity, I constructed and purified various domain-truncated mutants of RadA (Fig. 5*A*) and performed an EMSA using 60 bp of double-stranded DNA (dsDNA) (Fig. 10). As expected, the mutants containing the rLonC domain exhibited DNA-binding activity. In the presence of the rLonC domain, as the protein concentration increased, bands of DNA-protein complex appeared while those of free DNA molecules disappeared (Fig. 10*A*). Binding of the wild type, rLonC (ΔN261) and ΔN53 to DNA showed positive cooperativity and the dissociation constants of the wild type and rLonC were comparable (Fig. 10*B* and Table 2). On the other hand, the loss of the N-terminal ZF domain (ΔN53) resulted in considerably lower DNA-binding activity than the wild type.

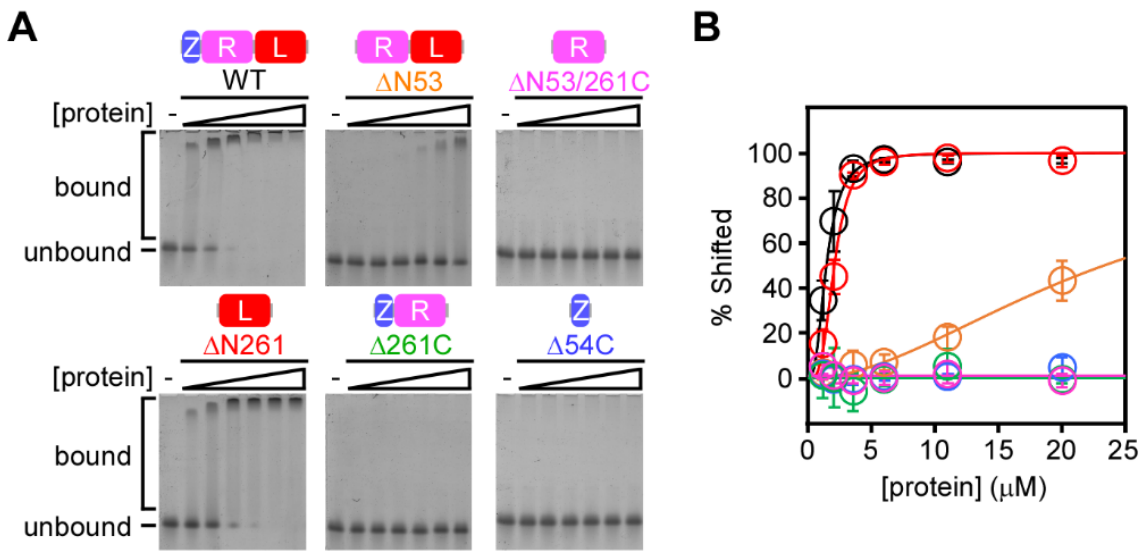

FIGURE 10. The DNA-binding activity of RadA domain-truncated mutants. (A and B) The DNA-binding activity of the domain-truncated mutants of RadA was assessed by EMSA. The representative gels are shown in (A). The protein concentrations of wild-type (WT) RadA and RadA mutants were 1.1, 2.0, 3.5, 6.0, 11, and 20 μM. Schematic representations of the domain organization of each mutant are shown above the panel as follows: *light blue*, a ZF domain (*Z*); *pink*, a RecA-like ATPase domain (*R*); and *red*, a LonC domain (*L*). The quantitative results of the protein concentration dependence in DNA binding are shown in (B). The WT and mutants are colored as follows: *black*, WT; *orange*, ΔN53; *pink*, ΔN53/260C; *red*, ΔN261; *green*, Δ260C; and *light blue*, Δ54C. The error bars indicate the standard deviation of the percent shifted from three independent experiments. The DNA-binding affinities estimated by the experiments are summarized in Table 3.

TABLE 3 The DNA-binding affinities determined from EMSA[a]

	K_d[b]	h[b,c]
	μM	
WT	1.4 ± 0.1	2.6 ± 0.3
Domain truncation		
ΔN53	23 ± 2	1.9 ± 0.3
ΔN53/260C	-[d]	-
ΔN261	2.0 ± 0.1	3.3 ± 0.3
Δ260C	-	-
Δ54C	-	-
Site-directed mutagenesis		
R286A	3.0 ± 0.1	3.0 ± 0.3
R305A	2.7 ± 0.1	2.9 ± 0.4
R314A	2.8 ± 0.1	3.3 ± 0.4
K345A	3.0 ± 0.2	2.7 ± 0.4
R385A	4.7 ± 0.2	2.7 ± 0.3
R392A	1.5 ± 0.1	2.8 ± 0.3
R395A	1.4 ± 0.1	2.1 ± 0.3
R399A	1.4 ± 0.0	2.4 ± 0.2
R404A	1.3 ± 0.0	2.6 ± 0.2
K412A/R413A	1.3 ± 0.0	2.4 ± 0.1
R286A/ R385A	10 ± 1	1.5 ± 0.2
R305A/R314A/K345A	11 ± 1	1.8 ± 0.2

[a]The results are shown in Figs. 10*B* and 12*B*.

[b]The errors are standard errors from the regression analysis.

[c]Hill coefficient.

[d]No binding activity was detected.

Interestingly, in the absence of rLonC domain, the bands corresponding to the DNA-protein complex were not observed: The RecA-like ATPase domain and/or the N-terminal ZF domain were insufficient for DNA binding of RadA (Fig. 10). Structural prediction and multiple sequence alignment analyses support this finding in the RecA-like ATPase domain of RadA (Fig. 11, *A* and *B*). The L1 and L2 loops of RecA have been thought to be important for DNA binding (Maraboeuf *et al*., 1995; Shinohara *et al*., 2015). However, the corresponding loops of RadA are expected to be shorter than those of RecA, and the residues that are important for DNA binding are not conserved (Fig. 11, *A* and *B*). It should be noted that the loss of the N-terminal ZF domain alone resulted in lower DNA-binding activity than the wild type, whereas the rLonC domain showed the DNA-binding activity comparable to the wild type (Fig. 10*B*). These data imply that the ZF domain is necessary for the DNA-binding activity of the full-length form of RadA. Further analysis is required to elucidate whether the N-terminal ZF domain and the ATPase domain have DNA-binding ability.

It has been shown that Lon protease activity is modulated by DNA binding and that the DNA-binding ability resides in the α subdomain of the N-terminal AAA^+ domain, but not in the pLonC domain (Lee *et al*., 2014; Lin *et al*., 2009; Lee *et al*., 2004). As mentioned above, rLonC has unique positively charged regions that are not found in pLonC. Thus, I expected that the positively charged regions at the surface of rLonC's structure have an important role in the DNA-binding ability.

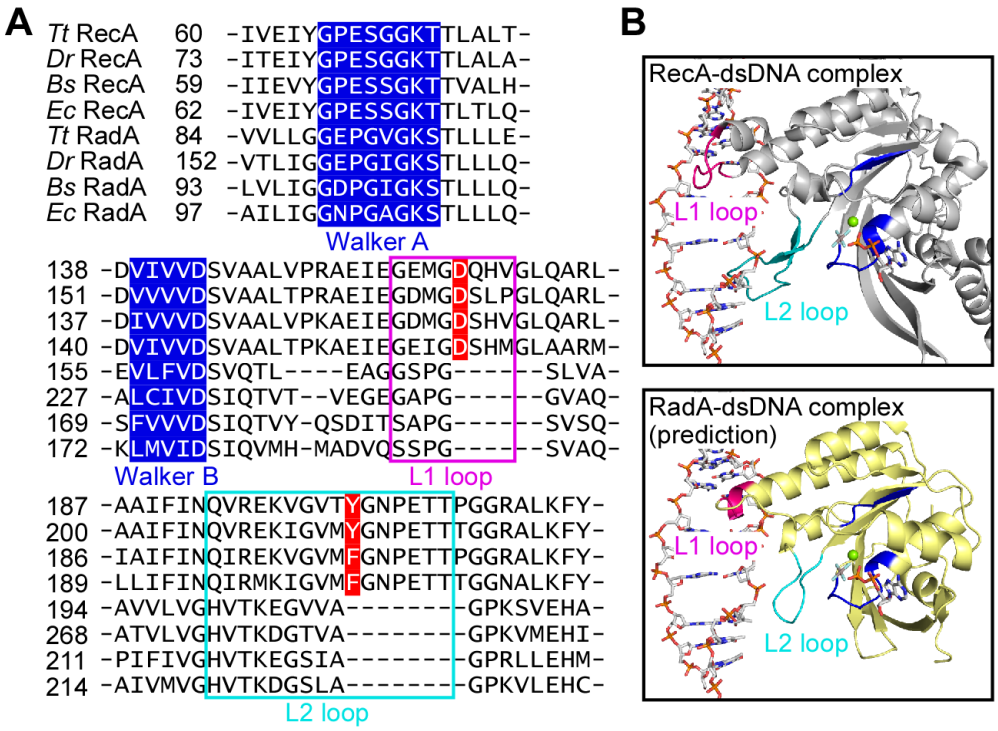

FIGURE 11. Comparison of the ATPase domain of RadA with that of RecA. (A) The multiple sequence alignment of the ATPase domain of RecA and RadA. The alignment was generated by Muscle (Edgar, 2004). The Walker-type ATPase motifs and the conserved residues that are important for DNA binding are highlighted with *blue* and *red backgrounds*, respectively. The L1 and L2 loops are boxed in *magenta* and *cyan*, respectively. The accession numbers of the sequences are as follows: *T. thermophilus* (*Tt*) RecA (YP_145084), *Deinococcus radiodurans* (*Dr*) RecA (NP_296061), *Bacillus subtilis* (*Bs*) RecA (NP_389576), *E. coli* (*Ec*) RecA (NP_417179), *Tt* RadA (YP_143807), *Dr* RadA (NP_294829), *Bs* RadA (NP_387968), and *Ec* RadA (NP_418806). (B) Structural comparison of the RecA-dsDNA complex (PDB ID, 3CMX) from *E. coli* (*top panel*) with the RecA-like ATPase domain of RadA from *T. thermophlius* HB8 (residues 54–260) (*bottom panel*). The structural model of the ATPase domain of RadA was created using the I-TASSER server (Yang *et al.*, 2015). The predicted structure of the ATPase domain of RadA complex with dsDNA was obtained by alignment with the RecA-dsDNA complex (r.m.s. deviation = 2.7 Å for Cα atoms). The L1 and L2 loops are colored in *magenta* and *cyan*, respectively.

Identification of the DNA-binding residues

To identify the residues responsible for DNA binding, I selected 11 basic surface residues and performed alanine-scanning mutagenesis using full-length RadA (Fig. 12*A*). Nine single mutants—R286A, R305A, R314A, K345A, R385A, R392A, R395A, R399A, and R404A—and one double mutant, K412A/R413A, were constructed and purified (Fig. 5*B*) to verify the DNA-binding activity by EMSA. Five mutants—R286A, R305A, R314A, K345A and R385A—exhibited slightly decreased DNA-binding activities compared to the wild type, while the others exhibited DNA-binding activities similar to the wild type (Fig. 12*B* and Table 3). The dissociation constants of R286A, R305A, R314A, K345A and R385A mutants were 1.9- to 3.4-fold higher than that of the wild type. Two residues—Arg286 and Arg385—and three residues—Arg305, Arg314, and Lys345—were located at the unique positively charged regions, namely the intersubunit cleft and the central hole, respectively (Figs. 9*B* and 12*B*). On the other hand, the other six residues were located at the outer wall and peripheral regions of the ring. These results suggest that both the intersubunit cleft and the central hole participate in DNA binding and that the outer wall and peripheral regions do not. Interestingly, both of the possible DNA-binding regions are located at subunit interface and formed by dimerization, suggesting the requirement of dimerization for DNA binding (Fig. 12*A*).

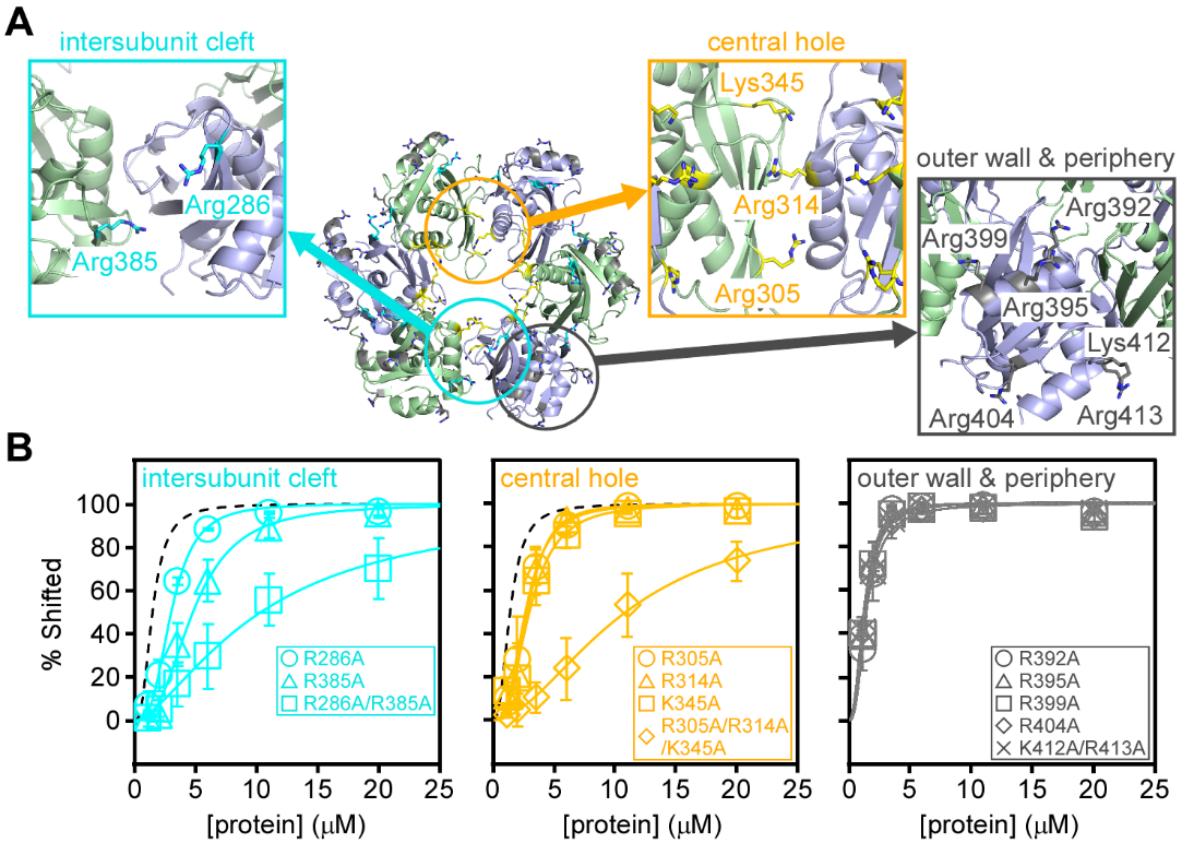

FIGURE 12. Identification of the DNA-binding sites of RadA. (A) Location of the residues for site-directed mutagenesis. An overall view of the hexameric ring and close-up views (*rectangles*) of three different regions (*circles*) are shown. The *bold arrows* indicate the close-up operation with small changes in viewing direction. The two molecules in the asymmetric unit are shown in different colors. The residues are shown as follows (related to Fig. 1*B*): *cyan*, the residues at the intersubunit cleft; *yellow*, the residues at the central hole; and *gray*, the residues at the outer wall and peripheral regions. (B) The DNA-binding activity of the site-directed mutants of RadA was assessed by EMSA. The mutants of the residues at the intersubunit cleft (*cyan*), central hole (*orange*), and outer wall and peripheral regions (*gray*) are shown in the *left*, *middle*, and *right panels*, respectively. The *insets* show the mutants and their respective symbols. The theoretical curves of the WT result are shown as a *black dotted line* (related to Fig. 10*B*). The error bars indicate the standard deviation of the percent shifted from three independent experiments. The DNA-binding affinities estimated by the experiments are summarized in Table 3.

Multiple sequence alignment shows that these possible DNA-binding residues, except for Arg314 (replaced with Asn in most organisms), are highly conserved in RadA orthologs (Fig. 6*B*, *cyan and yellow backgrounds*). By contrast, these basic residues are not conserved in Lon proteases, suggesting that the conservation of these residues is a unique feature of RadA. This finding supports my results that both the intersubunit cleft containing Arg286 and Arg385 and the central hole containing Arg305, Arg314 and Lys345 are important for the DNA-binding activity of RadA via rLonC (Fig. 12).

Furthermore, I constructed R286A/R385A and R305A/R314A/K345A mutants because each single mutation of these residues exhibited significant but only limited effects on DNA-binding activity. As a result, both R286A/R385A and R305A/R314A/K345A mutants exhibited lower DNA-binding activity than each single mutant (Fig. 12*B* and Table 2). The dissociation constants of R286A/R385A and R305A/R314A/K345A were 7.3- and 7.7-fold higher than that of the wild type. These data imply that each mutation has an independent effect on the DNA-binding activity. Based on these results, I concluded that both of the intersubunit cleft and the central hole are DNA-binding sites.

The DNA-binding activity of rLonC is essential for RadA-mediated DNA repair

One of the most important questions is whether the DNA-binding activity of rLonC is necessary for the DNA repair function of RadA *in vivo*. The previous results have shown that rLonC domain-truncated mutants cannot complement the phenotypes of disruptants of the *radA* genes in *E. coli* or *D. radiodurans* (Cooper *et al*., 2015; Qing *et al*., 2006). However, there is no evidence that the DNA-binding activity of rLonC is important for the DNA repair function

of RadA. To address the effect of DNA-binding activity of rLonC on RadA-mediated DNA repair *in vivo*, I performed an *in vivo* complementation assay using two DNA-binding-deficient mutants, namely R286A/R385A and R305A/R314A/K345A, in *T. thermophilus* HB8.

First, I constructed a Δ*radA* strain of *T. thermophilus* HB8 using a gene-disruption method via natural transformation and HR (Hashimoto *et al*., 2001) (Fig. 13*A* and *B*). To verify the involvement of RadA in DNA repair, the phenotype of Δ*radA* was characterized by measuring the sensitivity to UV and mitomycin C, which introduce DNA strand-breaks via DNA adducts and cross-links (Fig. 13*C*). Δ*radA* was approximately 30-fold more sensitive than the wild type to both UV and mitomycin C treatments. These results imply that RadA functions as a DNA-repair protein in *T. thermophilus* HB8 like it does in other organisms, such as *E. coli* and *D. radiodurans* (Cooper *et al*., 2015; Qing *et al*., 2006).

The DNA-damage sensitivities of Δ*radA* were recovered at the same level as those of wild type by introducing the expression plasmid of wild-type RadA (Fig. 13*C*). This indicates that the RadA protein expressed by the plasmid was fully functional *in vivo* and that the DNA-damage sensitivities of Δ*radA* were caused by its loss of RadA. On the other hand, overexpression of wild-type RadA in the wild-type strain showed no effect on sensitivity to UV and to mitomycin C (Fig. 13, *B* and *C*). It should be noted that in the western blotting analysis, the bands of degraded RadA were observed even in the presence of protease inhibitors (Fig. 13*B*). The His-tag-cleaved RadA (labeled with "RadA" in Fig. 13*B*) was observed in the strains that were complemented with plasmids encoding the His-tagged RadA.

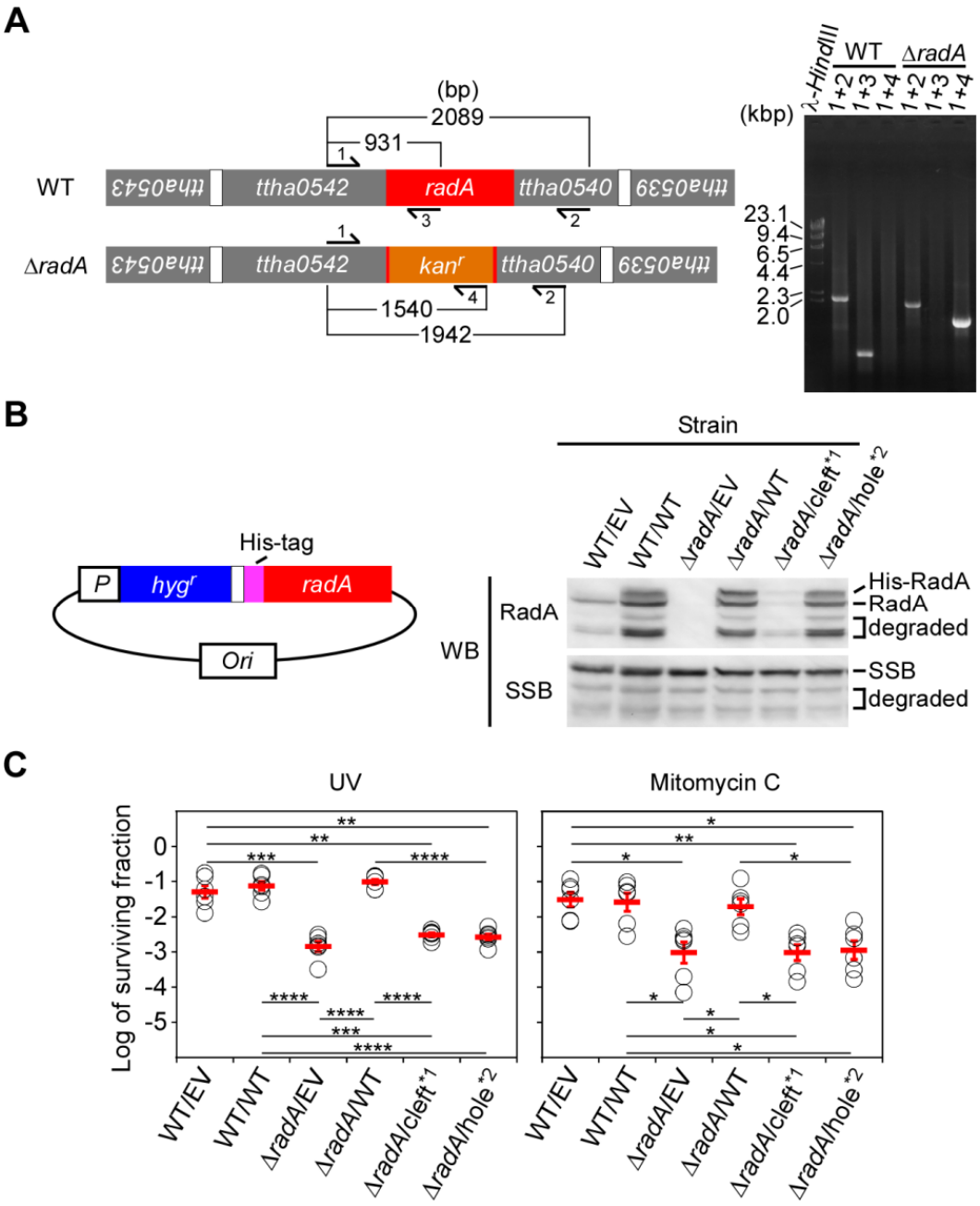

FIGURE 13. Gene disruption and *in vivo* complementation assays using the DNA-binding deficient mutants. (A) Schematic representations of the genomic regions around the *radA* gene of the wild type and Δ*radA* (*left panel*) and confirmation of gene disruption by genomic PCR. Agarose gel electrophoresis of the PCR products (*right panel*) is shown. The primers used in the PCR are indicated by *black arrows*. The theoretical lengths (bp) of the PCR products and the primer sets are shown. The sequences of the primers are listed in Table 1. (B) A schematic representation of plasmid construction for *in vivo* complementation (*left panel*) and estimation of protein expression levels in *T. thermophilus* HB8 cells by western blotting analysis using anti-RadA and anti-SSB antisera (*right panel*). (C) The effects of DNA-binding deficient

mutations to UV (*left panel*) and mitomycin C (*right panel*) sensitivity in *T. thermophilus* HB8 cells. The sensitivity is shown as log10-transformed survival fraction. The mean value (*black line*) of six independent experiments and each of their values (*gray circle*) are indicated. The error bar indicates the standard error of the mean. Statistical analysis was performed using Welch's *t*-test, and multiplicity was adjusted using Holm–Bonferroni method: **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. Statistical comparisons evaluated as "not significant" are not indicated. The "cleft[*1]" and "hole[*2]" show the R286A/R385A[*1] and R305A/R314A/K345A[*2] mutants, respectively. EV, empty vector; WT, wild type.

By contrast, neither the expressions of R286A/R385A nor of R305A/R314A/K345A mutants were able to rescue the DNA-damage sensitivities of Δ*radA* (Fig. 13*C*). These data suggest that these five residues located at the intersubunit cleft and the central hole are important for the proper function of RadA and that the DNA-binding ability of RadA via rLonC is essential for RadA-mediated DNA repair *in vivo*. It should be noted that the expression level of R286A/R385A mutant in Δ*radA* was apparently lower than that of the wild type and R305A/R314A/K345A mutant (Fig. 13*B*), implying the structural instability of the R286A/R385A mutant *in vivo*. The low expression level of R286A/R385A mutant might cause inability to rescue the DNA-damage sensitivities of Δ*radA*.

Biochemical characterization of the DNA-binding-deficient mutants

The next question is whether there are any functional differences between the two DNA-binding sites. To address this question, I examined the DNA-dependent ATPase activity of wild-type RadA and the DNA-binding-deficient mutants, namely R286A/R385A and R305A/R314A/K345A. In the case of wild type, the turnover rate was stimulated 4.5- and 2.2-fold in the presence of dsDNA and single-stranded DNA (ssDNA), respectively (Fig. 14*A* and Table 4). In *E. coli*, it has been shown that dsDNA and ssDNA stimulate the ATPase activity of RadA at the same level (Cooper & Lovett, 2016). My data suggest that dsDNA and ssDNA have different actions on the ATPase activity of RadA (Fig. 14, *A* and *B*). By contrast, the Michaelis constant for ATP was not affected by the addition of DNA, and the dissociation constants for dsDNA and ssDNA were comparable (Fig. 14, *A* and *B*; Tables 4 and 5).

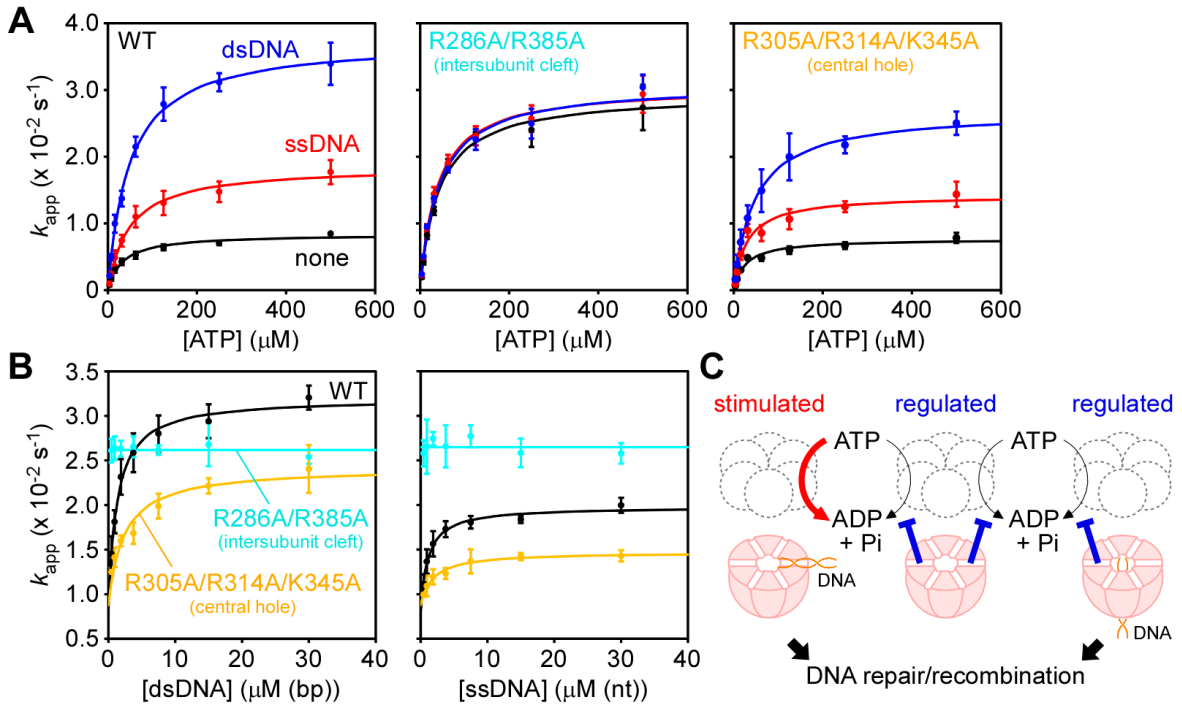

FIGURE 14. The DNA-dependent ATPase activity. (A) The ATP concentration dependence of the rate constants in the absence (*black*) and presence of 30 μM (bp) dsDNA (*blue*) and 30 μM (nt) ssDNA (*red*). The results for the wild type, R286A/R385A, and R305A/R314A/K345A are shown in the *left*, *middle*, and *right panels*, respectively. The kinetic parameters estimated by the experiments are summarized in Table 4. (B) The DNA concentration dependence of the rate constants in ATP hydrolysis of the wild type (*black*), R286A/R385A (*cyan*) and R305A/R314A/K345A (*yellow*) in the presence of 500 μM ATP. The results for dsDNA and ssDNA are shown in the *left* and *right panels*, respectively. The DNA-binding affinities estimated by the experiments are summarized in Table 5. The error bars indicate the standard deviation of the k_{app} values from three independent experiments. WT, wild type. (C) Summary of the DNA-binding sites of rLonC and their functions revealed by this study. The schematic representation of the hexameric ring structure of rLonC is colored in *purple*. The ATPase domain in each subunit is shown as a *gray circle* with a *dotted line*. The DNA molecule is colored in *orange*.

TABLE 4 The kinetic parameters of the DNA-dependent ATPase activity[a]

	K_m[b]	k_{cat}[b]	k_{cat}/K_m
	μM	s^{-1}	$M^{-1}\ s^{-1}$
WT	33 ± 3	0.0084 ± 0.0002	260
+ dsDNA	48 ± 4	0.037 ± 0.001	780
+ ssDNA	48 ± 6	0.019 ± 0.001	390
R286A/R385A	41 ± 4	0.029 ± 0.001	720
+ dsDNA	41 ± 4	0.031 ± 0.001	750
+ ssDNA	38 ± 3	0.031 ± 0.001	810
R305A/R314A/K345A	26 ± 4	0.0077 ± 0.0003	290
+ dsDNA	47 ± 7	0.027 ± 0.001	570
+ ssDNA	29 ± 5	0.014 ± 0.001	490

[a]The results are shown in Fig. 14*A*.

[b]The errors are standard errors from the regression analysis.

TABLE 5 The DNA-binding affinities determined from the DNA-dependent ATPase activity[a]

	K_d[b, c]
	μM
WT (dsDNA)	1.3 ± 0.2
WT (ssDNA)	1.1 ± 0.2
R305A/R314A/K345A (dsDNA)	2.3 ± 0.4
R305A/R314A/K345A (ssDNA)	1.9 ± 0.4

[a]The results are shown in Fig. 14*B*.

[b]The units for the K_d values are μM (bp) for dsDNA and μM (nt) for ssDNA, respectively.

[c]The errors are standard errors from the regression analysis.

Interestingly, the ATPase activity of R286A/R385A was not stimulated even at a high concentration of DNA (Fig. 14, *A* and *B*; Table 4). Furthermore, R286A/R385A exhibited a higher turnover rate in ATP hydrolysis than the wild type in the absence of DNA, and the turnover rate of R286A/R385A was comparable to that of the wild type in the presence of DNA. These data suggest that Arg286 and Arg385 have an important role in the DNA-dependent stimulation of ATP hydrolysis and that the R286A/R385A mutant mimics the stimulated state of wild-type RadA induced by DNA binding. In contrast to R286A/R385A, R305A/R314A/K345A showed similar patterns to the wild type in DNA-dependent ATPase activity, except for a slightly lower turnover rate and binding affinity to DNA than the wild type (Fig. 14, *A* and *B*; Tables 4 and 5). Based on these results, I concluded that these two DNA-binding sites have different functions in DNA-dependent ATPase activity (Fig. 14*C*).

Another LonC-like domain conserved in bacteria

Next, I performed an *in silico* search to identify novel rLonC-type LonC family proteins other than RadA in a sequence database. In the Pfam database (Eberhardt *et al*., 2016), most of the rLonC and pLonC have a ChlI (PF13541) signature as well as a Lon_C (PF05362) signature. This ChlI signature was originally found in an N-terminal domain of a functionally unknown protein annotated as magnesium chelatase family protein (Fig. 15*A*). This protein is widely conserved in bacteria and is also annotated as a ComM protein. Interestingly, like RadA and Lon protease, ComM also has an ATPase domain featuring Walker A and B motifs at the C-terminus (Fig. 15*A*). The *comM* gene has been shown to be involved in natural transformation via HR in genetic studies of *Haemophilus influenzae* and *Rhodobacter capsulatus* (Gwinn *et al*., 1998; Brimacombe *et al*., 2015). In addition, *comM* has also been identified as a SOS response gene under LexA control in *Sinorhizobium meliloti* (Erill, *et al*., 2004). Most of the ComM proteins also have the Lon_C signature in their N-terminal ChlI domain. Thus, I

considered ComM to be a novel LonC family protein.

To evaluate the evolutionary relationships between the LonC family proteins, I performed a phylogenetic analysis using the amino acid sequences of 26 Lon protease, 21 RadA and 17 ComM proteins. The results showed that the LonC family can be divided into three subfamilies: RadA, ComM, and Lon protease (Fig. 15*B*). Furthermore, the analysis also suggests that RadA and ComM are more closely related to each other than either one is to Lon protease. Interestingly, the catalytic Ser residue for the protease activity is not conserved in the N-terminal LonC domain of ComM, similar to RadA (Fig. 15*A*). The sequence alignment analysis was unable to show the conservation of the DNA-binding residues identified in this study among RadA and ComM homologs. However, structural prediction and structural alignment analyses showed that highly conserved Arg and Lys residues are located near Arg305, Arg314, and Lys345, the DNA-binding residues of rLonC (Fig. 15, *C* and *D*). By contrast, there are no conserved basic residues around Arg286 and Arg385. Considering the putative function of ComM in natural transformation via HR (Gwinn *et al*., 1998; Brimacombe *et al*., 2015), ComM might be involved in DNA processing via its LonC domain.

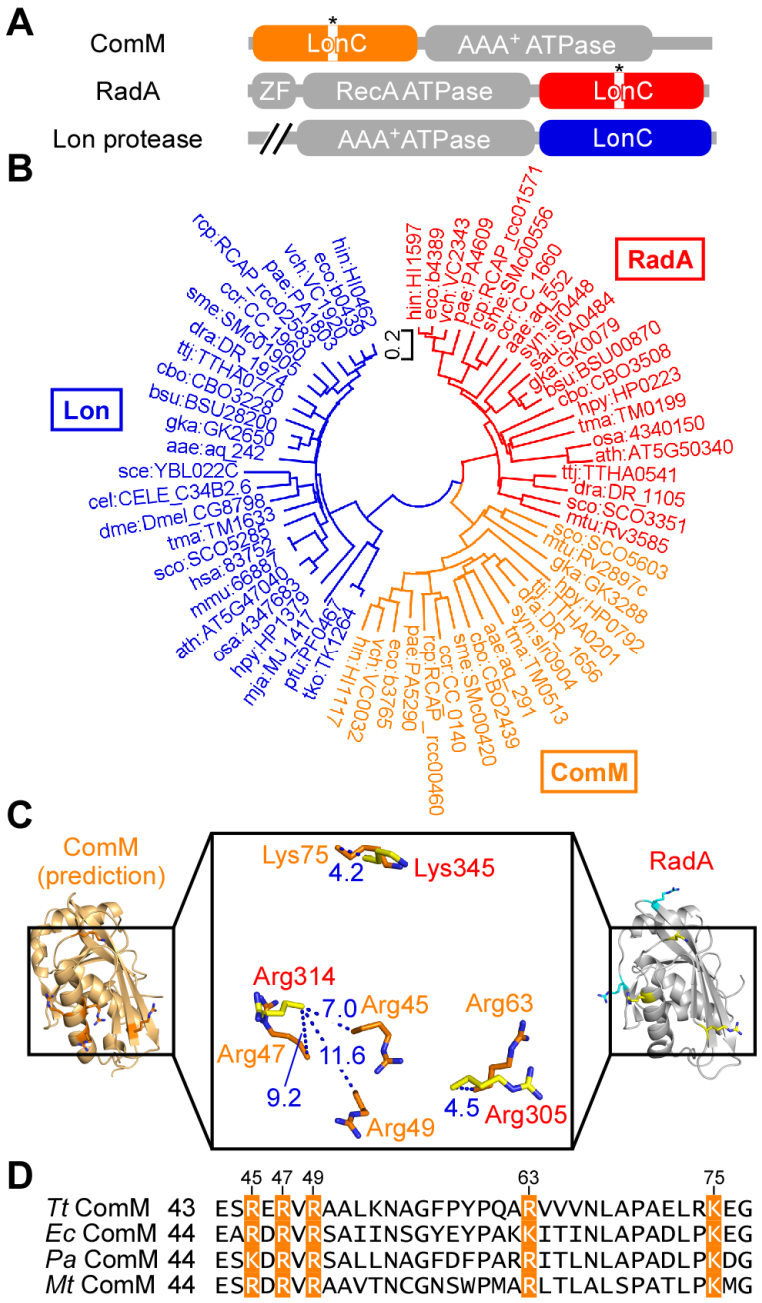

FIGURE 15. Another LonC domain-containing protein, ComM. (A) A schematic representation of the domain organization of three LonC-family proteins. Each LonC domain of ComM, RadA, and Lon protease is colored in *orange*, *red*, and *blue*, respectively. The asterisks indicate that the catalytic serine residue of Lon protease activity is not conserved in RadA or ComM (related to Fig. 2). (B) An unrooted phylogenetic tree of LonC family proteins. The sum of the branch lengths is 18.98. Proteins belonging to the ComM, RadA, and LonC subfamilies are colored in *orange*, *red*, and *blue*, respectively. (C) Structural comparison of the LonC domain of *T. thermophilus* HB8 ComM (TTHA0201; residues 1–180) with the structure of rLonC. The structural model of the LonC domain of ComM created using the I-TASSER

server (Yang *et al*., 2015) and the structure of rLonC are represented by *orange* and *gray cartoons*, respectively. The r.m.s. deviation for Cα atoms between the predicted structure of the LonC domain of ComM and the structure of rLonC was 1.86 Å. The DNA-binding residues of rLonC at the intersubunit cleft and central hole of the ring are represented in *stick form*, colored in *cyan* and *yellow*, respectively. The highly conserved basic residues of the LonC domain of ComM within 12 Å for Cα atoms from the DNA-binding residues (in *yellow sticks*) of rLonC are shown in *orange sticks*. The distances (Å) between the residues of the LonC domain of ComM and those of rLonC are shown as *blue dotted lines* (*right panel*). (D) The multiple sequence alignment of the LonC domain of ComM proteins. The highly conserved basic residues are highlighted with *orange backgrounds*. The accession numbers of the sequences are as follows: *T. thermophilus* (*Tt*) ComM (YP_143467), *E. coli* (*Ec*) ComM (NP_418214), *Pseudomonas aeruginosa* (*Pa*) ComM (NP_253977), and *Mycobacterium tuberculosis* (*Mt*) ComM (NP_217413).

DISCUSSION

In this study, I determined the X-ray crystal structure of rLonC, which represents the first report of the tertiary structure of the "non-protease-type" LonC domain (Fig. 6). I further revealed that rLonC possesses DNA-binding activity (Fig. 10). To the best of our knowledge, this is the first report on the tertiary structure and molecular function of pseudoprotease domains. The DNA-binding ability is exerted by two positively charged regions, namely the intersubunit cleft and central hole of the ring (Figs. 9 and 12). These data raise the hypothesis for the model of two DNA-binding modes: one is that the DNA molecule binds to the intersubunit cleft, and the other is that the DNA molecule passes through the central hole (Fig. 14*C*).

As for the latter mode, it should be noted that the loop containing 299-TPFPAP-304 forms a lid-like structure at the opposite side of the basic side of the ring (Figs. 7 and 9). The inner diameter between the two loops is approximately 17 Å, whereas the diameter of B-form dsDNA is 20 Å. Therefore, this loop could be steric hindrance based on the hypothesis that the DNA molecule passes through the central hole of the ring. However, the average *B*-factors of the Cα atoms of this loop and the overall structure were 92.0 and 51.2 $Å^2$, respectively, implying a high flexibility of this loop (Fig. 7*C*). These data raise the hypothesis that structural rearrangements in the loop region might occur upon binding to DNA through the central hole of the ring.

It should be also noted that RadA forms various types of oligomers that are smaller than a hexamer in solution (Fig. 8), implying the possibility that there are DNA-binding modes other than the hexameric ring structure. Both of the DNA-binding sites are located at the subunit interface and formed by dimerization (Fig. 14*C*). I speculate that dimerization, not hexamerization, is required and sufficient for DNA binding. On the other hand, there is the possibility that the hexameric structure is stabilized by DNA, nucleotide, or metal ion binding. For example, it has been shown that the hexameric structure of T7 DNA helicase is stabilized in the presence of DNA and nucleotide (Patel & Hingorani, 1993; Biswas & Biswas, 1999). It

has also been shown that *E. coli* DnaB helicase forms a hexamer in a way partially dependent on magnesium ion (Picha & Patel, 1998). Further biochemical analysis would be required to reveal the relationships between the oligomerization and DNA binding.

My biochemical analysis suggests the functional difference between the two DNA-binding sites in the DNA-dependent ATPase activity: the intersubunit cleft is essential to the DNA-dependent stimulation of the ATPase activity, while the central hole is not (Fig. 14*C*). Interestingly, the R286A/R385A mutant exhibited similar ATPase activity to the DNA-bound stimulated state of the wild type even in the absence of DNA. Thus, Arg286 and/or Arg385 would be required to regulate ATPase activity in the DNA-free state as well as to stimulate the activity upon DNA binding. I speculate that DNA binding to the intersubunit cleft of rLonC changes the ATPase domain in such a way as to stimulate its ATPase activity.

As shown in Fig. 16, I predicted that the hexameric structures of the ATPase and rLonC domains are oriented with each other. Interestingly, the "bottom" face of the hexameric model of the ATPase domain was predicted to be negatively charged (Fig. 16*B*), which would be favorable to interact with the basic face of rLonC domain. Thus, the regulation of the ATPase activity might be carried out through such electrostatic interactions in the absence of DNA. Upon DNA binding to the intersubunit cleft, the regulatory interactions between the two domains might be interrupted by a negative charge or a steric hindrance of the bound DNA molecule to stimulate the ATPase activity.

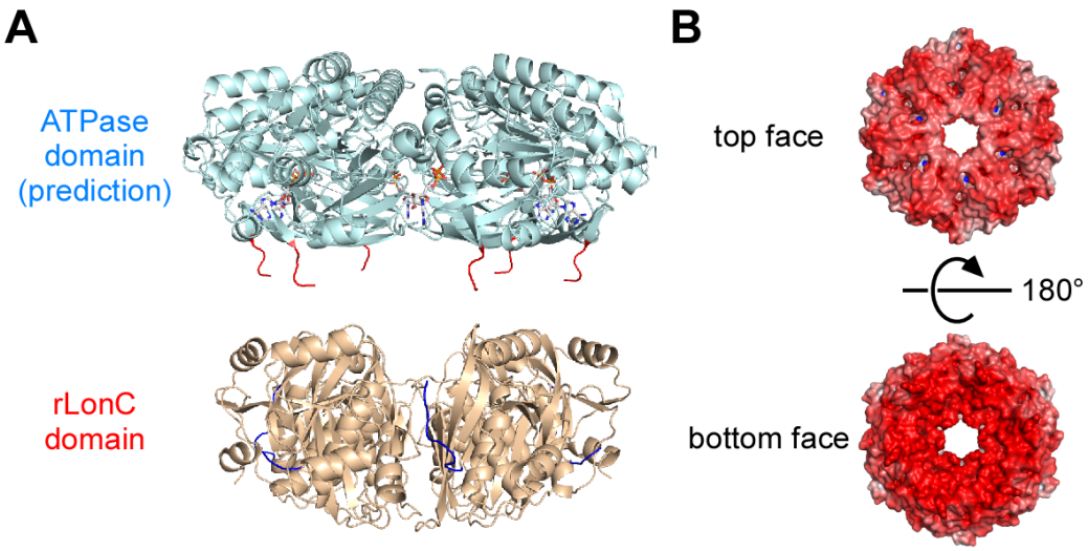

FIGURE 16. The prediction of the orientation and interface between the ATPase and rLonC domains of RadA. (A) The orientation of the hexameric structures of the ATPase and rLonC domains. The orientation was predicted based on the direction of the linker regions and orientation of Lon protease domains (PDB ID: 3K1J, related to Fig. 3). The hexamer of the ATPase domain was predicted by aligning the monomeric structure predicted using I-TASSER server (Yang *et al*., 2015) (related to Fig. 11) with the hexameric structure of an archaeal RecA-like ATPase (PDB ID: 2DR3). The C-terminal region of the ATPase domain and the N-terminal region of the rLonC domain are highlighted with *red* and *blue*. An ADP molecule bound to each ATPase domain is represented in CPK-colored *stick form*. (B) The surface electrostatic potential of the predicted hexamer of the ATPase domain. The top and bottom faces as in (A) are shown and contoured in the range from –10 kT/e (*red*) to +10 kT/e (*blue*).

It should be noted that RadA can bind to DNA in the absence of nucleotides (Fig. 10). Interestingly, it has been shown that *E. coli* RadA requires ADP for DNA-binding activity, suggesting the ATPase domain has an important role for DNA binding (Cooper & Lovett, 2016). Further biochemical analysis and structural information about full-length RadA and its DNA complex are necessary for a complete understanding of how the rLonC and ATPase domains coordinate to regulate DNA-binding activity.

The *in vivo* study suggests that the DNA-binding activity of rLonC is essential to the DNA-repair function of RadA (Fig. 13). Furthermore, it suggests that both DNA-binding sites are necessary for proper DNA repair. The HR pathway involves two DNA molecules and produces branched DNA structures as intermediates, such as the D-loop and Holliday junction (Cromie *et al*., 2001; West, 2003). It has been suggested that the two DNA-binding sites of RecA/Rad51 are necessary for efficient recombining of two DNA molecules. For instance, RecA has a unique C-terminal domain containing conserved basic residues that interact with dsDNA as well as the L1 and L2 loops in the ATPase domain (Kurumizaka *et al*., 1996). In addition, the N-terminal regions of Rad51, Dmc1, and archaeal RadA (not a bacterial RadA ortholog) contain a modified helix-hairpin-helix motif, which has a similar function to the C-terminal domain of RecA (Aihara *et al*., 1999; Kinebuchi *et al*., 2005; Chen *et al*., 2007). In *E. coli*, genetic and biochemical studies have shown that RadA might be involved in recombination intermediate processing *in vivo* and that RadA stimulates RecA-catalyzed branch migration *in vitro* (Beam *et al*., 2002; Cooper *et al*., 2015; Cooper & Lovett, 2016). Thus, the two DNA-binding sites of rLonC might be necessary for efficient recombination intermediate processing, such as branch migration.

In bacteria, RuvAB and RecG, well-known proteins with helicase functions at branch migration, processively unwind the branched DNA (Yamada *et al*., 2002; Singleton *et al*., 2001). In the RuvAB complex, RuvA binds to the branched-DNA structure and RuvB conducts ATP-

dependent translocation of DNA (Yamada *et al*., 2002). On the other hand, RecG, composed of three domains, has two DNA-binding sites located at different domains to efficiently bind the branched DNA and unwind dsDNA with an ATP-dependent motor domain (Singleton *et al*., 2001). In strand-exchange reactions by RecA *in vitro*, both RuvAB and RecG accelerate the reverse reaction back to substrate forms, resulting decreases of the recovery of full strand-exchange products (Whitby *et al*., 1993). Interestingly, in contrast to these branch-migration helicases, RadA has no helicase activity alone and can function in the context of the RecA synaptic filament, with the branched DNA migrated in the direction of RecA-promoted strand exchange (Cooper & Lovett, 2016). Thus, RadA might transiently destabilize or unwind dsDNA to support the strand-exchange and branch-migration activities of RecA (Fig. 17). In such a reaction, rLonC might bind to the D-loop structure through the two DNA binding sites and the ATPase activity might be used for transient destabilization of dsDNA to assist the RecA activity.

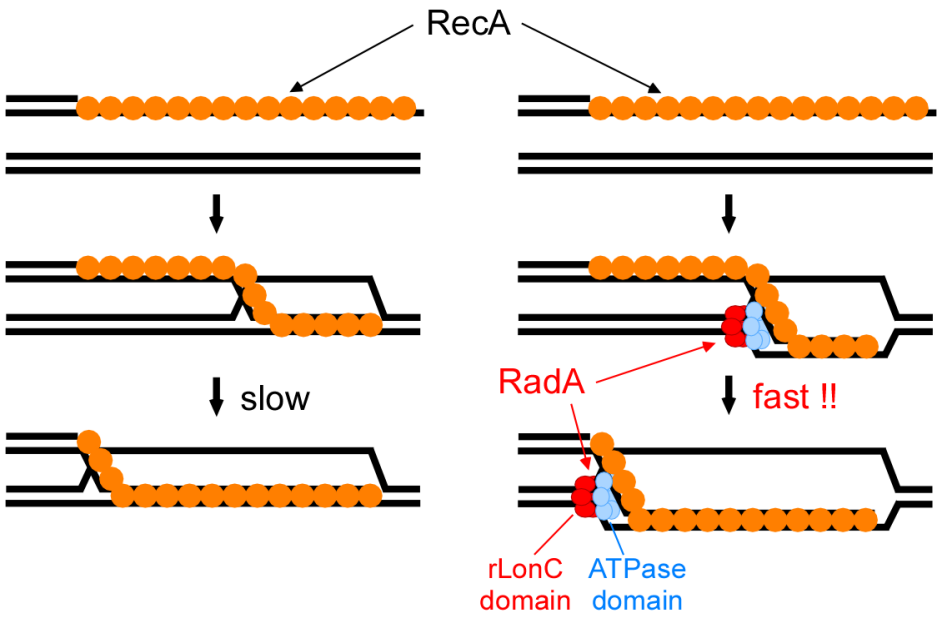

FIGURE 17. The proposed functional model of RadA in RecA-catalyzed strand exchange and branch migration. The pathways in the absence (*left panel*) and presence (*right panel*) of RadA are represented. The ZF domain is omitted from the figures.

The phylogenetic analysis suggests that the LonC family is divided into three subfamilies: RadA, ComM, and Lon protease (Fig. 15). Among this family, only Lon protease has the conserved catalytic Ser residue for the protease activity. Interestingly, the LonC family belongs to the ribosomal protein S5 domain 2-like superfamily, which includes various types of nucleic acid-binding proteins, such as ribosomal proteins S5 and S9, DNA mismatch repair protein MutL, DNA topoisomerases, ribonuclease P, and elongation factor G in the SCOP and CATH databases (Andreeva *et al*., 2007; Sillitoe *et al*., 2015). In particular, ribosomal proteins, MutL, and ribonuclease P utilize the domain belonging to this superfamily as a nucleic acid-binding module (Brodersen *et al*., 2002; Robertson *et al*., 2006; Jovanovic *et al*., 2002). Therefore, I speculate that the ancestor of the LonC family has a nucleic acid-binding role. Structural and biochemical characterization of the LonC domain of ComM will be necessary to better understand the evolutionary relationships of LonC's subfamilies.

REFERENCES

Aihara, H., Ito, Y., Kurumizaka, H., Yokoyama, S., and Shibata, T. (1999) The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR. *J. Mol. Biol.* **290**, 495−504

Altieri, F., Grillo, C., Maceroni, M., and Chichiarelli, S. (2008) DNA damage and repair: from molecular mechanisms to health implications. *Antioxid. Redox Signal.* **10**, 891−937

Amerik, A., Antonov, V. K., Gorbalenya, A. E., Kotova, S. A., Rotanova, T. V., and Shimbarevich, E. V. (1991) Site-directed mutagenesis of La protease. A catalytically active serine residue. *FEBS Lett.* **287**, 211−214

Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **36**, D419−D425

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037−10041

Beam, C. E., Saveson, C. J., and Lovett, S. T. (2002) Role for *radA/sms* in recombination intermediate processing in *Escherichia coli*. *J. Bacteriol.* **184**, 6836−6844

Bell, J. C., and Kowalczykowski, S. C. (2016) RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends Biochem. Sci.* **41**, 491−507

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252−W258

Biswas, E. E. and Biswas,S. B. (1999) Mechanism of DnaB Helicase of Escherichia coli: Structural Domains Involved in ATP Hydrolysis, DNA Binding, and Oligomerization.

Biochemistry **38**, 10919–10928

Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A., and Gustchina, A. (2004) The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J. Biol. Chem.* **279**, 8140−8148

Brigulla, M., and Wackernagel, W. (2010) Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. *Appl. Microbiol. Biotechnol.* **86**, 1027−1041

Brimacombe, C. A., Ding, H., Johnson, J. A., and Beatty, J. T. (2015) Homologues of Genetic Transformation DNA Import Genes Are Required for *Rhodobacter capsulatus* Gene Transfer Agent Recipient Capability Regulated by the Response Regulator CtrA. *J. Bacteriol.* **197**, 2653−2663

Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* **316**, 725−768

Carrasco, B., Cozar, M. C., Lurz, R., Alonso, J. C., and Ayora, S. (2004) Genetic recombination in *Bacillus subtilis* 168: contribution of Holliday junction processing functions in chromosome segregation. *J. Bacteriol.* **186**, 5557−5566

Cha, S. S., An, Y. J., Lee, C. R., Lee, H. S., Kim, Y. G., Kim, S. J., Kwon, K. K., De Donatis, G. M., Lee, J. H., Maurizi, M. R., and Kang, S. G. (2010) Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J.* **29**, 3520−3530

Chen, C. Y., Malchus, N. S., Hehn, B., Stelzer, W., Avci, D., Langosch, D., and Lemberg, M. K. (2014) Signal peptide peptidase functions in ERAD to cleave the unfolded protein

response regulator XBP1u. *EMBO J*. **33**, 2492−2506

Chen, L. T., Ko, T. P., Chang, Y. W., Lin, K. A., Wang, A. H. J., and Wang, T. F. (2007) Structural and Functional Analyses of Five Conserved Positively Charged Residues in the L1 and N-Terminal DNA Binding Motifs of Archaeal RadA Protein. *PLoS ONE* **2**, e858

Chen, Z., Yang, H., and Pavletich, N. P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489−494

Compton, S. A., Ozgür, S., Griffith, J. D. (2010) Ring-shaped Rad51 paralog protein complexes bind Holliday junctions and replication forks as visualized by electron microscopy. *J. Biol. Chem*. **285**, 13349−13356

Cooper, D. L., Boyle, D. C., and Lovett, S. T. (2015) Genetic analysis of *Escherichia coli* RadA: functional motifs and genetic interactions. *Mol. Microbiol.* **95**, 769−779

Cooper, D. L., and Lovett, S. T. (2016) Recombinational branch migration by the RadA/Sms paralog of RecA in *Escherichia coli*. *eLife* **5**, e10807

Cox, M. M. (2007) Regulation of bacterial RecA protein function. *Crit. Rev. Biochem. Mol. Biol.* **42**, 41−63

Cromie, G. A., Connelly, J. C., and Leach, D. R. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell* **8**, 1163−1174

Diver, W. P., Sargentini, N. J., and Smith, K. C. (1982) A mutation (*radA100*) in *Escherichia coli* that selectively sensitizes cells grown in rich medium to X-or UV-radiation, or methyl methanesulphonate. *Int. J. Radiat. Biol.* **42**, 339−346

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522−W525

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792−1797

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486−501

Erill, I., Jara, M., Salvador, N., Escribano, M., Campoy, S., and Barbe, J. (2004) Differences in LexA regulon structure among Proteobacteria through in vivo assisted comparative genomics. *Nucleic Acids Res.* **32**, 6617−6626

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279−D285

Fukui, K., Nishida, M., Nakagawa, N., Masui, R., and Kuramitsu, S. (2008) Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. *J. Biol. Chem.* **283**, 12136−12145

Greenblatt, E. J., Olzmann, J. A., and Kopito, R. R. (2011) Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant α-1 antitrypsin from the endoplasmic reticulum. *Nat*. *Struct*. *Mol*. *Biol*. **18**, 1147–1152

Gwinn, M. L., Ramanathan, R., Smith, H. O., and Tomb, J. F. (1998) A new transformation-deficient mutant of *Haemophilus influenzae* Rd with normal DNA uptake. *J. Bacteriol.* **180**, 746−748

Hashimoto, Y., Yano, T., Kuramitsu, S., and Kagamiyama, H. (2001) Disruption of *Thermus thermophilus* genes by homologous recombination using a thermostable kanamycin-resistant marker. *FEBS Lett.* **506**, 231−234

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65−70

Jovanovic, M., Sanchez, R., Altman, S., and Gopalan, V. (2002) Elucidation of structure-function relationships in the protein subunit of bacterial RNase P using a genetic complementation approach. *Nucleic Acids Res.* **30**, 5065−5073

Kevorkian, Y., Shirley, D. J., and Shen, A. (2016) Regulation of Clostridium difficile spore germination by the CspA pseudoprotease domain. *Biochimie* **122**, 243−254

Kinebuchi, T., Kagawa, W., Kurumizaka, H., and Yokoyama, S. (2005) Role of the N-terminal domain of the human DMC1 protein in octamer formation and DNA binding. *J. Biol. Chem.* **280**, 28382−28387

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012) Homologous recombination and its regulation. *Nucleic Acids Res.* **40**, 5795−5818

Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774−797

Krüger, E., Msadek, T., Ohlmeier, S., and Hecker, M. (1997) The *Bacillus subtilis clpC* operon encodes DNA repair and competence proteins. *Microbiology* **143**, 1309−1316

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870−1874

Kuramitsu, S., Hiromi, K., Hayashi, H., Morino, Y., and Kagamiyama, H. (1990) Pre-steady-state kinetics of *Escherichia coli* aspartate aminotransferase catalyzed reactions and thermodynamic aspects of its substrate specificity. *Biochemistry* **29**, 5469−5476

Kurumizaka, H., Aihara, H., Ikawa, S., Kashima, T., Bazemore, L. R., Kawasaki, K., Sarai, A., Radding, C. M., and Shibata, T. (1996) A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *J. Biol. Chem.* **271**, 33515−33524

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*

26, 283−291

Lee, A. Y., Chen, Y. D., Chang, Y. Y., Lin, Y. C., Chang, C. F., Huang, S. J., Wu, S. H., and Hsu, C. H. (2014) Structural basis for DNA-mediated allosteric regulation facilitated by the AAA^+ module of Lon protease. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 218−230

Lee, A. Y., Hsu, C. H., and Wu, S. H. (2004) Functional domains of *Brevibacillus thermoruber* Lon protease for oligomerization and DNA binding: role of N-terminal and sensor and substrate discrimination domains. *J. Biol. Chem.* **279**, 34903−34912

Lin, C. C., Su, S. C., Su, M. Y., Liang, P. H., Feng, C. C., Wu, S. H., and Chang, C. I. (2016) Structural Insights into the Allosteric Operation of the Lon AAA+ Protease. *Structure* **24**, 667−675

Lin, Y. C., Lee, H. C., Wang, I., Hsu, C. H., Liao, J. H., Lee, A. Y., Chen, C., and Wu, S. H. (2009) DNA-binding specificity of the Lon protease α-domain from *Brevibacillus thermoruber* WR-249. *Biochem. Biophys. Res. Commun.* **388**, 62−66

Lin, Z., Kong, H., Nei, M., and Ma, H. (2006) Origins and evolution of the *recA/RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10328−10333

Liu, J., Renault, L., Veaute, X., Fabre, F., Stahlberg, H., and Heyer, W. D. (2011) Rad51 paralogues Rad55-Rad57 balance the antirecombinase Srs2 in Rad51 filament formation. *Nature* **479**, 245−248

Liu, Y., Masson, J. Y., Shah, R., O'Regan, P., West, S. C. (2004) RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243−246

Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins* **50**, 437−450

Lovett, S. T. (2006) Replication arrest-stimulated recombination: Dependence on the RecA

paralog, RadA/Sms and translesion polymerase, DinB. *DNA Repair (Amst)* **5**, 1421−1427

Maraboeuf, F., Voloshin, O., Camerini-Otero, R. D., and Takahashi, M. (1995) The central aromatic residue in loop L2 of RecA interacts with DNA. Quenching of the fluorescence of a tryptophan reporter inserted in L2 upon binding to DNA. *J. Biol. Chem.* **270**, 30927−30932

Morita, R., Nakane, S., Shimada, A., Inoue, M., Iino, H., Wakamatsu, T., Fukui, K., Nakagawa, N., Masui, R., Kuramitsu, S. (2010) Molecular Mechanisms of the Whole DNA Repair System: A Comparison of Bacterial and Eukaryotic Systems. *J. Nucleic Acids* **2010**, 179594

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) *REFMAC*5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355−367

Murugasu-Oei, B., Rodrigues, V., Yang, X., and Chia, W. (1995) Masquerade: a novel secreted serine protease-like molecule is required for somatic muscle attachment in the Drosophila embryo. *Genes Dev*. **15**, 139−154.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307−326

Patel, S. S. and Hingorani, M. M. (1993) Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins. *J. Biol. Chem*. **268**, 10668–10675

Picha, K. M. and Patel, S. S. (1998) Bacteriophage T7 DNA helicase binds dTTP, forms hexamers, and binds DNA in the absence of Mg^{2+}. The presence of dTTP is sufficient for hexamer formation and DNA binding. *J. Biol. Chem*. **273**, 27315–27319

Pinti, M., Gibellini, L., Nasi, M., De Biasi, S., Bortolotti, C. A., Iannone, A., and Cossarizza, A. (2016) Emerging role of Lon protease as a master regulator of mitochondrial

functions. *Biochim. Biophys. Acta* **1857**, 1300−1306

Rasulova, F. S., Dergousova, N. I., Starkova, N. N., Melnikov, E. E., Rumsh, L. D., Ginodman, L. M., and Rotanova, T. V. (1998) The isolated proteolytic domain of *Escherichia coli* ATP-dependent protease Lon exhibits the peptidase activity. *FEBS Lett.* **432**, 179−181

Robertson, A., Pattishall, S. R., and Matson, S. W. (2006) The DNA binding activity of MutL is required for methyl-directed mismatch repair in *Escherichia coli*. *J. Biol. Chem.* **281**, 8399−8408

Rotanova, T. V., Melnikov, E. E., Khalatova, A. G., Makhovskaya, O. V., Botos, I., Wlodawer, A., and Gustchina, A. (2004) Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. *Eur. J. Biochem.* **271**, 4865−4871

Rudyak, S. G., Brenowitz, M., and Shrader, T. E. (2001) Mg^{2+}-linked oligomerization modulates the catalytic activity of the Lon (La) protease from *Mycobacterium smegmatis*. *Biochemistry* **40**, 9317−9323

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425

Schärer, O. D. (2003) Chemistry and biology of DNA repair. *Angew. Chem. Int. Ed. Engl.* **42**, 2946−2974

Shin, D. S., Chahwan, C., Huffman, J. L., and Tainer, J. A. (2004) Structure and function of the double-strand break repair machinery. *DNA Repair (Amst)* **3**, 863−873

Shinohara, T., Ikawa, S., Iwasaki, W., Hiraki, T., Hikima, T., Mikawa, T., Arai, N., Kamiya, N., and Shibata, T. (2015) Loop L1 governs the DNA-binding specificity and order for RecA-catalyzed reactions in homologous recombination and DNA repair. *Nucleic Acids Res.* **43**, 973−986

Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., Furnham, N., Laskowski, R. A., Lee, D., Lees, J. G., Lehtinen, S., Studer, R. A., Thornton, J., and Orengo, C. A. (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **43**, D376−D381

Singleton, M. R., Scaife, S., and Wigley, D. B. (2001) Structural Analysis of DNA Replication Fork Reversal by RecG. *Cell* **107**, 79–89

Slade, D., Lindner, A. B., Paul, G., and Radman, M. (2009) Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. *Cell* **136**, 1044−1055

Song, Y., and Sargentini, N. J. (1996) *Escherichia coli* DNA repair genes *radA* and *sms* are the same gene. *J. Bacteriol.* **178**, 5045−5048

Su, S. C., Lin, C. C., Tai, M. C., Chang, M. Y., Ho, M. R., Babu, C. S., Liao, J. H., Wu, S. H., Chang, Y. C., Lim, C., and Chang, C. I. (2016) Structural Basis for the Magnesium-Dependent Activation and Hexamerization of the Lon AAA+ Protease. *Structure* **24**, 676–686

Suwaki, N., Klare, K., and Tarsounas, M. (2011) RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin. Cell Dev. Biol.* **22**, 898−905

Taylor, M. R., Spirek, M., Chaurasiya, K. R., Ward, J. D., Carzaniga, R., Yu, X., Egelman, E. H., Collinson, L. M., Rueda, D., Krejci, L., and Boulton, S. J. (2015) Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell* **162**, 271−286

Thacker, J. (2005) The *RAD51* gene family, genetic instability and cancer. *Cancer Lett.* **219**, 125−135

Thomas, C. M., and Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711−721

Vagin, A., and Teplyakov, A. (1997) *MOLREP*: an automated program for molecular

replacement. *J. Applied. Crystallogr.* **30**, 1022−1025

Van Melderen, L., and Aertsen, A. (2009) Regulation and quality control by Lon-dependent proteolysis. *Res. Microbiol.* **160**, 645−651

Vieux, E. F., Wohlever, M. L., Chen, J. Z., Sauer, R. T., and Baker, T. A. (2013) Distinct quaternary structures of the AAA+ Lon protease control substrate degradation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2002−E2008

West, S. C. (2003) Molecular views of recombination proteins and their control. *Nat. Rev. Mol. Cell Biol.* **4**, 435−445

Whitby, M. C., Ryder, L., and Lloyd, R. G. (1993) Reverse branch migration of holliday junctions by RecG protein: A new mechanism for resolution of intermediates in recombination and DNA repair. *Cell* **75**, 341−350

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the *CCP*4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235−242

Yamada, K., Miyata, T., Tsuchiya, D., Oyama, T., Fujiwara, Y., Ohnishi, T., Iwasaki, H., Shinagawa, H., Ariyoshi, M., Mayanagi, K., and Morikawa, K. (2002) Crystal Structure of the RuvA-RuvB Complex: A Structural Basis for the Holliday Junction Migrating Motor Machinery. *Mol. Cell* **10**, 671–681

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7−8

Zhou, Q., Zhang, X., Xu, H., Xu, B., and Hua, Y. (2006) RadA: A protein involved in DNA damage repair processes of *Deinococcus radiodurans* R1. *Chin. Sci. Bull.* **51**, 2993−2999

ACKNOWLEDGEMENTS

I would like to express my great appreciation to Drs. Seiki Kuramitsu, Ryoji Masui, and Noriko Nakagawa for their guidance and many valuable discussions. I also would like to express my great appreciation to Drs. Genji Kurisu, Hisao Masukata, Kenji Matsuno, and Tetsuro Yonesaki for their critical advices in this study. I also thank Drs. Takato Yano and Kenji Fukui for their help in the X-ray crystallographic study. I thank my colleagues in Kuramitsu and Masui laboratories, my friends for their kind help and encouragement.

PUBLICATION LIST

Inoue, M., Fukui, K., Fujii, Y., Nakagawa, N., Yano, T., Kuramitsu, S., and Masui, R. (2017) The Lon protease-like domain in the bacterial RecA paralog RadA is required for DNA binding and repair. *J. Biol. Chem. in press*.

Masui, R., Takahata, Y., Inoue, M., Iio, Y., Okanishi, H., Kim, K., Nakagawa, N., Yura, K., and Kuramitsu, S. (2014) Structural insights of post-translational modification sites in the proteome of *Thermus thermophiles*. *J. Struct. Funct. Genomics* **15**, 137–151

Takahata, Y., Inoue, M., Kim, K., Iio, Y., Miyamoto, M., Masui, R., Ishihama, Y., and Kuramitsu, S. (2012) Close proximity of phosphorylation sites to ligand in the phosphoproteome of the extreme thermophile *Thermus thermophilus* HB8. *Proteomics* **12**, 1414–1430

Morita, R., Nakane, S., Shimada, A., Inoue, M., Iino, H, Wakamatsu, T., Fukui, K., Nakagawa, N., Masui, R., and Kuramitsu, S. (2010) Molecular Mechanisms of the Whole DNA Repair System: A Comparison of Bacterial and Eukaryotic Systems. *J. Nucleic Acids* **2010**, 179594