



HAL
open science

An alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease

Likui Zhang, Haoqiang Shi, Qi Gan, Yuxiao Wang, Mai Wu, Zhihui Yang,
Phil M. Oger, Jianting Zheng

► **To cite this version:**

Likui Zhang, Haoqiang Shi, Qi Gan, Yuxiao Wang, Mai Wu, et al.. An alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease. *DNA Repair*, 2020, pp.102734. 10.1016/j.dnarep.2019.102734 . hal-02334467

HAL Id: hal-02334467

<https://hal.science/hal-02334467v1>

Submitted on 17 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**An alternative pathway for repair of deaminated bases in DNA triggered by
archaeal NucS endonuclease**

Running title: Deaminated DNA repair by archaeal NucS endonuclease

Likui Zhang^{1,5#}, Haoqiang Shi¹, Qi Gan¹, Yuxiao Wang¹, Mai Wu¹, Zhihui Yang^{2#},
Philippe Oger^{3#} and Jianting Zheng⁴

¹Marine Science & Technology Institute, Yangzhou University, China

Department of Environmental Science and Engineering, Yangzhou University, China

²College of Plant Protection, Agricultural University of Hebei, Baoding City, Hebei
Province 071001, China

³Univ Lyon, INSA de Lyon, CNRS UMR 5240, Villeurbanne, France

⁴State Key Laboratory of Microbial Metabolism, School of Life Sciences
and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

⁵Guangling College, Yangzhou University

#Corresponding author: Dr. Likui Zhang

Tel: +86-514-89795882

Fax: +86-514-87357891

E-mail address: lkzhang@yzu.edu.cn

Corresponding author: Prof. Zhihui Yang

E-mail address: bdyzh@hebau.edu.cn

Corresponding author: Prof. Philippe Oger

E-mail address: philippe.oger@insa-lyon.fr

ABSTRACT

Recent studies show that NucS endonucleases participate in mismatch repair in several archaea and bacteria. However, the function of archaeal NucS endonucleases has not been completely clarified. Here, we describe a NucS endonuclease from the hyperthermophilic and radioresistant archaeon *Thermococcus gammatolerans* (Tga NucS) that can cleave uracil (U)- and hypoxanthine (I)-containing dsDNA at 80°C. Biochemical evidence shows that the cleavage sites of the enzyme are at the second phosphodiester on the 5'- site of U or I, and at the third phosphodiester on the 5'-site of the opposite base of U or I, creating a double strand break with a 4-nt 5'-overhang. The ends of the cleaved product of Tga NucS are ligatable, possessing 5'-phosphate and 3'-hydroxyl termini, which can be utilized by DNA repair proteins or enzymes. Tga NucS displays a preference for U/G- and I/T-containing dsDNA over other pairs with U or I, suggesting that the enzyme is responsible for repair of U and I in DNA that arise from deamination. Biochemical characterization of cleaving U- and I-containing DNA by Tga NucS was also investigated. The DNA-binding results show that the enzyme exhibits a higher affinity for normal, U- and I-containing dsDNA than for normal, U- and I-containing ssDNA. Therefore, we present an alternative pathway for repair of deaminated bases in DNA triggered by archaeal NucS endonuclease in hyperthermophilic archaea.

Keywords: Archaea; NucS endonuclease; Deamination; DNA repair

1. Introduction

Although double-stranded DNA (dsDNA) is stable, DNA is always damaged by endogenous and exogenous DNA damaging agents. Replication of damaged DNA can lead to serious consequences for a cell if it is not repaired, such as mutation and eventually cell death. Thus, efficient DNA repair is essential for all forms of life. Deamination is a typical damage in genomic DNA of a living cell. Cytosine and adenine are commonly deaminated into the corresponding base analogs: uracil (U) and hypoxanthine (I), respectively. On one hand, a cytosine to thymine transition can occur because of a pre-mutagenic pairing of U with G [1]. On the other hand, an A-T base pair can be subsequently converted to a G-C base pair after DNA replication if the I in DNA is not repaired [2]. Enhanced deamination rates occur more frequently at high temperature under which extremophiles thrive [3]. However, genetic data suggest that extremophilic microbes have spontaneous mutation rates similar to those of *E.coli* [4,5]. Therefore, it is expected that these microbes would possess more efficient repair pathways to counteract higher rates of mutations generated by enhanced deamination.

Archaea have been shown to encode eukaryotic-like DNA repair proteins. Hence, their analysis has allowed to reveal properties of eukaryotic DNA repair systems [6-9]. Hyperthermophilic archaea (HA) have been shown to possess DNA repair proteins from nucleotide excision repair (NER) [10], base excision repair (BER) [11], homologous recombination (HR) [12] and translesion synthesis [13] as observed in bacteria and eukaryotes. It should be noted that in archaea several proteins are missing in each repair pathway according to their genome content. For example, a

well-known mismatch repair (MMR) machinery is usually not present in archaea [14], suggesting that alternative DNA repair systems for repairing mismatched DNA might exist in HA. Therefore, a complete understanding of what DNA repair pathways are present in HA is definitely crucial.

Base excision repair is a key pathway present in bacteria, archaea and eukarya for restoring U and I in DNA [11]. HA might employ similar BER mechanism to remove U and I in DNA. However, if the numbers of U and I in genomic DNA exceed BER capability of HA, other repair pathways would help to repair enhanced DNA lesions created at high temperature.

Recently, an alternative excision repair (AER) has been reported as a novel repair pathway to restore two deaminated bases in DNA: U and I [15]. Generally, an AER is triggered by a specific endonuclease which introduces a nick in the strand containing a lesion. The recently reported *P. furiosus* EndoQ is efficient for U- and I-containing dsDNA [16]. A homologous protein with *P. furiosus* EndoQ, *Thermococcus kodakarensis* EndoQ cleaves the first and second phosphodiester bond at the 5'-site of I and U bases in DNA, respectively [16]. Another endonuclease V (EndoV), which is ubiquitous in bacteria, archaea and eukarya, is responsible for the cleavage of the second phosphodiester bond located at the 3'-side of the I base [17]. Thus, it is reasonable to assume that both EndoQ and EndoV in HA are involved in repairing I lesions created after adenine deamination [18].

As a novel archaeal endonuclease, the *Pyrococcus abyssi* NucS (nuclease for sssDNA) was originally described as a protein able to cleave flap and splayed DNA

substrates, suggesting that it could play a potential role in NER [10, 19-20]. Recently, Ishino and colleagues showed that the homolog of this enzyme in *T. kodakarensis* is able to recognize mismatched DNA, thus referring the enzyme as Tko EndoMS [21]. Furthermore, mutational analysis of bacterial NucS homologs shows an increase of mutation rates, a large majority of them being transitions corroborating the role of these proteins *in vivo* suggested by *in vitro* studies [22-24]. In addition, a recent study revealed that the NucS homolog of *Sulfolobus acidocaldarius* is responsible for DNA repair of helix-distorting DNA lesions, suggesting that this endonuclease is involved in NER [25]. Therefore, these data suggest that NucS endonucleases may be associated to different mechanisms or DNA metabolism pathways.

Thermococcus gammatolerans, a hyperthermophilic archaeon with an optimal growth temperature of 88°C, was isolated from a hydrothermal vent located in the Gulf of California [26]. This archaeon is thought as the most radioresistant archaeon to date, fully withstanding a 5.0 kGy dose of gamma irradiation without loss of viability [27]. The genome of *T. gammatolerans* was sequenced [27], encoding a putative NucS endonuclease. In this work, we characterized the NucS endonuclease from *T. gammatolerans*, capable of cleaving U- and I-containing dsDNA at high temperature. Thus, efficient cleavage of U- and I-containing DNA by Tga NucS provides an alternative pathway for repair of deaminated base in DNA in *Thermococcus* cells.

2. Materials and methods

2.1. Materials

Materials were purchased from the following companies: dNTPs, T4 DNA ligase, *NdeI*, *NotI*, and Pfu DNA polymerase, Thermo Fisher Scientific (Waltham, MA); Quikchange XL Site-directed Mutagenesis Kit, *Escherichia coli* BL21 (DE3) pLysS cells, Transgene (Beijing, China); pET-30a (+), Novagen (Merck, Darmstadt, Germany); Chemicals, Amresco (WA, USA); PCR Cycle Pure Kit, Omega (Guangzhou, China).

2.2. DNA substrates

All the oligonucleotides used in this work were synthesized by Sangon Company (Shanghai, China). The sequences of all the oligonucleotides are listed in Table S1. The Cy3-labeled dsDNA and Hex-labeled dsDNA were prepared by mixing the oligonucleotides and their complementary oligonucleotides in an annealing buffer containing 20 mM Tris-Cl (pH 8.0) and 100 mM NaCl. The mixture was heated at 100°C for 5 min and cooled slowly to room temperature at least 4 hours. All DNA substrates in this work are listed in Table S2.

2.3. Cloning of the gene encoding *Tga NucS*

The TGAM_RS00670 gene encoding the NucS endonuclease of *T. gammatolerans* was amplified by Pfu DNA polymerase with the forward primer (5'-GGG AAT TCC ATA TGC CCA AGG TTG AGC TTA GGG AG-3', the underlined nucleotides represent *NdeI* restriction site) and the reverse primer (5'-ATA AGA ATG CGG CCG CAA ACA GCC TGA GCT GTC TTC C-3', the underlined nucleotides represent *NotI* restriction site). The PCR product was digested by *NdeI* and *NotI*, and cloned into a pET-30a (+) vector at the same restriction sites. The sequence of the

recombinant plasmid was verified before it was transformed into *E. coli* BL21 (DE3) pLysS cells for expressing Tga NucS protein with a 6 x His-tag in its C-terminal.

2.4. Overexpression and purification of Tga NucS

The expression strain *E. coli* BL21 (DE3) pLysS harboring the recombinant plasmid was cultured overnight into LB medium containing 10 µg/mL kanamycin and 17 µg/mL chloramphenicol. The fresh culture was inoculated (1:100) into 1 L of fresh medium with the same concentrations of kanamycin and chloramphenicol, and cultured at 37°C until the OD₆₀₀ reached 0.6. Then, isopropyl thiogalactoside (IPTG) was added at a final concentration of 0.8 mM for 12 hours to induce the expression from the recombinant plasmid.

The cells were harvested by centrifugation (5000 g) at room temperature, resuspended in a Ni column buffer A containing 20 mM Tris-HCl (pH 8.0), 1 mM dithiothreitol (DTT), 500 mM NaCl, 50 mM imidazole and 10% glycerol, and disrupted by ultrasonication. The soluble cell extract was heated at 70°C for 20 min to remove non-thermostable proteins from *E.coli*. After centrifugation (16000 g) at 4°C, the supernatant was loaded into a HisTrap FF column (GE Healthcare, Uppsala, Sweden) and eluted with NCG™ Chromatography System (Bio-Rad, Hercules, CA, USA) by a linear gradient of 50–500 mM imidazole with a Ni column buffer B containing 20 mM Tris-HCl (pH 8.0), 1 mM DTT, 500 mM NaCl, 500 mM imidazole and 10% glycerol. Fractions harboring Tga NucS protein were collected and analyzed by migration on a 10% SDS-PAGE. The purified Tga NucS protein fractions were dialyzed against a storage buffer containing 50 mM Tris-HCl (pH 8.0), 1 mM DTT,

50% glycerol and 50 mM NaCl, and stored at -80°C. The Tga NucS protein concentration was quantitated by measuring the absorbance at 280 nm. The theoretical molar extinction coefficient of the enzyme protein is predicted to be 12950 M⁻¹ cm⁻¹.

2.5. Construction, overexpression and purification of Tga NucS mutant D163A

Using the wild type plasmid harboring the Tga NucS gene as a template, the site-directed mutagenesis was performed using a Quikchange XL Site-directed Mutagenesis Kit to construct the D163A mutant, following manufacturer instructions. The residue D163 is located in the conserved motif II of Tga NucS (Fig. 1A). The sequences of the mutagenic primers were as following: 5'-AAG GCA CGG CAT AGT TGC_CGT TTT GGG GGT TGA CA-3', and 5'-TGT CAA CCC CCA AAA CGG_CAA CTA TGC CGT GCC TT-3'. The resulting plasmids were sequenced to verify the presence of the single mutation. The Tga NucS mutant protein was overexpressed, purified and quantified, following the same protocols as described above.

2.6. DNA cleavage assays

The DNA cleavage activity of Tga NucS was assayed with reactions containing 200 nM Cy3- or Hex-labeled dsDNA or ssDNA, 20 mM Tris-HCl (pH 8.0), 1 mM DTT, 1 mM MgCl₂, 8% glycerol and Tga NucS at designated concentration. The reactions were performed at designated temperature for 30 min and quenched by the addition of 10 mM EDTA. The reaction products were analyzed after migration into a 15% native polyacrylamide gelelectrophoresis in 0.5 x TBE (Tris-borate-EDTA). Alternatively, the reactions were stopped by the addition of 10 µL of stop solution

containing 98% formamide and 20 mM EDTA. The reaction products were heated at 100°C for 3 min and chilled rapidly on ice for 5 min, and then separated by electrophoresis into a 15% denaturing polyacrylamide gel containing 8M urea in 0.5 x TBE. After electrophoresis, the gels were visualized with a Molecular Image analyzer, PharosFx System (Bio-Rad). All experiments of DNA cleavage assays were replicated three times.

2.7. DNA ligation assays

The cleaved products of Tga NucS were purified by a PCR Cycle Pure Kit to remove the enzyme protein and used as the substrates for DNA ligation reactions performed by T4 DNA ligase at 22°C for 2 hours. The ligation products were heated at 100°C for 3 min and chilled rapidly on ice for 5 min, and analyzed by denaturing electrophoresis and imaged with a Molecular Image analyzer (Bio-Rad). ImageQuant software was used for quantitative analysis. All experiments of DNA ligation assays were replicated three times.

2.8. Electrophoretic mobility shift assays

200 nM Cy3-labeled dsDNA or ssDNA were incubated Tga NucS with increasing concentrations in 20 mM Tris-HCl (pH 8.0), 1 mM DTT and 8% glycerol at room temperature for 10 min. Samples were analyzed by electrophoresis in a 4% native polyacrylamide gel in 0.1 x TBE buffer and visualized with a Molecular Image analyzer (Bio-Rad). ImageQuant software was used for quantitative analysis. All experiments of DNA binding assays were replicated three times.

3. Results

3.1. The genome of *T. gammatolerans* encodes a putative NucS endonuclease

A NucS endonuclease is encoded by the genome of *T. gammatolerans*. Blast analysis showed that Tga NucS displays 87%, 85%, 86%, 25% and 25% similarity to the homologs from *T. kodakarensis*, *P. furiosus*, *Pyrococcus abyssi*, *Mycolicibacterium smegmatis* and *Mycobacterium tuberculosis*, respectively. As shown in Fig. 1A, the sequence alignment of NucS endonucleases from several archaea and bacteria showed that Tga NucS contains the conserved motifs (Motif I, II, III and IV), and motif V that plays an important role for interaction with the replication clamp PCNA (proliferating cell nuclear antigen), a DNA polymerase processivity in archaea and eukarya [28]. Mutational analysis of Tko EndoMS suggests that Motif II and Motif III are likely part of a nuclease motif [21].

To characterize Tga NucS, we cloned its gene into a pET30a (+) expression vector, and overexpressed and purified the protein from *E. coli*. The recombinant Tga NucS protein was purified to near homogeneity after Ni-column affinity purification (Fig. 1B). The purified recombinant protein was found to have an approximate MW of 27 kDa (Fig. 1B), correlating with the deduced amino acid sequences. The gel filtration profile of the purified Tga NucS protein showed the predicted molecular mass of 54 kDa (Fig.S1), suggesting the protein is a homodimer in solution as observed for Tko EndoMS [21].

3.2. Cleavage U- and I-containing DNA by Tga NucS

To determine whether Tga NucS is able to cleave U- and I-containing DNA, the DNA substrates with U and I were incubated with the enzyme at various temperatures

(37, 60 and 80°C). Note that U-containing dsDNA contains U:G mispair since U:G mispair can be formed by cytosine deamination from C:G pair (Fig. 2A). As shown in Fig. 2B, the cleavage percentages of Tga NucS (300 nM) were measured to be $36 \pm 3.5\%$, $36 \pm 4.2\%$ and $91 \pm 0.7\%$ at 37, 60 and 80°C, respectively. Thus, the results suggest that the enzyme is capable of efficiently cleaving U-containing dsDNA at 80°C. By contrast, the enzyme displayed very low activity on normal dsDNA (C:G) since little smear DNA product was observed at this three temperatures (Fig. 2B).

Although the endonucleases in *E. coli* would be inactivated by the heat treatment at 70°C for 20 min, there was still a possibility of endonuclease contamination during the purification of Tga NucS, thus potentially affecting our results. To rule out this possibility, we constructed the D163A mutant of Tga NucS. Note that residue D163 of Tga NucS is analogous to residue D165 of Tko EndoMS, which is located in Motif II. The purification profile of Tga NucS D163A mutant are shown in Fig. S2. Compared to the wild type enzyme, the D163A mutant completely lost the activity to cleave U- and I-containing dsDNA at 100 nM and 200 nM concentrations (Fig. 3D). Therefore, the possibility of endonuclease contamination was ruled out, suggesting that Tga NucS is capable of cleaving U- and I-containing dsDNA. Furthermore, the D163A mutant retained $36 \pm 3.5\%$ and $12 \pm 2.8\%$ cleavage activity for cleaving U/G- and I/T-containing dsDNA, respectively, suggesting the residue D163 is one of key residues in active sites of Tga NucS.

Considering the instability of the tested dsDNA substrate (45 bp) at 80°C, we proposed that Tga NucS keeps it stable at this high temperature during cleavage. To

prove our hypothesis, we performed the cleavage reactions of the enzyme using U-containing and normal ssDNA as the substrates at 80°C in the presence of 50 nM, 100 nM and 300 nM Tga NucS. Surprisingly, we found that the enzyme is almost inactive to both U-containing and normal ssDNA at 80°C (Fig. S3), which confirms our above hypothesis. Thus, Tga NucS is capable of cleaving U-containing dsDNA rather than U-containing ssDNA at the close physiological temperature.

High temperature can lead to increased not only cytosine deamination, but also adenine deamination, which leads to the formation of I base in DNA. Next, we investigated whether Tga NucS can cleave I-containing dsDNA. We employed I-containing dsDNA (I:T) as the substrate since I:T mispair can be formed by adenine deamination of A:T pair (Fig. 2A). Under the reaction conditions as described for cleaving U-containing dsDNA, $28 \pm 3.5\%$, $25 \pm 2.1\%$ and $89 \pm 1.4\%$ cleavage percentages were observed for the enzyme to cleave I-containing dsDNA at 37, 55 and 80°C (Fig. 2C), respectively. However, the enzyme displayed no activity for normal dsDNA (A:T) (Fig. 2C). Therefore, these observations suggest that Tga NucS can efficiently cleave I-containing dsDNA at the close physiological temperature. In combination, efficient cleavage of U- and I-containing dsDNA by Tga NucS at the close physiological temperature might provide an alternative pathway for repair U and I bases of DNA in *Thermococcus* cells.

3.3. Substrate specificity of Tga NucS

To determine the substrate specificity of Tga NucS, we employed various pairs of U-containing dsDNA (U:N, N:A, T, C and G) and I-containing dsDNA (I:N) as

the substrates using varied enzyme concentration (100, 200 and 400 nM) (Fig. 3A). As shown in Fig. 3B, the cleavage percentage of Tga NucS was measured to be 25 ± 1.4 , 72 ± 1.4 , and $96 \pm 0.7\%$ at 100, 200 and 400 nM for U/G-containing dsDNA (Fig. 3B), respectively. Compared with cleaving U/G-containing dsDNA, the enzyme displayed reduced efficiencies for cleaving U/T-containing dsDNA at lower enzyme concentrations (100 and 200 nM) (Fig. 3B). However, similar cleavage efficiencies were obtained at 400 nM for cleaving U/G- and U/T-containing dsDNA (Fig. 3B). By contrast, the enzyme is inactive to both U/C- and U/A-containing dsDNA (Fig. 3B). Therefore, efficient cleavage of U/G-containing dsDNA by Tga NucS suggest that *in vivo* the enzyme preferably removes uracils from U/G mispair in DNA that arise from cytosine deamination.

On the other hand, Tga NucS displayed similar efficiencies for cleaving I/T- and I/G-containing dsDNA at these three enzyme concentrations, but had no activity for cleaving I/A- and I/C-containing dsDNA (Fig. 3C). Taken together, Tga NucS displays high efficiencies for cleaving U/G- and I/T-containing dsDNA substrates.

3.4. Biochemical characterization of cleaving U- and I-containing dsDNA by Tga NucS

The biochemical characteristics of cleaving U-containing dsDNA by Tga NucS were investigated as a function of temperature, pH, divalent cations and salt concentrations by using U/G-containing dsDNA as the substrates (Fig. 4A). As expected, the enzyme is capable of cleaving U-containing dsDNA at high temperature ranging from 55°C to 85°C with varied efficiencies (Fig. 4B). Maximal cleavage percentage ($96 \pm 2.1\%$) of the enzyme was observed at 75°C, suggesting that the

enzyme activity has an optimum temperature, ca. 75°C. The thermo-tolerance results showed that the enzyme is able to fully withstand 90°C for 30 min, and even retains a $24 \pm 3.5\%$ activity after a pre-incubation at 100°C for 30 min (Fig. 4C), suggesting that Tga NucS is an extremely thermostable endonuclease.

As shown in Fig. 4D, the enzyme cleaved U-containing dsDNA in a broad pH range from 6.0 to 9.0 with maximal cleavage efficiency (91~98%), similar to that of Tko EndoMS [21]. The DNA cleavage activity of Tga NucS is dependent a divalent cation, such as Mg^{2+} , Mn^{2+} , Fe^{2+} , Co^{2+} and Ni^{2+} , among which Mg^{2+} and Mn^{2+} are the preferred ion (~98% cleavage percentage) (Fig. 4E). Furthermore, the activity of Tga NucS is independent on NaCl, but is partially inhibited at NaCl concentrations ranging from 200 to 500 mM, and totally inhibited at 1M NaCl (Fig.4F), which is consistent with the observations of high salt concentration in the *Thermococcales* cells [21,29-30]. Tga NucS displays essentially similar characteristics for I-containing dsDNA (Fig. S4), with a few differences noticeably in terms of thermo-tolerance.

3.5. Identification of the cleavage site of Tga NucS

To clarify the position of cleavage site of Tga NucS on both strands of the U- and I-containing DNA, we synthesized the Cy3-labeled oligonucleotides (22 mer and 24 mer) and the Hex-labeled (18 mer) as the DNA markers, and prepared two types of U- and I-containing dsDNA substrates: one labeled strand and double labeled strands (Figs. S5A and 5A). The cleaved products of Tga NucS were analyzed after electrophoresis on a denaturing polyacrylamide sequencing gel. As shown in Fig. S5B, the cleavage of the labeled strand of U-and I-containing DNA by Tga NucS

yielded 23 mer fragments, indicating that the enzyme cleaves the second phosphodiester bound on the 5' side of the damaged base (U or I). By contrast, Tko EndoMS cleaves mismatched DNA to form 22 mer DNA fragments [21]. Similar to Tga NucS, Tko EndoMS can cleave I-containing dsDNA at 55°C, however, its cleavage position remains unclear [21].

Tko EndoMS is able to cleave mismatched DNA at both strands (25). Here, we determined whether Tga NucS can cleave U- and I-containing dsDNA at both strands. We employed U- and I-containing dsDNA with double labeled strands as the substrates to perform the cleavage reactions of the enzyme (Fig. 5A). By running a 15% denaturing polyacrylamide gel, we found that the two short DNA fragments were created by the enzyme when using double labeled strands containing U and I as the substrates (Fig. 5A), corresponding to 23 mer and 18 mer (Fig. 5B), suggesting that the enzyme is similar to Tko EndoMS, able of cleaving both strands of dsDNA [21]. The formation with 23 mer product of Tga NucS suggest that the enzyme cleaves the second phosphodiester bound on the 5'-side of the U or I base, as observed above. The formation with 18 mer product of Tga NucS showed that the enzyme cleaves the third phosphodiester on the 5'-side of the opposite base of the U and I base in the complementary strand. Taken together, Tga NucS cleaves damaged dsDNA on both strands to generate a double strand break with a 4-nt 5'-overhang, which is distinct from Tko EndoMS that cleaves DNA to form a 5-nt 5'-overhang [21].

To investigate whether reaction temperature changes the position of the cleavage site of the enzyme, we performed the DNA cleavage reactions at 55°C using the

enzyme with high concentration (800 nM) because the enzyme has the reduced activity at this temperature (Fig. 2A), found that there was no change of cleavage site of the enzyme (Fig. S6). Therefore, reaction temperature has no effect in the position of the cleavage sites of Tga NucS.

3.6. Re-ligation of the cleaved products of Tga NucS

To investigate whether Tga NucS cleaves dsDNA to produce a 5'-phosphate and 3'-hydroxyl termini, we performed re-ligation of the cleaved products of this endonuclease by T4 DNA ligase using U/G and I/T-containing dsDNA as the cleavage substrates (Fig. 5C). As shown in Fig. 5D, $88 \pm 8.4\%$ of the U-containing dsDNA substrate was cleaved by Tga NucS before ligation. After ligation by T4 DNA ligase, the cleavage efficiency was reduced to be $28 \pm 3.5\%$, suggesting that the cleaved product of the enzyme can be ligated. Similar, the cleaved product of I-containing dsDNA catalyzed by the enzyme was also ligated due to the decreased cleavage efficiency from 94% to 43% (Fig. 5D). Therefore, the enzyme cleaves U- and I-containing dsDNA on both strands to produce ligatable cohesive ends.

3.7. Binding of U- and I-containing DNA by Tga NucS

Next, we incubated normal, U- and I-containing dsDNA with Tga NucS to investigate whether the enzyme binds to DNA as normal, U- and I-containing ssDNA and dsDNA as the substrates (Fig. 6A). As shown in Figs. 6B and 6C, Tga NucS can bind to normal ssDNA and dsDNA with varied efficiencies, although the enzyme cannot cleave normal ssDNA and dsDNA at 80°C. The enzyme had higher efficiencies for binding normal dsDNA than for binding normal ssDNA.

Interestingly, we found that Tga NucS had $41 \pm 3.5\%$ and $87 \pm 7.8\%$ binding efficiencies for U-containing dsDNA at 1000 and 2000 nM (Fig. 6B), respectively. Similar binding efficiencies of the enzyme were observed for normal dsDNA. Furthermore, the U-containing ssDNA was also bound by the enzyme with clearly reduced binding efficiencies, suggesting that Tga NucS exhibits the preference for binding normal and U-containing dsDNA over normal and U-containing ssDNA. Likewise, Tga NucS displays higher binding efficiencies for normal and I-containing dsDNA than normal and I-containing ssDNA (Fig. 6C). Overall, these observations suggest that Tga NucS is able to efficiently bind to U- and I-containing dsDNA.

4. Discussion

In this work, we demonstrate a novel function of the NucS endonuclease from *T. gammatolerans* which can recognize and cleave DNA with deaminated bases (uracil and hypoxanthine) at 80°C (Fig. 2), which is close to the physiological temperature of its host. By contrast, the enzyme is almost inactive to U- and I-containing ssDNA at 80°C. Interestingly, the two bands were observed in Figs. 2-4, suggesting that this endonuclease can cleave the substrate at two sites under the optimal reaction conditions, such as at high concentration (400 nM), high temperature (80°C), pH 8.0 and 9.0, Mg^{2+} and Mn^{2+} . Since *T. gammatolerans* possesses double stranded genomic DNA, efficient cleavage of U- and I-containing dsDNA by Tga NucS provides an alternative pathway for repair of uacils and hypoxanines in DNA.

Despite 86% similarity, Tga NucS displays varied functions with Tko EndoMS. Firstly, Tga NucS cleaves U- and I-containing dsDNA on both strands to produce a

double strand break with a 4-nt 5'-overhang (Fig. 5), and not a 5-nt 5'-overhang as observed for Tko EndoMS on mismatched dsDNA substrate [21]. Furthermore, Tga NucS displays a better activity for cleaving U- and I-containing dsDNA at 80°C than at 55°C (Fig. 2). By contrast, Tko EndoMS exhibits higher efficiencies for cleaving mismatch DNA substrates at 55°C than at 80°C. Taken together, Tga NucS exhibits several biochemical properties distinct from Tko EndoMS that led us to propose that Tga NucS might be involved in deaminated base repair at the physiological temperature of its host.

T. gammatolerans cells are thought to contain enhanced levels of uracil in genomic DNA that may originate from increased cytosine deamination [26]. Genomic sequence of *T. gammatolerans* encodes 2 putative UDGs [27]. Although these UDGs have not been characterized, they might be capable of repairing uracil in DNA in *T. gammatolerans* cells. However, if the uracil levels in the genomic DNA of *T. gammatolerans* exceed repair capability of these UDGs, an alternative repair pathway would be required. *P. furiosus* EndoQ can nick U-containing dsDNA, triggering an alternative pathway for repair of uracil in DNA. Herein, Tga NucS can recognize and cleave U-containing dsDNA at the close physiological temperature of its host, providing a novel alternative excision repair for removal of enhanced uracils in HA.

Currently, hypoxanthine of DNA is recognized by EndoV, a ubiquitous enzyme in bacteria, archaea and eukarya. After nicking by EndoV, another endonuclease or 3'-5' exonuclease would be needed to process DNA at the 5' upstream of the hypoxanthine in the DNA or degrade the DNA strand to remove the hypoxanthine. For

instance, *P. furiosus* EndoQ can nick I-containing DNA at the 3' upstream of the I nucleotide, suggesting that the endonuclease is involved in I-containing DNA repair [16,29]. However, the detailed pathways for repairing hypoxanthine in DNA of archaea remains poorly understood. Similar to other HA, *T. gammatolerans* encodes a putative EndoV [27], which might be capable of repairing hypoxanthine, as well as a putative EndoQ (TGAM_0434) homologous to Pfu EndoQ [27], suggesting that these enzymes may be involved in hypoxanthine repair in *T. gammatolerans*. In this study, we revealed that Tga NucS can also participate in I-containing DNA repair by cleaving both strands with an hypoxanthine, providing a novel alternative pathway to remove hypoxanthine in HA.

Tga NucS is able to cleave both strands of U- and I- containing dsDNA, yielding double strand breaks. Since they are extremely severe to the archaeal cells, the generated DSBs need to be further repaired by HR. Luckily, the *T. gammatolerans* genome encodes the proteins of HR, which are conserved in bacteria and euarya. Furthermore, the ends of cleaved DNA product of Tga NucS possess 5'-phosphate and 3'-hydroxyl termini (Fig.6), which can be utilized by the proteins in HR. It has been proposed that *T. kodakarensis* cells and other euryarchaea contain multiple copies of genomic DNA (7-19 copies per cell) [30], which would be favorable for efficient HR. Many euryarchaea, such as *Thermococcus*, are highly polyploid [31], which would be advantageous for efficient HR, whereas the crenarchaea have eukaryal-like monoploid and diploid stages [32]. Taken together, efficient HR in *T. gammatolerans* cells would be expected to process a double strand break created by U- and I- containing DNA

cleavage of Tga NucS.

Like in *T. kodakarensis*, the Tga NucS CDS (NCBI gene ID: 7988711) in *T. gammatolerans* genome [27], starts a few nucleotides downstream the end of the RadA CDS (NCBI gene ID: 7988710), suggesting that they may belong to the same operon. RadA protein is conserved in bacteria and archaea [33-36], and as Rad51 in eukarya [37]. The co-expression of the Tga NucS and RadA would favor repair of double-strand breaks created by the enzyme.

In summary, we present important evidence that Tga NucS is capable of recognizing and cleaving U- and I-containing dsDNA at the close physiological temperature of its host, thus providing an novel alternative pathway for repair of damaged bases in DNA that arise from deamination. The U- and I-containing dsDNA would be recognized and cleaved by Tga NucS at both strands surrounding the lesion position, yielding a double strand break with cohesive ends with a 4-nt 5'-overhang that might be repaired by HR. Overall, NucS appears as a key protein for repairing deaminated bases in DNA in hyperthermophilic *Thermococcus* species that thrive in high temperature environment.

Supplementary data

Supplementary data are available at DNA Repair online.

Acknowledgement

We thank Prof. Fabrice Confalonieri at Radiorésistance des Bactéries et des Archaea Université Paris Sud for providing *T. gammatolerans* genomic DNA and critical reading. We also thank current group members in our lab for their work.

Funding

This work was supported by the Provincial Natural Science Foundation of Jiangsu (No. BK20191219), the Academic Leader of Middle and Young People of Yangzhou University Grant and State Key Laboratory of Microbial Metabolism, Shanghai JiaoTong University (No. MMLKF18-05) to L.Z.; the Practice Innovation Training Program for Postgraduate Students in Yangzhou University to H.S (No. XKYCX18_072).

Conflict of interest statement

All authors declare they have no conflict of interest.

Figure legends

Fig. 1. The genome of *T. gammatolerans* encodes a putative NucS endonuclease. **A.** Partial amino acid alignment of NucS endonucleases from euryarchaea and bacteria. Tga: *Thermococcus gammatolerans* (NCBI reference sequence: WP_015857754.1); Tko: *Thermococcus kodakarensis* (NCBI reference sequence: WP_011250849.1); Pfu: *Pyrococcus furiosus* (NCBI reference sequence: WP_014835498.1); Pab: *Pyrococcus abyssi* (PDB: 2VLD_B); Msm: *Mycobacterium smegmatis* (NCBI reference sequence: WP_003896320.1); Mtu: *Mycobacterium tuberculosis* (GenBank: SIP67590.1). The conserved amino acid residues are bolded. **B.** The overexpression and purification of the recombinant Tga NucS protein.

Fig. 2. Cleavage of U- and I-containing DNA by Tga NucS. **A.** The sequences of oligonucleotide duplex substrates (45 bp) containing normal, U or I base. **B.** The normal dsDNA (C:G) and U-containing dsDNA (U:G) (200 nM) cleavage reactions

of the enzyme (300 nM) were performed at 37, 60 and 80°C for 30 min, respectively.

C. The normal dsDNA (A:T) and U-containing dsDNA (I:T) (200 nM) cleavage reactions of the enzyme (300 nM) were performed at 37, 60 and 80°C for 30 min, respectively. The cleaved products were analyzed by a denaturing polyacrylamide gel. CK: the reaction without the enzyme.

Fig. 3. Substrate specificity of Tga NucS. **A.** The sequences of oligonucleotide duplex substrates (45 bp) containing normal, U or I base. **B.** The U-containing dsDNA (U:G, U:C, U:T and U:A) cleavage reactions of the enzyme with varied concentrations (100, 200 and 400 nM) were performed at 80°C for 30 min. **C.** The I-containing dsDNA (I:T, I:A, I:G and I:C) cleavage reactions of the enzyme with varied concentrations (100, 200 and 400 nM) were performed at 80°C for 30 min. **D.** The U-containing dsDNA (U:G) and I-containing dsDNA (I:T) cleavage reactions of the D163A mutant with varied concentrations (100, 200 and 400 nM) were performed at 80°C for 30 min. The cleaved products were analyzed by running a denaturing polyacrylamide gel. CK: the reaction without the enzyme.

Fig. 4. Biochemical properties of U-containing dsDNA cleavage of Tga NucS. **A.** The sequences of oligonucleotide duplex substrates (45 bp) containing the U base. **B.** The optimal temperature of the enzyme activity; **C.** The thermo-tolerance of the enzyme; **D.** The optimal pH of the enzyme activity; **E.** The effects of divalent cations on the enzyme activity; **F.** the effect of NaCl on the enzyme activity. The U-containing dsDNA cleavage reactions were performed at 80°C. The cleaved products were analyzed by a denaturing polyacrylamide gel. CK in panels B, D and F, and CK1 in

panels C and E: the reaction without the enzyme; CK2 in panel C: the reaction with unheated enzyme; CK2 in panel D: the reaction without a divalent cation.

Fig. 5. Cleavage site and re-ligation of the cleaved products of Tga NucS. **A.** The sequences of oligonucleotide duplex substrate (45 bp) containing a uracil or hypoxanthinebase. The 5'- Cy3-labeled strand containing U or I and the 5'- Hex-labeled complementary strand were annealed to create the U- and I- containing dsDNA substrates. **B.** The substrates were used in DNA cleavage reactions of the enzyme at 80°C for 30 min. The cleaved products were analyzed by a denaturing polyacrylamide gel. Lane 1: the DNA markers (24 mer, 22 mer and 18 mer); lane 2: the reaction without the enzyme; lane 3: U-containing dsDNA cleavage; lane 4: I-containing dsDNA cleavage. **C.** The sequences of oligonucleotide duplex substrates (45 bp) containing normal, U or I base. **D.** The DNA cleavage reactions of the enzyme using U-and I- containing DNA as the substrates were performed at 80°C for 30 min. The cleaved products were purified and ligated by T4 DNA ligase at 22°C. The ligated products were analyzed by running a denaturing polyacrylamide gel.

Fig. 6. DNA binding assays of Tga NucS. **A.** The sequences of oligonucleotide duplex substrates (45 bp) containing normal, U or I base. **B.** The normal and U-containing ssDNA and dsDNA substrates were incubated with the enzyme with various concentrations (500, 1000 and 2000 nM) at room temperature for 10 min. **C.** The normal and I-containing ssDNA and dsDNA substrates were incubated with the enzyme with various concentrations (500, 1000 and 2000 nM) at room temperature for 10 min. The retarded DNA products were analyzed by a native polyacrylamide

gel. CK: the binding reaction without the enzyme.

References

1. H.E. Krokan, F. Drablos, G. Slupphaug, Uracil in DNA - occurrence, consequences and repair, *Oncogene* 21 (2002), 8935-8948.
2. M. Hill-Perkins, M.D. Jones, P. Karran, Site-specific mutagenesis in vivo by single methylated or deaminated purine bases, *Mutat. Res.* 162 (1986) 153-163.
3. T. Lindahl, B. Nyberg, Heat-induced deamination of cytosine residues in deoxyribonucleic acid, *Biochemistry* 13 (1974) 3405-3410.
4. K.L. Jacobs, D.W. Grogan, Rates of spontaneous mutation in an archaeon from geothermal environments, *J. Bacteriol.* 179 (1997) 3298-3303.
5. D.W. Grogan, G.T. Carver, J.W. Drake, Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*, *Proc. Natl. Acad. Sci. USA* 98 (2001) 7928-7933.
6. Z. Kelman, M.F. White, Archaeal DNA replication and repair, *Curr. Opin. Microbiol.* 8 (2005) 669-676.
7. K.M. Zatopek, A.F. Gardner, Z. Kelman, Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms, *FEMS Microbiol. Rev.* 42 (2018) 477-488.
8. M.F. White, T. Allers, DNA repair in the archaea-an emerging picture, *FEMS Microbiol. Rev.* 42 (2018) 514-526.
9. Grogan, D.W. (1998) Hyperthermophiles and the problem of DNA instability. *Mol Microbiol*, 28, 1043-1049.
10. C. Rouillon, M.F. White, The evolution and mechanisms of nucleotide excision

- repair proteins, *Res. Microbiol.* 162 (2011) 19-26.
11. S. Grasso, G. Tell, Base excision repair in Archaea: back to the future in DNA repair, *DNA Repair (Amst)* 21 (2014) 148-157.
 12. S.J. Sandler, L.H. Satin, H.S. Samra, A.J. Clark, recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmcl proteins of the yeast *Saccharomyces cerevisiae*, *Nucleic Acids Res.* 24 (1996) 2125-2132.
 13. M.F. Goodman, R. Woodgate, Translesion DNA polymerases, Cold Spring Harb. *Perspect Biol.* 5 (2013) a010363.
 14. D.W. Grogan, Stability and repair of DNA in hyperthermophilic Archaea. *Curr. Issues Mol. Biol.* 6 (2004) 137-144.
 15. A. Yasui, Alternative excision repair pathways. Cold Spring Harb. *Perspect Biol.* 5 (2013).
 16. M. Shiraishi, S. Ishino, T. Yamagami, Y. Egashira, S. Kiyonari, Y. Ishino, A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*, *Nucleic Acids Res.* 43 (2015) 2853-2863.
 17. W. Cao, Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell Mol. Life Sci.* 70 (2013) 3145-3156.
 18. S. Ishino, N. Makita, M. Shiraishi, T. Yamagami, Y. Ishino, EndoQ and EndoV work individually for damaged DNA base repair in *Pyrococcus furiosus*, *Biochimie* 118 (2015) 264-269.
 19. B. Ren, J. Kuhn, L. Meslet-Cladiere, J. Briffotiaux, C. Norais, R. Lavigne, D.

- Flament, R. Ladenstein, H. Myllykallio, Structure and function of a novel endonuclease acting on branched DNA substrates, *EMBO J.* 28 (2009) 2479-2489.
20. C. Creze, A. Ligabue, S. Laurent, R. Lestini, S.P. Laptanok, J. Khun, M.H. Vos, M. Czjzek, H. Myllykallio, D. Flament, Modulation of the *Pyrococcus abyssi* NucS endonuclease activity by replication clamp at functional and structural levels, *J. Biol. Chem.* 287 (2012) 15648-15660.
21. S. Ishino, Y. Nishi, S. Oda, T. Uemori, T. Sagara, N. Takatsu, T. Yamagami, T. Shirai, Y. Ishino, Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea, *Nucleic Acids Res.* 44 (2016) 2977-2986.
22. N. Takemoto, I. Numata, M. Su'etsugu, T. Miyoshi-Akiyama, Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. *Nucleic Acids Res.* 46 (2018) 6152-6165.
23. S. Ishino, S. Skouloubris, H. Kudo, C. l'Hermitte-Stead, A. Es-Sadik, J.C. Lambry, Y. Ishino, H. Myllykallio, Activation of the mismatch-specific endonuclease EndoMS/NucS by the replication clamp is required for high fidelity DNA replication, *Nucleic Acids Res.* 46 (2018) 6206-6217.
24. A. Castaneda-Garcia, A.I. Prieto, J. Rodriguez-Beltran, N. Alonso, D. Cantillon, C. Costas, L. Perez-Lago, E.D. Zegeye, M. Herranz, P. Plocinski, et al, A non-canonical mismatch repair pathway in prokaryotes, *Nat Commun.* 8 (2017) 14246.
25. S. Suzuki, N. Kurosawa, Endonucleases responsible for DNA repair of helix-

- distorting DNA lesions in the thermophilic crenarchaeon *Sulfolobus acidocaldarius* in vivo. *Extremophiles* 23 (2019) 613-624.
26. E. Jolivet, S. L'Haridon, E. Corre, P. Forterre, D. Prieur, *Thermococcus gammatolerans* sp nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation, *Int. J. Syst. Evol. Microb.* 53 (2003) 847-851.
 27. Y. Zivanovic, J. Armengaud, A. Lagorce, C. Leplat, P. Guerin, M. Dutertre, V. Anthouard, P. Forterre, P. Wincker, F. Confalonieri,) Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea, *Genome Biol.* 10 (2009).
 28. G.L. Moldovan, B. Pfander, S. Jentsch, PCNA, the maestro of the replication fork, *Cell* 129 (2007) 665-679.
 29. K.I. Miyazono, S. Ishino, N. Makita, T. Ito, Y. Ishino, M. Tanokura, Crystal structure of the novel lesion-specific endonuclease Pfu EndoQ from *Pyrococcus furiosus*, *Nucleic Acids Res.* 46 (2018) 4807-4818.
 30. S.K. Spaans, J. van der Oost, S.W. Kengen, The chromosome copy number of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1, *Extremophiles* 19 (2015) 741-750.
 31. C. Hildenbrand, T. Stock, C. Lange, M. Rother, J. Soppa, Genome copy numbers and gene conversion in methanogenic archaea, *J. Bacteriol.* 193 (2011) 734-743.
 32. M. Lundgren, R. Bernander, Genome-wide transcription map of an archaeal cell cycle, *Proc. Natl. Acad. Sci. USA* 104 (2007) 2939-2944.

33. K. Komori, T. Miyata, J. DiRuggiero, R. Holley-Shanks, I. Hayashi, I.K. Cann, K. Mayanagi, H. Shinagawa, Y. Ishino, Both RadA and RadB are involved in homologous recombination in *Pyrococcus furiosus*, *J. Biol. Chem.* 275 (2000) 33782-33790.
34. S. Haldenby, M.F. White, T. Allers, RecA family proteins in archaea: RadA and its cousins, *Biochem. Soc. Trans.* 37 (2009) 102-107.
35. M. Prentiss, C. Prevost, C. Danilowicz, Structure/function relationships in RecA protein-mediated homology recognition and strand exchange, *Crit. Rev. Biochem. Mol. Biol.* 50 (2015) 453-476.
36. A. Peacock-Villada, D. Yang, C. Danilowicz, E. Feinstein, N. Pollock, S. McShan, V. Coljee, M. Prentiss, Complementary strand relocation may play vital roles in RecA-based homology recognition, *Nucleic Acids Res.* 40 (2012) 10441-10451.
37. S.K. Godin, M.R. Sullivan, K.A. Bernstein, Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication, *Biochem. Cell Biol.* 94 (2016) 407-418.

A

```

Tga 28 VLTIFARCRVHYDGRAKSELGPGDRVIIVKPDGSGFLIHQKEKR-EPVNWQ-PPGSVVRLELREKPVLSVRRKPRETLEVELEDEVYLITVF
Tko 30 MLTIFARCKVHYDGRAKSELGSGDRVIIVKPDGSGFLIHQSCKR-EPVNWQ-PPGSRVRLELRENPLVLSIRRKPRETLEVELEEVYMSVVF
Pfu 30 LLTIFARCTVYYEGRAKSELGEGDRIIIIKPDGSGFLIHQKKR-EPVNWQ-PPGSKVKMEGNS---LISIRRNPKETLKVVDIEEYAAVLF
Pab 28 VVTIFARCKVHYEGRAKSELGEGDRIIIIKPDGSGFLIHQNKRR-EPVNWQ-PPGSKVTF---KENSXISIRRRPYERLEVEIEPYSLVVF
Msm 1 MRLVIAQCTVDYVGRHTAHLPSARRLLLFKADGSGSVHADDRAYKPLNWMSPPCWVTEQDTEGTGVALWVVENKTGEQLRITVEDIE-----
Mtu 1 MRLVIAQCTVDYVGRHTAHLPSARRLLLFKADGSGSVHADDRAYKPLNWMSPPCWLTEESGGQAPVWV-VENKAGEQLRITIEGIE-----

```

Motif I

Motif II

Motif III

Motif IV

```

Tga 117 HAEDYE-----ELALTGSEAEAEELIFENPEVIELGFKPLYREKPIRHGIVDVLGVDRDGNLVLVLELKRRRADLHAVSQLKRYVETLREE
Tko 119 RAEDYE-----ELALTGSEAEAEELIFENPEVIEPGFKPLFREKAIGTIVDVLGRDSDGNIVVLELKRRRAELHAVRQLKSYVEILREE
Pfu 116 MAEDYE-----ELTLTGSEAEAEELIFQNPVIEEGFKPMFREKPIKHGIVDVLGVDRDGNLVLVLELKRRRADLHAVSQLKRYVDALKEE
Pab 114 LAEDYE-----ELALTGSEAEEXANLIFENPRVIEEGFKPIYREKPIRHGIVDVXGVDKDGNIVVLELKRRKADLHAVSQKRYVDSLKEE
Msm 87 HDSSHELGVDPGLVKDGEAHLQALLAEHVELLGAGYTLVRRREYPTPIGPVDLLCRDELGRSVAVEIKRRGEIDG-VEQLTRYLELLNRD
Mtu 86 HDSSHELGVDPGLVKDGEAHLQALLAEHIQLLGEYTLVRRREYMTAIGPVDLLCRDERGGSVAVEIKRRGEIDG-VEQLTRYLELLNRD

```

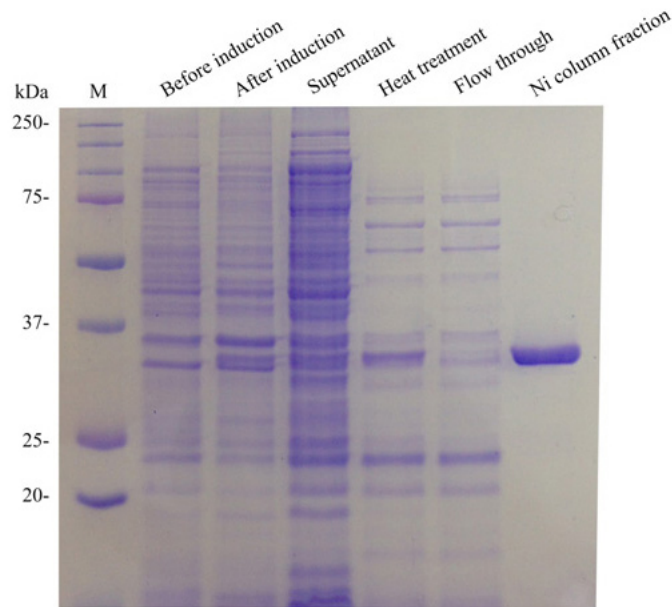
Motif V

```

Tga 202 H--EN-VRGILVAPSLTSGAKKLEKEGLEFRKLEPPKRDGKSRGRQLRLEF-----
Tko 204 YG-DK-VRGILVAPSLTSGAKRLEKEGLEFRKLEPPKRDGKSRGRQKTLF-----
Pfu 201 H--GNKVRGILVAPSLTEGAKKLEKLEGLEFRKLEPPKGGKSSKQKTLDFLNDTVRITGASPPPEAIQ
Pab 199 YG-EN-VRGILVAPSLTEGAKKLEKEGLEFRKLEPPKGGNEKRSKQKTLDFFTF-----
Msm 176 SLLAP-VAGVFQAQKIQARTLATDRGIRCVTLDDYDQMRGMD-SDEYRLF-----
Mtu 175 SVLAP-VKGVFAAQKIQARILATDRGIRCVTLDDYDQMRGMD-SGEYRLF-----

```

B



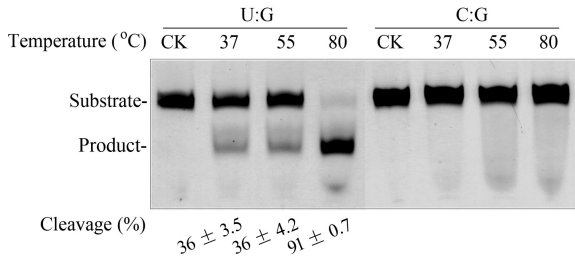
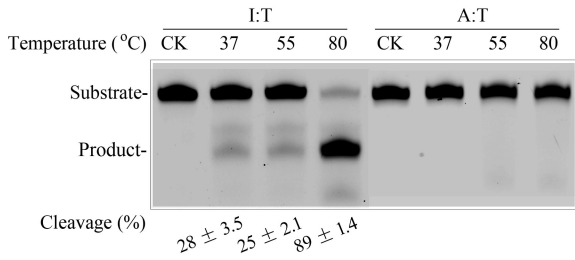
A

X:U, C, I or A

5' Cy3-CGAACTGCCTGGAATCCTGACGACX~~T~~GTAGCGAACGATCACCTCA

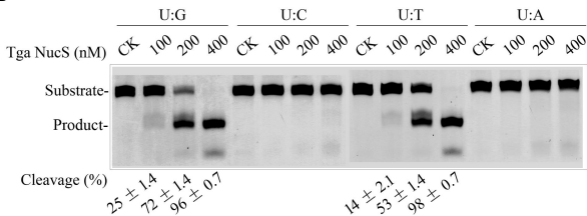
Y:G or T

3'-GCTTGACGGACCTTAGGACTGCTGYACATCGCTTGCTAGTGGAGT

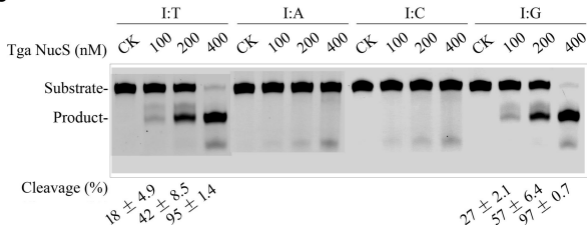
B**C**

A **X:U or I**
 5' Cy3-CGAACTGCCTGGAATCCTGACGAC~~X~~TGTAGCGAACGATCACCTCA
 3' -GCTTGACGGACCTTAGGACTGCTGYACATCGCTTGCTAGTGGAGT
Y:G,C,T or A

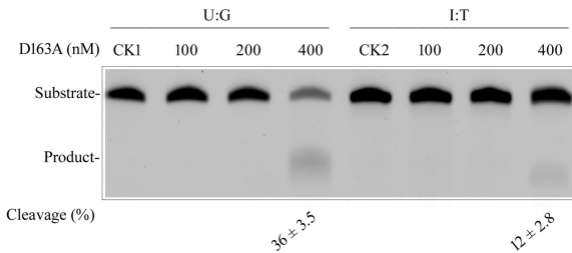
B



C

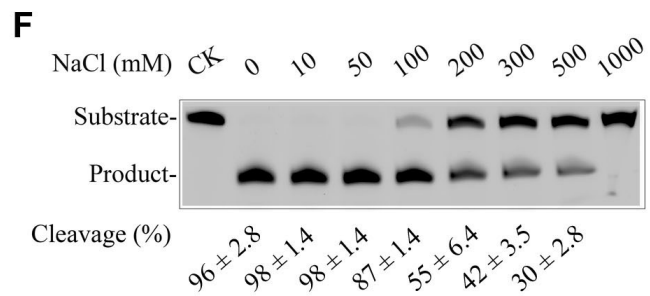
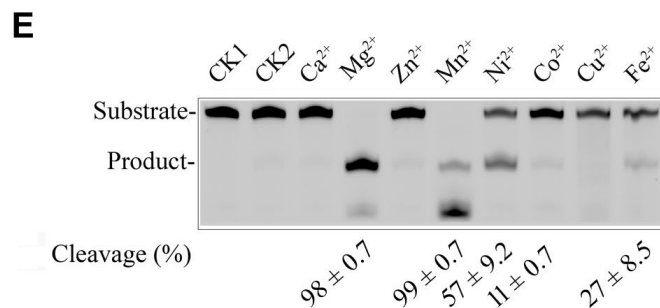
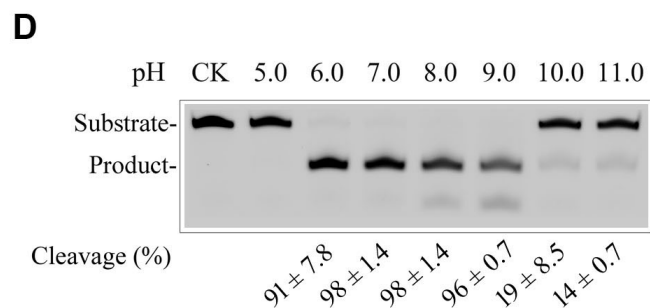
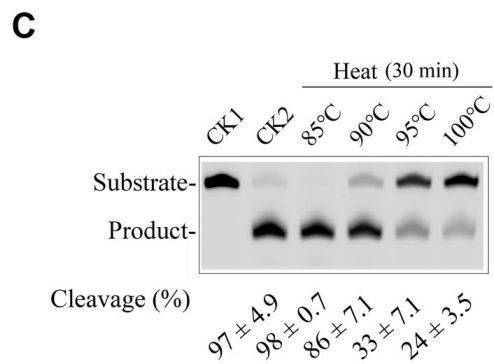
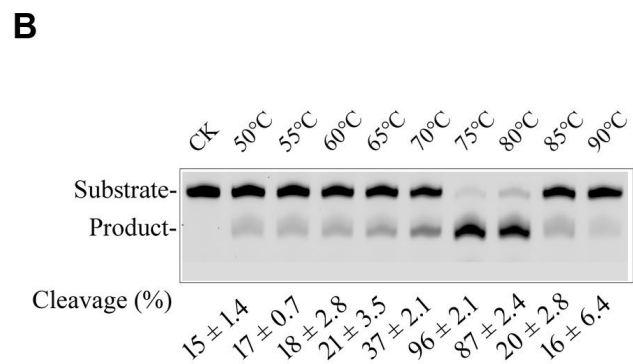


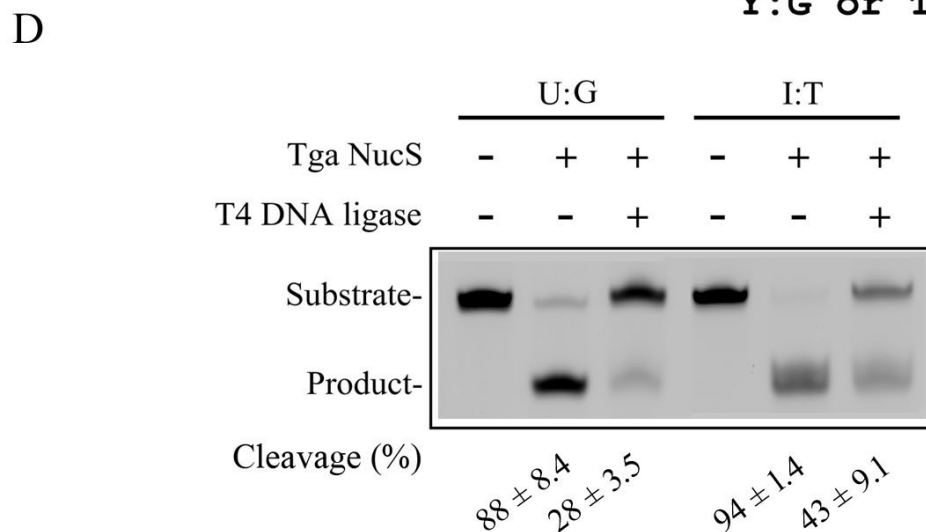
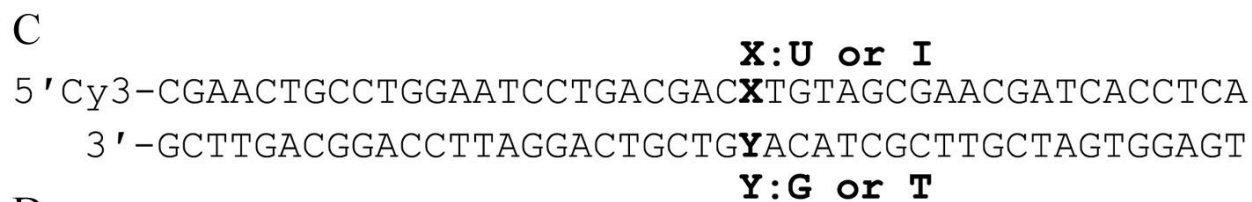
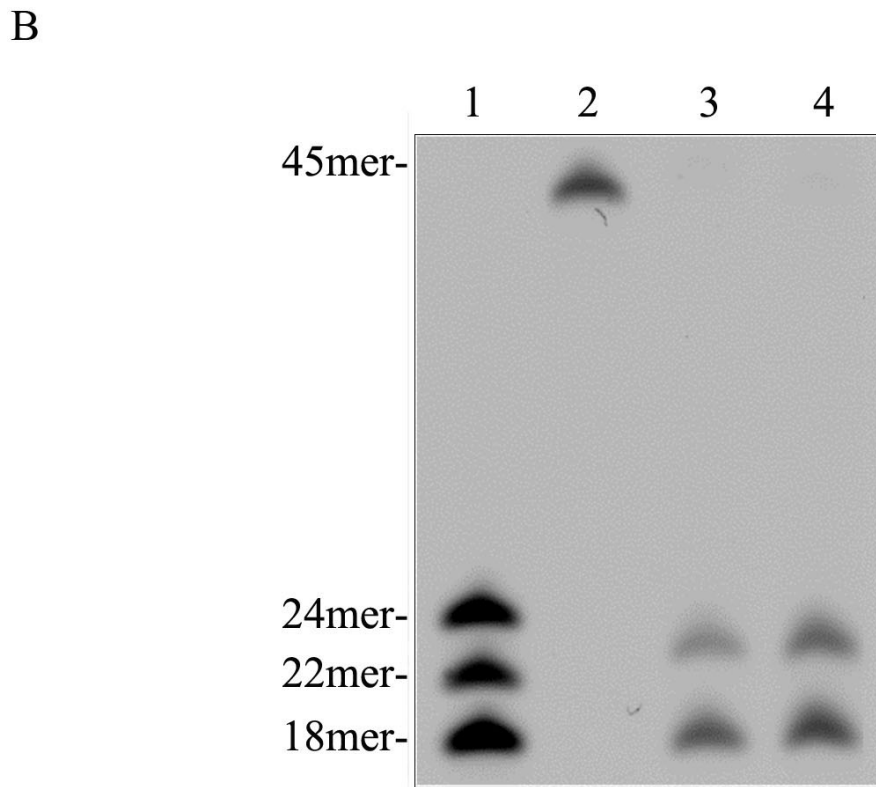
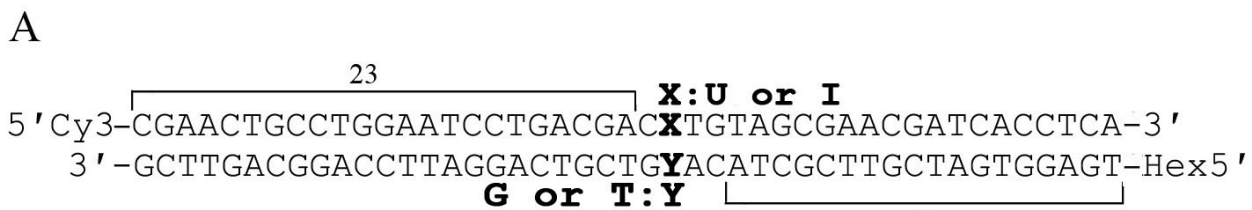
D



A

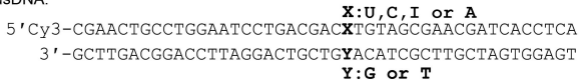
5' Cy3-CGAACTGCCTGGAATCCTGACGAC**U**TGTAGCGAACGATCACCTCA
 3' -GCTTGACGGACCTTAGGACTGCTG**G**ACATCGCTTGCTAGTGGAGT



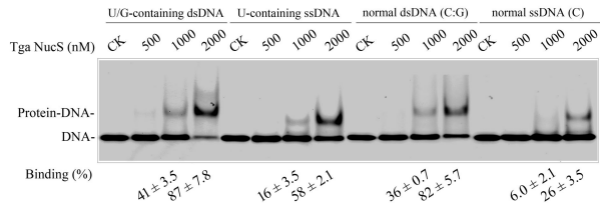


A

dsDNA:



ssDNA:

**B****C**