



HAL
open science

Biochemical characterization and mutational studies of the thermostable uracil DNA glycosylase from the hyperthermophilic euryarchaeon *Thermococcus barophilus* Ch5

Haoqiang Shi, Qi Gan, Donghao Jiang, Yuqi Wu, Youcheng Yin, Haiyue Hou, Hongxun Chen, Li Miao, Zhihui Yang, Phil M. Oger, et al.

► To cite this version:

Haoqiang Shi, Qi Gan, Donghao Jiang, Yuqi Wu, Youcheng Yin, et al.. Biochemical characterization and mutational studies of the thermostable uracil DNA glycosylase from the hyperthermophilic euryarchaeon *Thermococcus barophilus* Ch5. *International Journal of Biological Macromolecules*, 2019, 10.1016/j.ijbiomac.2019.05.073 . hal-02130096

HAL Id: hal-02130096

<https://hal.science/hal-02130096v1>

Submitted on 4 Nov 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.

1
2
3
4 **Biochemical characterization and mutational studies of the thermostable uracil**
5
6 **DNA glycosylase from the hyperthermophilic euryarchaeon *Thermococcus***
7 ***barophilus* Ch5**
8
9

10
11
12
13
14
15
16 Haoqiang Shi^a, Qi Gan^a, Haiyue Hou^a, Hongxun Chen^a, Yinuo Xu^a, Li Miao^a, Zhihui

17
18 Yang^{b#}, Philippe Oger^{c#} and Likui Zhang^{a#}
19
20
21
22

23 ^aMarine Science & Technology Institute

24
25 Department of Environmental Science and Engineering, Yangzhou University, China
26

27 ^bCollege of Plant Protection, Agricultural University of Hebei, Baoding City, Hebei

28
29 Province 071001, China
30
31

32 ^cUniversité de Lyon, INSA de Lyon, CNRS UMR 5240, Lyon, France
33
34

35 Corresponding author: Dr. Likui Zhang

36
37 E-mail address: lkzhang@yzu.edu.cn
38

39
40 Tel: +86-514-89795882
41

42
43 Fax: +86-514-87357891
44

45 Corresponding author: Prof. Zhihui Yang

46
47 E-mail address: bdybh@hebau.edu.cn
48

49 Corresponding author: Prof. Philippe Oger

50
51 E-mail address: philippe.oger@insa-lyon.fr
52
53
54
55
56
57
58
59

60
61
62
63 **Abstract**
64

65 Uracil DNA glycosylases (UDGs) play an important role in removing uracil from
66 DNA to initiate DNA base excision repair. Here, we first characterized biochemically
67 a thermostable UDG from the hyperthermophilic euryarchaeon *Thermococcus*
68 *barophilus* Ch5 (Tba UDG), and probed its mechanism by mutational analysis.
69
70 The recombinant Tba UDG cleaves specifically uracil-containing ssDNA and dsDNA
71 at 65°C. The enzyme displays an optimal cleavage activity at 55–75°C. Tba UDG
72
73 cleaves DNA over a wide pH spectrum ranging from 4.0 to 9.0 with an optimal pH of
74
75 5.0–8.0. In addition, the Tba UDG activity is independent on a divalent metal ion;
76
77 however, both Zn²⁺ and Cu²⁺ completely inhibits the enzyme activity. Furthermore,
78
79 the Tba UDG activity is also inhibited by high NaCl concentration. Tba UDG
80
81 removes uracil from DNA by the order: U≈U/G>U/T≈U/C>U/A. The mutational
82
83 studies showed that both the E118A and N159A mutants completely abolish the
84
85 cleavage activity and retain the compromised binding activity, suggesting that
86
87 residues E118 and N159 in Tba UDG are important for uracil recognition and removal.
88
89 Our work provides a basis for determining the role of Tba UDG in the base excision
90
91 repair pathway for uracil repair in *Thermococcus*.
92
93
94
95
96
97
98
99
100
101
102

103 **Keywords:** *Thermococcus barophilus*; Uracil DNA glycosylase; Base excision
104
105 repair
106
107
108
109
110
111
112
113
114
115
116
117
118

1. Introduction

Uracil bases in DNA are created by deamination of cytosine or by dUMP incorporation catalyzed by a DNA polymerase. It is estimated that the deamination of cytosine leads to up to 500 uracil residues in a single human cell each day [1-2]. The rate of deamination of cytosine is greatly enhanced at elevated temperatures [3], and thus hyperthermophilic organisms that live at temperatures above 80°C are facing a serious threat caused by cytosine deamination. Since uracil has a strong ability to form mismatch with adenine (A), a G-C base pair would be subsequently converted to an A-T base pair if DNA replication occurs before the uracil is repaired [4], potentially leading to mutations in the genome. However, the estimated spontaneous mutation rates in hyperthermophilic bacteria and archaea are similar to those observed in *Escherichia coli* [5-6], suggesting that hyperthermophilic microorganisms are more efficient in repairing hydrolytic and oxidative damage to DNA bases [7]. Increased GC to AT mutations by replicating uracil that originates from deamination of cytosine are detrimental to the cells, which may cause genome instability or even cancer occurrence. In response, cells have evolved a base excision repair (BER) pathway to counteract potential mutations generated by uracil replication. Uracil DNA glycosylase (UDG) is the first BER enzyme to remove uracil from DNA, which leads to apurinic/apyrimidinic (AP) site. The generated AP site is subsequently repaired by different enzymes: AP endonuclease, DNA deoxyribosephosphodiesterase, DNA polymerase and DNA ligase [8].

UDGs are ubiquitous in bacteria, archaea, eukarya and viruses, and can remove

178
179
180
181 uracil from DNA through hydrolyzing their glycosyl bonds. Based on sequence
182
183 similarity, UDGs have currently been classified into six families [9]. Family 1 UDGs
184
185 have been well studied in *Escherichia coli* and human [10-15], and can remove uracil
186
187 base efficiently from ssDNA and from ds DNA. Mismatch-specific DNA glycosylases
188
189 and single-stranded specific monofunctional UDGs form families 2 and 3,
190
191 respectively [16-17]. Family 4 UDGs are exclusively observed in the
192
193 hyperthermophilic microorganisms and possess a 4Fe - 4S cluster [18-20]. Family 5
194
195 UDGs have broad substrate specificity, but lack a polar residue at the active-site motif
196
197 [20, 21-22]. Last, family 6 UDGs can cleave hypoxanthine instead of uracil, and
198
199 thus belong to hypoxanthine DNA glycosylase family [23].
200
201
202
203
204

205 The *Archaeoglobus fulgidus* UDG is the first enzyme to be identified and
206
207 characterized from archaea [24]. Currently, another eight UDG homologues from
208
209 hyperthermophilic archaea have been reported from *Pyrobaculum aerophilum* [25],
210
211 *Pyrococcus furiosus* [26-27], *Methanococcus jannaschii* [28], *Aeropyrum pernix* [29],
212
213 *Sulfolobus solfataricus* [30], *Sulfolobus tokodaii* [31-32], *Sulfolobus acidocaldarius*
214
215 [33], and *Thermoplasma acidophilum* [34]. The *A. fulgidus* UDG is the most
216
217 thermostable, retaining activity even after 1.5 hr of heating at 95°C [24]. It is able to
218
219 remove uracil from dsDNA containing a U/A or U/G base pair, as well as from
220
221 ssDNA. This enzyme is inhibited by apurinic sites. *M. jannaschii* UDG possesses the
222
223 helix-hairpin-helix and [4Fe-4S]-binding cluster to recognize and bind uracil, and can
224
225 efficiently cleave uracil in ssDNA and dsDNA and 8-oxoG in DNA [28]. The crystal
226
227 structure of *S. tokodaii* UDG shows that residues Leu169, Tyr170 and Asn171 in the
228
229
230
231
232
233
234
235
236

237
238
239
240 leucine-intercalation loop of the enzyme play important roles in uracil-binding [32]. *T.*
241
242 *acidophilum* UDG is involved in mediating the transfer from long patch repair to short
243
244 patch repair [34]. Several studies on archaeal UDGs have shown that they can interact
245
246 with PCNA (proliferating cell nuclear antigen) [26, 30, 35], which is an important
247
248 component in DNA replication and repair in and eukarya and archaea [36].
249
250 Furthermore, due to their thermostability, they have been shown to have potential
251
252 applications to enhance PCR yield or to help in jump starting the reactions [29].
253
254
255
256

257 *Thermococcus* is an important branch of euryarchaea comprising more than 40
258
259 described species, which mostly thrive in the hottest deep-sea hydrothermal vent
260
261 systems. Thus, similar to other hyperthermophilic archaea and bacteria,
262
263 *Thermococcus* is also facing severe challenges due to increased deamination of
264
265 cytosine dependent on high-temperature [37]. It is expected that the
266
267 hyperthermophilic *Thermococcus* would have evolved a repair pathway to counteract
268
269 the mutational effects of cytosine deamination in order to maintain their genome
270
271 stability. *Thermococcus barophilus* is one of the most extreme member of the
272
273 *Thermococcus* genus, being hyperthermophilic, piezophilic and capable of
274
275 auxotrophic growth on carbon monoxide. Strain Ch5 was isolated from a deep-sea
276
277 hydrothermal field of the Mid-Atlantic Ridge (Logachev field chimney, 3,020 m
278
279 depth) [38]. *T. barophilus* Ch5 has a pressure optimum of 40 MPa and a temperature
280
281 optimum of 85°C [39]. The completed genome sequence of *T. barophilus* Ch5 shows
282
283 that this strain possesses two uracil DNA glycosylases (Tba UDGs) [40]. In this study,
284
285 we cloned one of the Tba UDGs gene, purified and characterized its product. In
286
287
288
289
290
291
292
293
294
295

296
297
298
299 addition, we probed the mechanism of cleaving uracil by Tba UDG through
300
301 mutational analysis. We report here that Tba UDG is a thermostable enzyme cleaving
302
303 specifically uracil-containing DNA, and its residues E118 and N159 are important for
304
305 its catalysis. To the best of our knowledge, this is the first report of the biochemical
306
307 characterization and mutational studies of a thermostable UDG from *Thermococcus*
308
309 species.
310
311
312

313 **2. Materials and methods**

314 *2.1. Cloning, expression and purification of Tba UDG*

315
316 The Tba UDG in this work is encoded by gene TBCH5v1_0629 (GenBank
317
318 accession number: CP013050.1). The genomic DNA of *T. barophilus* Ch5 was
319
320 extracted as described by Oger et al. [40] and then used as a template to amplify the
321
322 gene TBCH5v1_0629 using the Phusion DNA polymerase (Thermo Scientific,
323
324 Waltham, MA, USA) and the two primers (Tba UDG F and Tba UDG R, Table 1).
325
326 The amplified DNA product was inserted into the vector pET-30a (+) (Novagen,
327
328 Merck, Darmstadt, Germany). The recombinant plasmid harboring a sequence
329
330 encoded a 6 × His-tag at the C-terminus of Tba UDG was sequenced to verify the
331
332 accuracy of the sequence of the enzyme gene and transformed into *E. coli* BL21
333
334 (DE3) cells (Transgene, Beijing, China) for protein expression.
335
336
337
338
339
340
341

342 The expression strain *E. coli* harboring the recombinant plasmid was cultured
343
344 into LB medium with 100 µg/mL kanamycin at 37°C until an OD₆₀₀ of 0.6, at which
345
346 point protein expression was induced with isopropyl thiogalactoside (IPTG) at a final
347
348 concentration of 0.1 mM. The culture was further shaken for 10 hr at room
349
350
351
352
353
354

355
356
357
358 temperature until it reached an OD₆₀₀ of 1.1.
359

360 The cells were harvested by centrifugation (5,000 × g for 20 min at room
361 temperature). The resultant pellet was resuspended with Ni column buffer A (20 mM
362 Tris-HCl pH 8.0, 1 mM dithiothreitol (DTT), 500 mM NaCl, 50 mM imidazole and
363 10% glycerol). The cells were disrupted by sonication into an ice bath. After
364 centrifugation (16,000 × g for 30 min at 4°C), the supernatant was collected into a 50
365 mL tube and heated at 70°C for 20 min. The non-thermostable *E. coli* proteins were
366 almost removed by centrifugation (16,000 × g for 30 min at 4°C). The resulting
367 supernatant was loaded to a HisTrap FF column (GE Healthcare, Uppsala, Sweden)
368 and purified with NCG™ Chromatography System (Bio-Rad, Hercules, CA, USA). A
369 linear gradient of 50–500 mM imidazole was used to elute the Tba UDG protein.
370 Fractions of Tba UDG protein were harvested and run by a 12% sodium dodecyl
371 sulfate-polyacrylamide gel electrophoresis. The Tba UDG protein was stained and
372 visualized by Coomassie-staining method. Finally, the purified Tba UDG protein was
373 dialyzed in a storage buffer containing 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM
374 DTT and 50% glycerol, and was stored at –80°C. The protein concentration was
375 determined using the Bradford Protein Assay Kit (Bio-Rad).
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397

398 399 2.2. Construction, overexpression and purification of the Tba UDG mutants 400

401 Using the wild-type plasmid harboring the Tba UDG gene as a template, the site-
402 directed mutagenesis was performed by a SDM Kit to construct E118A and N159A
403 mutants, according to its manual instruction. Note that residues E118 and N159 in Tba
404 UDG are located in the conserved Motif B and Motif D, respectively (Fig. 1). The
405
406
407
408
409
410
411
412
413

414 sequences of mutagenic primers are listed in Table 1. The mutant plasmids were
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472

sequences of mutagenic primers are listed in Table 1. The mutant plasmids were verified by sequencing. Similar to the wild-type protein, the Tba UDG mutant proteins were overexpressed, purified and quantitated.

2.3. DNA substrate

Normal and uracil-containing deoxyoligonucleotides were synthesized by Sangon Biotech company, China. The sequences of these deoxyoligonucleotides are shown in Table 2. The Cy3-labeled deoxyoligonucleotide duplexes shown in Table 3 were prepared by annealing the Cy3-labeled deoxyoligonucleotides to the complementary deoxyoligonucleotides in a 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 at least 4 hours to room temperature.

2.4. Glycosylase assays

The standard assays of Tba UDG activity were carried out in the reactions (10 μ L) which contained 20 mM Tris-HCl pH 8.0, 5 mM DTT, 50 mM NaCl, 1 mM EDTA, 8% glycerol, 200 nM DNA, wild-type or mutant Tba UDG with varied concentrations. The reactions were performed at 75°C for 10 min for ssDNA cleavage and at 65°C for 10 min for dsDNA cleavage. 1 μ L 500 mM NaOH and 9 μ L formamide-EDTA (98% formamide and 20 mM EDTA) were added to stop the reactions. The reaction products were heated at 95°C for 5 min and chilled rapidly on ice for 5 min, and then loaded onto a denaturing 15% polyacrylamide gel with 8M urea. After electrophoresis, the gels were scanned and the Cy3-labeled DNA was visualized with a Molecular Image analyser (PharosFx System, Bio-Rad). The

473
474
475
476 ImageQuant software was used for the quantitative analysis. All experiments of the
477
478 glycosylase assays were replicated three times.
479
480

481 *2.5. Biochemical characterization assays*

482

483 The optimal temperature of Tba UDG to cleave DNA in the reactions (10 μ L)
484
485 contained 800 nM enzyme and 200 nM Cy3-labeled ssDNA with uracil as a target.
486
487 The reactions were performed at 35, 45, 55, 65, 75, 85 and 95°C for uracil-containing
488
489 ssDNA for 10 min.
490
491

492
493 To examine the thermostability of the enzyme, Tba UDG was heated at 80, 85,
494
495 90, 95 and 100°C for 30 min. The activity of the heated Tba UDG protein were
496
497 investigated under the same conditions but using 1 μ M of the heated enzyme protein.
498
499 Samples were treated as described above.
500
501

502 The effect of pH on the Tba UDG activity was evaluated by examining DNA
503
504 cleavage in similar 10 μ L reactions at 75°C under pHs ranging from 4.0 to 11.0 using
505
506 1 μ M of the enzyme protein. The varied pHs were adjusted with five different buffers
507
508 (all at 20 mM concentrations): acetate-sodium acetate (pH 4.0 and pH 5.0), sodium
509
510 phosphate-NaOH (pH 6.0 and pH 7.0), Tris-HCl (pH 8.0), Gly-NaOH (pH 9.0), and
511
512 NaHCO₃-NaOH (pH 10.0 and pH 11.0).
513
514
515
516

517 The effect of divalent metal ions on Tba UDG activity was investigated by
518
519 adding of 2 mM of Mg²⁺, Mn²⁺, Ca²⁺, Zn²⁺ or Cu²⁺ (analytical purity) to the reaction
520
521 mixture. Assays were performed at 75°C with 250 nM of the enzyme. Samples were
522
523 treated as described above.
524
525

526 To evaluate the effect of salinity on Tba UDG activity, glycosylase assays were
527
528
529
530
531

532
533
534
535 performed in the presence of various NaCl concentrations ranging from 50 to 1,000
536 mM using 1 μ M of the enzyme. Samples were treated as described above.
537
538

539 540 2.6. *Glycosylase single-turnover assays*

541
542 The reaction mixtures containing 800 nM wild-type or mutant Tba UDG and 200
543 nM DNA substrates were incubated at 65°C for various times. Samples were treated
544 as described above. Data from the DNA cleavage experiments under single-turnover
545 conditions were fit to a single-exponential decay equation:
546
547
548
549

$$550 \quad [\text{Product}] = A \exp(-k_{\text{endo}} t)$$

551
552 where A and k_{endo} represent the reaction amplitude and observed DNA cleavage
553 rate, respectively.
554
555
556

557 558 2.7. *Substrate specificity*

559
560 To investigate the substrate specificity of the enzyme, we employed normal
561 ssDNA and dsDNA, ssDNA with uracil and dsDNA with U (U/T, U/C, U/G or U/A),
562 and dsDNA with a mismatch (G/T) as the substrates to perform the glycosylase assays
563 at 65°C for 10 min using 800 nM enzyme. Samples were treated as described above.
564
565
566
567
568
569
570

571 2.8. *DNA-binding Assays*

572
573 Electrophoresis mobility shift assays (EMSA) were performed by incubating the
574 wild-type or mutant Tba UDG with uracil-containing ssDNA and dsDNA in a DNA
575 binding buffer (10 μ L) containing 20 mM Tris-HCl pH 8.0, 5 mM DTT, 8% glycerol,
576 200 nM DNA and the wild-type or mutant Tba UDG with varied concentrations at
577 25°C for 10 min. The samples were electrophoresed on a 4% native polyacrylamide
578 gel in 0.1 \times TBE (Tris-borate-EDTA) buffer. After electrophoresis, the gels were
579
580
581
582
583
584
585
586
587
588
589
590

591 scanned and Cy3-labeled DNA was visualized with a Molecular Image analyser (Bio-
592 Rad). ImageQuant software was used for the quantitative analysis.
593
594
595
596
597

598 **3. Results**

600 *3.1. Tba UDG is a thermostable uracil DNA glycosylase*

601
602 The alignment result of partial amino acid sequences of UDGs from
603 hyperthermophilic archaea and bacteria shows that Tba UDG possesses six conserved
604 motifs (A-F) that are characteristics of family 4 UDGs (Fig. 1A), suggesting that this
605 enzyme belongs to family 4 UDGs. On the other hand, Motif B and Motif F in Tba
606 UDG are conserved in all six family UDGs (Fig. 1B). Tba UDG displays 21%, 19%,
607 22%, 22%, 18%, 20%, 19%, and 21% similarities to those of from *P. furiosus*, *P.*
608 *horikoshii*, *A. fulgidus*, *P. aerophilum* UDGA, *A. pernix*, *S. solfataricus*, *S. tokodaii*
609 and *Thermotoga maritima*, respectively. The low similarity between Tba UDG and
610 other UDGs suggests that Tba UDG might be a novel glycosylase.
611
612
613
614
615
616
617
618
619
620
621
622
623
624

625 The Tba UDG gene from the hyperthermophilic archaeon *T. barophilus* Ch5 was
626 cloned into the pET-30a (+) expression vector, and expressed in *E. coli* BL21(DE3).
627 The recombinant Tba UDG protein was successfully expressed as a His-tag fusion
628 protein (Fig. 1C). By means of sonication, heat treatment (70°C for 20 min) and
629 purification by affinity chromatography with a Ni column, we purified the Tba UDG
630 protein (~27 kDa) (Fig. 1C).
631
632
633
634
635
636
637
638

639 We used the normal, uracil-containing ssDNA and dsDNA as the substrates to
640 investigate DNA cleavage by Tba UDG at 65°C. Using normal ssDNA and dsDNA as
641 the substrates, no product was formed by the enzyme, however, the cleaved product of
642
643
644
645
646
647
648
649

650
651
652
653 the enzyme was observed in the presence of uracil-containing ssDNA and dsDNA
654
655 (Fig. 1D). The results showed that Tba UDG is a thermostable glycosylase, capable of
656
657 removing uracil from ssDNA and dsDNA at 65°C.
658

659
660 The heating treatment (70°C for 20 min) during the purification of Tba UDG
661
662 protein can denature most of *E. coli* proteins (Fig. 1); however, there was a slight
663
664 possibility of *E. coli* UDG contamination, which would interfere with our results. To
665
666 test this possibility, we used cell extracts made from cells expressing the empty pET-
667
668 30a (+) vector. We could detect no cleavage product when using this heated
669
670 supernatant produced from the empty vector (data not shown), thus ruling out the
671
672 possibility of an *E. coli* UDG contamination during purification of Tba UDG. Overall,
673
674 our results suggest that Tba UDG can cleave uracil-containing DNA at high
675
676 temperature.
677
678
679
680

681 682 3.2. Biochemical characterization of Tba UDG 683

684 Since *T. barophilus* Ch5 thrives at high temperature (85°C) and we could show
685
686 that Tba UDG can cleave uracil-containing DNA at 75°C, we first investigated the
687
688 optimal temperature for the enzyme to cleave uracil-containing DNA by using the
689
690 uracil-containing ssDNA as the substrate. The cleavage percentage of Tba UDG
691
692 increased from 54% to 98% when increasing reaction temperature from 35 to 75°C
693
694 (Fig. 2A). Interestingly, even at the lowest tested temperatures, e.g. 35°C and 45°C,
695
696 Tba UDG displayed a significant activity, with 54% and 92% cleavage efficiencies
697
698
699
700
701 (Fig. 2A), respectively. At the temperatures higher than 75°C, the efficiency
702
703
704
705
706
707
708

709
710
711
712 decreased to reach 24% at 95°C (Fig. 2A), suggesting that the optimal activity of the
713
714 enzyme to cleave uracil-containing ssDNA is between 55°C and 75°C.
715

716
717 To further investigate the thermostability of the enzyme, we heated Tba UDG at
718
719 various temperatures prior to activity assessments. When heated at 80°C for 30 min,
720
721 Tba UDG retained about 90% of cleavage activity (Fig. 2B). The enzyme activity
722
723 rapidly decreased at higher temperatures to reach 23% for 85°C and no remaining
724
725 activity above 90°C (Fig. 2B). Overall, these observations suggest that Tba UDG is
726
727 thermostable.
728
729

730
731 We examined the impact of pH on the endonuclease activity of Tba UDG over a
732
733 wide pH range from 4.0 to 11.0 in the standard DNA cleavage reactions. No activity
734
735 could be detected at the highest pHs (pH=10 and pH=11, Fig. 2C). By contrast, we
736
737 could detect significant cleavage activity even at the lowest pH (pH=4, activity =
738
739 69%). The maximal activity was observed for pH ranging from 5 to 8, ranging from
740
741 96% to 91%, respectively (Fig. 2C). At pH 9.0, Tba UDG retained 43% cleavage
742
743 efficiency. These results suggest that Tba UDG cannot effectively cleave uracil-
744
745 containing DNA at high pHs (pH>10.0) and that the optimal pH for this enzyme to
746
747 cleave uracil-containing DNA was between 5.0 and 7.0.
748
749
750
751

752
753 To evaluate the effects of various divalent metal ions (Mg^{2+} , Mn^{2+} , Ca^{2+} , Zn^{2+}
754
755 and Cu^{2+}) on the DNA cleavage activity of Tba UDG, we reduced the concentration
756
757 of the enzyme in the reactions. In the absence of a divalent ion and in the presence of
758
759 EDTA, Tba UDG displayed about 70% cleavage activity (Fig. 2D), suggesting that a
760
761 divalent metal ion is not required for the enzyme to cleave uracil-containing DNA.
762
763
764
765
766
767

768
769
770
771 We observed no inhibition of Ca^{2+} or Mg^{2+} on the activity of Tba UDG, with cleavage
772
773 efficiencies ca. 77% in the presence of both ions (Fig. 2D). Two metals, e.g. Zn^{2+} or
774
775 Cu^{2+} , were found to totally inhibit the enzyme. Last, the activity of the enzyme was
776
777 partially inhibited in the presence of Mn^{2+} . Overall, our results suggest that a divalent
778
779 metal ion is not needed for Tba UDG to effectively cleave uracil-containing DNA.
780
781

782
783 To uncover the effect of salinity on the Tba UDG activity, we added NaCl with
784
785 various concentrations in the DNA cleavage reactions. Under the standard conditions
786
787 in the absence of NaCl, Tba UDG cleaved almost completely DNA substrate with
788
789 97% of cleavage efficiency (Fig. 2E). No impact of salinity was observed below 200
790
791 mM, at which salinity Tba UDG retained 96% of cleavage efficiency, which is similar
792
793 to the control reactions (Fig. 2E). However, only 13% cleavage efficiency of Tba
794
795 UDG activity was observed in the presence of 400 mM NaCl (Fig. 2E). No cleaved
796
797 DNA product was observed at NaCl concentrations from 600 to 1000 mM (Fig. 2E).
798
799 These results show that the Tba UDG is a salt-tolerant enzyme, inhibited only by high
800
801 NaCl concentrations (>400 mM).
802
803

804 805 806 807 *3.3. Substrate specificity of Tba UDG*

808
809 To evaluate the substrate specificity of the enzyme, we used the mismatched
810
811 DNA (G/T), four mismatched DNA with uracil, and ssDNA with uracil as the
812
813 substrates to examine the enzyme activity. As shown in Fig. 3A, the cleavage
814
815 efficiencies of Tba UDG were 95%, 88%, 30% and 85% when using the mismatched
816
817 DNA with U/G, U/C, U/A and U/T as the substrates, respectively. These results
818
819 suggest Tba UDG exhibit various cleavage efficiencies on uracil-containing dsDNA.
820
821
822
823
824
825
826

827
828
829
830 Furthermore, no cleavage product was found when using mismatched dsDNA (G/T)
831
832 (Fig. 3B). The cleavage efficiency of the enzyme was 93% when using uracil-
833
834 containing ssDNA as the substrate, which is close to that of using uracil-containing
835
836 dsDNA (U/G). Thus, Tba UDG has a preference for substrates with the order from
837
838 high to low: $U \approx U/G > U/T \approx U/C > U/A$.
839

840 841 842 *3.4. Kinetics of DNA cleavage by Tba UDG* 843

844 Here, we carried out time course of DNA cleavage activity of Tba UDG under
845
846 the optimal reaction condition as described above. As the reaction time extended,
847
848 DNA cleavage product of Tba UDG was gradually enhanced until the uracil-
849
850 containing ssDNA (Fig. 4A) and dsDNA (Fig. 4B) were almost cleaved. When
851
852 reaction time was 10 min, the percent of Tba UDG for cleaving uracil-containing
853
854 ssDNA and dsDNA reached 97% and 93%, respectively. These observations suggest
855
856 that the enzyme has a strong activity for cleaving uracil-containing DNA at high
857
858 temperature.
859
860
861
862

863 The molar amount of remaining DNA substrate in the DNA cleavage reactions
864
865 catalyzed by Tba UDG was plotted as a function of reaction time (Fig. 4C), and the
866
867 data were fit to the single-exponential decay equation to yield k_{exo} . The k_{exo} values
868
869 are $0.25 \pm 0.03 \text{ min}^{-1}$ and $0.31 \pm 0.04 \text{ min}^{-1}$ for uracil-containing ssDNA and dsDNA,
870
871 respectively. Therefore, Tba UDG displays similar rates for cleaving uracil-containing
872
873 ssDNA and dsDNA.
874
875
876
877

878 879 880 *3.5. Mutational analysis of Tba UDG* 881

882 As shown in Fig. 5A, the crystal structure of *S. tokodaii* UDG shows that
883
884
885

886
887
888
889 residues Glu42, His164, Asn82, Phe55 and Glu48 might be key for uracil recognition
890
891 [32]. Note that there is no a corresponding amino acid residue in Tba UDG for residue
892
893 Glu48 in *S. tokodaii*, and residues Glu42, His164, Asn82 and Phe55 in *S. tokodaii*
894
895 UDG correspond to residues Glu118, His216, Asn159 and Tyr127 in Tba UDG.
896
897 Sequence comparison shows that these conserved residues Glu118, His216, Asn159
898
899 and Tyr127 in Tba UDG are located in Motif B, C, D and F (Fig. 1A), respectively.
900
901 To investigate the function of these residues of Tba UDG, we mutated two of these
902
903 residues to alanine. The purification profiles of the Tba UDG E118A and N159A
904
905 mutants are shown in Fig. 5B.
906
907
908
909

910
911 In the control reaction with the wild-type Tba UDG, the enzyme can effectively
912
913 cleave the uracil-containing ssDNA (Fig. 6A) and dsDNA (Fig. 6D). When using 500
914
915 nM enzyme, the cleavage percent of Tba UDG reached approximate 90%. By
916
917 contrast, the E118A and N159A mutants had no cleaving activity, no matter what the
918
919 uracil-containing ssDNA or dsDNA was used (Figs. 6B, 6C, 6E and 6F). Therefore,
920
921 our data suggest that both mutations enable Tba UDG to abolish its activity, and thus
922
923 residues E118 and N159 in the enzyme play essential roles in cleaving uracil-
924
925 containing DNA.
926
927
928

929 930 *3.6. DNA-binding of the wild-type and mutant Tba UDGs*

931

932
933 To assess the effect of these two mutations on the affinity of the enzyme to
934
935 uracil-containing DNA, we investigated whether or not the wild-type and mutant Tba
936
937 UDGs binds to uracil-containing ssDNA or dsDNA by employing EMSA. As shown
938
939 in Figs. 7A and 7D, the free uracil-containing ssDNA and dsDNA were gradually
940
941
942
943
944

945
946
947
948 bound as increasing the enzyme concentrations. At $\geq 1,100$ nM Tba UDG, the uracil-
949
950 containing ssDNA and dsDNA was almost bound by the enzyme (Figs. 7A and 7D).
951

952
953 In contrast, the maximal binding percents of the E118A mutant only reached
954
955 35% for uracil-containing ssDNA and 25% for uracil-containing dsDNA even in the
956
957 presence of high enzyme concentration (1,500 nM) (Figs. 7B and 7D), suggesting that
958
959 the E118A mutant retains the compromised ability to bind to uracil-containing ssDNA
960
961 and dsDNA.
962
963

964
965 Compared with the wild-type protein, the N159A mutant displayed the clearly
966
967 reduced efficiencies for binding to uracil-containing ssDNA at lower concentration
968
969 ($< 1,100$ nM) (Fig. 7C). However, the binding percent of the N159A mutant was 92%
970
971 at 1,500 nM enzyme, which is similar to that of the wild-type protein. On the other
972
973 hand, the N159A mutant had the lower efficiency for binding to uracil-containing
974
975 dsDNA than the wild-type protein (Fig. 7F). Furthermore, the binding efficiencies of
976
977 the N159A mutant were higher than those of the E118A mutant. Thus, the residue
978
979 N159 in the Tba UDG is essential for catalysis, and also are involved in binding to
980
981 uracil.
982
983
984
985

986 987 **4. Discussion**

988
989 In this work, we characterized biochemically for the first time the thermostable
990
991 UDG from the hyperthermophilic archaeon *T. barophilus* Ch5, and revealed that Tba
992
993 UDG can specifically cleave uracil-containing DNA at temperatures ranging from 35
994
995 to 95°C. Similar to that of the closest homologue, *Pyrococcus furiosus* UDG, the
996
997 optimal temperature of the enzyme activity is 55–75°C, which is lower than that of
998
999

1004
1005
1006
1007 the optimal temperature of *T. barophilus* Ch5 [40]. Compared with *A. fulgidus* UDG
1008
1009 (80°C) [42], Tba UDG has an slightly lower optimal activity temperature. However,
1010
1011 the optimal temperature of Tba UDG activity is clearly higher than that of *S.*
1012
1013 *solfataricus* UDG [30], and similar to that of *A. pernix* UDG [29], *P. aerophilum*
1014
1015 UDG [41] and *T. maritima* UDG [43]. Thus, the optimal temperatures of archaeal
1016
1017 UDGs vary with hyperthermophilic organisms, which might be due to distinct living
1018
1019 environments. Furthermore, Tba UDG still retains the pronounced endonuclease
1020
1021 activity even when heated at 85°C for 30 min, suggesting that Tba UDG is a
1022
1023 thermostable endonuclease. Since the rate of deamination of cytosine increases with
1024
1025 temperature, significant amount of uracil might be generated at 85°C, which is the
1026
1027 optimal growth temperature of *T. barophilus* Ch5 [40]. Thus, the activity of Tba UDG
1028
1029 might be essential for mutation prevention in this organism in response to the known
1030
1031 mutagenic potential of uracil.
1032
1033
1034
1035
1036
1037

1038 DNA cleavage efficiencies by UDGs vary with pH. Tba UDG exhibits maximal
1039
1040 activity over a broad pH range from 5.0 to 7.0, which is close to that of the purified
1041
1042 recombinant *A. fulgidus* UDG that has a optimal pH 4.8 [42]. Interestingly, the native
1043
1044 *A. fulgidus* UDG displays maximal activity around pH 6.2 [42]. The difference pHs
1045
1046 for optimal activity of *A. fulgidus* UDG from native cells and expression cells might
1047
1048 be related to covalent modifications or accessory factors, or a different folding when
1049
1050 expressed in the native host. By contrast, the optimal pH for the *A. pernix* UDG
1051
1052 activity is estimated to be 8.0 to 10.5, with the highest removal of uracil from ssDNA
1053
1054 at pH 9.0 [29]. In addition, the optimal pH for Tba UDG strongly differs from that of
1055
1056
1057
1058
1059
1060
1061
1062

1063
1064
1065
1066 its closest homologue, that of *P. furiosus*, which is ca. pH 9 [27]. The rationale for this
1067
1068 strong divergences in optimal pHs is quite surprising since most of these Archaea
1069
1070 have near neutral intracellular pHs.
1071

1072
1073 The reported UDGs are independent on a divalent metal ion [27, 29], which is
1074
1075 also the case for Tba UDG. Similar to *A. pernix* UDG [29], Tba UDG is almost
1076
1077 inactive to cleave DNA in the presence of Zn²⁺ or Cu²⁺. However, both Mg²⁺ and
1078
1079 Mn²⁺ have no detectable effect on DNA cleavage of Tba UDG. However, Mn²⁺ shows
1080
1081 some inhibition of the activity of *A. pernix* UDG [29].
1082
1083

1084
1085 Tba UDG displays substrate specificity for cleaving DNA in the order:
1086
1087 U≈U/G>U/T≈U/C>U/A. By contrast, the *P. furiosus* UDG, as its closest homologue
1088
1089 of Tba UDG, removes uracil from various DNA substrates with the following order:
1090
1091 U/T≈U/C>U/G≈U/AP≈U/->U/U≈U/I≈U/A [27]. On the other hand, the *A. fulgidus*
1092
1093 UDG exhibits opposite base-dependent excision of uracil by the following order:
1094
1095 U>U/T>U/C=U/G=U/A [44]. Furthermore, the uracil-releasing activity of *M.*
1096
1097 *jannaschii* UDG is observed by the following order U/T>U/C>U/G>U/A [28]. In
1098
1099 addition, *A. pernix* UDG exhibits the uracil removal as follows:
1100
1101 U/C=U/G>U/T=U/AP=U/->U/U=U/I>U/A [29]. Moreover, *P. aerophilum* UDG
1102
1103 shows the substrate specificity by the order: G/U>A/U>ssU [25]. Overall, the
1104
1105 substrate specificities of archaeal UDGs vary with these organisms.
1106
1107
1108
1109
1110

1111
1112 On the other hand, Tba UDG has no detected activity on G/T mismatched DNA,
1113
1114 similar to *S. solfataricus* UDG [30]. By contrast, the *P. aerophilum* UDG h can cleave
1115
1116 normal mismatched DNA (G/T) and U/G [41]. Furthermore, the preferred substrates
1117
1118
1119
1120
1121

1122
1123
1124
1125 of *S. solfataricus* UDG and Tba UDG appear to be the G:U-containing double-
1126
1127 stranded oligonucleotide. In addition, both Tba UDG and *S. solfataricus* UDG can
1128
1129 cleave single-stranded DNA containing uracil; however, Tba UDG displays higher
1130
1131 efficiencies for this cleavage than *S. solfataricus* UDG.
1132
1133

1134 The uracil recognition mechanisms of several UDGs have been reported,
1135
1136 however, an complete understanding on how archaeal UDGs recognize and cleave
1137
1138 uracil-containing DNA remains elusive. The crystal structure of *S. tokodaii* UDG
1139
1140 suggest that this UDG has a special structure of the leucine-intercalation loop [32],
1141
1142 which is distinct from other UDGs, Further mutational analysis on the loop indicates
1143
1144 that Tyr170 in *S. tokodaii* UDG is critical for substrate DNA recognition and the
1145
1146 catalysis [32]. Mutational studies on the iron sulfur cluster loop motif in the *A.*
1147
1148 *fulgidus* uracil-DNA glycosylase suggest that the R86A, C85A and C101A mutants
1149
1150 exhibit reduced activity for uracil removal only within double-stranded DNA, while
1151
1152 the K100A mutant exhibits enhanced uracil excision activity [45]. In this work, we
1153
1154 did the mutational studies based on the *S. tokodaii* UDG structure by mutating
1155
1156 residues E118 and N159 in Tba UDG to alanine, which are the corresponding residues
1157
1158 E42 and N82 in *S. tokodaii* UDG. Our data show that residues E188 and N159 are key
1159
1160 for uracil recognition and removal, suggesting that the conserved motif B and Motif
1161
1162 D are important for uracil recognition and removal. Thus, our observations provide
1163
1164 new insight into understanding mechanism and function of archaeal UDGs.
1165
1166
1167
1168
1169
1170
1171
1172

1173 **5. Conclusion**

1174
1175 In summary, we present the biochemical characteristics and mechanism of the
1176
1177
1178
1179
1180

1181
1182
1183
1184 thermostable UDG from *T. barophilus* Ch5 in this work, which is first report on UDG
1185
1186 from *Thermococcus* species. The recombinant Tba UDG displays specifically uracil-
1187
1188 containing DNA cleavage activity with the highest efficiency at 55–75°C and with an
1189
1190 optimal pH of 5.0–7.0. A divalent metal ion is not required for the enzyme to cleave
1191
1192 uracil-containing DNA. Furthermore, the enzyme activity is inhibited by Zn²⁺ or Cu²⁺,
1193
1194 and high NaCl concentration. The enzyme exhibits the substrate specificity by the
1195
1196 order: U≈U/G>U/T≈U/C>U/G>U/A. Mutational studies suggest that residues E118
1197
1198 and N159 in Tba UDG are essential for uracil recognition and removal. Our work
1199
1200 provides a basis for determining the role of Tba UDG in the base excision repair
1201
1202 pathway for repairing potentially elevated uracils in *Thermococcus*.
1203
1204
1205
1206
1207
1208
1209

1210 **Acknowledgements**

1211
1212 This work was supported by the Academic Leader of Middle and Young People
1213
1214 of Yangzhou University Grant and Open Project of State Key Laboratory of Microbial
1215
1216 Metabolism, Shanghai Jiao Tong University (No. MMLKF18-05) to L.Z.; the practice
1217
1218 innovation training program for college students in Yangzhou University to H.S. (No.
1219
1220 XKYCX18_072); Open Project of Key Laboratory of Marine Medicine, Guangdong
1221
1222 Province and Key Laboratory of Tropical Marine Bio-resources and Ecology, Chinese
1223
1224 Academy of Sciences (2018011008) to L.M.
1225
1226
1227
1228
1229

1230 **Author contributions**

1240
1241
1242
1243 LZ, ZY and PO designed experiments; HS, QG, HH, HC, YX and LM performed
1244
1245 experiments; LZ and PO analyzed data; LZ, ZY and PO wrote and revised the paper.
1246
1247

1248 **References**

- 1251
1252 [1] T. Lindahl, Instability and decay of the primary structure of DNA, *Nature* 362
1253
1254 (1993) 709-715.
1255
1256 [2] J.C. Shen, W.M. Rideout 3rd, P.A. Jones, The rate of hydrolytic deamination of 5-
1257
1258 methylcytosine in double-stranded DNA, *Nucleic Acids Res* 22 (1994) 972-976.
1259
1260 [3] T. Lindahl, B. Nyberg, Heat-induced deamination of cytosine residues in
1261
1262 deoxyribonucleic acid, *Biochemistry* 13 (1974) 3405-3410.
1263
1264 [4] M. Hill-Perkins, M.D. Jones, P. Karran, Site-specific mutagenesis in vivo by
1265
1266 single methylated or deaminated purine bases, *Mutat Res* 162 (1986) 153-163.
1267
1268 [5] D.W. Grogan, G.T. Carver, J.W. Drake, Genetic fidelity under harsh conditions:
1269
1270 analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus*
1271
1272 *acidocaldarius*, *Proc Natl Acad Sci U S A* 98 (2001) 7928-7933.
1273
1274 [6] K.L. Jacobs, D.W. Grogan, Rates of spontaneous mutation in an archaeon from
1275
1276 geothermal environments, *J Bacteriol* 179 (1997) 3298-3303.
1277
1278 [7] A. Koulis, D.A. Cowan, L.H. Pearl, R. Savva, Uracil-DNA glycosylase activities
1279
1280 in hyperthermophilic micro-organisms, *FEMS Microbiol Lett* 143 (1996) 267-
1281
1282 271.
1283
1284 [8] S.S. Wallace, Base excision repair: a critical player in many games, *DNA Repair*
1285
1286 (Amst) 19 (2014) 14-26.
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298

- 1299
1300
1301
1302 [9] N. Schormann, R. Ricciardi, D. Chattopadhyay, Uracil-DNA glycosylases-
1303
1304 Structural and functional perspectives on an essential family of DNA repair
1305
1306 enzymes, *Protein Sci* 23 (2014) 1667-1685.
1307
1308
1309 [10] C.D. Mol, A.S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H.E. Krokan, J.A.
1310
1311 Tainer, Crystal structure and mutational analysis of human uracil-DNA
1312
1313 glycosylase: structural basis for specificity and catalysis, *Cell* 80 (1995) 869-878.
1314
1315
1316 [11] S.S. Parikh, C.D. Putnam, J.A. Tainer, Lessons learned from structural results on
1317
1318 uracil-DNA glycosylase, *Mutat Res* (2000) 183-199.
1319
1320
1321 [12] L.H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily,
1322
1323 *Mutat Res* (2000) 165-181.
1324
1325
1326 [13] R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, The structural basis of specific
1327
1328 base-excision repair by uracil-DNA glycosylase, *Nature* (1995) 487-493.
1329
1330
1331 [14] G. Slupphaug, C.D. Mol, B. Kavli, A.S. Arvai, H.E. Krokan, J.A. Tainer, A
1332
1333 nucleotide-flipping mechanism from the structure of human uracil-DNA
1334
1335 glycosylase bound to DNA, *Nature* 384 (1996) 87-92.
1336
1337
1338 [15] G.Y. Xiao, M. Tordova, J. Jagadeesh, A.C. Drohat, J.T. Stivers, G.L. Gilliland,
1339
1340 Crystal structure of *Escherichia coli* uracil DNA glycosylase and its complexes
1341
1342 with uracil and glycerol: Structure and glycosylase mechanism revisited, *Proteins*
1343
1344 35 (1999) 13-24.
1345
1346
1347 [16] T.E. Barrett, R. Savva, G. Panayotou, T. Barlow, T. Brown, J. Jiricny, L.H. Pearl,
1348
1349 Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch
1350
1351 recognition by complementary-strand interactions, *Cell* 92 (1998) 117-129.
1352
1353
1354
1355
1356
1357

- 1358
1359
1360
1361 [17] K.A. Haushalter, M.W. Todd Stukenberg, M.W. Kirschner, G.L. Verdine,
1362
1363 Identification of a new uracil-DNA glycosylase family by expression cloning
1364
1365 using synthetic inhibitors, *Curr Biol* 9 (1999) 174-185.
1366
1367
1368 [18] J.A. Hinks, M.C. Evans, Y. De Miguel, A.A. Sartori, J. Jiricny, L.H. Pearl, An
1369
1370 iron-sulfur cluster in the family 4 uracil-DNA glycosylases, *J Biol Chem* 277
1371
1372 (2002) 16936-16940.
1373
1374
1375 [19] J. Hoseki, A. Okamoto, R. Masui, T. Shibata, Y. Inoue, S. Yokoyama, S.
1376
1377 Kuramitsu, Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus*
1378
1379 *thermophilus* HB8, *J Mol Biol* 333 (2003) 515-526.
1380
1381
1382 [20] V. Starkuviene, H.J. Fritz, A novel type of uracil-DNA glycosylase mediating
1383
1384 repair of hydrolytic DNA damage in the extremely thermophilic eubacterium
1385
1386 *Thermus thermophilus*, *Nucleic Acids Res* 30 (2002) 2097-2102.
1387
1388
1389 [21] H. Kosaka, J. Hoseki, N. Nakagawa, S. Kuramitsu, R. Masui, Crystal structure of
1390
1391 family 5 uracil-DNA glycosylase bound to DNA, *J Mol Biol* 373 (2007) 839-
1392
1393 850.
1394
1395
1396 [22] A.A. Sartori, S. Fitz-Gibbon, H. Yang, J.H. Miller, J. Jiricny, A novel uracil-
1397
1398 DNA glycosylase with broad substrate specificity and an unusual active site,
1399
1400 *EMBO J* 21 (2002) 3182-3191.
1401
1402
1403 [23] H.W. Lee, B.N. Dominy, W. Cao, New family of deamination repair enzymes in
1404
1405 uracil-DNA glycosylase superfamily, *J Biol Chem* 286 (2011) 31282-31287.
1406
1407
1408 [24] M. Sandigursky, W.A. Franklin, Uracil-DNA glycosylase in the extreme
1409
1410 thermophile *Archaeoglobus fulgidus*, *J Biol Chem* 275 (2000) 19146-19149.
1411
1412
1413
1414
1415
1416

- 1417
1418
1419
1420 [25] A.A. Sartori, P. Schar, S. Fitz-Gibbon, J.H. Miller, J. Jiricny, Biochemical
1421
1422 characterization of uracil processing activities in the hyperthermophilic archaeon
1423
1424 *Pyrobaculum aerophilum*, J Biol Chem 276 (2001) 29979-29986.
1425
1426
1427 [26] S. Kiyonari, M. Uchimura, T. Shirai, Y. Ishino, Physical and functional
1428
1429 interactions between uracil-DNA glycosylase and proliferating cell nuclear
1430
1431 antigen from the euryarchaeon *Pyrococcus furiosus*, J Biol Chem 283 (2008)
1432
1433 24185-24193.
1434
1435
1436 [27] L.B. Lin, Y.F. Liu, X.P. Liu, J.H. Liu, Biochemical characterization of uracil-
1437
1438 DNA glycosylase from *Pyrococcus furiosus*, Chem Res Chinese U 28 (2012)
1439
1440 477-482.
1441
1442
1443 [28] J.H. Chung, E.K. Im, H.Y. Park, J.H. Kwon, S. Lee, J. Oh, K.C. Hwang, J.H.
1444
1445 Lee, Y. Jang, A novel uracil-DNA glycosylase family related to the helix-
1446
1447 hairpin-helix DNA glycosylase superfamily, Nucleic Acids Res 31 (2003) 2045-
1448
1449 2055.
1450
1451
1452 [29] X.P. Liu, J.H. Liu, Characterization of family IV UDG from *Aeropyrum pernix*
1453
1454 and its application in hot-start PCR by family B DNA polymerase, PLoS One 6
1455
1456 (2011) e27248.
1457
1458
1459 [30] I. Dionne, S.D. Bell, Characterization of an archaeal family 4 uracil DNA
1460
1461 glycosylase and its interaction with PCNA and chromatin proteins, Biochem J
1462
1463 387 (2005) 859-863.
1464
1465
1466 [31] A. Kawai, S. Higuchi, M. Tsunoda, K.T. Nakamura, S. Miyamoto, Purification,
1467
1468 crystallization and preliminary X-ray analysis of uracil-DNA glycosylase from
1469
1470
1471
1472
1473
1474
1475

- 1476
1477
1478
1479 *Sulfolobus tokodaii* strain 7, Acta Crystallogr F 68 (2012) 1102-1105.
1480
1481 [32] A. Kawai, S. Higuchi, M. Tsunoda, K.T. Nakamura, Y. Yamagata, S. Miyamoto,
1482
1483 Crystal structure of family 4 uracil-DNA glycosylase from *Sulfolobus tokodaii*
1484
1485 and a function of tyrosine 170 in DNA binding, FEBS Lett 589 (2015) 2675-
1486
1487 2682.
1488
1489
1490 [33] G.S. Yi, W.W. Wang, W.G. Cao, F.P. Wang, X.P. Liu, *Sulfolobus*
1491
1492 *acidocaldarius* UDG can remove dU from the RNA backbone: insight into the
1493
1494 specific recognition of uracil linked with deoxyribose, Genes (Basel) 8 (2017)
1495
1496 E38.
1497
1498
1499 [34] M.N. Moen, I. Knaevelsrud, G.T. Haugland, K. Grosvik, N.K. Birkeland, A.
1500
1501 Klungland, S. Bjelland, Uracil-DNA glycosylase of *Thermoplasma acidophilum*
1502
1503 directs long-patch base excision repair, which is promoted by deoxynucleoside
1504
1505 triphosphates and ATP/ADP, into short-patch repair, J Bacteriol 193 (2011)
1506
1507 4495-4508.
1508
1509
1510 [35] H. Yang, J.H. Chiang, S. Fitz-Gibbon, M. Lebel, A.A. Sartori, J. Jiricny, M.M.
1511
1512 Slupska, J.H. Miller, Direct interaction between uracil-DNA glycosylase and a
1513
1514 proliferating cell nuclear antigen homolog in the crenarchaeon *Pyrobaculum*
1515
1516 *aerophilum*, J Biol Chem 277 (2002) 22271-22278.
1517
1518
1519 [36] G.L. Moldovan, B. Pfander, S. Jentsch, PCNA, the maestro of the replication
1520
1521 fork, Cell 129 (2007) 665-679.
1522
1523
1524 [37] M. van Wolferen, M. Ajon, A.J.M. Driessen, S.V. Albers, How
1525
1526 hyperthermophiles adapt to change their lives: DNA exchange in extreme
1527
1528
1529
1530
1531
1532
1533
1534

- 1535
1536
1537
1538 conditions, *Extremophiles* 17 (2013) 545-563.
1539
- 1540 [38] Y.J. Kim, H.S. Lee, E.S. Kim, S.S. Bae, J.K. Lim, R. Matsumi, A.V. Lebedinsky,
1541
1542 T.G. Sokolova, D.A. Kozhevnikova, S.S. Cha, S.J. Kim, K.K. Kwon, T.
1543
1544 Imanaka, H. Atomi, E.A. Bonch-Osmolovskaya, J.H. Lee, S.G. Kang, Formate-
1545
1546 driven growth coupled with H₂ production, *Nature* 467 (2010) 352-355.
1547
1548
- 1549 [39] V.T. Marteinsson, J.L. Birrien, A.L. Reysenbach, M. Vernet, D. Marie, A.
1550
1551 Gambacorta, P. Messner, U.B. Sleytr, D. Prieur, *Thermococcus barophilus* sp.
1552
1553 nov., a new barophilic and hyperthermophilic archaeon isolated under high
1554
1555 hydrostatic pressure from a deep-sea hydrothermal vent, *Int J Syst Bacteriol* 49
1556
1557 (1999) 351-359.
1558
1559
- 1560 [40] P. Oger, T.G. Sokolova, D.A. Kozhevnikova, E.A. Taranov, P. Vannier, H.S.
1561
1562 Lee, K.K. Kwon, S.G. Kang, J.H. Lee, E.A. Bonch-Osmolovskaya, A.V.
1563
1564 Lebedinsky, Complete genome sequence of the hyperthermophilic and
1565
1566 piezophilic archaeon *Thermococcus barophilus* Ch5, capable of growth at the
1567
1568 expense of hydrogenogenesis from carbon monoxide and formate, *Genome*
1569
1570 *Announc* 4 (2016).
1571
1572
1573
1574
1575
- 1576 [41] H. Yang, S. Fitz-Gibbon, E.M. Marcotte, J.H. Tai, E.C. Hyman, J.H. Miller,
1577
1578 Characterization of a thermostable DNA glycosylase specific for U/G and T/G
1579
1580 mismatches from the hyperthermophilic archaeon *Pyrobaculum aerophilum*, *J*
1581
1582 *Bacteriol* 182 (2000) 1272-1279.
1583
1584
1585
- 1586 [42] I. Knaevelsrud, S. Kazazic, N.K. Birkeland, S. Bjelland, The pH optimum of
1587
1588 native uracil-DNA glycosylase of *Archaeoglobus fulgidus* compared to
1589
1590
1591
1592
1593

1594
1595
1596
1597 recombinant enzyme indicates adaption to cytosolic pH, *Acta Biochim Pol* 61
1598
1599 (2014) 393-395.
1600

1601
1602 [43] M. Sandigursky, A. Faje, W.A. Franklin, Characterization of the full length
1603 uracil-DNA glycosylase in the extreme thermophile *Thermotoga maritima*,
1604
1605 *Mutat Res* 485 (2001) 187-195.
1606
1607

1608
1609 [44] I. Knaevelsrud, P. Ruoff, H. Anensen, A. Klungland, S. Bjelland, N.K.
1610 Birkeland, Excision of uracil from DNA by the hyperthermophilic Afung protein
1611 is dependent on the opposite base and stimulated by heat-induced transition to a
1612 more open structure, *Mutat Res* 487 (2001) 173-190.
1613
1614
1615
1616
1617

1618
1619 [45] L.M. Engstrom, O.A. Partington, S.S. David, An iron-sulfur cluster loop motif in
1620 the *Archaeoglobus fulgidus* uracil-DNA glycosylase mediates efficient uracil
1621 recognition and removal, *Biochemistry* 51 (2012) 5187-5197.
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652

Figure legends

Fig. 1. Tba UDG can cleave uracil-containing ssDNA and dsDNA at high temperature. A. Partial amino acid alignment of UDGs from hyperthermophilic crenarchaea, euryarchaea and bacteria. Tba: *Thermococcus barophilus* (WP_056934618.1); Pfu: *Pyrococcus furiosus* (WP_011012532.1); Pho: *Pyrococcus horikoshii* (WP_048053599.1); Afu: *Archaeoglobus fulgidus* (GenBank: AIG99287.1); Pae: *Pyrobaculum aerophilum* (GenBank: AAL62921.1); Ape: *Aeropyrum pernix* (GenBank: BAA79385.2); Sso: *Sulfolobus solfataricus* (GenBank: AKA78326.1); Sto: *Sulfolobus tokodaii* (PDB: 4ZBY); Tma: *Thermotoga maritima* (PDB: 1L9G_A). B. The conserved Motif B and Motif F in six families of UDG. Family 1, Eco (*E. coli*) UDG (EMBL:J03725); Family 2, Human TDG (EMBL:U51166); Family 3, Human SUMG1 (EMBL: AF125182); Family 4, Tba UDG247 (NCBI reference sequence: WP_056934618.1); Family 5, *P. aerophilum* (Pae) UDGb (NP_559226); Family 6, Mba (*Methanosarcina barkeri*) HDG (YP_304295.1). C. Overexpression and purification of Tba UDG. M: Protein marker. D. DNA cleavage assays of Tba UDG. DNA cleavage reactions were performed by Tba UDG in the presence of normal and uracil-containing ssDNA and dsDNA at 65°C. CK: the reaction without the enzyme.

Fig. 2. Biochemical characterization of Tba UDG. A. The optimal temperature of the enzyme. B. The thermostability of the enzyme. C. The pH adaptation of the enzyme. D. Effects of divalent metal ions on the enzyme activity. E. Effect of NaCl on the enzyme activity. Reaction products were detected by electrophoresis through running

1712
1713
1714
1715 a 15% denaturing PAGE. CK: the reaction without the enzyme; CK1 in the panel B:
1716
1717 the reaction without the enzyme; CK2 in the panle B: the reaction with the unheated
1718
1719 enzyme.
1720

1721
1722
1723 **Fig. 3.** Substrate specificity of Tba UDG. DNA cleavage reactions of Tba UDG were
1724 performed using the uracil-containing ssDNA and dsDNA, and mismatched DNA
1725 (G/T) as the substrates. Reaction products were analyzed by electrophoresis through
1726 running a 15% denaturing PAGE. A. The substrates were ssDNA with U, and
1727 mismatched dsDNA with U/T, U/C, U/G, or U/A. B. The substrates were mismatched
1728 DNA (G/T). CK: the reaction without the enzyme.
1729
1730
1731
1732
1733
1734
1735
1736
1737

1738
1739 **Fig. 4.** Kinetic analysis of DNA cleavage of Tba UDG. DNA cleavage reactions by
1740 Tba UDG were performed under the optimal reaction condition at various time (10
1741 sec – 30 min). Reaction products were analyzed by electrophoresis through running a
1742 12% denaturing PAGE. A. Uracil-containing ss DNA cleavage; B. Uracil-containing
1743 ds DNA cleavage. CK: the reaction without the enzyme; C. Rate of DNA cleavage
1744 catalyzed by Tba UDG. By using the single-exponential decay equation, the amount
1745 of remaining substrate was plotted as a function of time to yeild the best fit (the solid
1746 lines). Tba UDG cleaved the uracil-containing ssDNA (\circ) and dsDNA (\square) at the rates
1747 of $0.25 \pm 0.03 \text{ min}^{-1}$ and $0.31 \pm 0.04 \text{ min}^{-1}$, respectively.
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759

1760
1761 **Fig. 5.** Possible uracil recognition mechanism of Tba UDG. A. Interactions between
1762 amino acid residues and uracil of Tba UDG. The residues E42, N82, H164 and F55 in
1763 *S. tokodaii* UDG that correspond to the residues E118, N159, H216 and Y127 in Tba
1764
1765
1766
1767
1768
1769
1770

1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829

UDG are depicted in blue, red, cyan and yellow sticks, respectively. The figure was adapted from the *S. tokodaii* UDG structure (PDB: 4zby) by Pymol [32]. Tba UDG residues are indicated in parentheses. The uracil is shown with dots. B. Purification of the wild-type, E118A and N159A Tba UDG mutant proteins. M: Protein marker.

Fig. 6. DNA cleavage assays of the wild-type and mutant Tba UDGs. DNA cleavage reactions of Tba UDG were performed using uracil-containing ssDNA and dsDNA as the substrates at 65°C for 10 min, respectively. Reaction products were analyzed by electrophoresis through running a 15% denaturing PAGE. A. Cleaving uracil-containing ssDNA by the wild-type; B. Cleaving uracil-containing ssDNA by the E118A mutant; C. Cleaving uracil-containing ssDNA by the N159A mutant; D. Cleaving uracil-containing dsDNA by the wild type; E. Cleaving uracil-containing dsDNA by the E118A mutant; F. Cleaving uracil-containing dsDNA by the N159A mutant. CK: the reaction without the enzyme.

Fig. 7. The binding assays of the wild-type and mutant Tba UDGs. The uracil-containing ssDNA and dsDNA (U:G) were employed as the substrates to examine the DNA-binding of the wild-type and mutant Tba UDGs. The wild-type and mutant Tba UDGs and DNA were incubated at 25°C for 10 min, and were run by electrophoresis on a 4% native polyacrylamide gel. A. Binding to uracil-containing ssDNA by the wild-type protein; B. Binding to uracil-containing ssDNA by the E118A mutant; C. Binding to uracil-containing ssDNA by the N159A mutant; D. Binding to uracil-containing dsDNA by the wild-type protein; E. Binding to uracil-containing dsDNA

1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888

by the E118A mutant; F. Binding to uracil-containing dsDNA by the N159A mutant.

CK: the binding assay without the enzyme.

Table 1 Sequences of the oligonucleotides used to clone the Tba UDG gene and construct its mutants

Name	Sequence (5'-3')
Tba UDG F	GGAATTCC <i>ATATGCTGCTGGAGTTTGAACGCC</i>
Tba UDG R	CCGCTCGAGTTTAGTAATATTTAAGCTTTTCC
E118A F	AAAGGTTGTTTTGGTCGGGG <u>GCGG</u> CTCCAGGAAGGAAAGGCT
E118A R	AGCCTTTCCTTCCTGGAGCC <u>GCC</u> CCCGACCAAACAACCTTT
N159A F	TTTTGTGTATATCACAGCTGTTGTAAAATGCAATC
N159A R	<u>AGCT</u> GTGATATACACAAAATCGGGGTTAATTCCGA

The italic nucleotides represent restriction sites.

The substitution bases are underlined.

1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

Table 2 Sequences of the oligonucleotides used in this work

Number	Sequence (5'-3')
1	CGAACTGCCTGGAATCCTGACGAC <u>U</u> TGTAGCGAACGATCACCTCA
2	CGAACTGCCTGGAATCCTGACGAC <u>C</u> TGTAGCGAACGATCACCTCA
3	CGAACTGCCTGGAATCCTGACGAC <u>G</u> TGTAGCGAACGATCACCTCA
4	TGAGGTGATCGTTCGCTACAG <u>G</u> TCGTCAGGATTCCAGGCAGTTTCG
5	TGAGGTGATCGTTCGCTACAC <u>G</u> TCGTCAGGATTCCAGGCAGTTTCG
6	TGAGGTGATCGTTCGCTACA <u>A</u> GTCGTCAGGATTCCAGGCAGTTTCG
7	TGAGGTGATCGTTCGCTACA <u>T</u> GTCGTCAGGATTCCAGGCAGTTTCG

The underlined base is used to prepare normal and uracil-containing dsDNA.

Table 3 DNA substrates prepared with the oligonucleotides in Table 2

Strand	labeling	Combination	Base pair
ssDNA	Cy3	1*	U/-
ssDNA	Cy3	2*	C/-
dsDNA	Cy3	1*+4	U/G
dsDNA	Cy3	1*+5	U/C
dsDNA	Cy3	1*+6	U/A
dsDNA	Cy3	1*+7	U/T
dsDNA	Cy3	2*+4	C/G
dsDNA	Cy3	3*+6	G/T

The symbol “*” indicates the labeled strand.

A

	<u>Motif A</u>	<u>Motif B</u>	<u>Motif C</u>	<u>Motif D</u>
Tba 101	GLPFANGVWGSKVVLVGEAP-G	---RKGCYTG	ICFYRDASGMLLRKALFSL-G	INPD-----FVYITNVVKCNPPENKLGVD--KREL
Pfu 29	PVP-GYGNDAKIMFVGEAP-GY	WEDQKGLPFVG	-----KAGKVLDELDDGI	GLTRE-----DVYITNVVKCRPPNRRDPT
Pho 23	PVP-GDGSYDTRKIMFVGEAP-GY	WEDQMGLPFVG	-----KAGKVLDELKLI	GLKRS-----EVYITNVVKCRPPNRRDPT
Afu 26	YVP-GVGNKAEIVFVGEAP-GR	EDLKGEPFVG	-----AAGKLLTEMLASI	GLRRE-----DVYITNVVKCRPPNRRDPT
Pae 25	AVP-GEGEMKLGVMIVGEAP-GA	SEDEAGRPFVG	-----AAGQLLTEALSRL	GVRRG-----DVYITNVVKCRPPNRRTPNR
Ape 28	AVP-GEGPGEAGVMVIVGEAP-GR	MEDRLGRPFVG	-----PAGKLLDLSLELA	GLSRG-----EVYITNVVKCRPPNRRDPRE
Sso 32	PVP-PNGQIDAEIVIVGLAPAG	NGGNRTGRMFTGD	-----ESSNNLANALYAV	GLSNQPFVSKDDGLKLFNVYITSAVKCAP
Sto 30	GYP-----KAEIMFVGEAP-GE	NEDKGRPFVG	-----AAGKLLTQMIKEIL	GLERD-----QVYITNVVKCRPPNRRDPE
Tma 30	VVV-GEGNLDTRIVFVGEAP-G	EEEDKTGRPFVG	-----RAGMLLTELRES	GIRRE-----DVYICNVVKCRPPNRRTP

Motif E

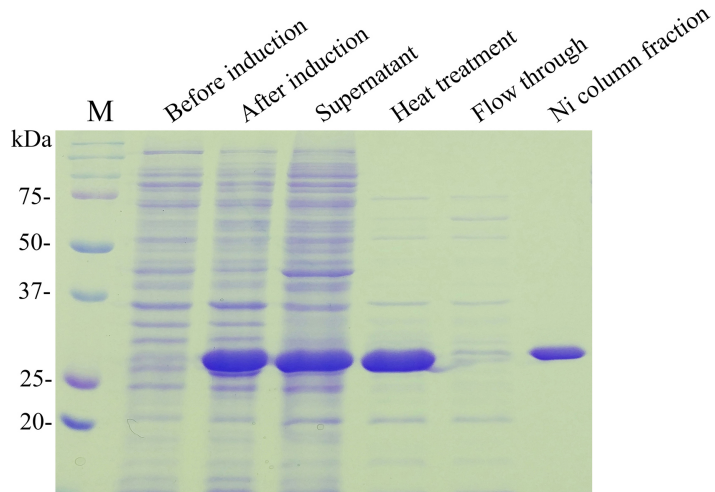
Motif F

Tba 179	SLLKKELEIFR-PKAI	FALGR	----	TAEKALKRVGFDA	-----	VYLR*	HPAWYVRRGLREPND	EMLEEY
Pfu 106	PYLDQQIDIIK-PKV	IVTLGR	----	HSTNYILKKFG	FDLEPISKIH	GKVF	KAKTLFGTLY	-IFPT
Pho 100	PYLDQQIDIIK-PKV	IVTLGR	----	FSTAYIMKKY	GFNVEPISKIH	GRVFE	ARTLFGKIY	-IVPM
Afu 103	NYLVRQLEAIR-PNV	IVCLGR	----	FAAQFIENL	FDLEFTTISR	VKGK	VEVERWGKVKV	-VIAI
Pae 102	PYLTIQQIGILK-PRR	IATLGL	----	ISAKALMELM	GRRAEKLGD	VKGK	CYQGR	IAGVQVE
Ape 105	PYLVEQISLIR-PR	LVIAGR	----	HAGRTLFR	LAGLRWPLAR	ARGR	VWRGR	IIGVVELL
Sso 123	VFLVEEVRI	LKNTKVI	TALGR	IAWDSL	LIYVFKK	IGYN	VNPVRF	YHGCALV
Sto 104	PYLDQQIDIIIM-PK	IIVTLGR	----	HSTKYIFSKM	GENFSSITK	VRGK	SYVWKYKEKE	-IIVFPT
Tma 107	HFLLAQIEIIN-PD	VIVALGA	----	TALSFFV	---	DGKKV	SITKVRGN	PID--

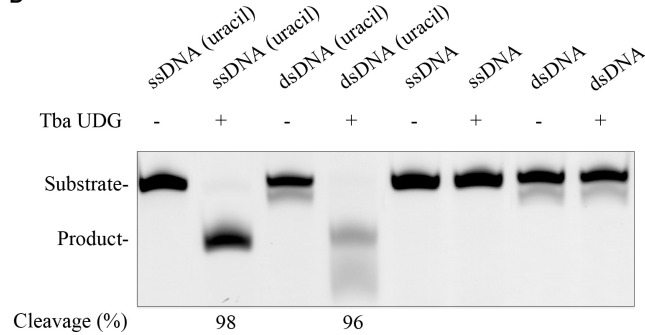
B

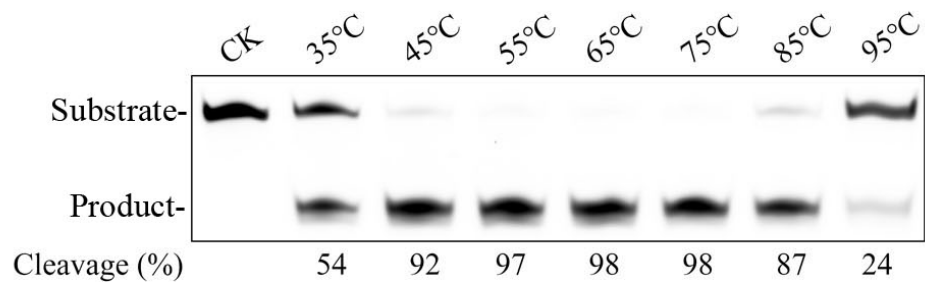
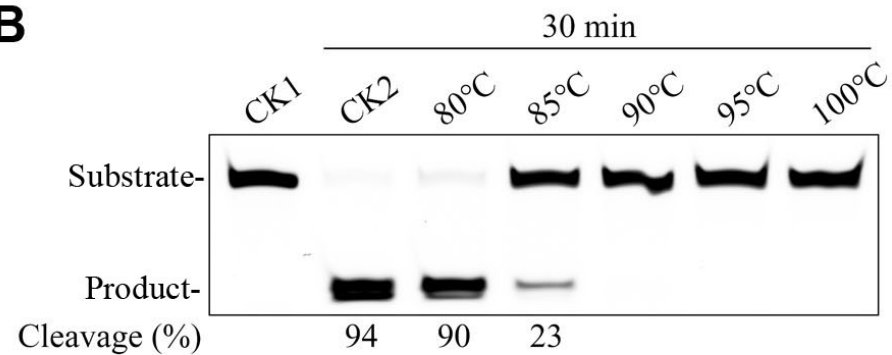
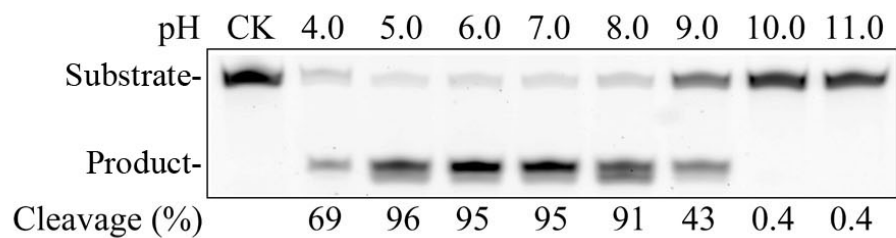
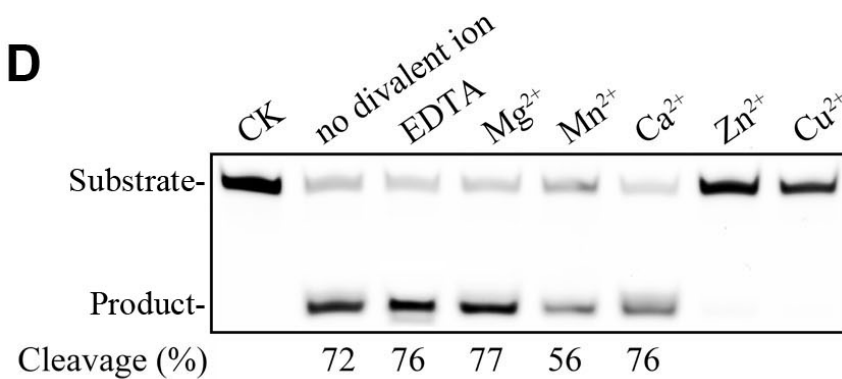
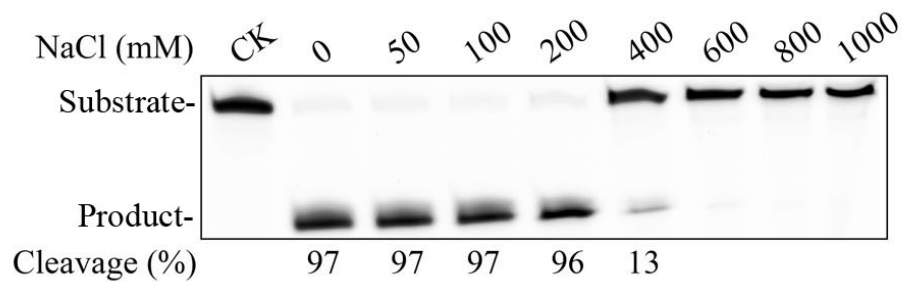
		<u>Motif B</u>	<u>Motif F</u>
Family 4	Tba UDG	113 V V L V G E A P G	216 H P A W Y V R
Family 1	Eco UDG	58 V V I L G Q D P Y	187 H P S P L S A
Family 2	Human TDG	134 I V I I G I N P G	269 M P S S S A R
Family 3	Human SMUG1	79 V L F L G M N P G	239 H P S P R N P
Family 5	Pae UDGb	62 V M V V G L A P G	196 H P S P L N V
Family 6	Mba HDG	42 W R L L G S I I G	136 S S S G A N R

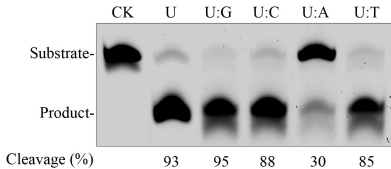
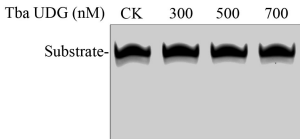
C

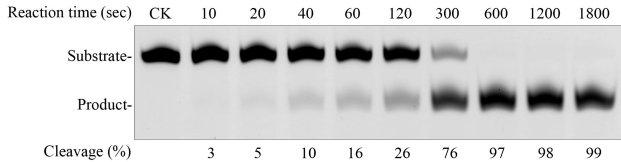
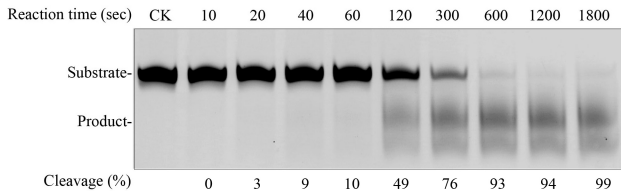
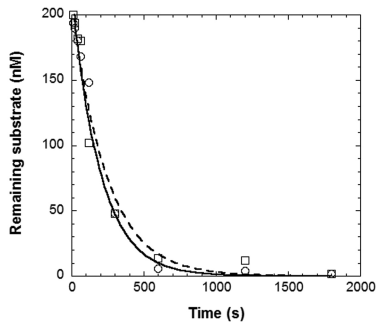


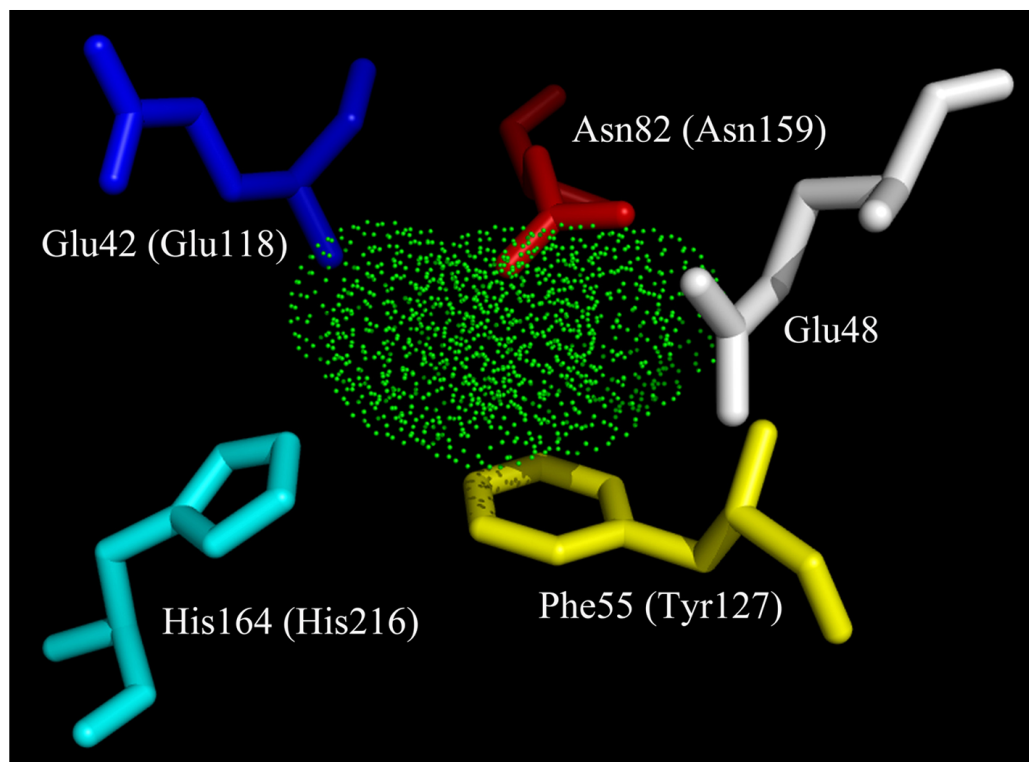
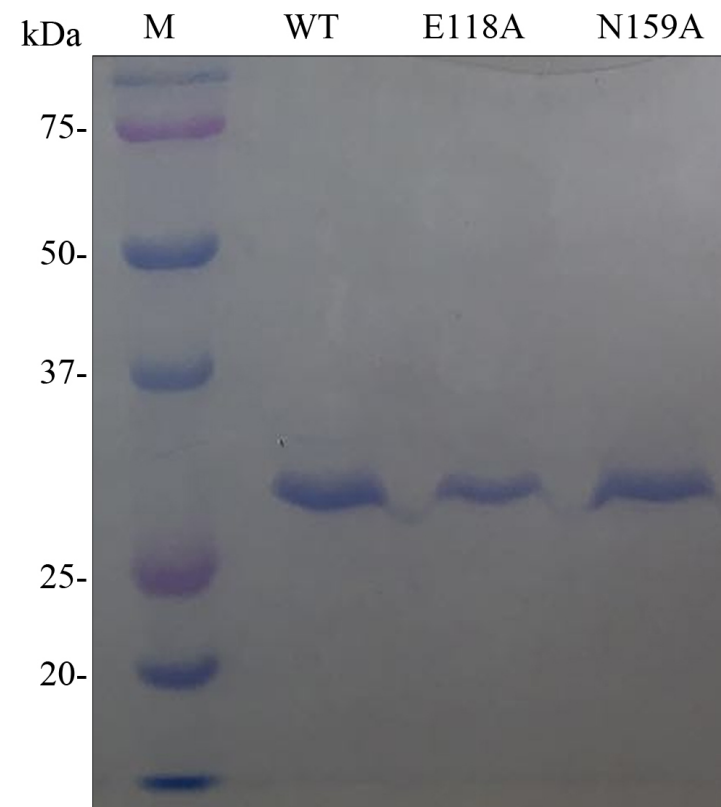
D

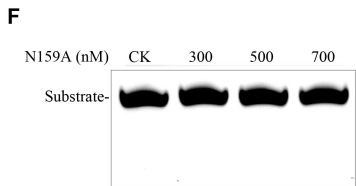
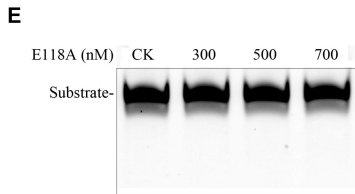
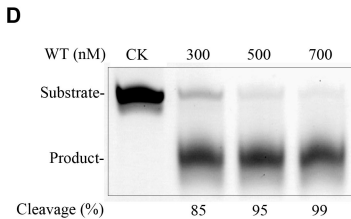
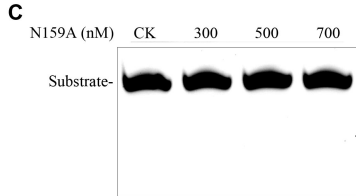
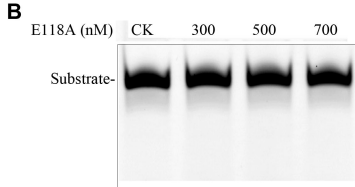
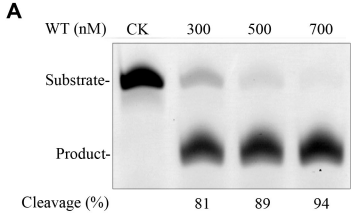


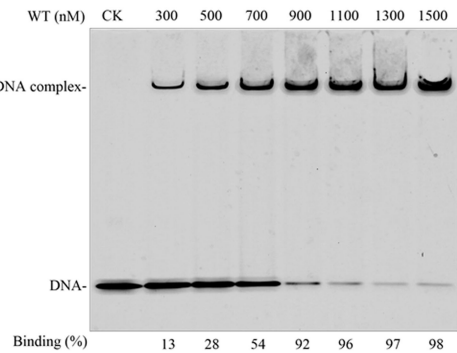
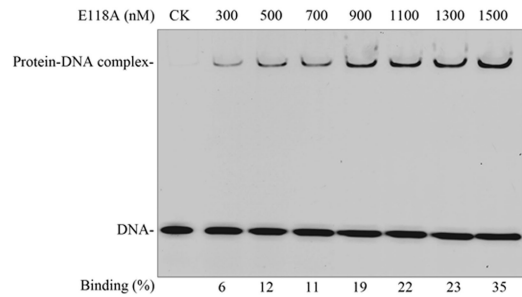
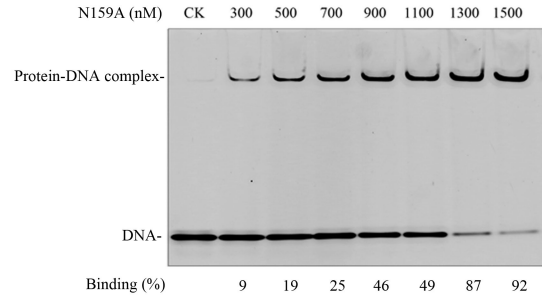
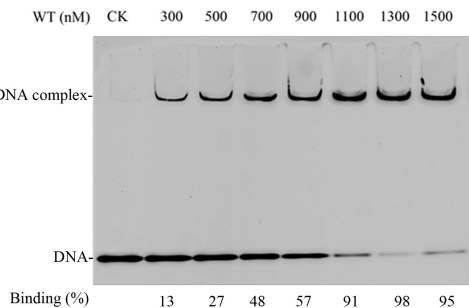
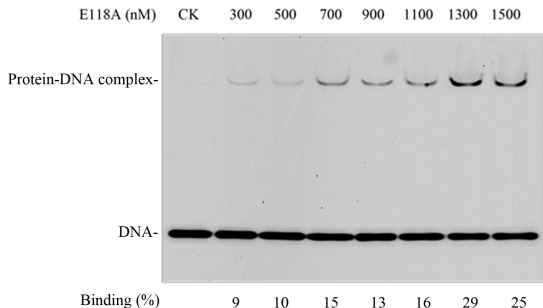
A**B****C****D****E**

A**B**

A**B****C**

A**B**



A**B****C****D****E****F**