Supplemental Material

Molecular basis of substrate recognition of endonuclease Q from the euryarchaeon *Pyrococcus furiosus*

Miyako Shiraishi[#] and Shigenori Iwai

Department of Materials Engineering Science, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyamacho, Toyonaka, Osaka, Japan

Running title: Identifying new genome maintenance functions of EndoQ

Address correspondence to Miyako Shiraishi (m.shiraishi@chem.es.osaka-u.ac.jp)

Table S1. Oligonucleotides used in this study.

Name	Length (nt)	Lesion/Modification	Position	Sequence (5′-3′)	Description
BsuEndoQ-F	39	none		ATTTTCAGGGCCATATGAAAACCATCTATGCGGATCTGC	For the cloning of the BsuEndoQ gene
BsuEndoQ-R	39	none		TGCTCGAGTGCGGCCGCTTACGGCTTGATTCTGCCGTAC	For the cloning of the BsuEndoQ gene
BsEndoQ-E90A-F	37	none		TCATCATAAATCTCCAGCGCGCTCCCCAATAAAAGAG	For the E90A mutation
BsEndoQ-E90A-R	37	none		CTCTTTTATTGGGGAGCGCGCTGGAGATTTATGATGA	For the E90A mutation
marker (13)	13	none		CTCGTCAGCATCT	
marker (14)	14	none		CTCGTCAGCATCTT	
T	30	none		CTCGTCAGCATCT**T**CATCATACAGTCAGTG	
U	30	uracil	14	CTCGTCAGCATCT**(U)**CATCATACAGTCAGTG	
Hx	30	hypoxanthine	14	CTCGTCAGCATCT**(Hx)**CATCATACAGTCAGTG	
5mC	30	5-methylcytosine	14	CTCGTCAGCATCT**(5mC)**CATCATACAGTCAGTG	
5hmU	30	5-hydroxymethyluracil	14	CTCGTCAGCATCT**(5hmU)**CATCATACAGTCAGTG	
cis CPD	30	*cis-sin* cyclobutane pyrimidine dimer	14,15	CTCGTCAGCATCT**(cisCPD)**CATCATACAGTCAGTG	
trans CPD	30	*trans-sin* cyclobutane pyrimidine dimer	14,15	CTCGTCAGCATC**(transCPD)**CATCATACAGTCAGTG	
6-4PP	30	(6-4) photoproduct	14,15	CTCGTCAGCATC**(6-4PP)**CATCATACAGTCAGTG	
DewarPP	30	Dewar photoproduct	14,15	CTCGTCAGCATC**(DewarPP)**CATCATACAGTCAGTG	
AP	30	abasic furan (dSpacer)	14	CTCGTCAGCATCT**(AP)**CATCATACAGTCAGTG	
5hU	30	5-hydroxyuracil	14	CTCGTCAGCATCT**(5hU)**CATCATACAGTCAGTG	
5hC	30	5-hydoxycytosine	14	CTCGTCAGCATCT**(5hC)**CATCATACAGTCAGTG	
DHT	30	5,6-dihydrothymine	14	CTCGTCAGCATCT**(DHT)**CATCATACAGTCAGTG	
DHU	30	5,6-dihydrouracil	14	CTCGTCAGCATCT**(DHU)**CATCATACAGTCAGTG	
*R*Tg	30	(5*R*)-thymine glycol	14	CTCGTCAGCATCT**(RTg)**CATCATACAGTCAGTG	
*S*Tg	30	(5*S*)-thymine glycol	14	CTCGTCAGCATCT**(STg)**CATCATACAGTCAGTG	
G:A	30	none	14	CTCGTCAGCATCT**G**CATCATACAGTCAGTG	G:A mismatch
C:A	30	none	14	CTCGTCAGCATCT**C**CATCATACAGTCAGTG	C:A mismatch
6mA	30	N^6-methyladenine	14	CTCGTCAGCATCT**(6mA)**CATCATACAGTCAGTG	
6mG	30	O^6-methylguanine	14	CTCGTCAGCATCT**(6mG)**CATCATACAGTCAGTG	
8oxoG	30	8-oxoguanine	14	CTCGTCAGCATCT**(8oxoG)**CATCATACAGTCAGTG	
rU	30	uridine	14	CTCGTCAGCATCT**(rU)**CATCATACAGTCAGTG	
rA	30	adenosine	14	CTCGTCAGCATCT**(rA)**CATCATACAGTCAGTG	
RNA	30	none		CUCGUCAGCAUCUUCAUCAUACAGUCAGUG	RNA
dC-RNA	30	deoxycytosine	12	CUCGUCAGCAU**(dC)**UUCAUCAUACAGUCAGUG	RNA
comp(A)	30	none		CACTGACTGTATGATGAAGATGCTGACGAG	complement DNA
comp(T)	30	none		CACTGACTGTATGATGTAGATGCTGACGAG	complement DNA
comp(G)	30	none		CACTGACTGTATGATGGAGATGCTGACGAG	complement DNA
comp(C)	30	none		CACTGACTGTATGATGCAGATGCTGACGAG	complement DNA

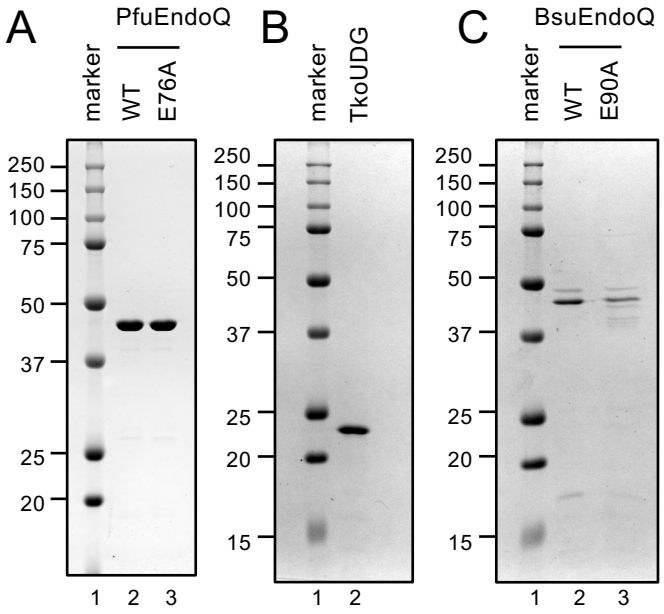

Figure S1. Purified recombinant proteins. Each protein was visualized by 12.5% SDS-PAGE, followed by Coomassie Brilliant Blue staining. Molecular weight (MW) standards are indicated on the left of the panels (lane 1, Bio-Rad, #1610374). (A) Lane 2, PfuEndoQ^{WT} (MW: 47639.03; 1 µg); lane 3, PfuEndoQ^{E76A} (MW: 47580.99; 1 µg). (B) Lane 2, TkoUDG (MW: 22463.3; 1µg). (C) Lane 2, BsuEndoQ^{WT} (MW: 44863.37; 0.5 µg); lane 3 BsuEndoQ^{E90A} (MW: 44808.33; 0.5 µg). Purification of TkoUDG: he expression vectors encoding TkoUDG (BAD86332) (pET21a-TkoUDG) was gifted by Dr. Y. Ishino (Kyushu University, Fukuoka, Japan). TkoUDG was overproduced in *E. coli* BL21 CodonPlus (DE3)-RIL (Agilent Technologies) cells containing pET21a-TkoUDG. The cells were grown with shaking in 1 L of Luria-Bertani (LB) medium, containing 50 µg/mL ampicillin and 34 µg/mL chloramphenicol (Cm) at 37°C until the optical density (OD) at 600 nm was 0.5. To produce proteins, the inducer isopropyl β-D-thiogalactopyranoside (IPTG, 0.1 mM) was added to the culture and the cells were further grown overnight at 18°C. The cells were collected by centrifugation and sonicated in buffer A (50 mM Tris-HCl, pH 8.0, 0.5 mM DTT, and 10% glycerol) containing 0.1 M NaCl and 1 mM phenylmethylsulfonyl fluoride (PMSF). The soluble fraction was obtained by centrifugation and subjected to a 1 mL HiTrap Heparin HP column (GE Healthcare) and eluted with a linear gradient of 50–1000 mM NaCl in buffer A. Fractions containing TkoUDG as observed on a 15% SDS-PAGE gel, were pooled and diluted 10-fold with buffer A, and subjected to a 1 mL HiTrap SP HP column (GE Healthcare). The column was developed with a linear gradient of 50–1000 mM NaCl in buffer A. The eluted protein fractions were stored at −20°C with 50% glycerol. The protein purities were evaluated with 12.5% SDS-PAGE, followed by Coomassie brilliant blue (CBB) staining. TkoUDG (17670 M−1 cm−1) were estimated as described in the main text.

Figure S1. Purified recombinant proteins (Continued). Cloning and purification of recombinant BsuEndoQ: The gene for BsuEndoQ (KIX84288, YqxK) was amplified directly by PCR from the *B. subtilis* genomic DNA using gene-specific primers (Table S1). The amplified gene was cloned into the expression vector pET28TEV, which is a modified plasmid of pET28a (Novagen) from a thrombin recognition site to the tobacco etch virus (TEV) protease recognition site, using the In-Fusion HD Cloning Kit (Clontech). The resulting plasmids were designated as pET-BsuEndoQ^{WT}. The E90A mutation in BsuEndoQ was generated using the primers (Table S1). The resulting plasmid was designated as pET-BsuEndoQ^{E90A}. The nucleotide sequences of the inserted regions of all plasmids were confirmed by sequencing. For the protein purification, BsuEndoQ was overproduced in *E. coli* BL21 CodonPlus (DE3)-RIL (Agilent Technologies) cells carrying pET28-BsuEndoQ^{WT} or BsuEndoQ^{E90A}. The cells were grown with shaking in 2 L of LB medium containing 20 μg/ml kanamycin and 34 μg/mL Cm at 37°C (OD_{600} = 0.3). Then, IPTG (0.1 mM) was added to the culture and the cells were further grown overnight at 18°C. The collected cells were sonicated in buffer B (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl) containing 30 mM imidazole and 1 mM PMSF. The soluble fraction was obtained by centrifugation and loaded onto a 1 mL HisTrap HP column (GE Healthcare) and eluted with a linear gradient of 40–300 mM imidazole in buffer B. Fractions containing BsuEndoQ as observed on a 12.5% SDS-PAGE gel, were pooled and diluted 5-folds with buffer B, and subjected to a 1 mL HisTrap HP column (GE Healthcare) again by a step elution with 300 mM imidazole. The pooled fractions containing BsuEndoQ were further subjected to a 1 mL HiTrap Heparin HP column (GE Healthcare). The protein was eluted by a step elution with 500 mM NaCl in buffer (50 mM Tris-HCl, pH 8.0, 0.5 mM DTT, 10% glycerol). The eluted protein fractions were stored at 4°C. Protein purity was evaluated with 12.5% SDS-PAGE, followed by CBB staining. The protein concentrations were determined by measuring the absorbance at 280 nm. The theoretical molar extinction coefficient of BsuEndoQ (28225 M^{-1} cm^{-1}) was estimated as described in the main text.

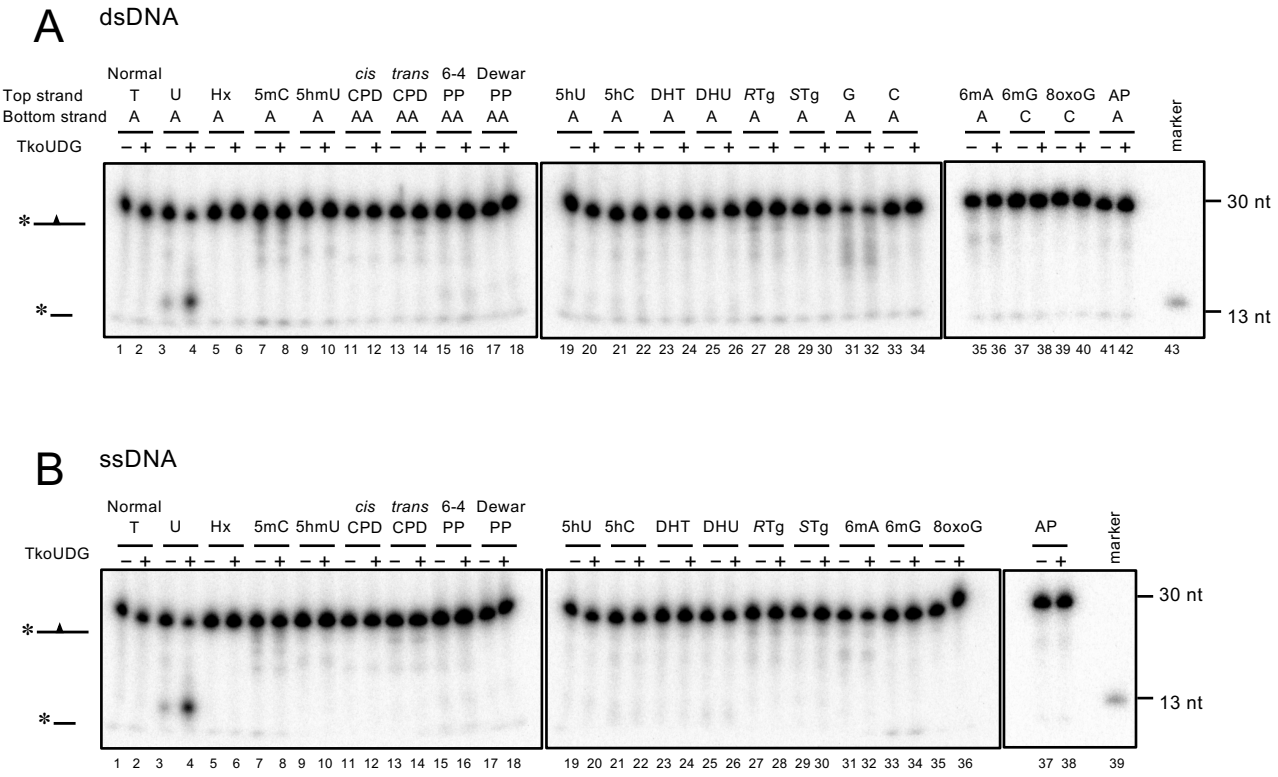

Figure S2. Substrate specificity of TkoUDG. DNA substrates containing various modified or damaged bases (Table S1) were incubated with TkoUDG (100 nM) at 60°C for 10 min in a 20 μL reaction mixture (50 mM Tris-HCl, pH 7.5, 1 mM DTT, 1 mM EDTA, 100 mM NaCl, and 5 nM ^{32}P-labeled DNA). The resulting AP site was visualized by cleaving the DNA backbone with 2 μL of 200 mM NaOH and heating at 95°C for 5 min, then 2 μL of 200 mM HCl was added for neutralization. The substrates were denatured with 10 μL of the stop solution, incubated at 95°C for 5 min, and immediately cooled on ice. DNA products were separated by 8 M urea-12% PAGE, and ^{32}P-labeled DNA strands were detected by autoradiography. (A) dsDNA. (B) ssDNA; −, no enzyme; +, 100 nM TkoUDG; marker, 13 nt.

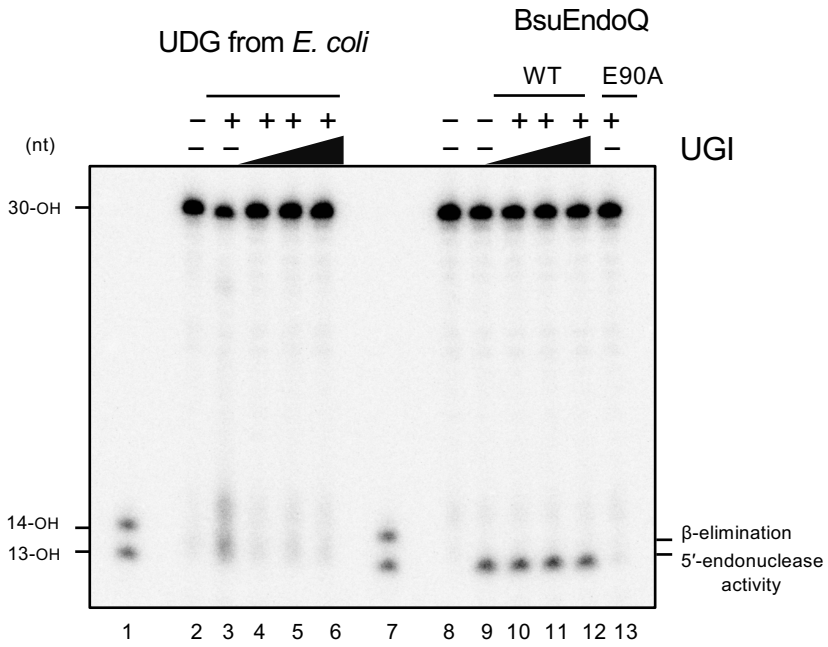

Figure S3. Activity of *B. subtilis* EndoQ is not inhibited by UGI from *Bacillus* bacteriophage PBS1. The cleavage reactions for *E. coli* UDG and BsuEndoQ were performed at 37°C for 30 min in a 20 μL reaction mixture (50 mM Tris-HCl, pH 8.0, 1 mM DTT, 1 mM $MnCl_2$, 0.01% Tween20, 50 mM NaCl, 5 nM ^{32}P-labeled U-containing dsDNA (U:A)) in the presence or absence of UDG from *E. coli* (purchased from New England Biolabs; 50 nM, lanes 3–6), and BsuEndoQ^{WT} (50 nM, lanes 9–11), BsuEndoQ^{E90A} (50 nM, lane 12) and UGI from *Bacillus* bacteriophage PBS1 (purchased from New England Biolabs; 50 nM, lanes 4 and 9; 100 nM, lanes 5 and 11; 200 nM, lanes 6 and 11). Glycosylase activity assay: lanes 2–5; endonuclease activity assay: lanes 8–13. DNA products were separated by 8 M urea-12% PAGE, and ^{32}P-labeled DNA strands were detected by autoradiography. Marker, 13 and 14 nt (Lane 1).

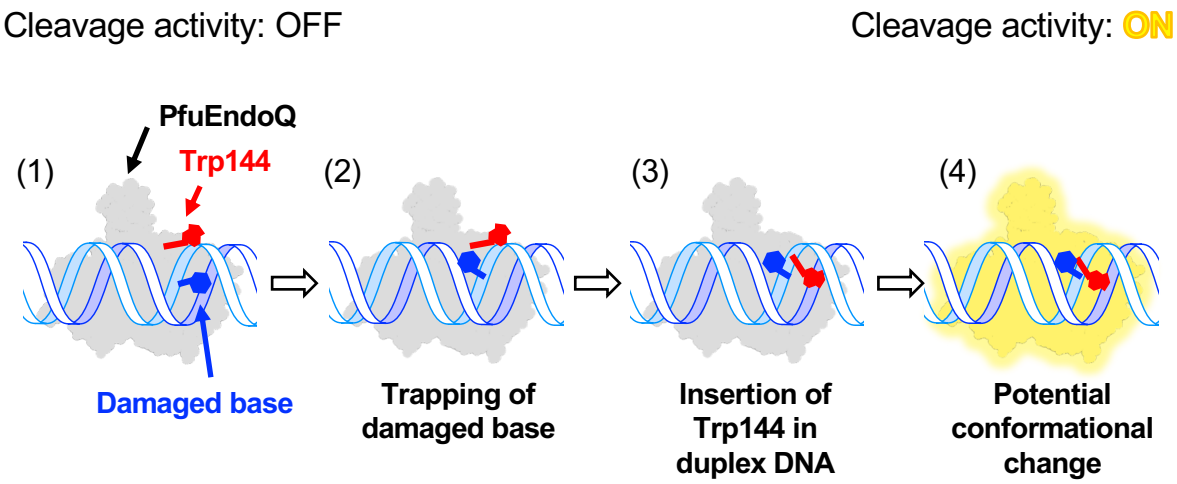

Figure S4. A model for cleavage activity by EndoQ. EndoQ has a relatively weak affinity to DNA (1), and may be used for searching for damaged bases in the genome (step 1). Once EndoQ finds a damaged base, the base is trapped (step 2) and a amino acid residue is inserted into the DNA duplex (step 3). This results in a conformational change in EndoQ (tight EndoQ-DNA binding), and leads to a cleavage activity (step 4). The structural image of PfuEndoQ was obtained from the protein data bank (5ZB8) (2).

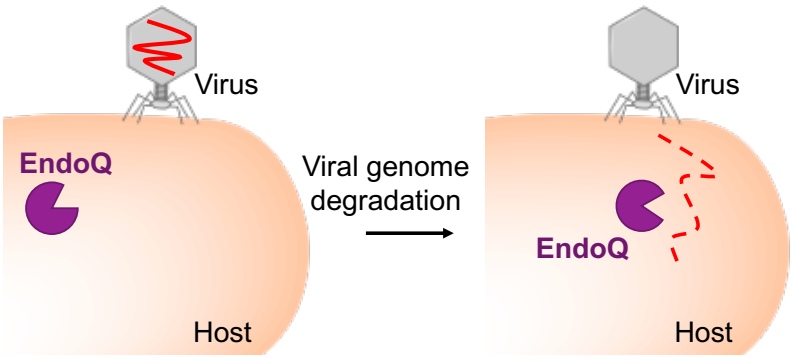

Figure S5. Possible role of EndoQ in restriction-modification systems. Modified viral genome (such as uracil and its derivatives) could be restricted by EndoQ.

References

1. Shiraishi M, Ishino S, Yamagami T, Egashira Y, Kiyonari S, Ishino Y. 2015. A novel endonuclease that may be responsible for damaged DNA base repair in *Pyrococcus furiosus*. Nucleic Acids Res 43:2853–2863.

2. Miyazono K, Ishino S, Makita N, Ito T, Ishino Y, Tanokura M. 2018. Crystal structure of the novel lesion-specific endonuclease PfuEndoQ from *Pyrococcus furiosus*. Nucleic Acids Res 46:4807–4818.