RESEARCH ARTICLE

Crystal structure of mimivirus uracil-DNA glycosylase

Eunju Kwon, Deepak Pathak, Hyeun Wook Chang, Dong Young Kim*

College of Pharmacy, Yeungnam University, Gyeongsan, Gyeongbuk, South Korea

* dyokim@ynu.ac.kr

Abstract

Cytosine deamination induced by stresses or enzymatic catalysis converts deoxycytidine into deoxyuridine, thereby introducing a G to A mutation after DNA replication. Base-excision repair to correct uracil to cytosine is initiated by uracil-DNA glycosylase (UDG), which recognizes and eliminates uracil from DNA. Mimivirus, one of the largest known viruses, also encodes a distinctive UDG gene containing a long N-terminal domain (N-domain; residues 1–130) and a motif-I (residues 327–343), in addition to the canonical catalytic domain of family I UDGs (also called UNGs). To understand the structural and functional features of the additional segments, we have determined the crystal structure of UNG from *Acanthamoeba polyphaga* mimivirus (mvUNG). In the crystal structure of mvUNG, residues 95–130 in the N-domain bind to a hydrophobic groove in the catalytic domain, and motif-I forms a short β-sheet with a positively charged surface near the active site. Circular dichroism spectra showed that residues 1–94 are in a random coil conformation. Deletion of the three additional fragments reduced the activity and thermal stability, compared to full-length mvUNG. The results suggested that the mvUNG N-domain and motif-I are required for its structural and functional integrity.

Introduction

Although uracil is a base component of RNA, it is occasionally introduced into DNA by hydrolytic deamination of cytosine [1]. Uracil generated by cytosine deamination causes a 'G to A mutation' in the complementary strand during DNA replication, which must be corrected to maintain genomic integrity [1]. Uracil in DNA can be repaired by the base excision repair (BER) pathway [2, 3]. In the BER pathway, uracil-DNA glycosylase (UDG) recognizes uracil which exists in DNA and creates an abasic site by cleaving the N-glycosidic bond between uracil and deoxyribose. Apurinic/apyrimidinic (AP) endonuclease then removes the sugar-phosphate backbone of the abasic site. Finally, DNA polymerase adds a correct nucleotide to the site [2, 3].

UDG genes are found ubiquitously in cellular organisms from prokaryotes to humans, and have also been identified in some large DNA viruses, such as poxviruses and herpesviruses [2, 4, 5]. UDG proteins are classified into six families, based on characteristic motifs related to their substrate specificity and catalytic activity [6]. The family-I UDGs, called UNGs, are

characterized by five conserved motifs that are arranged around the active site [7]. The motifs include the catalytic water-activating loop, the Pro-rich loop, the uracil-binding motif, the Gly-Ser loop, and the minor groove intercalation loop, corresponding to 143-GQDPYH-148, 165-PPPPS-169, 201-LLLN-204, 246-GS-247, and 268-HPSPLS-273 in human UNG1 (hUNG1), respectively [7–9].

Although UNG removes uracil included in single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA), it shows a preference for ssDNA than dsDNA as a substrate [10, 11]. The catalytic mechanism of UNG has been studied extensively through structural and bio-chemical analyses [12–16]. It is described as "pinch, push, plug, and pull". In human UNG, the Pro-rich, Gly-Ser, and minor groove intercalation loops lead to DNA bending by compressing the DNA backbone (pinch). Next, a Leu residue in the minor groove intercalation loop pene-trates into the DNA double helix (push), resulting in the base flipping of uracil, followed by uracil insertion into the binding pocket of the active site (plug). After removing uracil from DNA, the Leu in the minor groove intercalation loop is retracted (pull).

The UNG gene in *Acanthamoeba polyphaga* mimivirus (APMV) was identified recently by its genome sequencing [17]. APMV, which belongs to the *Mimiviridae* family of nucleocyto-plasmic large DNA viruses (NCLDV), is one of the largest known viruses, comparable to bacte-ria in size [18]. It has a large-sized virion (~750 nm) and harbors a complex genome (~1.2 Mb) that encodes approximately 1,000 putative genes. Some of these genes have been identified as core genes conserved among NCLDVs, whereas some are unique to cellular organisms and have never been identified in other viruses, blurring the boundary between cellular organisms and viruses [17, 19, 20]. Interestingly, APMV UNG (mvUNG) is longer than other known UNGs, suggesting that it contains domains or motifs in addition to the canonical catalytic domain. As a first step towards understanding of mvUNG, we determined its crystal structure and analyzed its enzymatic activity and thermal stability. The results suggested that the addi-tional segments are required for mvUNG's structural and functional integrity.

Materials and methods

Plasmid preparation, protein expression, and purification

The gene encoding mvUNG (residues 1–370; GenBank accession number AAV50521.1) was synthesized artificially after codon optimization (Cosmo Genetech, Seoul, South Korea) and inserted into pET-22b (Merck Millipore, Billerica, MA, USA) with DNA encoding His_{6x}-thioredoxin. The plasmid was then transformed into *Escherichia coli* strain BL21-star(DE3) (Thermo Fisher Scientific, Waltham, MA, USA), and the cells were grown in Luria-Bertani medium. Protein expression was induced using 0.4 mM isopropyl-β-D-thiogalactoside at 15˚C. After overnight induction, cells were harvested by centrifugation, and clarified cell lysates were prepared in buffer A (20 mM hydroxyethyl piperazineethanesulfonic acid (HEPES) pH 7.5, 0.5 M NaCl, 5% glycerol, and 0.2 mM tris(2-carboxyethyl)phosphine). The mvUNG protein was purified using immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). Proteins purified by IMAC were dialyzed twice in buffer A and then treated with thrombin to cleave the His_{x6}-thioredoxin tag from mvUNG. After complete cleavage of the tag, the protein solution was passed through IMAC resin to remove His_{6x}-thioredoxin. mvUNG was further purified using a Superdex-75 size exclusion column equilibrated with buffer A, followed by a Superdex-200 column equilibrated with buffer C (20 mM HEPES pH 7.5, 0.2 M NaCl, 5% glycerol, and 0.2 mM tris(2-carboxyethyl) phosphine). Purified mvUNG was concentrated to 10 mg/mL and estimated to be >95% pure by SDS-PAGE. Mutant proteins (mvUNG$_{95-370}$, mvUNG$_{122-370}$, and mvUNG$_{\Delta327-343}$) were expressed and purified by the same procedure as that used for native mvUNG.

Crystallization, data collection, and structure determination

Crystallization of mvUNG was performed using the micro-batch method at 20°C. The crystallization drop was prepared by mixing 1 μL of protein solution (10 mg/mL) and 1 μL of crystallization solution (0.1 M HEPES pH 7.5, 25% (v/v) Polyethylene glycol 3350, 4% (v/v) Isopropanol, and 0.1 M $CaCl_2$) under a layer of Al's oil (Hampton Research, Aliso Viejo, CA, USA). Needle-shaped crystals of mvUNG were fully grown in 2 months. Crystals were picked up using a cryo-loop (Hampton Research) and then flash-frozen in a cold nitrogen stream without the addition of a cryoprotectant. Diffraction data were collected at PLS-BL7A (Beam line 7A, Pohang Light Source, South Korea) and were indexed, integrated, and scaled using HKL2000 software [21].

The crystal structure of mvUNG was determined by the molecular replacement (MR) method using PHASER [22]. The structure of *Leishmania naiffi* UDG (PDB ID: 3CXM) was used as a template for MR. Two mvUNG proteins were found in an asymmetric unit. Cycles of refinement and model rebuilding were performed at 2.3 Å resolution using phenix.refine [23] and COOT [24]. Residues 95–370 in mvUNG were traced in the electron density. Final refinement with solvents resulted in R/R_{free} values of 17.3 / 21.6% without residues in the disallowed region of the Ramachandran plot. Data collection and refinement statistics are summarized in Table 1. The

Table 1. Data collection and refinement statistics of mvUNG.

Data collection		
Space group		$C222_1$
Unit cell		
	a, b, c (Å)	96.27, 95.62, 132.38
	α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)		50.0–2.30 (2.34–2.30)
Wavelength (Å)		0.97934
Unique reflections		27418 (1370)
Redundancy		7.2 (7.3)
Completeness (%)		99.9 (100.0)
I/σ		26.4 (5.3)
Rmerge (%)		11.9 (57.3)
Refinement		
Resolution		30.0–2.30
No. reflections, working/free		26005/1376
R_{work}/R_{free} (%)		17.3/21.6
No. atoms		
	Protein	4494
	Water	320
B factors		
	Protein	28.9
	Water	32.2
RMSD		
	Bond length (Å)	0.008
	Bond angle (°)	0.987
Ramachandran plot (%)		
	Favor	97.6
	Allowed	2.4
	Disallowed	0.0

figures were drawn using PyMOL [25, 26] and ALSCRIPT [27]. Structural alignment was analyzed using the DALI server [28].

Activity assay of mvUDG

The activity of mvUNG was measured using ssDNA and dsDNA containing a single uracil (5′-GTA AAA CGA CGG CCA GTG UAT TCG AGC TCG GTA CCC GGG G-3′) as a substrate (Cosmo Genetech) [29]. The reaction mixture, containing 10 μM DNA substrate, 20 mM Tris-HCl pH 8.0, 1 mM dithiothreitol, and 1 mM ethylenediaminetetraacetic acid, was incubated with mvUNG at 37˚C. After incubation for 2 h, mvUNG was inactivated at 95˚C for 5 min. Abasic sites generated by mvUNG were incised by treatment with 0.1 mM NaOH at 95˚C for 5 min. The reaction products were analyzed by electrophoresis on 16.7% (w/v) polyacrylamide gels containing 7–8 M urea and 1× Tris-borate-ethylenediaminetetraacetic acid (Biosesang, Seongnam, South Korea) and the substrates were visualized by in-gel silver staining (Thermo Fisher Scientific) [30]. The quantity of the uncleaved substrate was calculated using GelAnalyzer software, and data were analyzed using Minitab 17 (Minitab Inc.). The assay for the activity comparison of mvUNG mutants was repeated seven times and the activity values of activity were averaged.

SEC with multi-angle laser light scattering (SEC-MALLS)

To measure its molecular size, purified mvUNG was injected into a Superdex-200 analytical column connected to Purifier FPLC (GE Healthcare, Chicago, IL, USA) and the elution products were applied to inline miniDAWN TREOS MALLS and Optilab digital signal processing refractive index detectors (Wyatt Technology Corporation, Santa Barbara, CA, USA). Data were collected at 1.0 s intervals at a flow rate of 0.4 mL/min and were analyzed using the ASTRA V software package (Wyatt Technology Corporation).

Circular dichroism (CD) measurement

CD spectra were obtained using the JASCO J-810 spectropolarimeter (Jasco, Tokyo, Japan). For far-UV CD spectra, 10 μM protein in a 1 mm cuvette was exposed to a wavelength range of 200–260 nm, and the data were recorded at a 1 nm intervals with a scanning speed of 5 nm/min. Three spectra from a protein sample were collected and averaged. The spectra were smoothed after subtraction of the buffer spectrum. For the thermal melting spectra, 10 μM protein in a 1 mm cuvette was exposed to a wavelength of 220 nm. The thermal melting curves were recorded at 0.5˚C intervals in the temperature range of 20–90˚C. Temperature was increased at a rate of 2˚C/min.

Accession number

The final coordinate and structure factor for mvUNG were deposited into the Protein Data Bank (PDB ID: **5X55**).

Results

Structure determination of catalytically active mvUNG

mvUNG (residues 1–370) contains a catalytic domain conserved among UNG family proteins. Additionally, it has an N-terminal domain (N-domain; residues 1–130) and an insertional motif (motif-I; residues 327–343; Fig 1a). The N-domain in mvUNG is longer than those in other UNGs, and motif-I is not found in the catalytic domains of other known UNGs. For the structural and functional analysis of mvUNG, we prepared catalytically active mvUNG
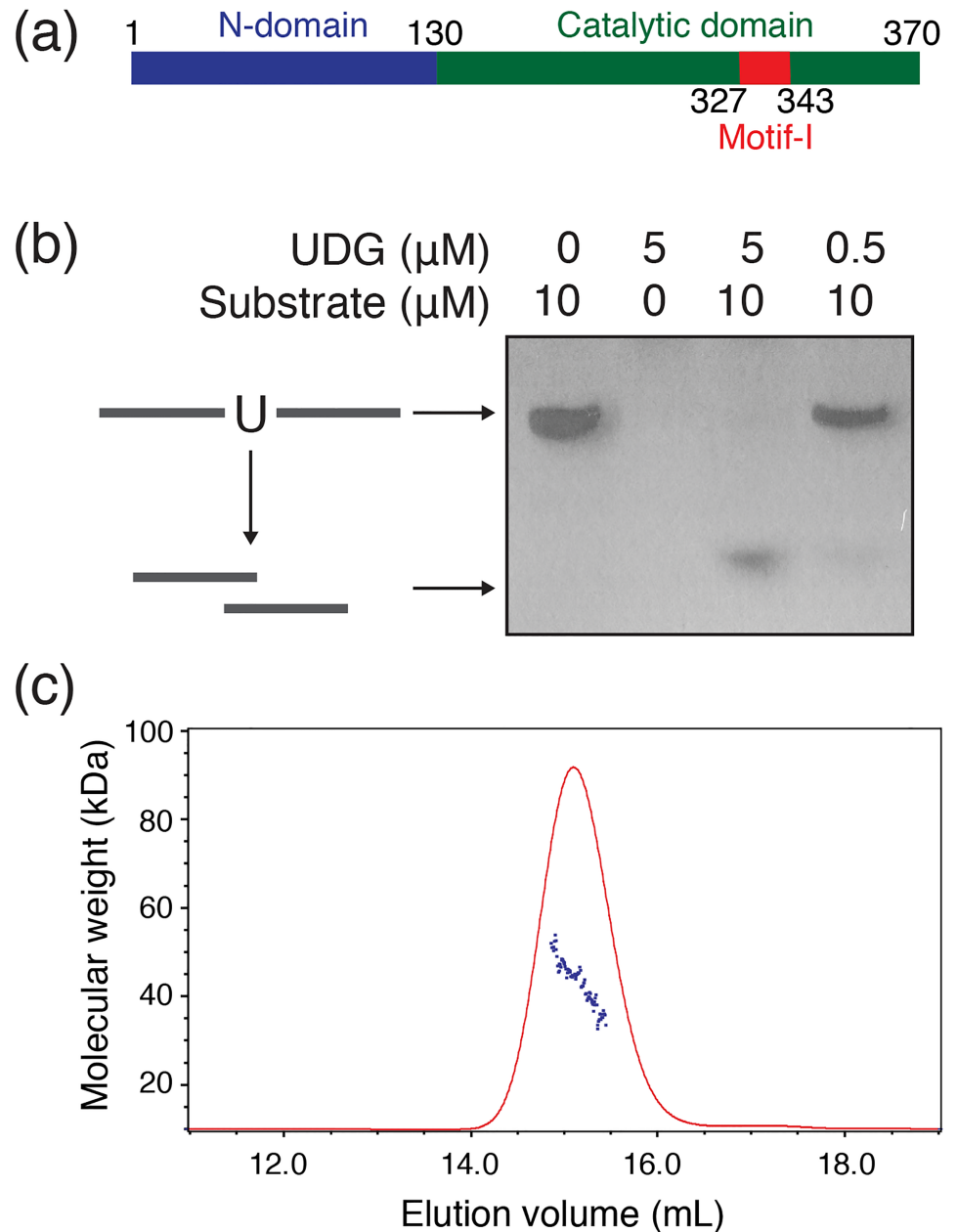
Fig 1. Characterization of mvUNG. (a) Overall domain organization of mvUNG. Green, blue, and red bars represent the relative lengths of the catalytic domain, N-terminal domain, and motif-I, respectively. The schematic shows that mvUNG contains a long N-terminal domain and motif-I, which is inserted in the catalytic domain. (b) Enzymatic activity of mvUNG. In the silver-stained gel, the top and bottom bands indicate uncleaved ssDNA containing a single uracil and cleaved fragments, respectively. This shows that mvUNG has the activity of uracil-DNA glycosylase activity. (c) Size measurement by SEC-MALLS. The molecular mass of purified mvUNG in solution was calculated as 43.4 kDa, which is close to that of the monomer (molecular weight of a monomer: 42.2 kDa).

protein. The recombinant protein was expressed in *E. coli* and purified by Ni-affinity chromatography and SEC. The activity of the purified mvUNG was measured by incubating it with ssDNA or dsDNA containing a single uracil. In the assay, the incubated mixture was treated with NaOH to cleave the abasic site in DNA, and the cleavage was then visualized using a

silver-stained gel. The gel showed that ssDNA substrate was cleaved in the presence of mvUNG, whereas dsDNA was not (Fig 1b and S1 Fig). This result indicates that the uracil in ssDNA is removed by mvUNG, while uracil in dsDNA does not seem to be a preferred substrate.

Needle-shaped crystals were grown from full-length mvUNG, and the diffraction data were collected up to 2.3 Å resolution. The phase was obtained by molecular replacement, using the structure of *Leishmania naiffi* UDG, which shows high sequence identity with the catalytic core domain of mvUNG. Residues 95–370 of two mvUNG proteins in the asymmetric unit were traced into the electron density (S2 Fig), and the final model was refined with R/R_{free} values of 17.3/21.6% (Table 1). Matthews's coefficient (V_m) and the solvent content under crystal packaging of full-length mvUNG were calculated as 1.8 Å^3/Da and 31.8%, respectively, which are out of the range commonly observed for protein crystals, while those of mvUNG$_{95-370}$ were 2.4 Å^3/Da and 48.8%, which falls within the range. A solvent space that could accommodate residues 1–94 was insufficient in the crystal. Moreover, mvUNG$_{95-370}$ facilitated the growth of the needle-shaped crystal under same crystallization conditions as full-length mvUNG. Thus, the N-terminal residues seem to be truncated before crystallization.

Two mvUNGs in the asymmetric unit were superimposed with a root-mean-square deviation (RMSD) of 0.4 Å for 276 Cα atoms, indicating that the two mvUNGs are in the same conformation. The oligomeric state in solution was determined by SEC. In SEC, mvUNG (molecular weight of a monomer: 42.2 kDa) was eluted as a monodisperse protein between ovalbumin (44 kDa) and conalbumin (75 kDa), which means that it was slightly larger size than a monomer of mvUNG (S3 Fig). To measure its oligomerization state precisely, mvUNG was applied to SEC-MALLS. The molecular mass of mvUNG calculated by SEC-MALLS was approximately 43.4 kDa (Fig 1c), indicating that the protein is a monomer in solution. Thus, the slightly faster elution in SEC seems to have resulted from a flexible region or an elongated shape of mvUNG.

Overall structure of mvUNG

Crystal structures of UNG proteins from diverse organisms and large viruses including human [8], *E. coli* [31], herpesviruses [32], and poxviruses [33] have been reported. Their catalytic domains share a highly conserved α/β fold containing a four-stranded parallel β-sheet surrounded by eight α-helices. The catalytic domain of mvUNG also forms a similar fold, comprising a central β-sheet (S2, S3, S5, and S6 Figs) sandwiched by α-helices (H5–H12). The N-termini of H5, H7, and H12 and the C-terminus of H7 extend to the 3$_{10}$-helix. Additionally, two short β-strands (S1 and S4 Figs) form an antiparallel β-sheet, separately from the central β-sheet (Fig 2). Searching for structural similarity using the DALI program, showed that mvUNG has high similarity to UNG proteins, such as human UNG (hUNG) [14] and herpesvirus UNG (hsvUNG) [32]. hUNG and hsvUNG were superimposed onto mvUNG with RMSD values of 1.2 Å for 220 Cα and 1.6 Å for 221 Cα positions, respectively, which indicated that the catalytic domain of mvUNG shares a conserved fold with other UNG proteins (Fig 2d).

In addition to the canonical fold of UNGs, the structure of mvUNG contains a long N-terminal domain (N-domain; residues 1–130) and an insertional motif (motif-I; residues 327–343) located between S6 and H12 (Fig 2). The N-domains in UNG proteins are highly variable in length and do not share conserved motifs. In the structure of mvUNG, residues 95–130 comprise four short helices, two α-helices (H1 and H3), and two 3$_{10}$-helices (H2 and H4). Residue R96 in the N-domain forms an ionic interaction with E169 in the catalytic domain, and the segment is fitted into a hydrophobic groove of the catalytic domain. Residues 95–130 of
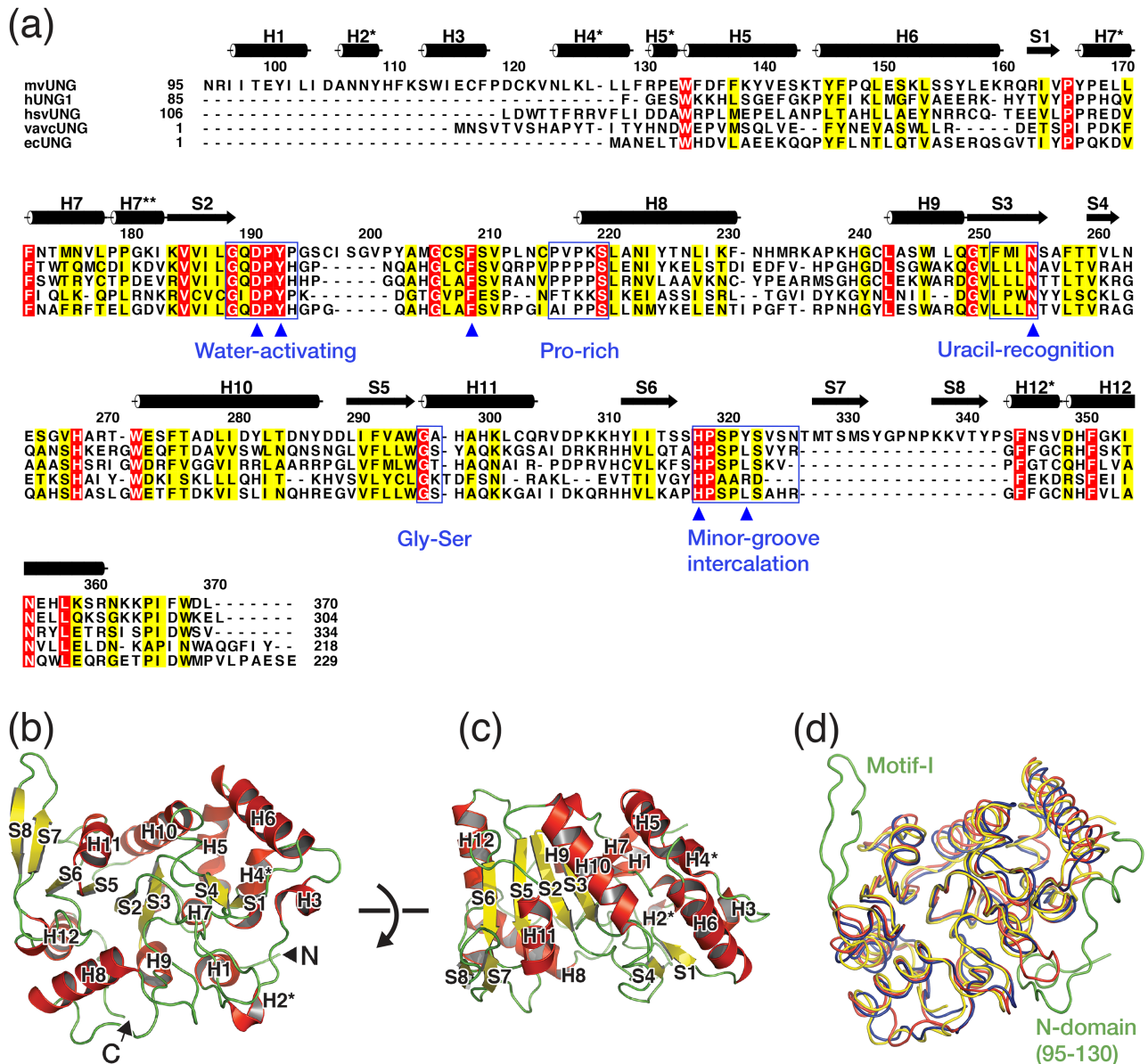
Fig 2. The structure of mvUNG. (a) Structure-based sequence alignment of UNGs. UNG sequences from *Acanthamoeba polyphaga* mimivirus (mvUNG), human (hUNG), herpes simplex virus (hsvUNG), vaccinia virus (vacvUNG) and *Escherichia coli* (EcUNG) are aligned based on a structural comparison. Identical and homologous residues are boxed in red and yellow, respectively. Residues of the active site are marked with blue triangles. The five conserved motifs in UNGs are boxed and labeled in blue. The secondary structure of mvUNG is displayed using a cylinders for helices and arrows for β-strands. Helices and β-strands are labeled with H and S, respectively. Helix labels with and without an asterisk (*) indicate an α-helix and 3_{10} helix, respectively. (b, c) Ribbon models of mvUNG shown at two different orientations. α-helix, β-strand, and loops are colored differently, and the secondary structures are labeled using the same scheme as in (a). (d) Structural comparison of UNGs. Structures of mvUNG, hUNG (PDB ID: 1AKZ), and hsvUNG (PDB ID: 1UDG) were superimposed, and Cα trace models were drawn in the same orientation as (b). Red, yellow, and blue represent mvUNG, hUNG, and hsvUNG, respectively. Additional segments in mvUNG (motif-I and residues 95–130 of the N-domain) are colored green and labeled.

the N-domain seem to stabilize the catalytic core domain by binding to its hydrophobic surface (S4 Fig). Motif-I forms a two-stranded antiparallel β-sheet (S7 and S8) (Figs 2 and 3a) and provides a positively charged surface near the active site, which might improve the interaction with the DNA backbone (S5 Fig). Taken together, the mvUNG structure suggests that the
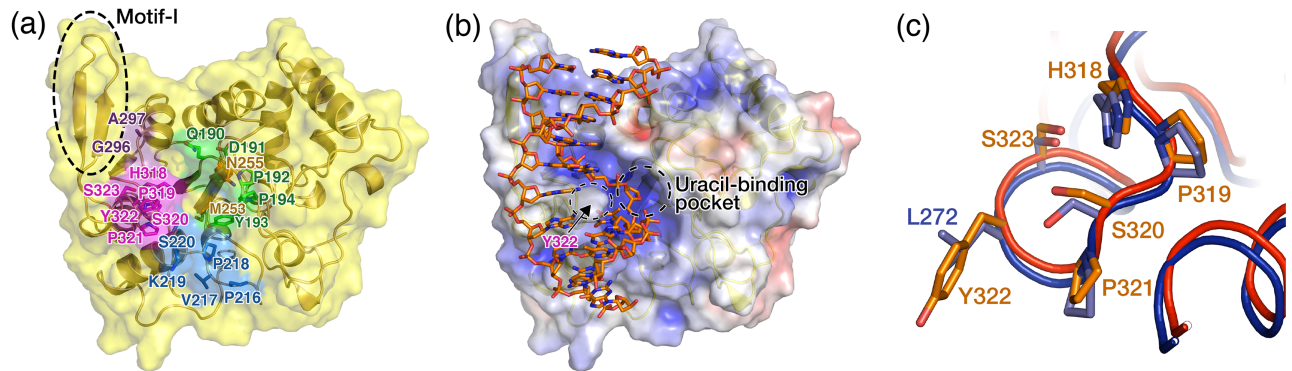
Fig 3. The active site of mvUNG. (a) Motifs of mvUNG conserved in UNG proteins. The ribbon and stick model covered by a transparent yellow surface is shown. Residues of the water-activating, Pro-rich, uracil-recognition, Gly-Ser, and minor-groove intercalation motifs are labeled and colored green, blue, orange, purple, and magenta, respectively. Motif-I (black dotted oval) is on the left side of the Gly-Ser loop (purple) and the minor-groove intercalation loop (magenta) in this figure. (b) The DNA-binding model of mvUNG. The model was generated by superimposing mvUNG on the structure of hUNG/DNA containing an abasic site (PDB ID: 2SSP). mvUNG was drawn as a surface model with charge distribution. DNA in the structure of hUNG/DNA complex is drawn as a stick model. The figure shows that DNA is intercalated by Y322 and binds to a positively charged groove formed by five motifs conserved in UNG family proteins. mvUNG structures in (a, b) are in the same orientation as in Fig 2b. (c) Comparison of active site residues. Structures of mvUNG and hUNG (PDB ID: 1AKZ) are superimposed. Red and blue Cα trace models represent mvUNG and hUNG, respectively. Residues of the active site are drawn as sticks and labeled. Leu residue in minor-groove intercalation loop of hUNG is replaced by Tyr (Y322) in mvUNG.

visible N-domain (residues 95–130) and motif-I (residues 327–343) might affect the structural and functional integrity of mvUNG.

The active site of mvUNG

UNG proteins from human, *E. coli*, and herpesviruses contain six conserved active site residues that are involved in uracil-binding specificity and catalysis (Fig 2a; D145, Y147, F158, N204, H268, and L272 in human UNG1) [8, 31, 32]. However, some UNG proteins do not contain a Leu residue in the minor groove intercalation loop [33, 34]. For example, the Leu residue is replaced by hydrophobic Phe in *Bacillus subtilis* UNG and by positively charged Arg in vaccinia virus UNG (vacvUNG). mvUNG also does not contain the Leu residue (Fig 2a). The catalytic residue (D191) and the other four residues for uracil-binding specificity in the binding pocket (Y193, F209, N255, and H318) are conserved as the canonical active site residues, whereas Leu in the minor groove intercalation loop is replaced by Y322 in mvUNG (Figs 2a and 3c). The bulky polar residue, Y322, does not induce a conformational change of the minor groove intercalation loop. The loop conformation, including amino acid side chains, is superposed well with that of human UNG (Fig 3c). The Leu residue (Phe in *B. subtilis* UNG) in the minor groove intercalation loop is a critical residue for the inhibition of UNGs by Ugi and p56, as well as for catalytic activity. The Leu or Phe residue is recognized by the hydrophobic pockets of Ugi [12, 35–37] and p56 [34]. The vacvUNG, which contains Arg instead of Leu in the minor groove intercalation loop, is not inhibited by Ugi [38]. This implies that the interaction between mvUNG and inhibitor proteins (Ugi and p56) might be prevented or weakened by the hydroxyl group of Y322 in mvUNG.

In addition to active site residues, UNGs share five conserved motifs required for the formation of a uracil-binding pocket and DNA-binding groove (Figs 2a and 3). Conserved motifs include the Pro-rich and Gly-Ser loops for the compression of DNA substrate, the minor groove intercalation loop (or Leu-intercalation loop) for the separation of ds DNA, and the uracil-recognition and water-activating loops for enzymatic catalysis. In mvUNG, 189-GQDPYP-194, 216-PVPKS-220, 252-FMIN-255, 296-GA-297, and 318-HPSPYSVSN-326

are structurally aligned as the water-activating, Pro-rich, uracil-recognition, Gly-Ser, and minor-groove intercalation loops, respectively (Figs 2a and 3). In the comparison of the conserved motifs, 16 and 11 out of 26 residues in the motifs of mvUNG are identical to those of hUNG and vacvUNG, respectively (Fig 2a).

Deletions of N-domain and motif-I reduce enzymatic activity

To evaluate the effects of the additional segments (residues 1–94, 95–130 and 327–343) on UNG activity, we performed activity assays with the mutant mvUNG proteins. A motif-I deletion mutant (mvUNG$_{\Delta327-343}$) and two N-terminal deletion mutants (mvUNG$_{95-370}$ and mvUNG$_{122-370}$) were prepared using the same procedure as the full-length mvUNG. The activity assays showed that UNG activity of the three deletion mutants was significantly lower than that of the full-length mvUNG (Fig 4).

Motif-I (residues 327–343) of mvUNG is located around the Gly-Ser and minor-groove intercalation loops, without disrupting the active site. Its surface distributes a weakly positive charge (Fig 3 and S5 Fig). In the activity assay, the deletion of motif-I reduced the activity of mvUNG to 59% in the reaction of 0.2 μM mvUNG$_{\Delta327-343}$ and 10 μM ss-DNA substrate, compared with the activity of full-length mvUNG (first and second bars in Fig 4b). This implied that motif-I might affect the efficiency of the active site by improving the interaction between the enzyme and DNA. The N-domain in eukaryotic and viral UNGs varies in length and its sequences are not conserved. It seems to contribute to diverse cellular functions of UNGs. For example, the N-domain of hUNG is involved in the subcellular localization of isoforms [39] and the interaction with the DNA repair factor RPA [40, 41]. In the activity assay of the N-domain deletion mutants, mvUNG$_{95-370}$ which residues 1–94 is deleted showed 55% activity, compared with full-length mvUNG (first and third bars in Fig 4b). Additional deletion of residues 95–121 reduced the activity to 15%, even though the deletion mutant contains the entire catalytic domain (first and fourth bars in Fig 4b). This result indicated that both invisible (residues 1–94) and visible (residues 95–130) segments of the N-domain are required for the intrinsic activity of mvUNG.

The flexible region of the N-domain (residues 1–94) affects the thermal stability of mvUNG

The N-domain is required for the intrinsic activity of mvUNG, as shown above (Fig 4). Residues 95–130 appear to affect the catalytic domain by binding to its hydrophobic surface, while residues 1–94 are not visible in the crystal structure of mvUNG, probably due to its truncation before crystallization. To evaluate the secondary structure of the invisible segment in the N-domain, we compared the CD spectra of mvUNG and mvUNG$_{95-370}$. In the far-UV CD spectra, the deletion of residues 1–94 did not change the pattern and scale of the CD spectrum in the range between 210–250 nm, indicating that residues 1–94 have a random coil conformation (Fig 5a). Consistent with the CD data, residues 1–40 and 50–90 of mvUNG were predicted to be disordered regions using a computational approach. Next, CD ellipticity at 220 nm was measured by increasing the temperature gradually from 20 to 90˚C to calculate melting temperature (T$_m$). The T$_m$ value of full-length mvUNG (55.9˚C) was 4.4˚C higher than that of mvUNG$_{95-370}$ (51.5˚C) (Fig 5b), indicating that the deletion of residues 1–94 decreases the thermal stability of mvUNG. Activity assays and CD spectra implied that the unstructured segment (residues 1–94) in the N-domain also contributes to thermal stability and catalytic activity of mvUNG.
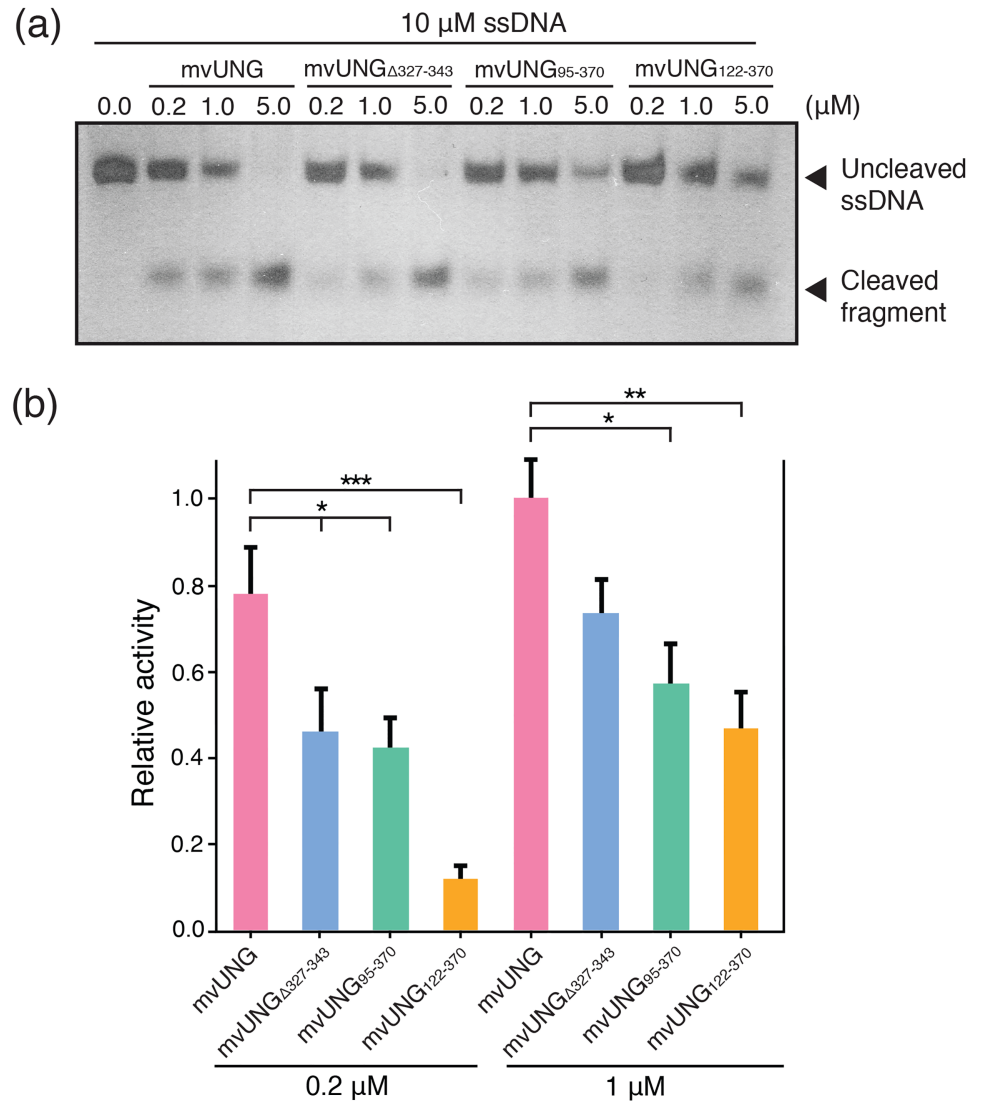
Fig 4. Activity comparison of mvUNG mutants. (a) Silver-stained gel showing the cleavage of ssDNA by mvUNG mutants. Full-length mvUNG, mvUNG$_{\Delta327-343}$, mvUNG$_{95-370}$, and mvUNG$_{122-370}$ were incubated with ssDNA to compare their UNG activities. (b) Relative activities of mvUNG mutants drawn as a histogram. The activity was calculated based on the amount of remaining uncleaved substrate. Seven independent measurements were averaged (S6 Fig). *P* values were calculated using a two-sample *t*-test. Error bars indicate mean ± standard error (*P< 0.05, **P < 0.01, ***P < 0.001, n = 7). In the reaction of 0.2 μM mvUNG and 10 μM ssDNA, the activities of mvUNG$_{\Delta327-343}$, mvUNG$_{95-370}$, and mvUNG$_{122-370}$ were reduced to 59, 55, and 15% of the activity of full-length mvUNG, respectively. In the reaction of 1 μM mvUNG and 10 μM ssDNA, the activities of mvUNG$_{\Delta327-343}$, mvUNG$_{95-370}$, and mvUNG$_{122-370}$ were reduced to 73, 58, and 47% of the activity of the full-length mvUNG, respectively.

Discussion

UNG is a ubiquitous enzyme for base excision repair. All UNGs, from human to large viruses, contain a conserved catalytic domain that removes uracil from DNA. Recently, a gene was identified from a mimivirus that encodes an UNG containing a long N-domain and motif-I in addition to the catalytic core domain [17]. To understand the structural features of the additional segments, we determined the crystal structure of mvUNG. In the crystal structure of
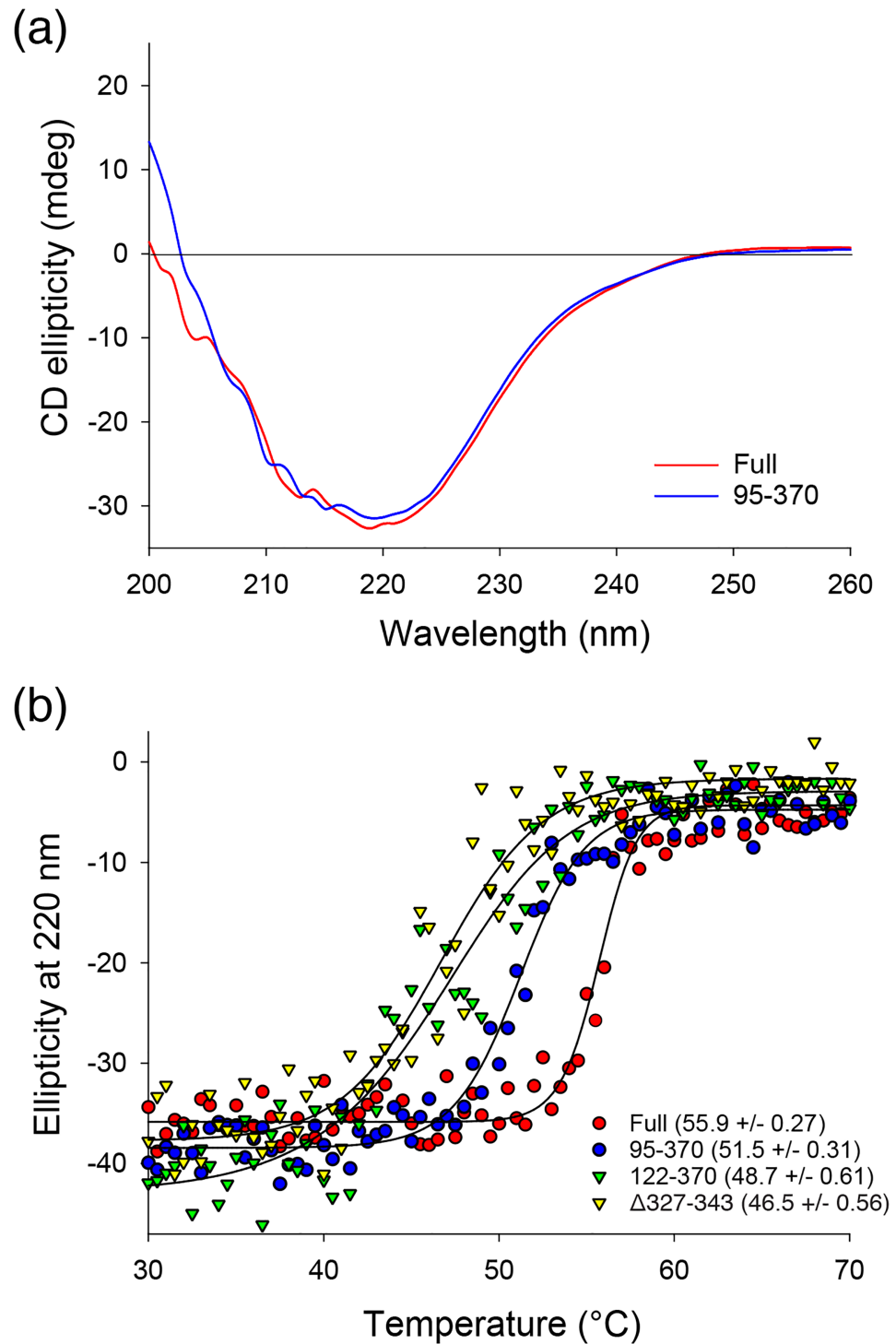
(a)

(b)

Fig 5. Circular dichroism (CD) spectra of mvUNGs. (a) Far-UV CD spectra of full-length mvUNG (Full) and mvUNG$_{95-370}$ (95–370). The two CD spectra show similar patterns and ellipticity scales, indicating that residues 1–94 of mvUNG are close to a random coil conformation. (b) Thermal melting curves. Melting curves were collected at a wavelength of 220 nm by increasing temperature to 90°C. Calculated T$_m$ values of full-length mvUNG, mvUNG$_{95-370}$, mvUNG$_{122-370}$, and mvUNG$_{\Delta 327-343}$ were 55.9, 51.5, 48.7, and 46.5°C, respectively, indicating that deletion of the additional segments decreased the thermal stability of mvUNG.

mvUNG, residues 95–130 in the N-domain bind to the hydrophobic groove of the catalytic domain (S4 Fig). The deletion of residues 95–121 decreased the activity (Fig 4b) and thermal stability (Fig 5b) of the crystallized protein (mvUNG$_{95-370}$). Thus, the structured part of the N-domain seems to stabilize the catalytic domain by binding directly to its hydrophobic surface. Motif-I, inserted into the catalytic domain, forms a short β-sheet near the active site (S5 Fig). The deletion of motif-I also decreased the activity (Fig 4b) and thermal stability of mvUNG (Fig 5b). The UNG activity of mvUNG$_{\Delta 327-343}$ was similar to that of mvUNG$_{95-370}$, while its thermal stability was lower than the N-domain deletion mutants. Although the structural features and activity assays indicated that motif-I might contribute to DNA binding, the possibility of reduced activity by decreased thermal stability cannot be excluded. Thus, the activity assay and thermal melting curves suggested that both the structured part of the N-domain (residues 95–130) and motif-I (residues 327–343) are required for the structural and functional integrity of mvUNG.

In contrast to the structured part of the N-domain and the motif-I, residues 1–94 in N-domain are not included in the crystal structure of mvUNG. They seem to be truncated before crystallization. Indeed, all reported crystal structures of UNG proteins display only a catalytic domain. CD spectra showed that residues 1–94 in mvUNG are close to a random coil (Fig 5a). Although the flexible fragment seems to be unnecessary for UNG activity and thermal stability of the core domain, the deletion of residues 1–94 in mvUNG decreased the activity and thermal stability (Figs 4 and 5). It is unclear how the flexible fragment affects the function of mvUNG. One possibility is an intramolecular interaction between the catalytic domain and the flexible N-domain. The intramolecular interaction might be not tight enough or might not be a specifically ordered interaction. Recently, it was reported that the disordered C-domain of human DNA glycosylase NEIL1 (hNEIL1) stabilizes the catalytic domain via electrostatic interactions [42, 43]. In a binding assay of hNEIL1, performed by His-tag pull-down followed by immunoblotting, the cluster of basic residues (355-KKGRRK-360) in an intrinsically disordered C-terminal domain weakly bound to the acidic cluster in the core domain. Notably, the N-domain of mvUNG also contains a cluster of basic residues (69-KKSKKSKKSKKSKKS-83), suggesting that the cluster might be associated with the integrity of mvUNG. Although the disordered domains between mvUNG and hNEIL1 are located at the opposite ends of amino acid sequences; they may function similarly because the N- and C-termini are exposed in the same direction. In addition, the N-domain of mvUNG and the C-domain of hNEIL1 share 15% sequence identity over about 100 amino acids, which is higher than that between the N-domains of other UNGs.

Supporting information

S1 Fig. Comparison of ssDNA and dsDNA as substrates for mvUNG. The UDG activity of mvUNG was measured using ssDNA and dsDNA as a substrate. Uracil was removed from ssDNA but not dsDNA.
(TIF)

S2 Fig. Two mvUNG proteins in an asymmetric unit (MolA and MolB). (a) Ribbon model of two monomers in an asymmetric unit. (b) Cα trace model of two monomers superimposed.
(TIF)

S3 Fig. Size exclusion chromatography of mvUNG. mvUNG was eluted as a monodisperse protein between ovalbumin (44 kDa; red triangle) and conalbumin (75 kDa; blue triangle) in both the Superdex-200 (a) and Superdex-75 (b) analytical columns.
(TIF)

S4 Fig. The structure of the N-domain in mvUNG. (a) The visible N-domain and catalytic domain in mvUNG are drawn as stick and surface models, respectively. (b) The mvUNG monomer (residues 95–370) was drawn as a surface model with the charge distribution. (a, b) The panels show that the ordered fragment (residues 95–130) in the N-domain binds to the hydrophobic surface of the catalytic domain.
(TIF)

S5 Fig. DNA-binding models of mvUNG (a) and mvUNG$_{\Delta 327-343}$ (b). The models were generated by superimposing mvUNG and mvUNG$_{\Delta 327-343}$ on the structure of hUNG/DNA containing an abasic site (PDB ID: 2SSP). mvUNGs and DNA were drawn as surface and backbone models, respectively. Motif-I (residues 327–343) does not disturb DNA-binding and provides a positively charged surface near the active site.
(TIF)

S6 Fig. Silver-stained polyacrylamide gels used to calculate mvUNG activity.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Eunju Kwon, Dong Young Kim.

Data curation: Eunju Kwon.

Formal analysis: Eunju Kwon, Deepak Pathak, Dong Young Kim.

Funding acquisition: Hyeun Wook Chang, Dong Young Kim.

Investigation: Eunju Kwon, Deepak Pathak, Dong Young Kim.

Methodology: Eunju Kwon, Deepak Pathak, Dong Young Kim.

Project administration: Eunju Kwon, Dong Young Kim.

Resources: Eunju Kwon, Deepak Pathak.

Supervision: Hyeun Wook Chang, Dong Young Kim.

Validation: Eunju Kwon, Dong Young Kim.

Visualization: Eunju Kwon, Dong Young Kim.

Writing – original draft: Eunju Kwon, Dong Young Kim.

Writing – review & editing: Eunju Kwon, Deepak Pathak, Dong Young Kim.

References

1. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993; 362(6422):709–15. https://doi.org/10.1038/362709a0 PMID: 8469282.

2. Lindahl T. DNA repair enzymes. Annu Rev Biochem. 1982; 51:61–87. https://doi.org/10.1146/annurev. bi.51.070182.000425 PMID: 6287922.

3. Sancar A, Sancar GB. DNA repair enzymes. Annu Rev Biochem. 1988; 57:29–67. https://doi.org/10. 1146/annurev.bi.57.070188.000333 PMID: 3052275.

4. Worrad DM, Caradonna S. Identification of the coding sequence for herpes simplex virus uracil-DNA glycosylase. J Virol. 1988; 62(12):4774–7. PMID: 2846888;

5. Upton C, Stuart DT, McFadden G. Identification of a poxvirus gene encoding a uracil DNA glycosylase. Proc Natl Acad Sci U S A. 1993; 90(10):4518–22. PMID: 8389453;

6. Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci. 2014; 23(12):1667–85. https://doi.org/10.1002/pro.2554 PMID: 25252105;

7. Pearl LH. Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res. 2000; 460(3–4):165–81. PMID: 10946227.

8. Mol CD, Arvai AS, Slupphaug G, Kavli B, Alseth I, Krokan HE, et al. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell. 1995; 80 (6):869–78. PMID: 7697717.

9. Parikh SS, Putnam CD, Tainer JA. Lessons learned from structural results on uracil-DNA glycosylase. Mutat Res. 2000; 460(3–4):183–99. PMID: 10946228.

10. Panayotou G, Brown T, Barlow T, Pearl LH, Savva R. Direct measurement of the substrate preference of uracil-DNA glycosylase. J Biol Chem. 1998; 273(1):45–50. PMID: 9417045.

11. Slupphaug G, Eftedal I, Kavli B, Bharati S, Helle NM, Haug T, et al. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. Biochemistry. 1995; 34(1):128–38. PMID: 7819187.

12. Mol CD, Arvai AS, Sanderson RJ, Slupphaug G, Kavli B, Krokan HE, et al. Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. Cell. 1995; 82 (5):701–8. PMID: 7671300.

13. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature. 1996; 384(6604):87–92. https://doi.org/10.1038/384087a0 PMID: 8900285.

14. Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J. 1998; 17(17):5214–26. https://doi.org/10.1093/emboj/17.17.5214 PMID: 9724657;

15. Jiang YL, Stivers JT. Mutational analysis of the base-flipping mechanism of uracil DNA glycosylase. Biochemistry. 2002; 41(37):11236–47. PMID: 12220189.

16. Wong I, Lundquist AJ, Bernards AS, Mosbaugh DW. Presteady-state analysis of a single catalytic turnover by Escherichia coli uracil-DNA glycosylase reveals a "pinch-pull-push" mechanism. J Biol Chem. 2002; 277(22):19424–32. https://doi.org/10.1074/jbc.M201198200 PMID: 11907039.

17. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004; 306(5700):1344–50. https://doi.org/10.1126/science.1101485 PMID: 15486256.

18. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Science. 2003; 299(5615):2033. https://doi.org/10.1126/science.1081867 PMID: 12663918.

19. Raoult D, La Scola B, Birtles R. The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent. Clin Infect Dis. 2007; 45(1):95–102. https://doi.org/10.1086/518608 PMID: 17554709.

20. Aherfi S, Colson P, La Scola B, Raoult D. Giant Viruses of Amoebas: An Update. Front Microbiol. 2016; 7:349. https://doi.org/10.3389/fmicb.2016.00349 PMID: 27047465;

21. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–26. PMID: 27754618.

22. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40(Pt 4):658–74.: https://doi.org/10.1107/S0021889807021206 PMID: 19461840;

23. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012; 68(Pt 4):352–67.: https://doi.org/10.1107/S0907444912001308 PMID: 22505256;

24. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66(Pt 4):486–501. https://doi.org/10.1107/S0907444910007493 PMID: 20383002;

25. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001; 98(18):10037–41. https://doi.org/10.1073/pnas.181342398 PMID: 11517324;

26. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

27. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. 1993; 6(1):37–40. PMID: 8433969.

28. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38(Web Server issue):W545–9. https://doi.org/10.1093/nar/gkq366 PMID: 20457744;

29. Bulgar AD, Weeks LD, Miao Y, Yang S, Xu Y, Guo C, et al. Removal of uracil by uracil DNA glycosylase limits pemetrexed cytotoxicity: overriding the limit with methoxyamine to inhibit base excision repair. Cell Death Dis. 2012; 3:e252. https://doi.org/10.1038/cddis.2011.135 PMID: 22237209;

30. Lu CC, Huang HT, Wang JT, Slupphaug G, Li TK, Wu MC, et al. Characterization of the uracil-DNA glycosylase activity of Epstein-Barr virus BKRF3 and its role in lytic viral DNA replication. J Virol. 2007; 81 (3):1195–208. https://doi.org/10.1128/JVI.01518-06 PMID: 17108049;

31. Xiao G, Tordova M, Jagadeesh J, Drohat AC, Stivers JT, Gilliland GL. Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited. Proteins. 1999; 35(1):13–24. PMID: 10090282.

32. Savva R, McAuley-Hecht K, Brown T, Pearl L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature. 1995; 373(6514):487–93. https://doi.org/10.1038/373487a0 PMID: 7845459.

33. Schormann N, Grigorian A, Samal A, Krishnan R, DeLucas L, Chattopadhyay D. Crystal structure of vaccinia virus uracil-DNA glycosylase reveals dimeric assembly. BMC Struct Biol. 2007; 7:45. https://doi.org/10.1186/1472-6807-7-45 PMID: 17605817;

34. Banos-Sanz JI, Mojardin L, Sanz-Aparicio J, Lazaro JM, Villar L, Serrano-Heras G, et al. Crystal structure and functional insights into uracil-DNA glycosylase inhibition by phage Phi29 DNA mimic protein p56. Nucleic Acids Res. 2013; 41(13):6761–73. https://doi.org/10.1093/nar/gkt395 PMID: 23671337;

35. Handa P, Roy S, Varshney U. The role of leucine 191 of Escherichia coli uracil DNA glycosylase in the formation of a highly stable complex with the substrate mimic, ugi, and in uracil excision from the synthetic substrates. J Biol Chem. 2001; 276(20):17324–31. https://doi.org/10.1074/jbc.M011166200 PMID: 11278852.

36. Savva R, Pearl LH. Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. Nat Struct Biol. 1995; 2(9):752–7. PMID: 7552746.

37. Putnam CD, Shroyer MJ, Lundquist AJ, Mol CD, Arvai AS, Mosbaugh DW, et al. Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J Mol Biol. 1999; 287(2):331–46. https://doi.org/10.1006/jmbi.1999.2605 PMID: 10080896.

38. Scaramozzino N, Sanz G, Crance JM, Saparbaev M, Drillien R, Laval J, et al. Characterisation of the substrate specificity of homogeneous vaccinia virus uracil-DNA glycosylase. Nucleic Acids Res. 2003; 31(16):4950–7. PMID: 12907738;

39. Otterlei M, Haug T, Nagelhus TA, Slupphaug G, Lindmo T, Krokan HE. Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucleic Acids Res. 1998; 26(20):4611–7. PMID: 9753728;

40. Nagelhus TA, Haug T, Singh KK, Keshav KF, Skorpen F, Otterlei M, et al. A sequence in the N-terminal region of human uracil-DNA glycosylase with homology to XPA interacts with the C-terminal part of the 34-kDa subunit of replication protein A. J Biol Chem. 1997; 272(10):6561–6. PMID: 9045683.

41. Mer G, Bochkarev A, Gupta R, Bochkareva E, Frappier L, Ingles CJ, et al. Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. Cell. 2000; 103 (3):449–56. PMID: 11081631.

42. Hegde ML, Hazra TK, Mitra S. Functions of disordered regions in mammalian early base excision repair proteins. Cell Mol Life Sci. 2010; 67(21):3573–87. https://doi.org/10.1007/s00018-010-0485-5 PMID: 20714778;

43. Hegde ML, Tsutakawa SE, Hegde PM, Holthauzen LM, Li J, Oezguen N, et al. The disordered C-terminal domain of human DNA glycosylase NEIL1 contributes to its stability via intramolecular interactions. J Mol Biol. 2013; 425(13):2359–71. https://doi.org/10.1016/j.jmb.2013.03.030 PMID: 23542007;