LETTERS

# Crystal structure of thymine DNA glycosylase conjugated to SUMO-1

Daichi Baba[1], Nobuo Maita[1,2], Jun-Goo Jee[3]†, Yasuhiro Uchimura[4], Hisato Saitoh[4], Kaoru Sugasawa[5,6], Fumio Hanaoka[5,6,7], Hidehito Tochio[1], Hidekazu Hiroaki[1] & Masahiro Shirakawa[1,3,8,9]

**Members of the small ubiquitin-like modifier (SUMO) family can be covalently attached to the lysine residue of a target protein through an enzymatic pathway similar to that used in ubiquitin conjugation[1], and are involved in various cellular events that do not rely on degradative signalling via the proteasome or lysosome[2–5]. However, little is known about the molecular mechanisms of SUMO-modification-induced protein functional transfer. During DNA mismatch repair, SUMO conjugation of the uracil/thymine DNA glycosylase TDG promotes the release of TDG from the abasic (AP) site created after base excision, and coordinates its transfer to AP endonuclease 1, which catalyses the next step in the repair pathway[6]. Here we report the crystal structure of the central region of human TDG conjugated to SUMO-1 at 2.1 Å resolution. The structure reveals a helix protruding from the protein surface, which presumably interferes with the product DNA and thus promotes the dissociation of TDG from the DNA molecule. This helix is formed by covalent and non-covalent contacts between TDG and SUMO-1. The non-covalent contacts are also essential for release from the product DNA, as verified by mutagenesis.**

TDG initiates base excision repair by releasing thymine or uracil from G·T and G·U mismatches arising from the hydrolytic deamination of methyl-cytosine and cytosine bases that are paired with guanines[6]. Both deamination products can be detrimental to the cell because, unless repaired, they induce a C-to-T transition after DNA replication. After it excises the base from these mismatches, TDG remains stably bound to the resultant AP site, protecting this harmful repair intermediate until it is transferred to AP endonuclease 1 (APE1) to enable the subsequent step in the repair pathway[7,8]. Conjugation of SUMO-1 and SUMO-2/3 to TDG markedly reduces the affinity of TDG for the AP site[9]. Therefore, SUMO conjugation probably constitutes the specific mechanism that releases TDG from the product DNA and coordinates its transfer to APE1.

Human TDG consists of a catalytic core domain (residues 123–300), which shares high sequence similarity with the *Escherichia coli* mismatch-specific uracil DNA glycosylase (MUG). The amino- and carboxy-terminal domains of human TDG are less conserved, with the C-terminal domain containing the SUMO conjugation site Lys 330, the ε-amino group of which can be linked with the C terminus of SUMO-1 via an isopeptide bond. The crystal structure shows that *E. coli* MUG binds stably to DNA by forming hydrogen bonds with the unpaired guanine residue opposite the AP site[10]. Mutational analyses have shown that structural features of the AP site binding are conserved between *E. coli* MUG and mammalian TDG[11].

To elucidate the molecular mechanism of the SUMO modification (SUMOylation)-directed release of TDG from product DNA, we have determined the crystal structure of the central region of human TDG (residues 112–339) conjugated to SUMO-1 (hereafter referred to as SUMO-1–TDG). A map of the electron densities at 2.1 Å resolution was obtained for all amino acids except for five and eighteen residues at the N terminus of TDG and SUMO-1, respectively, and the seven C-terminal residues of TDG (Fig. 1a; see also Supplementary Fig. 1). The structure shows that SUMO-1–TDG is comprised of two domains: a catalytic core domain of TDG comprising residues 117–300, and a SUMO-containing domain, consisting of the structured region of SUMO-1 and the C-terminal region (residues 307–330) of TDG (Fig. 1b). These domains are connected by a short crossover loop comprising residues 301–306 of TDG.

The catalytic core domain, but not the C-terminal segment, of TDG does not seem to undergo substantial structural rearrangements upon conjugation. The structure of the TDG core domain closely resembles those of mismatch DNA glycosylases, such as MUG from *E. coli* and mammalian uracil DNA glycosylases (UDGs). The root-mean-square deviation (r.m.s.d.) of 122 Cα atoms between the TDG core domain and *E. coli* MUG (Protein Data Bank (PDB) code 1MWI) is 1.4 Å. The alignment of these structures superimposes the pyrimidine-binding pocket of MUG onto a pocket of TDG, suggesting that the pocket of the catalytic core of TDG is also involved in the nucleotide flipping mechanism commonly used in UDG enzymes[10,12–14] (Fig. 1b). Residues that are important for DNA binding and the catalytic activity of MUG are also effectively superimposed onto the TDG core domain, indicating high structural conservation throughout evolution[11]. These observations also suggest that no major structural rearrangement of the TDG core domain occurs upon SUMO-1 conjugation. Furthermore, SUMO-1 does not undergo substantial conformational changes after conjugation, because the overall fold of the structured region of SUMO-1 (residues 19-97) in the SUMO-1–TDG complex is nearly identical to that in unconjugated SUMO-1 (PDB code 1A5R), as shown by a r.m.s.d. of the Cα atoms of 2.2 Å. The 18 N-terminal amino acids for which electron densities are not observed in the SUMO-1–TDG crystal have been shown to be flexible in unconjugated SUMO-1 in solution[15].

The major molecular interface between TDG and SUMO-1 is formed by the C-terminal segment (residues 307–330) of TDG. The elongated structure of this segment wraps around SUMO-1, forming the globular SUMO-containing domain (Fig. 1a, b). The segment

[1]Graduate School of Integrated Science, Yokohama City University, Yokohama 230-0045, Japan. [2]Japan Biological Informatics Consortium, Tokyo 104-0032, Japan. [3]RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan. [4]Department of Regeneration Medicine, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto 860-0811, Japan. [5]Cellular Physiology Laboratory, Discovery Research Institute, RIKEN, Wako 351-0198, Japan. [6]SORST, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [7]Graduate School of Frontier Biosciences, Osaka University, Suita 565-0871, Japan. [8]Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan. [9]CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan. †Present address: Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

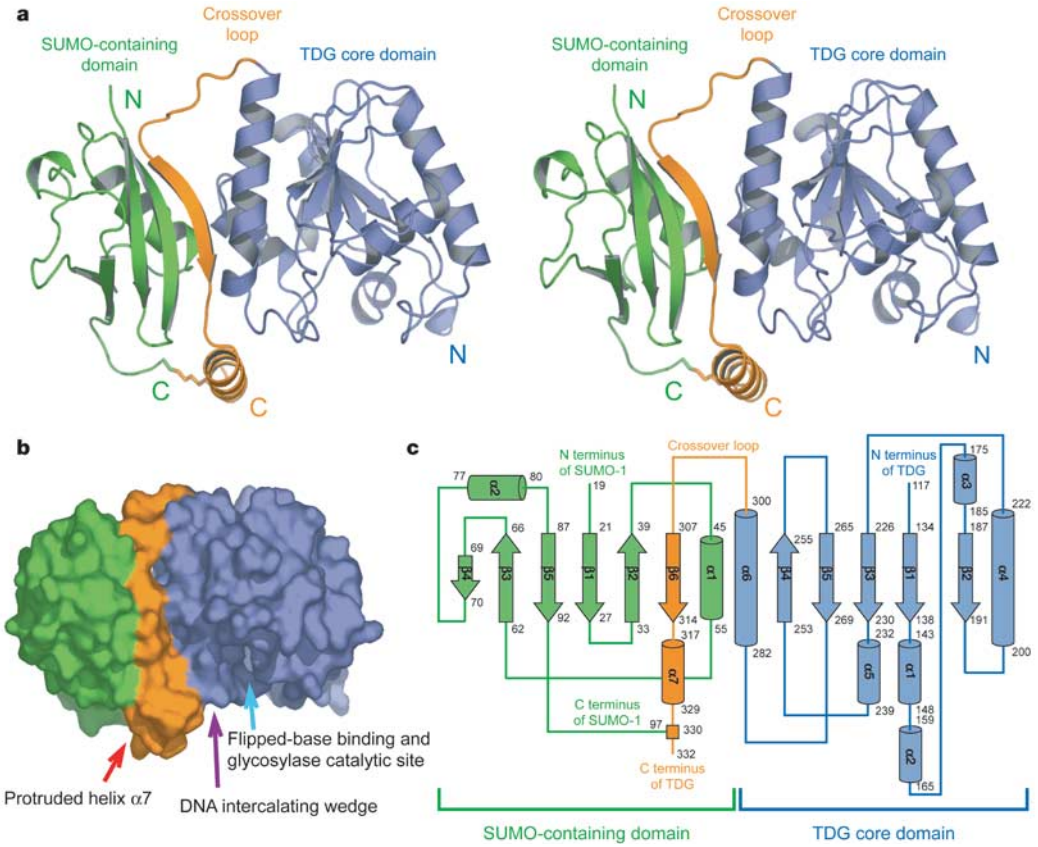


**Figure 1** | **Structure of SUMO-1–TDG.** **a**, Stereo ribbon diagram of the structure of SUMO-1–TDG. The catalytic core domain and the C-terminal segment of TDG are shown in purple and orange, respectively; SUMO-1 is shown in green. **b**, Surface of SUMO-1–TDG. The colouring and viewing angle are the same as in **a**. The positions of the protruded helix α7, putative base-binding pocket and DNA intercalating wedge are also shown. **c**, Topology of the secondary structure elements. Residue numbers and the N and C termini of TDG and SUMO-1 are indicated. The colouring is the same as in **a** and **b**.

makes both covalent and non-covalent contacts with SUMO-1. The covalent contact occurs at the C terminus of this segment through the isopeptide bond, and the non-covalent contacts occur at strand β6 (residues 307–314) at the N terminus of the segment (Fig. 1c). This strand forms an intermolecular antiparallel β-sheet with strand β2, an edge strand of the β-sheet of SUMO-1, resulting in a continuous six-stranded, mixed β-sheet. In addition to their main chain contacts, the side chains of residues from TDG β6 make extensive polar and hydrophobic contacts with those from β2 and β1 of SUMO-1 (Fig. 2a, b). Glu 310 of TDG makes bidentate hydrogen bonds to Arg 54 in SUMO-1, whereas the side chain of Val 308 packs into a pocket formed by the side chains of Phe 36, Val 38 and Leu 47 of SUMO-1. The intermolecular contacts also involve a hydrogen bond network formed between Arg 281, Asp 284, Tyr 313 and Asp 323 of TDG and Asp 30 of SUMO-1. The interface between SUMO-1 and TDG is largely confined to the edge of the β-sheet of SUMO-1.

The most prominent feature of SUMO-1–TDG is helix α7 of TDG, which contains the conjugation site. Buttressed by this conjugation

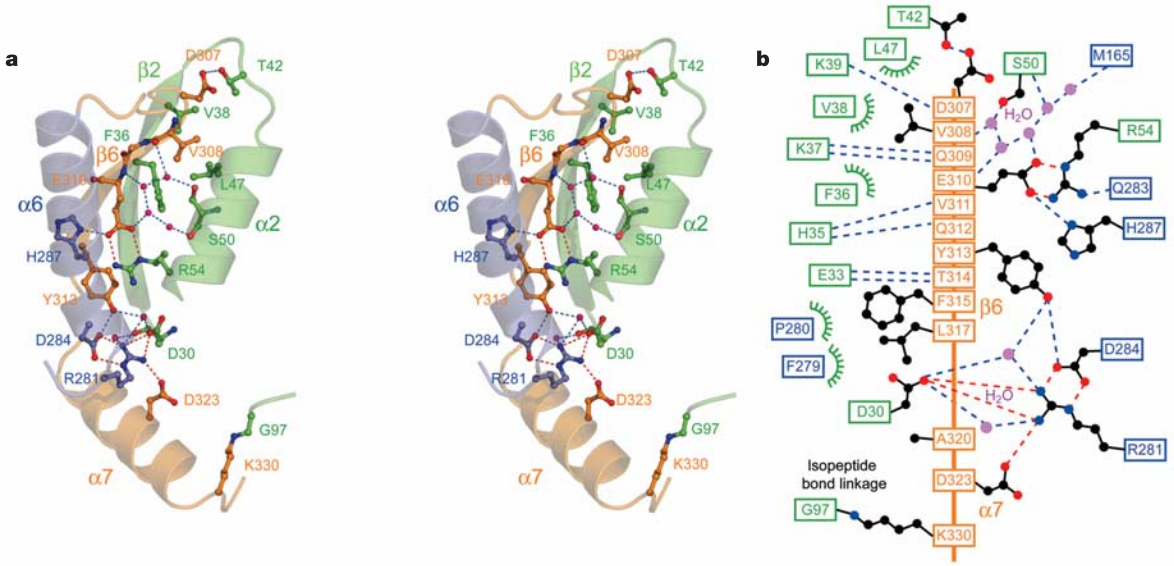


**Figure 2** | **Molecular interface between TDG and SUMO-1.** **a**, Detailed stereo view of contacts between TDG and SUMO-1. Residues from the C-terminal segment and core of TDG are shown in orange and purple, respectively; residues from SUMO-1 are shown in green. **b**, Schematic summary of contacts in SUMO-1–TDG. Hydrophobic, hydrogen bond and electrostatic interactions are shown in green, blue and red, respectively. Protein residues are coloured as in **a**.

site at its C terminus and the intermolecular β-sheet pairing of strand β6 flanking its N terminus, this helix is held up above SUMO-1 and forms a large protrusion on the protein surface (Fig. 1b). In particular, the C-terminal end of the helix makes almost no non-covalent contacts with other parts of the proteins, and thus is raised only by the isopeptide bond between Lys 330 and SUMO-1.

The strong conservation of tertiary structures and residues that are important for DNA binding and catalysis between the catalytic core domain of TDG and *E. coli* MUG suggests that these enzymes share a similar mode of DNA interaction and catalytic mechanism. We therefore constructed a model of the complex formed between TDG and product DNA by best-fit superposition of the coordinates of the TDG core domain to those of MUG bound to a product DNA containing an AP site[10] (see Methods). Notably, α7, the protruded helix at the C-terminal segment of TDG, is positioned such that it would encounter severe steric clash with the sugar-phosphate backbone of the bound DNA at positions +1 and +2 from the unpaired guanine in the TDG–DNA model (Fig. 3). Furthermore, it seems unlikely that the bound DNA could undergo a bend or deformation to avoid this apparent steric clash, because the model suggests that loops between strand β5 and helix α6, and between helices α1 and α2, make contact with the sugar-phosphate backbone of the AP-site-containing strand at positions 0 and −1, in very close proximity to the putative clash site. Therefore, it seems most likely that this steric clash between the protruded helix and the DNA backbone induces the dissociation of SUMO-1–TDG from the AP site on DNA. It is noteworthy that SUMO-1 is positioned so that it does not encounter steric interference from any part of the DNA in the model.

Structural features of the complex suggest that maintenance of α7, the protruded helix, requires both SUMO-1 conjugation and β-sheet pairing between β6 of TDG and β2 of SUMO-1 (Fig. 2). Except for these interactions, the helix makes only a few contacts with other parts of the proteins. Therefore, the residues that form this helix are probably unstructured when SUMO-1 is not conjugated. Consistent with this, unconjugated TDG is more sensitive to the proteases trypsin and thermolysin than is its SUMO-1-conjugated form (Supplementary Fig. 2). Therefore, the formation of this protruded helix is most probably the consequence of a structural rearrangement of the C-terminal segment of TDG, induced by SUMO-1 conjugation.

We carried out mutational analyses that showed that the non-covalent interactions between the C-terminal segment of TDG and SUMO-1 are indispensable for releasing SUMO-1–TDG from the product DNA. Glutamine substitution for Glu 310 of TDG, which makes bidentate hydrogen bonds with Arg 54 of SUMO-1, largely restored the ability of SUMO-1–TDG to bind to the product DNA (Fig. 4a). Similarly, substitution of Arg 281 of TDG—which is involved in the hydrogen bond network formed between TDG and SUMO-1—by alanine recovered the DNA binding of SUMO-1–TDG. Notably, these residues are essential for the previously observed[9], non-covalent binding of TDG to SUMO-1, as shown by glutathione *S*-transferase (GST) pull-down assays (Fig. 4b). Phe 315 is another residue for which alanine substitution resulted in a loss of SUMO-1 binding and recovery of the DNA binding of SUMO-1–TDG (Fig. 4a, b). Therefore, both the non-covalent and covalent interactions that fix the C-terminal segment of TDG seem to be essential for the release of DNA from SUMO-1–TDG.

Notably, strand β6, the major intermolecular contact region of TDG, contains a sequence (VQEV, residues 308–311; Fig. 2b) that is not identical, but is similar, to the recently proposed SUMO-binding motif (SBM)[16]. The SBM has the consensus sequence V/I-X-V/I-V/I and is found in proteins that bind to SUMO family proteins. Mutagenesis has confirmed that residues in the SBM-like sequence are important for non-covalent SUMO-1 binding (Fig. 4b). These observations suggest that the SBM found in other proteins may bind

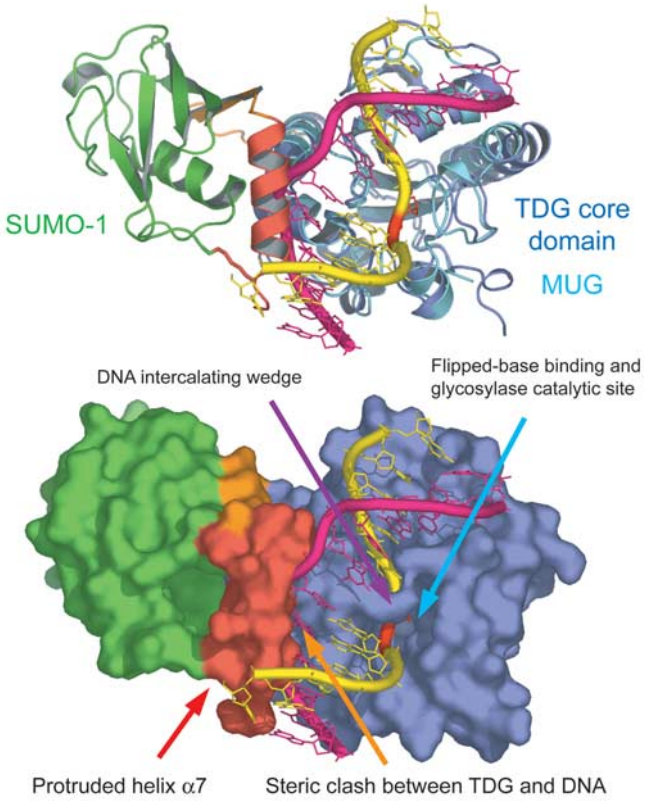


**Figure 3 | Model of a complex between SUMO-1–TDG and a DNA molecule containing an AP site.** The model was constructed by superimposing the structure of SUMO-1–TDG on that of MUG bound to a product DNA containing an AP site[10] (PDB code 1MWI). A canonical B-form conformation outside the AP sites has been assumed. The locations of the putative base-binding pocket and DNA intercalating wedge are indicated. The 'protruded helix' α7 is shown in red, and the site of putative steric clash between the helix and DNA backbone is also indicated.

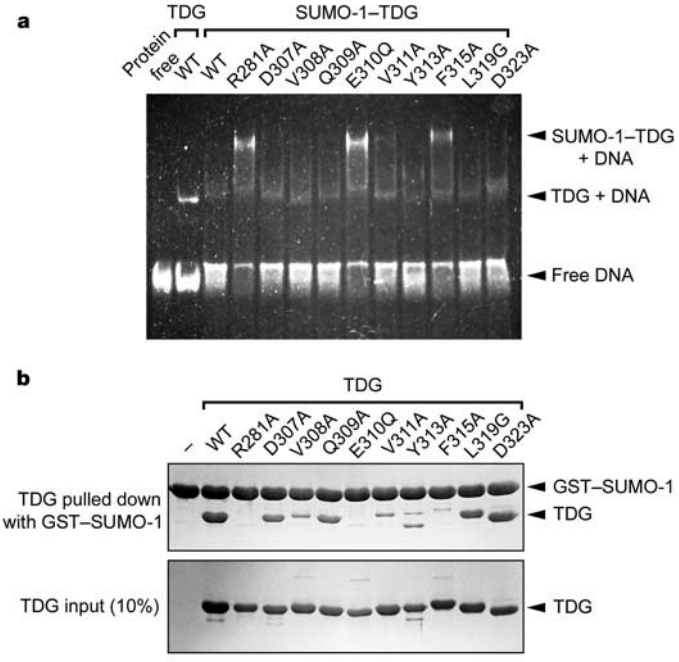


**Figure 4 | DNA and SUMO-1 binding activity of TDG. a**, Electrophoretic mobility shift assay examining the DNA-binding capacity of different TDG mutants conjugated to SUMO-1. **b**, GST pull-down assay examining the SUMO-1 binding capacity of different mutants of unconjugated TDG. The concentration of TDG and its mutants as well as SUMO-1 was 10 μM.

to SUMO family proteins through the formation of an intermolecular β-sheet, similar to that seen in SUMO-1–TDG. Consistent with this, it has been shown that residues from strand β2 of SUMO-1 display marked perturbations in chemical shifts upon binding to an SBM-containing peptide[16]. Of note, many SUMOylated proteins contain an SBM consensus sequence[16]. This raises the possibility that protein segments between the SBM and the SUMOylation site of these proteins may undergo a conformational rearrangement similar to that seen in SUMO-1–TDG. In conclusion, the structure of SUMO-1–TDG suggests that SUMO-1 conjugation induces the formation of the protruded helix in TDG, which allows its dissociation from the product AP site. It cannot be ruled out that formation of the helix may also induce larger conformational changes in the enzyme that facilitate its functional transfer.

## METHODS

**Crystallization and structure determination.** Proteins are expressed and purified as described[17] (see also Supplementary Methods). Crystals of SUMO-1–TDG were grown in 25% PEG3350, 0.2 M $MgCl_2$ and 0.1 M Tris-HCl (pH 8.5) at 20 °C by using a micro-seeding technique. The crystals belong to space group $P2_12_12_1$ with unit-cell dimensions of $a = 42.2$ Å, $b = 70.4$ Å and $c = 106.4$ Å. X-ray diffraction data were collected at beamline BL6A of the Photon Factory (Tsukuba, Japan) with a Quantum R4 CCD detector (ADSC) at wavelength 1.0 Å. The diffraction data were processed with MOSFLM[18] and scaled with SCALA[19]. The 2.1 Å resolution structure of SUMO-1–TDG was solved by molecular replacement with MolRep[20] using *E. coli* MUG[10] (PDB code 1MUG) and yeast Smt3 (ref. 21) (PDB code 1EUV) as search models. The model building was done with program O[22] and refined with CNS[23]. The Ramachandran plot of the final model shows that 93.8% of the residues are in the most favoured regions, and 6.2% in additionally favoured regions. The current model gives $R_{work}$ and $R_{free}$ values of 20.5% and 24.5%, respectively. The data collection and refinement statistics are summarized in Supplementary Table 1. Figures 1a, b, 2a and 3 were made with the program PyMOL[24].

**Model building of a SUMO-1–TDG–DNA complex.** The model of a complex between SUMO-1–TDG and a DNA molecule containing an AP site was constructed by superimposing the structure of SUMO-1–TDG on that of MUG bound to a product DNA containing an AP site[10] (PDB code 1MWI).

We assumed that the DNA adopts a canonical B-form conformation, because DNA molecules in MUG complexes have been supposed to maintain B-form conformations outside the AP sites[10,25]. We found that most of the protein residues that make DNA contacts and form the pyrimidine-binding pocket are located at similar positions relative to the DNA in the MUG–DNA complex and in the model, supporting the validity of our model.

**Biochemical assays.** A fluorescein isothiocyanate (FITC)-labelled 35-mer double-stranded oligonucleotide substrate containing a U·G mismatch (5′-GGCAATCAGTTCACTTCGAGCCCAGGTATTTAGCC-3′; 5′-FITC-GGCTAAATACCTGGGCTUGAAGTGAACTGATTGCC-3′) was chemically synthesized and annealed in a buffer containing 10 mM Tris-HCl, 5 mM $MgCl_2$ and 0.1 mM EDTA (pH 7.5). An oligonucleotide containing an AP site was generated by incubating 10 pmol of the duplex containing the U·G base pair with 1 unit of uracil DNA glycosylase (New England Biolabs) in a buffer containing 20 mM Tris-HCl, 1 mM EDTA and 1 mM dithiothreitol (DTT) (pH 8.0) for 2 h at 37 °C. The accuracy and completion of AP site formation were analysed by NaOH treatment and denaturing gel electrophoresis. For the electrophoretic mobility shift assay, 4 pmol of TDG proteins were incubated in 10 μl of reaction mixture (25 mM HEPES-NaOH, 150 mM NaCl, 5 mM $MgCl_2$, 0.01% Triton X-100, 1 mM EDTA, 5% glycerol, 1 mM DTT and 0.1 mg $ml^{-1}$ BSA; pH 7.5) with 1 pmol of the oligonucleotide duplex containing an AP site and 5 pmol of the non-labelled homoduplex competitor. After incubation for 30 min at 30 °C, the samples were immediately loaded onto 8% native polyacrylamide gels in 0.5 × TBE at 100 V at room temperature. The fluorescent probes were visualized by using a luminescent image analyser LAS-1000 Plus (Fuji Photo Film) in the fluorescence mode.

The non-covalent interactions between SUMO-1 and TDG were analysed by GST pull-down experiments. Purified GST–SUMO-1(1–97) was immobilized onto glutathione-sepharose beads. Various TDG mutants were added to the beads and incubated at 4 °C for 1 h in a buffer containing 25 mM HEPES, 150 mM NaCl, 1 mM DTT and 0.05% NP-40 (pH 7.5). The beads were then washed three times with the same buffer and analysed by SDS–polyacrylamide gel electrophoresis with Coomassie brilliant blue stain. The standard condition used 50 μl of the beads and 100 μl of the GST fusion protein and various TDG mutants (10 μM).

**Received 6 December 2004; accepted 14 April 2005.**


1. Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–382 (2004).
2. Gill, G. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev.* **18,** 2046–2059 (2004).
3. Seeler, J. S. & Dejean, A. Nuclear and unclear functions of SUMO. *Nature Rev. Mol. Cell Biol.* **4,** 690–699 (2003).
4. Kim, K. I., Baek, S. H. & Chung, C. H. Versatile protein tag, SUMO: its enzymology and biological function. *J. Cell. Physiol.* **191,** 257–268 (2002).
5. Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88,** 97–107 (1997).
6. Hardeland, U. *et al.* Thymine DNA glycosylase. *Prog. Nucleic Acid Res. Mol. Biol.* **68,** 235–253 (2001).
7. Waters, T. R. & Swann, P. F. Kinetics of the action of thymine DNA glycosylase. *J. Biol. Chem.* **273,** 20007–20014 (1998).
8. Waters, T. R., Gallinari, P., Jiricny, J. & Swann, P. F. Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **274,** 67–74 (1999).
9. Hardeland, U., Steinacher, R., Jiricny, J. & Schar, P. Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J.* **21,** 1456–1464 (2002).
10. Barrett, T. E. *et al.* Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell* **92,** 117–129 (1998).
11. Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **275,** 33449–33456 (2000).
12. Roberts, R. J. On base flipping. *Cell* **82,** 9–12 (1995).
13. Mol, C. D. *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* **82,** 701–708 (1995).
14. Mol, C. D. *et al.* Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80,** 869–878 (1995).
15. Bayer, P. *et al.* Structure determination of the small ubiquitin-related modifier SUMO-1. *J. Mol. Biol.* **280,** 275–286 (1998).
16. Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G. & Chen, Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc. Natl Acad. Sci. USA* **101,** 14373–14378 (2004).
17. Uchimura, Y., Nakamura, M., Sugasawa, K., Nakao, M. & Saitoh, H. Overproduction of eukaryotic SUMO-1- and SUMO-2-conjugated proteins in *Escherichia coli*. *Anal. Biochem.* **331,** 204–206 (2004).
18. Leslie, A. G. Integration of macromolecular diffraction data. *Acta Crystallogr. D* **55,** 1696–1702 (1999).
19. Collaborative Computational Project No. 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
20. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).
21. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5,** 865–876 (2000).
22. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
23. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
24. DeLano, W. L. *PyMOL* ⟨http://www.pymol.org⟩ (2002).
25. Barrett, T. E. *et al.* Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J.* **18,** 6599–6609 (1999).


**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** This work was supported by grants to M.S. from the Japanese Ministry of Education, Science, Sports and Culture, and Japan Science and Technology Agency.

**Author Information** Atomic coordinates of SUMO-1–TDG have been deposited in the Protein Data Bank under the accession number 1WYW. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S. (shirakawa@moleng.kyoto-u.ac.jp).